H. Sugiyarto, S.Si., M.Si., Ph.D.

PENGANTAR STATISTIKA MATEMATIKA 2

PENGANTAR STATISTIKA MATEMATIKA 2

Penyusun : H. Sugiyarto, M.Si., Ph.D
Layout dan Desain Cover : Rizal Redita Putra
Galang Suryaputra
M. Yahya Firza Afitian
Imas Abu Yazid

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI TERAPAN
UNIVERSITAS AHMAD DAHLAN

Kata Pengantar

Bismillahirrahmanirrahiim
Puji syukur Alhamdulillah penyusun panjatkan kepada Allah SWT atas berkat rahmat dan hidayah-Nya sehingga penyusun dapat menyelesaikan buku dengan judul "Pengantar Statistika Matematika 2". Buku ini digunakan sebagai salah satu pegangan untuk mata kuliah Statistika Matematika 2.

Buku ini terdiri dari 7 bab yaitu :

1. Ekspektasi Bersyarat Dari Variabel Acak Bivariat
2. Transformasi Variabel Acak dan Distribusinya
3. Distribusi Diskrit Bivariat
4. Distribusi Kontinu Bivariat
5. Beberapa Teknik Dalam Menemukan Estimasi Titik Dari Parameter
6. Kriteria Untuk Mengevaluasi Kualitas Dari Estimator
7. Beberapa Teknik Dalam Mencari Estimator Interval Parameter

Penyusun menyadari bahwa buku ini masih banyak kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi kesempurnaan buku ini. Akhir kata, penyusun berharap semoga buku ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan khususnya pada bidang Matematika

Yogyakarta, 2021

Penyusun

Daftar ISI

BAB 7

BEBERAPA TEKNIK DALAM MENCARI ESTIMATOR INTERVAL PARAMETER...............318

BAB I

EKSPEKTASI BERSYARAT DARI VARIABEL ACAK BIVARIAT

Bab ini akan membahas mean bersyarat dan varians bersyarat terkait dengan dua variabel acak. Mean bersyarat sangat berguna dalam estimasi parameter Bayesian dengan fungsi kuadrat.

1.1. Nilai Harapan Bersyarat

Misalkan X dan Y adalah dua variabel acak dengan kepadatan gabungan $f(x,y)$. Ingatlah bahwa kepadatan probabilitas bersyarat dari X, dengan syarat kejadian $Y=y$, didefinisikan sebagai berikut

$$g(x\,|\,y)=\frac{f(x,y)}{f_2(y)}, \qquad f_2(y)>0$$

di mana $f_2(y)$ adalah kepadatan probabilitas marginal Y. Demikian pula, kepadatan probabilitas bersyarat dari Y, dengan syarat kejadian $X=x$, didefinisikan sebagai berikut

$$h(y\,|\,x)=\frac{f(x,y)}{f_1(x)}, \qquad f_1(x)>0$$

di mana $f_1(x)$ adalah kepadatan probabilitas marginal dari X.

Definisi 1.1.

Mean bersyarat dari X dengan syarat $Y=y$ didefinisikan sebagai

$$\mu_{X\,|\,y}=E(X\,|\,y)$$

Dimana

$$E(X\,|\,y)=\begin{cases}\displaystyle\sum_{x\in R_X} x\,g(x\,|\,y) & \textit{jika X adalah Diskrit}\\[2ex] \displaystyle\int_{-\infty}^{\infty} x\,g(x\,|\,y)\,dx & \textit{jika X adalah Kontinu}\end{cases}$$

Demikian pula, mean bersyarat dari Y. dengan syarat $X=x$ didefinisikan sebagai

$$\mu_{Y|x} = E(Y|x)$$

Dimana

$$E(Y|x) = \begin{cases} \sum_{x \in R_Y} y\, h(y \mid x) & \textit{jika Y adalah Diskrit} \\ \\ \int_{-\infty}^{\infty} y\, h(y \mid x)\, dy & \textit{jika Y adalah kontinu} \end{cases}$$

■ **Contoh 1.1.** Jika X dan Y adalah variabel acak diskrit dengan fungsi kepadatan probabilitas gabungan

$$f(x, y) = \begin{cases} \frac{1}{21}(x+y) & \textit{untuk } x = 1,2,3; \quad y = 1,2 \\ \\ 0 & \textit{untuk yang lainnya} \end{cases}$$

Berapa mean bersyarat dari X dengan syarat $Y = y$, yakni $E(X|y)$?

Jawab :

Untuk menghitung mean bersyarat dari X dengan syarat $Y = y$, kita memerlukan kepadatan bersyarat $g(x \mid y)$ dari X dengan syarat $Y = y$. Namun, untuk mencari $g(x \mid y)$, kita perlu mengetahui marginal dari Y, yaitu $f_2(y)$. Jadi, kita mulai dengan

$$f_2(y) = \sum_{x=1}^{3} \frac{1}{21}(x+y)$$

$$= \frac{1}{21}(6+3y)$$

Oleh karena itu, kepadatan bersyarat dari X dengan syarat $Y = y$ diberikan oleh

$$g(x \mid y) = \frac{f(x, y)}{f_2(y)}$$

$$= \frac{x+y}{6+3y}, \qquad x = 1,2,3.$$

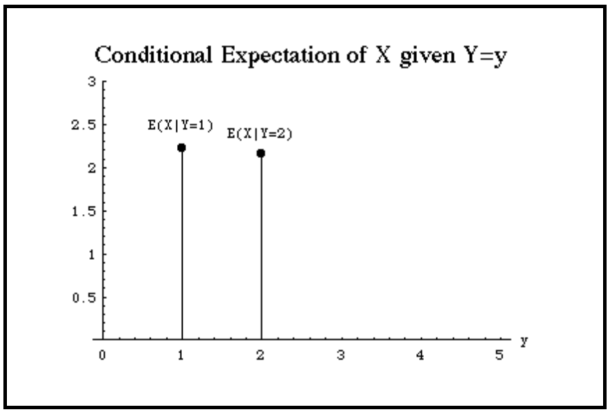
Nilai harapan (ekspektasi) bersyarat dari X dengan syarat kejadian $Y = y$

$$E(X|y) = \sum_{x \in R_X} x\, g(x\,|y)$$

$$= \sum_{x=1}^{3} x \frac{x+y}{6+3y}$$

$$= \frac{1}{6+3y}\left[\sum_{x=1}^{3} x^2 + y\sum_{x=1}^{3} x\right]$$

$$= \frac{14+6y}{6+3y}, \qquad y = 1,2.$$



Catatan 1.1. Perhatikan bahwa mean bersyarat dari X dengan syarat $Y = y$ adalah bergantung hanya pada y, yakni $E(X|y)$ adalah fungsi ϕ dari y. Dalam contoh di atas, fungsi ϕ ini merupakan fungsi rasional,

yaitu $\phi(y) = \dfrac{14+6y}{6+3y}$.

Contoh 1.2. Misalkan X dan Y memiliki fungsi densitas gabungan

$$f(x,y) = \begin{cases} x+y & \textit{untuk}\ 0 < x,\ \ y < 1 \\ 0 & \textit{untuk yang lainnya} \end{cases}$$

Berapa mean bersyarat dari $E\left(Y \mid X = \frac{1}{3}\right)$?

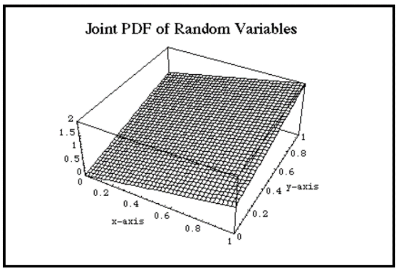
Jawab :

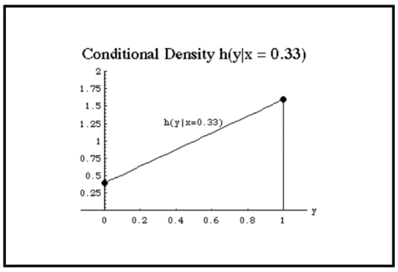
$$f_1(x) = \int_0^1 (x+y)\,dy$$

$$= \left[xy + \frac{1}{2}y^2\right]_0^1$$

$$= x + \frac{1}{2}$$

$$h(y|x) = \frac{f(x,y)}{f_1(x)} = \frac{x+y}{x+\frac{1}{2}}$$





$$E\left(Y \mid X = \frac{1}{3}\right) = \int_0^1 y\, h(y|x)\,dy$$

$$= \int_0^1 y\,\frac{x+y}{x+\frac{1}{2}}\,dy$$

$$= \int_0^1 y \frac{\frac{1}{3} + y}{\frac{5}{6}} dy$$

$$= \frac{6}{5} \int_0^1 \left(\frac{1}{3} y + y^2 \right) dy$$

$$= \frac{6}{5} \left[\frac{1}{6} y^2 + \frac{1}{3} y^3 \right]_0^1$$

$$= \frac{6}{5} \left[\frac{1}{6} + \frac{1}{2} \right]$$

$$= \frac{6}{5} \left(\frac{3}{6} \right)$$

$$= \frac{3}{5}$$

Mean dari variabel acak Y adalah bilangan deterministik. Mean bersyarat dari Y dengan syarat $X = x$, yaitu $E(Y|x)$ adalah fungsi $\phi(x)$ dari variabel x. Dengan menggunakan fungsi tersebut dapat membentuk $\phi(X)$ Fungsi $\phi(X)$ ini adalah variabel acak. Jadi mulai dari fungsi deterministik $E(Y|x)$, kita telah membentuk variabel acak $E(Y|X) = \phi(X)$. Sifat-sifat penting terkait dengan ekspektasi bersyarat diberikan dengan mengikuti teorema berikut ini

Teorema 1.1.

Nilai harapan dari variabel acak $E(Y|X)$ adalah sama dengan nilai harapan dari Y, yaitu

$$E_x\left(E_{y|x}(Y|X)\right) = E_y(Y)$$

dimana $E_x(X)$ adalah ekspektasi dari X dengan distribusi dari X dan $E_{y|x}(Y|X)$ adalah nilai harapan (ekspektasi) dari Y dengan kepadatan bersyarat $h(y|X)$.

Bukti :

Bukti teorema ini untuk variable kontinu

$$E_x\left(E_{y|x}\left(Y|X\right)\right)=E_x\left[\int_{-\infty}^{\infty} y\, h\left(y\mid X\right)dy\right]$$

$$=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} y\, h(y|x)dy\right) f_1\left(x\right)dx$$

$$=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} y\, h(y|x)\, f_1\left(x\right)dy\right)dx$$

$$=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} h(y|x)\, f_1\left(x\right)dx\right)ydy$$

$$=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f\left(x,y\right)dx\right)ydy$$

$$=\int_{-\infty}^{\infty} y\, f_2\left(y\right)dy$$

$$=E_y\left(Y\right)$$

Contoh 1.3. Seekor serangga meletakan sejumlah Y telur di mana Y merupakan distribusi Poisson dengan parameter λ. Jika probabilitas setiap telur bertahan adalah p, lalu berapa rata-rata telur yang akan bertahan?

Jawab :

Misalkan X menunjukkan jumlah telur yang bertahan hidup. Kemudian diberikan bahwa $Y=y$ (dengan ketentuan bahwa serangga telah meletakan y telur) variabel acak X memiliki distribusi binomial dengan parameter y dan p. Jadi

$$X\mid Y \sim BIN\left(Y,p\right)$$

$$Y \sim POI\ \left(\lambda\right)$$

Oleh karena itu, ekspektasi jumlah telur yang bertahan hidup diberikan oleh

$$E_x(X) = E_y(E_{x|y}(X|Y)$$

$$= E_y(pY) \qquad \text{(karena } X|Y \sim BIN(Y,p))$$

$$= pE_y(Y)$$

$$= p\lambda \qquad \text{(karena } Y \sim POI(\lambda)$$

Definisi 1.2.

Variabel acak X dikatakan memiliki distribusi campuran jika distribusi X bergantung pada kuantitas yang juga memiliki distribusi.

Contoh 1.4.

Koin yang adil dilemparkan. Jika gambar pada koin muncul, 1 dadu digulirkan. Jika angka pada koin muncul, 2 dadu digulirkan. Misalkan Y menjadi total pada dadu. Berapakah nilai harapan (ekspektasi) dari Y ?

Jawab :

Misalkan X menunjukkan hasil dari melempar koin. Kemudian $X \sim BER(p)$, dimana probabilitas keberhasilannya adalah $p = \frac{1}{2}$

$$E_y(Y) = E_x\left(E_{y|x}(Y|X)\right)$$

$$= \frac{1}{2}E_{y|x}(Y|X=0) + \frac{1}{2}E_{y|x}(Y|X=1)$$

$$= \frac{1}{2}\left(\frac{1+2+3+4+5+6}{6}\right) + \frac{1}{2}\left(\frac{2+6+12+20+30+42+40+36+30+22+12}{36}\right)$$

$$= \frac{1}{2}\left(\frac{126}{36} + \frac{252}{36}\right)$$

$$= \frac{378}{72}$$

$$= 5,25$$

Perhatikan bahwa ekspektasi jumlah titik yang muncul ketika 1 dadu digulirkan adalah $\frac{126}{36}$, dan jumlah titik yang diharapkan muncul ketika 2 dadu digulirkan adalah $\frac{252}{36}$

Teorema 1.2.

Jika X dan Y adalah dua variabel acak dengan mean μ_X dan μ_Y, dan standar deviasi σ_X dan σ_Y. Jika ekspektasi bersyarat Y dengan syarat $X = x$ linier dalam x, maka

$$E\left(Y|X = x\right) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}\left(x - \mu_X\right)$$

di mana ρ menunjukkan koefisien korelasi X dan Y.

Bukti :

Asumsikan bahwa variabel acak X dan Y adalah kontinu. Jika diskrit, bukti teorema mengikuti dengan cara yang sama persis mengganti Integral $\left(\int\right)$ dengan Sigma $\left(\sum\right)$. Diketahui bahwa $E(Y \mid X = x)$ adalah linier dalam x, yaitu

$$E\left(Y|X = x\right) = ax + b \qquad (1.0)$$

dimana a dan b adalah dua konstanta. Maka dari itu, dari atas kita dapatkan

$$\int_{-\infty}^{\infty} y\,h\left(y \mid x\right)dy = ax + b$$

yang berarti

$$\int_{-\infty}^{\infty} y\frac{f\left(x,y\right)}{f_1\left(x\right)}dy = ax + b$$

Mengalikan kedua sisi dengan $f_1\left(x\right)$, kita dapatkan

$$\int_{-\infty}^{\infty} y\,f\left(x,y\right)dy = \left(ax + b\right)f_1\left(x\right) \qquad (1.1)$$

Sekarang mengintegrasikan terhadap x, kita dapatkan

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f\left(x, y\right) dy\, dx = \int_{-\infty}^{\infty} \left(ax + b\right) f_1\left(x\right) dx$$

Hasilnya

$$\mu_X = a\mu_X + b \qquad (1.2)$$

Sekarang, kita mengalikan persamaan (1.1) dengan x dan kemudian mengintegrasikan pernyataan yang dihasilkan x untuk mendapatkan

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f\left(x, y\right) dy\, dx = \int_{-\infty}^{\infty} \left(ax^2 + bx\right) f_1\left(x\right) dx$$

Dari sini kita dapatkan

$$E\left(XY\right) = aE\left(X^2\right) + b\mu_X \qquad (1.3)$$

Menyelesaikan persamaan (1.2) dan persamaan (1.3) untuk a dan b yang tidak diketahui, kita dapatkan

$$a = \frac{E\left(XY\right) - \mu_X\, \mu_Y}{\sigma_X^2}$$

$$= \frac{\sigma_{XY}}{\sigma_X^2}$$

$$= \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \frac{\sigma_Y}{\sigma_X}$$

$$= \rho \frac{\sigma_Y}{\sigma_X}$$

Begitu pula yang kita dapatkan

$$b = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}\ \mu_X$$

Misalkan a dan b menjadi (1.0) kita mendapatkan hasil yang tegas dan bukti dari teorema sekarang telah selesai.

Contoh 1.5. Misalkan X dan Y adalah variabel acak dengan $E(Y \mid X = x) = -x + 3$ dan $E(X \mid Y = y) = -\frac{1}{y}y + 5$. Berapakah koefisien korelasi X dan Y?

Jawab :

Dari Teorema 1.2, kita dapatkan

$$\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X) = -x + 3$$

Oleh karena itu, dengan menyamakan koefisien dari suku-suku x, kita dapatkan

$$\rho \frac{\sigma_Y}{\sigma_X} = 1 \qquad (1.4)$$

Begitu pula saat

$$\mu_X + \rho \frac{\sigma_Y}{\sigma_X}(y - \mu_Y) = -\frac{1}{4}y + 5$$

kita punya

$$\rho \frac{\sigma_Y}{\sigma_X} = -\frac{1}{4} \qquad (1.5)$$

Mengalikan (1.4) dengan (1.5), kita dapatkan

$$\rho \frac{\sigma_Y}{\sigma_X} \rho \frac{\sigma_Y}{\sigma_X} = (-1)\left(-\frac{1}{4}\right)$$

yang mana

$$\rho^2 = \frac{1}{4}$$

$$\rho = \pm \frac{1}{2}$$

Saat $\rho \frac{\sigma_Y}{\sigma_X} = -1$ dan $\rho \frac{\sigma_Y}{\sigma_X} > 0$, kita dapatkan

$$\rho = -\frac{1}{2}$$

1.2. Varians Bersyarat

Varians dari fungsi kepadatan probabilitas $f(y\,|x)$ disebut varians bersyarat dari Y yang diberikan $X = x$. Variansi bersyarat ini didefinisikan sebagai berikut :

Definisi 1.3.

Misalkan X dan Y adalah dua variabel acak dengan kepadatan gabungan $f(x,y)$ dan $f(y\,|x)$ adalah kepadatan bersyarat dari Y yang diberikan $X = x$. Varians bersyarat dari Y yang diberikan $X = x$, dilambangkan dengan $Var\ (Y\,|x)$, didefinisikan sebagai

$$Var(Y|x) = E(Y^2\,|x) - (E(Y\,|x))^2$$

Dimana $(Y\,|x)$ menunjukkan mean bersyarat dari Y yang diberikan $X = x$.

Contoh 1.6. Misalkan X dan Y adalah variabel acak kontinu dengan fungsi kepadatan probabilitas gabungan

$$f(x) = \begin{cases} e^{-y}, & untuk\, 0 < x < y < \infty \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Berapa varians bersyarat dari Y dengan syarat $X = x$?

Jawab :

Densitas marginal $f_1(x)$ diberikan oleh

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \int_{-\infty}^{\infty} e^{-y}\,dy$$

$$= \left[-e^{-y}\right]_x^{\infty}$$

$$= e^{-x}$$

Jadi, kepadatan bersyarat dari Y yang diberikan $X = x$ adalah

$$h(y \mid x) = \frac{f(x, y)}{f_1(x)}$$

$$= \frac{e^{-y}}{e^{-x}}$$

$$= e^{-(y-x)} \qquad \textit{untuk } y > x$$

Jadi, diberikan $X = x$, Y memiliki distribusi eksponensial dengan parameter $\theta = 1$ dan parameter lokasi x. Rata-rata bersyarat dari Y yang diberikan $X = x$ adalah

$$E(Y \mid x) = \int_{-\infty}^{\infty} y\, h(y \mid x)\, dy$$

$$= \int_x^{\infty} y\, e^{-(y-x)}\, dy$$

$$= \int_0^{\infty} (z + x) e^{-z}\, dz \qquad \textit{dimana } z = y - x$$

$$= x \int_0^{\infty} e^{-z}\, dz + \int_0^{\infty} z\, e^{-z}\, dz$$

$$= x\Gamma(1) + \Gamma(2)$$

$$= x + 1$$

Demikian pula, menghitung momen kedua dari distribusi $h(y \mid x)$

$$E(Y^2 \mid x) = \int_{-\infty}^{\infty} y^2 \, h(y \mid x) \, dy$$

$$= \int_{x}^{\infty} y^2 e^{-(y-x)} \, dy$$

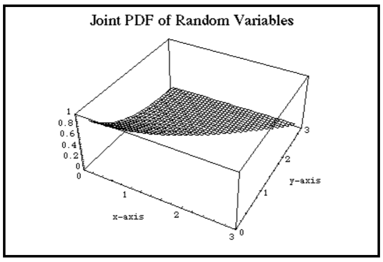
$$= \int_{0}^{\infty} (z+x)^2 \, e^{-z} \, dz \qquad dimana\, z = y - x$$

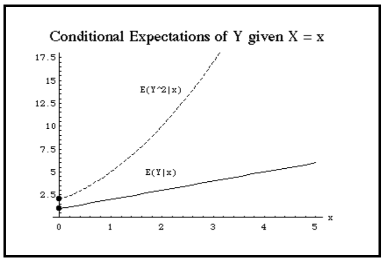
$$= x^2 \int_{0}^{\infty} e^{-z} \, dz + \int_{0}^{\infty} z^2 e^{-z} \, dz + 2x \int_{0}^{\infty} z \, e^{-z} \, dz$$

$$= x^2 \, \Gamma(1) + \Gamma(3) + 2x \Gamma(2)$$

$$= x^2 + 2 + 2x$$

$$= (1+x)^2 + 1$$





Karena itu

$$Var(Y \mid x) = E(Y^2 \mid x) - (E(Y \mid x))^2$$

$$= (1+x)^2 + 1 - (1+x)^2$$

$$= 1$$

Catatan 1.2. Varians dari Y adalah 2. Hal ini dapat dilihat sebagai berikut :

Karena, marginal Y diberikan oleh

$$f_2(y) = \int_0^y e^{-y}\,dx = ye^{-y}$$

Nilai harapan dari Y adalah

$$E(Y) = \int_0^\infty y^2 e^{-y}\,dy = \Gamma(3) = 2$$

Dan

$$E(Y^2) = \int_0^\infty y^3 e^{-y}\,dy = \Gamma(4) = 6$$

Jadi, varians dari Y adalah

$$Var(Y) = 6 - 4 = 2$$

Akan tetapi, diberikan $X = x$, varians dari Y adalah 1.

Teorema 1.3. Misalkan X dan Y adalah dua variabel acak dengan masing-masing dengan mean μ_x dan μ_y, dan standar deviasi σ_X dan σ_Y. Jika ekspektasi bersyarat Y dengan syarat $X = x$ adalah linear pada x, maka

$$E_x\left(Var(Y|x)\right) = \left(1-\rho^2\right)Var(Y)$$

di mana ρ menunjukkan koefisien korelasi dari X dan Y.

Contoh 1.7. Misalkan $E(Y\,|\,X = x) = 2x$ dan $E(Y\,|\,X = x) = 4x^2$ dan misalkan X memiliki distribusi seragam pada interval dari 0 sampai 1. Berapakah varians dari Y?

Jawab :

Jika $E(Y\,|\,X = x)$ adalah fungsi linier dari x, maka

$$E(Y|X = x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

Dan

$$E_x\left(Var\left(Y|x\right)\right)=\sigma_Y^2\left(1-\rho^2\right)$$

Kita dapatkan

$$\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}\left(x-\mu_X\right)=2x$$

Karenanya, dengan menyamakan koefisien dari suku-suku x, kita dapatkan

$$\rho\frac{\sigma_Y}{\sigma_X}=2$$

yang mana

$$\rho=2\frac{\sigma_Y}{\sigma_X} \qquad (1.6)$$

Maka

$$Var\left(Y|X\right)=4x^2$$

Saat $X \sim UNIF\left(0,1\right)$, kita mendapatkan kepadatan X menjadi $f\left(x\right)=1$ di atas interval $\left(0,1\right)$ Oleh karena itu,

$$E_x\left(Var\left(Y|X\right)\right)=\int_{-\infty}^{\infty}Var\left(Y|X=x\right)f\left(x\right)dx$$

$$=\int_0^1 4x^2dx$$

$$=4\left[\frac{x^3}{3}\right]_0^1$$

$$=\frac{4}{3}$$

Dari teorema 1.3,

$$\frac{4}{3} = E_x\left(Var\left(Y|X\right)\right)$$

$$= \sigma_Y^2\left(1-\rho^2\right)$$

$$= \sigma_Y^2\left(1-\frac{4\sigma_X^2}{\sigma_Y^2}\right)$$

$$= \sigma_Y^2 - 4\sigma_X^2$$

Karena itu

$$\sigma_Y^2 = \frac{4}{3} + 4\sigma_X^2$$

Karena $X \sim UNIF\left(0,1\right)$, varians dari X diberikan oleh $\sigma_X^2 = \frac{1}{12}$. Oleh karena itu, varians Y diberikan oleh

$$\sigma_Y^2 = \frac{4}{3} + \frac{4}{12} = \frac{16}{12} + \frac{4}{12} = \frac{20}{12} = \frac{5}{3}$$

Contoh 1.8. Misalkan $E\left(X|Y=y\right) = 3y$ dan $Var\left(X|Y=y\right) = 2$, dan misalkan Y memiliki fungsi kepadatan

$$f\left(y\right) = \begin{cases} e^{-y} & jika\ y > 0 \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Berapakah Varians X ?

Jawab :

Dengan menggunakan Teorema 1.3 didapatkan

$$Var\left(X|Y=y\right) = \sigma_X^2\left(1-\rho^2\right) = 2 \qquad (1.7)$$

Dan

$$\mu_X + \rho \frac{\sigma_X}{\sigma_Y}\left(y - \mu_Y \right) = 3y$$

Sehingga

$$\rho = 3\frac{\sigma_Y}{\sigma_X}$$

Karenanya dari (1.7), kita mendapatkan $E_Y\left(Var\left(X|Y \right) \right) = 2$ dan

$$\sigma_X^2 \left(1 - 9\frac{\sigma_Y^2}{\sigma_X^2} \right) = 2$$

$$\sigma_X^2 = 9\sigma_Y^2 + 2$$

Sekarang kita akan hitung varians Y. Untuk itu, kita memerlukan $E\left(Y \right)$ dan $E\left(Y^2 \right)$

$$E\left(Y \right) = \int_0^\infty yf\left(x \right) dy$$

$$= \int_0^\infty ye^{-y} dy$$

$$= \Gamma\left(2 \right)$$

$$= 1$$

Demikian pula

$$E\left(Y^2 \right) = \int_0^\infty y^2 f\left(x \right) dy$$

$$= \int_0^\infty y^2 e^{-y} dy$$

$$= \Gamma(3)$$

$$= 2$$

Karena itu

$$Var(Y) = E(Y^2) - [E(Y)]^2 = 2 - 1 = 1$$

Oleh karena itu, varians X dapat dihitung

$$\sigma_X^2 = 9\sigma_Y^2 + 2$$

$$= 9(1) + 2$$

$$= 11$$

Catatan 1.3. Perhatikan bahwa, pada Contoh 1.8, kita menghitung varians Y secara langsung menggunakan bentuk $f(y)$. Mudah untuk dihitung bahwa $f(y)$ memiliki bentuk kepadatan eksponensial dengan parameter $\theta = 1$ dan oleh karena itu variansnya adalah kuadrat dari parameter tersebut sehingga $\sigma_Y^2 = 1$

1.3. Kurva Regresi dan Kurva Skedastik

Salah satu tujuan utama dalam studi statistik adalah untuk mengetahui hubungan antara dua atau lebih variabel acak. Misalnya, perusahaan ingin mengetahui hubungan antara potensi penjualan produk baru dalam hal harganya.

Definisi 1.4. Misalkan X dan Y adalah dua variabel acak dengan fungsi kepadatan probabilitas gabungan $f(x, y)$ dan misalkan $h(y\,|x)$ adalah kepadatan bersyarat dari Y yang diberikan $X = x$. Sehingga rata-rata (mean) bersyaratnya

$$E(Y|X = x) = \int_{-\infty}^{\infty} y\, h(y|x)dy$$

disebut fungsi regresi Y pada X. Grafik dari fungsi regresi Y pada X ini dikenal sebagai kurva regresi Y pada X.

Contoh 1.9. Misalkan X dan Y adalah dua variabel acak dengan kepadatan gabungan

$$f(x,y)=\begin{cases} xe^{-x(1+y)} & dimana\ x>0\ ;\ y>0 \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Berapa fungsi regresi Y pada X ?

Jawab :

$$f_1(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$$

$$=\int_0^{\infty} xe^{-x(1+y)}dy$$

$$=\int_0^{\infty} xe^{-x}e^{-xy}dy$$

$$=xe^{-x}\int_0^{\infty} e^{-xy}dy$$

$$=xe^{-x}\left[-\frac{1}{x}e^{-xy}\right]_0^{\infty}$$

$$=e^{-x}$$

Kepadatan bersyarat dari Y dengan syarat $X=x$ adalah

$$h(y|x)=\frac{f(x,y)}{f_1(x)}$$

$$=\frac{xe^{-x(1+y)}}{e^{-x}}$$

$$= xe^{-xy}$$

Rata-rata bersyarat dari Y dengan syarat $X = x$ adalah

$$E(Y|X = x) = \int_{-\infty}^{\infty} y\, h(y|x) dy$$

$$= \int_{0}^{\infty} yxe^{-xy} dy$$

$$= \frac{1}{x}\int_{0}^{\infty} ze^{-z} dz$$

$$= \frac{1}{x}\Gamma(2)$$

$$= \frac{1}{x}$$

Jadi, fungsi regresi (atau persamaan) dari Y pada X diberikan oleh

$$E(Y|x) = \frac{1}{x} \qquad untuk\; 0 < x < \infty$$

Definisi 1.5.

Misalkan X dan Y adalah dua variabel acak dengan fungsi kepadatan probabilitas gabungan $f(x, y)$ dan misalkan $E(Y \mid X = x)$ menjadi fungsi regresi dari Y pada X. Jika fungsi regresi ini linier, maka $E(Y \mid X = x)$ disebut regresi linier dari Y pada X. Sebaliknya disebut regresi nonlinier dari Y pada X.

Contoh 1.10.

Diberikan garis regresi $E(Y|X = x) = x + 2$ dan $E(X|Y = y) = 1 + \frac{1}{2}y$ berapa nilai harapan (ekspektasi) dari X ?

Jawab :

Karena ekspektasi bersyarat $E\,(Y \mid X = x)$ linear dalam x, Dapatkan

$$\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}\left(x - \mu_X\right) = x + 2$$

Oleh karena itu, menyamakan koefisien x dan konstanta, kita dapatkan

$$\rho \frac{\sigma_Y}{\sigma_X} = 1 \qquad (1.8)$$

Dan

$$\mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X = 2 \qquad (1.9)$$

Dengan menggunakan persamaan (1.8) terhadap persamaan (1.9) kita dapatkan

$$\mu_Y - \mu_X = 2$$

Demikian pula, karena $E\left(X | Y = y\right)$ linear dalam y, kita dapatkan

$$\rho \frac{\sigma_X}{\sigma_Y} = \frac{1}{2}$$

Dan

$$\mu_X - \rho \frac{\sigma_X}{\sigma_Y} \mu_Y = 1$$

$$2\mu_X - \mu_Y = 2$$

$$\mu_X = 4$$

Catatan 1.4. Dalam statistik, regresi linear pada umumnya berarti ekspektasi bersyarat $E(Y|x)$ adalah linear dalam parameter, tetapi tidak dalam x. Oleh karena itu, $E(Y\,|x) = \alpha + \theta x^2$ akan menjadi model linear, sedangkan $E(Y\,|x) = \alpha x^{\theta}$ bukan merupakan model regresi linear.

Definisi 1.6.

Misalkan X dan Y adalah dua variabel acak dengan fungsi kepadatan probabilitas gabungan $f(x,y)$ dan misalkan $h(y|x)$ adalah kepadatan bersyarat Y dengan syarat $X=x.$ Kemudian varians bersyarat

$$Var(Y|X=x)=\int_{-\infty}^{\infty} y^2 h(y|x)\,dy$$

disebut fungsi skedastik Y pada X. Grafik fungsi skedastik Y pada X ini disebut sebagai kurva skedastik Y pada X.

LATIHAN SOAL BAB 1

1. Diketahui garis regresi $E(Y|X=x)=x+2$ dan $E(X|Y=y)=1+\frac{1}{2}y$, Berapa nilai ekspektasi dari Y?

Jawab :

$$E\ (Y|X=x)=x+2$$

$$E[Y]=E_x[E[Y|X=x]]$$

$$=E_x[x+2]$$

$$=E[X]+2 ....(1)$$

Dan

$$E(X|Y=y)=1+\frac{1}{2}y$$

$$E[X]=E_y[E[X|Y=y]]$$

$$=E_y\left[1+\frac{1}{2}y\right]$$

$$=1+\frac{1}{2}E[y]\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Substitusi persamaan (2) ke persamaan (1)

$$E[Y]=1+\frac{1}{2}E[Y]+2$$

$$E[Y]=3+\frac{1}{2}E[Y]$$

$$E[Y]-\frac{1}{2}E[Y]=3$$

$$\frac{1}{2}E[Y]=3$$

$$E[Y]=3\times 2$$

$$E[Y]=6$$

Jadi, nilai harapan dari Y adalah 6

2. Jika kepadatan gabungan dari X dan Y adalah

$$f(x,y)=\begin{cases} k & jika\ -1<x<1; x^2<y<1 \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Dengan k adalah konstanta, berapakah $E(Y|X=x)$?

Jawab :

$$f(x)=\int_{x^2}^{1} f(x,y)\,dy$$

$$= \int_{x^2}^{1} k\, dy$$

$$= k\left[y\right]_{x^2}^{1}$$

$$= k\left[1 - x^2\right]$$

Fungsi distribusi kumulatif dari $Y \mid X = x$

$$f\left(Y|X = x\right) = \frac{f\left(x, y\right)}{f\left(X = x\right)}$$

$$f\left(Y|X = x\right) = \frac{k}{k\left[1 - x^2\right)}$$

$$f(Y \mid X = x) = \left[1 - x^2\right]^{-1} \qquad untuk\, x^2 < y < 1$$

\

$$E\left(Y|X = x\right) = \int_{y}^{1} y\, f\left(X = x\right) dy$$

$$= \int_{x^2}^{1} y\left(\left[1 - x^2\right]^{-1}\right) dy$$

$$= \left[1 - x^2\right]^{-1}\left[\frac{y^2}{2}\right]_{x^2}^{1}$$

$$= \frac{1}{2}\left[1 - x^2\right]^{-1}\left[1^2 - \left(x^2\right)^2\right]$$

$$= \frac{1}{2}\left[1 - x^2\right]^{-1}\left[\left(1 - x^2\right)\left(1 + x^2\right)\right] menggunakan\, a^2 + b^2 = \left(a - b\right)\left(a + b\right)$$

$$=\frac{1}{2}\left(1+x^2\right)$$

Jadi, $E\left(Y|X=x\right)=\frac{1}{2}\left(1+x^2\right)$

3. Jika kepadatan gabungan dari X dan Y didefinisikan oleh

$$f\left(x,y\right)=\begin{cases}10xy^2 & jika\, 0<x<y<1\\ \\ 0 & untuk\ yang\ lainnya\end{cases}$$

Berapakah $E(X^2|Y=y)$?

Jawab :

Densitas marginal dari $f_y\left(y\right)$ diberikan oleh

$$f_y\left(y\right)=\int_{-\infty}^{\infty}f\left(x,y\right)dy$$

$$=\int_0^1 10xy^2 \quad dy$$

$$=10\int_0^1 xy^2 \quad dx$$

$$=10\left[\frac{1}{3}xy^3\right]_0^1$$

$$=10\left(\frac{1}{2}.x.(1)^3\right)$$

$$=5x$$

Selanjutnya menentukan densitas bersyarat dari X diberikan $Y=y$

$$g(x|y)=\frac{f\left(x,y\right)}{f_y\left(y\right)}$$

$$= \frac{10xy^2}{5x}$$

$$= 2y^2$$

Maka diperoleh

$$E\left(X^2 | Y = y\right) = \int_{-\infty}^{\infty} x^2 g(x|y) dx$$

$$= \int_0^1 x^2 \, 2y^2 \, dx$$

$$= 2\int_0^1 x^2 . y^2 \, dx$$

$$= 2\int_0^1 x^3 \, y^2 dx$$

$$= 2\left[\frac{1}{4}x^4 y^2\right]_0^1$$

$$= 2\left(\frac{1}{4}(1)^4 \, y^2\right)$$

$$= \frac{1}{2}y^2$$

Jadi, $E\left(X^2 | Y = y\right) = \frac{1}{2}y^2$

4. Misalkan X dan Y fungsi densitas gabungan

$$f(x,y) = \begin{cases} 2e^{-2(x+y)} & jika \, 0 < x < y < \infty \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Berapa nilai harapan dari Y diberikan $X = x$, untuk $x > 0$?

Jawab :

$$E(Y|X = x) = \int_{-\infty}^{\infty} y.f_{\frac{y}{x}}(y)\,dy$$

$$= \int_{-\infty}^{\infty} y \frac{f(x,y)}{f(x)}\,dy$$

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \int_{0}^{\infty} 2e^{-2(x+y)}\,dy$$

$$= \left[2e^{-2x}\frac{e^{-2y}}{-2}\right]_0^{\infty}$$

$$= -e^{-2x}\left(e^{-\infty} - e^{0}\right)$$

$$f(x) = e^{-2x} \quad untuk\, 0 < x < y$$

$$E(Y|X = x) = \int_{-\infty}^{\infty} y \frac{f(x,y)}{f(x)}\,dy$$

$$= \int_{0}^{\infty} y \frac{2e^{-2(x+y)}}{2ye^{-2x}}\,dy$$

$$= 2\int_0^\infty y \frac{e^{-2x}e^{-2y}}{2ye^{-2x}}\,dy$$

$$= 2\int_0^\infty ye^{-2y}\,dy$$

$$= 2\left[y\frac{e^{-2y}}{-2}\right]_0^\infty - \int_0^\infty 1\frac{e^{-2y}}{-2}\,dy$$

$$= 2\left[0 + \frac{1}{2}\frac{e^{-2y}}{-2}\right]_0^\infty$$

$$= 2\left[-\frac{1}{4}\left(e^{-\infty} - e^0\right)\right]$$

$$= \frac{2}{4} + x$$

$$= \frac{1}{2} + x$$

Jadi, $E\left(Y|X = x\right) = \frac{1}{2} + x$

5. Misalkan X dan Y fungsi densitas gabungan

$$f\left(x,y\right) = \begin{cases} 8xy & jika\, 0 < x < 1; 0 < y < x \\ 0 & untuk\ yang\, lainnya \end{cases}$$

Berapa kurva regresi y pada x, yaitu $E(Y|X = x)$?

Jawab :

$$f_x\left(x\right) = \int_{-\infty}^{\infty} f\left(x,y\right) dy$$

$$= \int_0^1 8xy\, dy$$

$$= 8y \int_0^1 x\, dy$$

$$= 8y \left[\frac{x^2}{2} \right]_0^1$$

$$= 8y \left[\frac{1}{2} - 0 \right]$$

$$= 4y$$

$$\mathrm{E}\left(\mathrm{Y} \mid X = x\right) = \frac{f\left(x, y\right)}{f\left(x\right)}$$

$$= \frac{8xy}{4y}$$

$$= 2x$$

Jadi, kurva regresi y pada x, yaitu $E(Y \mid X = x)$ adalah $2x$

6. Misalkan X dan Y adalah variabel acak dengan mean masing-masing μ_X dan μ_Y, dan

$$E\left(Y|X = x\right) = -\frac{1}{3}x + 10 \;\; dan \quad E\left(X|Y = y\right) = -\frac{3}{4}y + 2$$

Berapa nilai μ_X dan μ_Y?

Jawab :

X dan Y adalah variabel random dengan mean yaitu μ_X dan μ_Y

$$E\left(Y \mid X = x\right) = -\frac{1}{3}x + 10$$

$$E\left(X \mid Y = y\right) = -\frac{3}{4}y + 2$$

Dari probabilitas bersyarat,

$$E\left(X\right) = E\left(E\left(X|Y\right)\right)$$

Sekarang menghitung,

$$E\left(Y\right) = E\left(E\left(Y|X = x\right)\right)$$

$$E\left(Y\right) = E\left(-\frac{1}{3}x + 10\right)$$

$$E\left(Y\right) = -\frac{1}{3}E\left(x\right) + 10 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots..(1)$$

$$E\left(X\right) = E\left(E\left(X|Y = Y\right)\right)$$

$$E\left(X\right) = E\left(-\frac{3}{4}y + 2\right)$$

$$E\left(X\right) = -\frac{3}{4}E\left(y\right) + 2 \ldots\ldots\ldots\ldots\ldots..(2)$$

Substitusikan persamaan 1 ke persamaan 2

$$\mu_X \,/\, E\left(X\right) = -\frac{3}{4}\left[-\frac{1}{3}E\left(x\right) + 10\right] + 2$$

$$E\left(X\right) = \frac{1}{4}E\left(X\right) - \frac{10\left(3\right)}{4} + 2$$

$$E(X)=\frac{1}{4}E(X)-\frac{30}{4}+2$$

$$E(X)=\frac{E(X)-30+8}{4}$$

$$4E(X)=E(X)-22$$

$$E(X)=-\frac{22}{3}$$

Substitusikan $E(X)$ ke persamaan $E(Y)$

$$\mu_Y \,/\, E(Y)=-\frac{1}{3}E(x)+10$$

$$=-\frac{1}{3}\left[-\frac{22}{3}\right]+10$$

$$=\frac{22}{9}+10$$

$$=\frac{22}{9}+\frac{90}{9}$$

$$=\frac{112}{9}$$

Jadi mean dari X yaitu $\mu_X=-\frac{22}{3}$ dan mean dari Y yaitu $\mu_Y=\frac{112}{9}$

7. Misalkan X dan Y memiliki densitas gabungan

$$f(x,y)=\begin{cases}\frac{24}{5}(x+y) & untuk\, 0\le 2y\le x\le 1\\ \\ 0 & untuk\ yang\, lainnya\end{cases}$$

Berapa ekspektasi bersyarat dari X diberikan $Y=y$?

Jawab :

Fungsi kepadatan probabilitas marginal dari Y adalah

$$f_y(y)=\int_{-\infty}^{\infty} f(x,y)\,dx$$

$$=\int_{2y}^{1}\frac{24}{5}(x+y)\,dx$$

$$=\frac{24}{5}\left[\frac{x^2}{2}+xy\right]_{2y}^{1}$$

$$=\frac{24}{5}\left[\left(\frac{1}{2}+y\right)-\left(2y^2+2y^2\right)\right]$$

$$=\frac{24}{5}\left[\frac{1}{2}+y-4y^2\right]$$

Jadi,

$$f_y(y)=\begin{cases}\frac{24}{5}\left[\frac{1}{2}+y-4y^2\right] & untuk\ \ 0\le y\le\frac{1}{2}\\ \\ 0 & untuk\ yang\, lainnya\end{cases}$$

Fungsi kepadatan probabillitas bersyarat dari $(X \mid Y = y)$ adalah

$$g(x \mid y) = \frac{f(x,y)}{f_y}$$

$$= \frac{\frac{24}{5}(x+y)}{\frac{24}{5}\left(\frac{1}{2}+y-4y^2\right)}$$

$$= \frac{(x+y)}{\left(\frac{1}{2}+y-4y^2\right)}$$

$$g(x,y) = \begin{cases} \dfrac{(x+y)}{\left(\frac{1}{2}+y-4y^2\right)} & \textit{untuk} \quad 2y \le x \le 1 \\ 0 & \textit{untuk yang lainnya} \end{cases}$$

Sekarang, kita memerlukan

$$E[X \mid Y = y] = \int_{2y}^{1} x\, g(x \mid y)\, dx$$

$$= \frac{1}{\left(\frac{1}{2}+y-4y^2\right)} \int_{2y}^{1} (x^2 + xy)\, dx$$

$$= \frac{1}{\left(\frac{1}{2}+y-4y^2\right)} \left[\frac{x^3}{3} + \frac{x^2 y}{2} \right]_{2y}^{1}$$

$$= \frac{1}{\left(\frac{1}{2}+y-4y^2\right)}\left[\left(\frac{1}{3}+\frac{y}{2}\right)-\left(\frac{8y^3}{3}+\frac{4y^3}{3}\right)\right]$$

$$= \frac{1}{\left(\frac{1}{2}+y-4y^2\right)}\left[\frac{1}{3}+\frac{y}{2}-\frac{14y^3}{3}\right]$$

$$= \frac{6\left(\frac{1}{3}+\frac{y}{2}-\frac{14y^3}{3}\right)}{6\left(\frac{1}{2}+y-4y^2\right)}$$

$$= \frac{1\left(2+3y-28y^3\right)}{3\left(1+2y-8y^2\right)}$$

8. Misalkan X dan Y memiliki densitas gabungan

$$f(x,y)=\begin{cases} cxy^2 & \textit{untuk } 0\le y\le 2x; 1\le x\le 5 \\ 0 & \textit{untuk yang lainnya} \end{cases}$$

Berapa ekspektasi bersyarat dari Y dengan $X=x$?

Jawab :

Fungsi kepadatan probabilitas marginal dari X adalah

$$f_x(x)=\int_0^{2x} cxy^2\,dy$$

$$= cx\int_0^{2x} y^2\,dy$$

$$= cx\left[\frac{y^3}{3}\right]_0^{2x}$$

$$= cx\frac{(2x)^3}{3}$$

$$= \frac{cx(8x^3)}{3}$$

$$= \frac{8cx^4}{3}$$

Fungsi kepadatan probabilitas bersyarat dari Y dengan syarat $X = x$ adalah

$$f_{Y|x}(Y|X) = \frac{f(x,y)}{f_x(x)}$$

$$= \frac{cxy^2}{\frac{8cx^4}{3}}$$

$$= cxy^2 \times \frac{3}{8cx^4}$$

$$= \frac{3y^2}{8x^3}$$

Nilai ekspektasi bersyarat dari Y dengan syarat $X = x$ adalah

$$E(Y|X = x) = \int_0^{2x} y\, f_{Y|X}(Y|X)\,dy$$

$$= \int_0^{2x} \frac{3y^3}{8x^3}\,dy$$

$$= \left[\frac{3y^4}{(8)(4)x^3}\right]_0^{2x}$$

$$= \frac{3(2x)^4}{32x^3}$$

$$= \frac{(3)(16x^4)}{32x^3}$$

$$= \frac{48x^4}{32x^3}$$

$$= \frac{3x}{2}$$

9. Misalkan X dan Y memiliki kepadatan gabungan

$$f(x,y) = \begin{cases} e^{-y} & \textit{untuk } y \geq x \geq 0 \\ 0 & \textit{untuk yang lainnya} \end{cases}$$

Berapa ekspektasi bersyarat dari X dengan syarat $Y = y$?

Jawab :

Fungsi kepadatan probabilitas marginal pada Y

$$f(y) = \int_0^y e^{-y} dx$$

$$= e^{-y} [x]_0^y$$

$$= e^{-y} [y - 0]$$

$$= ye^{-y} \qquad \textit{dimana } y \geq 0$$

Fungsi kepadatan probabilitas bersyarat dari X dengan syarat $Y = y$

$$g(x|y)=\frac{f\left(x,y\right)}{f\left(y\right)}$$

$$=\frac{e^{-y}}{ye^{-y}}$$

$$=\frac{1}{y} \qquad dimana\, 0\le x<y$$

Ekspektasi bersyarat dari X dengan syarat $Y=y$

$$E(X|Y=y)=\int_0^y x\,g(x|y)dx$$

$$=\int_0^y x\frac{1}{y}dx$$

$$=\frac{1}{y}\left[\frac{x^2}{2}\right]_0^y$$

$$=\frac{1}{2y}\left[y^2-0\right]$$

$$=\frac{y^2}{2y}$$

$$=\frac{y}{2}$$

Jadi, $E(X|Y=y)=\dfrac{y}{2}$

10. Misalkan X dan Y memiliki kepadatan gabungan

$$f(x,y)=\begin{cases} 2xy & untuk\, 0 \le y \le 2x < 2 \\ \\ 0 & untuk\; yang\; lainnya \end{cases}$$

Berapa ekspektasi bersyarat dari X dengan syarat $Y = y$?

Jawab :

Fungsi kepadatan probabilitas marginal dari X adalah

$$f_x(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$$

$$=\int_0^{2x} 2xy\,dy$$

$$=2x\left[\frac{y^2}{2}\right]_0^{2x}$$

$$=2x\,\frac{(2x)^2}{2}$$

$$\left[0 \le 2x \le 2 \quad \leftrightarrow \quad 0 \le x \le 1\right]$$

$$f_x(x)=\begin{cases} 4x^3, & untuk\, 0 \le x \le 1 \\ \\ 0 & untuk\; yang\; lainnya \end{cases}$$

Fungsi kepadatan probabilitas bersyarat dari $(Y \mid X = x)$ adalah

$$f_{Y\,|\,X=x}(y) = \frac{f(x,y)}{f_x(x)}$$

$$= \frac{2xy}{4x^3}$$

$$= \frac{y}{2x^2}$$

Jadi

$$f_{Y\,|\,X=x}(y) = \begin{cases} \dfrac{y}{2x^2}; & \textit{untuk}\; 0 \le y \le 2x \\ 0; & \textit{untuk yang lainnya} \end{cases}$$

Sekarang, kita peroleh

$$E[Y\,|\,X=x] = \int_0^{2x} y\; f_{Y|X=x}(y)\,dy$$

$$= \frac{1}{2x^2}\int_0^{2x} y^2 dy$$

$$= \frac{1}{2x^2}\left[\frac{y^3}{3}\right]_0^{2x}$$

$$= \frac{1}{6x^2}\left[(2x)^3\right]$$

$$= \frac{8x^3}{6x^2}$$

$$= \frac{4}{3}x$$

Jadi, ekspektasi bersyarat dari X dengan syarat $Y = y$ adalah $\frac{4}{3}x$

11. Misalkan $E(Y|X = x) = 2 + 5x$, $Var(Y|X = x) = 3$, dan jika X memiliki fungsi kepadatan

$$f(x) = \begin{cases} \frac{1}{4}xe^{-\frac{x}{2}} & untuk\, 0 < x < \infty \\ \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Berapa mean dan varians dari variabel acak Y ?

Jawab :

Diketahui

$$f(x) = \begin{cases} \frac{1}{4}xe^{-\frac{x}{2}} & untuk\, 0 < x < \infty \\ \\ 0 & untuk\ yang\ lainnya \end{cases}$$

$$E((y\ |x)) = 2 + 5x$$

$$Var(Y|x) = 3$$

Sehingga,

$$E(y) = E\ |E(y\ |x = x) = \int_0^\infty (2x + 5)\, fx\, dx$$

$$= \frac{1}{4}\int_0^\infty \left(2xe^{-\frac{x}{2}} + 5x^2e^{-\frac{x}{2}}\right)dx$$

$$= \frac{1}{4}\left[-4\left(x+2\right)e^{-\frac{x}{2}} - 10(x^2 + 4x + 8)e^{-\frac{x}{2}}\right]_0^\infty$$

$$= \frac{1}{4}[88]$$

$$E\left(y\right) = 22$$

Sekarang :

$$Var\left(Y|X = x\right) = E(Y^2|X = x) - (E(\text{Y} \mid X = x))^2$$

$$3 = E(Y^2|X = x) - \left[2 + 5x\right]^2$$

$$3 = E(Y^2|X = x) - 4 - 25x^2 - 20x$$

$$E(Y^2|X = x) = 7 - 25x^2 - 20x$$

$$E\left(y^2\right) = E\left|E\left(y\right|x = x\right) = \int_0^\infty \left(7 - 25x^2 - 20x\right) fx\, dx$$

$$= \frac{1}{4}\int_0^\infty \left(7xe^{-\frac{x}{2}} + 25x^3 e^{-\frac{x}{2}} + 20x^2 e^{-\frac{x}{2}}\right) dx$$

$$= \frac{1}{4}\left[-14\left(x+2\right)e^{-\frac{x}{2}} - 50(x^3 + 6x^2 + 24x + 48)e^{-\frac{x}{2}} - 40(x^2 + 4x + 8)e^{-\frac{x}{2}}\right]_0^\infty$$

$$= \frac{1}{4}\left[14\left(2\right) - 50\left(48\right) + 40\left(8\right)\right]$$

$$= 687$$

Varians dari y adalah

$$Var\left(Y|X=x\right)=E(Y^2|X=x)-(E(\mathrm{Y}\;|X=x))^2$$

$$=687-22^2$$

$$=203$$

12. Misalkan X dan Y memiliki densitas gabungan

$$f\left(x,y\right)=\begin{cases} 2 & untuk\, 0<y<1; dan\;\; 0<x<1 \\ 0 & untuk\; yang\, lainnya \end{cases}$$

Berapa Varians bersyarat dari Y dengan syarat $X=x$

Jawab :

$$Var\left(Y|X=x\right)=E\left(Y^2|X\right)-E(Y|X=2)^2$$

Menentukan fungsi kepadatan probabilitas bersyarat dari $g\left(y\;|x\right)$

$$g(y|x)=\frac{f\left(x,y\right)}{f_x\left(x\right)}$$

$$f_x\left(x\right) marginal\, dari\, x$$

$$f_x\left(x\right)=\int_{-\infty}^{\infty} f\left(x,y\right)dy$$

$$=\int_0^{1-x} 2\,dy$$

$$=2\left[y\right]_0^{1-x}$$

$$= 2(1-x)$$

$$g(y|x) = \frac{2}{2(1-x)}$$

$$g(y|x) = \frac{1}{1-x}$$

Jadi, fungsi kepadatan probabilitas bersyaratnya adalah

$$g(y|x) = \frac{1}{1-x}$$

Menentukan nilai ekspektasi

$$E(Y|X) = \int_{-\infty}^{\infty} y\, g(y|x)\, dy$$

$$= \int_{0}^{1-x} y \frac{1}{1-x} dy$$

$$= \frac{1}{1-x}\left[\frac{y^2}{2}\right]_0^{1-x}$$

$$= \frac{1}{1-x}\frac{(1-x)^2}{2}$$

$$= \frac{(1-x)^2}{2(1-x)}$$

$$= \frac{(1-x)}{2}$$

$$E(Y^2|X) = \int_{-\infty}^{\infty} y^2\, g(y|x)\, dy$$

$$= \int_0^{1-x} y^2 \frac{1}{1-x} dy$$

$$= \frac{1}{1-x}\left[\frac{y^3}{3}\right]_0^{1-x}$$

$$= \frac{1}{1-x}\frac{(1-x)^3}{3}$$

$$= \frac{(1-x)^3}{3(1-x)}$$

$$= \frac{(1-x)^2}{3}$$

$$Var\left(Y|X=x\right) = E\left(Y^2|X\right) - E(Y|X=2)^2$$

$$= \frac{(1-x)^2}{3} - \left[\frac{(1-x)}{2}\right]^2$$

$$= \frac{(1-x)^2}{3} - \frac{(1-x)^2}{4}$$

$$= \frac{4(1-x)^2 - 3(1-x)^2}{12}$$

$$= \frac{(1-x)^2}{12}$$

$$= \frac{1}{12}(1-x)^2$$

13. Misalkan X dan Y meliliki densitas gabungan

$$f(x,y)=\begin{cases}12x & untuk\, 0<y<2x<1\\ \\ 0 & untuk\ yang\ lainnya\end{cases}$$

Berapa varians bersyarat dari Y dengan syarat $X=0,5$?

Jawab :

$$f(y|x)=\frac{f(x,y)}{f_x(x)}$$

$$f(x,y)=12x \qquad 0<y<2x<1$$

$$f_x(x)=\int_{-\infty}^{\infty} f(x,y)\ dy$$

$$=\int_{0}^{2x} 12x\ dy$$

$$=\left[12xy\right]_0^{2x}$$

$$=24x^2 \qquad 0<x<0,5$$

$$f(y|x)=\frac{f(x,y)}{f_x(x)}$$

$$=\frac{12x}{24x^2}$$

$$=\frac{1}{2x}, \qquad\qquad 0<y<1$$

$$f\left(y|x=0,5\right)=\frac{1}{2\times 0,5} \qquad 0<y<1$$

Varians bersyarat dari Y dengan $X=0,5$

$$Var\left(Y|X=0,5\right)=E\left(Y^2|X=0,5\right)-E(Y|X=2,5)^2$$

$$E\left(Y|X=2,5\right)=\int_{-\infty}^{\infty} y\, f\left(y|x=0,5\right)dy$$

$$=\int_0^1 y\times 1\, dy$$

$$=\left[\frac{y^2}{2}\right]_0^1$$

$$=\frac{1}{2}$$

$$E\left(Y^2|X=2,5\right)=\int_{-\infty}^{\infty} y^2\, f\left(y|x=0,5\right)dy$$

$$=\int_0^1 y^2\times 1\, dy$$

$$=\left[\frac{y^3}{3}\right]_0^1$$

$$=\frac{1}{3}$$

$$Var\left(Y|X=0,5\right)=E\left(Y^2|X=0,5\right)-E(Y|X=2,5)^2$$

$$=\frac{1}{3}-\left(\frac{1}{2}\right)^2$$

$$=\frac{1}{3}-\frac{1}{4}$$

$$=\frac{4-3}{12}$$

$$=\frac{1}{12}$$

=

BAB II

TRANSFORMASI VARIABEL ACAK DAN DISTRIBUSINYA

Dalam penerapan ilmu statistik, diberikan distribusi probabilitas variabel acak univariat X, dimana akan dicari distribusi probabilitas dari selain variabel acak univariat $Y=\phi(X)$, di mana ϕ adalah beberapa fungsi yang diketahui. Contohnya, jika kita mengetahui distribusi probabilitas dari variabel acak X, kita akan mengetahui distribusi dari $Y=\ln(X)$. Untuk variabel acak univariat X, beberapa variabel acak transformasi Y dari X yang sering digunakan adalah :

$$Y=X^2,\ Y=|X|,\ Y=\sqrt{|X|},\ Y=\ln(X),\ Y=\frac{X-\mu}{\sigma}\ ,\ dan\quad Y=\frac{X-\mu}{\sigma}^2$$

Demikian pula untuk variabel acak bivariat (X,Y), beberapa transformasi X dan Y yang paling umum adalah :

$$X+Y,\ XY,\ \frac{X}{Y},\ \min\{X,Y\},\ \max\{X,Y\},\ atau\quad \sqrt{X^2+Y^2}$$

Kita mulai dengan contoh untuk variabel acak diskrit univariat

Contoh 2.1. Fungsi kepadatan probabilitas dari variabel acak X ditunjukkan pada tabel di bawah.

X	-2	-1	0	1	2	3	4
$f(x)$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$

Berapakah fungsi kepadatan probabilitas dari variabel acak $Y=X^2$?

Jawab:

Ruang dari variabel acak X adalah $R_X=\{-2,-1,0,1,2,3,4\}$ Maka ruang variabel acak Y adalah $R_Y=\{x^2\,|x\in R_x\}$ Jadi, $R_Y=\{0,1,4,9,16\}$. Sekarang kita menghitung fungsi kepadatan probabilitas $g(y)$ untuk y di R_Y

$$g(0)=P(Y=0)=P(X^2=0)=P(X=0)=\frac{1}{10}$$

$$g(1) = P(Y = 1) = P(X^2 = 1) = P(X = -1) + P(X = 1) = \frac{3}{10}$$

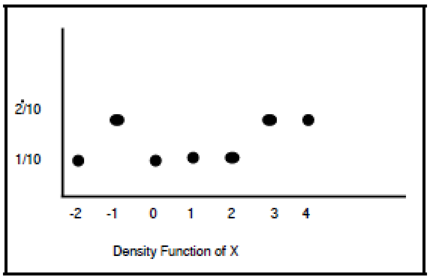
$$g(4) = P(Y = 4) = P(X^2 = 4) = P(X = -2) + P(X = 2) = \frac{2}{10}$$

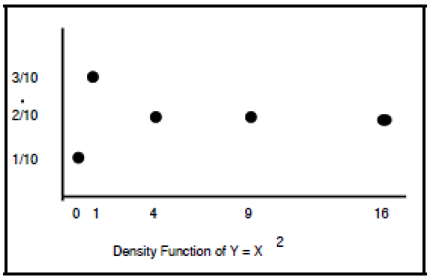
$$g(9) = P(Y = 9) = P(X^2 = 9) = P(X = 3) = \frac{2}{10}$$

$$g(16) = P(Y = 16) = P(X^2 = 16) = P(X = 4) = \frac{2}{10}$$

Kita ringkas distribusi Y dalam tabel berikut :

Y	0	1	4	9	16
$g(y)$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{2}{10}$





Contoh 2.2. Fungsi kepadatan probabilitas dari variabel acak X ditunjukkan pada tabel di bawah ini :

x	1	2	3	4	5	6
$f(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Berapakah fungsi kepadatan probabilitas dari variabel acak $Y = 2X + 1$?

Jawab:

Ruang dari variabel acak X adalah $R_X = \{1,2,3,4,5,6\}$. Maka ruang dari variabel acak Y adalah $R_Y = \{2x+1 \mid x \in R_X\}$. Jadi, $R_Y = \{3,5,7,9,11,13\}$. Selanjutnya kita menghitung fungsi kepadatan probabilitas $g(y)$ untuk y

di R_Y. Fungsi Kepadatan Probabilitas $g(y)$ diberikan oleh

$$g(3) = P(Y = 3) = P(2X + 1 = 3) = P(X = 1) = \frac{1}{6}$$

$$g(5) = P(Y = 5) = P(2X + 1 = 5) = P(X = 2) = \frac{1}{6}$$

$$g(7) = P(Y = 7) = P(2X + 1 = 7) = P(X = 3) = \frac{1}{6}$$

$$g(9) = P(Y = 9) = P(2X + 1 = 9) = P(X = 4) = \frac{1}{6}$$

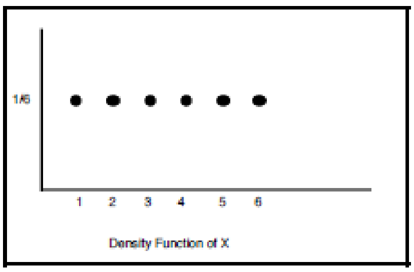
$$g(11) = P(Y = 11) = P(2X + 1 = 11) = P(X = 5) = \frac{1}{6}$$

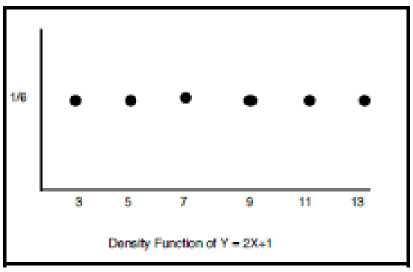
$$g(13) = P(Y = 13) = P(2X + 1 = 13) = P(X = 6) = \frac{1}{6}$$

Kita ringkas distribusi Y dalam tabel berikut.

Y	3	5	7	9	11	13
$g(y)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Distribusi X dan $2X + 1$ diilustrasikan di bawah ini.





Pada contoh 2.1, kita menghitung distribusi (yakni, fungsi kepadatan probabilitas) dari variabel acak yang ditransformasikan $Y = \phi(X)$, di mana $\phi(x) = x^2$. Transformasi ini tidak meningkat atau menurun (yaitu, monotonik) dalam ruang, R_X, dari variabel acak X. Oleh karena itu, distribusi Y ternyata sangat berbeda dengan distribusi X. Pada contoh 2.2, bentuk distribusi dari transformasi variabel acak $Y = \phi(X)$, dimana $\phi(x) = 2x+1$, pada dasarnya sama. Hal ini terutama disebabkan oleh fakta bahwa $\phi(x) = 2x+1$ monotonik di R_X.

Dalam bab ini, kita akan memeriksa fungsi kepadatan probabilitas dari transformasi variabel acak dengan mengetahui fungsi kepadatan dari variabel acak aslinya. Ada beberapa metode untuk menemukan fungsi kepadatan probabilitas dari variabel acak yang ditransformasikan. Beberapa dari metode tersebut adalah:

(1) Metode Fungsi Distribusi

(2) Metode Transformasi

(3) Metode Konvolusi

(4) Metode Fungsi Pembangkit Momen.

Di antara keempat metode ini, Metode Transformasi adalah yang paling banyak digunakan. Metode Konvolusi adalah kasus khusus dari metode ini. Metode Transformasi diturunkan menggunakan Metode Fungsi Distribusi.

2.1. Metode Fungsi Distribusi

Contoh 2.3. Sebuah kotak harus dibuat sedemikian rupa sehingga tingginya 4 inci dan dasarnya adalah X inci kali X inci. Jika X memiliki distribusi normal standar, berapa distribusi volume kotaknya?

Jawab :

Volume kotak adalah variabel acak, karena X adalah variabel acak. Variabel acak V ini diberikan oleh $V = 4X^2$. Untuk mencari kepadatan dari fungsi V, pertama kita tentukan bentuk dari fungsi distribusi $G(v)$ dari V dan kemudian kita turunkan $G(v)$ untuk mencari fungsi kepadatan V. Fungsi distribusi V diberikan oleh

$$G(v) = P(V \le v)$$

$$= P(4X^2 \le v)$$

$$= P\left(-\frac{1}{2}\sqrt{v} \le X \le \frac{1}{2}\sqrt{v}\right)$$

$$= \int_{-\frac{1}{2}\sqrt{v}}^{\frac{1}{2}\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

$$= 2\int_{0}^{\frac{1}{2}\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

Oleh karena itu, dengan Teorema Dasar Kalkulus, kita dapatkan

$$g(v) = \frac{dG(v)}{dv}$$

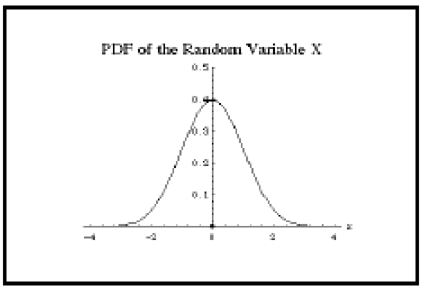
$$= \frac{d}{dv}\left(2\int_{0}^{\frac{1}{2}\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \right)$$

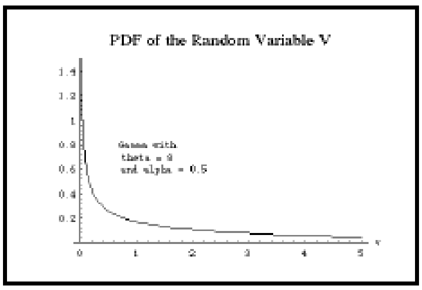
$$= 2\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{1}{2}\sqrt{v}\right)^2} \left(\frac{1}{2}\right)\frac{d\sqrt{v}}{dv}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{8}v} \frac{1}{\sqrt{2v}}$$

$$= \frac{1}{\Gamma\left(\frac{1}{2}\right)\sqrt{8}} v^{\frac{1}{2}-1} e^{-\frac{v}{8}}$$

$$= V \sim GAM\left(8, \frac{1}{2}\right)$$





Contoh 2.4. Jika fungsi kepadatan X adalah

$$f(x) = \begin{cases} \dfrac{1}{2} & untuk\ -1 < x < 1 \\ \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Berapa fungsi kepadatan probabilitas dari $Y = X^2$?

Jawab:

Pertama, kita cari fungsi distribusi kumulatif dari Y dan kemudian dengan diferensiasi, kita dapatkan kepadatan Y. Fungsi distribusi $G(y)$ dari Y diberikan oleh

$$G(y) = P(Y \le y)$$

$$= P(X^2 \le y)$$

$$= P(-\sqrt{y} \le X \le \sqrt{y})$$

$$= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2} dx$$

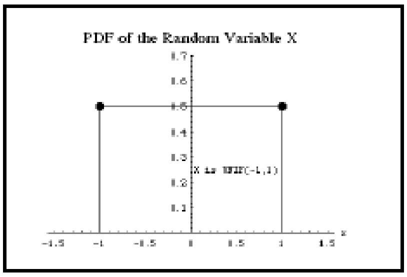
$$= \sqrt{y}$$

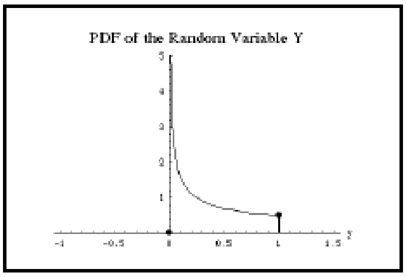
Oleh karena itu, fungsi kepadatan Y diberikan oleh

$$g(y) = \frac{dG(y)}{dy}$$

$$= \frac{d\sqrt{y}}{dy}$$

$$= \frac{1}{2\sqrt{y}} \qquad \textit{untuk}\ 0 < y < 1$$





2.2. Metode Transformasi Kasus Univariat

Teorema berikut adalah dasar dari metode transformasi

Teorema 2.1.

Misalkan X adalah variabel acak kontinu dengan fungsi kepadatan probabilitas $f(x)$. Misalkan $y = T(x)$ menjadi fungsi peningkatan (atau penurunan). Kemudian fungsi kepadatan dari variabel acak $Y = T(X)$ diberikan oleh

$$g(y) = \left|\frac{dx}{dy}\right| f\left(W(y)\right)$$

di mana $x = W(y)$ adalah fungsi invers dari $T(x)$.

Bukti :

Misalkan $y = T(x)$ adalah fungsi yang meningkat. Fungsi distribusi $G(y)$ dari Y diberikan oleh

$$G(y) = P(Y \le y)$$

$$= P\left(T(X) \le y\right)$$

$$= P\left(X \le W(y)\right)$$

$$= \int_{-\infty}^{W(y)} f(x)\,dx$$

Kemudian, dengan turunan kita dapatkan fungsi kepadatan dari Y, yaitu

$$g(y) = \frac{dG(y)}{dy}$$

$$= \frac{d}{dy}\left(\int_{-\infty}^{W(y)} f(x)\,dx \right)$$

$$= f(W(y))\frac{dW(y)}{dy}$$

$$= f(W(y))\frac{dx}{dy} \qquad (saat \quad x = W(Y))$$

Sebaliknya, jika $y = T(x)$ adalah fungsi penurunan, maka fungsi distribusi Y diberikan oleh

$$G(y) = P(Y \leq y)$$

$$= P(T(X) \leq y)$$

$$= P(X \leq W(y)) \qquad (saat \quad T(x) menurun)$$

$$= 1 - P(X \leq W(y))$$

$$= 1 \int_{-\infty}^{W(y)} f(x)\,dx$$

Seperti sebelumnya, dengan melakukan diferensiasi kita mendapatkan fungsi kepadatan dari Y, yaitu

$$g(y) = \frac{dG(y)}{dy}$$

$$= \frac{d}{dy}\left(1 \int_{-\infty}^{W(y)} f(x)\,dx \right)$$

$$= f\left(W(y)\right)\frac{dW(y)}{dy}$$

$$= -f\left(W(y)\right)\frac{dx}{dy} \qquad \left(sejak\; x = W(y)\right)$$

Oleh karena itu, dengan menggabungkan kedua kasus tersebut, kami dapatkan

$$g(y) = \left|\frac{dx}{dy}\right| f\left(W(y)\right)$$

Contoh 2.5. Misalkan $Z = \dfrac{X-\mu}{\sigma}$ Jika $X \sim N\left(\mu, \sigma^2\right)$ berapa fungsi kepadatan probabilitas Z ?

Jawab :

$$z = U(x) = \frac{x-\mu}{\sigma}$$

Oleh karena itu, invers dari U yaitu

$$W(z) = x$$

$$= \sigma z + \mu$$

Karena itu

$$\frac{dx}{dz} = \sigma$$

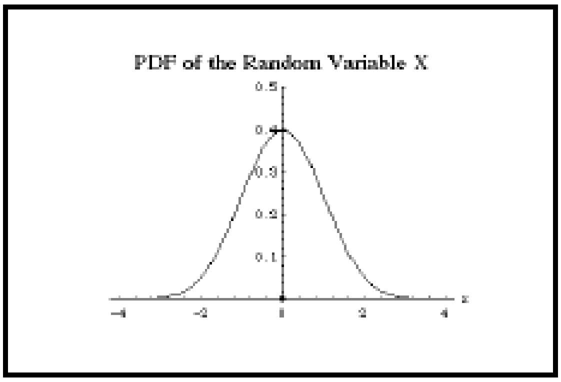
Oleh karena itu, dengan menggunakan Teorema 2.1, kepadatan dari Z yaitu

$$g(y) = \left|\frac{dx}{dy}\right| f(W(y))$$

$$= \sigma \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{W(z)-\mu}{\sigma}\right)^2}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z\sigma+\mu-\mu}{\sigma}\right)^2}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$



Contoh 2.6. Misalkan $Z = \dfrac{X-\mu}{\sigma}$. Jika $X \sim N(\mu, \sigma^2)$, tunjukkan bahwa Z^2 adalah chi-square dengan satu derajat kebebasan, yaitu $Z^2 \sim \chi^2(1)$

Jawab :

$$y = T(x) = \left(\frac{x-\mu}{\sigma}\right)^2$$

$$x = \mu + \sigma\sqrt{y}$$

$$W(y) = \mu + \sigma\sqrt{y}, \qquad y > 0$$

$$\frac{dx}{dy} = \frac{\sigma}{2\sqrt{y}}$$

Kepadatan Y adalah

$$g(y) = \left|\frac{dx}{dy}\right| f\left(W(y)\right)$$

$$= \sigma\frac{1}{2\sqrt{y}} f\left(W(y)\right)$$

$$= \sigma\frac{1}{2\sqrt{y}}\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{W(y)-\mu}{\sigma}\right)^2}$$

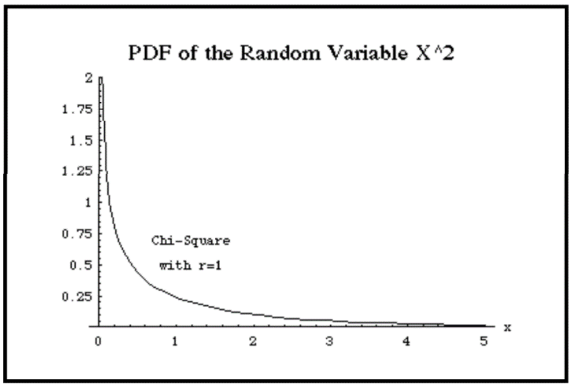
$$= \frac{1}{2\sqrt{2\pi y}} e^{-\frac{1}{2}\left(\frac{\sqrt{y}\sigma+\mu-\mu}{\sigma}\right)^2}$$

$$= \frac{1}{2\sqrt{2\pi y}} e^{-\frac{1}{2}y}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{2}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y}$$

$$= \frac{1}{2\Gamma\left(\frac{1}{2}\right)\sqrt{2}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y}$$

Oleh karena itu $Y \sim \chi 2(1)$.



Contoh 2.7. Misalkan $Y = -\ln X$. Jika $X \sim UNIF(0,1)$ lalu berapakah fungsi kepadatan Y dimana bukan nol?

Jawab :

Kita dapatkan

$$y = T(x) = -\ln x.$$

Oleh karena itu, invers dari $y = T(x)$ yaitu

$$W(y) = x$$

$$= e^{-y}$$

Karena itu

$$\frac{dx}{dy} = -e^{-y}$$

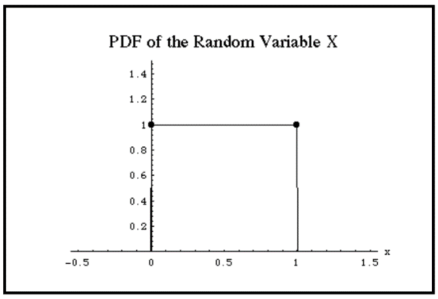
maka dengan menggunakan Teorema 2.1, kepadatan probabilitas Y adalah

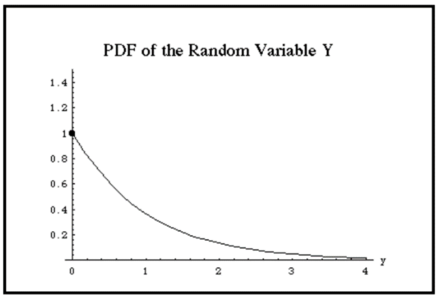
$$g(y) = \left|\frac{dx}{dy}\right| f(W(y))$$

$$= e^{-y} f(W(y))$$

$$= e^{-y}$$

Jadi $Y \sim EXP(1)$. Jadi, jika $X \sim UNIF(0,1)$, maka variabel acak $-\ln X \sim EXP(1)$.





2.3. Metode Transformasi untuk Kasus Bivariat

Pada bagian ini, kita perluas Teorema 2.2 ke kasus bivariat dan menyajikan beberapa contoh untuk menggambarkan pentingnya ekstensi ini. Kita nyatakan teorema ini tanpa bukti

Teorema 2.2. Misalkan X dan Y adalah dua variabel acak kontinu dengan kepadatan gabungan $f(x,y)$. Misalkan $U = P(X,Y)$ dan $V = Q(X,Y)$ adalah fungsi dari X dan Y. Jika fungsi $P(x,y)$ dan $Q(x,y)$ memiliki nilai invers tunggal, katakanlah $X = R(U,V)$ dan $Y = S(U,V)$, maka densitas gabungan $g(u,v)$ dari U dan V diberikan oleh

$$g(u,v) = |J|\, f(R(u,v), S(u,v))$$

dimana J menunjukkan Jacobian dan diberikan oleh

$$J = \det\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}$$

$$= \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial u}\frac{\partial y}{\partial v}$$

Contoh 2.8. Misalkan X dan Y memiliki fungsi kepadatan probabilitas gabungan

$$f(x,y)=\begin{cases} 8xy & untuk\, 0<x<y<1 \\ \\ 0 & untuk\ yang\, lainnya \end{cases}$$

Berapa kepadatan gabungan $U=\dfrac{X}{Y}$ dan $V=Y$?

Jawab :

Karena

$$\left.\begin{aligned} U &= \frac{X}{Y} \\ V &= Y \end{aligned}\right\}$$

Kita dapatkan dengan menyelesaikan X dan Y

$$\left.\begin{aligned} X &= UY = UV \\ Y &= V \end{aligned}\right\}$$

Oleh karena itu, transformasi Jacobian diberikan oleh

$$J=\frac{\partial x}{\partial u}\frac{\partial y}{\partial v}-\frac{\partial x}{\partial u}\frac{\partial y}{\partial v}$$

$$=v.1-u.0$$

$$=v$$

Fungsi kepadatan gabungan dari U dan V adalah

$$g(u,v)=|J|\ \ f\left(R(u,v),S(u,v)\right)$$

$$=|v|\,f(uv,v)$$

$$=v8(uv)v$$

$$= 8uv^3$$

Perhatikan, saat

$$0 < x < y < 1$$

kita punya

$$0 < uv < v < 1$$

Pertidaksamaan terakhir menghasilkan

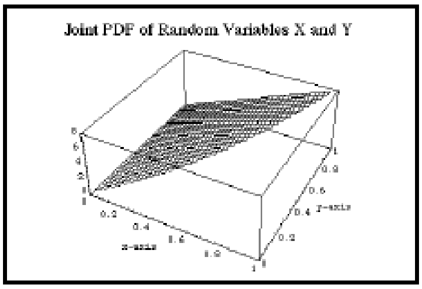
$$\left.\begin{array}{l} 0 < uv < v \\ 0 < v < 1 \end{array}\right\}$$

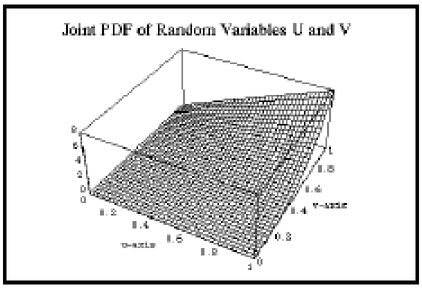
Oleh karena itu, kita dapatkan

$$\left.\begin{array}{l} 0 < u < 1 \\ 0 < v < 1 \end{array}\right\}$$

Jadi, kepadatan gabungan dari U dan V diberikan oleh

$$g(u,v) = \begin{cases} 8uv^3 & \textit{untuk}\ 0 < u < 1; 0 < v < 1 \\ 0 & \textit{untuk yang lainnya} \end{cases}$$





Contoh 2.9. Jika setiap variabel acak independen X dan Y mempunyai fungsi kepadatan.

$$f(x)=\begin{cases} e^{-x} & untuk\; 0<x<\infty \\ \\ 0 & untuk\; yang\; lainnya \end{cases}$$

Berapa kepadatan gabungan dari $U = X$ dan $V = 2X + 3Y$ dan dimana domain kepadatannya adalah positif ?

Jawab :

Saat

$$\left.\begin{aligned} U &= X \\ V &= 2X+3Y \end{aligned}\right\}$$

Kita dapat menyelesaikannya untuk X dan Y

$$\left.\begin{aligned} X &= U \\ Y &= \frac{1}{3}V-\frac{2}{3}U \end{aligned}\right\}$$

Oleh karena itu, transformasi Jacobian diberikan oleh

$$J=\frac{\partial x}{\partial u}\frac{\partial y}{\partial v}-\frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

$$=1.\left(\frac{1}{3}\right)-0.\left(-\frac{2}{3}\right)$$

$$=\frac{1}{3}.$$

Fungsi kepadatan gabungan dari U dan V adalah

$$g(u,v)=|J|\,f\big(R(u,v),S(u,v)\big)$$

$$= \left|\frac{1}{3}\right| f\left(u, \frac{1}{3}v - \frac{2}{3}u\right)$$

$$= \frac{1}{3} e^{-u}\, e^{-\frac{1}{3}v + \frac{2}{3}u}$$

$$= \frac{1}{3} e^{-\left(\frac{u+v}{3}\right)}$$

Karena

$$0 < x < \infty$$

$$0 < y < \infty,$$

Kita dapat

$$0 < u < \infty$$

$$0 < y < \infty$$

Selanjutnya, saat $v = 2u + 3y$ dan $3y > 0,$ kita mempunyai

$$v > 2u.$$

Oleh karena itu, domain dari $g(u,v)$ dimana bukan nol diberikan oleh

$$0 < 2u < v < \infty.$$

Kepadatan gabungan $g(u,v)$ dari variabel acak U dan V diberikan oleh

$$g(u,v) = \begin{cases} \frac{1}{3} e^{-\left(\frac{u+v}{3}\right)} & \text{untuk } 0 < 2u < v < \infty \\ 0 & \text{untuk yang lainnya} \end{cases}$$

Contoh 2.10. Jika X dan Y menjadi variabel acak independen, masing-masing dengan fungsi kepadatan

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & untuk\, 0 < x < \infty \\ \\ 0 & untuk\ yang\, lainnya \end{cases}$$

Dimana $\lambda > 0$. Jika $U = X + 2Y$ dan $V = 2X + Y$. Berapa kepadatan gabungan dari U dan V?

Jawab :

Saat

$$\left.\begin{aligned} U &= X + 2Y \\ V &= 2X + Y \end{aligned}\right\}$$

Kita dapat menyelesaikan X dan Y

$$\left.\begin{aligned} X &= -\frac{1}{3}U + \frac{2}{3}V \\ Y &= \frac{2}{3}U - \frac{1}{3}V \end{aligned}\right\}$$

Oleh karena itu, transformasi Jacobian diberikan oleh

$$J = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

$$= \left(-\frac{1}{3}\right)\left(-\frac{1}{3}\right) - \left(\frac{2}{3}\right)\left(\frac{2}{3}\right)$$

$$= \frac{1}{9} - \frac{4}{9}$$

$$= -\frac{1}{3}$$

Fungsi kepadatan gabungan dari U dan V adalah

$$g(u,v) = |J| f(R(u,v), S(u,v))$$

$$= \left|-\frac{1}{3}\right| f(R(u,v)) f(S(u,v))$$

$$= \frac{1}{3}\lambda e^{\lambda R(u,v)} \lambda e^{\lambda S(u,v)}$$

$$= \frac{1}{3}\lambda^2 e^{\lambda[R(u,v)+S(u,v)]}$$

$$= \frac{1}{3}\lambda^2 e^{-\lambda\left(\frac{u+v}{3}\right)}.$$

Oleh karena itu, kepadatan gabungan $g(u,v)$ dari variabel acak U dan V diberikan oleh

$$g(u,v) = \begin{cases} \frac{1}{3}\lambda^2 e^{-\lambda\left(\frac{u+v}{3}\right)} & \text{untuk } 0 < u < \infty; 0 < v < \infty \\ 0 & \text{untuk yang lainnya} \end{cases}$$

Contoh 2.11. Jika X dan Y menjadi variabel acak independen, masing-masing dengan fungsi kepadatan

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad -\infty < x < \infty.$$

Jika $U = \frac{X}{Y}$ dan $V = Y$. Berapa kepadatan gabungan dari U dan V? Juga, berapa kepadatan dari U?

Jawab :

Saat

$$\left.\begin{aligned} U &= \frac{X}{Y} \\ V &= Y \end{aligned}\right\}$$

Kita dapatkan dengan menyelesaikan X dan Y

$$\left.\begin{aligned} X &= UV \\ Y &= V \end{aligned}\right\}$$

Oleh karena itu, transformasi Jacobian diberikan oleh

$$J = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

$$= v.(1) - u.(0)$$

$$= v.$$

Fungsi kepadatan gabungan dari U dan V adalah

$$g(u,v) = |J|\, f\left(R(u,v), S(u,v)\right)$$

$$= |v|\, f\left(R(u,v)\right) f\left(S(u,v)\right)$$

$$= |v| \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}R^2(u,v)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}S^2(u,v)}$$

$$= |v| \frac{1}{2\pi} e^{-\frac{1}{2}\left[R^2(u,v)+S^2(u,v)\right]}$$

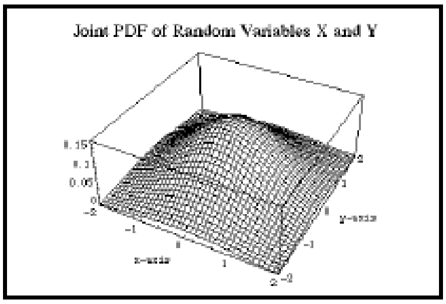
$$= |v| \frac{1}{2\pi} e^{-\frac{1}{2}\left[u^2v^2+v^2\right]}$$

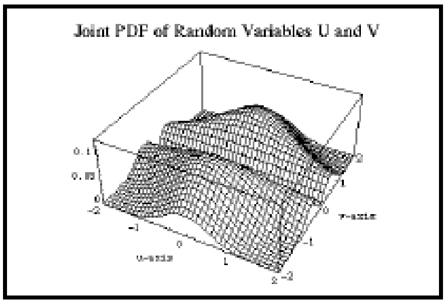
$$= |v| \frac{1}{2\pi} e^{-\frac{1}{2}v^2(u^2+1)}$$

Sehingga kepadatan gabungan $g(u,v)$ dari variabel acak U dan V diberikan oleh

$$g(u,v) = |v| \frac{1}{2\pi} e^{-\frac{1}{2}v^2(u^2+1)}$$

Dimana $-\infty < u < \infty$ dan $-\infty < v < \infty$.





Selanjutnya, kita ingin menemukan kepadatan dari U. Kita bisa memperolehnya dengan menemukan marginal dari U dari kepadatan gabungan U dan V. Oleh karena itu, marginal $g_1(u)$ dari U diberikan oleh

$$g_1(u) = \int_{-\infty}^{\infty} g(u,v)\,dv$$

$$= \int_{-\infty}^{\infty} |v| \frac{1}{2\pi} e^{-\frac{1}{2}v^2(u^2+1)} dv$$

$$= \int_{-\infty}^{0} -v \frac{1}{2\pi} e^{-\frac{1}{2}v^2(u^2+1)} dv + \int_{0}^{\infty} v \frac{1}{2\pi} e^{-\frac{1}{2}v^2(u^2+1)} dv$$

$$= \frac{1}{2\pi}\left(\frac{1}{2}\right)\left[\frac{2}{u^2+1} e^{-\frac{1}{2}v^2(u^2+1)}\right]_{-\infty}^{0} + \frac{1}{2\pi}\left(\frac{1}{2}\right)\left[\frac{-2}{u^2+1} e^{-\frac{1}{2}v^2(u^2+1)}\right]_{0}^{\infty}$$

$$= \frac{1}{2\pi}\frac{1}{u^2+1} + \frac{1}{2\pi}\frac{1}{u^2+1}$$

$$= \frac{1}{\pi\left(u^2+1\right)}.$$

Jadi, $U \sim CAU(1)$.

Catatan 2.1. Jika X dan Y adalah independen dan variabel acak normal standar, maka hasil bagi $\frac{X}{Y}$ selalu merupakan sebuah variabel acak Cauchy. Akan tetapi, sebaliknya tidaklah benar. Sebagai contoh, jika X dan Y independen dan masing-masing memiliki fungsi kepadatan yang sama yakni

$$f(x) = \frac{\sqrt{2}}{\pi}\frac{x^2}{1+x^4}, \qquad -\infty < x < \infty,$$

Kemudian itu dapat ditunjukkan bahwa variabel acak adalah $\frac{X}{Y}$ variabel acak Cauchy. Laha (1959) dan Kotlarski (1960) menjelaskan secara lengkap tentang kelompok dari semua kemungkinan fungsi kepadatan f mengenai hasil bagi dari $\frac{X}{Y}$ mengikuti distribusi standar Cauchy kapanpun X dan Y adalah independen dan distribusi identik variabel acak dengan kepadatan gabungan f.

Contoh 2.12. Jika X mempunyai sebuah distribusi Poisson dengan rata-rata λ. Temukan sebuah transformasi $T(x)$ sehingga $Var(T(X))$ bebas untuk λ, untuk nilai yang besar dari λ.

Jawab :

Kita kembangkan fungsi $T(x)$ dengan deret Taylor tentang λ. Lalu, abaikan urutan tertinggi untuk nilai terbesar dari λ, kita peroleh

$$T(x) = T(\lambda) + (x-\lambda)T'(\lambda) + \ldots$$

Dimana $T'(\lambda)$ menunjukkan turunan dari $T(x)$ dari $x=\lambda$. Sekarang, kita hitung varians dari $T(X)$.

$$Var(T(X)) = Var(T(\lambda)) + (X-\lambda)T'(\lambda) + \ldots)$$

$$= Var(T(\lambda)) + Var((X-\lambda)T'(\lambda))$$

$$= 0 + \left[T'(\lambda)\right]^2 Var(X)$$

$$= \left[T'(\lambda)\right]^2 Var(X)$$

$$= \left[T'(\lambda)\right]^2 \lambda.$$

Kita buat $Var(T(X))$ untuk menjadi bebas dari λ. Oleh karena itu, kita dapat

$$\left[T'(\lambda)\right]^2 \lambda = k,$$

Dengan k adalah konstanta. Maka dari itu kita dapatkan

$$T'(\lambda) = \frac{c}{\sqrt{\lambda}}$$

Dimana $c = \sqrt{k}$. Kita selesaikan dengan persamaan diferensial, kita dapat

$$T(\lambda) = c \int \frac{1}{\sqrt{\lambda}} d\lambda$$

$$= 2c\sqrt{\lambda}.$$

Maka dari itu, transformasi $T(x) = 2c\sqrt{x}$ akan bebas $Var(T(X))$ dari λ jika variabel acak $X \sim POI(\lambda)$.

Contoh 2.13. Jika $X \sim POI(\lambda_1)$ dan $Y \sim POI(\lambda_1)$. Berapa fungsi kepadatan probabilitas dari $X + Y$ jika X dan Y adalah independen?

Jawab :

Akan kita tunjukkan $U = X + Y$ dan $V = X$. Pertama, kita temukan kepadatan gabungan dari U dan V lalu jumlah dari kepadatan gabungan kita tentukan marginal dari U dimana fungsi kepadatan dari $X + Y$ Sekarang, tulis X dan Y dalam U dan V, kita dapat

$$\left.\begin{aligned} X &= V \\ Y &= U - X = U - V \end{aligned}\right\}$$

Oleh karena itu, transformasi Jacobian diberikan oleh

$$J = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

$$= (0)(-1) - (1)(1)$$

$$= -1.$$

Fungsi densitas gabungan dari U dan V adalah

$$g(u,v) = |J|\, f\left(R(u,v), S(u,v)\right)$$

$$= |-1|\, f(v, u-v)$$

$$= f(v)\, f(u-v)$$

$$= \left(\frac{e^{-\lambda_1}\lambda_1^{v}}{v!}\right)\left(\frac{e^{-\lambda_2}\lambda_2^{u-v}}{(u-v)!}\right)$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{v}\lambda_2^{u-v}}{(v)!(u-v)!},$$

Dimana $v = 0,1,2,\ldots,u$ dan $u = 0,1,2,\ldots,\infty$. Oleh karena itu, kepadatan marginal dari U diberikan oleh

$$g_1(u) = \sum_{v=0}^{u} \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{v}\lambda_2^{u-v}}{(v)!(u-v)!}$$

$$= e^{-(\lambda_1+\lambda_2)} \sum_{v=0}^{u} \frac{\lambda_1^{v}\lambda_2^{u-v}}{(v)!(u-v)!}$$

$$= e^{-(\lambda_1+\lambda_2)} \sum_{v=0}^{u} \frac{1}{u!}\binom{u}{v}\lambda_1^{v}\lambda_2^{u-v}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{u!}(\lambda_1+\lambda_2)^u$$

Lalu, fungsi kepadatan dari $U = X + Y$ diberikan oleh

$$g_1 = \begin{cases} \dfrac{e^{-(\lambda_1+\lambda_2)}}{u!}(\lambda_1+\lambda_2)^u & \textit{untuk } u = 0,1,2,\ldots,\infty \\ 0 & \textit{untuk yang lain} \end{cases}$$

Contoh ini memberitahukan kita bahwa jika $X \sim POI(\lambda_1)$ dan $Y \sim POI(\lambda_2)$ dan masing-masing independen, lalu $X + Y \sim POI(\lambda_1 + \lambda_2)$.

Teorema 2.3.

Jika kepadatan gabungan dari variabel acak X dan Y menjadi $f(x, y)$.

Probabilitas fungsi kepadatan dari $X + Y, XY,$ dan $\frac{Y}{X}$ masing-masing diberikan oleh

$$h_{X+Y}(v) = \int_{-\infty}^{\infty} f(u, v-u)\,du$$

$$h_{XY}(v) = \int_{-\infty}^{\infty} \frac{1}{|u|} f\left(u, \frac{v}{u}\right) du$$

$$h_{\frac{X}{Y}}(v) = \int_{-\infty}^{\infty} |u|\, f(u, vu)\,du,$$

Bukti :

Jika $U = X$ dan $V = X + Y$. Sehingga $X = R(U,V) = U$, dan $Y = S(U,V) = V - U$. Oleh karena itu, transformasi Jacobian diberikan oleh

$$J = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u} = 1.$$

Fungsi kepadatan gabungan dari U dan V adalah

$$g(u,v) = |J|\, f\left(R(u,v), S(u,v)\right)$$

$$= f\left(R(u,v), S(u,v)\right)$$

$$= f(u, v-u)$$

Oleh karena itu, kepadatan marginal dari $V = X + Y$ diberikan oleh

$$h_{X+Y}(v) = \int_{-\infty}^{\infty} f(u, v-u)\,du$$

Dengan cara yang sama, dapat kita peroleh dua fungsi kepadatan lainnya.

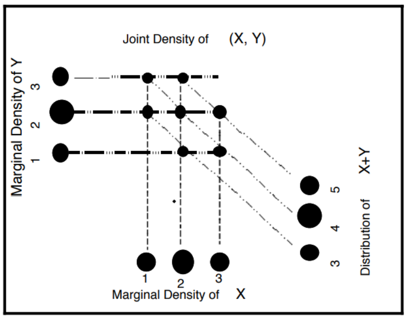
Selain itu, jika variabel acak X dan Y pada teorema 2.3 adalah independen dan memiliki probabilitas fungsi kepadatan $f(x)$ dan $g(y)$, maka kita miliki

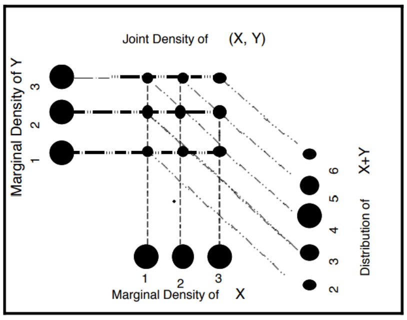
$$h_{X+Y}(z) = \int_{-\infty}^{\infty} g(y)\, f(z-y)\,dy$$

$$h_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|y|} g(y)\, f\left(\frac{z}{y}\right) dy$$

$$h_{\frac{X}{Y}}(z) = \int_{-\infty}^{\infty} |y|\, g(y)\, f(zy)\,dy.$$

Setiap angka berikut ini menunjukkan bagaimana distribusi variabel acak $X + Y$ didapatkan dari distribusi gabungan dari (X, Y).





Contoh 2.14. Lempar dadu sebanyak dua kali. Jika X menunjukkan hasil di lemparan pertama dan Y menunjukkan hasil di lemparan kedua, berapakah distribusi variabel acak $Z = \max\{X,Y\}$?

Jawab :

Ruang sampel X adalah $R_X = \{1,2,3,4,5,6\}$. Begitu pula dengan ruang sampel Y adalah $R_Y = \{1,2,3,4,5,6\}$. Oleh karena itu, ruang variabel acak (X,Y) adalah $R_X \times R_Y$. Tabel berikut menunjukkan distribusi (X,Y).

1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
	1	2	3	4	5	6

Ruang sampel variabel acak $Z = max\{X,Y\}$ adalah $R_Z = \{1,2,3,4,5,6\}$.

Jadi $Z = 1$ hanya jika $(X,Y) = (1,1)$. Oleh karena itu, $P(Z = 1) = \frac{1}{36}$.

Demikian pula, $Z = 2$ hanya jika $(X,Y) = (1,2),(2,2)$ atau $(2,1)$.

Karenanya, $P(Z = 2) = \frac{3}{36}$. Selanjutnya dengan cara yang sama, kita mendapatkan distribusi Z yang dirangkum dalam tabel di bawah ini.

z	1	2	3	4	5	6
$h(z)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

Dalam contoh ini, variabel acak Z dapat dideskripsikan sebagai yang terbaik dari dua lemparan. Perhatikan bahwa kepadatan probabilitas Z juga dapat dinyatakan sebagai

$$h(z) = \frac{2z-1}{36}, \quad untuk\ z \in \{1,2,3,4,5,6\}.$$

2.4. Metode Konvolusi untuk Jumlah Variabel Acak

Pada bagian ini, kita ilustrasikan bagaimana teknik konvolusi dapat digunakan untuk menemukan distribusi jumlah variabel acak ketika distribusinya independen. Teknik konvolusi ini tidak dapat digunakan jika variabel acak tidak independen.

Definisi 2.1.

Misalkan f dan g adalah dua fungsi bernilai real. Konvolusi dari f dan g, dilambangkan dengan $f \star g$, didefinisikan sebagai

$$(f \star g)(z) = \int_{-\infty}^{\infty} f(z-y)\,g(y)\,dy$$

$$= \int_{-\infty}^{\infty} g(z-x)\,f(x)\,dx.$$

Dari definisi tersebut terbukti bahwa $f \star g = g \star f$

Misalkan X dan Y adalah dua variabel acak independen dengan fungsi kepadatan probabilitas $f(x)$ dan $g(y)$. Sehingga dengan teorema 2.3 kita dapatkan

$$h(z) = \int_{-\infty}^{\infty} g(z-y)\,g(y)\,dy.$$

Oleh karena itu, hasil ini menunjukkan bahwa densitas variabel random $Z = X + Y$ adalah konvolusi dari kepadatan X dengan kepadatan dari Y.

Contoh 2.15. Berapa kepadatan probabilitas dari jumlah dua variabel acak independen, yang masing-masing terdistribusi secara uniform pada interval $[0,1]$?

Jawab :

Misalkan $Z = X + Y$, dengan $X \sim UNIF(0,1)$ dan $Y \sim UNIF(0,1)$. Oleh karena itu, fungsi kepadatan $f(x)$ dari variabel acak X diberikan oleh

$$f(x) = \begin{cases} 1 & untuk\, 0 \le x \le 1 \\ \\ 0 & untuk\ yang\ lain \end{cases}$$

Fungsi kepadatan $g(y)$ dari variabel acak Y diberikan oleh

$$f(x) = \begin{cases} 1 & untuk\, 0 \le x \le 1 \\ \\ 0 & untuk\ yang\ lain \end{cases}$$

Karena X dan Y independen, fungsi kepadatan dari Z dapat diperoleh dengan metode konvolusi Karena, jumlah $z = x + y$ adalah antara 0 dan 2, kita memperhitungkan dua kasus. Pertama, misalkan $0 \le z \le 1$, maka

$$h(z) = (f \star g)(z)$$

$$= \int_{-\infty}^{\infty} f(z-x)\, g(x)\, dx$$

$$= \int_{0}^{1} f(z-x)\, g(x)\, dx$$

$$= \int_{0}^{z} f(z-x)\, g(x)\, dx + \int_{z}^{1} f(z-x)\, g(x)\, dx$$

$$= \int_{0}^{z} f(z-x)\, g(x)\, dx + 0 \qquad \left(Karena\ f(z-x) = 0\, \text{antara}\ z\, \text{dan}\, 1\right)$$

$$= \int_{0}^{z} dx$$

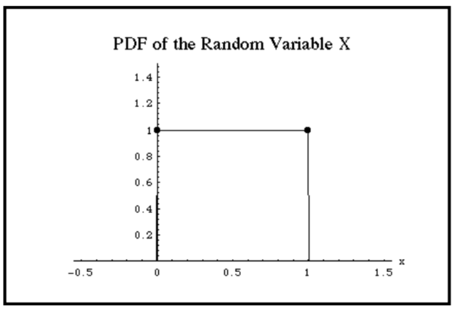
$$= z.$$

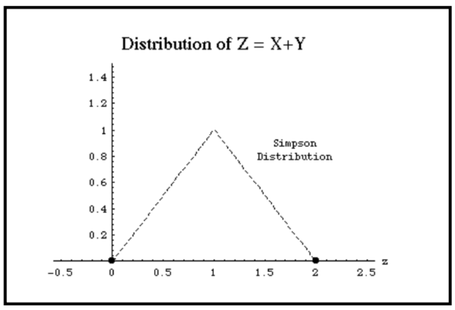
Dengan cara yang sama, jika $1 \le z \le 2$, maka

$$
\begin{aligned}
h(z) &= (f \star g)(z) \\
&= \int_{-\infty}^{\infty} f(z-x)\, g(x)\, dx \\
&= \int_{0}^{1} f(z-x)\, g(x)\, dx \\
&= \int_{0}^{z-1} f(z-x)\, g(x)\, dx + \int_{z-1}^{1} f(z-x)\, g(x)\, dx \\
&= 0 + \int_{z-1}^{1} f(z-x)\, g(x)\, dx \qquad \left(karena\ f(z-x) = 0 \text{ antara } 0 \text{ dan } z-1\right) \\
&= \int_{z-1}^{1} dx \\
&= 2 - z.
\end{aligned}
$$

Jadi, fungsi densitas $Z = X + Y$ diberikan oleh

$$
h(z) = \begin{cases} 0 & untuk\ -\infty < z \le 0 \\ z & untuk\ 0 \le z \le 1 \\ 2-z & untuk\ 1 \le z \le 2 \\ 0 & untuk\ 2 < z < \infty \end{cases}
$$





Grafik fungsi densitas ini tampak seperti tenda dan disebut fungsi tenda. Namun dalam literatur, fungsi kepadatan ini dikenal dengan sebutan distribusi Simpson.

Contoh 2.16. Berapakah kepadatan probabilitas dari penjumlahan dua variabel random independen yang masing-masing adalah gamma dengan parameter $\alpha = 1$ dan $\theta = 1$?

Jawab :

Misalkan $Z = X + Y$, dimana $X \sim GAM(1,1)$ dan $X \sim GAM(1,1)$ Oleh karena itu, fungsi kepadatan $f(x)$ dari variabel acak X diberikan oleh

$$f(x) = \begin{cases} e^{-x} & \mathit{untuk}\ 0 < x < \infty \\ 0 & \mathit{untuk\ yang\ lainnya} \end{cases}$$

Demikian pula, fungsi kepadatan $g(y)$ dari Y diberikan oleh

$$f(x) = \begin{cases} e^{-y} & \mathit{untuk}\ 0 < y < \infty \\ 0 & \mathit{untuk\ yang\ lainnya} \end{cases}$$

Karena X dan Y adalah independen, fungsi kepadatan Z dapat diperoleh dengan metode konvolusi. Perhatikan bahwa jumlah $z = x + y$ berada di antara 0 dan ∞, dan $0 < x < z$. Oleh karena itu, fungsi kepadatan Z diberikan oleh

$$h(z) = (f \star g)(z)$$

$$= \int_{-\infty}^{\infty} f(z-x)\, g(x)\, dx$$

$$= \int_{0}^{\infty} f(z-x)\, g(x)\, dx$$

$$= \int_{0}^{z} e^{-(z-x)} e^{-x}\, dx$$

$$= \int_{0}^{z} e^{-z+x} e^{-x}\, dx$$

$$= \int_{0}^{z} e^{-z}\, dx$$

$$= ze^{-z}$$

$$= \frac{1}{\Gamma(2)1^2} z^{2-1} e^{-\frac{z}{1}}$$

Karenanya $Z \sim GAM(1,2)$. Jadi, jika $X \sim GAM(1,1)$ dan $Y \sim GAM(1,1)$, lalu $X+Y \sim GAM(1,2), X+Y$ adalah gamma dengan $\alpha=2$ dan $\theta=1$. Ingat bahwa variabel acak gamma dengan $\alpha=1$ disebut sebagai variabel acak eksponensial dengan parameter θ. Jadi, dengan melihat contoh di atas, kita melihat bahwa jumlah dua variabel acak eksponensial independen belum tentu merupakan variabel eksponensial.

Contoh 2.17. Berapa kepadatan probabilitas dari jumlah dua variabel acak independen, yang masing-masing adalah normal standar?

Jawab :

Misalkan $Z = X + Y$, di mana $X \sim N(0,1)$ dan $Y \sim N(0,1)$. Oleh karena itu, fungsi kepadatan $f(x)$ dari variabel acak X diberikan oleh

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Demikian pula, fungsi kepadatan $g(y)$ dari Y diberikan oleh

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

Karena X dan Y adalah independen, fungsi kepadatan Z dapat diperoleh dengan metode konvolusi. Perhatikan bahwa jumlah $z = x + y$ berada di antara $-\infty$ dan ∞. Oleh karena itu, fungsi kepadatan Z diberikan oleh

$$h(z) = (f \star g)(z)$$

$$= \int_{-\infty}^{\infty} f(z-x) g(x) dx$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\frac{-(z-x)}{2}} e^{\frac{x^2}{2}} dx$$

$$= \frac{1}{2\pi} e^{\frac{z^2}{4}} \int_{-\infty}^{\infty} e^{-\left(x-\frac{z}{2}\right)^2} dx$$

$$= \frac{1}{2\pi} \mathrm{e}^{-\frac{z^2}{4}} \sqrt{\pi} \left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-\left(x-\frac{z}{2}\right)^2} dx \right]$$

$$= \frac{1}{2\pi} \mathrm{e}^{-\frac{z^2}{4}} \left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-w^2} dw \right], \quad dimana\, w = x - \frac{z}{2}$$

$$= \frac{1}{\sqrt{4\pi}} e^{-\frac{z^2}{4}}$$

$$= \frac{1}{\sqrt{4\pi}} e^{-\frac{1}{2}\left(\frac{z-0}{\sqrt{2}}\right)^2}.$$

Integral dalam tanda kurung sama dengan satu, karena integrand adalah fungsi kepadatan normal dengan mean $\mu = 0$ dan varians $\sigma^2 = \frac{1}{2}$. Oleh karena itu, jumlah dua variabel acak normal standar sekali lagi merupakan variabel acak normal dengan mean 0 dan varians 2.

Contoh 2.18. Berapa kepadatan probabilitas dari jumlah dua variabel acak independen, yang masing-masing adalah Cauchy?

Jawab :

Misalkan $Z = X + Y$, di mana $X \sim N(0,1)$ dan $Y \sim N(0,1)$. Oleh karena itu, fungsi kepadatan $f(x)$ dari variabel acak X dan Y masing-masing diberikan oleh

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{dan} \quad g(y) = \frac{1}{\pi(1+y^2)},$$

Karena X dan Y adalah independen, fungsi kepadatan Z dapat diperoleh dengan metode konvolusi. Perhatikan bahwa jumlah $z = x + y$ berada di antara $-\infty$ dan ∞. Oleh karena itu, fungsi kepadatan Z diberikan oleh

$$h(z) = (f \star g)(z)$$

$$=\int_{-\infty}^{\infty} f\left(z-x\right)g\left(x\right)dx$$

$$=\int_{-\infty}^{\infty}\frac{1}{\pi\left(1+\left(z-x\right)^2\right)}\frac{1}{\pi\left(1+x^2\right)}dx$$

$$=\frac{1}{\pi^2}\int_{-\infty}^{\infty}\frac{1}{\left(1+\left(z-x\right)^2\right)}\frac{1}{1+x^2}dx$$

Untuk mengintegrasikan integral di atas, kita dekomposisi integral tersebut menggunakan dekomposisi pecahan parsial. Karenanya

$$\frac{1}{1+\left(z-x\right)^2}\ \frac{1}{1+x^2}=\frac{2Ax+B}{1+x^2}+\frac{2C\left(z-x\right)+D}{1+\left(z-x\right)^2}$$

Dimana

$$A=\frac{1}{z\left(4-z\right)^2}=C \quad dan \quad B=\frac{1}{4+z^2}=D$$

Sekarang hasil integrasi

$$\frac{1}{\pi^2}\int_{-\infty}^{\infty}\frac{1}{\left(1+\left(z-x\right)^2\right)}\frac{1}{1+x^2}dx$$

$$=\frac{1}{\pi^2 z^2\left(4+z^2\right)}\left[z\ln\left(\frac{1+x^2}{1+\left(z-x\right)^2}\right)+\ z^2\ \tan^{-1}x-z^2\tan^{-1}\left(z-x\right)\right]_{-\infty}^{\infty}$$

$$=\frac{1}{\pi^2 z^2\left(4+z^2\right)}\left[0+z^2\pi+z^2\pi\right]$$

$$=\frac{2}{\pi\left(4+z^2\right)}$$

Karenanya jumlah dari dua variabel acak Cauchy independen bukanlah variabel acak Cauchy.

Jika $X \sim CAU(0)$ dan $Y \sim CAU(0)$, maka dapat dengan mudah ditunjukkan dengan menggunakan Contoh 2.18 bahwa variabel acak $Z = \dfrac{X+Y}{2}$ sekali lagi adalah Cauchy, yaitu $Z \sim CAU(0)$.

Misalkan X dan Y adalah dua variabel acak diskrit yang keduanya mengambil nilai yang merupakan bilangan bulat. Misalkan $Z = X + Y$ adalah jumlah dari dua variabel acak. Oleh karena itu Z mengambil nilai pada himpunan bilangan bulat. Misalkan $X = n$ di mana n adalah bilangan bulat. Maka $Z = z$ jika dan hanya jika $Y = z - n$. Jadi kejadian $(Z = z)$ adalah gabungan dari pasangan kejadian disjoint $(X = n)$ dan $(Y = z - n)$ dimana n adalah bilangan bulat. Fungsi distribusi kumulatif $H(z)$ dari Z dapat diperoleh sebagai berikut:

$$P = (Z = z) = \sum_{n=-\infty}^{\infty} P(X = n) P = (Y = z - n)$$

Yang mana

$$h(z) = \sum_{n=-\infty}^{\infty} f(n) g(z-n)$$

di mana $F(x)$ dan $G(y)$ masing-masing adalah fungsi distribusi kumulatif dari X dan Y.

Definisi 2.2.

Misalkan X dan Y adalah dua variabel acak diskrit independen bernilai bilangan bulat, dengan fungsi kepadatan probabilitas masing-masing $f(x)$ dan $g(y)$. Maka konvolusi $f(x)$ dan $g(y)$ adalah fungsi distribusi kumulatif $h = f \star g$ diberikan oleh

$$h(m) = \sum_{n=-\infty}^{\infty} f(n) g(m-n)$$

untuk $m = -\infty, \ldots, -2, -1, 0, 1, 2, \ldots . \infty$. Fungsi $h(z)$ adalah fungsi distribusi probabilitas dari variabel acak diskrit $Z = X + Y$.

Contoh 2.19. Misalkan masing-masing variabel acak X dan Y mewakili hasil dari dadu bersisi enam. Berapakah fungsi kepadatan kumulatif dari jumlah X dan Y?

Jawab :

Karena range dari X dan Y adalah $\{1,2,3,4,5,6\}$, maka range $Z = X + Y$ adalah $R_Z = \{2,3,4, ...,11,12\}$. Fungsi kepadatan probabilitas dari Z diberikan oleh

$$h(2)=f(1)g(1)=\frac{1}{6}.\frac{1}{6}=\frac{1}{36}$$

$$h(3)=f(1)g(2)+f(2)g(1)=\frac{1}{6}.\frac{1}{6}+\frac{1}{6}.\frac{1}{6}=\frac{2}{36}$$

$$h(4)=f(1)g(3)+h(2)g(2)+f(3)g(1)=\frac{1}{6}.\frac{1}{6}+\frac{1}{6}.\frac{1}{6}+\frac{1}{6}.\frac{1}{6}=\frac{3}{36}$$

Melanjutkan dengan cara ini kita memperoleh

$$h(5)=\frac{4}{36},\ h(6)=\frac{5}{36},\ h(7)=\frac{6}{36},\ h(8)=\frac{5}{36},$$
$$h(9)=\frac{4}{36},\ h(10)=\frac{3}{36},\ h(11)=\frac{2}{36},\ h(12)=\frac{1}{36}$$

Letakkan ini ke persamaan yang kita miliki

$$h(z)=\sum_{n=1}^{z-1} f(n)g(z-n)$$

$$=\frac{6-|z-7|}{36},\ z=2,3,4, ...,12$$

Mudah untuk dicatat bahwa operasi konvolusi bersifat komutatif dan asosiatif. Menggunakan asosiativitas operasi konvolusi, kita dapat menghitung fungsi distribusi kumulatif dari variabel acak $S_n = X_1 + X_2 + \ldots + X_n$, di mana $X_1 + X_2 + \ldots + X_n$ adalah variabel acak yang masing-masing memiliki fungsi distribusi kumulatif $f(x)$ yang sama. Maka fungsi distribusi kumulatif dari S_1 adalah $f(x)$. Karena $S_n = S_{n-1} + X_n$ dan fungsi distribusi kumulatif dari X_n adalah $f(x)$, fungsi distribusi kumulatif S_n dapat diperoleh dengan induksi.

2.5. Metode Fungsi Pembangkit Momen

Kita tahu bahwa jika X dan Y adalah variabel acak independen, maka

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

Hasil ini dapat digunakan untuk menemukan jumlah $X+Y$. Seperti metode konvolusi, metode ini dapat digunakan untuk menemukan distribusi $X+Y$ jika X dan Y adalah variabel acak independen. Secara singkat diilustrasikan metode menggunakan contoh berikut.

Contoh 2.20. Misalkan $X \sim POI(\lambda_1)$ dan $Y \sim POI(\lambda_2)$. Berapakah fungsi kepadatan probabilitas dari $X+Y$ jika X dan Y independen?

Jawab :

Karena, $X \sim POI(\lambda_1)$ dan $Y \sim POI(\lambda_2)$, kita dapatkan

$$M_X(t) = e^{\lambda_1(e^t - 1)}$$

Dan

$$M_Y(t) = e^{\lambda_2(e^t - 1)}$$

Selanjutnya, karena X dan Y adalah independen, kita punya

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

$$= e^{\lambda_1(e^t-1)} e^{\lambda_2(e^t-1)}$$

$$= e^{\lambda_1(e^t-1)} + e^{\lambda_2(e^t-1)}$$

$$= e^{(\lambda_1+\lambda_2)(e^t-1)}$$

yaitu, $X+Y \sim POI(\lambda_1+\lambda_2)$. Oleh karena itu fungsi kepadatan $h(z)$ dari $Z = X+Y$ diberikan oleh

$$h(z)=\begin{cases}\dfrac{e^{-(\lambda_1+\lambda_2)}}{z!}(\lambda_1+\lambda_2)^z & \textit{untuk } z=0,1,2,3\ldots\\ \\ 0 & \textit{untuk yang lain}\end{cases}$$

Bandingkan contoh ini dengan Contoh 2.13. akan terlihat bahwa metode momen memiliki keuntungan yang besar dibandingkan metode konvolusi. Namun, jika menggunakan metode momen pada Contoh 2.15, maka akan kesulitan mengidentifikasi bentuk fungsi kepadatan variabel acak $X+Y$. Dengan demikian, sulit untuk mengatakan metode mana yang selalu berhasil.

Contoh 2.21. Berapakah fungsi kepadatan probabilitas dari penjumlahan dua variabel acak independen yang masing-masing adalah gamma dengan parameter θ dan α?

Jawab :

Misalkan X dan Y adalah dua variabel acak gamma independen dengan parameter θ dan α, yaitu $X \sim GAM(\theta,\alpha)$ dan $Y \sim GAM(\theta,\alpha)$. Dari Teorema 6.3 (Pengantar Statistika Matematika 1), fungsi pembangkit momen X dan Y diperoleh masing-masing sebagai $M_X(t)=(1-\theta)^{-\alpha}$ dan $M_Y(t)=(1-\theta)^{-\alpha}$. Karena, X dan Y adalah independen.

$$M_{X+Y}(t)=M_X(t)M_Y(t)$$

$$=(1-\theta)^{-\alpha}(1-\theta)^{-\alpha}$$

$$=(1-\theta)^{-2\alpha}$$

Dengan demikian $X+Y$ memiliki fungsi pembangkit momen dari variabel acak gamma dengan parameter θ dan 2α. Karena itu

$$X+Y \sim GAM(\theta,2\alpha).$$

LATIHAN BAB 2

1. Misalkan X adalah variabel acak dengan fungsi kepadatan

$$f(x)=\begin{cases}\frac{3}{8}x^2 & untuk\, 0<x<2 \\ \\ 0 & untuk\ yang\, lainnya\end{cases}$$

Misalkan $Y=mX^2$, di mana m adalah bilangan positif tetap. Berapakah fungsi kepadatan Y di mana bukan nol?

Jawab :

$$g(y)=f\left(W(y)\right)\left|\frac{dx}{dy}\right|, \ Teorema\, 2.1$$

Transformasi antara x dan y adalah satu-satu

$$Y=mX^2$$

$$\frac{y}{m}=x^2$$

$$\sqrt{\frac{y}{m}}=x$$

$$\left|\frac{dx}{dy}\right|=\left|\frac{\sqrt{m}}{2m\sqrt{y}}\right|$$

$$\left|\frac{dx}{dy}\right|=\frac{\sqrt{m}}{2m\sqrt{y}}$$

$$=\frac{1}{8}\times\left(\frac{y}{m}\right)\frac{3}{2}$$

$$\therefore f_y\left(y\right)=\frac{3}{2}\times\frac{1}{8}\times\left(\frac{y}{m}\right)^{\frac{3}{2}-1}\times\frac{1}{m}$$

$$=\frac{3}{16m}(\frac{y}{m})^{\frac{1}{2}}$$

$$=\frac{3}{16m}\sqrt{\frac{y}{m}} \qquad y;\left(0,\quad 4m\right)$$

Fungsi distribusi kumulatif dari Y

$$f_y\left(y\right)=\begin{cases}\frac{3}{16m}\sqrt{\frac{y}{m}} & untuk\, 0\le y\le 4m\\ \\ 0 & untuk\ yang\, lainnya\end{cases}$$

2. Misalkan X adalah variabel acak kontinu dengan fungsi kepadatan

$$f\left(x\right)=\begin{cases}2\mathrm{e}^{-2x} & untuk\, x>0\\ \\ 0 & untuk\ yang\, lain\end{cases}$$

dan misalkan $Y=e^{-X}$. Berapakah fungsi kepadatan $g\left(y\right)$ dari Y dimana bukan nol?

Jawab:

Langkah 1 = misalkan $Y=e^{-x}$

$$X=-\log X$$

Langkah 2 = saat $x>0$ => $y>0$

Langkah 3 = jacobian J adalah $J=\dfrac{dx}{dy}=\dfrac{-1}{y}$

Langkah 4 = PDF dari Y adalah

$$g(y) = f(x)|y|$$

$$= 2e^{-2x} \times \frac{1}{y}$$

$$= 2e^{-2(-logy)} \times \frac{1}{y}$$

$$= 2e^{-2x} \times \frac{1}{y}$$

$$= \frac{2y^2}{y}$$

$$= 2y$$

$$g(y) = \begin{cases} 2y & untuk\, 0 \le y \le 1 \\ 0 & untuk\ yang\ lainnya \end{cases}$$

3. Misalkan X dan Y adalah variabel acak dengan fungsi kepadatan gabungan

$$f(x,y) = \begin{cases} e^{-x} & untuk\, 0 < x < \infty,\ 0 < y < 1 \\ 0 & untuk\ yang\ lain \end{cases}$$

Jika $Z = X + 2Y$, lalu berapa kepadatan gabungan X dan Z di mana bukan nol?

Jawab:

$$f(x,y) = \begin{cases} e^{-x} & untuk\, 0 < x < \infty,\ 0 < y < 1 \\ 0 & untuk\ yang\ lain \end{cases}$$

$$Z = X + 2Y$$

Jika $U = X$

Maka, transformasi (x, y) ke (U, Z)

Kita mempunyai, $X = U$ dan $Y = \dfrac{Z - U}{2}$

$$J\frac{(x,y)}{(U,Z)} = \begin{vmatrix} \frac{dx}{du} & \frac{dy}{du} \\ \frac{dx}{dz} & \frac{dy}{dz} \end{vmatrix} = \begin{vmatrix} 1 & -\frac{1}{2} \\ 0 & \frac{1}{2} \end{vmatrix}$$

$$|J| = \frac{1}{2}$$

Untuk $g(Z, U) = f(x, y).|J|$

$$= e^{-U}.\frac{1}{2}$$

Untuk

$$0 < U < \infty \qquad \{\therefore U = x)$$

$$0 < Z - x < 2 \qquad \{\therefore Z - x = 2y\}$$

$$x < Z < 2 + x$$

$$\therefore g(Z, x) = \begin{cases} \frac{1}{2}e^{-x} & 0 < x < Z < 2 + x < \infty \\ 0 & untuk\ yang\ lainnya \end{cases}$$

4. Misalkan X adalah variabel acak kontinu dengan fungsi kepadatan

$$f(x) = \begin{cases} \dfrac{2}{x^2} & untuk\ 1 < x < 2 \\ 0 & untuk\ yang\ lain \end{cases}$$

Jika $Y = \sqrt{X}$, maka berapakah fungsi kepadatan Y untuk $1 < y < \sqrt{2}$?

Jawab :

Fungsi distribusi kumulatif dari $y = f\left(y\right)$

$$= P\left(Y < y\right)$$

$$= P\left(\sqrt{x} < y\right)$$

$$= P\left(X < y^2\right)$$

$$= \int_1^{y^2} f\left(x\right) dx$$

$$= \int_1^{y^2} \left(\frac{2}{x^2}\right) dx$$

$$= \left[-\frac{2}{x^2}\right]_1^{y^2}$$

$$= 2\left(1 - \frac{1}{y^2}\right)$$

oleh karena itu fungsi kepadatan probabilitas dari $y : f\left(y\right)\left(\frac{d}{dy}\right)\left(2\left(1 - \frac{1}{y^2}\right)\right)$

$$f\left(y\right) = \frac{4}{y^3}$$

$$f\left(y\right) = \begin{cases} \dfrac{4}{y^3} & untuk\, 0 < y < \sqrt{2} \\ 0 & untuk\ yang\, lainnya \end{cases}$$

5. Lemparkan dadu yang tidak bias 3 kali. Jika U menunjukkan hasil di lemparan pertama, V menunjukkan hasil di lemparan kedua, dan W menunjukkan hasil dari lemparan ketiga, Berapa distribusi variabel acak $Z = max\{U, V, W\}$?

   Jawab :

   Dibawah ini adalah distribusi probabilitas Z

$$P(Z=1) = P(semuanya\ ada\ 1) = \left(\frac{1}{6}\right)\left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{1}{216}$$

$$P(Z=2) = P(semuanya\ paling\ banyak\ 2) - P(Z=1) = \left(\frac{2}{6}\right)\left(\frac{2}{6}\right)\left(\frac{2}{6}\right) - \frac{1}{216} = \frac{7}{216}$$

$$P(Z=3) = \left(\frac{3}{6}\right) - \left(\frac{3}{6}\right) - \left(\frac{3}{6}\right) - \frac{1}{216} - \frac{7}{216} = \frac{19}{216}$$

$$P(Z=4) = \frac{37}{216}$$

$$P(Z=5) = \frac{61}{216}$$

$$P(Z=6) = \frac{91}{216}$$

$$atau\ P(Z=z) = \frac{3z^2 - 2z + 1}{216}, \qquad z = 1, 2, 3, 4, 5, 6.$$

6. Kepadatan probabilitas V, kecepatan molekul gas, menurut hukum Maxwell-Boltzmann diberikan oleh

$$f(v) = \begin{cases} \frac{4\mathrm{h}^3}{\sqrt{\pi}} \mathrm{v}^2 \mathrm{e}^{-\mathrm{h}^2 v^2} & untuk\ 0 \le v < \infty \\ 0 & untuk\ yang\ lain \end{cases}$$

   dengan h adalah konstanta Plank. Jika m mewakili massa molekul gas, maka berapakah densitas probabilitas energi kinetik $Z = \frac{1}{2} mV^2$?

Jawab :

Misalkan V adalah sebagai kecepatan dari molekul gas.

Probabilitas kepadatan dari $V, f_V(v) = \frac{4h^3}{\sqrt{\pi}} v^2 e^{-h^2v^2}$; $v \in R$

$$Z = \frac{1}{2} mV^2$$

$$v = \sqrt{\frac{2z}{m}}$$

$$v \in R \Rightarrow z \in R^+$$

$$dz = \frac{1}{2} m \times 2v\, dv$$

Jacobian dari transformasi adalah $\left|\frac{dv}{dz}\right| = \frac{1}{mv} = \frac{1}{m} \times \sqrt{\frac{m}{2z}}$

Probabilitas kepadatan dari Z adalah $f_Z(z)$

$$= \frac{4h^3}{\sqrt{\pi}} \frac{2z}{m} e^{-\frac{2h^2z}{m}} \times \frac{1}{m} \times \sqrt{\frac{m}{2z}}$$

$$= \frac{4h^3}{m\sqrt{\pi}} \sqrt{\frac{2z}{m}} e^{-\frac{2h^2z}{m}}$$

7. Jika variabel acak X dan Y memiliki kepadatan gabungan

$$f(x,y)=\begin{cases}\frac{6}{7}x & untuk\ 1\le x+y\le 2, x\ge 0, y\ge 0\\ 0 & untuk\ yang\ lain\end{cases}$$

berapa kepadatan gabungan $U=2X+3Y$ dan $V=4X+Y$?

Jawab :

Diberikan fungsi kepadatan gabungan

$$f(x)=\begin{cases}\frac{6}{7}x & untuk\ 1\le x+y\le 2, x\ge 0, y\ge 0\\ 0 & untuk\ yang\ lainnya\end{cases}$$

Sekarang

$$U=2x+3y \qquad \ldots.1$$

$$V=4x+y \qquad \ldots.2$$

dengan menguraikan 1 dan 2

$$1 \qquad \rightarrow U=2x+3y$$

$$2x3 \qquad \rightarrow 3V=2x+3y$$

$$\ldots=\ldots \quad \ldots$$

$$U-3V=-10x$$

$$x=\frac{3V-U}{10}$$

$$\rightarrow U=V-4x=V-4\left(\frac{3V-U}{10}\right)=\frac{10V-12V+4U}{10}$$

$$y=\frac{2U-V}{10}$$

Baru

$$J = \begin{vmatrix} \dfrac{dx}{du} & \dfrac{dx}{dv} \\ \dfrac{dy}{du} & \dfrac{dy}{dv} \end{vmatrix} . \begin{vmatrix} -\dfrac{1}{10} & \dfrac{3}{10} \\ \dfrac{2}{5} & -\dfrac{1}{5} \end{vmatrix} = \left(-\frac{1}{10}\right)\left(-\frac{1}{5}\right) - \left(\frac{3}{10}\right)\left(\frac{2}{5}\right)$$

$$= \frac{1}{50} - \frac{6}{50} = -\frac{5}{50} = -\frac{1}{10}$$

Menghitung dengan kisaran U dan V

$$1 \le x + y \le 2, \quad x \ge 0,\ y \ge 0$$

$$x = 1,\ y = 1 => x + y = 2$$

$$U = 2x + 3y = y \ge 0\ ;\ V = 4x + 3y = 5 \ge 0$$

$$U \ge 0, V \ge 0$$

$$1 \le x + y \le 2$$

$$=> \quad 1 \le \frac{3V - U}{10} + \frac{2U - V}{10} \le 2$$

$$=> \quad 1 \le \frac{3U + V}{10} \le 2 => \quad 1 \le 3U + V \le 20\ ; U \ge 0\ ;\ V \ge 0$$

Sekarang mari kita cari kepadatan gabungan dari (U,V)

$$h(U,V) = f(x,y)\ .\ |\mathrm{J}|$$

$$= \frac{6}{7}x \times \left|-\frac{1}{10}\right| = \frac{6}{70}x$$

$$= \frac{6}{70} \times \frac{3V - U}{10} = \frac{3}{350}(3V - U)$$

Kepadatan gabungan dari (U, V) adalah

$$h(U,V) = \begin{cases} \frac{3}{350}(3V - U) & untuk\, 1 \le 3U + V \le 20\ ; U \ge 0\ ; V \ge 0 \\ \\ 0 & untuk\ yang\ lainnya \end{cases}$$

8. Misalkan X dan Y memiliki fungsi kepadatan probabilitas gabungan

$$f(x,y) = \begin{cases} \frac{5}{16} xy^2 & untuk\, 0 < x < y < 2 \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Berapakah fungsi kepadatan gabungan dari $U = 3X - 2Y$ dan $V = X + 2Y$ di mana bukan nol?

Jawab :

Diberikan

$$f(x,y) = \begin{cases} \frac{5}{16} xy^2 & untuk\, 0 < x < y < 2 \\ \\ 0 & untuk\ yang\ lain \end{cases}$$

$$u = 3x - 2y, v = x + 2y$$

$$4x = u + v, x = \frac{u+v}{4}, y = \frac{3v - u}{8}$$

$$J = \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dv} \\ \frac{dy}{du} & \frac{dy}{dv} \end{vmatrix} = \begin{vmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{8} & \frac{3}{8} \end{vmatrix}$$

$$= \frac{3}{32} + \frac{1}{32} = \frac{1}{8}$$

$$J = \frac{1}{8}$$

$$0 < y < y < 2$$

$$0 < \frac{u+v}{4} < \frac{3v-u}{8} < 2$$

$$0 < 2u < 2v < 3v - u < 16$$

Karenanya fungsi kepadatan probabilitas dari $(u,v).15, g(u,v) = f(x,y)$ .15

$$g(u,v) = \begin{cases} \frac{5}{16}\left(\frac{u+v}{4}\right)\left(\frac{3v-u}{8}\right)^2\left(\frac{1}{8}\right) & untuk\ 0 < 2u < 2v < 3v - u < 16 \\ 0 & untuk\ yang\ lainnya \end{cases}$$

$$g(u,v) = \begin{cases} \frac{5\left(9v^3 - 5u^2 + 3uv^2 + u^3\right)}{32768} & untuk\ 0 < 2u < 2v < 3v - u < 16 \\ 0 & untuk\ yang\ lainnya \end{cases}$$

9. Misalkan X dan Y memiliki fungsi kepadatan probabilitas gabungan

$$f(x,y) = \begin{cases} 4x & untuk\ 0 < x < \sqrt{y} < 1 \\ 0 & untuk\ yang\ lain \end{cases}$$

Berapakah fungsi kepadatan gabungan dari $U = 5X - 2Y$ dan $V = 3X + 2Y$ di mana bukan nol?

Jawab :

$$f(x,y)=\begin{cases} 4x & 0<x<\sqrt{y}<1 \\ 0 & \textit{untuk yang lainnya} \end{cases}$$

$$U=5x-2y$$

$$V=3x+2y$$

$$U+V=8x$$

$$x=\frac{U+V}{8}$$

$$V=3\left(\frac{U+V}{8}\right)+2y$$

$$V=\frac{3U+3V}{8}+2y$$

$$V=\frac{3U+3V+16y}{8}$$

$$8V=3U+3V+16y$$

$$3U+3V+8V=16y$$

$$5V-3U=16y$$

$$y=\frac{5V-3U}{16}$$

Untuk Jacobian $|J|=\begin{vmatrix} \frac{1}{8} & \frac{1}{8} \\ -\frac{3}{16} & \frac{5}{16} \end{vmatrix}=\frac{5}{128}+\frac{3}{128}=\frac{8}{128}$

$$|J| = \frac{8}{128} \text{ dengan } 0 < x < \sqrt{y} < 1$$

$$\Rightarrow 0 < \frac{U+V}{8} < \sqrt{\frac{5V-3U}{16}} < 1$$

$$\Rightarrow 0 < \frac{U+V}{8} < \frac{1}{4}\sqrt{5V-3U} < 1$$

$$\Rightarrow 0 < U+V < 2\sqrt{5V-3U} < 8$$

Karena fungsi kepadatan gabungan dari (U,V) adalah $g(U,V) = f(x,y).|J|$

Maka

$$g(U,V) = \begin{cases} 4\left(\frac{U+V}{8}\right).\frac{8}{128} & \textit{untuk } 0 < x < \sqrt{y} < 1 \\ 0 & \textit{untuk yang lainnya} \end{cases}$$

$$g(U,V) = \begin{cases} \left(\frac{U+V}{32}\right) & \textit{untuk } 0 < U+V < 2\sqrt{5V-3U} < 8 \\ 0 & \textit{untuk yang lainnya} \end{cases}$$

BAB III

DISTRIBUSI DISKRIT BIVARIAT

Dalam bab ini, kita akan mempelajari beberapa fungsi kepadatan probabilitas diskrit bivariat. Bab ini akan berfokus pada beberapa distribusi diskrit bivariat terkenal yang distribusi marginalnya merupakan distribusi univariat yang terkenal.

3.1. Distribusi Bernoulli Bivariat

Kita definisikan variabel acak Bernoulli bivariat dengan menentukan bentuk distribusi probabilitas gabungan.

Definisi 3.1.

Variabel acak diskrit bivariat (X,Y) dikatakan memiliki distribusi Bernoulli bivariat jika kepadatan probabilitas gabungannya berbentuk

$$f(x,y)=\begin{cases} \dfrac{1}{x!\,y!(1-x-y)}\,p_1^x\,p_2^y(1-p_1-p_2)^{1-x-y}, & \textit{jika } x,y=0,1 \\ 0 & \textit{untuk yang lainnya} \end{cases}$$

Di mana $0<p_1,\ p_2\ ,p_1+p_2<1$ dan $x+y\le 1$. Kita menunjukkan variabel acak Bernoulli bivariat dengan menuliskan $(X,Y)\sim BER(p_1,\ p_2)$

Dalam teorema berikut, disajikan nilai yang diharapkan dan varians X dan Y, kovarian antara X dan Y, dan fungsi pembangkit momen gabungannya. Ingatlah bahwa fungsi pembangkit momen gabungan dari X dan Y didefinisikan sebagai $M(s,t):=E\left(e^{sX+tY}\right)$.

Teorema 3.1.

Jika $(X,Y)\sim BER(p_1,p_2)$, di mana p_1 dan p_2 adalah parameter. Maka

$$E(X)=p_1$$

$$E(Y)=p_2$$

$$Var(X)=p_1(1-p_1)$$

$$Var(Y)=p_2(1-p_2)$$

$$Cov(X,Y) = -p_1 p_2$$

$$M(s,t) = 1 - p_1 - p_2 + p_1 e^s + p_2 e^t$$

Bukti :

Pertama, kita peroleh fungsi penghasil momen gabungan dari X dan Y, kemudian menetapkan hasil darinya. Fungsi pembangkit momen gabungan X dan Y diberikan oleh

$$M(s,t) = E\left(e^{sX+tY}\right)$$

$$= \sum_{x=0}^{1}\sum_{y=0}^{1} f(x,y)\, e^{sx+ty}$$

$$= f(0,0) + f(1,0)e^s + f(0,1)e^t + f(1,1)e^{t+s}$$

$$= 1 - p_1 - p_2 + p_1 e^s + p_2 e^t + 0e^{t+s}$$

$$= 1 - p_1 - p_2 + p_1 e^s + p_2 e^t$$

Nilai ekspektasi dari X diberikan oleh

$$E(X) = \left.\frac{\partial M}{\partial s}\right|_{(0,0)}$$

$$= \left.\frac{\partial}{\partial s}\left(1 - p_1 - p_2 + p_1 e^s + p_2 e^t\right)\right|_{(0,0)}$$

$$= \left. p_1 e^s \right|_{(0,0)}$$

$$= p_1$$

Demikian pula, nilai ekspektasi dari Y diberikan oleh

$$E(Y) = \left.\frac{\partial M}{\partial s}\right|_{(0,0)}$$

$$= \left.\frac{\partial}{\partial s}\left(1 - p_1 - p_2 + p_1 e^s + p_2 e^t\right)\right|_{(0,0)}$$

$$= \left. p_2 e^t \right|_{(0,0)}$$

$$= p_2$$

Hasil kali momen dari X dan Y adalah

$$E(XY) = \left.\frac{\partial^2 M}{\partial t \partial s}\right|_{(0,0)}$$

$$= \left.\frac{\partial^2}{\partial t \partial s}\left(1 - p_1 - p_2 + p_1 e^s + p_2 e^t\right)\right|_{(0,0)}$$

$$= \left.\frac{\partial}{\partial t}\left(p_1 e^s\right)\right|_{(0,0)}$$

$$= 0$$

Oleh karena itu kovariansi dari X dan Y adalah

$$Cov(X,Y) = E(XY) - E(X)E(Y) = -p_1 p_2$$

Demikian pula, dapat ditunjukkan bahwa

$$E\left(X^2\right) = p_1 \qquad \text{dan} \qquad E\left(Y^2\right) = p_2.$$

Jadi,

$$Var(X) = E(X^2) - E(X)^2 = p_1 - p_1^2 = p_1(1-p_1)$$

dan

$$Var(X) = E(Y^2) - E(Y)^2 = p_2 - p_2^2 = p_2(1-p_2)$$

Teorema berikutnya menyajikan beberapa informasi mengenai distribusi bersyarat $f(x|y)$ dan $f(y|x)$.

Teorema 3.2. Jika $(X,Y) \sim BER(p_1, p_2)$, di mana p_1 dan p_2 adalah parameter. Kemudian distribusi bersyarat $f(x|y)$ dan $f(y|x)$ juga Bernoulli dan

$$E(Y|x) = \frac{p_2(1-x)}{1-p_1}$$

$$E(X|y) = \frac{p_1(1-y)}{1-p_2}$$

$$Var(Y|x) = \frac{p_2(1-p_1-p_2)(1-x)}{(1-p_1)^2}$$

$$Var(X|y) = \frac{p_1(1-p_1-p_2)(1-y)}{(1-p_2)^2}$$

Bukti :

Perhatikan bahwa

$$f(y|x) = \frac{f(x,y)}{f_1(x)}$$

$$= \frac{f(x,y)}{\sum_{y=0}^{1} f(x,y)}$$

$$= \frac{f(x,y)}{f(x,0)+f(x,1)}, \qquad x=0,1; y=0,1; 0 \le x+y \le 1$$

Oleh karena itu

$$f(1\,|0) = \frac{f(0,1)}{f(0,0)+f(0,1)}$$

$$= \frac{p_2}{1-p_1-p_2+p_2}$$

$$= \frac{p_2}{1-p_1}$$

Dan

$$f(1\,|0) = \frac{f(1,1)}{f(1,0)+f(1,1)}$$

$$= \frac{0}{p_1 - 0} = 0$$

Sekarang kita hitung ekspektasi bersyarat $E(Y\,|x)$ untuk $x = 0,1$. Maka

$$E(Y\,|x=0) = \sum_{y=0}^{1} y f(y\,|0)$$

$$= f(1\,|0)$$

$$= \frac{p_2}{1-p_1}$$

Dan

$$E(Y\,|x=1) = f(1\,|1) = 0.$$

Menggabungkan keduanya, kita dapatkan

$$E(Y \mid x) = \frac{p_2(1-x)}{1-p_1}$$

Demikian pula,

$$E\left(Y^2 \mid x=0\right) = \sum_{y=0}^{1} y^2 f(y \mid 0)$$

$$= f(1 \mid 0)$$

$$= \frac{p_2}{1-p_1}$$

Dan

$$E\left(Y^2 \mid x=1\right) = f(1 \mid 1) = 0.$$

Oleh karena itu

$$Var(Y \mid x=0) = E\left(Y^2 \mid x=0\right) - E(Y \mid x=0)^2$$

$$= \frac{p_2}{1-p_1} - \left(\frac{p_2}{1-p_1}\right)^2$$

$$= \frac{p_2(1-p_1) - p_2^2}{(1-p_1)^2}$$

$$= \frac{p_2(1-p_1-p_2)}{(1-p_1)^2}$$

Dan

$$Var(Y \mid x=1) = 0.$$

Menggabungkan keduanya, diperoleh

$$Var\left(Y \mid x\right) = \frac{p_2\left(1 - p_1 - p_2\right)\left(1 - x\right)}{\left(1 - p_1\right)^2}, \quad x = 0,1.$$

Ekspektasi bersyarat $E\left(X \mid y\right)$ dan varians bersyarat $Var\left(X \mid y\right)$ dapat diperoleh dengan cara yang sama.

3.2. Distribusi Binomial Bivariat

Variabel acak binomial bivariat ditentukan dengan menentukan bentuk distribusi probabilitas gabungan.

Definisi 3.2.

Variabel acak bivariat diskrit $\left(X,Y\right)$ dikatakan berdistribusi binomial bivariat dengan parameter n, p_1, p_2 jika kepadatan probabilitas gabungannya berbentuk

$$f\left(x,y\right) = \begin{cases} \dfrac{n!}{x!\,y!(n-x-y)!} p_1^x p_2^y (1-p_1-p_2)^{n-x-y}, & jika\ x, y = 0,1,\ldots,n \\ 0 & untuk\ yang\ lainnya \end{cases}$$

dimana $0 < p_1,\ p_2,\ p_1 + p_2 < 1,\ x + y \le n$ dan n adalah bilangan bulat positif. Kita menunjukkan variabel acak binomial bivariat dengan menuliskan $\left(X,Y\right) \sim BIN\left(n, p_1, p_2\right)$.

Distribusi binomial bivariat juga dikenal sebagai distribusi trinomial. Akan ditunjukkan dalam bukti Teorema 3.4 bahwa distribusi marginal dari X dan Y masing-masing adalah $BIN\left(n, p_1\right)$ dan $BIN\left(n, p_2\right)$.

Dua contoh berikut mengilustrasikan penerapan distribusi bivariat binomial.

Contoh 3.1.

Di kota Yogyakarta pada Jumat malam, stasiun radio A memiliki 50 persen pendengar, stasiun radio B memiliki 30 persen pendengar, dan stasiun radio C memiliki 20 persen pendengar. Berapa probabilitas bahwa di antara 8 pendengar radio di kota Yogyakarta, yang dipilih secara acak pada Jumat malam, 5 akan mendengarkan stasiun A, 2 akan mendengarkan stasiun B, dan 1 akan mendengarkan stasiun C?

Jawab :

Jika X menyatakan jumlah pendengar yang mendengarkan stasiun A, dan Y menyatakan pendengar yang mendengarkan stasiun B. Kemudian

distribusi gabungan dari X dan Y adalah binomial bivariat dengan

$n = 8, p_1 = \frac{5}{10}$, dan $p_2 = \frac{3}{10}$. Probabilitas bahwa di antara 8 pendengar di kota Yogyakarta, yang dipilih secara acak pada Jumat malam, 5 akan mendengarkan stasiun A, 2 akan mendengarkan stasiun B, dan 1 akan mendengarkan stasiun C diberikan oleh

$$P(X = 5, Y = 2) = f(5,2)$$

$$= \frac{n!}{x!\,y!\,(n-x-y)!} p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y}$$

$$= \frac{8!}{5!2!1!} \left(\frac{5}{10}\right)^5 \left(\frac{3}{10}\right)^2 \left(\frac{2}{10}\right)$$

$$= 0.0945.$$

Contoh 3.2. Sebuah permainan tertentu melibatkan pelemparan sepasang dadu yang adil dan mengamati lemparan berjumlah 4 dan 5. Berapakah probabilitas bahwa dalam 10 lemparan dadu, satu 4 dan tiga 5 akan muncul?

Jawab :

Jika X menunjukkan dadu berjumlah 4 dan Y menunjukan dadu berjumlah 5. Maka distribusi gabungan X dan Y adalah binomial bivariat dengan $n = 10$, $p_1 = \frac{1}{6}$, $p_2 = \frac{1}{6}$ dan $1 - p_1 - p_2 = \frac{4}{6}$. Oleh karena itu, probabilitas bahwa dalam 10 lemparan sepasang dadu, muncul 4 satu kali dan 5 tiga kali adalah

$$P(X = 4, Y = 5) = f(1,3)$$

$$= \frac{n!}{x!\,y!\,(n-x-y)!} p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y}$$

$$= \frac{10!}{1!3!(10-1-3)!} \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^1 \left(1 - \frac{1}{6} - \frac{1}{6}\right)^{10-1-3}$$

$$= \frac{10!}{1!3!(10-1-3)!}\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{4}{6}\right)^6$$

$$= \frac{573440}{10077696}$$

$$= 0.0569.$$

Dengan menggunakan metode transformasi yang telah dibahas pada bab 2, dapat ditunjukkan bahwa jika X_1, X_2, dan X_3 adalah variabel acak binomial independen, maka distribusi gabungan dari variabel acak

$$X = X_1 + X_2 \ dan \ Y = X_1 + X_3$$

adalah binomial bivariat. Pendekatan ini dikenal sebagai teknik reduksi tri-variat untuk membangun distribusi bivariat.

Untuk menetapkan teorema berikutnya, diperlukan suatu generalisasi dari teorema binomial yang telah dibahas pada Bab 1 (Pengantar Statistika Matematika 1). Hasil berikut ini menggeneralisasi teorema binomial dan dapat disebut dengan teorema trinomial. Mirip dengan bukti teorema binomial, kita dapat menyusun

$$(a+b+c)^n = \sum_{x=0}^{n}\sum_{y=0}^{n}\binom{n}{x,y} a^x b^y c^{n-x-y},$$

Dimana $0 \le x + y \le n$ dan

$$\binom{n}{x,y} = \frac{n!}{x!\,y!(n-x-y)!}$$

Dalam teorema berikut, disajikan nilai harapan dari X dan Y, variansnya, kovariansi antara X dan Y, dan fungsi pembangkit momen gabungan.

Teorema 3.3.

Jika $(X,Y) \sim BIN(n, p_1, p_2)$, dimana n, p_1 dan p_2 adalah parameternya. Maka

$$E(X) = n\,p_1$$

$$E(Y) = n\,p_2$$

$$Var(X) = n\,p_1(1-p_1)$$

$$Var(Y) = n\,p_2(1-p_2)$$

$$Cov(X,Y) = -n\,p_1 p_2$$

$$M(s,t) = \left(1 - p_1 - p_2 + p_1 e^s + p_2 e^t\right)^n$$

Bukti :

Pertama, kita temukan fungsi pembangkit momen gabungan X dan Y. Fungsi pembangkit momen $M(s,t)$ diberikan oleh

$$M(s,t) = E\left(e^{sX+tY}\right)$$

$$= \sum_{x=0}^{n}\sum_{y=0}^{n} e^{sx+ty} f(x,y)$$

$$= \sum_{x=0}^{n}\sum_{y=0}^{n} e^{sx+ty} \frac{n!}{x!\,y!\,(n-x-y)!} p_1^x p_2^y \left(1-p_1-p_2\right)^{n-x-y}$$

$$= \sum_{x=0}^{n}\sum_{y=0}^{n} \frac{n!}{x!\,y!\,(n-x-y)!} \left(e^s p_1\right)^x \left(e^t p_2\right)^y \left(1-p_1-p_2\right)^{n-x-y}$$

$$= \left(1 - p_1 - p_2 + p_1 e^s + p_2 e^t\right)^n, \qquad (dengan\, teorema\, trinomial)$$

Nilai harapan dari X diberikan oleh

$$E(X) = \left.\frac{\partial M}{\partial s}\right|_{(0,0)}$$

$$= \left.\frac{\partial}{\partial s}\left(1 - p_1 - p_2 + p_1 e^s + p_2 e^t\right)^n\right|_{(0,0)}$$

$$= n\left(1 - p_1 - p_2 + p_1 e^s + p_2 e^t\right)^{n-1} p_1 e^s \Big|_{(0,0)}$$

$$= n\, p_1$$

Demikian pula, nilai harapan dari Y diberikan oleh

$$E\left(Y\right) = \frac{\partial M}{\partial t}\bigg|_{(0,0)}$$

$$= \frac{\partial}{\partial t}\left(1 - p_1 - p_2 + p_1 e^s + p_2 e^t\right)^n \bigg|_{(0,0)}$$

$$= n\left(1 - p_1 - p_2 + p_1 e^s + p_2 e^t\right)^{n-1} p_2 e^t \Big|_{(0,0)}$$

$$= n\, p_2$$

Hasil kali momen dari X dan Y adalah

$$E\left(XY\right) = \frac{\partial^2 M}{\partial t \partial s}\bigg|_{(0,0)}$$

$$= \frac{\partial^2}{\partial t \partial s}\left(1 - p_1 - p_2 + p_1 e^s + p_2 e^t\right)^n \bigg|_{(0,0)}$$

$$= \frac{\partial}{\partial t}\left(n\left(1 - p_1 - p_2 + p_1 e^s + p_2 e^t\right)^{n-1} p_2 e^s\right)\bigg|_{(0,0)}$$

$$= n\left(n-1\right) p_1 p_2$$

Oleh karena itu kovariansi dari X dan Y adalah

$$Cov\left(X,Y\right) = E\left(XY\right) - E\left(X\right)E\left(Y\right) = n\left(n-1\right) p_1 p_2 - n^2 p_1 p_2 = -n p_1 p_2.$$

Demikian pula, dapat ditunjukkan bahwa

$$E\left(X^2\right)=n\left(n-1\right)p_1^2+np_1 \qquad \text{dan} \qquad E\left(Y^2\right)=n\left(n-1\right)p_2^2+np_2.$$

Jadi,

$$Var\left(X\right)=E\left(X^2\right)-E\left(X\right)^2$$

$$=n\left(n-1\right)p_2^2+np_2-n^2p_1^2$$

$$=np_1\left(1-p_1\right)$$

dan demikian pula

$$Var\left(Y\right)=E\left(Y^2\right)-E\left(Y\right)^2=np_2\left(1-p_2\right).$$

Hasil berikut diperlukan untuk teorema berikutnya dan dapat disusun menggunakan teorema binomial yang dibahas dalam bab 1 (Pengantar Statistika Matematika 1). Untuk bilangan real a dan b, kita peroleh

$$\sum_{y=0}^{m} y\binom{m}{y}a^y b^{m-y}=m\,a\left(a+b\right)^{m-1} \qquad \left(3.1\right)$$

Dan

$$\sum_{y=0}^{m} y^2\binom{m}{y}a^y b^{m-y}=m\,a\left(m\,a+b\right)\left(a+b\right)^{m-2} \qquad \left(3.2\right)$$

dimana m adalah bilangan bulat positif.

Contoh 3.3. Jika X sama dengan angka satu dan Y sama dengan angka dua dan tiga ketika sepasang dadu yang adil digulirkan, lalu berapa koefisien korelasi dari X dan Y?

Jawab :

Kepadatan gabungan X dan Y adalah binomial bivariat dan dinyatakan oleh

$$f(x,y)=\frac{2!}{x!y!(2-x-y)!}\left(\frac{1}{6}\right)^{x}\left(\frac{2}{6}\right)^{y}\left(\frac{3}{6}\right)^{2-x-y}, \quad 0\le x+y\le 2$$

di mana x dan y adalah bilangan bulat nonnegatif. Dengan menggunakan Teorema 3.3, diperoleh

$$Var(X)=np_1(1-p_1)=2\frac{1}{6}\left(1-\frac{1}{6}\right)=\frac{10}{36}$$

$$Var(Y)=np_2(1-p_2)=2\frac{2}{6}\left(1-\frac{2}{6}\right)=\frac{16}{36}$$

dan

$$Cov(X,Y)=-np_1p_2=-2\frac{1}{6}\cdot\frac{2}{6}=-\frac{4}{36}$$

Karena itu

$$Corr(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

$$=-\frac{4}{4\sqrt{10}}$$

$$=-0.3162$$

Teorema berikutnya menyajikan beberapa informasi tentang distribusi bersyarat $f(x|y)$ dan $f(y|x)$.

Teorema 3.4.

Jika $(X,Y)\sim BIN(n,p_1,p_2)$, dimana n, p_1 dan p_2 adalah parameter. Kemudian distribusi bersyarat $f(y|x)$ dan $f(x|y)$ juga binomial dan

$$E(Y|x)=\frac{p_2(n-x)}{1-p_1}$$

$$E(X\,|y)=\frac{p_1(n-y)}{1-p_2}$$

$$Var(Y\,|x)=\frac{p_2(1-p_1-p_2)(n-x)}{(1-p_1)^2}$$

$$Var(X\,|y)=\frac{p_1(1-p_1-p_2)(n-y)}{(1-p_2)^2}$$

Bukti :

Saat

$$f(y\,|x)=\frac{f(x,y)}{f_1(x)}$$

Pertama kita temukan kepadatan marginal dari X. Kepadatan marginal $f_1(x)$ dari X diberikan oleh

$$f_1(x)=\sum_{y=0}^{n-x}\frac{n!}{x!y!(n-x-y)!}p_1^x p_2^y(1-p_1-p_2)^{n-x-y}$$

$$=\frac{n!p_1^x}{x!(n-x)!}\sum_{y=0}^{n-x}\frac{(n-x)!}{y!(n-x-y)!}p_2^y(1-p_1-p_2)^{n-x-y}$$

$$=\binom{n}{x}p_1^x(1-p_1-p_2+p_2)^{n-x} \qquad \text{(oleh teorema binomial)}$$

$$=\binom{n}{x}p_1^x(1-p_1)^{n-x}$$

Untuk menghitung ekspektasi bersyarat, dibutuhkan kepadatan bersyarat $f(x,y)$. Kepadatan bersyarat dari Y dengan syarat $X=x$ adalah

$$f(y\,|x)=\frac{f(x,y)}{f_1(x)}$$

$$= \frac{f(x,y)}{\binom{n}{x} p_1^x (1-p_1)^{n-x}}$$

$$= \frac{(n-x)!}{y!(n-x-y)!} p_2^y (1-p_1-p_2)^{n-x-y} (1-p_1)^{x-n}$$

$$= (1-p_1)^{x-n} \binom{n-x}{y} p_2^y (1-p_1-p_2)^{n-x-y}$$

Oleh karena itu ekspektasi bersyarat dari Y dengan syarat kejadian $X = x$ adalah

$$E(Y \mid x) = \sum_{y=0}^{n-x} y (1-p_1)^{x-n} \binom{n-x}{y} p_2^y (1-p_1-p_2)^{n-x-y}$$

$$= (1-p_1)^{x-n} \sum_{y=0}^{n-x} y \binom{n-x}{y} p_2^y (1-p_1-p_2)^{n-x-y}$$

$$= (1-p_1)^{x-n} \; p_2 (n-x)(1-p_1)^{n-x-1}$$

$$= \frac{p_2(n-x)}{1-p_1}$$

Selanjutnya, kita temukan varians bersyarat dari Y dengan syarat kejadian $X = x$. Untuk itu kita memerlukan ekspektasi bersyarat $E(Y^2 \mid x)$, yang diberikan oleh

$$E(Y^2 \mid x) = \sum_{y=0}^{n-x} y^2 f(x,y)$$

$$= \sum_{y=0}^{n-x} y^2 (1-p_1)^{x-n} \binom{n-x}{y} p_2^y (1-p_1-p_2)^{n-x-y}$$

$$= \left(1-p_1\right)^{x-n} \sum_{y=0}^{n-x} y^2 \binom{n-x}{y} p_2^y \left(1-p_1-p_2\right)^{n-x-y}$$

$$= \left(1-p_1\right)^{x-n} p_2 \left(n-x\right)\left(1-p_1\right)^{n-x-2} \left[\left(n-x\right)p_2 + 1 - p_1 - p_2\right]$$

$$= \frac{p_2\left(n-x\right)\left[\left(n-x\right)p_2 + 1 - p_1 - p_2\right]}{(1-p_1)^2}$$

Oleh karena itu, varians bersyarat dari Y dengan syarat $X = x$ adalah

$$Var\left(Y \mid x\right) = E(Y^2 \mid x) - E(Y \mid x)^2$$

$$= \frac{p_2\left(n-x\right)\left[\left(n-x\right)p_2 + 1 - p_1 - p_2\right]}{(1-p_1)^2} - \left(\frac{p_2\left(n-x\right)}{1-p_1}\right)^2$$

$$= \frac{p_2\left(1-p_1-p_2\right)\left(n-x\right)}{(1-p_1)^2}$$

Demikian pula, kita dapat menyusun

$$E\left(X \mid y\right) = \frac{p_1\left(n-x\right)}{1-p_2} \qquad dan \qquad Var\left(X \mid y\right) = \frac{p_1\left(1-p_1-p_2\right)\left(n-y\right)}{(1-p_2)^2}$$

Perhatikan bahwa $f(y \mid x)$ pada teorema di atas adalah fungsi kepadatan probabilitas binomial univariat. Untuk melihatnya, perhatikan

$$\left(1-p_1\right)^{x-n} \binom{n-x}{y} p_2^y \left(1-p_1-p_2\right)^{n-x-y} = \binom{n-x}{y}\left(\frac{p_2}{1-p_1}\right)^y \left(1-\frac{p_2}{1-p_1}\right)^{n-x-y}$$

Oleh karena itu, $f(y|x)$ adalah fungsi kepadatan probabilitas dari variabel acak binomial dengan parameter $n-x$ dan $\frac{p_2}{1-p_1}$.

Kepadatan marginal $f_2\left(y\right)$ dari Y dapat diperoleh dengan cara yang sama

$$f_2\left(y\right) = \binom{n}{y} p_2^y \left(1-p_2\right)^{n-y}$$

dimana $y=0,1, ..., n$. Bentuk dari kepadatan ini menunjukkan bahwa marginalnya adalah distribusi binomial bivariat.

Contoh 3.4. Jika W sama dengan berat dari sabun dalam 1 Kg kotak didistribusikan di Bantul. Misalkan $P(W<1)=0,02$ dan $P(W>1,072)=0,08$ Sebut sekotak sabun ringan, bagus, atau berat tergantung apakah $W\left\langle 1,1\leq W\leq 1.072, atau\, W\right\rangle 1.072$, masing-masing. Dalam sampel acak 50 kotak, jika X sama dengan jumlah kotak ringan dan Y jumlah kotak yang bagus. Berapa kurva regresi dan skedastik dari Y pada X ?

Jawab :

Fungsi kepadatan probabilitas gabungan dari X dan Y diberikan oleh

$$f\left(x,y\right)=\frac{50!}{x!y!\left(50-x-y\right)!}p_1^x p_2^y\left(1-p_1-p_2\right)^{50-x-y}, \qquad 0\leq x+y\leq 50$$

di mana x dan y adalah bilangan bulat nonnegatif. Misalkan $\left(X,Y\right)\sim BIN\left(n,p_1,p_2\right)$, dengan $n=50, p_1=0.02$ dan $p_2=0.90$. Kurva regresi Y pada X adalah diberikan oleh

$$E\left(Y\mid x\right)=\frac{p_2\left(n-x\right)}{1-p_1}$$

$$=\frac{0.90\left(50-x\right)}{1-0.02}$$

$$=\frac{45}{49}\left(50-x\right)$$

Kurva skedastik dari Y pada X adalah varians bersyarat dari Y dengan syarat $X=x$ dan itu sama dengan

$$Var\left(Y\mid x\right)=\frac{p_2\left(1-p_1-p_2\right)\left(n-x\right)}{(1-p_1)^2}$$

$$=\frac{0.90\left(0.08\right)\left(50-x\right)}{(1-0.02)^2}$$

$$= \frac{180}{2401}(50 - x)$$

Perhatikan bahwa jika $n = 1$, maka distribusi binomial bivariat berkurang menjadi menjadi distribusi Bernoulli bivariat.

3.3. Distribusi Geometri Bivariat

Ingatlah bahwa jika variabel acak X menunjukkan jumlah percobaan ke-berhasilan pertama, maka X adalah geometri univariat. Fungsi kepadatan probabilitas dari variabel geometri univariat adalah

$$f(x) = p^{x-1}(1-p), \qquad x = 1, 2, 3, \ldots, \infty$$

Dengan p adalah probabilitas kegagalan dalam sebuah percobaan Bernoulli. Distribusi geometri univariat ini dapat digeneralisasikan menjadi kasus bivariat.

Definisi 3.3.

Variabel acak bivariat diskrit (X, Y) dikatakan memiliki distribusi geometri bivariat dengan parameter p_1 dan p_2 jika bentuk kepadatan probabilitas gabungan

$$f(x,y) = \begin{cases} \dfrac{(x+y)!}{x!\,y!} p_1^x p_2^y (1 - p_1 - p_2), & jika\ x, y = 0, 1, \ldots, \infty \\ 0 & untuk\ yang\ lainnya \end{cases}$$

dimana $0 < p_1, p_2, p_1 + p_2 < 1$. Kita tunjukkan bahwa variabel acak geometri

bivariat dengan menulis $(X, Y) \sim GEO(p_1, p_2)$.

Contoh 3.5.

Kendaraan bermotor yang tiba di sebuah perempatan dapat belok ke kanan atau ke kiri atau terus lurus ke depan. Dalam suatu studi tentang pola persimpangan lalu lintas pada saat ini selama periode waktu yang lama, insinyur telah mencatat bahwa 40 persen kendaraan bermotor belok kiri, 25 persen belok kanan, dan sisanya terus lurus ke depan. Untuk 10 mobil berikutnya yang memasuki persimpangan, berapa probabilitas 5 mobil berbelok ke kiri, 4 mobil berbelok ke kanan, dan mobil terakhir akan berjalan lurus ke depan?

Jawab :

Misalkan X menunjukkan banyaknya mobil yang belok kiri dan Y menunjukan jumlah mobil yang belok kanan. Karena, mobil terakhir akan lurus ke depan, distribusi gabungan X dan Y geometri dengan parameter $p_1 = 0{,}4, p_2 = 0{,}25$ dan $p_3 = 1 - p_1 - p_2 = 0{,}35$. Untuk 10 mobil berikutnya memasuki persimpangan, kemungkinan 5 mobil berbelok ke kiri, 4 mobil akan berbelok ke kanan, dan mobil terakhir yang akan lurus di depan diberikan oleh

$$P(X=5,Y=4)=f(5,4)$$

$$=\frac{(x+y)!}{x!y!}p_1^x p_2^y(1-p_1-p_2)$$

$$=\frac{(5+4)!}{5!4!}(0.4)^5(0.25)^4(1-0.4-0.25)$$

$$=\frac{9!}{5!4!}(0.4)^5(0.25)^4(0.35)$$

$$=0.00677$$

Hasil teknis berikut ini penting untuk membuktikan teorema berikut. Jika a dan b adalah bilangan real positif dengan $0<a+b<1$, maka

$$\sum_{x=0}^{\infty}\sum_{y=0}^{\infty}\frac{(x+y)!}{x!y!}a^x b^y=\frac{1}{1-a-b} \qquad (3.3)$$

Dalam teorema berikut, disajikan nilai harapan dan varians dari X dan Y, kovarian antara X dan Y, dan fungsi pembangkit momen.

Teorema 3.5. Jika $(X,Y)\sim GEO(p_1,p_2)$, dimana p_1 dan p_2 adalah parameter. Sehingga

$$E(X)=\frac{p_1}{1-p_1-p_2}$$

$$E(Y)=\frac{p_2}{1-p_1-p_2}$$

$$Var(X)=\frac{p_1(1-p_2)}{(1-p_1-p_2)^2}$$

$$Var(Y) = \frac{p_2(1-p_1)}{(1-p_1-p_2)^2}$$

$$Cov(X,Y) = \frac{p_1 p_2}{(1-p_1-p_2)^2}$$

$$M(s,t) = \frac{1-p_1-p_2}{1-p_1e^s-p_2e^t}$$

Bukti :

Diketahui fungsi pembangkit momen gabungan $M(s,t)$ dari X dan Y. Fungsi pembangkit momen gabungan $M(s,t)$ diberikan oleh

$$M(s,t) = \left(Ee^{sX+tY}\right)$$

$$= \sum_{x=0}^{n}\sum_{y=0}^{n} e^{sx+ty} f(x,y)$$

$$= \sum_{x=0}^{n}\sum_{y=0}^{n} e^{sx+ty} \frac{(x+y)!}{x!\,y!} p_1^x p_2^y (1-p_1-p_2)$$

$$= (1-p_1-p_2)\sum_{x=0}^{n}\sum_{y=0}^{n} \frac{(x+y)!}{x!\,y!}(p_1e^s)^x(p_2e^t)^y$$

$$= \frac{(1-p_1-p_2)}{1-p_1e^s-p_2e^t} \qquad (dengan(3.3))$$

Hasil berikut diperlukan untuk teorema berikutnya. Misalkan a adalah bilangan real positif kurang dari satu. Sehingga

$$\sum_{y=0}^{\infty} \frac{(x+y)!}{x!\,y!} a^y = \frac{1}{(1-a)^{x+1}} \qquad (3.4)$$

$$\sum_{y=0}^{\infty} \frac{(x+y)!}{x!\,y!} y a^y = \frac{a(1+x)}{(1-a)^{x+2}} \qquad (3.5)$$

Dan

$$\sum_{y=0}^{\infty}\frac{(x+y)!}{x!y!}y^2a^y=\frac{a(1+x)}{(1-a)^{x+3}}\left[a(1+x)+1\right] \qquad (3.6)$$

Teorema berikutnya menyajikan beberapa informasi tentang kepadatan bersyarat $f(x\,|\,y)$ dan $f(y\,|\,x)$.

Teorema 3.6.

Jika $(X,Y)\sim GEO(p_1,p_2)$, dimana p_1 dan p_2 adalah parameter. Kemudian distribusi bersyarat $f(y\,|\,x)$ dan $f(x\,|\,y)$ juga geometri dan

$$E(Y\,|\,x)=\frac{p_2(1+x)}{1-p_2}$$

$$E(X\,|\,y)=\frac{p_1(1+y)}{1-p_1}$$

$$Var(Y\,|\,x)=\frac{p_2(1+x)}{(1-p_2)^2}$$

$$Var(X\,|\,y)=\frac{p_1(1+y)}{(1-p_1)^2}$$

Bukti :

Seperti sebelumnya, pertama-tama kita temukan kepadatan probabilitas bersyarat dari Y dengan syarat kejadian $X=x$. Kepadatan marginal $f_1(x)$ diberikan oleh

$$f_1(x)=\sum_{y=0}^{\infty}f(x,y)$$

$$=\sum_{y=0}^{\infty}\frac{(x+y)!}{x!y!}p_1^x p_2^y(1-p_1-p_2)$$

$$=(1-p_1-p_2)\,p_1^x\sum_{y=0}^{\infty}\frac{(x+y)!}{x!y!}p_2^y$$

$$= \frac{\left(1 - p_1 - p_2\right) p_1^x}{(1 - p_2)^{x+1}} \qquad \left(dengan\left(3.4\right)\right)$$

Oleh karena itu, kepadatan bersyarat dari Y dengan syarat kejadian $X = x$ adalah

$$f\left(y \mid x\right) = \frac{f\left(x, y\right)}{f_1\left(x\right)} = \frac{\left(x + y\right)!}{x!\,y!} p_2^y (1 - p_2)^{x+1}$$

Ekspektasi bersyarat dari Y dengan syarat kejadian $X = x$ adalah

$$E\left(Y \mid x\right) = \sum_{y=0}^{\infty} yf\left(x, y\right)$$

$$= \sum_{y=0}^{\infty} y \frac{\left(x + y\right)!}{x!\,y!} p_2^y (1 - p_2)^{x+1}$$

$$= \frac{p_2\left(1 + x\right)}{\left(1 - p_2\right)} \qquad \left(dengan\left(3.5\right)\right)$$

Demikian pula, dapat ditunjukkan

$$E(X \mid y) = \frac{p_1\left(1 + y\right)}{\left(1 - p_1\right)}$$

Untuk menghitung varians bersyarat dari Y dengan syarat kejadian $X = x$, pertama kita harus mencari $E(Y^2 \mid x)$, yang diberikan oleh

$$E(Y^2 \mid x) = \sum_{y=0}^{\infty} y^2 f(y \mid x)$$

$$= \sum_{y=0}^{\infty} y^2 \frac{\left(x + y\right)!}{x!\,y!} p_2^y (1 - p_2)^{x+1}$$

$$= \frac{p_2\left(1 + x\right)}{(1 - p_2)^2}\left[p_2\left(1 + x\right) + 1\right] \qquad \left(dengan\left(3.6\right)\right)$$

Karena itu

$$Var\left(Y^2 \mid x\right) = E(Y^2 \mid x) - E(Y \mid x)^2$$

$$= \frac{p_2\left(1+x\right)}{(1-p_2)^2}\left[p_2\left(1+x\right)+1\right] - \left(\frac{p_2\left(1+x\right)}{\left(1-p_2\right)}\right)^2$$

$$= \frac{p_2\left(1+x\right)}{(1-p_2)^2}$$

Momen lainnya dapat ditentukan dengan cara yang sama.

3.4. Distribusi Binomial Negatif Bivariat

Distribusi binomial negatif univariat dapat digeneralisasikan ke kasus bivariat.

Definisi 3.4.

Sebuah variabel acak diskrit bivariat $\left(X,Y\right)$ dikatakan memiliki distribusi binomial negatif bivariat dengan parameter k, p_1 dan p_2 jika kepadatan probabilitas gabungannya adalah dalam bentuk

$$f\left(x,y\right) = \begin{cases} \dfrac{(x+y+k-1)!}{x!\,y!(k-1)!} p_1^x p_2^y (1-p_1-p_2)^k, & jika\ x,y = 0,1,\ldots,\infty \\ 0 & untuk\ yang\ lainnya \end{cases}$$

dimana $0 < p_1, p_2,\ p_1 + p_2 < 1$ dan k adalah bilangan bulat positif bukan nol. Kita tunjukkan variabel acak binomial negatif bivariat dengan menuliskan $\left(X,Y\right) \sim NBIN\left(k, p_1, p_2\right)$.

Contoh 3.6. Eksperimen terdiri dari pemilihan kelereng secara acak dan dengan pengembalian dari kotak yang berisi 10 kelereng putih, 15 kelereng hitam dan 5 kelereng hijau. Berapa probabilitas yang dibutuhkan tepat 11 percobaan untuk mendapatkan 5 kelereng putih, 3 kelereng hitam dan 3 kelereng hijau pada percobaan ke-11?

Jawab :

Jika X menunjukkan banyaknya kelereng putih dan Y menunjukkan jumlah kelereng hitam. Distribusi gabungan dari X dan Y adalah binomial negatif bivariat dengan parameter $p_1 = \frac{1}{3},\ p_2 = \frac{1}{2}, dan\, k = 3$. Oleh karena itu probabilitas dibutuhkannya tepat 11 percobaan untuk mendapatkan 5 putih, 3 hitam dan ketiga kelereng hijau pada percobaan ke-11 ini

$$P(X=5,Y=3)=f(5,3)$$

$$=\frac{(x+y+k-1)!}{x!y!(k-1)!}p_1^x p_2^y (1-p_1-p_2)^k$$

$$=\frac{(5+3+3-1)!}{5!3!(3-1)!}(0.33)^5(0.5)^3(1-0.33-0.5)^3$$

$$=\frac{10!}{5!3!2!}(0.33)^5(0.5)^3(0.17)^3$$

$$=0.0000503$$

Teorema binomial negatif yang dibahas dalam bab 5 bisa jadi digeneralisasikan menjadi

$$\sum_{x=0}^{\infty}\sum_{y=0}^{\infty}\frac{(x+y+k-1)!}{x!y!(k-1)!}p_1^x p_2^y=\frac{1}{(1-p_1-p_2)^k} \qquad (3.7)$$

Dalam teorema berikut, kami menyajikan nilai ekspektasi dan varians dari X dan Y, kovarian antara X dan Y, dan fungsi pembangkit momen.

Teorema 3.7

Jika $(X,Y)\sim NBIN(k,p_1,p_2)$, dimana k, p_1 dan p_2 adalah parameter. Kemudian

$$E(X)=\frac{kp_1}{1-p_1-p_2}$$

$$E(Y)=\frac{kp_1}{1-p_1-p_2}$$

$$Var(X)=\frac{kp_1(1-p_2)}{(1-p_1-p_2)^2}$$

$$Var(Y)=\frac{kp_2(1-p_2)}{(1-p_1-p_2)^2}$$

$$Cov(X,Y)=\frac{kp_1p_2}{(1-p_1-p_2)^2}$$

$$M(s,t)=\frac{(1-p_1-p_2)^k}{(1-p_1e^s-p_2e^t)^k}$$

Bukti :

Kita hanya mencari fungsi penghasil momen gabungan $M(s,t)$ dari variabel acak X dan Y. Fungsi pembangkit momen diberikan oleh

$$M(s,t)=E\left(e^{sX+tY}\right)$$

$$=\sum_{x=0}^{\infty}\sum_{y=0}^{\infty}e^{sX+tY}f(x,y)$$

$$=\sum_{x=0}^{\infty}\sum_{y=0}^{\infty}e^{sX+tY}\frac{(x+y+k-1)!}{x!y!(k-1)!}p_1^xp_2^y(1-p_1-p_2)^k$$

$$=(1-p_1-p_2)^k\sum_{x=0}^{\infty}\sum_{y=0}^{\infty}\frac{(x+y+k-1)!}{x!y!(k-1)!}\left(e^sp_1\right)^x\left(e^tp_2\right)^y$$

$$=\frac{(1-p_1-p_2)^k}{(1-p_1e^s-p_2e^t)^k}\qquad(dengan(3.7))$$

Untuk menyusun teorema berikutnya, kita membutuhkan dua hasil berikut. Jika a adalah konstanta real positif dalam interval $(0,1)$, maka

$$\sum_{y=0}^{\infty}\frac{(x+y+k-1)!}{x!y!(k-1)!}a^y=\frac{1(x+k)}{(1-a)^{x+k}}\qquad(3.8)$$

$$\sum_{y=0}^{\infty} y \frac{(x+y+k-1)!}{x!y!(k-1)!} a^y = \frac{a(x+k)}{(1-a)^{x+k+1}} \tag{3.9}$$

Dan

$$\sum_{y=0}^{\infty} y^2 \frac{(x+y+k-1)!}{x!y!(k-1)!} a^y = \frac{a(x+k)}{(1-a)^{x+k+2}} \left[1+(x+k)a\right] \tag{3.10}$$

Teorema berikutnya menyajikan beberapa informasi tentang kepadatan bersyarat $f(x\,|y)$ dan $f(y\,|x)$

Teorema 3.8

Jika $(X,Y) \sim NBIN(k, p_1, p_2)$, dimana k, p_1 dan p_2 adalah parameter. Maka kepadatan bersyarat $f(y\,|x)$ dan $f(x\,|y)$ juga binomial negatif dan

$$E(Y\,|x) = \frac{p_2(k+x)}{1-p_2}$$

$$E(X\,|y) = \frac{p_1(k+y)}{1-p_1}$$

$$Var(Y\,|x) = \frac{p_2(k+x)}{(1-p_2)^2}$$

$$Var(X\,|y) = \frac{p_1(k+y)}{(1-p_1)^2}$$

Bukti :

Pertama, kita temukan kepadatan marginal X. Kepadatan marginal $f_1(x)$ diberikan oleh

$$f_1(x) = \sum_{y=0}^{\infty} f(x,y)$$

$$= \sum_{y=0}^{\infty} \frac{(x+y+k-1)!}{x!y!(k-1)!} p_1^x p_2^y$$

$$= \left(1-p_1-p_2\right)^k p_1^x \frac{\left(x+y+k-1\right)!}{x!y!\left(k-1\right)!} p_2^y$$

$$= \left(1-p_1-p_2\right)^k p_1^x \frac{1}{\left(1-p_2\right)^{x+k}} \qquad \left(dengan\left(3.8\right)\right)$$

Kepadatan bersyarat dari Y dengan syarat kejadian $X = x$ adalah

$$f\left(y\,|x\right) = \frac{f\left(x,y\right)}{f_1\left(x\right)}$$

$$= \frac{\left(x+y+k-1\right)!}{x!y!\left(k-1\right)!} p_2^y \left(1-p_2\right)^{x+k}$$

Ekspektasi bersyarat $E(Y\,|x)$ diberikan oleh

$$E\left(Y\,|x\right) = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} y \frac{\left(x+y+k-1\right)!}{x!y!\left(k-1\right)!} p_2^y \left(1-p_2\right)^{x+k}$$

$$= \left(1-p_2\right)^{x+k} \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} y \frac{\left(x+y+k-1\right)!}{x!y!\left(k-1\right)!} p_2^y$$

$$= \left(1-p_2\right)^{x+k} \frac{p_2\left(x+k\right)}{\left(1-p_2\right)^{x+k+1}} \qquad \left(dengan\left(3.9\right)\right)$$

$$= \frac{p_2\left(x+k\right)}{(1-p_2)}$$

Ekspektasi bersyarat $E(Y^2\,|x)$ dapat dihitung sebagai berikut

$$E\left(Y^2\,|x\right) = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} y^2 \frac{\left(x+y+k-1\right)!}{x!y!\left(k-1\right)!} p_2^y \left(1-p_2\right)^{x+k}$$

$$= \left(1-p_2\right)^{x+k} \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} y^2 \frac{\left(x+y+k-1\right)!}{x!y!\left(k-1\right)!} p_2^y$$

$$=\left(1-p_2\right)^{x+k}\frac{p_2\left(x+k\right)}{\left(1-p_2\right)^{x+k+2}}\left[1+\left(x+k\right)p_2\right] \qquad \left(dengan\left(3.10\right)\right)$$

$$=\frac{p_2\left(x+k\right)}{\left(1-p_2\right)^2}\left[1+\left(x+k\right)p_2\right]$$

Varians bersyarat dari Y dengan syarat $X=x$ adalah

$$Var\left(Y\mid x\right)=E\left(Y^2\mid x\right)-E(Y\mid x)^2$$

$$=\frac{p_2\left(x+k\right)}{\left(1-p_2\right)^2}\left[1+\left(x+k\right)p_2\right]-\left(\frac{p_2\left(x+k\right)}{(1-p_2)}\right)^2$$

$$=\frac{p_2\left(x+k\right)}{(1-p_2)^2}$$

Nilai ekspektasi bersyarat $E(X\mid y)$ dan varians bersyarat $Var\left(X\mid y\right)$ dapat dihitung dengan cara yang sama. Perhatikan bahwa jika $k=1$, maka distribusi binomial negatif bivariat berkurang menjadi distribusi geometri bivariat.

3.5. Distribusi Hipergeometri Bivariat

Distribusi hipergeometrik univariat dapat digeneralisasikan ke kasus bivariat.

Definisi 3.5.

Variabel acak bivariat diskrit $\left(X,Y\right)$ dikatakan memiliki distribusi hipergeometrik bivariat dengan parameter $r,\ n_1,\ n_2,\ n_3$ jika distribusi probabilitas gabungannya berbentuk

$$f\left(x,y\right)=\begin{cases}\dfrac{\dbinom{n_1}{x}\dbinom{n_2}{y}\dbinom{n_3}{r-x-y}}{\dbinom{n_1+n_2+n_3}{r}}, & jika\ x,y=0,1,\ldots,r\\ 0 & untuk\ yang\ lainnya\end{cases}$$

dimana $x \le n_1$, $y \le n_2$, $r - x - y \le n_3$ dan r adalah bilangan bulat positif kurang dari atau sama dengan $n_1 + n_2 + n_3$. Variabel acak hipergeometrik bivariat dapat dituliskan sebagai $(X,Y) \sim HYP(r, n_1, n_2, n_3)$.

Contoh 3.7. Panel calon Senat Universitas terdiri dari 6 orang dengan gelar Doktor, 4 orang dengan gelar Magister dan 9 orang Profesor. Jika pemilihannya secara acak, berapakah probabilitas bahwa juri akan terdiri dari 4 orang Doktor, 3 orang Magister dan 5 orang Profesor?

Jawab :

Diketahui $n_1 = 7, n_2 = 3$ dan $n_3 = 9$ sehingga $n = 19$. Totalnya 12 Senat Universitas akan dipilih sehingga $r = 12$. Dalam contoh ini $x = 4, y = 3$ dan $r - x - y = 5$. Oleh karena itu kemungkinan bahwa juri akan terdiri dari 4 Doktor, 3 Magister dan 5 Profesor adalah

$$f(4,3) = \frac{\binom{7}{4}\binom{3}{3}\binom{9}{5}}{\binom{19}{12}} = \frac{4410}{50388} = 0.0875$$

Contoh 3.8. Di antara 25 mahasiswa berprestasi, ada 15 dari Pulau Jawa, 7 dari Pulau Sumatra, dan 3 dari Pulau Sulawesi. Jika 5 dari Mahasiswa Berprestasi ini pilih secara acak, berapa probabilitas untuk mendapatkan 4 dari Pulau Jawa dan 1 dari Sumatra?

Jawab :

Disini $n = 25, r = 5$ dan $n_1 = 15, n_2 = 7, n_3 = 3$. Probabilitas mendapatkan 4 dari Pulau Jawa dan 1 dari Pulau Sumatra adalah

$$f(4,1) = \frac{\binom{15}{4}\binom{7}{1}\binom{3}{0}}{\binom{25}{5}} = \frac{9555}{53130} = 0.1798$$

Dalam teorema berikut, kita menyajikan nilai ekspektasi dan varians dari X dan Y, dan kovarian antara X dan Y

Teorema 3.9

Misalkan $(X,Y) \sim HYP\ (r,n_1,n_2,n_3)$, dimana r,n_1,n_2 dan n_3 adalah parameter. Sehingga

$$E(X) = \frac{rn_1}{n_1+n_2+n_3}$$

$$E(Y) = \frac{rn_2}{n_1+n_2+n_3}$$

$$Var(X) = \frac{r\,n_1(n_2+n_3)}{(n_1+n_2+n_3)^2}\left(\frac{n_1+n_2+n_3-r}{n_1+n_2+n_3-1}\right)$$

$$Var(Y) = \frac{r\,n_2(n_1+n_3)}{(n_1+n_2+n_3)^2}\left(\frac{n_1+n_2+n_3-r}{n_1+n_2+n_3-1}\right)$$

$$Cov(X,Y) = -\frac{r\,n_1n_2}{(n_1+n_2+n_3)^2}\left(\frac{n_1+n_2+n_3-r}{n_1+n_2+n_3-1}\right)$$

Bukti :

Kita hanya menemukan mean dan varians dari X. Mean dan varians dari Y dapat ditemukan dengan cara yang sama. Kovarians dari X dan Y dapat diselesaikan secara individu. Untuk mencari nilai ekspektasi dari X, kita membutuhkan kepadatan marginal $f_1(x)$ dari X. Marginal X diberikan oleh

$$f_1(x) = \sum_{y=0}^{r-x} f(x,y)$$

$$= \sum_{y=0}^{r-x} \frac{\binom{n_1}{x}\binom{n_2}{y}\binom{n_3}{r-x-y}}{\binom{n_1+n_2+n_3}{r}}$$

$$= \frac{\binom{n_1}{x}}{\binom{n_1+n_2+n_3}{r}} \sum_{y=0}^{r-x} \binom{n_2}{y} \binom{n_3}{r-x-y}$$

$$= \frac{\binom{n_1}{x}}{\binom{n_1+n_2+n_3}{r}} \binom{n_2+n_3}{r-x} \qquad (dengan\, teorema\, 1.3)$$

Hal ini menunjukkan bahwa $X \sim HYP\ (r, n_1, n_2, n_3)$. Oleh karena itu, dengan menggunakan Teorema 5.7 (Pengantar Statistika Matematika 1), kita dapatkan

$$E(X) = \frac{r\, n_1}{n_1 + n_2 + n_3}$$

Dan

$$Var(X) = \frac{r\, n_1 (n_2 + n_3)}{(n_1 + n_2 + n_3)^2} \left(\frac{n_1 + n_2 + n_3 - r}{n_1 + n_2 + n_3 - 1} \right)$$

Demikian pula, variabel random $Y \sim HYP\ (r, n_1, n_2, n_3)$. Oleh karena itu, dengan Teorema 5.7 diperoleh

$$E(Y) = \frac{r\, n_2}{n_1 + n_2 + n_3}$$

Dan

$$Var(Y) = \frac{r\, n_2 (n_1 + n_3)}{(n_1 + n_2 + n_3)^2} \left(\frac{n_1 + n_2 + n_3 - r}{n_1 + n_2 + n_3 - 1} \right)$$

Teorema berikutnya menyajikan beberapa informasi tentang kepadatan bersyarat $f(x \mid y)$ dan $f(y \mid x)$

Teorema 3.10

Misalkan $(X,Y) \sim HYP\,(r,n_1,n_2,n_3)$, dimana r,n_1,n_2 dan n_3 adalah parameter. Maka distribusi bersyarat $f\,(y\,|x)$ dan $f\,(x\,|\,y)$ merupakan hipergeometri dan

$$E(Y\,|x) = \frac{n_2(r-x)}{n_2+n_3}$$

$$E(X\,|\,y) = \frac{n_1(r-y)}{n_1+n_3}$$

$$Var(Y\,|x) = \frac{n_2 n_3}{n_2+n_3-1}\left(\frac{n_1+n_2+n_3-x}{n_2+n_3}\right)\left(\frac{x-n_1}{n_2+n_3}\right)$$

$$Var(X\,|\,y) = \frac{n_1 n_3}{n_1+n_3-1}\left(\frac{n_1+n_2+n_3-y}{n_1+n_3}\right)\left(\frac{y-n_2}{n_1+n_3}\right)$$

Bukti :

Untuk mencari $E(Y\,|x)$, kita membutuhkan kepadatan bersyarat $f(y\,|x)$ dari Y dengan syarat kejadian $X = x$. Densitas bersyarat $f(y\,|x)$ yaitu

$$f(y\,|x) = \frac{f(x,y)}{f_1(x)}$$

$$= \frac{\binom{n_2}{y}\binom{n_3}{r-x-y}}{\binom{n_2+n_3}{r-x}}$$

Oleh karena itu, variabel acak Y diberi $X = x$ adalah variabel acak hipergeometri dengan parameter n_2, n_3 dan $r - x$, yaitu

$$Y\,|x \sim HYP(n_2,\, n_3, r-x)$$

Oleh karena itu, dengan Teorema 5.7, kita dapatkan

$$E(Y\mid x)=\frac{n_2(r-x)}{n_2+n_3}$$

Dan

$$Var(Y\mid x)=\frac{n_2n_3}{n_2+n_3-1}\left(\frac{n_1+n_2+n_3-x}{n_2+n_3}\right)\left(\frac{x-n_1}{n_2+n_3}\right)$$

Demikian pula, kita dapat menemukan $E(X\mid y)$ dan $Var(X\mid y)$.

3.6. Distribusi Poisson Bivariat

Distribusi Poisson univariat dapat digeneralisasikan menjadi bivariate kasus. Pada tahun 1934, Campbell pertama kali memperoleh distribusi ini. Namun, pada tahun 1944, Aitken memberikan rumus eksplisit untuk fungsi distribusi Poisson bivariat. Pada tahun 1964, Holgate juga mencapai di distribusi Poisson bivariat sehingga memperoleh distribusi gabungan dari $X=X_1+X_3$ dan $Y=X_2+X_3$, dimana X_1,X_2,X_3 adalah variabel acak Poisson independen. Berbeda dengan sebelumnya distribusi bivariat, distribusi bersyarat dari distribusi Poisson bivariat bukan Poisson. Faktanya, Seshadri dan Patil (1964), menunjukkan hal itu tidak ada distribusi bivariat yang memiliki distribusi marginal dan bersyarat dari bentuk Poisson.

Definisi 3.6.

Variabel acak bivariat diskrit (X,Y) dikatakan berdistribusi Poisson bivariat dengan parameter $\lambda_1,\lambda_2,\lambda_3$ jika kepadatan probabilitasnya berbentuk

$$f(x,y)=\begin{cases}\dfrac{e^{(-\lambda_1-\lambda_2+\lambda_3)}(\lambda_1-\lambda_3)^x(\lambda_2-\lambda_3)^y}{x!\,y!}\psi(x,y), & untuk\ x,y=0,1,\ldots,\infty\\ 0 & untuk\ yang\ lainnya\end{cases}$$

Dimana

$$\psi(x,y):=\sum_{r=0}^{\min(x,y)}\frac{x^{(r)}y^{(r)}\lambda_3^r}{(\lambda_1-\lambda_3)^r(\lambda_2-\lambda_3)^r\,r!}$$

Dengan

$$x^{(r)}:=x(x-1)\ldots(x-r+1)$$

dan $\lambda_1>\lambda_3>0,\ \lambda_2>\lambda_3>0$ adalah parameter. Kita menunjukan variabel acak Poisson bivariat dengan menuliskan $(X,Y)\sim POI(\lambda_1,\lambda_2,\lambda_3)$.

Dalam teorema berikut, kita menyajikan nilai ekspektasi dan varians dari X dan Y, dan kovarian antara X dan Y, dan fungsi pembangkit momen gabungan

Teorema 3.11.

Misalkan $(X,Y) \sim POI(\lambda_1,\lambda_2,\lambda_3)$, di mana λ_1,λ_2 dan λ_3 adalah parameternya. Maka

$$E(X)=\lambda_1$$

$$E(Y)=\lambda_2$$

$$Var(X)=\lambda_1$$

$$Var(Y)=\lambda_2$$

$$cov(X,Y)=\lambda_3$$

$$M(s,t)=e^{-\lambda_1-\lambda_2-\lambda_3+\lambda_1 e^{s}+\lambda_2 e^{t}+\lambda_3 e^{s+t}}$$

Teorema berikutnya menyajikan beberapa karakteristik khusus dari kepadatan bersyarat $f(x\mid y)$ dan $f(y\mid x)$.

Teorema 3.12.

Misalkan $(X,Y) \sim POI(\lambda_1,\lambda_2,\lambda_3)$, di mana λ_1,λ_2 dan λ_3 adalah parameter. Maka

$$E(Y\mid x)=\lambda_2-\lambda_3+\left(\frac{\lambda_3}{\lambda_1}\right)x$$

$$E(X\mid y)=\lambda_1-\lambda_3+\left(\frac{\lambda_3}{\lambda_2}\right)y$$

$$Var(Y\mid x)=\lambda_2-\lambda_3+\left(\frac{\lambda_3(\lambda_1-\lambda_3)}{\lambda_1^2}\right)x$$

$$Var(Y \mid x) = \lambda_2 - \lambda_3 + \left(\frac{\lambda_3 (\lambda_1 - \lambda_3)}{\lambda_1^2} \right) x$$

Latihan BAB 3

1. Satu kotak berisi 10 kelereng putih, 15 kelereng hitam dan 5 kelereng hijau. Jika 10 kelereng dipilih secara acak dan tanpa pengembalian, berapakah probabilitas terpilihnya 5 kelereng putih, 3 kelereng hitam dan 2 kelereng hijau?

   Penyelesaian :

   Diketahui :

$$n_1 \,(kelereng\, merah) = 10$$

$$n_2 \,(kelereng\, hitam) = 15$$

$$n_3 \,(kelereng\, hijau) = 5$$

$$n_{total} \,(jumlah\, satu\, kotak\, kelereng\,) = n_1 + n_2 + n_3 = 30$$

$$Jumlah\, kelereng\, dipilih\, secara\, acak\; r = 10$$

$$x = 5, y = 3, r - x - y = 2$$

   Ditanyakan :

   Probabilitas terpilihnya 5 kelereng putih, 3 kelereng hitam dan 2 kelereng hijau adalah

$$f(x, y) = \frac{\binom{n_1}{x}\binom{n_2}{y}\binom{n_3}{r - x - y}}{\binom{n_{total}}{r}}$$

$$f(5,3)=\frac{\binom{10}{5}\binom{15}{3}\binom{5}{2}}{\binom{30}{10}}$$

$$=\frac{\left(\frac{10!}{(10-5)!5!}\right)\left(\frac{15!}{(15-3)!3!}\right)\left(\frac{5!}{(5-2)!2!}\right)}{\frac{30!}{(30-10)!10!}}$$

$$=\frac{\left(\frac{10!}{5!5!}\right)\left(\frac{15!}{12!3!}\right)\left(\frac{5!}{3!2!}\right)}{\frac{30!}{20!10!}}$$

$$=\frac{(252)(455)(30)}{30045015}$$

$$=\frac{3439800}{30045015}$$

$$=0{,}1145$$

Jadi probabilitas terpilihnya 5 kelereng putih, 3 kelereng hitam dan 2 kelereng hijau adalah 0,1145

2. Sebuah guci berisi 3 bola merah, 2 bola hijau dan 1 bola kuning. Tiga bola dipilih secara acak tanpa pengembalian. Berapakah probabilitas bahwa setidaknya 1 warna tidak dipilih?

   Penyelesaian :

$$bola\,merah(M)=3$$

$$bola\,hijau(H)=2$$

$$bola\,kuning\,(K) = 1$$

$$Total\,bola = 3 + 2 + 1 = 6$$

Tiga bola dipilih secara acak tanpa pengembalian $\binom{6}{3} = 20$

$$Banyak\ cara\ pemilihan\ 1\ bola\ tiap\ warna = \binom{3}{1} \times \binom{2}{1} \times \binom{1}{1}$$

$$= \left(\frac{3!}{(3-1)!3!}\right)\left(\frac{2!}{(2-1)!1!}\right)\left(\frac{1!}{(1-1)!1!}\right)$$

$$= 3 \times 2 \times 1$$

$$= 6$$

$$P(1\,bola\,dari\,setiap\,warna\,ditarik) = \frac{6}{20} = \frac{3}{10}$$

$$P(1\,warna\,tidak\,ditarik) = 1 - P(1\,bola\,dari\,setiap\,warna\,ditarik)$$

$$= 1 - \frac{3}{10}$$

$$= \frac{7}{10}$$

Jadi probabilitas bahwa setidaknya 1 warna tidak ditarik adalah $\frac{7}{10}$

3. Sebuah guci berisi 4 bola merah, 8 bola hijau dan 2 bola kuning. 5 bola dipilih secara acak tanpa pengembalian. Berapakah kemungkinan 1 bola merah, 2 bola hijau, dan 2 bola kuning akan dipilih?

Penyelesaian :

$$bola\, hijau = n_1 = 8$$

$$bola\, kuning = n_2 = 2$$

$$bola\, merah = n_3 = 4$$

$$n_{total} = 8 + 2 + 4 = 14$$

$$Jumlah\, total\, cara = \binom{n}{x, y}$$

$$= \frac{n!}{x!\,y!\,(n - x - y)!}$$

$$= \frac{5!}{2!2!(5 - 2 - 2)!}$$

$$= \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 2 \times 1}$$

$$= 30$$

$$p_1 = P(n_1) = \frac{8}{14} = \frac{4}{7}$$

$$p_2 = P(n_2) = \frac{2}{14} = \frac{1}{7}$$

$$p_3 = P(n_3) = \frac{4}{14} = \frac{2}{7}$$

$$P(X=2, Y=2) = f(2,2)$$

$$= \frac{n!}{x!\,y!(n-x-y)!} \bullet p_1^x\, p_2^y\, (1-p_1-p_2)^{n-x-y}$$

$$= 30\left(\frac{4}{7}\right)^2 \left(\frac{1}{7}\right)^2 \left(1-\frac{4}{7}-\frac{1}{7}\right)^{5-2-2}$$

$$= 30\left(\frac{4}{7}\right)\left(\frac{4}{7}\right)\left(\frac{1}{7}\right)\left(\frac{1}{7}\right)\left(\frac{2}{7}\right)$$

$$= \frac{960}{7^5}$$

4. Di antara banyak pelamar untuk posisi tertentu, 60 persen hanya memiliki pendidikan Sarjana, 30 persen memiliki pendidikan Magister, dan 10 persen memiliki pendidikan Doktoral. Jika 5 pelamar secara acak dipilih untuk diwawancarai, berapa probabilitas setidaknya satu orang memiliki gelar Doktoral?

   Penyelesaian :

   Probabilitas bahwa seseorang terpilih memiliki gelar Doktoral $p = 0{,}10$. Jumlah pelamar yang dipilih $n = 5$.

   Asumsikan bahwa pemilihan dilakukan secara independen.

   Jumlah pendaftar yang telah memiliki gelar Doktoral, X mengikuti distribusi binomial dengan parameter $n = 5$ dan $p = 0{,}10$ maka fungsi kepadatan probabilitas dari X sebagai berikut :

$$p(x) = \binom{5}{x}(0{,}10)^x (1-0{,}10)^{5-x}$$

   Probabilitas bahwa setidaknya satu orang terpilih memiliki gelar Doktoral.

$$P(X \geq 1) = 1 - P(X = 0)$$

$$= 1 - \left[\binom{5}{0} (0,10)^{0} (1-0,10)^{5-0} \right]$$

$$= 1 - \left[\frac{5!}{(5-0)!0!} (1)(0,9)^{5} \right]$$

$$= 1 - \left[(1)(1)(0,5905) \right]$$

$$= 1 - 0,5905$$

$$= 0,4095$$

5. Dalam populasi sebanyak 200 mahasiswa yang baru saja menyelesaikan kursus pertama mata kuliah Statistika, 50 mahasiswa mendapatkan nilai A, 80 mahasiswa mendapatkan nilai B dan sisanya mendapatkan F. Sebuah sampel berukuran 25 diambil secara acak tanpa penggantian dari populasi ini. Berapa probabilitas 10 mahasiswa mendapatkan nilai A, 12 mahasiswa mendapatkan nilai B dan 3 mahasiswa mendapatkan nilai F?

   Penyelesaian :

   Diketahui :

$$n_1 \left(jumlah\, mendapatkan\, nilai\; A \right) = 50$$

$$n_2 \left(jumlah\, mendapatkan\, nilai\; B \right) = 80$$

$$n_3 \left(jumlah\, mendapatkan\, nilai\; F \right) = 70$$

$$n_{total} \left(jumlah\, satu\, kotak\, kelereng \right) = n_1 + n_2 + n_3 = 200$$

$$Jumlah\, kelereng\, dipilih\, secara\, acak = r = 25$$

$$x = 10,\; y = 12,\; r - x - y = 3$$

Ditanyakan :

Berapa probabilitas 10 mahasiswa mendapatkan nilai A, 12 mahasiswa mendapatkan nilai B dan 3 mahasiswa mendapatkan nilai F adalah

$$f(x,y)=\frac{\binom{n_1}{x}\binom{n_2}{y}\binom{n_3}{r-x-y}}{\binom{n_{total}}{r}}$$

$$f(x,y)=\frac{\binom{50}{x}\binom{80}{y}\binom{70}{25-x-y}}{\binom{200}{25}} \qquad untuk\; x>0, y>0, x+y\le 25$$

$$f(10,12)=\frac{\binom{50}{10}\binom{80}{12}\binom{70}{3}}{\binom{200}{25}}$$

$$=\frac{\left(\frac{50!}{(50-10)!10!}\right)\left(\frac{80!}{(80-12)!12!}\right)\left(\frac{70!}{(70-3)!3!}\right)}{\frac{200!}{(200-25)!25!}}$$

$$=\frac{\left(\frac{50!}{40!10!}\right)\left(\frac{80!}{68!12!}\right)\left(\frac{70!}{67!3!}\right)}{\frac{200!}{175!25!}}$$

$$=\frac{608164100127151395625}{811744530737959007797446}$$

$$=0.0007492062799$$

Jadi probabilitas 10 siswa mendapatkan A, 12 siswa mendapatkan B dan 3 siswa mendapatkan F adalah 0.0007492062799

6. Jika fungsi pembangkit momen gabungan X dan Y adalah $M(s,t)=k\left(\frac{4}{7-e^s-2e^t}\right)^5$, lalu berapakah nilai konstanta k? Berapa koefisien korelasi antara X dan Y?

Penyelesaian :

$$M_{XY}(s,t)=K\left(\frac{4}{7-e^s-2e^t}\right)^5$$

Diketahui bahwa $M_{XY}(0,0)=1$

Sehingga

$$K\left(\frac{4}{7-1-2}\right)^5=1 \qquad \{e^0=1\}$$

$$K=1$$

$$\therefore M_{XY}(s,t)=1024\left(7-e^s-2e^t\right)^{-5}$$

Marginal fungsi pembangkit momen dari x

$$M_X(s)=M_{XY}(s,0)$$

$$=1024\ \left(7-e^s-2e^t\right)^{-5}$$

$$=1024\ \left(5-e^5\right)^{-5}$$

$$E(X)=\frac{d}{ds}M_X(s)\Big|_{s=0}$$

$$=1024(-5)\left(5-e^s\right)^{-6}\left(0-e^s\right)\Big|_{s=0} \quad \ldots(1)$$

$$=1024(5)(5-1)^{-6} \qquad \gg \{e^{0}=1\}$$

$$=\frac{5}{4}$$

$$E\left(X^{2}\right)=\frac{d^{2}}{ds^{2}}M_{X}(s)\Big|_{s=0}$$

$$=1024(5)\left(5-e^{s}\right)^{-6}e^{s}\Big|_{s=0} \qquad \textit{berdasarkan persamaan}(1)$$

$$=1024(5)\Big[\left(s-e^{s}\right)^{-6}e^{s}+e^{s}(-6)\left(s-e^{s}\right)^{-7}\left(0-e^{s}\right)\Big]\Big|_{s=0}$$

$$=1024(5)\Big[(5-1)^{-6}+6(5-1)^{-7}\Big]$$

$$=\frac{5}{4}+\frac{30}{16}$$

$$=\frac{25}{8}$$

$$Var(X)=E\left(X^{2}\right)-\left[E(X)\right]^{2}$$

$$=\frac{25}{8}-\left[\frac{5}{4}\right]^{2}$$

$$=\frac{25}{8}-\frac{25}{16}$$

$$=\frac{25}{16}$$

Marginal fungsi pembangkit momen dari y

$$M_Y(t) = M_{XY}(0,t)$$

$$= 1024\ \left(7-1-2e^t\right)^{-5}$$

$$M_Y(t) = 1024\ \left(6-2e^t\right)^{-5}$$

$$E(Y) = \frac{d}{dt} M_Y(t)\Big|_{t=0}$$

$$= 1024(-5)\left(6-2e^t\right)^{-6}\left(0-2e^t\right)\Big|_{t=0} \quad \ldots(2)$$

$$= 1024(10)(6-2)^{-6}$$

$$= \frac{10}{4}$$

$$= \frac{5}{2}$$

$$E\left(X^2\right) = \frac{d^2}{dt^2} M_Y(t)\Big|_{t=0}$$

$$= \frac{d}{dt} 1024(10)\left(6-2e^t\right)^{-6} e^t\Big|_{t=0} \qquad \textit{berdasarkan persamaan}(2)$$

$$= 1024(10)\left[e^t(-6)\left(6-2e^t\right)^{-7}\left(-2e^t\right) + \left(6-2e^t\right)^{-6} e^t\right]\Big|_{t=0}$$

$$= 1024(10)\left[12(6-2)^{-7} + (6-2)^{-7}\right]$$

$$= \frac{120}{16} + \frac{10}{4} = \frac{160}{16}$$

$$= 10$$

$$Var(Y) = E(Y^2) - \left[E(Y)\right]^2$$

$$= 10 - \frac{25}{4}$$

$$= \frac{15}{4}$$

$$E(XY) = \frac{d^2}{dsdt} M_{XY}(s,t)\Big|_{s=t=0}$$

$$= \frac{d^2}{dsdt} 1024 \left(7 - e^s - 2e^t\right)^{-5}\Big|_{s=t=0}$$

$$= 1024 \frac{d}{ds}(-5)\left(7 - e^s - 2e^t\right)^{-6}\left(-2e^t\right)\Big|_{s=t=0}$$

$$= 1024(10)e^t(-6)\left(7 - e^s - 2e^t\right)^{-7}\left(-e^s\right)\Big|_{s=t=0}$$

$$= 1024(10)(6)(7 - 1 - 2)^{-7}$$

$$= \frac{60}{16}$$

$$= \frac{15}{4}$$

$$Cov(XY) = E(XY) - E(X)E(Y)$$

$$= \frac{15}{4} - \left(\frac{5}{4}\right)\left(\frac{5}{2}\right)$$

$$= \frac{5}{8}$$

$$Corr(XY) = \frac{Cov(XY)}{\sqrt{Var(X)Var(Y)}}$$

$$= \frac{\frac{5}{8}}{\sqrt{\frac{25}{16}\left(\frac{15}{4}\right)}}$$

$$= \frac{5}{8} \times \frac{4 \times 2}{5\sqrt{15}}$$

$$= \frac{1}{\sqrt{15}}$$

$$= 0.2582$$

7. Sebuah dadu dengan 1 titik pada tiga sisi, 2 titik pada dua sisi, dan 3 titik di satu sisi digulirkan 15 kali. Berapa probabilitas kita akan mendapatkan delapan 1, enam 2, dan 3 di lemparan terakhir?

   Penyelesaian :

$$W_1 = 8 \qquad n = 15$$

$$W_2 = 6 \qquad x = 8$$

$$W_3 = 3 \qquad\qquad y = 6$$

Berapakah probabilitas yang akan didapatkan dari $W_1 = 8$, $W_2 = 6$, dan $W_3 = 3$?

$$P(X = 8, Y = 6) = f(8,6)$$

$$= \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y}$$

$$= \frac{15!}{8!6!1!} \left(\frac{8}{17}\right)^8 \left(\frac{6}{17}\right)^6 \left(\frac{3}{17}\right)$$

$$= 0.03695.$$

8. Output mesin dinilai sangat baik 80 persen, baik 15 persen, dan cacat 5 persen. Berapa probabilitas bahwa sampel acak berukuran 15 memiliki output 10 sangat baik, 3 baik, dan 2 cacat?

Penyelesaian :

$$p_1 = 0,8 \qquad\qquad x = 10$$

$$p_2 = 0,15 \qquad\qquad y = 3$$

$$1 - p_1 - p_2 = 0,05 \qquad n - x - y = 2$$

$$n = 15$$

Probabilitas bahwa sampel acak berukuran 15 memiliki output 10 sangat baik, 3 baik, dan 2 cacat item adalah

$$P(X = 10, Y = 3) = f(10,3)$$

$$= \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y}$$

$$= \frac{15!}{10!3!2!}(0.80)^{10}(0.15)^{3}(0.05)^{2}$$

$$= 0.0272$$

9. Penilaian ujian Statistika dinilai dengan 80 persen mahasiswa mendapatkan nilai A, 15 persen mahasiswa mendapatkan nilai B, dan 5 persen mahasiswa mendapatkan nilai C. Sampel acak berukuran 15 diambil dari mahasiswa ujian Statistika. Berapa probabilitas 10 mahasiswa mendapatkan nilai A, 3 mahasiswa mendapatkan nilai B dan 2 mahasiswa mendapatkan nilai C?

Penyelesaian :

$$p_1 = 0,8 \qquad x = 10$$

$$p_2 = 0,15 \qquad y = 3$$

$$1 - p_1 - p_2 = 0,05 \qquad n = 15$$

probabilitas 10 mahasiswa mendapatkan nilai A, 3 mahasiswa mendapatkan nilai B dan 2 mahasiswa mendapatkan nilai C

$$P(X = 10, Y = 3) = f(10,3)$$

$$= \frac{(x+y)!}{x!y!} p_1^x p_2^y (1 - p_1 - p_2)$$

$$= \frac{(10+3)!}{10!3!} p_1^x p_2^y (1 - p_1 - p_2)$$

$$= \frac{13!}{10!3!}(0.80)^{10}(0.15)^{3}(0.05)$$

$$= 0.00518$$

BAB IV

DISTRIBUSI KONTINU BIVARIAT

Dalam bab ini, kita akan mempelajari beberapa fungsi kepadatan probabilitas bivariat kontinu yang umum digunakan. Pertama, kita ingat terlebih dahulu terkait dengan fungsi kepadatan probabilitas univariat yang dibahas dalam bab 6 (Pengantar Statistika Matematika 1). Kemudian akan diberikan beberapa distribusi bivariat lainnya yang telah dibahas dalam bab ini. Kita mulai bab ini dengan distribusi uniform bivariat.

4.1. Distribusi Uniform Bivariat

Pada bagian ini, akan dibahas distribusi uniform bivariat Morgenstern secara rinci. Marginal dari distribusi uniform bivariat Morgenstern adalah uniform. Dalam pengertian ini, ini adalah perluasan dari distribusi uniform univariat. Distribusi uniform bivariat lainnya akan ditunjukkan tanpa pembahasan lebih lanjut.

Distribusi bivariat yang marginal univariatnya berdistribusi uniform adalah dengan rumus berikut :

$$f(x,y) = f_1(x) f_2(y)\ (1+\alpha\left[2F_1(x)-1\right]\left[2F_2(y)-1\right]$$

Dimana $\alpha \in [-1,1]$ adalah parameter. Jika kita mengasumsikan bahwa fungsi distribusi kumulatif $F_i(x) = x$ dan fungsi kepadatan probabilitas $f_i(x) = 1\ (i = 1,2)$, kemudian kita telah sampai pada distribusi uniform Morgenstern pada satuan persegi. Fungsi kepadatan probabilitas gabungan $f(x,y)$ dari distribusi uniform Morgenstern pada satuan persegi diberikan oleh

$$f_{(x,y)} = 1+\alpha(2x-1)(2y-1), \qquad 0 < x, y < 1, \quad -1 \le \alpha \le 1$$

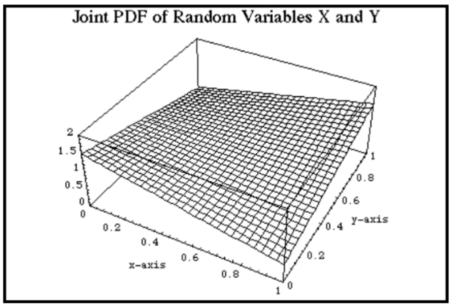
Selanjutnya, kita definisikan distribusi seragam Morgenstern pada persegi panjang yang berubah-ubah $[a,b]\times[c,d]$.

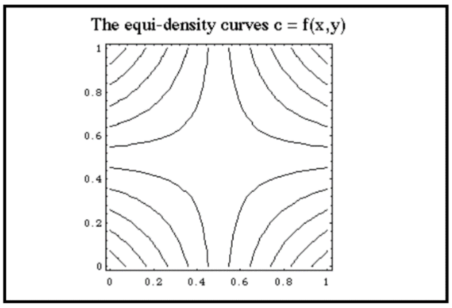
Definisi 4.1.

Variabel acak kontinu bivariat (X,Y) dikatakan memiliki distribusi uniform bivariat pada persegi panjang $[a,b]\times[c,d]$ jika fungsi kepadatan probabilitasnya berbentuk

$$f(x,y)=\begin{cases}\dfrac{1+a\left(\dfrac{2x-2a}{b-a}-1\right)\left(\dfrac{2y-2c}{d-c}-1\right)}{(b-a)(d-c)}, & \text{untuk } x\in[a,b]\ y\in[c,d]\\ 0, & \text{untuk yang lainnya}\end{cases}$$

di mana α adalah parameter yang berubah-ubah di $[-1,1]$. Kita menunjukkan variabel acak uniform bivariat Morgenstern pada persegi panjang $[a,b]\times[c,d]$ dengan menuliskan $(X,Y)\sim UNIF(a,b,c,d,\alpha)$.





Gambar berikut menunjukkan grafik dan kurva equi-density $f(x,y)$ pada satuan persegi dengan $\alpha=0,5$

Dalam teorema berikut, kita sajikan nilai ekspektasi, varians dari variabel acak X dan Y, dan kovarians antara X dan Y.

Teorema 4.1.

Misalkan $(X,Y)\sim UNIFM(a,b,c,d,\alpha)$, dengan a,b,c,d dan α adalah parameter. Sehingga

$$E(X)=\frac{b+a}{2}$$

$$E(Y)=\frac{d+c}{2}$$

$$Var(X)=\frac{(b-a)^2}{12}$$

$$Var(Y)=\frac{(d-c)^2}{12}$$

$$Cov(Y,X)=\frac{1}{36}\alpha(b-a)(d-c)$$

Bukti :

Menentukan kepadatan marginal dari X yaitu

$$f_1(x)=\int_c^d f(x,y)\,dy$$

$$=\int_c^d \frac{1+\alpha\left(\frac{2x-2a}{b-a}-1\right)\left(\frac{2y-2c}{d-c}-1\right)}{(b-a)(d-c)}dy$$

$$=\frac{1}{b-a}$$

Dengan demikian, kepadatan marginal dari X adalah seragam pada interval dari a ke b. Dimana $X \sim UNIF(a,b)$. Karenanya dengan Teorema 6.1, kita memiliki

$$E(X)=\frac{b+a}{2} \quad dan \quad Var(X)=\frac{(b-a)^2}{12}$$

Hal ini menunjukkan bahwa $Y \sim UNIF(c,d)$ sehingga dengan Teorema 6.1

$$E(Y)=\frac{d+c}{2} \quad dan \quad Var(Y)=\frac{(d-c)^2}{12}$$

Hasil kali momen dari X dan Y adalah

$$E(XY)=\int_a^b\int_c^d xy\,f(x,y)\,dx\,dy$$

$$= \int_a^b \int_c^d xy \frac{1+\alpha\left(\frac{2x-2a}{b-a}-1\right)\left(\frac{2y-2c}{d-c}-1\right)}{(b-a)(d-c)} dx\,dy$$

$$= \frac{1}{36}\alpha(b-a)(d-c) + \frac{1}{4}(b+a)(d+c)$$

Kemudian kovarians dari X dan Y adalah

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

$$= \frac{1}{36}\alpha(b-a)(d-c) + \frac{1}{4}(b+a)(d+c) - \frac{1}{4}(b+a)(d+c)$$

$$= \frac{1}{36}\alpha(b-a)(d-c)$$

Dalam teorema berikutnya, diberikan beberapa informasi yang berkaitan dengan kerapatan kondisional $f(y\,|x)$ dan $f(x\,|\,y)$.

Teorema 4.2.

Misalkan $(X,Y) \sim UNIF(a,b,c,d,\alpha)$, dimana a,b,c,d dan α adalah parameter. Kemudian

$$E(Y\,|x) = \frac{d+c}{2} + \frac{\alpha}{6(b-a)}\left(c^2+4cd\right)+d^2)\left(\frac{2x-2a}{b-a}-1\right)$$

$$E(X|\,y) = \frac{b+a}{2} + \frac{\alpha}{6(b-a)}\left(a^2+4ab+b^2\right)\left(\frac{2y-2c}{d-c}-1\right)$$

$$Var(Y|x) = \frac{1}{36}\left(\frac{d-c}{b-a}\right)^2\left[\alpha^2(a+b)(4x-a-b)+3(b-a)^2-4\alpha^2x^2\right]$$

$$Var(X|y)=\frac{1}{36}\left(\frac{b-a}{d-c}\right)^2\left[\alpha^2(c+d)(4y-c-d)+3(d-c)^2-4\alpha^2y^2\right]$$

Bukti :

Pertama, kita tentukan fungsi kepadatan bersyarat $f(y|x)$. Ingatlah bahwa $f_1(x)=\frac{1}{b-a}$. Karenanya

$$f(y|x)=\frac{1}{d-c}\left[1+a\left(\frac{2x-2a}{b-a}-1\right)\left(\frac{2y-2c}{d-c}-1\right)\right]$$

Ekspektasi bersyarat $E(Y|x)$ diberikan oleh

$$E(Y|x)=\int_c^d y\,f(y|x)dy$$

$$=\frac{1}{d-c}\int_c^d y\left[1+a\left(\frac{2x-2a}{b-a}-1\right)\left(\frac{2y-2c}{d-c}-1\right)\right]dy$$

$$=\frac{d+c}{2}+\frac{\alpha}{6(d-c)^2}\left(\frac{2x-2a}{b-a}-1\right)\left[d^3-c^3+3dc^2-3cd^2\right]$$

$$=\frac{d+c}{2}+\frac{\alpha}{6(d-c)^2}\left(\frac{2x-2a}{b-a}-1\right)\left[d^2-4dc+c^2\right]$$

Ekspektasi bersyarat $E(Y^2|x)$ diberikan oleh

$$E(Y^2|x)=\int_c^d y^2 f(y|x)dy$$

$$=\frac{1}{d-c}\int_c^d y^2\left[1+\alpha\left(\frac{2x-2a}{b-a}-1\right)\left(\frac{2y-2c}{d-c}-1\right)\right]dy$$

$$= \frac{1}{d-c} + \left[\frac{d^2 - c^2}{2} + \frac{\alpha}{d-c} \left(\frac{2x-2a}{b-a} - 1 \right) \frac{1}{6} \left(d^2 c^2 \right) \left(d-c \right)^2 \right]$$

$$= \frac{d+c}{2} + \frac{1}{6} \alpha \left(d^2 - c^2 \right) \left(\frac{2x-2a}{b-a} - 1 \right)$$

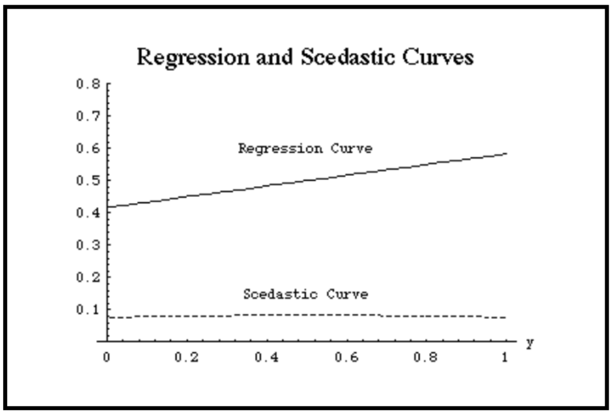
$$= \frac{d+c}{2} \left[1 + \frac{\alpha}{3} \left(d-c \right) \left(\frac{2x-2a}{b-a} - 1 \right) \right]$$

Oleh karena itu, varians bersyarat dari Y dengan syarat kejadian $X = x$ adalah

$$Var\left(Y \mid x\right) = E\left(Y^2 \mid x\right) - E(Y|x)^2$$

$$= \frac{1}{36} \left(\frac{d-c}{b-a} \right)^2 \left[\alpha^2 \left(a+b \right) \left(4x - a - b \right) + 3\left(b-a \right)^2 - 4\alpha^2 x^2 \right]$$

Ekspektasi bersyarat $E(X \mid y)$ dan varians bersyarat $Var(X \mid y)$ dapat ditemukan dengan cara yang sama.



Gambar berikut mengilustrasikan kurva regresi dan skedastik dari fungsi distribusi uniform Morgenstern pada satuan persegi dengan $\alpha = 0,5$.

Selanjutnya, diberikan definisi dari distribusi uniform bivariat umum lainnya

Definisi 4.2.

Jika $S \subset \mathbb{R}^2$ adalah suatu daerah pada bidang Euclid $\mathbb{R}^2$ dengan luas A Variabel acak X dan Y dikatakan uniform bivariat terhadap S jika kepadatan gabungan X dan Y berbentuk

$$f(x,y) = \begin{cases} \dfrac{1}{A} & \text{untuk}\,(x,y) \in S \\ \\ 0 & \text{untuk yang lainnya} \end{cases}$$

Pada tahun 1965, Plackett menyusun kelas distribusi bivariat $F(x,y)$ yang menyatakan bahwa marginal $f_1(x)$ dan $f_2(y)$ sebagai akar kuadrat dari persamaan

$$(\alpha-1)F(x,y)^2 - \{1+(\alpha-1)[F_1(x)+F_2(y)]\}F(x,y) + \alpha F_1(x)F_2(y) = 0$$

(Dimana $0 < \alpha < \infty$) yang memenuhi pertidaksamaan Fréchet

$$\max\{F_1(x)+F_2(y)-1,0\} \le F(x,y) \le \min\{F_1(x),F_2(y)\}$$

Kelas fungsi kepadatan gabungan bivariat yang disusun oleh Plackett adalah sebagai berikut

$$f(x,y) = \alpha f_1(x) f_2(y) \frac{\left[(\alpha-1)\{F_1(x)+F_2(y)-2F_1(x)F_2(y)\}+1\right]}{\left[S(x,y)^2 - 4\alpha(\alpha-1)F_1(x)F_2(y)\right]^{\frac{3}{2}}}$$

Dimana

$$S(x,y) = 1+(\alpha-1)\left(F_1(x)+F_2(y)\right)$$

Jika kita mengambil $F_i(x) = x$ dan $f_i(x) = 1$ (untuk $i = 1,2$), maka fungsi kepadatan gabungan yang dibangun oleh Plackett berkurang menjadi

$$f(x,y) = \alpha \frac{\left[(\alpha-1)\{x+y-2xy\}+1\right]}{\left[\{1+(\alpha-1)(x+y)\}^2 - 4\alpha(\alpha-1)xy\right]^{\frac{3}{2}}}$$

dimana $0 \le x,\ y \le 1$, dan $\alpha > 0$.

4.2. Distribusi Cauchy Bivariat

Ingatlah bahwa distribusi probabilitas Cauchy univariat didefinisikan pada bab 3 yakni

$$f(x) = \frac{\theta}{\pi\left[\theta + (x-\alpha)^2\right]} \qquad -\infty < x < \infty$$

di mana $\alpha > 0$ dan θ adalah parameter real. Parameter α disebut parameter lokasi. Dalam Bab 4, ditunjukkan bahwa setiap variabel acak yang fungsi kepadatan probabilitasnya adalah Cauchy tidak memiliki momen. Variabel acak ini, lebih lanjut tidak memiliki fungsi pembangkit momen. Distribusi Cauchy banyak digunakan untuk tujuan instruksional selain penggunaan statistiknya. Tujuan utama dari bagian ini adalah untuk menggeneralisasi distribusi univariat Cauchy menjadi kasus bivariat dan mempelajari berbagai sifat intrinsiknya. Kita definisikan variabel acak Cauchy bivariat dengan menggunakan bentuk fungsi kepadatan probabilitas gabungannya.

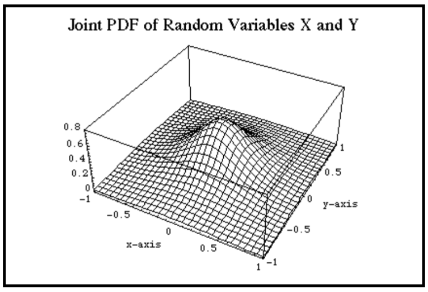
Definisi 4.3.

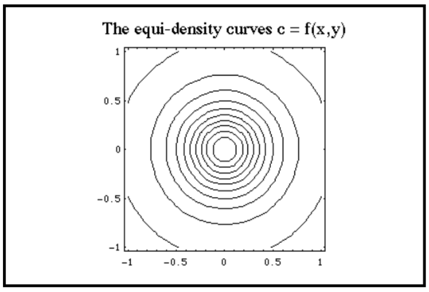
Variabel acak bivariat kontinu (X,Y) dikatakan memiliki distribusi Cauchy bivariat jika fungsi kepadatan probabilitas gabungannya berbentuk

$$f(x,y) = \frac{\theta}{2\pi\left[\theta^2 + (x-a)^2 + (y-\beta)^2\right]^{\frac{2}{3}}} \qquad -\infty < x < \infty$$

dimana θ adalah parameter positif dan α dan β adalah parameter lokasi. Kita tunjukan variabel acak bivariat cauchy dengan menulis $(X,Y) \sim CAU(\theta,\alpha,\beta)$

Gambar berikut menunjukkan grafik dan kurva equi-density fungsi kerapatan Cauchy $f(x,y)$ dengan parameter $\alpha = 0 = \beta$ dan $\theta = 0,5$.





Teorema berikut menunjukkan bahwa jika variabel acak bivariat (X,Y) adalah Cauchy, maka tidak ada momen seperti variabel acak Cauchy univariat. Selanjutnya, untuk variabel acak Cauchy bivariat (X,Y), kovarian (dan karenanya korelasi) antara X dan Y tidak ada.

Teorema 4.3.

Misalkan $(X,Y) \sim CAU(\theta,\alpha,\beta)$, di mana $\theta > 0,\alpha$ dan β adalah parameter. Maka momen $E(X),E(Y),Var(X),Var(Y)$, dan $Cov(X,Y)$ tidak ada.

Bukti :

Untuk menemukan momen X dan Y, kita membutuhkan distribusi marginalnya. Mencari margin X yang diberikan oleh

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \int_{-\infty}^{\infty} \frac{\theta}{2\pi\left[\theta^2 + (x-a)^2 + (y-\beta)^2\right]^{\frac{2}{3}}}\,dy$$

Untuk mengevaluasi integral di atas, kita gunakan substitusi trigonometri

$$y = \beta + \sqrt{\left[\theta^2 + (x-a)^2\right]}\tan\psi$$

Kemudian

$$dy = \sqrt{\left[\theta^2 + (x-a)^2\right]}sec^2\psi\, d\psi$$

Dan

$$\left[\theta^2 + (x-a)^2 + (y-\beta)^2\right]^{\frac{2}{3}} = \left[\theta^2 + (x-a)^2\right]^{\frac{2}{3}}\left(1 + tan^2\psi\right)^{\frac{2}{3}}$$

$$= \left[\theta^2 + (x-a)^2\right]^{\frac{2}{3}} sec^2\psi$$

Menggunakan ini dalam integral di atas, kita dapatkan

$$\int_{-\infty}^{\infty} \frac{\theta}{2\pi\left[\theta^2 + (x-\alpha)^2 + (y-\beta)^2\right]^{\frac{3}{2}}}\,dy = \frac{\theta}{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sqrt{\left[\theta^2 + (x-\alpha)^2\right]}\,sec^2\psi\, d\psi}{\left[\theta^2 + (x-\alpha)^2\right]^{\frac{3}{2}} sec^3\psi}$$

$$= \frac{\theta}{2\pi\left[\theta^2 + (x-\alpha)^2\right]} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} cos\,\psi\, d\psi$$

$$= \frac{\theta}{\pi\left[\theta^2 + (x-\alpha)^2\right]}$$

Oleh karena itu, marginal X adalah distribusi Cauchy dengan parameter θ dan α. Jadi, untuk variabel acak X, nilai ekspektasi $E(X)$ dan varians $Var(X)$ tidak ada (lihat Contoh 4.2 Pengantar Statistika Matematika 1). Dengan cara yang sama, dapat ditunjukkan bahwa distribusi marginal dari Y juga Cauchy dengan parameter θ dan β sehingga $E(Y)$ dan $Var(Y)$ tidak ada. Berdasarkan

$$Cov(X,Y) = E(XY) - E(X)E(Y),$$

Mudah untuk dicatat bahwa $Cov(X,Y)$ juga tidak ada.

Distribusi bersyarat dari Y dengan syarat $X = x$ diberikan oleh

$$f(y\,|x) = \frac{f(x,y)}{f_1(x)} = \frac{1}{2}\frac{\theta^2 + (x-\alpha)^2}{\left[\theta^2 + (x-\alpha)^2 + (y-\beta)^2\right]^{\frac{3}{2}}}$$

Demikian pula, distribusi bersyarat dari X dengan syarat kejadian $Y = y$ adalah

$$f(y\,|x) = \frac{1}{2}\frac{\theta^2 + (y-\beta)^2}{\left[\theta^2 + (x-\alpha)^2 + (y-\beta)^2\right]^{\frac{3}{2}}}$$

Teorema berikutnya menyatakan beberapa sifat dari kepadatan bersyarat $f(y\,|x)$ dan $f(x\,|y)$.

Teorema 4.4.

Jika $(X,Y) \sim CAU(\theta,\alpha,\beta)$, dimana $\theta > 0, \alpha$ dan β merupakan parameternya. Maka ekspektasi bersyarat

$$E(Y\,|x) = \beta$$

$$E(X\,|\,y)=\alpha,$$

Dan varians bersyarat $Var(Y\,|\,x)$ dan $Var(X\,|\,y)$ tidak ada.

Bukti :

Pertama, kita akan menunjukkan bahwa $E(Y\,|\,x)$ adalah β. Ekspektasi bersyarat dari Y dengan syarat kejadian $X=x$ dapat dihitung sebagai

$$E(Y\,|\,x)=\int_{-\infty}^{\infty} y\,f(y\,|\,x)\,dy$$

$$=\int_{-\infty}^{\infty} y\,\frac{1}{2}\,\frac{\theta^2+(x-\alpha)^2}{\left[\theta^2+(x-\alpha)^2+(y-\beta)^2\right]^{\frac{3}{2}}}\,dy$$

$$=\frac{1}{4}\left[\theta^2+(x-\alpha)^2\right]\int_{-\infty}^{\infty}\frac{d\,(\theta^2+(x-\alpha)^2+(y-\beta)^2}{\left[\theta^2+(x-\alpha)^2+(y-\beta)^2\right]^{\frac{3}{2}}}$$
$$+\frac{\beta}{2}\left[\theta^2+(x-\alpha)^2\right]\int_{-\infty}^{\infty}\frac{dy}{\left[\theta^2+(x-\alpha)^2+(y-\beta)^2\right]^{\frac{3}{2}}}$$

$$=\frac{1}{4}\left[\theta^2+(x-\alpha)^2\right]\left[-\frac{2}{\sqrt{\theta^2+(x-\alpha)^2+(y-\beta)^2}}\right]_{-\infty}^{\infty}$$
$$+\;\frac{\beta}{2}\left[\theta^2+(x-\alpha)^2\right]\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{cos\,\psi\,d\psi}{[\theta^2+(x-\alpha)^2]}$$

$$=0+\beta$$

$$=\beta$$

Demikian pula, dapat ditunjukkan bahwa $E(X\,|\,y)=\alpha$. Selanjutnya, kita tunjukkan bahwa varians bersyarat dari Y dengan syarat $X=x$ tidak ada. Untuk menunjukkan ini, kita membutuhkan $E(Y^2\,|\,x)$, yang diberikan oleh

$$E\left(Y^2 \mid x\right)=\int_{-\infty}^{\infty} y^2 \frac{1}{2} \frac{\theta^2+(x-\alpha)^2}{\left[\theta^2+(x-\alpha)^2+(y-\beta)^2\right]^{\frac{3}{2}}} dy$$

Integral di atas tidak ada dan karenanya momen kedua bersyarat dari Y dengan syarat $X=x$ tidak ada. Akibatnya, $Var\left(Y \mid x\right)$ juga tidak ada. Demikian pula, varians X dengan syarat kejadian $Y=y$ juga tidak ada.

4.3. Distribusi Gamma Bivariat

Pada bagian ini, kita sajikan tiga fungsi kepadatan probabilitas gamma bivariat yang berbeda dan mempelajari beberapa sifat intrinsiknya

Definisi 4.4.

Variabel acak kontinu bivariat (X,Y) dikatakan memiliki distribusi gamma bivariat jika fungsi kepadatan probabilitas gabungannya berbentuk

$$f(x,y)=\begin{cases} \dfrac{(xy)^{\frac{1}{2}(\alpha-1)}}{(1-\theta)\Gamma(\alpha)(\theta)^{\frac{1}{2}(\alpha-1)}} e^{-\frac{x+y}{1-\theta}} I_{\alpha-1}\left(\dfrac{2\sqrt{\theta xy}}{1-\theta}\right), & \textit{untuk } 0 \le x,y<\infty \\ 0, & \textit{untuk yang lainnya} \end{cases}$$

Dimana $\theta \in [0,1)$ dan $\alpha>0$ merupakan parameternya

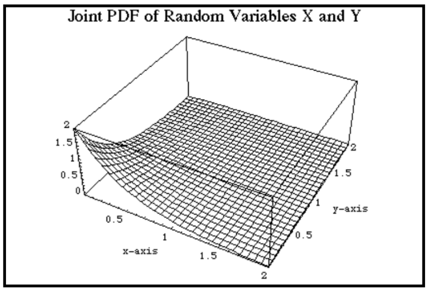
$$I_k(\mathcal{Z}) := \sum_{r=0}^{\infty} \frac{\left(\frac{1}{2}\mathcal{Z}\right)^{k+2r}}{r!\,\Gamma(k+r+1)}$$

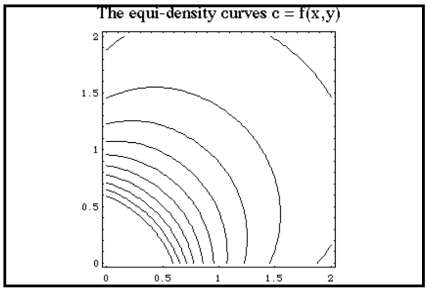
Penulisan variabel acak gamma bivariat ini adalah $(X,Y) \sim GAMK(\alpha,\theta)$.

Fungsi $I_k(\mathcal{Z})$ disebut fungsi Bessel modifikasi dari jenis pertama susunan k. Dalam bentuk eksplisit $f(x,y)$ diberikan oleh

$$f(x,y)=\begin{cases} \dfrac{1}{\theta^{\alpha-1}\Gamma(\alpha)} e^{-\frac{x+y}{1-\theta}} \sum_{k=0}^{\infty} \dfrac{(\theta xy)^{\alpha+k-1}}{k!\,\Gamma(\alpha+k)(1-\theta)^{\alpha+2k}}, & \textit{untuk } 0 \le x,y<\infty \\ 0, & \textit{untuk yang lainnya} \end{cases}$$

Gambar berikut menunjukkan grafik fungsi kepadatan gabungan $f(x, y)$ dari variabel acak gamma bivariat dengan parameter $\alpha = 1$ dan $\theta = 0.5$ dan kurva equi-density $f(x, y)$.





Pada tahun 1941, Kibble menemukan fungsi kepadatan gamma bivariat ini. Akan tetapi, Wicksell pada tahun 1933 telah menyusun fungsi karakteristik dari fungsi kepadatan gamma bivariat ini tanpa mengetahui bentuk eksplisit dari fungsi kepadatan ini. Jika $\{(X_i, Y_i) | i = 1, 2, \ldots, n\}$ adalah sampel acak dari distribusi normal bivariat dengan mean nol, kemudian variabel acak bivariat (X, Y), dimana

$$X = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \quad dan \quad Y = \frac{1}{n}\sum_{i=1}^{n} Y_i^2,$$

memiliki distribusi gamma bivariat. Fakta ini dibuktikan oleh Wicksell dengan mencari ciri khasnya fungsi (X, Y).

Teorema 4.5

Jika variabel random $(X, Y) \sim GAMK(\alpha, \theta)$, dimana $0 < \alpha < \infty$ dan $0 \leq \theta < 1$ merupakan parameternya. Maka marginal dari X dan Y merupakan gamma univariat dan

$$E(X) = \alpha$$

$$E(Y) = \alpha$$

$$Var(X) = \alpha$$

$$Var(Y) = \alpha$$

$$Cov(X,Y)=\alpha\theta$$

$$M(s,t)=\frac{1}{\left[(1-s)(1-t)-\theta st\right]^{\alpha}}$$

Bukti :

Pertama, menunjukkan bahwa distribusi marginal dari X adalah gamma univariat dengan parameternya α (dan $\theta=1$). Kepadatan marginal dari X adalah sebagai berikut

$$f_1(x)=\int_0^{\infty} f(x,y)\,dy$$

$$=\int_0^{\infty}\frac{1}{\theta^{\alpha-1}\Gamma(\alpha)}e^{-\frac{x+y}{1-\theta}}\sum_{k=0}^{\infty}\frac{(\theta xy)^{\alpha+k-1}}{k!\Gamma(\alpha+\mathrm{k})(1-\theta)^{\alpha+2k}}dy$$

$$=\sum_{k=0}^{\infty}\frac{1}{\theta^{\alpha-1}\Gamma(\alpha)}e^{-\frac{x}{1-\theta}}\frac{(\theta xy)^{\alpha+k-1}}{k!\Gamma(\alpha+\mathrm{k})(1-\theta)^{\alpha+2k}}\int_0^{\infty}y^{\alpha+k-1}e^{-\frac{y}{1-\theta}}dy$$

$$=\sum_{k=0}^{\infty}\frac{1}{\theta^{\alpha-1}\Gamma(\alpha)}e^{-\frac{x}{1-\theta}}\frac{(\theta xy)^{\alpha+k-1}}{k!\Gamma(\alpha+\mathrm{k})(1-\theta)^{\alpha+2k}}(1-\theta)^{\alpha+k}\,\Gamma(\alpha+k)$$

$$=\sum_{k=0}^{\infty}\left(\frac{\theta}{1-\theta}\right)^{k}\frac{1}{k!\Gamma(\alpha)}x^{\alpha+k-1}e^{-\frac{x}{1-\theta}}$$

$$=\frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-\frac{x}{1-\theta}}\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{x\theta}{1-\theta}\right)^{k}$$

$$=\frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-\frac{x}{1-\theta}}e^{\frac{x\theta}{1-\theta}}$$

$$=\frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-x}$$

Jadi, distribusi marginal dari X adalah gamma dengan parameternya α dan $\theta = 1$. Oleh karena itu, berdasarkan Teorema 6.3 (Pengantar Statistika Matematika 1), kita dapatkan

$$E(X) = \alpha, \qquad Var(X) = \alpha$$

Demikian pula, kita bisa menunjukkan bahwa kepadatan marginal dari Y adalah gamma dengan parameternya α dan $\theta = 1$. Sehingga didapat

$$E(Y) = \alpha, \qquad Var(Y) = \alpha$$

Fungsi pembangkit momen dapat dihitung dengan cara yang sama

Hasil berikut diperlukan untuk teorema berikutnya.

$$e^{z} = \sum_{k=0}^{\infty} k \frac{\mathcal{Z}^k}{k!} \qquad (4.1)$$

Dan deret tak hingga di sebelah kanan menyatu untuk semua $\mathcal{Z} \in \mathbb{R}$. Diferensiasi kedua sisi dari (4.1) lalu mengalikan pernyataan yang dihasilkan dengan $\mathcal{Z}$, satu memperoleh

$$\mathcal{Z}e^{z} = \sum_{k=0}^{\infty} k \frac{\mathcal{Z}^k}{k!} \qquad (4.2)$$

Jika kita diferensiasi (4.2) lagi sehubungan dengan $\mathcal{Z}$ dan mengalikan pernyataan yang dihasilkan dengan $\mathcal{Z}$, maka akan didapat

$$\mathcal{Z}e^{z} + \mathcal{Z}^2 e^{z} = \sum_{k=0}^{\infty} k^2 \frac{\mathcal{Z}^k}{k!} \qquad (4.3)$$

Teorema 4.6.

Jika variabel random $(X,Y) \sim GAMK(\alpha, \theta)$, dimana $0 < \alpha < \infty$ dan $0 \le \theta < 1$ merupakan parameternya. Maka

$$E(Y \mid x) = \theta x + (1-\theta)\alpha$$

$$E(X \mid y) = \theta y + (1-\theta)\alpha$$

$$Var(Y\,|x)=(1-\theta)\left[2\theta x+(1-\theta)\alpha\right]$$

$$Var(X\,|x)=(1-\theta)\left[2\theta y+(1-\theta)\alpha\right]$$

Bukti :

Pertama, kita mencari fungsi kepadatan probabilitas bersyarat Y dengan syarat $X=x,$ yang diberikan oleh

$$f(y\,|x)=\frac{f(x,y)}{f_1(x)}$$

$$=\frac{1}{\theta^{\alpha-1}x^{\alpha-1}e^{-x}}e^{-\frac{x+y}{1-\theta}}\sum_{k=0}^{\infty}\frac{(\theta xy)^{\alpha+k-1}}{k!\Gamma(\alpha+k)(1-\theta)^{\alpha+2k}}$$

$$=e^{x-\frac{x}{1-\theta}}\sum_{k=0}^{\infty}\frac{1}{\Gamma(\alpha+k)(1-\theta)^{\alpha+2k}}\frac{(\theta x)^k}{k!}y^{\alpha+k-1}e^{-\frac{y}{1-\theta}}$$

Selanjutnya, kita menghitung ekspektasi bersyarat dari Y dengan syarat kejadian $X=x$. Ekspektasi bersyarat $E(Y\,|x)$ diberikan oleh

$$E(Y\,|x)=\int_0^{\infty}yf(y\,|x)\,dy$$

$$=\int_0^{\infty}ye^{x-\frac{x}{1-\theta}}\sum_{k=0}^{\infty}\frac{1}{\Gamma(\alpha+k)(1-\theta)^{\alpha+2k}}\frac{(\theta x)^k}{k!}y^{\alpha+k-1}e^{-\frac{y}{1-\theta}}\,dy$$

$$=e^{x-\frac{x}{1-\theta}}\sum_{k=0}^{\infty}\frac{1}{\Gamma(\alpha+k)(1-\theta)^{\alpha+2k}}\frac{(\theta x)^k}{k!}\int_0^{\infty}y^{\alpha+k}e^{-\frac{y}{1-\theta}}\,dy$$

$$=e^{x-\frac{x}{1-\theta}}\sum_{k=0}^{\infty}\frac{1}{\Gamma(\alpha+k)(1-\theta)^{\alpha+2k}}\frac{(\theta x)^k}{k!}(1-\theta)^{\alpha+k+1}\Gamma(\alpha+k)$$

$$=(1-\theta)e^{x-\frac{x}{1-\theta}}\sum_{k=0}^{\infty}(\alpha+k)\frac{1}{k!}\left(\frac{\theta x}{1-\theta}\right)^{k}$$

$$=(1-\theta)e^{x-\frac{x}{1-\theta}}\left[\alpha e^{\frac{\theta x}{1-\theta}}+\frac{\theta x}{1-\theta}e^{\frac{\theta x}{1-\theta}}\right] \quad (\text{berdasarkan}(4.1)\ \text{dan}(4.2))$$

$$=(1-\theta)\alpha+\theta x$$

Untuk menentukan varians bersyarat dari Y dengan syarat kejadian $X=x$ kita membutuhkan ekspektasi bersyarat dari Y^2 yang diberikan kejadian $X=x$

Ekspektasi bersyarat dapat dievaluasi sebagai berikut:

$$E\left(Y^2 \mid x\right)=\int_0^{\infty} y^2 f(y \mid x)\,dy$$

$$=\int_0^{\infty} y^2 e^{x-\frac{x}{1-\theta}}\sum_{k=0}^{\infty}\frac{1}{\Gamma(\alpha+k)(1-\theta)^{\alpha+2k}}\frac{(\theta x)^k}{k!}y^{\alpha+k-1}e^{-\frac{y}{1-\theta}}\,dy$$

$$=e^{x-\frac{x}{1-\theta}}\sum_{k=0}^{\infty}\frac{1}{\Gamma(\alpha+k)(1-\theta)^{\alpha+2k}}\frac{(\theta x)^k}{k!}\int_0^{\infty} y^{\alpha+k+1}e^{-\frac{y}{1-\theta}}\,dy$$

$$=e^{x-\frac{x}{1-\theta}}\sum_{k=0}^{\infty}\frac{1}{\Gamma(\alpha+k)(1-\theta)^{\alpha+2k}}\frac{(\theta x)^k}{k!}(1-\theta)^{\alpha+k+2}\,\Gamma(\alpha+k+2)$$

$$=(1-\theta)^2 e^{x-\frac{x}{1-\theta}}\sum_{k=0}^{\infty}(\alpha+k+1)(\alpha+k)\frac{1}{k!}\left(\frac{\theta x}{1-\theta}\right)^{k}$$

$$=(1-\theta)^2 e^{x-\frac{x}{1-\theta}}\sum_{k=0}^{\infty}\left(\alpha^2+2\alpha k+k^2+\alpha+k\right)\frac{1}{k!}\left(\frac{\theta x}{1-\theta}\right)^{k}$$

$$=\left(1-\theta\right)^{2}\left[\alpha^{2}+\alpha+\left(2\alpha+1\right)\frac{\theta x}{1-\theta}+\frac{\theta x}{1-\theta}+e^{x-\frac{x}{1-\theta}}\sum_{k=0}^{\infty}\frac{k^{2}}{k!}\left(\frac{\theta x}{1-\theta}\right)^{k}\right]$$

$$=\left(1-\theta\right)^{2}\left[\alpha^{2}+\alpha+\left(2\alpha+1\right)\frac{\theta x}{1-\theta}+\frac{\theta x}{1-\theta}+\left(\frac{\theta x}{1-\theta}\right)^{2}\right]$$

$$=\left(\alpha^{2}+\alpha\right)\left(1-\theta\right)^{2}+2\left(\alpha+1\right)\theta\left(1-\theta\right)x+\theta^{2}x^{2}$$

Variansi bersyarat dari Y dengan syarat $X=x$ adalah

$$Var\left(Y\mid x\right)=E\left(Y^{2}\mid x\right)-E\left(Y\mid x\right)^{2}$$

$$=\left(\alpha^{2}+\alpha\right)\left(1-\theta\right)^{2}+2\left(\alpha+1\right)\theta\left(1-\theta\right)x+\theta^{2}x^{2}$$
$$-\ [\left(1-\theta\right)^{2}\alpha^{2}+\theta^{2}x^{2}+2\alpha\theta\left(1-\theta\right)x]$$

$$=\left(1-\theta\right)\left[\alpha\left(1-\theta\right)+2\theta x\right]$$

Karena fungsi kerapatan $f\left(x,y\right)$ simetris, yaitu $f\left(x,y\right)=f\left(y,x\right)$, ekspektasi bersyarat $E\left(X\mid y\right)$ dan varians bersyarat $Var\left(X\mid y\right)$ dapat diperoleh dengan menukar x dengan y dalam rumus $E\left(Y\mid x\right)$ dan $Var\left(Y\mid x\right)$.

Pada tahun 1941, Cherian membangun distribusi gamma bivariat yang fungsi kepadatan probabilitasnya diberikan oleh

$$f(x,y)=\begin{cases}\dfrac{e^{-(x+y)}}{\prod_{i=1}^{3}\Gamma\left(\alpha_{1}\right)}\displaystyle\int_{0}^{\min\{x,y\}}\frac{z^{\alpha_{3}}\left(x-z\right)^{\alpha_{1}}\left(y-z\right)^{\alpha_{2}}}{z\left(x-z\right)\left(y-z\right)}e^{z}\,dz, & untuk\ 0<x,y<\infty\\ 0 & untuk\ yang\ lainnya\end{cases}$$

Dimana $\alpha_{1},\alpha_{2},\alpha_{3}\in\left(0,\infty\right)$ merupakan parameternya. Jika variabel acak bivariat $\left(X,Y\right)$ memiliki fungsi kepadatan probabilitas gamma bivariat Cherian dengan parameter $\alpha_{1},\ \alpha_{2}\ dan\ \alpha_{3}$, hal ini dapat ditulis $\left(X,Y\right)\sim GAMC\left(\alpha_{1},\alpha_{2},\alpha_{3}\right)$.

Dapat ditunjukkan bahwa margin dari $f(x,y)$ diberikan oleh

$$f_1(x)=\begin{cases}\dfrac{1}{\Gamma(\alpha_1+\alpha_3)}x^{\alpha_1+\alpha_3-1}e^{-x}, & untuk\ \ 0<x<\infty\\ 0 & untuk\ yang\ lainnya\end{cases}$$

Dan

$$f_2(x)=\begin{cases}\dfrac{1}{\Gamma(\alpha_2+\alpha_3)}x^{\alpha_2+\alpha_3-1}e^{-y}, & untuk\ \ 0<y<\infty\\ 0 & untuk\ yang\ lainnya\end{cases}$$

Teorema 4.7.

Jika $(X,Y)\sim GAMC(\alpha,\beta,\gamma)$, maka

$$E(X)=\alpha+\gamma$$

$$E(Y)=\beta+\gamma$$

$$Var(X)=\alpha+\gamma$$

$$Var(Y)=\beta+\gamma$$

$$E(XY)=\gamma+(\alpha+\gamma)(\beta+\gamma)$$

Teorema berikut dapat ditetapkan dengan terlebih dahulu menghitung fungsi kepadatan probabilitas bersyarat

Teorema 4.8.

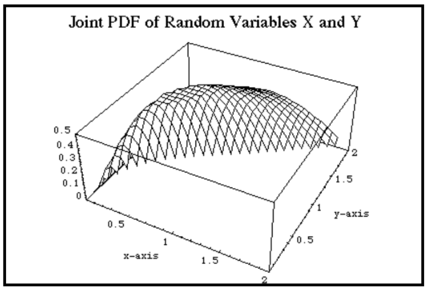
Jika $(X,Y)\sim GAMC(\alpha,\beta,\gamma)$, maka

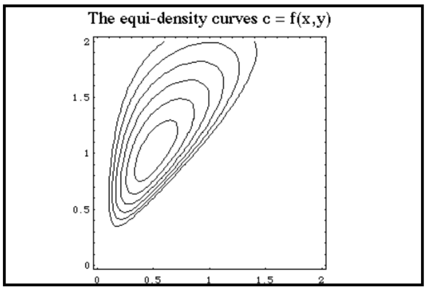
$$E(Y\mid x)=\beta+\frac{\gamma}{\alpha+\gamma}\quad dan\quad E(X\mid y)=\alpha+\frac{\gamma}{\beta+\gamma}y$$

Pada tahun 1934, McKay memberikan distribusi gamma bivariat lain yang fungsi kepadatan probabilitasnya berbentuk

$$f(x,y)=\begin{cases}\dfrac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(y-x)^{\beta-1}e^{-\theta y}, & untuk\ 0<x<y<\infty\\ 0, & untuk\ yang\ lainnya\end{cases}$$

di mana $\theta,\alpha,\beta\in(0,\infty)$ adalah parameter. Jika berupa kepadatan gabungan variabel acak (X,Y) mirip dengan fungsi kepadatan dari distribusi gamma bivariat dari McKay, kemudian kita tulis $(X,Y)\sim GAMM(\theta,\alpha,\beta)$. Grafik fungsi kepadatan probabilitas $f(x,y)$ dari distribusi gamma bivariat McKay untuk $\theta=\alpha=\beta=2$ ditunjukkan di bawah ini. Gambar lainnya mengilustrasikan kurva densitas ekuivalen dari fungsi densitas gabungan $f(x,y)$.





Dapat ditunjukkan bahwa jika $(X,Y)\sim GAMM(\theta,\alpha,\beta)$, maka marginal $f_1(x)$ dari X dan marginal $f_2(y)$ dari Y diberikan oleh

$$f_1(x)=\begin{cases}\dfrac{\theta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\theta x}, & untuk\ 0\le x\le\infty\\ 0 & untuk\ yang\ lainnya\end{cases}$$

Dan

$$f_2(y)=\begin{cases}\dfrac{\theta^{\alpha+\beta}}{\Gamma(\alpha+\beta)}x^{\alpha+\beta-1}e^{-\theta x}, & untuk\ 0\le x\le\infty\\ 0 & untuk\ yang\ lainnya\end{cases}$$

Karenanya $X\sim GAM\left(\alpha,\dfrac{1}{\theta}\right)$ dan $Y\sim GAM\left(\alpha+\beta,\dfrac{1}{\theta}\right)$.

Teorema 4.9.

Jika $(X,Y)\sim GAMM(\theta,\alpha,\beta)$ maka

$$E(X)=\frac{\alpha}{\theta}$$

$$E(Y)=\frac{\alpha+\beta}{\theta}$$

$$Var(X)=\frac{\alpha}{\theta^2}$$

$$Var(X)=\frac{\alpha+\beta}{\theta^2}$$

$$M(s,t)=\left(\frac{\theta}{\theta-s-t}\right)^{\alpha}\left(\frac{\theta}{\theta-t}\right)^{\beta}$$

Kami menyatakan berbagai sifat kepadatan bersyarat $f(x,y)$, tanpa bukti, dalam teorema berikut.

Teorema 4.10.

Jika $(X,Y)\sim GAMM(\theta,\alpha,\beta)$ maka

$$E(Y\,|x)=x+\frac{\beta}{\theta}$$

$$E(X\,|y)=\frac{\alpha\beta}{\alpha+\beta}$$

$$Var(Y\,|x)=\frac{\beta}{\theta^2}$$

$$Var(X\,|y)=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}y^2$$

Kita tahu bahwa distribusi eksponensial univariat adalah kasus khusus dari distribusi gamma univariat. Begitu pula dengan distribusi eksponensial bivariat adalah kasus khusus dari distribusi gamma bivariat. Saat mengambil parameter indeks menjadi kesatuan dalam distribusi gamma bivariat Kibble dan Cherian yang diberikan di atas, kita memperoleh distribusi eksponensial bivariat. Sesuai fungsi kepadatan probabilitas eksponensial bivariat untuk bivariat distribusi gamma dari Kibble diberikan oleh

$$f(x,y)=\begin{cases} e^{-\left(\frac{x+y}{1-y}\right)}\sum_{k=0}^{\infty}\dfrac{(\theta xy)^k}{k!\Gamma(k+1)(1-\theta)^{2k+1}}, & \textit{untuk } 0<x,y<\infty \\ 0, & \textit{untuk yang lainnya} \end{cases}$$

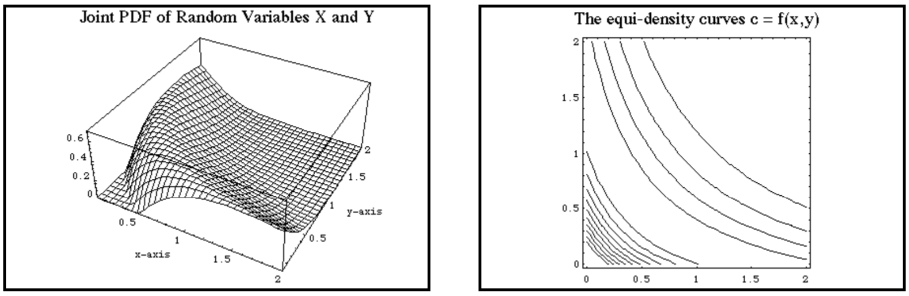
di mana $\theta\in(0,1)$ adalah parameter. Distribusi eksponensial bivariat yang sesuai untuk distribusi bivariat Cherian adalah sebagai berikut :

$$f(x,y)=\begin{cases} \left[e^{\min\{x,y\}}-1\right]e^{-(x+y)}, & \textit{untuk } 0<x,y<\infty \\ 0, & \textit{untuk yang lainnya} \end{cases}$$

Pada tahun 1960, Gumble telah mempelajari distribusi eksponensial bivariat berikut yang fungsi kepadatannya diberikan oleh

$$f(x,y)=\begin{cases} \left[(1+\theta x)(1+\theta y)-\theta\right]e^{-(x+y+\theta xy)}, & \textit{untuk } 0<x,y<\infty \\ 0, & \textit{untuk yang lainnya} \end{cases}$$

di mana $\theta>0$ adalah parameter



Pada tahun 1967, Marshall dan Olkin memperkenalkan distribusi eksponensial bivariat berikut

$$f(x,y)=\begin{cases} 1-e^{-(\alpha+\gamma)x}-e^{-(\beta+\gamma)y}+e^{-(\alpha x+\beta y+\gamma\max\{x,y\})}, & \textit{untuk } x,y<0 \\ 0, & \textit{untuk yang lainnya} \end{cases}$$

di mana $\alpha,\beta,\gamma>0$ adalah parameter. Fungsi distribusi eksponensial dari Marshall dan Olkin memenuhi sifat kekurangan memori

$$P(X>x+t,Y>y+t\,/X>t,Y>t)=P(X>x,Y>y).$$

4.4. Distribusi Beta Bivariat

Distribusi beta bivariat (juga dikenal sebagai distribusi Dirichlet) merupakan salah satu distribusi dasar dalam statistik. Distribusi beta bivariat digunakan dalam geologi, biologi, dan kimia untuk menangani data komposisi yang didasarkan pada batasan nonnegativitas dan jumlah konstan. Distribusi ini juga digunakan saat ini dengan meningkatnya frekuensi dalam pemodelan statistik, teori distribusi dan statistik Bayesian. Misalnya, digunakan dalam mencontohkan proporsi pemilih yang memilih calon di sebuah pemilihan dua kandidat. Dalam statistik Bayesian, distribusi beta sangat popular sebagai prior karena menghasilkan distribusi beta sebagai posterior. Di dalam bagian ini, diberikan beberapa fakta dasar tentang distribusi beta bivariat.

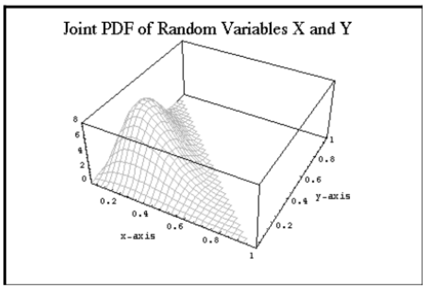
Definisi 4.5.

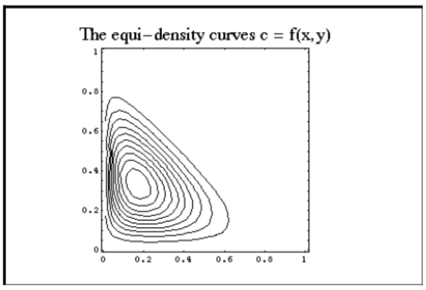
Sebuah variabel acak bivariat kontinu (X,Y) dikatakan memiliki distribusi beta bivariat jika fungsi kepadatan probabilitas gabungannya dalam bentuk

$$f(x,y)=\begin{cases}\dfrac{\Gamma(\theta_1+\theta_2+\theta_3)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)}x^{\theta_1-1}y^{\theta_2-1}(1-x-y)^{\theta_3-1}, & \text{untuk } 0<x,y,x+1<1\\ 0, & \text{untuk yang lainnya}\end{cases}$$

Dimana $\theta_1,\theta_2,\theta_3$ adalah parameter positif. Kita akan menunjukkan variabel acak beta bivariat (X,Y) dengan parameter positif $\theta_1,\theta_2,\theta_3$ dengan cara menuliskan $(X,Y)\sim Beta(\theta_1,\theta_2,\theta_3)$.

Gambar berikut menunjukkan grafik dan kurva kepadatan ekuivalen $f(x,y)$ pada domain dari definisinya.





Dalam teorema berikut, diberikan nilai ekspektasi, varians variabel acak X dan Y, dan korelasi antara X dan Y

Teorema 4.11.

Jika $(X,Y)\sim Beta(\theta_1,\theta_2,\theta_3)$, di mana $\theta_1,\theta_2,\theta_3$ adalah parameter yang dipilih apriori positif, maka $X\sim Beta(\theta_1,\theta_2+\theta_3)$ dan $Y\sim Beta(\theta_2,\theta_1+\theta_3)$ dan

$$E(X)=\frac{\theta_1}{\theta},\qquad Var(X)=\frac{\theta_1(\theta-\theta_1)}{\theta^2(\theta+1)}$$

$$E(Y)=\frac{\theta_2}{\theta},\qquad Var(Y)=\frac{\theta_2(\theta-\theta_2)}{\theta^2(\theta+1)}$$

$$Cov(X,Y) = -\frac{\theta_1\theta_2}{\theta^2(\theta+1)}$$

dimana $\theta = \theta_1 + \theta_2 + \theta_3$.

Bukti :

Pertama, kita tunjukkan bahwa $X \sim Beta(\theta_1, \theta_2 + \theta_3)$ dan $Y \sim Beta(\theta_2, \theta_1 + \theta_3)$. Karena $(X,Y) \sim Beta(\theta_2, \theta_1, \theta_3)$, kepadatan gabungan dari (X,Y) diberikan oleh

$$f(x,y) = \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} x^{\theta_1 - 1} y^{\theta_2 - 1} (1-x-y)^{\theta_3 - 1},$$

dimana $\theta = \theta_1 + \theta_2 + \theta_3$.

Kemudian kepadatan marginal dari X diberikan oleh

$$f_1(x) = \int_0^1 f(x,y)\,dy$$

$$= \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} x^{\theta_1 - 1} \int_0^{1-x} y^{\theta_2 - 1} (1-x-y)^{\theta_3 - 1}\, dy$$

$$= \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} x^{\theta_1 - 1} (1-x)^{\theta_3 - 1} \int_0^{1-x} y^{\theta_2 - 1} \left(1 - \frac{y}{1-x}\right)^{\theta_3 - 1} dy$$

Sekarang kita substitusikan $u = 1 - \frac{y}{1-x}$ dalam integral di atas. Kemudian kita dapatkan

$$f_1(x) = \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} x^{\theta_1 - 1} (1-x)^{\theta_2 + \theta_3 - 1} \int_0^1 u^{\theta_2 - 1} (1-u)^{\theta_3 - 1}\, du$$

$$= \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} x^{\theta_1 - 1} (1-x)^{\theta_2 + \theta_3 - 1} B(\theta_2, \theta_3)$$

$$= \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} x^{\theta_1 - 1}(1-x)^{\theta_2+\theta_3-1}$$

Sehingga

$$\int_0^1 u^{\theta_2-1}(1-u)^{\theta_3-1}\,du = B(\theta_2,\theta_3) = \frac{\Gamma(\theta_2)\Gamma(\theta_3)}{\Gamma(\theta_2+\theta_3)}$$

Ini membuktikan bahwa variabel acak $X \sim Beta(\theta_1, \theta_2+\theta_3)$. Demikian pula, kita dapat menunjukkan bahwa variabel acak $Y \sim Beta(\theta_2, \theta_1+\theta_3)$. Sekarang menggunakan Teorema 6.5, kita tahu bahwa

$$E(X) = \frac{\theta_1}{\theta}, \qquad Var(X) = \frac{\theta_1(\theta-\theta_1)}{\theta^2(\theta+1)}$$

$$E(Y) = \frac{\theta_2}{\theta}, \qquad Var(Y) = \frac{\theta_2(\theta-\theta_2)}{\theta^2(\theta+1)}$$

dimana $\theta = \theta_1+\theta_2+\theta_3$.

Selanjutnya, kita menghitung hasil kali momen dari X dan Y.

$$E(XY) = \int_0^1 \left[\int_0^{1-x} xy\, f(x,y)\,dy\right] dx$$

$$= \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} \int_0^1 \left[\int_0^{1-x} xy\, x^{\theta_1-1} y^{\theta_2-1}(1-x-y)^{\theta_3-1}\,dy\right] dx$$

$$= \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} \int_0^1 \left[\int_0^{1-x} x^{\theta_1} y^{\theta_2}(1-x-y)^{\theta_3-1}\,dy\right] dx$$

$$= \frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} \int_0^1 x^{\theta_1}(1-x)^{\theta_3-1} \left[\int_0^{1-x} y^{\theta_2}\left(1-\frac{y}{1-x}\right)^{\theta_3-1} dy\right] dx$$

Sekarang kita substitusikan $u = \dfrac{y}{1-x}$ ke dalam integral di atas

$$E(XY)=\frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)}\int_0^1 x^{\theta_1}(1-x)^{\theta_2+\theta_3}\left[\int_0^{1-x} u^{\theta_2}(1-u)^{\theta_3-1}\,du\right]dx$$

Karena

$$\int_0^{1-x} u^{\theta_2}(1-u)^{\theta_3-1}\,du=B(\theta_2+1,\theta_3)$$

Dan

$$\int_0^1 x^{\theta_1}(1-x)^{\theta_2+\theta_3}\,dx=B(\theta_1+1,\theta_2+\theta_3+1)$$

Kita mempunyai

$$E(XY)=\frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)}B(\theta_2+1,\theta_3)B(\theta_1+1,\theta_2+\theta_3+1)$$

$$=\frac{\Gamma(\theta)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)}\frac{\theta_1\Gamma(\theta_1)(\theta_2+\theta_3)\Gamma(\theta_2+\theta_3)}{\theta(\theta+1)\Gamma(\theta)}\frac{\theta_2\Gamma(\theta_2)\Gamma(\theta_3)}{(\theta_2+\theta_3)\Gamma(\theta_2+\theta_3)}$$

$$=\frac{\theta_1\theta_2}{\theta(\theta+1)}\qquad \text{dimana } \theta=\theta_1+\theta_2+\theta_3$$

Sekarang kita menghitung kovarians X dan Y

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

$$=\frac{\theta_1\theta_2}{\theta(\theta+1)}-\frac{\theta_1}{\theta}\frac{\theta_2}{\theta}$$

$$=\frac{\theta_1\theta_2}{\theta^2(\theta+1)}$$

Koefisien korelasi X dan Y dapat dihitung menggunakan kovarian sebagai berikut

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = -\sqrt{\frac{\theta_1\theta_2}{(\theta_1+\theta_2)(\theta_2+\theta_3)}}$$

Teorema berikutnya menyatakan beberapa sifat dari fungsi kepadatan bersyarat $f(x\,|\,y)$ dan $f(y|x)$.

Teorema 4.12.

Jika $(X,Y) \sim Beta(\theta_1,\theta_2,\theta_3)$, dimana $\theta_1,\theta_2,\theta_3$ adalah parameter positif. Maka

$$E(Y\,|x) = \frac{\theta_2(1-x)}{\theta_2+\theta_3}, \qquad Var(Y\,|x) = \frac{\theta_2\theta_3(1-x)^2}{(\theta_2+\theta_3)^2(\theta_2+\theta_3+1)}$$

$$E(X\,|y) = \frac{\theta_1(1-y)}{\theta_1+\theta_3}, \qquad Var(X\,|y) = \frac{\theta_1\theta_3(1-y)^2}{(\theta_1+\theta_3)^2(\theta_1+\theta_3+1)}$$

Bukti :

Kita tahu bahwa jika $(X,Y) \sim Beta(\theta_1,\theta_2,\theta_3)$, variabel acak $X \sim Beta(\theta_1,\theta_2+\theta_3)$. Maka

$$f(y\,|x) = \frac{f(x,y)}{f_1(x)}$$

$$= \frac{1}{1-x}\frac{\Gamma(\theta_2+\theta_3)}{\Gamma(\theta_2)\Gamma(\theta_3)}\left(\frac{y}{1-x}\right)^{\theta_2-1}\left(1-\frac{y}{1-x}\right)^{\theta_3-1}$$

Untuk setiap $0<y<1-x$. Kemudian variabel acak $\left.\frac{Y}{1-x}\right|_{X=x}$ adalah variabel acak beta dengan parameter θ_2 dan θ_3.

Sekarang kita menghitung ekspektasi bersyarat dari $Y\,|x$.

$$E(Y\,|x) = \int_0^{1-x} y\,f(y\,|x)\,dy$$

$$= \frac{1}{1-x}\frac{\Gamma(\theta_2+\theta_3)}{\Gamma(\theta_2)\Gamma(\theta_3)}\left(\frac{y}{1-x}\right)^{\theta_2-1}\left(1-\frac{y}{1-x}\right)^{\theta_3-1}$$

Sekarang kita substitusikan $u = \dfrac{y}{1-x}$ ke dalam integral di atas

$$E(Y \mid x) = \frac{\Gamma(\theta_2+\theta_3)}{\Gamma(\theta_2)\Gamma(\theta_3)}(1-x)\int_0^1 u^{\theta_2}(1-u)^{\theta_3-1}\,du$$

$$= \frac{\Gamma(\theta_2+\theta_3)}{\Gamma(\theta_2)\Gamma(\theta_3)}(1-x)B(\theta_2+1,\theta_3)$$

$$= \frac{\Gamma(\theta_2+\theta_3)}{\Gamma(\theta_2)\Gamma(\theta_3)}(1-x)\frac{\theta_2\Gamma(\theta_2)\Gamma(\theta_3)}{(\theta_2+\theta_3)\Gamma(\theta_2+\theta_3)}$$

$$= \frac{\theta_2}{\theta_2+\theta_3}(1-x)$$

Selanjutnya kita hitung $E(Y^2 \mid x)$.

$$E(Y^2 \mid x) = \int_0^{1-x} y^2 f(y|x)dy$$

$$= \frac{1}{1-x}\frac{\Gamma(\theta_2+\theta_3)}{\Gamma(\theta_2)\Gamma(\theta_3)}\int_0^{1-x} y^2\left(\frac{y}{1-x}\right)^{\theta_2-1}\left(1-\frac{y}{1-x}\right)^{\theta_3-1}dy$$

$$= \frac{\Gamma(\theta_2+\theta_3)}{\Gamma(\theta_2)\Gamma(\theta_3)}(1-x)^2\int_0^1 u^{\theta_2+1}(1-u)^{\theta_3-1}\,du$$

$$= \frac{\Gamma(\theta_2+\theta_3)}{\Gamma(\theta_2)\Gamma(\theta_3)}(1-x)^2 B(\theta_2+2,\theta_3)$$

$$= \frac{\Gamma(\theta_2+\theta_3)}{\Gamma(\theta_2)\Gamma(\theta_3)}(1-x)^2\frac{(\theta_2+1)\theta_2\,\Gamma(\theta_2)\Gamma(\theta_3)}{(\theta_2+\theta_3+1)(\theta_2+\theta_3)\Gamma(\theta_2+\theta_3)}$$

$$= \frac{(\theta_2+1)\theta_2}{(\theta_2+\theta_3+1)(\theta_2+\theta_3)}(1-x)^2$$

$$Var\left(Y\mid x\right)=E(Y^2|x)-E(Y|x)^2\,\frac{\theta_2\theta_3\left(1-x\right)^2}{\left(\theta_2+\theta_3\right)^2\left(\theta_2+\theta_3+1\right)}$$

Distribusi Dirichlet dapat diperluas dari satuan persegi $\left(0,1\right)^2$ ke persegi panjang sembarang $\left(a_{1,}b_1\right)\times\left(a_{2,}b_2\right)$

Definisi 4.6.

Sebuah variabel acak bivariat kontinu $\left(X_{1,}X_2\right)$ dikatakan memiliki distribusi beta bivariat umum jika fungsi kepadatan probabilitas gabungannya berbentuk

$$f\left(x_1,x_2\right)=\frac{\Gamma\left(\theta_1+\theta_2+\theta_3\right)}{\Gamma\left(\theta_1\right)\Gamma\left(\theta_2\right)\Gamma\left(\theta_3\right)}\prod_{k=1}^{2}\left(\frac{x_k-a_k}{b_k-a_k}\right)^{\theta_k-1}\left(1-\frac{x_k-a_k}{b_k-a_k}\right)^{\theta_3-1}$$

di mana $0<x_1,x_2,x_1+x_2<1$ dan $\theta_1,\theta_2,\theta_3,a_1,b_1,a_2,b_2$ adalah parameter. Kita akan menunjukkan variabel acak beta umum bivariat $\left(X,Y\right)$ dengan parameter positif θ_1,θ_2 dan θ_3 dengan menuliskan $\left(X,Y\right)\sim GBeta\left(\theta_1,\theta_2,\theta_3,a_1,b_1,a_2,b_2\right)$

Dapat ditunjukkan bahwa jika $X_k=\left(b_k-a_k\right)Y_k+a_k$ (untuk $k=1,2$) dan masing-masing $\left(Y_1,Y_2\right)\sim Beta\left(\theta_1,\theta_2,\theta_3\right)$, lalu $\left(X_1,X_2\right)\sim GBeta\left(\theta_1,\theta_2,\theta_3,a_1,b_1,a_2,b_2\right)$

Teorema 4.13.

Jika $\left(X,Y\right)\sim GBeta(\theta_1,\theta_2,\theta_3,a_1,b_1,a_2,b_2)$, di mana θ_1,θ_2 dan θ_3 adalah parameter terpilih apriori positif. Maka $X\sim Beta\left(\theta_1,\theta_2+\theta_3\right)$ dan $Y\sim Beta\left(\theta_2,\theta_1+\theta_3\right)$ dan

$$E\left(X\right)=\left(b_1-a_1\right)\frac{\theta_1}{\theta}+a_1,\qquad Var\left(X\right)=\left(b_1-a_1\right)^2\frac{\theta_1\left(\theta-\theta_1\right)}{\theta^2\left(\theta+1\right)}$$

$$E\left(Y\right)=\left(b_2-a_2\right)\frac{\theta_2}{\theta}+a_2,\qquad Var\left(Y\right)=\left(b_2-a_2\right)^2\frac{\theta_2\left(\theta-\theta_2\right)}{\theta^2\left(\theta+1\right)}$$

$$Cov\left(X,Y\right)=-\left(b_1-a_1\right)\left(b_2-a_2\right)\frac{\theta_1\theta_2}{\theta^2\left(\theta+1\right)}$$

Dimana $\theta=\theta_1+\theta_2+\theta_3$.

Generalisasi lain dari distribusi beta bivariat adalah sebagai berikut:

Definisi 4.7.

Sebuah variabel acak bivariat kontinu (X_1, X_2) dikatakan memiliki distribusi beta bivariat umum jika kepadatan probabilitas gabungan fungsinya berbentuk

$$f(x_1,x_2)=\frac{1}{B(\alpha_1,\beta_1)B(\alpha_2,\beta_2)}x^{\alpha_1-1}(1-x_1)^{\beta_1-\alpha_2-\beta_2}\,x_2^{\alpha_2-1}(1-x_1x_2)^{\beta_2-1}$$

dimana $0<x_1,x_2,x_1+x_2<1$ dan $\alpha_1,\alpha_2,\beta_1,\beta_2$ adalah parameter.

4.5. Distribusi Normal Bivariat

Distribusi normal bivariat merupakan generalisasi dari distribusi normal univariat. Pada bagian ini, diberikan perlakuan mendalam tentang distribusi bivariat normal.

Definisi 4.8.

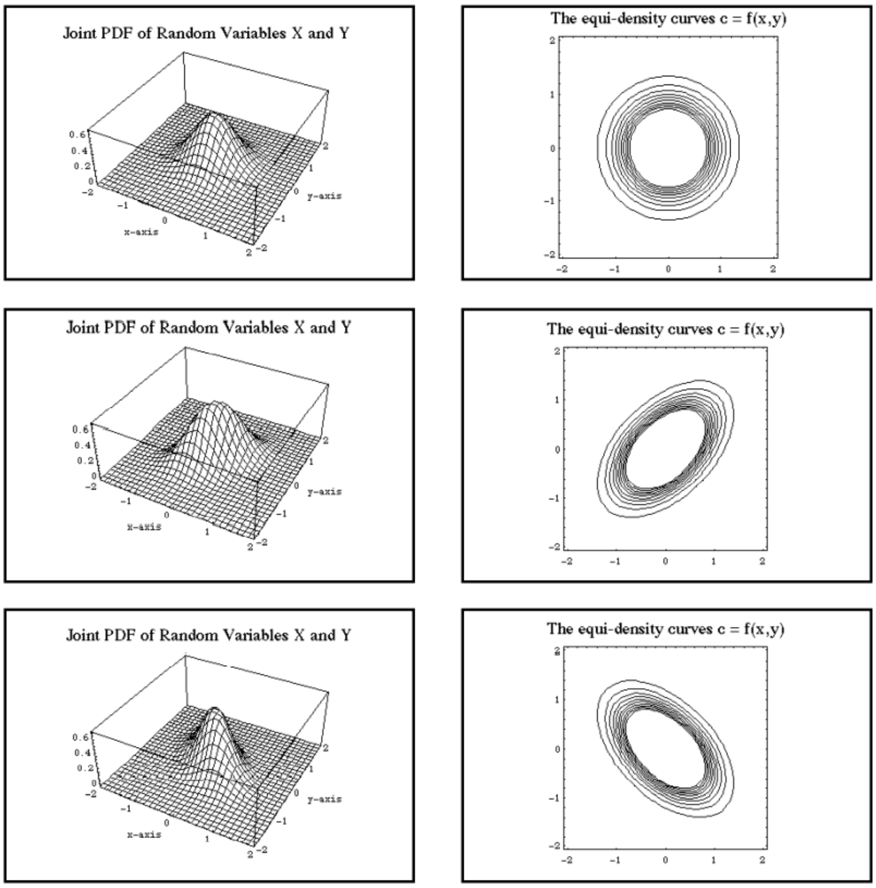
Sebuah variabel acak bivariat kontinu (X,Y) dikatakan memiliki distribusi normal bivariat jika fungsi kepadatan probabilitas gabungannya berbentuk

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2}Q(x,y)} \qquad ,-\infty<x,y<\infty$$

dimana $\mu_1,\mu_2\in\mathbb{R}$, $\sigma_1,\sigma_2\in(0,\infty)$ dan $\rho\in(-1,1)$ adalah parameter, dan

$$Q(x,y):=\frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

Seperti biasa, kita menunjukkan variabel acak normal bivariat ini dengan menulis $(X,Y)\sim N\ (\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$. Grafik $f(x,y)$ berbentuk "gunung". Pasangan (μ_1,μ_2) memberi tahu kita di mana letak pusat gunung dalam bidang (x,y), sedangkan σ_1^2 dan σ_2^2 mengukur penyebaran gunung ini pada arah x dan arah y, masing-masing. Parameter ρ menentukan bentuk dan orientasi pada bidang (x,y) dari gunung. Mengikuti gambar menunjukkan grafik distribusi normal bivariat dengan perbedaan nilai koefisien korelasi ρ. Dua gambar pertama menggambarkan grafik distribusi normal bivariat dengan $\rho=0,\mu_1=\mu_2=0$, dan $\sigma_1=\sigma_2=1$ dan plot equi-density. Dua gambar berikut mengilustrasikan grafik dari distribusi normal bivariat dengan $\rho=0,5$, $\mu_1=\mu_2=0$, dan $\sigma_1=\sigma_2=0,5$ dan plot equi-density. Dua gambar terakhir mengilustrasikan grafik dari distribusi normal bivariat dengan $\rho=-0,5$, $\mu_1=\mu_2=0$, dan $\sigma_1=\sigma_2=0,5$ dan plot equi-density.



Salah satu ciri khusus dari distribusi normal bivariat yaitu jika secara vertikal mengiris grafik $f(x,y)$ sepanjang segala arah, maka akan didapatkan distribusi normal univariat. Secara khusus, jika kita memotong grafik secara vertikal dari $f(x,y)$ sepanjang sumbu x, maka didapatkan distribusi normal univariat. Artinya marginal dari $f(x,y)$ kembali normal. Kita dapat menunjukkan bahwa margin dari $f(x,y)$ diberikan oleh

$$f_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}$$

Dan

$$f_2(y) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}$$

Teorema 4.14.

Jika $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1, \sigma_{2,} \rho)$, kemudian

$$E(X) = \mu_1$$

$$E(Y) = \mu_2$$

$$Var(X) = \sigma_1^2$$

$$Var(Y) = \sigma_2^2$$

$$Corr(X,Y) = \rho$$

.

$$M(s,t) = e^{\mu_1 s + \mu_2 t + \frac{1}{2}\left(\sigma_1^2 s^2 + 2\rho\sigma_1\sigma_2 st + \sigma_2^2 t^2\right)}$$

Bukti :

Mudah membuktikan formula untuk $E(X), E(Y), Var(X)$ dan $Var(Y)$ Kita hanya akan membuktikan fungsi pembangkit momen. Berdasarkan $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, kita memiliki $X \sim N\ (\mu_{1,}\sigma_1^2)$ dan $Y \sim N$ (μ_2, σ_2^2). Selanjutnya, untuk setiap s dan t, variabel acak $W = sX + tY$ kembali normal dengan

$$\mu w = s\mu_1 + t\mu_2 \qquad \text{dan} \qquad \sigma_w^2 = s^2\sigma_1^2 + 2st$$

Karena W adalah variabel acak normal, fungsi pembangkit momennya yaitu

$$M(\tau) = e^{\mu w \tau + \frac{1}{2}\tau^2\sigma^2 w}$$

Fungsi pembangkit momen gabungan (X,Y) adalah

$$M(s,t) = E\left(e^{sX+tY}\right)$$

$$= e^{\mu w + \frac{1}{2}\sigma^2 w}$$

$$= e^{\mu_1 s + \mu_2 t + \frac{1}{2}\left(\sigma_1^2 s^2 + 2\rho\sigma_1\sigma_2 st + \sigma_2^2 t^2\right)}$$

Dapat ditunjukkan bahwa kepadatan bersyarat Y dengan syarat $X = x$ adalah

$$f(y|x) = \frac{1}{\sigma_2\sqrt{2\pi\left(1-\rho^2\right)}} e^{-\frac{1}{2}\left(\frac{y-b}{\sigma_2\sqrt{1-\rho^2}}\right)^2}$$

Dimana

$$b = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}\left(x-\mu_1\right)$$

Demikian pula, kepadatan bersyarat $f(x|y)$ adalah

$$f(x|y) = \frac{1}{\sigma_1\sqrt{2\pi\left(1-\rho^2\right)}} e^{-\frac{1}{2}\left(\frac{x-c}{\sigma_1\sqrt{1-\rho^2}}\right)}$$

Dimana

$$c = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}\left(y-\mu_2\right)$$

Mengingat bentuk $f(y|x)$ dan $f(x|y)$, teorema berikut ini sudah jelas.

Teorema 4.15.

Jika $\left(X,Y\right) \sim N\left(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho\right)$, kemudian

$$E(Y|x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}\left(x-\mu_1\right)$$

$$E(X|y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}\left(y-\mu_2\right)$$

$$Var(Y|x) = \sigma_2^2\left(1-\rho^2\right)$$

$$Var(X|y) = \sigma_2^2\left(1-\rho^2\right)$$

Kita dapat melihat bahwa jika $\left(X,Y\right)$ adalah distribusi normal bivariat, sehingga distribusi dari X dan Y adalah juga normal. Namun, kebalikannya adalah tidak benar. Yaitu jika X dan Y memiliki distribusi normal sebagai marginnya, maka distribusi gabungannya belum tentu normal bivariat. Sekarang diberikan beberapa teorema karakterisasi tentang distribusi normal bivariat. Teorema pertama berasal dari Cramer (1941).

Teorema 4.16.

Variabel acak X dan Y memiliki distribusi normal gabungan bivariat jika dan hanya jika setiap kombinasi linier dari X dan Y memiliki distribusi normal univariat.

Teorema 4.17.

Variabel acak X dan Y dengan satuan varians dan koefisien korelasi ρ memiliki distribusi normal bivariat gabungan jika dan hanya jika

$$\frac{\partial}{\partial p}E\left[g\left(X,Y\right)\right]=E\left[\frac{\partial^2}{\partial X\partial Y}g\left(X,Y\right)\right]$$

memiliki fungsi arbitrer $g\left(x,y\right)$ dari dua variabel.

4.6. Distribusi Logistik Bivariat

Pada bagian ini, dipelajari dua distribusi logistik bivariat. Sebuah Distribusi logistik univariat sering dianggap sebagai alternatif dari distribusi normal univariat. Distribusi logistik univariat memiliki bentuk yang sangat mendekati distribusi normal univariat. Distribusi ini juga digunakan sebagai alternatif Distribusi Weibull univariat dalam pengujian kehidupan. Distribusi logistik univariat memiliki fungsi kepadatan probabilitas berikut

$$f\left(x\right)=\frac{\pi}{\sigma\sqrt{3}}\frac{e^{-\frac{\pi}{\sqrt{3}}\left(\frac{x-\mu}{\sigma}\right)}}{\left[1+e^{-\frac{\pi}{\sqrt{3}}\left(\frac{x-\mu}{\sigma}\right)}\right]^2}\qquad -\infty<x<\infty,$$

di mana $-\infty<\mu<\infty$ dan $\sigma>0$ adalah parameter. Parameter μ adalah mean dan parameter σ adalah standar deviasi dari distribusi. Sebuah variabel acak X dengan distribusi logistik di atas akan dilambangkan dengan $X\sim LOG\left(\mu,\sigma\right)$. Diketahui bahwa fungsi pembangkit momen distribusi logistik univariat yaitu

$$M\left(t\right)=e^{\mu t}\Gamma\left(1+\frac{\sqrt{3}}{\pi}\sigma t\right)\Gamma\left(1-\frac{\sqrt{3}}{\pi}\sigma t\right)$$

Untuk $|t| < \frac{\pi}{\sigma\sqrt{3}}$, diberikan bukti singkat dari hasil di atas untuk $\mu = 0$

dan $\sigma = \frac{\pi}{\sqrt{3}}$. Kemudian dengan asumsi tersebut, fungsi kepadatan logistik berkurang menjadi

$$f(x) = \frac{e^{-x}}{\left(1+e^{-x}\right)^2}$$

Fungsi pembangkit momen sehubungan dengan fungsi kerapatan ini adalah

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

$$= \int_{-\infty}^{\infty} e^{tx} \frac{e^{-x}}{\left(1+e^{-1}\right)^2} dx$$

$$= \int_{-\infty}^{\infty} \left(e^{-x}\right)^{-t} \frac{e^{-x}}{\left(1+e^{-1}\right)^2} dx$$

$$= \int_0^1 \left(z^{-1}-1\right)^{-t} dz \quad dimana\, z = \frac{1}{1+e^{-x}}$$

$$= \int_0^1 z^t \left(1-z\right)^{-t} dz$$

$$= B\left(1+t, 1-t\right)$$

$$= \frac{\Gamma(1+t)\Gamma(1-t)}{\Gamma(1+t+1-t)}$$

$$= \frac{\Gamma(1+t)\Gamma(1-t)}{\Gamma(2)}$$

$$=\Gamma(1+t)\Gamma(1-t)$$

$$=t\,cosec(t)$$

Ingatlah bahwa marginal dan persyaratan dari distribusi normal bivariat adalah normal univariat. Sifat distribusi normal bivariat ini ternyata kurang dari distribusi bivariat lain yang telah dibahas selama ini. Jika kita tidak dapat mendefinisikan distribusi logistik bivariat sehingga syarat dan marginal merupakan logistik univariat, maka setidaknya memiliki satu dari distribusi logistik marginal dan distribusi bersyarat dari variabel logistik lainnya

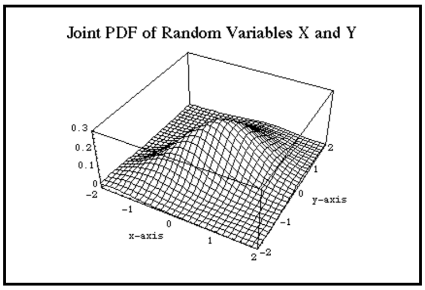
Definisi 4.9.

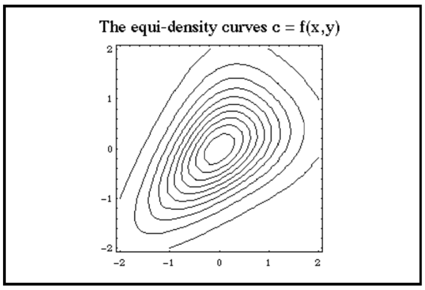
Sebuah variabel acak bivariat kontinu (X,Y) dikatakan memiliki distribusi logistik bivariat jenis pertama jika probabilitas gabungannya fungsi kerapatan berbentuk

$$f(x,y)=\frac{2\pi^2 e^{-\frac{\pi}{\sqrt{3}}\left(\frac{x-\mu_1}{\sigma_1}+\frac{y-\mu_2}{\sigma_2}\right)}}{3\sigma_1\sigma_{2\left[1+e^{-\frac{\pi}{\sqrt{3}}\left(\frac{x-\mu_1}{\sigma_1}\right)}+e^{-\frac{\pi}{\sqrt{3}}\left(\frac{y-\mu_2}{\sigma_2}\right)}\right]}}$$

di mana $-\infty<\mu_1,\mu_2<\infty$, dan $0<\sigma_1,\sigma_2<\infty$ adalah parameter.

Jika acak variabel (X,Y) memiliki distribusi logistik bivariat jenis pertama, maka dapat ditulis $(X,Y)\sim LOGF$ $(\mu_1,\mu_2,\sigma_1,\sigma_2)$. Gambar-gambar berikut menunjukkan grafik $f(x,y)$ dengan $\mu_1=0=\mu_2$ dan $\sigma_1=1=\sigma_2$ serta plot kesetaraan.





Dapat ditunjukkan bahwa secara garis besar, X adalah variabel acak logistik yaitu $X\sim LOG(\mu_1,\sigma_1)$. Demikian pula, $Y\sim LOG(\mu_2,\sigma_2)$. Fakta-fakta ini mengarahkan pada teorema berikut.

Teorema 4.18. Jika variabel acak $(X,Y) \sim LOGF(\mu_1,\mu_2,\sigma_1,\sigma_2)$,

Maka

$$E(X) = \mu_1$$

$$E(Y) = \mu_2$$

$$Var(X) = \sigma_1^2$$

$$Var(Y) = \sigma_2^2$$

$$E(XY) = \frac{1}{2}\sigma_1\sigma_2 + \mu_1\mu_2$$

dan fungsi pembangkit momen diberikan oleh

$$M(s,t) = e^{\mu_1 s + \mu_2 t}\Gamma\left(1 + \frac{(\sigma_1 s + \sigma_2 t)}{\pi}\right)\Gamma\left(1 - \frac{\sigma_1 s\sqrt{3}}{\pi}\right)\Gamma\left(1 - \frac{\sigma_2 t\sqrt{3}}{\pi}\right)$$

Untuk $|s| < \frac{\pi}{\sigma_1\sqrt{3}}$ dan $|t| < \frac{\pi}{\sigma_2\sqrt{3}}$.

Ini adalah latihan yang mudah untuk menghitung jika variabel acak X dan Y berdistribusi logistik bivariat gabungan, kemudian korelasi antara X dan Y adalah $\frac{1}{2}$. Ini dapat dianggap sebagai salah satu kelemahan dari distribusi ini, artinya membatasi ketergantungan antara variabel acak X dan Y.

Kepadatan bersyarat dari Y dengan syarat $X = x$ adalah

$$f(y|x) = \frac{2\pi}{\sigma_2\sqrt{3}} e^{-\frac{\pi}{\sqrt{3}}\left(\frac{y-\mu_2}{\sigma_2}\right)} \frac{\left[1 + e^{-\frac{\pi}{\sqrt{3}}\left(\frac{x-\mu_1}{\sigma_1}\right)}\right]^2}{\left[1 + e^{-\frac{\pi}{\sqrt{3}}\left(\frac{x-\mu_1}{\sigma_1}\right)} + e^{-\frac{\pi}{\sqrt{3}}\left(\frac{y-\mu_2}{\sigma_2}\right)}\right]^3}$$

Demikian pula kepadatan kondisional X dengan syarat $Y = y$

$$f(x|y) = \frac{2\pi}{\sigma_1\sqrt{3}} e^{-\frac{\pi}{\sqrt{3}}\left(\frac{x-\mu_1}{\sigma_1}\right)} \frac{\left[1+e^{-\frac{\pi}{\sqrt{3}}\left(\frac{y-\mu_2}{\sigma_2}\right)}\right]^2}{\left[1+e^{-\frac{\pi}{\sqrt{3}}\left(\frac{x-\mu_1}{\sigma_1}\right)}+e^{-\frac{\pi}{\sqrt{3}}\left(\frac{y-\mu_2}{\sigma_2}\right)}\right]^3}$$

Menggunakan kepadatan ini, teorema berikutnya menjelaskan berbagai sifat kondisional dari distribusi logistik bivariat.

Teorema 4.19.

Jika variabel acak $(X,Y) \sim LOGF(\mu_1,\mu_2,\sigma_1,\sigma_2)$

Maka

$$E(Y|x) = 1 - ln\left(1+e^{-\frac{\pi}{\sqrt{3}}\left(\frac{x-\mu_1}{\sigma_1}\right)}\right)$$

$$E(X|y) = 1 - ln\left(1+e^{-\frac{\pi}{\sqrt{3}}\left(\frac{y-\mu_2}{\sigma_2}\right)}\right)$$

$$Var(Y|x) = \frac{\pi^3}{3} - 1$$

$$Var(X|y) = \frac{\pi^3}{3} - 1$$

Definisi 4.10.

Sebuah variabel acak bivariat kontinu (X,Y) dikatakan berdistribusi logistik bivariat jenis kedua jika bentuk fungsi kepadatan probabilitas gabungan adalah

$$f(x,y) = \frac{\left[\phi_\alpha(x,y)\right]^{1-2\alpha}}{\left[1+\phi_\alpha(x,y)\right]^2}\left(\frac{\phi_\alpha(x,y)-1}{\phi_\alpha(x,y)+1}+\alpha\right)e^{-\alpha(x+y)},\ -\infty < x, y < \infty$$

di mana $\alpha > 0$ adalah parameternya, dan $\phi_\alpha(x,y) := \left(e^{-\alpha x} + e^{-\alpha y}\right)^{\frac{1}{\alpha}}$. Seperti sebelumnya, kita menunjukkan variabel acak logistik bivariat jenis kedua (X,Y) dengan menulis $(X,Y) \sim LOGS(\alpha)$.

Kepadatan marginal dari X dan Y adalah logistik dan diberikan oleh

$$f_1(x) = \frac{e^{-x}}{\left(1+e^{-x}\right)^2}, -\infty < x < \infty$$

Dan

$$f_2(y) = \frac{e^{-y}}{\left(1+e^{-y}\right)^2}, -\infty < y < \infty$$

Hal ini ditunjukkan oleh Oliveira (1961) bahwa jika $(X,Y) \sim LOGS(\alpha)$, maka korelasi antara X dan Y adalah

$$\rho(X,Y) = 1 - \frac{1}{2\alpha^2}$$

LATIHAN SOAL BAB 4

1. Jika $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ dengan $Q(x,y) = x^2 + 2y^2 - 2xy + 2x - 2y + 1$, lalu berapakah nilai varians bersyarat dari Y dengan syarat kejadian $X = x$?

   Jawab :

   $$Q(x,y) = x^2 + 2y^2 - 2xy + 2x - 2y + 1$$

   Jika $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ maka

$$\left(Y \mid X=x\right) \sim N\left[\mu_2+\rho\frac{\sigma_2}{\sigma_1}\left(x-\mu_1\right),\sigma_2^2\left(1-\rho^2\right)\right]$$

Misalkan

$$p = koefisien\, dari\, x^2 = 1 = \frac{1}{\left(1-\rho^2\right)\sigma_1^2}$$

$$q = koefisien\, dari\, y^2 = 2 = \frac{1}{\left(1-\rho^2\right)\sigma_2^2}$$

$$r = koefisien\, dari\, xy = -2$$

$$m = \frac{r^2}{pq} = \frac{4}{2} = 2$$

$$\rho = \frac{\sqrt{m}}{2} = \frac{\sqrt{2}}{2} \quad maka \quad \rho^2 = \frac{1}{2} \quad dan \quad \frac{1}{\left(1-\rho^2\right)\sigma_2^2} = 2$$

Sehingga

$$\sigma_2^2 = \frac{1}{2.\left(1-\frac{1}{2}\right)} = \frac{1}{2.\frac{1}{2}} = 1$$

Varians dari $\left(Y \mid X=x\right) = \left(1-\rho^2\right)\sigma_2^2 = \left(1-\frac{1}{2}\right).1 = \frac{1}{2}\;.1 = \frac{1}{2}$

2. Jika $\left(X,Y\right) \sim N\left(\mu_1,\mu_2,\ \sigma_1,\ \sigma_2,\rho\right)$ dengan

$$Q\left(x,y\right) = -\frac{1}{102}\left[\left(x+3\right)^2 - 16\left(x-3\right)\left(y-2\right) + 4\left(y-2\right)^2\right]$$

lalu berapakah nilai ekspektasi bersyarat dari Y dengan syarat $X = x$?

Jawab :

Diketahui $(X,Y) \sim N(\mu_1,\mu_2,\ \sigma_1,\ \sigma_2,\rho)$

$$Q(x,y) = -\frac{1}{102}\left[(x+3)^2 - 16(x-3)(y-2) + 4(y-2)^2\right]$$

Distribusi normal bivariat $(X,Y) \sim N(\mu_1,\mu_2,\ \sigma_1^2,\sigma_2^2,\rho)$

maka

$$f(x,y) = \frac{1}{2\mu\sigma_1\sigma_2\sqrt{1-\rho^2}}\ exp\left\{-\frac{1}{2(1-\rho^{2)}}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2p\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

Kita tahu bahwa

$$Q(x,y) = \frac{1}{(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2p\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

Sehingga $(1-\rho^2) = -102$

$$1-\rho^2 = 1{,}02$$

$$\rho^2 = 1-1{,}02 = -0{,}02$$

Maka kita dapat memiliki $\mu_1,\mu_2,\ \sigma_1,\ \sigma_2,\rho$

$$\mu_1 = -3, \mu_2 = 2,\ \sigma_1^2 = 1 \quad maka \quad \sigma_1 = 1,\ \sigma_2^2 = \frac{1}{4} \quad maka \quad \sigma_2 = \frac{1}{2},\ \rho = -0{,}02$$

$$f_x(x) = \frac{1}{\sqrt{2\mu}} exp\left\{-\frac{(x+3)^2}{2}\right\}$$

$$f(Y \mid X = x) = \frac{1}{\sigma_2\sqrt{1-\rho^2}\sqrt{2\mu}} exp\left\{-\frac{(y-\beta_x)^2}{2\sigma_2(1-\rho^2)}\right\}$$

Dimana $\beta_x = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$ adalah mean

$$E(Y \mid X = x) = \beta_x = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$$

dengan $\sigma_2^2(1-\rho^2)$ adalah varians

Sehingga :

$$Var = \frac{1}{4}(1+0,02)$$

$$= \frac{1}{4}.1,02$$

$$= \frac{1,02}{4} = 0,255$$

3. Jika $(X,Y) \sim N(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$, maka berapa koefisien korelasi dari variabel acak U dan V di mana $U = 2X+3Y$ dan $V = 2X-3Y$?

Jawab :

Diketahui $(X,Y) \sim N(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$

$$U = 2X+3Y$$

$$V = 2X-3Y$$

$$Corr(uv) = \frac{Cov(u,v)}{\sqrt{v(u)}\sqrt{v(v)}} \quad \ldots\ldots\ldots\ldots\ldots(1)$$

$$Cov(u,v) = Cov(2x+3y, 2x-3y)$$

$$= 4v(x) - 6\,Cov(x, y) + 6\,Cov(x, y) - 9v(y)$$

$$= 4v(x) - 9v(y)$$

$$= 4\sigma_1^2 - 9\sigma_2^2$$

$$v(u) = v(2x + 3y)$$

$$= 4v(x) + 9v(y) + 12\,Cov(x, y)$$

$$= 4\sigma_1^2 + 9\sigma_2^2 + 12\sigma_1\sigma_2\rho$$

$$v(v) = v(2x - 3y)$$

$$= 4v(x) + 9v(y) - 12\sigma_1\sigma_2\rho$$

$$Corr(uv) = \frac{Cov(u, v)}{\sqrt{v(u)}\sqrt{v(v)}}$$

$$= \frac{4\sigma_1^2 - 9\sigma_2^2}{\sqrt{4\sigma_1^2 + 9\sigma_2^2 + 12\sigma_1\sigma_2\rho}\sqrt{4\sigma_1^2 + 9\sigma_2^2 - 12\sigma_1\sigma_2\rho}}$$

Koefisien korelasi dari variabel random U dan V adalah

$$Corr(uv) = \rho_{u,v} = \frac{4\sigma_1^2 - 9\sigma_2^2}{\sqrt{4\sigma_1^2 + 9\sigma_2^2 + 12\sigma_1\sigma_2\rho}\sqrt{4\sigma_1^2 + 9\sigma_2^2 - 12\sigma_1\sigma_2\rho}}$$

4. Misalkan variabel acak X dan Y masing-masingmenunjukkan tinggi dan berat kalkun liar. Jika variabel random X dan Y memiliki bivariat normal distribusi dengan $\mu_1 = 18$ inci, $\mu_2 = 15$ pound, $\sigma_1 = 3$ inci, $\sigma_2 = 2$ pound, dan $\rho = 0,75$, lalu berapa bobot yang diharapkan dari salah satu hewan kalkun liar setinggi 17 inci?

   Jawab :

$$X = tinggi\left(inci\right)$$

$$Y = berat\left(pound\right)$$

$$\left(X,Y\right) \sim N\left(\mu_1 = 18, \mu_2 = 15, \sigma_1^2 = 32, \sigma_2^2 = 22, \rho_{12} = 0,75\right)$$

   Distribusi normal bivariat

$$E(Y \mid X = x) = \mu_y + \rho \frac{\sigma_y}{\sigma_x}\left(x - \mu_1\right)$$

$$E(Y \mid X = x) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}\left(x - \mu_1\right)$$

   Nilai ekspektasi dari Y (berat) dengan tinggi $X = 17$ inci

$$E(Y \mid X = 17) = 15 + 0,75 \bullet \frac{2}{3}\left(17 - 18\right)$$

$$= 15 + \frac{3}{4} \bullet \frac{2}{3} \bullet -1$$

$$= 15 - 0,5$$

$$E(Y \mid X = 17) = 14,5 \text{ pound}$$

   bobot yang diharapkan dari salah satu hewan kalkun liar setinggi 17 inci adalah 14,5 pound

5. Jika $\left(X,Y\right) \sim N\left(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho\right)$, berapakah fungsi pembangkit momen dari variabel acak U dan V, dimana $U = 7X + 3Y$ dan $V = 7X - 3Y$?

Jawab :

$$\left(X,Y\right)\sim N\left(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho\right)$$

$U=7X+3Y$ dan $V=7X-3Y$

Fungsi pembangkit momen normal bivariat

$$M_{x,y}\left(t_1,t_2\right)=exp\left[t_1\mu_1+t_2\mu_2+\frac{1}{2}(t_1^2\sigma_1^2+t_2^2\sigma_2^2+2\rho t_1t_2\sigma_1\sigma_2\right]$$

Berdasarkan definisi MGF dari variabel random adalah $M_x\left(t\right)=E\left(e^{tX}\right)$

MGF dari $U=7X+3Y$ adalah

$$M_U\left(t\right)=E\left(etU\right)$$

$$=E\left(e^{t(7X+3Y)}\right)$$

$$=E\left(e^{(7t)X+(3t)Y}\right)$$

$$=exp\left[\left(7t\right)\mu_1+\left(3t\right)\mu_2+\frac{1}{2}\left[\left(7t\right)^2\sigma_1^2+\left(3t\right)^2\sigma_2^2+2\rho\left(7t\right)\left(3t\right)\sigma_1\sigma_2\right]\right]$$

$$=exp\left[t\left(7\mu_1+3\mu_2\right)+\frac{t^2}{2}\left[7^2\sigma_1^2+3^2\sigma_2^2+2\rho\left(7\right)\left(3\right)\sigma_1\sigma_2\right]\right]$$

Dimana fungsi pembangkit momen dari distribusi normal dengan parameter $\mu=7\mu_1+3\mu_2$ dan $\sigma^2=7^2\sigma_1^2+3^2\sigma_2^2+2\rho\left(7\right)\left(3\right)\sigma_1\sigma_2$

Untuk itu $U=7X+3Y$ berdistribusi normal dengan $\mu=7\mu_1+3\mu_2$ dan $\sigma^2=7^2\sigma_1^2+3^2\sigma_2^2+2\rho\left(7\right)\left(3\right)\sigma_1\sigma_2$

MGF dari $V = 7X - 3Y$ adalah

$$M_V(t) = E\left(e^{tV}\right)$$

$$= E\left(e^{t(7X-3Y)}\right)$$

$$= E\left(e^{(7t)X+(-3t)Y}\right)$$

$$= exp\left[(7t)\mu_1 + (-3t)\mu_2 + \frac{1}{2}\left[(7t)^2\sigma_1^2 + (-3t)^2\sigma_2^2 + 2\rho(7t)(-3t)\sigma_1\sigma_2\right]\right]$$

$$= exp\left[t(7\mu_1 - 3\mu_2) + \frac{t^2}{2}\left[7^2\sigma_1^2 + 3^2\sigma_2^2 - 2\rho(7)(3)\sigma_1\sigma_2\right]\right]$$

Dimana MGF dari distribusi normal dengan parameter $\mu = 7 - 3\mu_2$ dan $\sigma^2 = 7^2\sigma_1^2 + 3^2\sigma_2^2 - 2\rho(7)(3)\sigma_1\sigma_2$

Untuk itu $U = 7X + 3Y$ berdistribusi normal dengan $\mu = 7\mu_1 - 3\mu_2$ dan $\sigma^2 = 7^2\sigma_1^2 + 3^2\sigma_2^2 - 2\rho(7)(3)\sigma_1\sigma_2$

6. Misalkan (X,Y) memiliki distribusi normal bivariat. Rata-rata dari X adalah 10 dan varians dari X adalah 12. Rata-rata dari Y adalah −5 dan varians dari Y adalah 5. Jika kovarians dari X dan Y adalah 4, berapakah probabilitas $X + Y$ adalah lebih dari 10?

Jawab :

$$E(X) = 10 \text{ dan } Var(X) = 12$$

$$E(Y) = -5 \text{ dan } Var(Y) = 5$$

$$Cov(X,Y) = 4$$

Ditanyakan $P(X + Y > 10)$

$(X+Y)$ berdistribusi normal

$$E(X+Y)=E(X)+E(Y)$$

$$E(X+Y)=10-5=5$$

$$Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$$

$$=12+5+2\bullet 4$$

$$=17+8$$

$$=25$$

$$P(X+Y>10)=1-P(X+Y\leq 10)$$

$$=1-P\left(\frac{(X+Y)-E(X+Y)}{\sqrt{Var(X+Y)}}\leq\frac{10-5}{5}\right)$$

$$=1-P(Z\leq 1)$$

$$=1-0,8413$$

$$=0,1587$$

Jadi probabilitas $P(X+Y>10)$ adalah sebersar $0,1587$

7. Misalkan X dan Y memiliki distribusi normal bivariat dengan mean $\mu_X=5$ dan $\mu_Y=6$ simpangan baku $\sigma_X=3$ dan $\sigma_Y=2$, dan kovarians $\sigma_{XY}=2$. Misalkan Φ menunjukkan fungsi distribusi kumulatif normal dari sebuah variabel acak dengan mean 0 dan varians 1. Berapakah $P(2\leq X-Y\leq 5)$ dalam pengertian Φ ?

Jawab :

Diberikan

$$\mu_X = 5 \qquad X \sim N(5,3)$$

$$\mu_Y = 6 \qquad Y \sim N(6,2)$$

Standar deviasi

$$\sigma_X = 3$$

$$\sigma_Y = 2$$

Kovarians

$$Cov(X,Y) = 2$$

Jadi

$$\mu = 2X - Y$$

$$\mu \sim (\mu, \sigma^2)$$

$$\mu = 2\mu_X - \mu_X$$

$$= 2(5) - 6$$

$$= 4$$

Kemudian

$$\sigma^2 = 4V(X) + V(Y) - 4cov(X,Y)$$

$$= 4(3) + 5 - (2)(4)$$

$$= 6$$

Sehingga

$$P\left(2X - 4 < 5\right) = P\left(X < 5\right)$$

$$= P\left(2 < \frac{5-2}{\sqrt{6}}\right)$$

$$= P\left(2 < \frac{3}{\sqrt{6}}\right)$$

$$= \Phi\left(\frac{3}{\sqrt{6}}\right)$$

BAB V

BEBERAPA TEKNIK DALAM MENEMUKAN ESTIMASI TITIK PARAMETER

Sebuah statistik populasi terdiri dari semua pengukuran penting dalam sebuah investigasi statistik. Biasanya sebuah populasi digambarkan dengan variabel acak X. Jika kita dapat membangun beberapa pengetahuan tentang fungsi kepadatan probabilitas $f(x;\theta)$ dari X kemudian ita juga membangun pengetahuan mengenai populasi pada investigasi.

Sampel adalah bagian dari populasi yang biasanya dipilih dengan metode sampling acak dan juga himpunan variabel acak $X_1, X_2, \ldots, X_n$ dengan beberapa probabilitas fungsi kepadatan $f(x;\theta)$ sebagai populasi. Setelah pengambilan sampel selesai, kita dapatkan

$$X_1 = x_1 \; , X_2 = x_2, \ldots, X_n = x_n$$

dimana $x_1, x_2, \ldots, x_n$ adalah data sampel.

Setiap metode statistik memakai sebuah variabel acak untuk mendapatkan informasi dari sebuah populasi. Ketika populasi telah menunjukan probabilitas fungsi kepadatan $f(x;\theta)$ dalam membuat kesimpulan statistik tentang distribusi populasi $f(x;\theta)$ berdasarkan informasi sampel. Sebuah kesimpulan statistik adalah pernyataan berdasarkan informasi sampel tentang populasi. Ada tiga jenis kesimpulan statistik :

a) Estimasi

b) Pengujian hipotesis

c) Prediksi

Pada poin estimasi, kita mencari parameter θ dari distribusi populasi $f(x;\theta)$ dari informasi sampel. Jadi, parametrik titik estimasi di asumsikan membentuk fungsi kepadatan probabilitas $f(x;\theta)$ untuk diketahui dan hanya memperkirakan parameter θ yang tidak diketahui dari populasi menggunakan informasi yang tersedia dari sampel.

Definisi 5.1.

Misalkan X adalah populasi dengan fungsi kepadatan $f(x;\theta)$ dimana θ parameter yang tidak diketahui. Himpunan semua nilai yang dapat diterima dari θ disebut ruang parameter dan dinotasikan sebagai Ω, bahwa

$$\Omega = \left\{ \theta \in \mathbb{R}^n \,\middle|\, f(x;\theta) \text{ adalah pdf} \right\}$$

Untuk semua bilangan asli m

Contoh 5.1. Jika $X \sim EXP(\theta)$, lalu berapakah ruang parameter dari θ ?

Jawab :

Saat $X \sim EXP(\theta)$ maka fungsi kepadatan dari X diberikan oleh

$$f(x;\theta) = \frac{1}{\theta} e^{-\left(\frac{x}{\theta}\right)}$$

Jika θ adalah nol atau negatif maka $f(x;\theta)$ bukanlah fungsi kepadatan. Jadi, nilai yang dapat diterima dari θ adalah semua bilangan asli positif. Oleh karena itu

$$\Omega = \{\theta \in \mathbb{R} \mid 0 < \theta < \infty\}$$

$$= \mathbb{R}_+$$

Contoh 5.2. Jika $X \sim N(\mu, \sigma^2)$, berapakah ruang parameternya ?

Jawab :

Ruang parameter Ω diberikan oleh

$$\Omega = \{\theta \in \mathbb{R}^2 \mid f(x;\theta) \sim N(\mu, \sigma^2)\}$$

$$= \{(\mu, \sigma) \in \mathbb{R}^2 \mid -\infty < \mu < \infty, 0 < \sigma < \infty\}$$

$$= \mathbb{R} \text{ x } \mathbb{R}_+$$

$$= \text{bidang setengah atas}$$

Umumnya, sebuah ruang parameter adalah subset dari $\mathbb{R}^2$ .mengenai statistik dengan estimasi parameter θ yang tidak diketahui dari sampel acak $X_1, X_2, \ldots, X_n$. Mengingat bahwa statistik adalah fungsi dari $X_1, X_2, \ldots, X_n$ dan bebas dari populasi parameter θ.

Definisi 5.2.

Misalkan $X \sim f(x;\theta)$ dan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari populasi X. . Beberapa statistik bisa juga digunakan untuk menduga parameter θ yang disebut estimator dari θ. Nilai numerik dari statistik disebut perkiraan dari θ. Estimator dari parameter θ disebut $\hat{\theta}$.

Salah satu hal yang mendasar ialah bagaimana mencari sebuah estimator dari populasi parameter θ. Ada beberapa metode dalam mencari estimator dari θ, yaitu :

a) Metode Momen

b) Metode Likelihood Maksimum

c) Metode Bayes

d) Metode Kuadrat Terkecil

e) Metode Chi-square Minimum

f) Metode Jarak Minimum

Pada bagian ini kita hanya akan membahas tiga metode pertama dari estimasi populasi parameter

5.1. Metode Momen

Misalkan $X_1, X_2, \ldots, X_n$ adalah sebuah sampel acak dari populasi X dengan probabilitas fungsi kepadatan $f\left(x;\theta_1,\theta_2,\ldots,\theta_m\right)$ dimana $\theta_1,\theta_2,\ldots,\theta_m$ adalah parameter m tidak diketahui. Diberikan

$$E\left(X^k\right) = \int_{-\infty}^{\infty} x^k \, f\left(x;\theta_1,\theta_2,\ldots,\theta_m\right) dx$$

Menjadi populasi momen ke-k sekitar 0. Selanjutnya, misalkan

$$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Menjadi sampel momen ke-k sekitar 0.

Dalam metode momen, kita mencari estimator untuk parameter $\theta_1,\theta_2,\ldots,\theta_m$ dengan menyamakan populasi momen pertama (jika ada) ke sampel momen pertama, yakni

$$E\left(X\right) = M_1$$

$$E\left(X^2\right) = M_2$$

$$E\left(X^3\right) = M_3$$

$$\vdots$$

$$E\left(X^m\right) = M_m$$

Metode momen adalah salah satu metode klasik untuk memperkirakan parameter dan alasan datang dari fakta bahwa sampel momen adalah beberapa perkiraan yang masuk akal untuk populasi momen. Metode momen pertama ditemukan oleh orang inggris yang bernama Karl Pearson pada tahun 1902. Sekarang kita memberikan contoh ilustrasi metode momen.

Contoh 5.3. Misalkan $X \sim N\left(\mu, \sigma^2\right)$ dan $X_1, X_2, \ldots, X_n$ adalah sampel acak berukuran n dari populasi X. Berapakah nilai estimator dari populasi parameter μ dan σ^2 jika kita menggunakan metode momen ?

Jawab :

Saat populasi berdistribusi normal, yakni

$$X \sim N\left(\mu, \sigma^2\right)$$

Kita tahu bahwa

$$E\left(X\right) = \mu$$

$$E\left(X^2\right) = \sigma^2 + \mu^2$$

Oleh karena itu

$$\mu = E\left(X\right)$$

$$= M_1$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$= \overline{X}$$

Oleh karena itu, estimator dari parameter μ adalah $\overline{X}$, sehingga

$$\widehat{\mu} = \overline{X}$$

Selanjutnya, kita cari estimator dari σ^2 yang menyamakan $E\left(X^2\right)$ dengan M_2. Catatan bahwa

$$\sigma^2 = \sigma^2 + \mu^2 - \mu^2$$

$$= E\left(X^2\right) - \mu^2$$

$$= M_2 - \mu^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

Pada baris terakhir sesuai dengan fakta tersebut bahwa

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\overline{X} + \overline{X}^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\sum_{i=1}^{n} 2X_i\overline{X} + \frac{1}{n}\sum_{i=1}^{n}\overline{X}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}\frac{1}{n}\sum_{i=1}^{n} X_i + \overline{X}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}\overline{X} + \overline{X}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$$

Jadi, penaksir dari σ^2 adalah $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$, yaitu

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

Contoh 5.4. Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak berukuran n dari populasi X dengan fungsi kepadatan probabilitas

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & jika\ 0 < x < 1 \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Dimana $0 < \theta < \infty$ adalah parameter yang tidak diketahui. Dengan menggunakan metode momen, temukan estimator dari θ?. Jika $x_1 = 0.2, x_2 = 0.6$, $x_3 = 0.5$, $x_4 = 0.3$ adalah sampel acak berukuran 4, maka berapakah penaksir dari θ

Jawab :

Untuk menemukan estimator, kita akan menyamakan populasi momen terhadap sampel momen. Populasi momen $E(X)$ diberikan oleh

$$E(X) = \int_0^1 x\, f(x;\theta)\, dx$$

$$= \int_0^1 x\,\theta\, x^{\theta-1} dx$$

$$= \theta\int_0^1 x^{\theta} dx$$

$$= \frac{\theta}{\theta+1}\left[x^{\theta+1}\right]_0^1$$

$$= \frac{\theta}{\theta+1}$$

Kita tahu bahwa $M_1 = \overline{X}$. Sekarang, menyusun M_1 disamakan dengan $E(X)$ dan penyelesaian untuk θ, kita dapatkan

$$\overline{X} = \frac{\theta}{\theta + 1}$$

Sehingga

$$\theta = \frac{\overline{X}}{1 - \overline{X}}$$

Dimana $\overline{X}$ adalah sampel mean. Jadi, statistik $\frac{\overline{X}}{1-\overline{X}}$ adalah estimator dari parameter θ. Oleh karena itu

$$\hat{\theta} = \frac{\overline{X}}{1 - \overline{X}}$$

Saat $x_1 = 0.2, x_2 = 0.6$, $x_3 = 0.5$, $x_4 = 0.3$, kita mempunyai $\overline{X} = 0.4$ and

$$\hat{\theta} = \frac{0.4}{1 - 0.4} = \frac{2}{3}$$

Merupakan penaksir dari θ.

Contoh 5.5. Apa prinsip dasar dari metode momen?

Jawab :

Untuk memilih nilai parameter populasi yang tidak diketahui dengan data yang diamati memiliki momen yang sama dengan populasi.

Contoh 5.6. Misalkan $X_1, X_2, \ldots, X_7$ adalah sampel acak dari populasi X dengan fungsi kepadatan

$$f(x;\beta) = \begin{cases} \dfrac{x^6 e^{-\frac{x}{\beta}}}{\Gamma(7)\beta^7}, & \textit{jika } 0 < x < \infty \\ 0, & \textit{untuk yang lainnya} \end{cases}$$

Temukan estimator dari β dengan menggunakan metode momen

Jawab :

Karena kita hanya memiliki satu parameter, kita hanya perlu menghitung populasi momen pertama $E(X)$ sekitar 0. Jadi,

$$E(X) = \int_0^\infty x\, f(x;\beta)\, dx$$

$$= \int_0^\infty x \frac{x^6 e^{-\left(\frac{x}{\beta}\right)}}{\Gamma(7)\beta^7} dx$$

$$= \frac{1}{\Gamma(7)} \int_0^\infty \left(\frac{x}{\beta}\right)^7 e^{-\left(\frac{x}{\beta}\right)} dx$$

$$= \beta \frac{1}{\Gamma(y)} \int_0^\infty y^7 e^{-y}\, dy$$

$$= \beta \frac{1}{\Gamma(7)} \Gamma(8)$$

$$= 7\beta$$

Saat $M_1 = \overline{X}$, menyamakan $E(X)$ dengan M_1, kita dapatkan

$$7\beta = \overline{X}$$

Sehingga

$$\beta = \frac{1}{7}\overline{X}$$

Oleh sebab itu, estimator dari β dengan metode momen adalah diberikan sebagai berikut

$$\widehat{\beta} = \frac{1}{7}\overline{X}$$

Contoh 5.7. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari populasi X dengan fungsi kepadatan

$$f(x;0) = \begin{cases} \dfrac{1}{\theta}, & jika\ 0 < x < \theta \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Carilah sebuah estimator dari θ menggunakan metode momen.

Jawab :

Periksa fungsi kepadatan dari populasi X, kita lihat bahwa $X \sim UNIF(0,\theta)$. Oleh karena itu

$$E(X) = \frac{\theta}{2}$$

Sekarang dengan menyamakan populasi momen ini dengan sampel momen, kita peroleh

$$\frac{\theta}{2} = E(X) = M_1 = \overline{X}$$

Oleh karena itu estimator dari θ adalah

$$\hat{\theta} = 2\overline{X}$$

Contoh 5.8. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari sebuah populasi X dengan fungsi kepadatan

$$f(x;\alpha;\beta) = \begin{cases} \dfrac{1}{\beta - \alpha}, & jika\ \alpha < x < \beta \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Carilah estimator dai α dan β dengan menggunakan metode momen

Jawab :

Periksa fungsi kepadatan dari populasi X, kita lihat bahwa $X \sim UNIF(\alpha, \beta)$. Karena distribusi memiliki dua parameter yang tidak diketahui, kita memerlukan dua populasi momen pertama. Oleh karena itu

$$E(X) = \frac{\alpha+\beta}{2} \qquad dan \qquad E(X^2) = \frac{(\beta-\alpha)^2}{12} + E(X)^2$$

Menyamakan momen-momen ini dengan momen sampel yang sesuai, kita dapatkan

$$\frac{\alpha+\beta}{2} = E(X) = M_1 = \overline{X}$$

Sehingga

$$\alpha + \beta = 2\overline{X} \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Dan

$$\frac{(\beta-\alpha)^2}{12} + E(X)^2 = E(X^2) = M_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Yang mana

$$(\beta-\alpha)^2 = 12\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - E(X)^2\right]$$

$$= 12\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right]$$

$$= 12\left[\frac{1}{n}\sum_{i=1}^{n} \left(X_i^2 - \overline{X}\right)^2\right]$$

Oleh karena itu, kita dapatkan

$$\beta - \alpha = \sqrt{\frac{12}{n}\sum_{i=1}^{n}\left(X_i^2 - \overline{X}\right)^2} \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Menambahkan persamaan (1) terhadap persamaan (2), kita peroleh

$$2\beta = 2\overline{X} \pm 2\sqrt{\frac{3}{n}\sum_{i=1}^{n}\left(X_i^2 - \overline{X}\right)^2}$$

Sehingga

$$\beta = \overline{X} \mp \sqrt{\frac{3}{n}\sum_{i=1}^{n}\left(X_i^2 - \overline{X}\right)^2}$$

Karena $\alpha < \beta$, kita lihat bahwa estimator dari α dan β adalah

$$\hat{\alpha} = \overline{X} - \sqrt{\frac{3}{n}\sum_{i=1}^{n}\left(X_i^2 - \overline{X}\right)^2} \quad \text{dan} \quad \beta = \overline{X} + \sqrt{\frac{3}{n}\sum_{i=1}^{n}\left(X_i^2 - \overline{X}\right)^2}$$

5.2. Metode Maksimum Likelihood

Metode likelihood maksimum pertama kali digunakan oleh Ronald Fisher pada tahun 1912 untuk mencari estimator dari parameter yang tidak diketahui. Bagaimanapun, metode ini berasal dari Gauss dan Bernoulli. Selanjutnya, kita akan membahas materinya secara detail.

Definisi 5.3.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari populasi X dengan fungsi kepadatan probabilitas $f(x;\theta)$ dimana θ adalah parameter yang belum diketahui. Fungsi likelihood, $L(\theta)$ adalah distribusi dari sampel. Yakni

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

Definisi ini menyatakan bahwa fungsi likelihood dari sebuah sampel acak $X_1, X_2, \ldots, X_n$ adalah gabungan dari kepadatan sampel acak $X_1, X_2, \ldots, X_n$.

θ yang memaksimalkan fungsi likelihood $L(\theta)$ disebut estimator maksimum likelihood dari θ dan dinotasikan sebagai $\hat{\theta}$. Oleh karena itu

$$\hat{\theta} = Arg \sup_{\theta \in \Omega} L(\theta)$$

Dimana Ω merupakan ruang parameter dari θ sehingga $L(\theta)$ adalah gabungan kepadatan dari sampel.

Metode likelihood maskimum dalam arti lain ialah memilih dari semua nilai kemungkinan dari θ yang paling mungkin menghasilkan observasi yang diberikan $x_1, x_2, \ldots, x_n$. Metode penjumlahan ini diantaranya :

a) Memperoleh sampel acak $x_1, x_2, \ldots, x_n$ dari distribusi populasi X dengan pdf $f(x;\theta)$;

b) Definisikan fungsi likelihood untuk sampel $x_1, x_2, \ldots, x_n$ dengan $L(\theta) = f(x_1;\theta) f(x_2;\theta) \ldots f(x_n;\theta)$.

c) Carilah pernyataan untuk θ bahwa $L(\theta)$ maksimal. Ini bisa diselesaikan langsung atau dengan $\ln L(4)$ maksimal.

d) Ganti θ dengan $\hat{\theta}$ untuk memperoleh sebuah pernyataan bahwa estimator likelihood θ maksimum.

e) Carilah nilai yang diamati dari estimator ini untuk sampel yang diberikan.

Contoh 5.9. Jika $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan fungsi kerapatan

$$f(x;\theta) = \begin{cases} (1-\theta)x^{-\theta}, & \textit{jika } 0 < x < 1 \\ 0, & \textit{untuk yang lainnya} \end{cases}$$

Carilah estimator maksimum likelihood θ ?

Jawab :

Fungsi likelihood dari sampel diberikan dengan

$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta).$$

Oleh karena itu

$$\ln L(\theta) = \ln\left(\prod_{i=1}^{n} f(x_i \;;\theta)\right)$$

$$= \sum_{i=1}^{n} \ln f(x_i;\theta)$$

$$= \sum_{i=1}^{n} \ln\left[(1-\theta)x_i^{-\theta}\right]$$

$$= n \ln(1-\theta) - \theta \sum_{i=1}^{n} \ln x_i$$

Sekarang kita maksimalkan $\ln L(\theta)$ terhadap θ

$$\frac{d \ln L(\theta)}{d\theta} = \frac{d}{d\theta}\left(n \ln(1-\theta) - \theta \sum_{i=1}^{n} \ln x_i \right)$$

$$= -\frac{n}{1-\theta} - \sum_{i=1}^{n} \ln x_i$$

Selesaikan turunan $\frac{d \ln L(\theta)}{d\theta}$ untuk 0, kita dapatkan

$$\frac{d \ln L(\theta)}{d\theta} = -\frac{n}{1-\theta} - \sum_{i=1}^{n} \ln x_i$$

$$= 0$$

Sehingga

$$\frac{1}{1-\theta} = -\frac{1}{n} \sum_{i=1}^{n} \ln x_i$$

Atau

$$\frac{1}{1-\theta} = -\frac{1}{n} \sum_{i=1}^{n} \ln x_i$$

$$= -\overline{\ln x}$$

Atau

$$\theta = 1 + \frac{1}{\overline{\ln x}}$$

Disini θ bisa juga menunjukan menjadi maksimum dengan menguji turunan kedua. Sehingga, estimator dari θ adalah

$$\hat{\theta} = 1 + \frac{1}{\overline{\ln\ x}}$$

Contoh 5.10. Jika $X_1, X_2, \ldots, X_n$ adalah samppel acak dari distribusi dengan fungsi kepadatan

$$f(x;\beta) = \begin{cases} \dfrac{x^6 e^{-\frac{x}{\beta}}}{\Gamma(7)\beta^7}, & jika\ 0 < x < \infty \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Kemudian berapakah estimator dari maksimum likelihood dari β ?

Jawab :

Fungsi likelihood dari sampel diberikan oleh

$$L(\beta) = \prod_{i=1}^{n} f(x_i;\beta)$$

Jadi,

$$\ln L(\beta) = \sum_{i=1}^{n} \ln f(x_i;\beta)$$

$$= 6\sum_{i=1}^{n} \ln x_i - \frac{1}{\beta}\sum_{i=1}^{n} \ln x_i - n\ln(6!) - 7n\ln(\beta)$$

Oleh karena itu

$$\frac{d}{d\beta}\ln L(\beta) = \frac{1}{\beta^2}\sum_{i=1}^{n} \ln x_i - \frac{7n}{\beta}$$

Selesaikan turunan $\dfrac{d}{d\beta}\ln L(\beta)$ hingga 0, kita punya

$$\frac{1}{\beta^2}\sum_{i=1}^{n} \ln x_i - \frac{7n}{\beta} = 0$$

Yang mengasilkan

$$\beta = \frac{1}{7n}\sum_{i=1}^{n} \ln x_i$$

Disini β bisa juga menunjukan menjadi maksimum dengan menguji turunan kedua. Oleh karena itu, estimator dari β diberikan oleh

$$\hat{\beta} = \frac{1}{7}\overline{X}$$

Catatan 5.1.

Perhatikan bahwa *MLE* dari β adalah sama seperti yang digunakan untuk metode momen pada Contoh 5.6. Namun, secara umum penaksir dengan metode yang berbeda adalah berbeda, seperti berikut mengilustrasikan contoh

Contoh 5.11. Jika $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan fungsi kepadatan

$$f(x;\theta) = \begin{cases} \dfrac{1}{\theta}, & \textit{jika } 0 < x < \theta \\ 0, & \textit{untuk yang lainnya} \end{cases}$$

Lalu carilah estimator maksimum likelihood dari θ ?

Jawab :

Fungsi likelihood dari sampel diberikan oleh

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

$$= \prod_{i=1}^{n}\left(\frac{1}{\theta}\right) \qquad \theta > x_i \quad (i = 1, 2, 3, \ldots, n)$$

$$= \left(\frac{1}{\theta}\right)^n \qquad \theta > \max\{x_1, x_2, \ldots, x_n\}$$

Karena ruang parameter dari θ dengan memperhatikan $L(\theta)$ diberikan oleh

$$\Omega = \{\theta \in \mathbb{R} | x_{\max} < \theta < \infty\} = (x_{\max}, \infty)$$

Sekarang kita maksimalkan $L(\theta)$ pada Ω. Pertama, kita hitung $\ln L(\theta)$ lalu turunkan sehingga

$$\ln L(\theta) = -n \ln(\theta)$$

Dan

$$\frac{d}{d\theta} \ln L(\theta) = -\frac{n}{\theta} < 0$$

Oleh karena itu $\ln L(\theta)$ adalah fungsi menurun dari θ dan demikian juga maksimum dari $\ln L(\theta)$ terjadi di titik akhir kiri pada interval $(x_{\max}, \infty)$. Oleh karena itu, $\theta = x_{\max}$ fungsi likelihood sampai maksimum. Karena estimator likelihood dari θ adalah

$$\hat{\theta} = X_{(n)}$$

Dimana $X_{(n)}$ dinotasikan sebagai urutan statistik ke-n dari sampel yang diberikan.

Jadi, contoh 5.6 dan contoh 5.11 mengatakan bahwa jika kita mengestimasikan parameter θ dari distribusi dengan kepadatan uniform pada interval $(0, \theta)$ lalu estimator maksimum likelihood diberikan oleh

$$\hat{\theta} = X_{(n)}$$

Dimana

$$\hat{\theta} = 2\overline{X}$$

Adalah estimator yang didapatkan dari metode momen. Karena, umumnya ada 2 metode yang tidak memberikan estimator yang sama dari parameter yang tidak diketahui.

Contoh 5.12. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan fungsi kepadatan

$$f(x;\theta) = \begin{cases} \sqrt{\dfrac{2}{\pi}} e^{-\frac{1}{2}(x-\theta)^2}, & jika\ x \geq \theta \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Berapa estimator maksimum likelihood dari θ ?

Jawab :

Fungsi likelihood $L(\theta)$ diberikan oleh

$$L(\theta)=\left(\sqrt{\frac{2}{\pi}}\right)^n \prod_{i=1}^{n} e^{-\frac{1}{2}(x_i-\theta)^2} \qquad x_i \geq \theta (i=1,2,3,\ldots,n)$$

Karena ruang parameter dari θ diberikan oleh

$$\Omega=\{\theta \in \mathbb{R} | 0 \leq \theta \leq x_{\min}\}=[0,x_{\min}]$$

Dimana $x_{\min}=\min\{x_1,x_2,\ldots,x_n\}$. Sekarang kita evaluasi logaritma dari fungsi likelihood.

$$\ln L(\theta)=\frac{n}{2}\ln\left(\frac{2}{\pi}\right)-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2$$

Dimana θ adalah interval pada $[0,x_{\min}]$. Sekarang kita maksimalkan subjek $\ln L(\theta)$ untuk kondisi $0 \leq \theta \leq x_{\min}$. Dengan pengambilan turunan kita dapatkan

$$\frac{d}{d\theta}\ln L(\theta)=-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)2(-1)$$

$$=\sum_{i=1}^{n}(x_i-\theta)$$

Pada contoh ini, jika kita samakan turunan dengan nol, lalu kita dapatkan $\theta=\bar{x}$. Akan tetapi nilai dari θ tidak ada pada ruang parameter Ω. Jadi, $\theta=\bar{x}$ bukanlah solusinya. Karena untuk mencari solusi dari proses optimasi ini, kita menguji $\ln L(\theta)$ pada interval $[0,x_{\min}]$. Ingatlah bahwa

$$\frac{d}{d\theta}\ln L(\theta)=-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)2(-1)=\sum_{i=1}^{n}(x_i-\theta)>0$$

Ketika $x_i \geq \theta$. Oleh karena itu, fungsi $\ln L(\theta)$ adalah fungsi peningkatan pada interval $[0,x_{\min}]$ dan demikian pula ini akan didapatkan maksimum pada titik akhir kanan pada interval $[0,x_{\min}]$. Oleh karena itu, estimator maksimum likelihood dari θ adalah

$$\hat{X} = X_{(1)}$$

Dimana $X_{(1)}$ dinotasikan sebagai pengamatan terkecil pada sampel acak $X_1, X_2, \ldots, X_n$.

Contoh 5.13. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari sebuah distribusi normal dengan rata-rata μ dan varians σ^2. Berapa estimator maksimum likelihood dari μ dan σ^2?

Jawab :

Saat $X \sim N(\mu, \sigma^2)$, fungsi kepadatan probabilitas dari X diberikan oleh

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Fungsi likelihood dari sampel diberikan oleh

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

Karena, logaritma fungsi likelihood ini diberikan oleh

$$\ln L(\mu, \sigma) = -\frac{\pi}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$$

Dengan menggunakan turunan parsial dari $\ln L(\mu, \sigma)$ untuk μ dan σ, kita dapatkan

$$\frac{\partial}{\partial\mu}\ln L(\mu, \sigma) = -\frac{1}{2\pi}\sum_{i=1}^{n}(x_i - \mu)(-2)$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$$

Dan

$$\frac{\partial}{\partial\mu}\ln L(\mu, \sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2$$

Selesaikan turunan $\frac{\partial}{\partial\mu}\ln L(\mu,\sigma)=0$ dan $\ln L(\mu,\sigma)=0$, dan carilah μ dan σ yang belum diketahui, kita dapatkan

$$\mu=\frac{1}{n}\sum_{i=1}^{n}x_i=\overline{x}$$

Jadi, estimator maksimum likelihood dari μ adalah

$$\widehat{\mu}=\overline{X}$$

Demikian pula

$$-\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2=0$$

Implikasi

$$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

Sekali lagi μ dan σ^2 yang ditemukan pada uji turunan pertama menunjukan maksimum menggunakan uji turunan kedua fungsi dua variabel. Karena menggunakan estimator dari μ dalam pernyataan tersebut kita punya estimator dari σ^2 sebagai berikut

$$\widehat{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

Contoh 5.14. Misalkan $X_1,X_2,\ldots,X_n$ adalah sampel acak dari distribusi dengan fungsi kepadatan

$$f(x;\alpha,\beta)=\begin{cases}\dfrac{1}{\beta-\alpha}, & jika\ \alpha<x<\beta\\ 0, & untuk\ yang\ lainnya\end{cases}$$

Carilah estimator dari α dan β dengan menggunakan metode maksimum likelihood

Jawab :

Fungsi likelihood dari sampel adalah

$$L(\alpha,\beta)=\prod_{i=1}^{n}\frac{1}{\beta-\alpha}$$

$$=\left(\frac{1}{\beta-\alpha}\right)^{n}$$

Untuk setiap $\alpha \le x_i$ dimana $(i=1,2,\ldots,n)$ dan untuk setiap $\beta \ge x_i$ dimana $(i=1,2,\ldots,n)$. Sehingga domain dari fungsi likelihood adalah

$$\Omega=\left\{(\alpha,\beta)|0<\alpha<x_i \; dan\, x_{(n)}\le\beta\le\infty\right\}$$

Ini menunjukan bahwa $L(\alpha,\beta)$ maksimum jika $\alpha=x_i$ dan $\beta=x_{(n)}$. Oleh karena itu, estimator maksimum likelihood dari α dan β adalah

$$\widehat{\alpha}=X_{(1)} \quad dan \quad \widehat{\beta}=X_{(n)}$$

Estimator maksimum likelihood $\hat{\theta}$ dari parameter θ mempunyai sifat yang luarbiasa yang dikenal sebagai sifat tak seragam (invarian). Sifat invarian mengatakan bahwa jika $\hat{\theta}$ adalah estimator maksimum likelihood dari θ lalu $g(\hat{\theta})$ adalah estimator likelihood maksimum dari $g(\theta)$ dimana g adalah fungsi dari $\mathbb{R}^k \subseteq \mathbb{R}^m$.

Teorema 5.1.

Misalkan $\hat{\theta}$ adalah estimator likelihood maksimum dari parameter θ dan misalkan $g(\theta)$ adalah fungsi dari θ. Kemudian estimator maksimum likelihood dari $g(\theta)$ diberikan oleh $g(\hat{\theta})$.

Contoh 5.15.

Misalkan $X_1,X_2,\ldots,X_n$ adalah sampel acak dari populasi normal dengan rata-rata μ dan varians σ^2. Carilah estimator likelihood maksimum dari σ dan $\mu-\sigma$?

Jawab :

Dari contoh 5.13 kita punya estimator maksimum likelihood dari μ dan σ^2 adalah

$$\widehat{\mu}=\overline{X}$$

Dan

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 =: \Sigma^2\left(\text{katakanlah}\right)$$

Sekarang, dengan menggunakan sifat invarian dari estimator likelihood maksimum kita dapatkan

$$\widehat{\sigma} = \Sigma$$

Dan

$$\widehat{\mu - \sigma} = \overline{X} - \Sigma$$

Contoh 5.16. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan fungsi kepadatan

$$f\left(x;\alpha,\beta\right) = \begin{cases} \dfrac{1}{\beta - \alpha}, & jika \ \alpha < x < \beta \\ 0, & untuk \ yang \ lainnya \end{cases}$$

Carilah estimator dari $\sqrt{\alpha^2 + \beta^2}$ dengan menggunakan metode maksimum likelihood.

Jawab :

Dari contoh 5.14 kita punya estimator maksimum likelihood dari α dan β masing-masing adalah

$$\widehat{\alpha} = X_{(1)} \quad dan \quad \widehat{\beta} = X_{(n)}$$

Sekarang, dengan menggunakan sifat invarian dari estimator maksimum likelihood kita tunjukan bahwa estimator likelihood maksimum dari $\sqrt{\alpha^2 + \beta^2}$ adalah $\sqrt{X_{(1)}^2 + X_{(n)}^2}$.

Definisi 5.4. Misalkan X adalah observasi dari sebuah populasi dengan fungsi kepadatan probabilitas $f\left(x;\theta\right)$. Jika $f\left(x;\theta\right)$ kontinu, terdiferensiasi dua kali dan tidak bergantung pada θ. Kemudian informasi Fisher, $I\left(\theta\right)$, sebuah pengamatan tunggal dari X yang diberikan oleh

$$I\left(\theta\right) = \int_{-\infty}^{\infty}\left[\frac{d \ln f\left(x;\theta\right)}{d\theta}\right]^2 f\left(x;\theta\right)dx$$

Jadi, $I(\theta)$ adalah kuadrat nilai ekspektasi dari variabel acak $\dfrac{d\ln f(x;\theta)}{d\theta}$
Sehingga

$$I(\theta)=E\left(\left[\frac{d\ln f(x;\theta)}{d\theta}\right]\right)$$

Lemma 5.1. Informasi Fisher memuat pengamatan tunggal mengenai parameter θ yang tidak diketahui sehingga bisa diberikan alternatif sebagai berikut

$$I(\theta)=-\int_{-\infty}^{\infty}\left[\frac{d^2\ln f(x;\theta)}{d\theta^2}\right]f(x;\theta)\,dx$$

Bukti :

Ketika $f(x;\theta)$ probabilitas fungsi kepadatan

$$\int_{-\infty}^{\infty}f(x;\theta)\,dx=1 \qquad (3)$$

Menggunakan turunan persamaan (3) dengan mengingat θ, maka

$$\frac{d}{d\theta}\int_{-\infty}^{\infty}f(x;\theta)\,dx=0$$

Dengan menulis ulang persamaan terakhir, kita punya

$$\int_{-\infty}^{\infty}\frac{df(x;\theta)}{d\theta}\frac{1}{f(x;\theta)}f(x;\theta)\,dx=0$$

Karena

$$\int_{-\infty}^{\infty}\frac{d\,ln\,f(x;\theta)}{d\theta}f(x;\theta)\,dx=0 \qquad (4)$$

Menggunakan turunan persamaan (4) dengan memperhatikan θ, maka

$$\int_{-\infty}^{\infty}\left[\frac{d^2\ln f(x;\theta)}{d\theta^2}f(x;\theta)+\frac{d\ln f(x;\theta)}{d\theta}\frac{df(x;\theta)}{d\theta}\right]dx=0$$

Dengan menulis ulang persamaan terakhir, kita punya

$$\int_{-\infty}^{\infty}\left[\frac{d^2 \ln f(x;\theta)}{d\theta^2} f(x;\theta)+\frac{d \ln f(x;\theta)}{d\theta}\frac{df(x;\theta)}{d\theta}\frac{1}{f(x;\theta)} f(x;\theta)\right] dx = 0$$

Dengan

$$\int_{-\infty}^{\infty}\left(\frac{d^2 \ln f(x;\theta)}{d\theta^2}\left[\frac{d \ln f(x;\theta)}{d\theta}\right]^2\right) f(x;\theta) dx = 0$$

Persamaan terakhir mengimplikasikan bahwa

$$\int_{-\infty}^{\infty}\left[\frac{d \ln f(x;\theta)}{d\theta}\right]^2 f(x;\theta) dx = -\int_{-\infty}^{\infty}\left[\frac{d^2 \ln f(x;\theta)}{d\theta^2}\right] f(x;\theta) dx$$

Karena dengan menggunakan definisi informasi fisher, kita punya

$$I(\theta)=\int_{-\infty}^{\infty}\left[\frac{d^2 \ln f(x;\theta)}{d\theta^2}\right] f(x;\theta) dx$$

Contoh 5.17. Misalkan X adalah observasi tunggal yang diambil dari normal populasi dengan rata-rata μ tidak diketahui dan varians σ^2 diketahui. Carilah informasi Fisher μ dalam observasi tunggal X.

Jawab :

Ketika $X \sim N\left(\mu, \sigma^2\right)$ probabilitas kepadatan dari X diberikan oleh

$$f(x;\mu)=\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Karena

$$\ln f(x;\mu)=-\frac{1}{2}\ln\left(2\pi\sigma^2\right)-\frac{(x-\mu)^2}{2\sigma^2}$$

Oleh sebab itu

$$\frac{d \ln f(x;\mu)}{d\mu} = \frac{x-\mu}{\sigma^2}$$

Dan

$$\frac{d^2 \ln f(x;\mu)}{d\mu^2} = -\frac{1}{\sigma^2}$$

Karena

$$I(\mu) = -\int_{-\infty}^{\infty} \left(-\frac{1}{\sigma^2}\right) f(x;\mu)\, dx = \frac{1}{\sigma^2}$$

Contoh 5.18. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari normal populasi dengan rata-rata μ tidak diketahui dan varians σ^2 diketahui. Carilah informasi fisher μ pada sampel berukuran n.

Jawab :

Misalkan $I_n(\mu)$ adalah yang diperlukan dalam informasi fisher. Kemudian, dari definisi kita punya

$$I_n(\mu) = -E\left(\frac{d^2 \ln f(X_1, X_2, \ldots, X_n;\mu)}{d\mu^2}\right)$$

$$= -E\left(\frac{d^2}{d\mu^2}\left\{\ln f(X_1;\mu) + \ldots + \ln f(X_n;\mu)\right\}\right)$$

$$= -E\left(\frac{d^2 \ln f(X_1;\mu)}{d\mu^2}\right) - \ldots - E\left(\frac{d^2 \ln f(X_n;\mu)}{d\mu^2}\right)$$

$$= I(\mu) + \ldots + I(\mu)$$

$$= n\, I(\mu)$$

$$= n\frac{1}{\sigma^2}$$

Contoh diatas menunjukan bahwa $X_1, X_2, \ldots, X_n$ adalah sebuah sampel acak dari populasi $X \sim f(x;\theta)$, lalu informasi Fisher $\mathrm{I_n}(\theta)$ dalam sampel dari parameter θ berukuran n sama dengan n kali informasi Fisher X pada θ. Jadi

$$I_n(\theta) = nI(\theta)$$

Jika X adalah variabel acak dengan fungsi kepadatan $f(x;\theta)$ dimana $\theta = (\theta_1, \ldots, \theta_n)$ adalah parameter vektor yang tidak diketahui lalu informasi fisher, $I_n(\theta)$ adalah matriks $n\,x\,n$ yaitu

$$I(\theta) = \left(I_{ij}(\theta)\right)$$

$$= \left(-E\left(\frac{\partial^2 \ln f(X;\theta)}{\partial\theta_i \partial\theta_j}\right)\right)$$

Contoh 5.19. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi populasi normal dengan rata-rata μ dan varians σ^2. Berapa informasi Fisher matriks $I_n(\mu, \sigma^2)$ dari sampel berukuran n parameter μ dan σ ?

Jawab :

Misalkan $\theta_1 = \mu$ dan $\theta_2 = \sigma^2$. Informasi Fisher, $I_n(\theta)$, pada sampel dengan ukuran n tentang parameter (θ_1, θ_2) adalah sama dengan n-kali informasi Fisher pada populasi tentang (θ_1, θ_2) maka

$$I_n(\theta_1, \theta_2) = nI(\theta_1, \theta_2)$$

Saat terdapat dua parameter θ_1 dan θ_2, informasi fisher matriks $I_n(\mu, \sigma^2)$ adalah matriks berordo 2×2

$$I(\theta_1, \theta_2) = \begin{pmatrix} I_{11}(\theta_1, \theta_2) & I_{12}(\theta_1, \theta_2) \\ I_{21}(\theta_1, \theta_2) & I_{22}(\theta_1, \theta_2) \end{pmatrix}$$

Dimana

$$I_{ij}(\theta_1, \theta_2) = -E\left(\frac{\partial^2 \ln f(X;\theta_1, \theta_2)}{\partial\theta_i \partial\theta_j}\right)$$

Untuk $i = j = 1, 2$. Selanjutnya kita hitung I_{ij}

$$I\left(x;\theta_1,\theta_2\right)=\frac{1}{\sqrt{2\pi\theta_2}}e^{-\frac{\left(x-\theta_1\right)^2}{2\theta_2}}$$

Kita punya

$$\ln f\left(x;\theta_1,\theta_2\right)=-\frac{1}{2}\ln\left(2\pi\theta_2\right)-\frac{\left(x-\theta_1\right)^2}{2\theta_2}$$

Dengan menggunakan turunan parsial dari $\ln f\left(x;\theta_1,\theta_2\right)$ kita punya

$$\frac{\partial\ln f\left(x;\theta_1,\theta_2\right)}{\partial\theta_1}=\frac{x-\theta_1}{\theta_2}$$

$$\frac{\partial\ln f\left(x;\theta_1,\theta_2\right)}{\partial\theta_2}=-\frac{1}{2\theta_2}+\frac{\left(x-\theta_1\right)^2}{2\theta_2^2}$$

$$\frac{\partial^2\ln f\left(x;\theta_1,\theta_2\right)}{\partial\theta_1^2}=-\frac{1}{\theta_2}$$

$$\frac{\partial^2\ln f\left(x;\theta_1,\theta_2\right)}{\partial\theta_2^2}=-\frac{1}{2\theta_2^2}+\frac{\left(x-\theta_1\right)^2}{2\theta_2^3}$$

$$\frac{\partial^2\ln f\left(x;\theta_1,\theta_2\right)}{\partial\theta_1\partial\theta_2}=-\frac{x-\theta_1}{\theta_2^2}$$

Oleh karena itu

$$I_{11}\left(\theta_1,\theta_2\right)=-E\left(-\frac{1}{\theta_2}\right)$$

$$=\frac{1}{\theta_2}$$

$$=\frac{1}{\sigma^2}$$

Dengan cara yang sama

$$I_{21}(\theta_1,\theta_2) = I_{12}(\theta_1,\theta_2) = -E\left(-\frac{X-\theta_1}{\theta_2^2}\right)$$

$$= \frac{E(X)}{\theta_2^2} - \frac{\theta_1}{\theta_2^2}$$

$$= \frac{\theta_1}{\theta_2^2} - \frac{\theta_1}{\theta_2^2}$$

$$= 0$$

Dan

$$I_{22}(\theta_1,\theta_2) = -E\left(-\frac{(X-\theta_1)^2}{\theta_2^3} + \frac{1}{2\theta_2^2}\right)$$

$$= \frac{E\left((X-\theta_1)^2\right)}{\theta_2^3} - \frac{1}{2\theta_2^2}$$

$$= \frac{\theta_2}{\theta_2^3} - \frac{1}{2\theta_2^2}$$

$$= \frac{1}{2\theta_2^2}$$

$$= \frac{1}{2\sigma^4}$$

Jadi dapat disimpulkan bahwa

$$I_n(\theta_1,\theta_2)=n\begin{pmatrix}\frac{1}{\sigma^2} & 0\\ 0 & \frac{1}{2\sigma^4}\end{pmatrix}$$

$$=\begin{pmatrix}\frac{n}{\sigma^2} & 0\\ 0 & \frac{n}{2\sigma^4}\end{pmatrix}$$

Teorema 5.2.

Kondisi dimana $f(x;\theta)$ adalah estimator maksimum likelihood $\hat{\theta}$ dari θ didasari oleh sampel acak berukuran n dari populasi X dengan kepadatan probabilitas $f(x;\theta)$ terdistribusi normal asimtotik dengan rata-rata μ dan varians $\frac{1}{nI(\theta)}$. Sehingga

$$\hat{\theta}_{ML} \sim N\left(\theta,\frac{1}{nI(\theta)}\right) \qquad dimana\, n\to\infty$$

Contoh 5.20.

Misalkan $X_1,X_2,\ldots,X_n$ adalah sampel acak dari distribusi dengan fungsi kepadatan

$$f(x;\theta)=\begin{cases}\frac{1}{2}, & jika\ 0-1\le x\le 0+1\\ 0, & untuk\ yang\ lainnya\end{cases}$$

Carilah estimator maksimum likelihood dari θ ?

Jawab :

Fungsi likelihood dari sampel adalah

$$L(\theta)=\begin{cases}\left(\frac{1}{2}\right)^n, & jika\ \max\{x_1,\ldots,x_n\}-1\le\theta\le\min\{x_1,\ldots,x_n\}+1\\ 0, & untuk\ yang\ lainnya\end{cases}$$

Ketika fungsi likelihood konstan, maka apapun nilai pada interval $\left[max\{x_1,..,x_n\}-1, min\{x_1,..,x_n\}+1\right]$ adalah estimator maksimum likelihood dari θ.

Contoh 5.21. Apa prinsip dasar dari estimator maksimum likelihood?

Jawab :

Untuk memilih nilai dari parameter data yang diamati memiliki probabilitas besar/tinggi atau kepadatannya memungkinkan. Dengan kata lain, estimator maksimum likelihood adalah nilai parameter dimana data tersebut mempunyai probabilitas yang sangat tinggi.

5.3. Metode Bayes

Pada pendekatan sederhana, parameter θ diasumsikan tidak diketahui namun jumlahnya tetap. Sebuah sampel acak $X_1, X_2, \ldots, X_n$ diambil dari populasi dengan probabilitas fungsi kepadatan $f(x;\theta)$ dan didasari pada nilai yang diamati pada sampel.

Pada pendekatan Bayes, θ ditinjau dari jumlah dengan variasi yang dapat dijelaskan dengan probabilitas distribusi (dikenal sebagai distribusi prior). Subjek distribusi ini, berdasarkan asumsi pada percobaan dan diformulasikan sebelum data terlihat (dan karenanya dinamakan distribusi prior). Sebuah sampel diambil dari populasi dimana θ adalah parameter dan distribusi prior diperbarui dengan informasi sampel. Pembaruan prior dinamakan distribusi posterior.

Pada bagian ini, kita akan menotasikan populasi kepadatan $f(x;\theta)$ menjadi $f(x|\theta)$, yakni kepadatan dari populasi X diberikan parameter θ.

Definisi 5.5.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan kepadatan $f(x|\theta)$ dimana θ adalah parameter yang tidak diketahui untuk di estimasi. Fungsi kepadatan probabilitas dari variabel acak θ dinamakan distribusi prior dari θ dan biasanya dinotasikan $h(\theta)$.

Definisi 5.6.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan kepadatan $f(x|\theta)$ dimana θ adalah parameter yang tidak diketahui untuk di estimasi. Kepadatan bersyarat $k(\theta | x_1, x_2, \ldots, x_n)$ dari θ yang diberikan sampel $x_1, x_2, \ldots x_n$ disebut distribusi posterior dari θ.

Contoh 5.22.

Misalkan $X_1 = 1, X_2 = 2$ adalah sampel acak berukuran 2 dari sebuah distribusi dengan fungsi kepadatan

$$f(x|\theta) = \binom{3}{x} \theta^x (1-\theta)^{3-x}, \qquad x = 0,1,2,3.$$

Jika kepadatan dari prior θ adalah

$$h(\theta)=\begin{cases} k & jika\,\dfrac{1}{2}<\theta<1 \\ 0 & untuk\ yang\ lainnya \end{cases}$$

carilah distribusi posterior dari θ?

Jawab :

Ketika $h(\theta)$ adalah probabilitas kepadatan dari θ, kita punya

$$\int_{\frac{1}{2}}^{1} h(\theta)\,d\theta=1$$

Dengan implikasi

$$\int_{\frac{1}{2}}^{1} k\,d\theta=1$$

Oleh karena itu, $k=2$. Kepadatan gabungan dari sampel dan parameter diberikan oleh

$$u(x_1,x_2,\theta)=f(x_1\,|\,\theta)\,f(x_2\,|\,\theta)\,h(\theta)$$

$$=\binom{3}{x_1}\theta^{x_1}(1-\theta)^{3-x_1}\binom{3}{x_2}\theta^{x_2}(1-\theta)^{3-x_2}\,2$$

$$=2\binom{3}{x_1}\binom{3}{x_2}\theta^{x_1+x_2}(1-\theta)^{6-x_1-x_2}$$

Karenanya

$$u(1,2,\theta)=2\binom{3}{1}\binom{3}{2}\theta^{3}(1-\theta)^{3}$$

$$=18\theta^{3}(1-\theta)^{3}$$

Distribusi marjinal dari sampel

$$g(1,2) = \int_{\frac{1}{2}}^{1} u(1,2,\theta)\, d\theta$$

$$= \int_{\frac{1}{2}}^{1} 18\theta^3 (1-\theta)^3 \, d\theta$$

$$= 18\int_{\frac{1}{2}}^{1} \theta^3 \left(1 + 3\theta^3 - 3\theta - \theta^3\right) d\theta$$

$$= 18\int_{\frac{1}{2}}^{1} \left(\theta^3 + 3\theta^5 - 3\theta^4 - \theta^6\right) d\theta$$

$$= \frac{9}{140}$$

Distribusi bersyarat dari parameter θ yang diberikan sampel $X = 1, X_2 = 2$ adalah

$$k(\theta \mid x_1 = 1, x_2 = 2) = \frac{u(1,2,\theta)}{g(1,2)}$$

$$= \frac{18\theta^3 (1-\theta)^3}{\frac{9}{140}}$$

$$= 280\theta^3 (1-\theta)^3$$

Oleh karena itu, distribusi posterior dari θ ialah

$$k(\theta | x_1 = 1, x_2 = 2) = \begin{cases} 280\ \theta^3 (1-\theta)^3, & \textit{jika } \frac{1}{2} < \theta < 1 \\ 0, & \textit{untuk yang lainnya} \end{cases}$$

Catatan 5.2. Jika $X_1, X_2, \ldots, X_n$ adalah sampel acak dari populasi dengan kepadatan $f(x|\theta)$ lalu kepadatan gabungan dari sampel dan parameter diberikan oleh

$$u(x_1, x_2, \ldots, x_n, \theta) = h(\theta)\prod_{i=1}^{n} f(x_i \mid \theta)$$

Diberikan kepadatan gabungan dari sampel bisa dihitung menggunakan formula

$$g(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{\infty} h(\theta)\prod_{i=1}^{n} f(x_i \mid \theta)\, d\theta$$

Sekarang gunakan aturan Bayes, distribusi posterior dari θ bisa dihitung menggunakan

$$k(\theta \mid x_1, x_2, \ldots, x_n) = \frac{h(\theta)\prod_{i=1}^{n} f(x/\theta)}{\int_{-\infty}^{\infty} h(\theta)\prod_{i=1}^{n} f(x_i/\theta)\, d\theta}$$

Definisi 5.7.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan kepadatan $f(x \mid \theta)$ dimana θ adalah parameter yang tidak diketahui. Misal $\widehat{\theta}$ adalah estimator dari θ. Fungsi dari

$$\mathcal{L}_2(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$$

dinamakan *squared error loss*. Fungsi dari

$$\mathcal{L}_1(\hat{\theta}, \theta) = \left|\hat{\theta} - \theta\right|$$

dinamakan *absolute eror loss*.

Definisi 5.8

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan kepadatan $f(x \mid \theta)$ dimana θ adalah parameter yang tidak diketahui. Misalkan $\hat{\theta}$ adalah estimator dari θ dan misalkan $\mathcal{L}(\hat{\theta}, \theta)$ adalah fungsi *loss*. Maka nilai ekspektasi dari fungsi *loss* mengenai distribusi populasi $f(x \mid \theta)$ adalah

$$R_{\mathcal{L}}(\theta) = \int \mathcal{L}(\hat{\theta}, \theta)\, f\left(\frac{x}{\theta}\right) dx$$

Ini dinamakan resiko.

Kepadatan posterior dari parameter θ yang diberikan sampel $x_1, x_2, \ldots, x_n$ adalah

$$k(\theta \mid x_1, x_2, \ldots, x_n)$$

Memuat semua informasi θ. Dalam metode Bayes estimasi dari parameter dipilih satu estimasi $\hat{\theta}$ untuk θ sehingga

$$k(\hat{\theta} \mid x_1, x_2, \ldots, x_n)$$

Maksimum subjek untuk fungsi *loss*. Secara matematis, persamaan ini meminimalkan integral

$$\int_{\Omega} \mathcal{L}\left(\hat{\theta}, \theta\right) k(\theta \mid x_1, x_2, \ldots, x_n) d\theta$$

Dengan $\hat{\theta}$ dimana Ω dinotasikan sebagai pendukung kepadatan prior $h(\theta)$ dari parameter θ.

Contoh 5.23. Misalkan satu pengamatan diambil dari variabel acak X yang menghasilkan nilai 2. Kepadatan fungsi dari X adalah

$$f\left(x \middle| \theta\right) = \begin{cases} \dfrac{1}{2} & \text{jika } 0 < x < \theta \\ 0 & \text{untuk yang lainnya} \end{cases}$$

Dan distribusi prior untuk parameter θ adalah

$$h(\theta) = \begin{cases} \dfrac{3}{\theta^4} & \text{jika } 1 < \theta < \infty \\ 0 & \text{untuk yang lainnya} \end{cases}$$

jika fungsi *loss* adalah $\mathcal{L}(z, \theta) = (z - \theta)^2$, lalu berapakah estimasi Bayes untuk θ?

Jawab :

Kepadatan prior dari variabel acak θ adalah

$$h(\theta) = \begin{cases} \dfrac{3}{\theta^4} & \text{jika } 1 < \theta < \infty \\ 0 & \text{untuk yang lainnya} \end{cases}$$

Karenanya, fungsi kepadatan probabilitas gabungan dari sampel dan parameter adalah

$$u(x,\theta) = h(\theta) f(x/\theta)$$

$$= \frac{3}{\theta^4}\frac{1}{\theta}$$

$$= \begin{cases} 3\theta^{-5} & jika\, 0 < x < \theta \quad dan\ 1 < \theta < \infty \\ 0 & untuk\ yang\, lainnya \end{cases}$$

Kepadatan marginal dari sampel adalah

$$g(x) = \int_x^\infty u(x,\theta)\, d\theta$$

$$= \int_x^\infty 3\theta^{-5} d\theta$$

$$= \frac{3}{4}x^{-4}$$

$$= \frac{3}{4x^4}$$

Jadi, jika $x = 2, g(2) = \dfrac{3}{64}$. Kepadatan posterior dari θ ketika $x = 2$ adalah

$$f(\theta \mid x = 2) = \frac{u(2,\theta)}{g(2)}$$

$$= \frac{64}{3} 3\theta^{-5}$$

$$= \begin{cases} 64\theta^{-5} & jika\, 2 < \theta < \infty \\ 0 & untuk\ yang\, lainnya \end{cases}$$

Sekarang, kita cari estimator Bayes dengan meminimalkan pernyataan $E\,|\,\mathcal{L}(\theta, z)\,/\,x = 2)$. Bahwa

$$\hat{\theta} = Arg \max_{z \in \Omega} \int_{\Omega} \mathcal{L}(\theta, z)\, k(\theta \,|\, x = 2)\, d\theta$$

Kemudian

$$\psi(z) = \int_{\Omega} \mathcal{L}(\theta, z)\, k(\theta \,|\, x = 2)\, d\theta$$

$$= \int_{2}^{\infty} (z - \theta)^2\, k(\theta \,|\, x = 2)\, d\theta$$

$$= \int_{2}^{\infty} (z - \theta)^2\, 64\theta^{-5}\, \theta$$

Kita akan mencari nilai dari z yang menghasilkan minimum dari $\psi(z)$. Ini bias dilakukan dengan mengambil turunan dari $\psi(z)$ dan mengevaluasi dimana turunannya adalah 0.

$$\frac{d}{dz}\psi(z) = \frac{d}{dz}\int_{2}^{\infty} (z - \theta)^2\, 64\theta^{-5}\, d\theta$$

$$= 2\int_{2}^{\infty} (z - \theta)\, 64\theta^{-5} d\theta$$

$$= 2\int_{2}^{\infty} z\, 64\theta^{-5}\, d\theta - 2\int_{2}^{\infty} \theta\, 64\theta^{-5}\, d\theta$$

$$= 2z - \frac{16}{3}$$

Selesaikan turunan dari $\psi(z)$ ke 0 dan menyelesaikan z, kita punya

$$2z - \frac{16}{3} = 0$$

$$\Rightarrow z = \frac{8}{3}$$

Saat $\frac{d^2\psi(z)}{dz^2} = 2$, fungsi $\psi(z)$ minimum di $z = \frac{8}{3}$. Oleh karena itu, estimasi Bayes dari θ adalah $\frac{8}{3}$

Pada contoh 5.23, kita mencari estimasi Bayes dari θ langsung diminimalkan $\int_\Omega \mathcal{L}(\hat{\theta},\theta) k\left(\frac{\theta}{x_1}, x_2, \ldots, x_n\right) d\theta$ dengan menyelesaikan $\hat{\theta}$. Hasil selanjutnya sangat berguna ketika mencari estimasi Bayes dengan menggunakan fungsi *loss* kuadrat. Ingat bahwa jika $\mathcal{L}(\hat{\theta},\theta) = (\theta - \hat{\theta})^2$, lalu $\mathcal{L}(\hat{\theta},\theta) = (\theta - \hat{\theta}) \int_\Omega k(\theta | x_1, x_2, \ldots x_n) d\theta$ adalah $E((\theta - \hat{\theta}^2 \mid x_1, x_2, \ldots x_n)$. Berdasarkan teorema pada kenyataan bahwa fungsi $\varnothing$ mendefinisikan sebagai $\varnothing(c) = E\left[(X-c)^2\right]$ mencapai minimum jika $c = E[X]$

Teorema 5.3.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan kepadatan $f(x \mid \theta)$ dimana θ adalah parameter yang tidak diketahui. Jika fungsi *loss* adalah kuadrat eror, estimator $\hat{\theta}$ Bayes dari parameter θ adalah

$$\hat{\theta} = E(\theta \mid x_1, x_2, \ldots, x_n)$$

Dimana ekspektasi diambil dengan menghitung kepadatan $k(\theta \mid x_1, x_2, \ldots, x_n)$

Contoh 5.24.

Misalkan distribusi prior dari θ seragam pada interval $(0,1)$. Diberikan θ, populasi dari X seragam pada interval $(0,\theta)$. Jika fungsi *squared loss eror* digunakan, carilah estimator Bayes $\hat{\theta}$ berdasarkan sampel berukuran 1.

Jawab :

Kepadatan prior dari θ adalah

$$h(\theta) = \begin{cases} 1 & jika\, 0 < \theta < 1 \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Kepadatan dari populasi adalah

$$f(x \mid \theta) = \begin{cases} \frac{1}{\theta} & jika\, 0 < x < \theta \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Kepadatan gabungan dari sampel dan parameter adalah

$$u(x,\theta) = h(\theta) f(x \mid \theta)$$

$$= 1\left(\frac{1}{\theta}\right)$$

$$= \begin{cases} \dfrac{1}{\theta} & jika\, 0 < x < \theta \\ 0 & untuk\ yang\, lainnya \end{cases}$$

Kepadatan marjinal dari sampel adalah

$$g(x) = \int_x^1 u(x,\theta)\, d\theta$$

$$= \int_x^1 \frac{1}{\theta} d\theta$$

$$= \begin{cases} -\ln x & jika\, 0 < x < 1 \\ 0 & untuk\ yang\, lainnya \end{cases}$$

Syarat kepadatan dari θ diberikan sampel

$$k(\theta \mid x) = \frac{u(x,\theta)}{g(x)}$$

$$= \begin{cases} -\dfrac{1}{\theta \ln x} & jika\, 0 < x < \theta < 1 \\ 0 & untuk\ yang\, lainnya \end{cases}$$

Saat fungsi *loss* adalah kuadrat eror, karena itu estimator Bayes dari θ adalah

$$\hat{\theta} = E[\theta \mid x]$$

$$= \int_x^1 \theta k(\theta \mid x) d\theta$$

$$= \int_x^1 \theta \frac{(-1)}{\theta \ln x} d\theta$$

$$= \frac{1}{\ln x} \int_x^1 d\theta$$

$$= \frac{(x-1)}{\ln x}$$

Jadi, estimator Bayes dari θ berdasarkan 1 observasi dari X adalah

$$\hat{\theta} = \frac{X-1}{\ln X}$$

Contoh 5.25. Diberikan θ, variabel acak X mempunyai distribusi binomial dengan $n = 2$ dan probabilitas sukses θ. Jika kepadatan dari prior adalah

$$h(\theta) = \begin{cases} k & jika\, \frac{1}{2} < \theta < 1 \\ 0 & untuk\ yang\ lainnya \end{cases}$$

berapa estimator Bayes dari θ untuk *squared error loss* jika $X = 1$?

Jawab :

Ingat bahwa θ adalah uniform pada interval $\left(\frac{1}{2}, 1\right)$ karena $k = 2$ karenanya kepadatan prior dari θ adalah

$$h(\theta) = \begin{cases} 2 & jika\, \frac{1}{2} < \theta < 1 \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Kepadatan populasi

$$f(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

$$= \binom{2}{x} \theta^x (1-\theta)^{2-x} \quad , \qquad x = 0,1,2$$

Gabungan kepadatan dari sampel dan parameter θ adalah

$$u(x,\theta) = h(\theta) f(x|\theta)$$

$$= 2\binom{2}{x} \theta^x (1-\theta)^{2-x}$$

Dimana $\frac{1}{2} < \theta < x$ dan $x = 0,1,2$. Kepadatan marjinal dari sampel adalah

$$g(x) = \int_{\frac{1}{2}}^{1} u(x,\theta)\, d\theta$$

Integral ini mudah dievaluasi jika kita masukan $X = 1$, karenanya

$$g(1) = \int_{\frac{1}{2}}^{1} 2\binom{2}{1} \theta(1-\theta)\, d\theta$$

$$= \int_{\frac{1}{2}}^{1} \left(4\theta - 4\theta^2\right) d\theta$$

$$= 4\left[\frac{\theta^2}{2} - \frac{\theta^3}{3}\right]_{\frac{1}{2}}^{1}$$

$$= \frac{2}{3}\left[3\theta^2 - 2\theta^3\right]_{\frac{1}{2}}^{1}$$

$$= \frac{2}{3}\left[(3-2) - \left(\frac{3}{4} - \frac{2}{8}\right)\right]$$

$$= \frac{1}{3}$$

Oleh karena itu, kepadatan posterior dari θ diberikan $x = 1$ adalah

$$k(\theta / x = 1) = \frac{u(1,\theta)}{g(1)}$$

$$= 12(\theta - \theta^2)$$

Dimana $\frac{1}{2} < \theta < 1$. Ketika fungsi *loss* adalah kuadrat eror, karenanya estimasi Bayes dari θ adalah

$$\hat{\theta} = E[\theta / x = 1]$$

$$= \int_{\frac{1}{2}}^{1} \theta k(\theta / x = 1) d\theta$$

$$= \int_{\frac{1}{2}}^{1} 12\theta(\theta - \theta^2) d\theta$$

$$= \left[4\theta^3 - 3\theta^4\right]_{\frac{1}{2}}^{1}$$

$$= 1 - \frac{5}{16}$$

$$= \frac{11}{16}$$

Karena berdasarkan sampel berukuran n dengan $X = 1$, maka estimasi Bayes dari θ adalah $\frac{11}{16}$.

Teorema 5.4.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan kepadatan $f(x \mid \theta)$ dimana θ belum diketahui parameternya. Jika fungsi *loss* adalah absolut eror, estimator $\hat{\theta}$ Bayes dari parameter θ, diberikan oleh

$$\hat{\theta} = median\, dari\, k\left(\theta \middle| x_1, x_2, \ldots, x_3\right)$$

Dimana $k\left(\theta \middle| x_1, x_2, \ldots, x_3\right)$ adalah distribusi posterior dari θ.

Contoh 5.26.

Diberikan θ, variabel acak X adalah binomial distribusi dengan $n = 3$ dan probabilitas sukses θ. Jika kepadatan prior dari θ adalah

$$h(\theta) = \begin{cases} k & jika\,\frac{1}{2} < \theta < 1 \\ 0 & untuk\ yang\, lainnya \end{cases}$$

Berapakah estimasi Bayes dari θ adalah absolut *loss* eror yang berbeda jika sampel terdiri dari 1 pengamatan $x = 3$?

Jawab :

Ketika kepadatan prior dari θ adalah

$$h(\theta) = \begin{cases} 2 & jika\,\frac{1}{2} < \theta < 1 \\ 0 & untuk\ yang\, lainnya \end{cases}$$

kepadatan populasi

$$f(x \mid \theta) = \binom{3}{x} \theta^x (1-\theta)^{3-x}$$

Kepadatan gabungan dari sampel dan parameter adalah

$$u(3,\theta) = h(\theta) f(3 \mid \theta)$$

$$= 2\theta^3$$

Dimana $\frac{1}{2} < x < 1$. Marjinal kepadatan dari sampel (titik $x = 3$) adalah

$$g(3) = \int_{\frac{1}{2}}^{1} u(3,\theta)\, d\theta$$

$$= \int_{\frac{1}{2}}^{1} 2\theta^3 d\theta$$

$$= \left[\frac{\theta^4}{2}\right]_{\frac{1}{2}}^{1}$$

$$= \frac{15}{32}$$

Karena itu, kepadatan bersyarat dari θ adalah $X = 3$, maka

$$k(\theta \mid x = 3) = \frac{u(3,\theta)}{g(3)}$$

$$= \begin{cases} \frac{64}{15}\theta^3 & jika\, \frac{1}{2} < \theta < 1 \\ 0 & untuk\ yang\, lainnya \end{cases}$$

Saat fungsi *loss* adalah absolut eror, maka estimator Bayes adalah median dari fungsi kepadatan probabilitas $k\left(\theta \middle| x = 3\right)$.

$$\frac{1}{2} = \int_{\frac{1}{2}}^{\hat{\theta}} \frac{64}{15}\theta^3 d\theta$$

$$= \frac{64}{60}\left[\theta^4\right]_{\frac{1}{2}}^{\hat{\theta}}$$

$$= \frac{64}{60}\left[\left(\hat{\theta}\right)^4 - \frac{1}{16}\right]$$

$$\hat{\theta} = \sqrt[4]{\frac{17}{32}}$$

$$= 0,8537$$

Contoh 5.27. Misalkan distribusi prior dari θ adalah uniform pada interval $(2,5)$. Diberikan θ, X adalah uniform pada interval $(0,\theta)$. Berapakah estimator Bayes dari θ untuk *loss* absolut jika $X = 1$?

Jawab :

ketika kepadatan prior dari θ adalah

$$h(\theta) = \begin{cases} \dfrac{1}{3} & jika\, 2 < \theta < 5 \\ 0 & untuk\ yang\ lainnya \end{cases}$$

dan kepadatan populasi

$$f(x/\theta) = \begin{cases} \dfrac{1}{\theta} & jika\, 0 < x < \theta \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Kepadatan gabungan dari sampel dan parameter adalah

$$u(x,\theta) = h(\theta) f(x|\theta)$$

$$= \frac{1}{3\theta}$$

Dimana $2 < \theta < 5$ dan $0 < x < \theta$ kepadatan marginal dari sampel pada $x = 1$ adalah

$$g(1) = \int_1^5 u(1,\theta)\, d\theta$$

$$= \int_1^2 u(1,\theta)\,d\theta + \int_2^5 u(1,\theta)\,d\theta$$

$$= \int_2^5 \frac{1}{3\theta}\,d\theta$$

$$= \frac{1}{3}\ln\left(\frac{5}{2}\right)$$

Oleh karena itu, kepadatan bersyarat dari θ sampel $x=1$ adalah

$$k(\theta|x=1) = \frac{u(1,\theta)}{g(1)}$$

$$= \frac{1}{\theta\ln\left(\frac{5}{2}\right)}$$

Saat fungsi *loss* adalah absolut eror, estimasi Bayes dari θ adalah median dari $k(\theta|x=1)$ karenanya

$$\frac{1}{2} = \int_2^{\hat{\theta}} \frac{1}{\theta\ln\left(\frac{5}{2}\right)}\,d\theta$$

$$= \frac{1}{\ln\left(\frac{5}{2}\right)}\ln\left(\frac{\hat{\theta}}{2}\right)$$

Penyelesaian untuk $\hat{\theta}$, kita punya

$$\hat{\theta} = \sqrt{10} = 3{,}16$$

■ **Contoh 5.28.** Apa prinsip dasar dari estimasi Bayes?

Jawab :

Prinsip dasar estimasi metode Bayes terdiri dari memilih nilai dari parameter θ dengan data yang diamati tinggi dan sebuah probabilitas posterior $\hat{\theta} = Arg\sup_{\theta\in\Omega} L(\theta)$ dari θ yang memungkinkan subjek untuk sebuah fungsi *loss*.

Latihan Soal BAB 5

1. Tentukan *MME* (Metode Momen Estimator) dari θ yang bergantung pada sampel acak $x_1, x_2, x_3, \ldots, x_n$ dengan

$$f(x;\theta)\begin{cases} \theta x^{\theta-1}; & untuk\, 0<x<1 \\ \\ 0; & untuk\; yang\; lainnya,\; 0<\theta \end{cases}$$

Penyelesaian :

$$f(x;\theta)=\theta x^{\theta-1};0\cup x\langle 1;0\rangle\theta$$

$$E(x)=\int_0^1 x.f(x;\theta)\,dx$$

$$=\int_0^1 x.\theta x^{\theta-1}dx$$

$$=\theta\int_0^1 x^{\theta}dx$$

$$=\theta\left[\frac{1}{\theta+1}x^{\theta+1}\right]$$

$$=\theta\left[\frac{1}{\theta+1}.1\right]=\frac{\theta}{\theta+1};\theta>0$$

$$M_1^{'}=E(x)$$

$$\frac{\sum x_i}{n}=E(x)$$

$$\overline{x}=\frac{\theta}{\theta+1}$$

$$\theta\overline{x}+\overline{x}=\theta$$

$$\theta-\theta\overline{x}=\overline{x}$$

$$\theta\left(1-\overline{x}\right)=\overline{x}$$

$$\hat{\theta}=\frac{\overline{x}}{1-\overline{x}}$$

Jadi *MME* (Metode Momen Estimator) dari θ adalah $\hat{\theta}=\frac{\overline{x}}{1-\overline{x}}$

2. Dapatkan *MME* (Metode Momen Estimator) dari sampel random berukuran n dari distribusi $x_i \sim GAM\left(2,k\right)$

   Penyelesaian :

$$E\left(x\right)=\mu=k\theta=k.2=2k$$

$$M_1^{'}=\frac{\sum x_i}{n}=\overline{x}$$

$$M_1^{'}=2k$$

$$\overline{x}=2k$$

$$\hat{k}=\frac{\overline{x}}{2}$$

Jadi *MME* (Metode Momen Estimator) dari $x_i \sim GAM\left(2,k\right)$ adalah $\hat{k}=\frac{\overline{x}}{2}$

3. Dapatkan MME (Metode Momen Estimator) dari sampel random berukuran n dari distribusi $x_i \sim DE\left(\theta,\eta\right)$ dengan θ dan η tidak diketahui

   Penyelesaian :

$$E\left(x\right)=\mu=\eta$$

$$M_1^{'} = E(x) = \overline{x}$$

$$\eta = \overline{x}$$

$$\hat{\eta} = \overline{x}$$

$$var(x) = 2\theta^2$$

$$E(x^2) = \sigma^2 + \mu^2$$

$$= 2\theta^2 + \eta^2$$

$$= 2\theta^2 + \overline{x}^2$$

$$M_2^{'} = E(x^2) = \frac{\sum x_i^2}{n}$$

Persamaan $a = b$

$$2\theta^2 + \overline{x}^2 = \frac{\sum x_i^2}{n}$$

$$2\theta^2 = \frac{\sum x_i^2}{n} - \overline{x}^2$$

$$2\theta^2 = \frac{\sum (x_i - \overline{x})^2}{n}$$

$$\theta^2 = \frac{\sum (x_i - \overline{x})^2}{2n}$$

$$\theta = \sqrt{\frac{\sum\left(x_i - \overline{x}\right)^2}{2n}}$$

$$\hat{\theta} = \sqrt{\frac{\sum\left(x_i - \overline{x}\right)^2}{2n}} = \sqrt{\frac{\left(\frac{n-1}{n}\right)S^2}{2}}$$

Jadi *MME* (Metode Momen Estimator) dari $x_i \sim DE(\theta, \eta)$ adalah

$$\hat{\theta} = \sqrt{\frac{\sum\left(x_i - \overline{x}\right)^2}{2n}} = \sqrt{\frac{\left(\frac{n-1}{n}\right)S^2}{2}}$$

4. Dapatkan *MME* (Metode Momen Estimator) dari sampel random berukuran n dari distribusi $X_i \sim NB(3, p)$

   Penyelesaian :

$$E(x) = \mu = \frac{r}{p} = \frac{3}{p}$$

$$M_1^{'} = \frac{\sum x_i}{n} = \overline{x}$$

$$(1) = (2), maka \frac{3}{p} = \overline{x} \rightarrow p = \frac{3}{\overline{x}}$$

$$E(x) = \frac{r}{p}$$

$$M_1^{'} = \frac{3}{p}$$

$$\overline{x} = \frac{3}{p}$$

$$\hat{p} = \frac{3}{\overline{x}}$$

5. Dapatkan *MME* (Metode Momen Estimator) dari sampel random berukuran n dari distribusi $X_i \sim WEI\left(\theta, \frac{1}{2}\right)$

Penyelesaian :

$$E(x) = \theta\,\Gamma\left(1+\frac{1}{\beta}\right) = \theta\,\Gamma\left(1+\frac{1}{2}\right) = \theta\,\Gamma\left(\frac{3}{2}\right)$$

$$M_1^{'} = \frac{\sum x_i}{n} = \overline{x}$$

$$(1) = (2),\ \text{maka}\ \theta\,\Gamma\left(\frac{3}{2}\right) = \overline{x}$$

$$\overline{\theta} = \frac{\overline{x}}{\Gamma\left(\frac{3}{2}\right)}$$

6. Temukan MLE berdasarkan pada sampel acak $X_1, \ldots, X_n$ dari distribusi Pareto $X_i \sim PAR(1, k)$ dimana k tidak diketahui.

Penyelesaian :

Karena

$$f(x;k) = k(1+x)^{-x-1},\ x > 0$$

Selanjutnya

$$lnL(k) = n\ln k - (k+1)\sum \ln(1+x_i)$$

Dan persamaan ML adalah

$$\frac{d}{dk}\ln L(k) = \frac{n}{k} - \sum \ln(1+x_i) = 0$$

Didapat MLE

$$\hat{k} = \frac{n}{\sum \ln(1+x_i)}$$

Untuk mencari CRLB *(Cramer Rao Lower Bound)*

$$\ln f(x;k) = \ln k - (k+1)\ln(1+x)$$

$$\frac{\partial}{\partial k}\ln f(x;k) = \frac{1}{k} - \ln(1+x)$$

Sehingga

$$CRLB = \frac{1}{n}E\left[\frac{1}{k} - \ln(1+x)\right]^2$$

Untuk mengevaluasi persamaan diatas dimisalkan dengan transformasi

$$Y = \ln(1+x) \sim EXP\left(\frac{1}{k}\right)$$

Sehingga

$$E\left[\ln(1+x)\right] = \frac{1}{k}$$

$$E\left[\frac{1}{k} - \ln(1+x)\right]^2 = Var\left[\ln(1+x)\right] = \frac{1}{k^2}$$

$$Var(\hat{k}) = CRLB = \frac{k^2}{n}$$

$$\hat{k} \sim N\left(k, \frac{k^2}{n}\right)$$

7. Diberikan $Y_1, Y_2, \ldots, Y_n$ yang menjadi sampel acak dari distribusi Normal dengan mean μ dan varians σ^2. Temukan MLE dari μ dan σ^2.

Penyelesaian :

Karena $Y_1, Y_2, \ldots, Y_n$ adalah variabel acak kontinu, $L(\mu, \sigma^2)$ adalah gabungan densitas dari sampel. Sehingga $L(\mu, \sigma^2) = f(y_1, y_2, \ldots, y_n \mid \mu, \sigma^2)$. Pada kasus ini,

$$L(\mu, \sigma^2) = f(y_1, y_2, \ldots, y_n \mid \mu, \sigma^2)$$

$$= f(y_1 \mid \mu, \sigma^2) \times (y_2 \mid \mu, \sigma^2) \times \ldots \times (y_n \mid \mu, \sigma^2)$$

$$= \left\{ \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(y_1 - \mu)^2}{2\sigma^2} \right] \right\} \times \ldots \times \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(y_n - \mu)^2}{2\sigma^2} \right]$$

$$= \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp\left[\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 \right]$$

Lebih lanjut

$$\ln\left[L(\mu, \sigma^2) \right] = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2$$

MLE dari μ dan σ^2 adalah nilai yang membuat $\ln\left[L(\mu, \sigma^2) \right]$ menjadi maksimum. Ambil turunan dengan memperhatikan μ dan σ^2, kita peroleh

$$\frac{\partial \left\{ \ln\left[L(\mu, \sigma^2) \right] \right\}}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \mu)$$

Dan

$$\frac{\partial \left\{ \ln\left[L(\mu, \sigma^2) \right] \right\}}{\partial \mu} = -\left(\frac{n}{2} \right) \left(\frac{1}{\sigma^2} \right) + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (y_i - \mu)^2$$

Atur turunan ini dengan di sama dengankan nol dan selesaikan, lebih lanjut dari persamaan pertama

$$\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(y_i-\hat{\mu})=0,\ atau \qquad \sum_{i=1}^{n}y_i-n\hat{\mu}=0,\ dan \qquad \hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}y_i=\bar{y}$$

Substitusikan $\bar{y}$ terhadap $\hat{\mu}$ pada persamaan kedua dan temukan $\hat{\sigma}^2$, kita dapatkan

$$-\left(\frac{n}{\hat{\sigma}^2}\right)+\frac{1}{\hat{\sigma}^4}\sum_{i=1}^{n}(y_i-\bar{y})^2=0,\ atau \qquad \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

Jadi, $\bar{Y}$ dan $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$ adalah MLE dari masing-masing μ dan σ^2. Perhatikan bahwa $\bar{Y}$ adalah tidak bias untuk μ. Meskipun $\hat{\sigma}^2$ adalah bukan tidak bias untuk σ^2, ini dapat disesuaikan dengan estimator tidak bias S^2

8. Misalkan $Y_1, Y_2, \ldots, Y_n$ menjadi sampel acak dari distribusi Bernoulli dimana $P(Y_i=1)=p$ dan $P(Y_i=0)=1-p$ dan asumsikan bahwa distribusi prior dari p adalah beta (α,β). Temukan distribusi posterior dari p

Penyelesaian :

Karena fungsi probabilitas Bernoulli, dapat dituliskan sebagai berikut

$$p(y_i \mid p)=p^{y_i}(1-p)^{1-y_i},\ y_i=0,1$$

Likelihood $L(y_1, y_2, \ldots, y_n \mid p)$ adalah

$$L(y_1, y_2, \ldots, y_n \mid p)=p(y_1, y_2, \ldots, y_n \mid p)$$

$$=p^{y_1}(1-p)^{1-y_1}\times p^{y_2}(1-p)^{1-y_2}\times\ldots\times p^{y_n}(1-p)^{1-y_n}$$

$$=p^{\Sigma y_i}(1-p)^{n-\Sigma y_i},\ y_i=0,1\ dan\ 0<p<1$$

Kemudian,

$$f(y_1, y_2, \ldots, y_n, p)=L(y_1, y_2, \ldots, y_n \mid p)\times g(p)$$

$$=p^{\Sigma y_i}(1-p)^{n-\Sigma y_i}\times\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\Sigma y_i + \alpha - 1} (1-p)^{n - \Sigma y_i + \beta - 1}$$

Dan

$$m(y_1, y_2, \ldots, y_n) = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\Sigma y_i + \alpha - 1} (1-p)^{n - \Sigma y_i + \beta - 1} dp$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\Sigma y_i + \alpha)\Gamma(n - \Sigma y_i + \beta)}{\Gamma(n+\alpha+\beta)}$$

Akhirnya, densitas posterior dari p adalah

$$g^\star(p | y_1, y_2, \ldots, y_n) = \frac{\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\Sigma y_i + \alpha - 1} (1-p)^{n - \Sigma y_i + \beta - 1}}{\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \dfrac{\Gamma(\Sigma y_i + \alpha)\Gamma(n - \Sigma y_i + \beta)}{\Gamma(n+\alpha+\beta)}}, \; 0 < p < 1$$

$$= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\Sigma y_i + \alpha)\Gamma(n - \Sigma y_i + \beta)} \times p^{\Sigma y_i + \alpha - 1} (1-p)^{n - \Sigma y_i + \beta - 1}, \; 0 < p < 1$$

Densitas *beta* dengan parameter $\alpha^\star = \Sigma y_i + \alpha$ dan $\beta^\star = n - \Sigma y_i + \beta$

9. Misalkan $Y_1, Y_2, \ldots, Y_n$ menyatakan sampel acak dari populasi normal dengan mean μ tidak diketahui dan varians σ_o^2 yang diketahui. Distribusi prior konjugasi untuk μ adalah distribusi Normal dengan mean η yang diketahui dan varians δ^2 yang diketahui. Temukan distribusi posterior dan estimator Bayes untuk μ.

Penyelesaian :

Karena $U = \Sigma Y_i$ adalah statistic kecukupan dari μ dan diketahui memiliki distribusi Normal dengan mean $n\mu$ dan varians $n\sigma_o^2$,

$$L(u | \mu) = \frac{1}{\sqrt{2\pi n\sigma_o^2}} \exp\left[\frac{1}{2n\sigma_o^2}(u - n\mu)^2\right], \; -\infty < u < \infty$$

Dan gabungan densitas dari U dan μ adalah

$$f(u,\mu) = L(u \mid \mu) \times g(\mu)$$

$$= \frac{1}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}} \exp\left[\frac{1}{2n\sigma_o^2}(u-n\mu)^2 - \frac{1}{2\delta^2}(\mu-\eta)^2\right], -\infty < u < \infty, -\infty < \mu < \infty$$

Mari kita lihat kuantitas dalam eksponen di atas

$$-\frac{1}{2n\sigma_o^2}(u-n\mu)^2 - \frac{1}{2\delta^2}(\mu-\eta)^2 = -\frac{1}{2n\sigma_o^2\delta^2}\left[\delta^2(u-n\mu)^2 + n\sigma_o^2(\mu-\eta)^2\right]$$

$$= -\frac{1}{2n\sigma_o^2\delta^2}\Big[\delta^2u^2 - 2\delta^2un\mu + \delta^2n^2\mu^2 + n\sigma_o^2\mu^2 - 2n\sigma_o^2\mu\eta + n\sigma_o^2\eta^2\Big]$$

$$= -\frac{1}{2n\sigma_o^2\delta^2}\Big[\left(n^2\delta^2 + n\sigma_o^2\right)\mu^2 - 2\left(n\delta^2u + n\sigma_o^2\eta\right)\mu + \delta^2u^2 + n\sigma_o^2\eta^2\Big]$$

$$= -\frac{1}{2n\sigma_o^2\delta^2}\left[\left(n\delta^2 + \sigma_o^2\right)\mu^2 - 2\left(\delta^2u + \sigma_o^2\eta\right)\mu\right] - \frac{1}{2n\sigma_o^2\delta^2}\left(\delta^2u^2 + n\sigma_o^2\eta^2\right)$$

$$= -\frac{n\delta^2 + \sigma_o^2}{2n\sigma_o^2\delta^2}\left[\mu^2 - 2\left(\frac{\delta^2u + \sigma_o^2\eta}{n\delta^2 + \sigma_o^2}\right)\mu + \left(\frac{\delta^2u + \sigma_o^2\eta}{n\delta^2 + \sigma_o^2}\right)^2\right] - \frac{1}{2n\sigma_o^2\delta^2}\left[\delta^2u^2 + n\sigma_o^2\eta^2 - \frac{n\left(\delta^2u + \sigma_o^2\eta\right)^2}{n\delta^2 + \sigma_o^2}\right]$$

Akhirnya, kita peroleh :

$$-\frac{1}{2n\sigma_o^2}(u-n\mu)^2 - \frac{1}{2\delta^2}(\mu-\eta)^2 = -\frac{n\delta^2 + \sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu - \frac{\delta^2u + \sigma_o^2\eta}{n\delta^2 + \sigma_o^2}\right)^2 - \frac{1}{2\left(n^2\delta^2 + n\sigma_o^2\right)}(u-n\eta)^2$$

Oleh karena itu,

$$f(u,\mu)=\frac{1}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}}\exp\left[-\frac{1}{2n\sigma_o^2}(u-n\mu)^2-\frac{1}{2\delta^2}(\mu-\eta)^2\right]$$

$$=\frac{1}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}}\exp\left[-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu-\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2\right]$$
$$\times\exp\left[-\frac{1}{2\left(n^2\delta^2+n\sigma_o^2\right)}(u-n\eta)^2\right]$$

Dan

$$m(u)=\frac{\exp\left[-\frac{1}{2\left(n^2\delta^2+n\sigma_o^2\right)}(u-n\eta)^2\right]}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}}\int_{-\infty}^{\infty}\exp\left[-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu-\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2\right]d\mu$$

$$=\frac{\exp\left[-\frac{1}{2\left(n^2\delta^2+n\sigma_o^2\right)}(u-n\eta)^2\right]}{\sqrt{2\pi n\sigma_o^2}\sqrt{2\pi\delta^2}}\int_{-\infty}^{\infty}\frac{\exp\left[-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu-\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2\right]}{\sqrt{\frac{2\pi\sigma_o^2\delta^2}{n\delta^2+\sigma_o^2}}}d\mu$$

Integral diatas sebagai fungsi densitas normal dan karenanya sama dengan 1, kita peroleh bahwa fungsi densitas marginal dari U adalah normal dengan mean $n\eta$ dan varians $\left(n^2\delta^2+n\delta_o^2\right)$ Lebih lanjut, densitas posterior dari μ diberikan $U=u$ adalah

$$g^{\star}(\mu\,|u)=\frac{f(u,\mu)}{m(u)}=\frac{1}{\sqrt{\frac{2\pi\sigma_o^2\delta^2}{n\delta^2+\sigma_o^2}}}\exp\left[-\frac{n\delta^2+\sigma_o^2}{2\sigma_o^2\delta^2}\left(\mu-\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)^2,-\infty<\mu<\infty\right]$$

Densitas normal dengan mean

$$\eta^{\star}=\left(\frac{\delta^2u+\sigma_o^2\eta}{n\delta^2+\sigma_o^2}\right)$$

Dengan varians

$$\delta^{\star 2}=\left(\frac{\sigma_o^2\delta^2}{n\delta^2+\sigma_o^2}\right)$$

Oleh karena itu, estimator Bayes untuk μ adalah

$$\hat{\mu}_B = \frac{\delta^2 U + \sigma_o^2}{n\delta^2 + \sigma_o^2} = \frac{n\delta^2}{n\delta^2 + \delta_o^2}\bar{Y} + \frac{\sigma_o^2}{n\delta^2 + \sigma_o^2}\eta$$

Sekali lagi, estimator Bayes ini adalah rata-rata terbobot dari *MLE*, $\bar{Y}$, mean sampel, dan mean dari prior η. Saat ukuran sampel n meningkat, bobot yang diberikan ke mean sampel $\bar{Y}$ meningkat sedangkan bobot yang diberikan ke mean prior η menurun.

BAB VI

KRITERIA UNTUK MENGEVALUASI KUALITAS DARI ESTIMATOR

Kita telah melihat di Bab 5 bahwa, secara umum, metode estimasi parameter yang berbeda menghasilkan penduga/penaksir/perkiraan yang berbeda. Misalnya, jika $X \sim UNIF(0,\theta)$ dan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari populasi X, kemudian penduga/penaksir/perkiraan dari θ diperoleh oleh metode momen adalah

$$\hat{\theta}_{MM} = 2\overline{X}$$

Sedangkan penduga yang diperoleh dengan metode kemungkinan maksimum adalah

$$\hat{\theta}_{ML} = X_{(n)}$$

Dimana $\overline{X}$ dan $X_{(n)}$ masing-masing adalah rata-rata sampel dan statistik terurut ke-n. Sekarang pertanyaannya muncul : manakah dari dua penduga/penaksir/perkiraan yang lebih baik? Jadi, diperlukan beberapa kriteria untuk menilai kualitas dari sebuah estimator (penaksir). Beberapa kriteria yang diketahui untuk mengevaluasi kualitas sebuah estimator adalah :

1. Ketidakbiasan
2. Efisiensi dan Efisiensi Relatif
3. Ketidakbiasan Varians Minimum Seragam
4. Kecukupan
5. Konsistensi

Dalam bab ini, kita hanya akan memeriksa empat kriteria pertama secara rinci. Konsep ketidakberpihakan, efisiensi dan kecukupan diperkenalkan oleh Sir Ronald Fisher

6.1. Estimator Tidak Bias

Misalkan $X_1, X_2, \ldots, X_n$ adalah sebuah variabel acak dengan ukuran n dari sebuah populasi dengan fungsi kepadatan probabilitas $f(x;\theta)$. Penaksir $\hat{\theta}$ dari θ adalah fungsi dari variabel acak $X_1, X_2, ..., X_n$ yang bebas dari parameter θ. Sebuah Estimasi adalah nilai realisasi penaksir yang diperoleh saat sampel benar-benar diambil

Definisi 6.1.

Penaksir $\hat{\theta}$ dari θ adalah dikatakan menjadi penaksir tidak bias dari θ jika dan hanya jika

$$E\left(\hat{\theta}\right) = \theta$$

Jika $\hat{\theta}$ tidak bias, kemudian itu disebut estimasi bias dari θ

Penaksir dari sebuah parameter mungkin tidak sama dengan nilai aktual parameter untuk setiap realisasi sampel $X_1, X_2, ..., X_n$, tetapi jika tidak bias maka rata-rata akan sama dengan parameter.

Contoh 6.1.

Misalkan $X_1, X_2, ..., X_n$ adalah sampel acak dari sebuah populasi normal dengan mean μ dan varians $\sigma^2 > 0$. Apakah sampel mean $\overline{X}$ adalah penaksir tidak bias dari parameter μ ?

Jawab :

Saat, masing-masing $X_i \sim N\left(\mu, \sigma^2\right)$, kita mempunyai

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Maka sampel mean adalah normal dengan mean μ dan varians $\frac{\sigma^2}{n}$. Sehingga

$$E\left(\overline{X}\right) = \mu$$

Oleh karena itu, sampel mean $\overline{X}$ adalah penaksir tidak bias dari μ.

Contoh 6.2.

Misalkan $X_1, X_2, ..., X_n$ adalah sampel acak dari sebuah populasi normal dengan mean μ dan varians $\sigma^2 > 0$. Berapakah kemungkinan maksimum penaksir dari σ^2 ?. Apakah maksimum likelihood penaksir merupakan penaksir tidak bias dari parameter σ^2

Jawab :

Pada contoh 5.13, kita dapat tunjukan bahwa maksimum likelihood penaksir dari σ^2 adalah

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

Sekarang, kita periksa ketidakbiasan dari penaksir ini

$$E\left[\widehat{\sigma^2}\right]=E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right]$$

$$=E\left[\frac{n-1}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right]$$

$$=\frac{n-1}{n}E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right]$$

$$=\frac{n-1}{n}E\left[S^2\right]$$

$$=\frac{\sigma^2}{n}E\left[\frac{n-1}{\sigma^2}S^2\right]\quad\left(saat\,\frac{n-1}{\sigma^2}S^2\sim\chi^2\left(n-1\right)\right)$$

$$=\frac{\sigma^2}{n}E\left[\chi^2\left(n-1\right)\right]$$

$$=\frac{\sigma^2}{n}\left(n-1\right)$$

$$=\frac{n-1}{n}\sigma^2$$

$$\neq\sigma^2$$

Oleh karena itu, penaksir maksimum likelihood dari σ^2 adalah penaksir bias Selanjutnya, pada contoh berikut, kita tunjukkan bahwa sampel varians S^2 diberikan dengan pernyataan

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$

adalah penaksir yang tidak bias dari varians populasi σ^2 terlepas dari distribusi populasi.

Contoh 6.3. Misalkan $X_1, X_2, ..., X_n$ adalah sampel acak dari sebuah populasi normal dengan mean μ dan varians $\sigma^2 > 0$. Apakah sampel varians S^2 merupakan penaksir tidak bias dari populasi varians σ^2 ?

Jawab :

Catatan bahwa distribusi dari populasi tidak diberikan. Namun, kita diberikan $E(X_i) = \mu$ dan $E\left[(X_i - \mu)^2\right] = \sigma^2$. Pada urutan untuk menemukan $E(S^2)$, kita memerlukan $E(\bar{X})$ dan $E(\bar{X}^2)$. Jadi kami melanjutkan untuk menemukan dua nilai ekspektasi. Perhitungkan

$$E(\bar{X}) = E\left(\frac{X_1, X_2, ..., X_n}{n}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

Demikian pula,

$$Var(\bar{X}) = Var\left(\frac{X_1, X_2, \ldots, X_n}{n}\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}$$

Oleh karena itu

$$E(\bar{X}^2) = Var(\bar{X}) + E(\bar{X})^2 = \frac{\sigma^2}{n} + \mu^2$$

Perhitungkan

$$E(S^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]$$

$$= \frac{1}{n-1}E\left[\sum_{i=1}^{n}\left(X_i^2 - 2\bar{X}X_i + \bar{X}^2\right)\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right]$$

$$= \frac{1}{n-1} \left\{\sum_{i=1}^{n} E\left[X_i^2\right] - E\left[n\overline{X}^2\right]\right\}$$

$$= \frac{1}{n-1} \left[n\left(\sigma^2 + \mu^2\right) - n\left(\mu^2 + \frac{\sigma^2}{n}\right)\right]$$

$$= \frac{1}{n-1} \left[(n-1)\sigma^2\right]$$

$$= \sigma^2$$

Oleh karena itu, sampel varians S^2 adalah penaksir tidak bias dari populasi varians σ^2.

Contoh 6.4. Misalkan X adalah variabel acak dengan mean 2. Misalkan $\hat{\theta}_1$ dan $\hat{\theta}_2$ adalah penaksir tidak bias masing-masing momen kedua dan momen ketiga dari X tentang asalnya. Temukan penaksir tidak bias momen ketiga dari X mengenai mean dalam pengertian $\hat{\theta}_1$ dan $\hat{\theta}_2$.

Jawab :

Karena $\hat{\theta}_1$ dan $\hat{\theta}_2$ adalah penaksir tidak bias momen kedua dan momen ketiga dari X tentang asal, kita dapatkan

$$E\left(\hat{\theta}_1\right) = E\left(X^2\right) \quad dan \quad E\left(\hat{\theta}_2\right) = E\left(X^3\right)$$

Penaksir tidak bias momen ketiga dari X mengenai mean adalah

$$E\left[(X-2)^3\right] = E\left[X^3 - 6X^2 + 12X - 8\right]$$

$$= E\left[X^3\right] - 6E\left[X^2\right] + 12E[X] - 8$$

$$= \hat{\theta}_2 - 6\hat{\theta}_1 + 24 - 8$$

$$= \hat{\theta}_2 - 6\hat{\theta}_1 + 16$$

Jadi, penaksir tidak bias momen ketiga dari X mengenai mean adalah $\hat{\theta}_2 - 6\hat{\theta}_1 + 16$.

Contoh 6.5. Misalkan $X_1, X_2, ..., X_5$ merupakan sampel berukuran 5 dari distribusi seragam pada interval $(0, \theta)$, dimana θ tidak diketahui. Misalkan penaksir dari θ merupakan $k\,X_{max}$, dimana k adalah konstanta dan X_{max} adalah pengamatan terbesar. Dalam urutan $k\,X_{max}$ menjadi penaksir tidak bias, berapakah nilai dari konstanta k

Jawab :

Fungsi kepadatan probabilitas dari X_{max} diberikan oleh

$$g(x) = \frac{5!}{4!0!}\left[F(x)\right]^4 f(x)$$

$$= 5\left(\frac{x}{\theta}\right)^4 \frac{1}{\theta}$$

$$= \frac{5}{\theta^5} x^4$$

Jika $k\,X_{max}$ adalah penaksir tidak bias dari θ, maka

$$\theta = E\left(k\,X_{max}\right)$$

$$= k\,E\left(X_{max}\right)$$

$$= k\int_0^\theta x\,g(x)\,dx$$

$$= k\int_0^\theta \frac{5}{\theta^5} x^5\,dx$$

$$= \frac{5}{6} k\,\theta$$

Maka dari itu,

$$k = \frac{6}{5}$$

Contoh 6.6. Misalkan $X_1, X_2, \ldots, X_n$ merupakan sampel dengan ukuran n dari distribusi dengan mean tidak diketahui $-\infty < \mu < \infty$, dan varians tidak diketahui $\sigma^2 > 0$. Tunjukkan bahwa statistik $\bar{X}$ dan $Y = \dfrac{X_1 + 2X_2 + \ldots + nX_n}{\frac{n(n-1)}{2}}$

keduanya adalah penaksir tidak bias dari μ. Lebih lanjut, tunjukkan bahwa $Var(\bar{X}) < Var(Y)$.

Jawab :

Pertama, kita tunjukan bahwa $\bar{X}$ adalah penaksir tidak bias dari μ

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

Karenanya, sampel mean $\bar{X}$ adalah penaksir tidak bias dari populasi mean terlepas dari distribusi X. Selanjutnya, kita tunjukkan bahwa Y adalah juga penaksir tidak bias dari μ.

$$E(Y) = E\left(\frac{X_1 + 2X_2 + \ldots + nX_n}{\frac{n(n-1)}{2}}\right)$$

$$= \frac{2}{n(n+1)}\sum_{i=1}^{n} i\,E(X_i)$$

$$= \frac{2}{n(n+1)} \sum_{i=1}^{n} i\,\mu$$

$$= \frac{2}{n(n+1)} \mu \frac{n(n+1)}{2}$$

$$= \mu$$

Karenanya, $\overline{X}$ dan Y keduanya adalah penaksir tidak bias dari populasi mean terlepas dari distribusi populasi. Varians dari $\overline{X}$ diberikan oleh

$$Var\left[\overline{X}\right] = Var\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right]$$

$$= \frac{1}{n^2} Var\left[X_1 + X_2 + \ldots + X_n\right]$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} Var\left[X_i\right]$$

$$= \frac{\sigma^2}{n}$$

Demikian pula, varians dari Y dapat dihitung dengan mengikuti :

$$Var(Y) = Var\left[\frac{X_1 + 2X_2 + \ldots + nX_n}{\frac{n(n-1)}{2}}\right]$$

$$= \frac{4}{n^2 (n+1)^2} Var\left[1X_1 + 2X_2 + \ldots + n\,X_n\right]$$

$$= \frac{4}{n^2 (n+1)^2} \sum_{i=1}^{n} Var\left[i\,X_i\right]$$

$$= \frac{4}{n^2 (n+1)^2} \sum_{i=1}^{n} i^2 \, Var[X_i]$$

$$= \frac{4}{n^2 (n+1)^2} \sigma^2 \sum_{i=1}^{n} i^2$$

$$= \sigma^2 \frac{4}{n^2 (n+1)^2} \frac{n(n+1)(2n+1)}{6}$$

$$= \frac{2}{3} \frac{2n+1}{(n+1)} \frac{\sigma^2}{n}$$

$$= \frac{2}{3} \frac{2n+1}{(n+1)} Var\left[\overline{X}\right]$$

Saat $\frac{2}{3} \frac{2n+1}{(n+1)} Var\left[\overline{X}\right] > 1$ untuk $n \geq 2$, kita dapat lihat bahwa

$Var\left[\overline{X}\right] \leq Var[Y]$. Ini menunjukan bahwa meskipun penaksir $\overline{X}$ dan Y keduanya adalah penaksir tidak bias dari μ, namun varians dari sampel mean $\overline{X}$ adalah lebih kecil dari varians Y.

6.2. Estimator Efisien Relatif

Kita dapat lihat pada Contoh 6.6 bahwa sampel mean

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

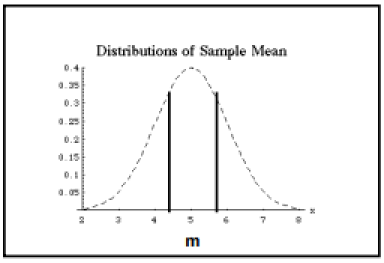
Dan statistik

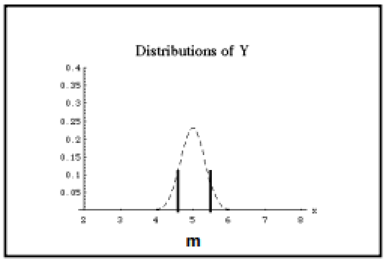
$$Y = \frac{X_1 + 2X_2 + \ldots + nX_n}{1 + 2 + \ldots + n}$$

Keduanya adalah penaksir tidak bias dari populasi mean. Akan tetapi, kita selalu lihat bahwa

$$Var\left(\bar{X}\right) < Var\left(Y\right)$$

Gambar grafik berikut mengilustrasikan bentuk distribusi dari kedua penaksir tidak bias





Jika penaksir yang tidak bias memiliki varians atau dispersi yang lebih kecil, maka ia memiliki peluang lebih besar untuk mendekati parameter θ sebenarnya. Karena itu saat dua penaksir dari θ keduanya tidak bias, maka kita harus memilih satu dengan varians yang lebih kecil.

Definisi 6.2.

Misalkan $\hat{\theta}_1$ dan $\hat{\theta}_2$ menjadi dua penaksir tidak bias dari θ. Penaksir $\hat{\theta}_1$ adlah dikatakan menjadi lebih efisien daripada $\hat{\theta}_2$ jika

$$Var\left(\hat{\theta}_1\right) < Var\left(\hat{\theta}_2\right)$$

Rasio η diberikan oleh

$$\eta\left(\hat{\theta}_1, \hat{\theta}_2\right) = \frac{Var\left(\hat{\theta}_2\right)}{Var\left(\hat{\theta}_1\right)}$$

Adalah disebut efisien relatif dari $\hat{\theta}_1$ dengan memperhatikan $\hat{\theta}_2$

Contoh 6.7.

Misalkan X_1, X_2, X_3 menjadi sampel random dengan ukuran 3 dari populasi dengan mean μ dan varians $\sigma^2 > 0$. Jika statistik $\bar{X}$ dan Y diberikan oleh

$$Y = \frac{X_1 + 2X_2 + 3X_3}{6}$$

Adalah dua penaksir tidak bias dari populasi mean μ, kemudian mana yang lebih efisien?

Jawab :

Saat $E\left(X_i\right)=\mu$ dan $Var\left(X_i\right)=\sigma^2$, kita dapatkan

$$E\left(\overline{X}\right)=E\left(\frac{X_1+X_2+X_3}{3}\right)$$

$$=\frac{1}{3}\left(E\left(X_1\right)+E\left(X_2\right)+E\left(X_3\right)\right)$$

$$=\frac{1}{3}3\mu$$

$$=\mu$$

Dan

$$E\left(Y\right)=E\left(\frac{X_1+X_2+X_3}{6}\right)$$

$$=\frac{1}{6}\left(E\left(X_1\right)+E\left(X_2\right)+E\left(X_3\right)\right)$$

$$=\frac{1}{6}6\mu$$

$$=\mu$$

Oleh karena itu $\overline{X}$ dan Y keduanya adalah tidak bias. Selanjutnya kita tentukan varians dari kedua penaksir. Varians dari penaksir tersebut diberikan oleh

$$Var\left(\overline{X}\right)=Var\left(\frac{X_1+X_2+X_3}{3}\right)$$

$$=\frac{1}{9}\left[Var\left(X_1\right)+Var\left(X_2\right)+Var\left(X_3\right)\right]$$

$$= \frac{1}{9} 3\sigma^2$$

$$= \frac{12}{36} \sigma^2$$

Dan

$$Var(Y) = Var\left(\frac{X_1 + 2X_2 + 3X_3}{6} \right)$$

$$= \frac{1}{36} \left[Var(X_1) + 4Var(X_2) + 9Var(X_3) \right]$$

$$= \frac{1}{36} 14\sigma^2$$

$$= \frac{14}{36} \sigma^2$$

Oleh karena itu

$$\frac{12}{36} \sigma^2 = Var(\overline{X}) < Var(Y) = \frac{14}{36} \sigma^2$$

Karenanya, $\overline{X}$ adalah lebih efisien daripada penaksir Y. Lebih lanjut, efisiensi relatif dari $\overline{X}$ dengan memperhatikan Y adalah diberikan oleh

$$\eta(\overline{X}, Y) = \frac{14}{12} = \frac{7}{6}$$

Contoh 6.8. Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak berukuran n dari populasi dengan kepadatan

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & untuk\, 0 \le x < \infty \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Dimana $\theta > 0$ adalah parameter. Apakah penaksir X_1 dan $\overline{X}$ tidak bias?. Berikan X_1 dan $\overline{X}$, manakah penaksir yang lebih efisien dari θ?

Jawab :

Saat populasi X adalah eksponensial dengan parameter θ, yaitu $X_1 \sim EXP(\theta)$, mean dan varians darinya adalah diberikan oleh

$$E(X) = \theta \qquad dan \qquad Var(X) = \theta^2$$

Saat $X_1, X_2, \ldots, X_n$ adalah sampel acak dari X, kita lihat bahwa statistik $X_1 \sim EXP(\theta)$. Karena itu, nilai ekspektasi dari X_1 adalah θ dan itu adalah penaksir tidak bias dari parameter θ. Demikian pula, sampel mean adalah penaksir tidak bias dari θ sejak

$$E(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_1)$$

$$= \frac{1}{n} n\theta$$

$$= \theta$$

Selanjutnya, kita perhitungkan varians dari penaksir tidak bias X_1 dan $\overline{X}$

$$Var(X_1) = \theta^2$$

Dan

$$Var(\overline{X}) = Var\left(\frac{X_1 + X_2 + \ldots, X_n}{n}\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$$

$$= \frac{1}{n^2} n\theta^2$$

$$=\frac{\theta^2}{n}$$

Oleh karena itu

$$\frac{\theta^2}{n}=Var\left(\overline{X}\right)<Var\left(X_1\right)=\theta^2$$

Jadi $\overline{X}$ adalah lebih efisien daripada X_1 dan efisiensi relatif dari $\overline{X}$ dengan memperhatikan X_1 adalah

$$\eta\left(\overline{X},X_1\right)=\frac{\theta^2}{\frac{\theta}{n}}=n$$

Contoh 6.9. Misalkan $X_1+X_2+X_3$ menjadi sampel acak berukuran 3 dari populasi dengan kepadatan

$$f\left(x;\lambda\right)=\begin{cases}\dfrac{\lambda^x e^{-\lambda}}{x!} & jika\, x=0,1,2,\ldots,\infty\\ 0 & untuk\ yang\ lainnya\end{cases}$$

Dimana λ adalah parameter. Apakah penaksir yang diberikan

$$\widehat{\lambda_1}=\frac{1}{4}\left(X_1+2X_2+X_3\right)\qquad dan\qquad \widehat{\lambda_2}=\frac{1}{9}\left(4X_1+3X_2+2X_3\right)$$

Tidak bias?. Jika diberikan $\widehat{\lambda_1}$ dan $\widehat{\lambda_2}$, manakah penaksir yang lebih efisien dari λ?. Temukan penaksir tidak bias dari λ yang mana variansnya lebih kecil dari $\widehat{\lambda_1}$ dan $\widehat{\lambda_2}$

Jawab :

Karena setiap $X_i \sim POI\left(\lambda\right)$, kita dapatkan

$$E\left(X_i\right)=\lambda\qquad dan\qquad Var\left(X_i\right)=\lambda$$

Dapat diperhitungkan sebagai berikut

$$E\left(\widehat{\lambda_1}\right) = \frac{1}{4}\left(E\left(X_1\right) + 2E\left(X_2\right) + E\left(X_3\right)\right)$$

$$= \frac{1}{4}4\lambda$$

$$= \lambda$$

Dan

$$E\left(\widehat{\lambda_2}\right) = \frac{1}{9}\left(4E\left(X_1\right) + 3E\left(X_2\right) + 2E\left(X_3\right)\right)$$

$$= \frac{1}{9}9\lambda$$

$$= \lambda$$

Dengan demikian, $\widehat{\lambda_1}$ dan $\widehat{\lambda_2}$ keduanya adalah penaksir tidak bias dari λ. Sekarang kita perhitungkan variansnya untuk mengetahui manakah penaksir yang lebih efisien

$$Var\left(\widehat{\lambda_1}\right) = \frac{1}{16}\left(Var\left(X_1\right) + 4Var\left(X_2\right) + Var\left(X_3\right)\right)$$

$$= \frac{1}{16}6\lambda$$

$$= \frac{6}{16}\lambda$$

Dan

$$Var\left(\widehat{\lambda_2}\right) = \frac{1}{81}\left(16Var\left(X_1\right) + 9Var\left(X_2\right) + 4Var\left(X_3\right)\right)$$

$$= \frac{1}{81}29\lambda$$

$$= \frac{29}{81}\lambda$$

Dengan menyamakan penyebutnya, kita bandingkan manakah yang lebih kecil antara $Var\left(\widehat{\lambda_1}\right)$ dan $Var\left(\widehat{\lambda_2}\right)$ yakni dengan mengalikan $Var\left(\widehat{\lambda_1}\right)$ dengan 81 dan mengalikan $Var\left(\widehat{\lambda_2}\right)$ dengan 16, maka diperoleh $Var\left(\widehat{\lambda_1}\right) = \frac{486}{1296}\lambda$ dan $Var\left(\widehat{\lambda_2}\right) = \frac{464}{1296}\lambda$. Dengan demikian $Var\left(\widehat{\lambda_2}\right) < Var\left(\widehat{\lambda_1}\right)$, penaksir $\widehat{\lambda_2}$ adalah efisien dibandingkan dengan penaksir $\widehat{\lambda_1}$. Kita dapat melihat pada bagian 6.1 bahwa sampel mean adalah selalu menjadi penaksir tidak bias dari populasi mean terlepas dari populasi distribusi. Varians dari sampel mean adalah selalu sama dengan $\frac{1}{n}$ kali populasi varians, dimana n menyatakan ukuran sampel. Oleh karena itu, kita dapatkan

$$Var\left(\overline{X}\right) = \frac{\lambda}{3}$$

Kita samakan terlebih dahulu penyebut dari $Var\left(\overline{X}\right)$ agar sesuai dengan penyebut dari $Var\left(\widehat{\lambda_1}\right)$ dan $Var\left(\widehat{\lambda_2}\right)$ yaitu dengan mengalikan penyebut $Var\left(\overline{X}\right)$ dengan 432 maka diperoleh

$$Var\left(\overline{X}\right) = \frac{\lambda}{3} = \frac{432}{1296}\lambda$$

Oleh karena itu, kita dapatkan

$$Var\left(\overline{X}\right) < Var\left(\widehat{\lambda_2}\right) < Var\left(\widehat{\lambda_1}\right)$$

Jadi, sampel mean memiliki varians yang lebih kecil dari pada dua penaksir tidak bias yang diberikan pada contoh ini.

Mengingat contoh ini, sekarang kita menghadapi masalah baru, yakni bagaimana menemukan penaksir yang tidak bias yang memiliki varians terkecil diantara semua penaksir yang tidak bias dari parameter tertentu.

6.3. Estimator Tidak Bias Varians Minimum Seragam (UMVUE)

Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak berukuran n dari sebuah populasi dengan fungsi kepadatan probabilitas $f(x;\theta)$. Ingat bahwa penaksir $\hat{\theta}$ dari θ adalah fungsi dari variabel acak $X_1, X_2, \ldots, X_n$ yang bergantung pada θ.

Definisi 6.3.

Penaksir tidak bias $\hat{\theta}$ dari θ dikatakan menjadi penaksir tidak bias varians minimum seragam dari θ jika dan hanya jika

$$Var\left(\hat{\theta}\right) \le Var\left(\hat{T}\right)$$

Untuk setiap penaksir tidak bias $\hat{T}$ dari $\hat{\theta}$

Jika penaksir $\hat{\theta}$ adalah tidak bias maka mean dari penaksir ini sama dengan parameter θ, yakni

$$E\left(\hat{\theta}\right) = \theta$$

Dan varians dari $\hat{\theta}$ adalah

$$Var\left(\hat{\theta}\right) = E\left[\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)^2\right]$$

$$= E\left[\left(\hat{\theta} - \theta\right)^2\right]$$

Varians ini, jika ada, adalah fungsi dari penaksir tidak bias $\hat{\theta}$ dan itu minimum di kelas semua penaksir yang tidak bias dari θ. Oleh karena itu kita memiliki definisi alternatif dari penaksir tidak bias varians minimum yang seragam.

Definisi 6.4.

Penaksir tidak bias $\hat{\theta}$ dari θ dikatakan menjadi penaksir tidak bias varians minimum seragam dari θ jika meminimalkan varians $E\left[\left(\hat{\theta} - \theta\right)^2\right]$

Contoh 6.10. Misalkan $\widehat{\theta_1}$ dan $\widehat{\theta_2}$ menjadi penaksir tidak bias dari θ. Misalkan $Var\left(\widehat{\theta_1}\right) = 1$, $Var\left(\widehat{\theta_2}\right) = 2$ dan $Cov\left(\widehat{\theta_1}, \widehat{\theta_2}\right) = \frac{1}{2}$. Berapa nilai dari c_1 dan c_2 yang mana $c_1\widehat{\theta_1} + c_2\widehat{\theta_2}$ adalah penaksir tidak bias dari θ dengan varians minimum antara penaksir tidak bias dari tipe ini?

Jawab :

Kita ingin $c_1\widehat{\theta_1} + c_2\widehat{\theta_2}$ menjadi varians minimum penaksir tidak bias dari θ. Kemudian

$$E\left[c_1\widehat{\theta_1}+c_2\widehat{\theta_2}\right]=\theta$$

$$\Rightarrow\ c_1\,E\left[\widehat{\theta_1}\right]+c_2\,E\left[\widehat{\theta_2}\right]=\theta$$

$$\Rightarrow\ c_1\theta+c_2\theta=\theta$$

$$\Rightarrow\ c_1+c_2=1$$

$$\Rightarrow\ c_2=1-c_1$$

Oleh karena itu

$$Var\left[c_1\widehat{\theta_1}+c_2\widehat{\theta_2}\right]=c_1^2Var\left[\widehat{\theta_1}\right]+c_2^2Var\left[\widehat{\theta_2}\right]+2c_1c_2Cov\left(\widehat{\theta_1},\widehat{\theta_1}\right)$$

$$=c_1^2+2c_2^2+c_1c_2$$

$$=c_1^2+2\left(1-c_1\right)^2+c_1\left(1-c_1\right)$$

$$=2\left(1-c_1\right)^2+c_1$$

$$=2+2c_1^2-3c_1$$

Karenanya, varians $Var\left[c_1\widehat{\theta_1}+c_2\widehat{\theta_2}\right]$ adalah fungsi dari c_1. Selanjutnya kita tunjukkan bahwa fungsi dengan $\phi\left(c_1\right)$, yaitu

$$\phi\left(c_1\right):=Var\left[c_1\widehat{\theta_1}+c_2\widehat{\theta_2}\right]=2+2c_1^2-3c_1$$

Ambil turunan dari $\phi\left(c_1\right)$ terhadap c_1, kita dapatkan

$$\frac{d}{dc_1}\phi\left(c_1\right)=4c_1-3$$

Untuk menemukan c_1, turunan tadi disamadengankan nol

$$4c_1 - 3 = 0 \quad \Rightarrow \quad c_1 = \frac{3}{4}$$

Oleh karena itu

$$c_2 = 1 - c_1 = 1 - \frac{3}{4} = \frac{1}{4}$$

Pada contoh 6.10, kita melihat bahwa $\widehat{\theta_1}$ dan $\widehat{\theta_2}$ masing-masing adalah dua estimator tidak bias dari θ, sehingga $c\widehat{\theta_1} + (1-c)\widehat{\theta_2}$ juga merupakan estimator tidak bias dari θ untuk setiap $c \in \mathbb{R}$. Karenanya diberikan dua estimator $\widehat{\theta_1}$ dan $\widehat{\theta_2}$

$$C = \left\{ \hat{\theta} \mid \hat{\theta} = c\widehat{\theta_1} + (1-c)\widehat{\theta_2},\ c \in \mathbb{R} \right\}$$

membentuk kelas estimator tak bias yang tak terhitung dari θ. Ketika varians dari $\widehat{\theta_1}$ dan $\widehat{\theta_2}$ diketahui bersama dengan kovariansnya, maka dalam Contoh 6.10 kami dapat menentukan penduga tak bias varians minimum di kelas C. Jika varians dari penduga $\widehat{\theta_1}$ dan $\widehat{\theta_2}$ tidak diketahui, maka sangat sulit untuk menemukan penduga varians minimum bahkan dalam kelas penduga C. Perhatikan bahwa C adalah himpunan bagian dari kelas semua tak bias estimator dan menemukan estimator tak bias varians minimum di kelas ini adalah tugas yang sulit.

Salah satu cara untuk menemukan penduga tak bias varians minimum yang seragam untuk parameternya adalah menggunakan batas bawah Cramer-Rao atau ketaksamaan fisher.

Teorema 6.1. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak berukuran n dari populasi X dengan kepadatan probabilitas $f(x;\theta)$, dimana θ adalah parameter skalar. Misal $\widehat{\theta}$ merupakan estimator tak bias dari θ. Misalkan fungsi likelihood $L(\theta)$ adalah fungsi yang berbeda dari fungsi θ dan memenuhi

$$\frac{d}{d\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h(x_1,\ldots,x_n)\, L(\theta)\, dx_1,\ldots,dx_n$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1,\ldots,x_n) \frac{d}{d\theta} L(\theta)\, dx_1,\ldots,dx_n \qquad (1)$$

Untuk setiap $h(x_1,\ldots,x_n)$ dengan $E(h(x_1,\ldots,x_n)) < \infty$. Maka

$$Var(\hat{\theta}) \geq \frac{1}{E\left[\dfrac{(\partial \ln L(\theta))^2}{\partial \theta}\right]} \qquad (CR1)$$

Bukti :

Ketika $L(\theta)$ sebuah fungsi kepadatan probabilitas bersama dari sampel $X_1, X_2, \ldots, X_n$,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} L(\theta)\, dx_1,\ldots,dx_n = 1 \qquad (2)$$

Turunkan pernyataan (2) terhadap θ kita punya

$$\frac{d}{d\theta} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} L(\theta)\, dx_1,\ldots,dx_n = 0$$

Dan gunakan pernyataan (1) dengan $h(x_1,\ldots,x_n) = 1$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{d}{d\theta} L(\theta)\, dx_1,\ldots,dx_n = 0 \qquad (3)$$

Tulis pernyataan (3) seperti

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{dL(\theta)}{d\theta} \frac{1}{L(\theta)} L(\theta)\, dx_1,\ldots,dx_n = 0$$

Sehingga

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{d \ln L(\theta)}{d\theta} L(\theta)\, dx_1,\ldots,dx_n = 0$$

Karenanya

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta \frac{d \ln L(\theta)}{d\theta} L(\theta)\, dx_1,\ldots,dx_n = 0 \qquad (4)$$

Saat $\hat{\theta}$ adalah estimator tak bias dari θ maka

$$E\left(\hat{\theta}\right)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\hat{\theta}L(\theta)\,dx_1,\ldots,dx_n=0 \qquad (5)$$

Diferensiasi pernyataan (5) terhadap θ sehingga kita punya

$$\frac{d}{d\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\hat{\theta}L(\theta)\,dx_1,\ldots,dx_n=1$$

Gunakan pernyataan (1) lagi dengan $h(X_1,\ldots,X_n)=1$, sehingga kita punya

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\hat{\theta}\frac{d}{d\theta}L(\theta)\,dx_1,\ldots,dx_n=1 \qquad (6)$$

Tulis ulang pernyataan (6) seperti

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\hat{\theta}\frac{dL(\theta)}{d\theta}L(\theta)\,dx_1,\ldots,dx_n=1$$

Kita punya

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\hat{\theta}\frac{d\ln L(\theta)}{d\theta}L(\theta)\,dx_1,\ldots,dx_n=1 \qquad (7)$$

Dari pernyataan (4)dan(7) kita memperoleh

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\hat{\theta}-\theta\right)\frac{d\ln L(\theta)}{d\theta}L(\theta)\,dx_1,\ldots,dx_n=1 \qquad (8)$$

dengan ketaksamaan Cauchy-Schwarz kita punya

$$1=\left(\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\hat{\theta}-\theta\right)\frac{d\ln L(\theta)}{d\theta}L(\theta)\,dx_1,\ldots,dx_n\right)^2$$

$$\leq\left(\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\hat{\theta}-\theta\right)^2L(\theta)\,dx_1,\ldots,dx_n\right)$$

$$\bullet\left(\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\frac{d\ln L(\theta)}{d\theta}\right)^2L(\theta)\,dx_1,\ldots,dx_n\right)$$

$$= Var\left(\hat{\theta}\right) E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)^2\right]$$

Oleh karena itu

$$Var\left(\hat{\theta}\right) \geq \frac{1}{E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)^2\right]}$$

Jika $L(\theta)$ di diferensiasi 2 kali terhadap θ, maka ketaksamaan $(CR1)$ bisa dinyatakan secara ekuivalen sebagai

$$Var\left(\hat{\theta}\right) \geq \frac{1}{E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)^2\right]} \qquad (CR2)$$

Ketaksamaan $(CR1)$dan$(CR2)$ dikenal sebagai batas bawah Cramer-Rao dari varians $\hat{\theta}$ atau ketaksamaan fisher. Pernyataan (1) menukar urutan integrasi dan diferensiasi. Karena itu, beberapa distribusi yang range nya bergantung pada nilai parameter tidak dicakup pada teorema ini. Karenanya, distribusi seperti distribusi Uniform tidak bisa dianalisis menggunakan batas bawah Cramer-Rao.

Jika estimator $\hat{\theta}$ merupakan varians minimum selain tak bias, maka persamaan berhenti. Kita nyatakan bahwa ini adalah teorema diluar bukti.

Teorema 6.2.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampe acak berukuran n dari populasi X dengan kepadatan probabilitas $f(x;\theta)$ dimana θ adalah parameter. Jika $\hat{\theta}$ adalah estimator tak bias dari θ dan

$$Var\left(\hat{\theta}\right) = \frac{1}{E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)^2\right]}$$

Maka $\hat{\theta}$ adalah UMVUE dari θ. Sebaliknya dari ini salah.

Definisi 6.5.

Estimator tak bias $\hat{\theta}$ dinamakan estimator efisien jika memenuhi batas bawah Cramer-Rao, sehingga

$$Var\left(\hat{\theta}\right) = \frac{1}{E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)^2\right]}$$

Kesimpulannya, teorema ini mudah untuk dicatat bahwa estimator efisien dari parameter selalu sebuah UMVUE dari sebuah parameter. Namun, setiap UMVUE dari parameter merupakan efisien. Dengan kata lain, setiap estimator tak bias varians minimum seragam dari parameter memenuhi batas bawah Cramer-Rao

$$Var\left(\hat{\theta}\right) \geq \frac{1}{E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)^2\right]}$$

Contoh 6.11. Misalkan $X_1, X_2, \ldots, X_n$ merupakan sampel acak berukuran n dari distribusi dengan fungsi densitas

$$f(x;\theta) = \begin{cases} 3\theta x^2 e^{-\theta x^3} & jika\, 0 < x < \infty \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Berapakah CRLB untuk varians dari estimator tak bias parameter θ?

Jawab :

Misalkan $\hat{\theta}$ merupakan estimator tak bias dari θ, maka batas bawah Cramer-Rao diberikan oleh

$$Var\left(\hat{\theta}\right) \geq \frac{-1}{E\left[\frac{\partial^2 \ln L(\theta)}{\partial \theta^2}\right]}$$

Dimana $L(\theta)$ dinotasikan sebagai fungsi likelihood dari sampel acak $X_1, X_2, \ldots, X_n$. Karena fungsi likelihood dari sampel adalah

$$L(\theta) = \prod_{i=1}^{n} 3\theta x^2 e^{-x_i^3}$$

Kita peroleh

$$\ln L(\theta) = n \ln \theta + \sum_{i=1}^{n} \ln\left(3x_i^2\right) - \theta \sum_{i=1}^{n} x_i^3$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} x_i^3$$

Dan

$$\frac{\partial^2 \ln L(\theta)}{\partial \theta^2} = -\frac{n}{\theta^2}$$

Karenanya, dengan menggunakan ketaksamaan Cramer-Rao ini maka kita punya

$$Var\left(\hat{\theta}\right) \geq \frac{\theta^2}{n}$$

Jadi, batas bawah Cramer-Rao (CRLB) untuk varians dari estimator tak bias θ adalah $\dfrac{\theta^2}{n}$

Contoh 6.12. Misalkan $X_1, X_2, \ldots, X_n$ merupakan sampel acak dari populasi normal dengan rata-rata μ tidak diketahui dan varians $\sigma^2 > 0$. Berapa estimator maksimum likelihood dari μ? Apakah estimator maksimum likelihood dari μ efisien ?

Jawab :

fungsi kepadatan probabilitas dari populasi adalah

$$f(x;\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Jadi,

$$\ln f(x;\mu) = -\frac{1}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}(x-\mu)^2$$

Dan karenanya

$$\ln L(\mu) = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2$$

Dengan menggunakan turunan dari $L(\mu)$ terhadap μ, kita punya

$$\frac{d\ln L(\mu)}{d\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)$$

Hitung turunan tersebut sampai nol dan selesaikan μ, kita lihat bahwa $\hat{\mu} = \bar{X}$. Varians dari $\bar{X}$ diberikan oleh

$$Var\left(\bar{X}\right) = Var\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right)$$

$$= \frac{\sigma^2}{n}$$

Selanjutnya tentukan batas bawah Cramer-Rao dari estimator $\bar{X}$. Kita sudah tahu bahwa

$$\frac{d\ln L(\mu)}{d\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)$$

Dan oleh karena itu

$$\frac{d^2\ln L(\mu)}{d\mu^2} = -\frac{n}{\sigma^2}$$

Sehingga

$$E\left(\frac{d^2\ln L(\mu)}{d\mu^2}\right) = -\frac{n}{\sigma^2}$$

Dan

$$-\frac{1}{E\left(\frac{d^2\ln L(\mu)}{d\mu^2}\right)} = \frac{\sigma^2}{n}$$

Jadi

$$Var\left(\bar{X}\right) = -\frac{1}{E\left(\frac{d^2 \ln L(\mu)}{d\mu^2}\right)}$$

Dan $\left(\bar{X}\right)$ adalah estimator efisien dari μ. Karena setiap estimator efisien adalah sebuah UMVUE, oleh karena itu $\left(\bar{X}\right)$ adalah UMVUE dari μ.

Contoh 6.13. Misalkan $X_1, X_2, \ldots, X_n$ merupakan sampel acak dari populasi normal dengan rata-rata μ diketahui dan varians $\sigma^2 > 0$ tidak diketahui. Berapakah estimator maksimum likelihood dari σ^2? Apakah estimator maksimum likelihood sebuah UMVUE dari σ^2?

Jawab :

Mari kita tulis $\theta = \sigma^2$. Maka

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{\theta}(x-\mu)^2}$$

Dan

$$\ln L(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\theta) - \frac{1}{2\theta}\sum_{(i=1)}^{n}(x_i - \mu)^2$$

Diferensiasi $\ln L(\theta)$ terhadap θ kita punya

$$\frac{d}{d\theta}\ln L(\theta) = -\frac{n}{2}\frac{1}{\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Hitung turunan sampai nol dan selesaikan θ, sehingga

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Selanjutnya kita tunjukan bahwa estimator tersebut tak bias. Sehingga kita berpendapat bahwa

$$E\left(\hat{\theta}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right)$$

$$= \frac{\sigma^2}{n} E\left(\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \right)$$

$$= \frac{\theta}{n} E\left(\mathcal{X}^2(n) \right)$$

$$= \frac{\theta}{n} n$$

$$= \theta$$

Oleh karena itu $\hat{\theta}$ merupakan estimator tak bias dari θ. Varians dari $\hat{\theta}$ bisa kita peroleh sebagai berikut :

$$Var\left(\hat{\theta}\right) = Var\left(\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 \right)$$

$$= \frac{\sigma^4}{n} Var\left(\sum_{i=1}^{n} (\frac{X_i - \mu}{\sigma^2} \right)$$

$$= \frac{\theta^2}{n^2} Var\left(\mathcal{X}^2(n) \right)$$

$$= \frac{\sigma^2}{n^2} 4 \frac{n}{2}$$

$$= \frac{2\theta^2}{n}$$

$$= \frac{2\sigma^4}{n}$$

Terakhir kita tentukan batas bawah Cramer-Rao untuk varians dari $\hat{\theta}$. Turunan kedua dari $\ln L(\theta)$ terhadap θ adalah

$$\frac{d^2 \ln L(\theta)}{d\theta^2} = \frac{n}{2\theta^2} - \frac{1}{\theta^3} \sum_{i=1}^{n} (x_i - \mu)^2$$

Karenanya

$$E\left(\frac{d^2 \ln L(\theta)}{d\theta^2}\right) = \frac{n}{2\theta^2} - \frac{1}{\theta^3} \left(\sum_{i=1}^{n} (X_i - \mu)^2\right)$$

$$= \frac{n}{2\theta^2} - \frac{\theta}{\theta^3} E\left(\mathcal{X}^2(n)\right)$$

$$= \frac{n}{2\theta^2} - \frac{n}{\theta^2}$$

$$= -\frac{n}{2\theta^2}$$

Jadi

$$-\frac{1}{E\left(\frac{d^2 \ln L(\theta)}{d\theta^2}\right)} = \frac{\theta^2}{n}$$

$$= \frac{2\sigma^4}{n}$$

Oleh karena itu

$$Var\left(\hat{\theta}\right) = -\frac{1}{E\left(\frac{d^2 \ln L(\theta)}{d\theta^2}\right)}$$

Oleh karena itu, $\hat{\theta}$ merupakan estimator efisien dari θ. Karena setiap estimator efisien merupakan UMVUE, karena itu

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$$

Merupakan UMVUE dari σ^2.

Contoh 6.14. Misalkan $X_1, X_2, \dots, X_n$ merupakan sampel acak berukuran n dari populasi normal dimana rata-rata μ diketahui dan varians $\sigma^2 > 0$. Tunjukan bahwa

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Merupakan estimator tak bias dari σ^2. Selanjutnya, tunjukan bahwa S^2 tidak dapat menghasilkan batas bawah Cramer-Rao.

Jawab :

Dari contoh 6.2, kita tahu bahwa S^2 merupakan estimator tak bias dari σ^2. Varians dari S^2 bisa dihitung dengan

$$Var\left(S^2\right) = Var\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right)$$

$$= \frac{\sigma^4}{(n-1)^2} Var\left(\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2\right)$$

$$= \frac{\sigma^4}{(n-1)^2} Var\left(\mathcal{X}^2(n-1)\right)$$

$$= \frac{\sigma^4}{(n-1)^2} 2(n-1)$$

$$= \frac{2\sigma^4}{n-1}$$

Selanjutnya, misalkan $\theta = \sigma^2$ dan tentukan batas bawah Cramer-Rao dari varians S^2. Turunan kedua dari $\ln L(\theta)$ terhadap θ adalah

$$\frac{d^2 \ln L(\theta)}{d\theta^2} = \frac{n}{2\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^{n}\left(x_i - \mu\right)^2$$

Karena itu

$$E\left(\frac{d^2 \ln L(\theta)}{d\theta^2}\right) = \frac{n}{2\theta^2} - \frac{1}{\theta^3} E\left(\sum_{i=1}^{2} (X_i - \mu^2\right)$$

$$= \frac{n}{2\theta^2} - \frac{\theta}{\theta^3} E\left(\mathcal{X}^2(n)\right)$$

$$= \frac{n}{2\theta^2} - \frac{n}{\theta^2}$$

$$= -\frac{n}{2\theta^2}$$

Jadi

$$-\frac{1}{E\left(\frac{d^2 \ln L(\theta)}{d\theta^2}\right)} = \frac{\theta^2}{n}$$

$$= \frac{2\sigma^4}{n}$$

Oleh karena itu

$$\frac{2\sigma^4}{n-1} = Var\left(S^2\right) > -\frac{1}{E\left(\frac{d^2 \ln L(\theta)}{d\theta^2}\right)}$$

$$= \frac{2\sigma^4}{n}$$

Ini menunjukan bahwa S^2 tidak dapat menghasilkan batas bawah Cramer-Rao.

Kekurangan dari batas bawah Cramer-Rao meliputi :

1. Tidak setiap fungsi kepadatan $f(x;\theta)$ memenuhi asumsi dari teorema Cramer-Rao
2. Tidak setiap estimator mencakup batas bawah Cramer-Rao.

Oleh karena itu dalam salah satu dari situasi ini, kita tidak tahu apakah estimator adalah UMVUE atau tidak

6.4 Estimator Kecukupan

Dalam beberapa situasi, kita tidak bisa dengan mudah menemukan distribusi dari estimator $\hat{\theta}$ untuk parameter θ meskipun kita mengetahui distribusi dari populasi. Oleh karena itu, kita punya beberapa cara untuk mengetahui apakah estimator $\hat{\theta}$ kita bias atau tak bias. Karenanya, kita gunakan beberapa kriteria lain untuk menilai sifat dari estimator. Kecukupan adalah salah satu dari beberapa kriteria dalam menilai sifat estimator.

Mengingat kembali bahwa estimator dari populasi parameter adalah fungsi dari nilai sampel bahwa tidak ada parameter yang diperoleh. Sebuah estimator penjumlahan mencari informasi dalam parameter sampel. Jika estimator penjumlahan hanya sedikit informasi mengenai parameter yang sedang diestimasikan pada sampel, maka estimator disebut estimator kecukupan.

Definisi 6.6.

Misalkan $X \sim f(x;\theta)$ merupakan populasi dan misalkan $X_1, X_2, \ldots, X_n$ merupakan sampel acak berukuran n dari populasi X. Estimator $\hat{\theta}$ dari parameter θ dikatakan estimator kecukupan dari θ jika kondisi distribusi dari sampel diberikan estimator $\hat{\theta}$ tidak bergantung pada parameter θ.

Contoh 6.15.

Jika $X_1, X_2, \ldots X_n$ merupakan sampel acak dari distribusi dengan fungsi kepadatan

$$f(x;\theta) = \begin{cases} \theta^x (1-\theta)^{1-x} & jika\, x = 0,1 \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Dimana $0 < \theta < 1$. Tunjukan bahwa $Y = \sum_{i-1}^{n} X_i$ adalah statistik kecukupan dari θ

Jawab :

Pertama, cari distribusi dari sampel, sehingga

$$f\left(x_1,x_2,\ldots,x_n\right)=\prod_{i=1}^{n}\theta^{x_i}\left(1-\theta\right)^{1-x_i}$$

$$=\theta^{y}\left(1-\theta\right)^{n-y}$$

Karena, setiap $X_i \sim BER\left(\theta\right)$ kita punya

$$Y=\sum_{i=1}^{n}X_i \sim BIN\left(n,\theta\right)$$

Oleh karena itu, fungsi kepadatan probabilitas dari Y diberikan oleh

$$g\left(y\right)=\binom{n}{y}\theta^{y}\left(1-\theta\right)^{n-y}$$

Selanjutnya, karena setiap $X_i \sim BER\left(\theta\right)$, ruang dari X_i diberikan oleh

$$R_{X_i}=\left\{0,1\right\}$$

Oleh karena itu, ruang variabel acak $Y=\sum_{i=1}^{n}X_i$ diberikan oleh

$$R_Y=\left\{0,1,2,\ldots,n\right\}$$

Misalkan A adalah kejadian $\left(X_1=x_1,X_2=x_2,\ldots,X_n=x_n\right)$ dan B dinotasikan kejadian $\left(Y=y\right)$. Maka, $A\subset B$ dan oleh karena itu $A\cap B=A$. Sekarang, kita cari kepadatan bersyarat dari sampel yang diberikan estimator Y, Sehingga

$$f(x_1,x_2,\ldots x_n|Y=y)=P(X_1=x_1,X_2=x_2,\ldots,X_n=x_n \;\; |Y=y)$$

$$=P(A|B)$$

$$=\frac{P\left(A\cap B\right)}{P\left(B\right)}$$

$$=\frac{P\left(A\right)}{P\left(B\right)}$$

$$= \frac{f(x_1, x_2, \ldots, x_n)}{g(y)}$$

$$= \frac{\theta^y (1-\theta)^{n-y}}{\binom{n}{y} \theta^y (1-\theta)^{n-y}}$$

$$= \frac{1}{\binom{n}{y}}$$

Karena itu, kepadatan bersyarat dari sampel statistik Y merupakan parameter θ independen. Oleh karena itu, dengan menggunakan definisi, Y adalah estimator kecukupan.

Contoh 6.16. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan probabilitas fungsi kepadatan

$$f(x;\theta) = \begin{cases} e^{-(x-\theta)} & jika\, \theta < x < \infty \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Dimana $-\infty < \theta < \infty$. Berapakah estimator maksimum likelihood dari θ Apakah estimator maksimum likelihood terrmasuk estimator kecukupan dari θ ?

Jawab :

Kita lihat Bab 5 bahwa estimator maksimum likelihood dari θ adalah $Y = X_{(1)}$, itu adalah statistik terurut pertama dari sampel. Misalkan kita cari fungsi kepadatan probabilitas dari statistik ini, maka

$$g(y) = \frac{n!}{(n-1)!}\left[F(y)\right]^0 f(y)\left[1-F(y)\right]^{n-1}$$

$$= n\, f(y)\left[1-F(y)\right]^{n-1}$$

$$= n\, e^{-(y-\theta)}\left[1-\left\{1-e^{-(y-\theta)}\right\}\right]^{n-1}$$

$$= n e^{n\theta} e^{-ny}$$

Probabilitas kepadatan dai sampel acak adalah

$$f(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} e^{-(x_i - \theta)}$$

$$= e^{n\theta} e^{-n(\bar{x})}$$

Dimana $n\bar{x} = \sum_{i=1}^{n} x_i$. Misalkan A adalah kejadian $(X_1 = x_1, X_2, = x2, \ldots, X_n = x_n)$ dan B dinotasikan kejadian $Y = y$. Maka, $A \subset B$ dan oleh karena itu $A \cap B = A$. Sekarang kita cari kepadatan bersyarat dari sampel yang diberikan estimator Y, sehingga

$$f(x_1, x_2, \ldots x_n \mid Y = y) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid Y = y)$$

$$= P(A|B)$$

$$= \frac{P(A \cap B)}{P(B)}$$

$$= \frac{P(A)}{P(B)}$$

$$= \frac{f(x_1, x_2, \ldots, x_n)}{g(y)}$$

$$= \frac{e^{n\theta} e^{-(n\bar{x})}}{n e^{n\theta} e^{-(ny)}}$$

$$= \frac{e^{-n\bar{x}}}{n e^{-ny}}$$

Karena kepadatan bersyarat dari sampel yang diberikan statistik Y adalah independen dari parameter θ. Oleh karena itu, dengan menggunakan definisi, Y merupakan statistik kecukupan.

Kita lihat bahwa apakah estimator merupakan kecukupan atau bukan, kita harus memeriksa kepadatan bersyarat dari estimator sampel. Untuk menghitung kepadatan bersyarat kita harus gunakan kepadatan dari estimator. Kepadatan dari estimator tidak mudah dicari. Karena, memverifikasi kecukupan dari estimator dengan menggunakan definisi tidaklah mudah. Teorema faktorisasi dari Fisher dan Neyman membantu dalam menentukan estimator kecukupan.

Teorema 6.3.

Misalkan $X_1, X_2, \ldots, X_n$ menunjukan sampel acak dengan fungsi kepadatan probabilitas $f(x_1, x_2, \ldots, x_n; \theta)$ yang bergantung pada populasi parameter

θ. Estimator $\hat{\theta}$ merupakan kecukupan dari θ jika dan hanya jika

$$f(x_1, x_2, \ldots, x_n; \theta) = \phi(\hat{\theta}, \theta) h(x_1, x_2, \ldots x_n)$$

Dimana ϕ bergantung pada $x_1, x_2, \ldots, x_n$ hanya melalui $\hat{\theta}$ dan $h(x_1, x_2, \ldots, x_n)$ tidak bergantung pada θ.

Contoh 6.17.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan fungsi kepadatan

$$f(x;\lambda) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!} & jika\, x = 0, 1, 2, \ldots, \infty \\ 0 & untuk\ yang\ lainnya \end{cases}$$

Dimana $\lambda > 0$ adalah parameter. Carilah estimator maksimum likelihood dari λ dan tunjukan bahwa estimator maksimum likelihood dari λ adalah estimator kecukupan dari parameter λ.

Jawab :

Pertama, cari kepadatan dari sampel atau fungsi likelihood dari sampel. Fungsi likelihood dari sampel diberikan oleh

$$L(\lambda) = \prod_{i=1}^{n} f(x_i; \lambda)$$

$$= \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$= \frac{\lambda^{n\overline{X}} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

Pengambilan logaritma dari fungsi likelihood kita punya

$$\ln L(\lambda) = n\overline{x} \ln \lambda - n\lambda - \ln \prod_{i=1}^{n} (x_i!)$$

Karenanya

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{1}{\lambda} n\overline{x} - n$$

Hitung turunan sampai nol dan selesaikan λ, kita punya

$$\lambda = \overline{x}$$

Uji turunan kedua diatas meyakinkan kita bahwa λ adalah rata-rata sampel $\overline{X}$. Selanjutnya, kita tunjukan bahwa $\overline{X}$ adalah kecukupan, dengan menggunakan teroema faktorisasi Fisher dan Neyman. Kita faktorkan kepadatan bersama dari sampel sebagai

$$L(\lambda) = \frac{\lambda^{n\overline{x}} e^{-n\lambda}}{\prod_{i=1}^{n} (x_i!)}$$

$$= \left\{ \lambda^{n\overline{x}} e^{-n\lambda} \right\} \frac{1}{\prod_{i=1}^{n} x_i!}$$

$$= \phi\left(\overline{X}, \lambda\right) h\left(x_1, x_2, \ldots, x_n\right)$$

Oleh karena itu, estimator $\overline{X}$ adalah estimator kecukupan dari λ.

Contoh 6.18. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi normal dengan fungsi kepadatan

$$f(x;\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}$$

Dimana $-\infty < \mu < \infty$ adalah parameter. Carilah estimator maksimum likelihood dari μ dan tunjukan bahwa estimator maksimum likelihood dari μ adalah estimator kecukupan.

Jawab :

Kita tahu bahwa estimator maksimum likelihood dari μ adalah rata-rata sampel $\overline{X}$. Selanjutnya, kita tunjukan bahwa estimator maksimum likelihood $\overline{X}$ ini adalah estimator kecukupan dari μ. Kepadatan dari sampel diberikan oleh

$$f\left(x_1, x_2, \ldots, x_n; \mu\right) = \prod_{i=1}^{n} f\left(x_i; \mu\right)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(x_i - \mu\right)^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\left(\sum_{i=1}^{n}\left[\left(x_i - \overline{x}\right) + \left(\overline{x} - \mu\right)\right]^2\right)}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\left(\sum_{i=1}^{n}\left[\left(x_i - \overline{x}\right)^2 + 2\left(x_i - \overline{x}\right)\left(\overline{x} - \mu\right) + \left(\overline{x} - \mu\right)\right]\right)}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\left(\sum_{i=1}^{n}\left[\left(x_i - \overline{x}\right)^2 + \left(\overline{x} - \mu\right)^2\right]\right)}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{n}{2}\left(\overline{x} - \mu\right)^2} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}$$

Karenanya, dengan teorema faktorisasi, $\overline{X}$ merupakan estimator kecukupan dari rata-rata populasi.

Ingat bahwa probabilitas fungsi kepadatan dari contoh 6.17 adalah

$$f(x;\lambda)=\begin{cases}\dfrac{\lambda^{x}e^{-\lambda}}{x!} & jika\; x=0,1,2,\ldots,\infty\\ 0 & untuk\ yang\ lainnya\end{cases}$$

Bisa ditulis seperti

$$f(x;\lambda)=e^{\{x\ln\lambda-\ln x!-\lambda\}}$$

Untuk $x=0,1,2,\ldots,\infty$. Fungsi kepadatan ini berbentuk

$$f(x;\lambda)=e^{\{K(x)A(\lambda)+S(x)+B(\lambda)\}}$$

Sama halnya fungsi kepadatan probabilitas dari contoh 6.12, yaitu

$$f(x;\lambda)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^{2}}$$

Bisa juga kita tulis

$$f(x;\mu)=e^{\left\{x\mu-\frac{x^{2}}{2}-\frac{\mu^{2}}{2}-\frac{1}{2}\ln(2\pi)\right\}}$$

Fungsi kepadatan probabilitas ini berbentuk

$$f(x;\mu)=e^{\{K(x)A(\mu)+S(x)+B(\mu)\}}$$

Kita juga bisa lihat bahwa kedua contoh estimator kecukupan adalah rata-rata sampel $\bar{X}$ yang bisa ditulis sebagai $\frac{1}{n}\sum_{i=1}^{n}X_i$

Teorema kita selanjutnya memberikan hasil yang umum pada petunjuk ini. Teorema ini disebut teorema Pitman-Koopman.

Teorema 6.4.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan fungsi kepadatan dari bentuk eksponensial

$$f(x;\theta)=e^{\{K(x)A(\theta)+S(x)+B(\theta)\}}$$

Pada kebebasan support dari θ. Maka, statistik dari $\sum_{i=1}^{n} K(X_i)$ adalah statistik kecukupan dari parameter θ.

Bukti :

Densitas gabungan dari sampel adalah

$$f(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x; \theta)$$

$$= \prod_{i=1}^{n} e^{\{K(x_i)A(\theta)+S(x_i)B(\theta)\}}$$

$$= e^{\left\{\sum_{i=1}^{n} K(x_i)A(\theta)+\sum_{i=1}^{n} S(x_i)+n\,B(\theta)\right\}}$$

$$= e^{\left\{\sum_{i=1}^{n} K(x_i)A(\theta)+n\,B(\theta)\right\}}\, e^{\left[\sum_{i=1}^{n} S(X_i)\right]}$$

Karena dengan teorema faktorisasi estimator dari $\sum_{i=1}^{n} K(X_i)$ adalah statistik kecukupan dari parameter θ.

Contoh 6.19. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan fungsi kepadatan

$$f(x;\theta) = \begin{cases} \theta\, x^{\theta-1} & \textit{untuk}\; 0 < x < 1 \\ 0 & \textit{untuk yang lainnya} \end{cases}$$

Dimana $\theta > 0$ adalah parameter. Gunakan teorema Pitman-Koopman untuk mencari estimator kecukupan dari θ.

Jawab :

Teorema Pitman-Koopman mengatakan bahwa jikafungsi kepadatan probabilitas bisa di ekspresikan dengan bentuk

$$f(x;\theta) = e^{\{K(x_i)A(\theta)+S(x_i)+B(\theta)\}}$$

Maka $\sum_{i=1}^{n} K(X_i)$ adalah statistik kecukupan dari θ. Diberikan populasi kepadatan bisa ditulis sebagai

$$f(x;\theta) = \theta x^{\theta-1}$$

$$= e^{\{\ln[\theta x^{\theta-1}]\}}$$

$$= e^{\{\ln\theta + (\theta-1)\ln x\}}$$

Jadi, $K(x) = \ln x, A(\theta) = \theta$,$S(x) = -\ln\theta$ dan $B(\theta) = \ln\theta$.

Karenanya dengan teorema Pitman-Koopman

$$\sum_{i=1}^{n} K(X_i) = \sum_{i=1}^{n} \ln X_i$$

$$= \ln \prod_{i=1}^{n} X_i$$

Jadi, $\ln \prod_{i=1}^{n} X_i$ merupakan statistik kecukupan dari θ.

Catatan 6.1. Perhatikan bahwa $\prod_{i=1}^{n} X_i$ juga merupakan statistik kecukupan dari θ, karena diketahui $\ln\left(\prod_{i=1}^{n} X_i\right)$ dan kita juga tahu bahwa $\prod_{i=1}^{n} X_i$

Contoh 6.20. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan fungsi kepadatan

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \textit{untuk } 0 < x < \infty \\ 0 & \textit{untuk yang lainnya} \end{cases}$$

Dimana $0 < \theta < \infty$ adalah parameter. Carilah estimator kecukupan dari θ.

Jawab :

Pertama, kita tulis ulang kepadatan populasi pada bentuk eksponensial, sehingga

$$f(x;\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$$

$$= e^{\ln\left[\frac{1}{\theta} e^{-\left(\frac{x}{\theta}\right)}\right]}$$

$$= e^{-\ln\theta - \frac{x}{\theta}}$$

Jadi, $K(x) = x,\ A(\theta) = -\frac{1}{\theta},\ S(x) = \theta,$ dan $B(\theta) = -\ln\theta$.

Jadi, $n\overline{X}$ adalah statistik kecukupan dari θ. Karena $n\overline{X}$ diketahui, kita juga tahu $\overline{X}$. Estimator $\overline{X}$ juga estimator kecukupan dari θ.

Contoh 6.21. Misalkan $X_1, X_2, \ldots X_n$ adalah sampel acak dari distribusi dengan fungsi kepadatan

$$f(x;\theta) = \begin{cases} e^{-(x-\theta)} & \textit{untuk}\ \theta < x < \infty \\ 0 & \textit{untuk yang lainnya} \end{cases}$$

Dimana $-\infty < \theta < \infty$ adalah parameter. Dapatkah teorema Pitman-Koopman digunakan untuk mencari statistik kecukupan dari θ?

Jawab :

Tidak. Kita tidak bisa gunakan teorema Pitman-Koopman untuk mencari statistik kecukupan dari θ ketika domain dimana populasi kepadatan bukanlah nol jadi ini tak bebas dari θ.

Selanjutnya, kita sajikan hubungan antara teorema Pitman-Koopman untuk mencari estimator maksimum likelihood dan estimator kecukupan. Jika ada estimator kecukupan dari parameter θ dan jika estimator maksimum likelihood dari θ unik, maka estimator maksimum likelihood adalah sebuah fungsi dari estimator kecukupan. Sehingga

$$\hat{\theta}_{ML} = \psi\left(\hat{\theta}_S\right)$$

Dimana ψ adalah fungsi bilangan asli, $\hat{\theta}_{ML}$ adalah estimator maksimum likelihood dari θ dan $\hat{\theta}_S$ adalah estimator kecukupan dari θ.

Demikian pula, koneksi dapat dibuat antara UMVUE dan estimator kecukupan dari parameter θ. Jika ada estimator kecukupan untuk parameter θ dan jika UMVUE dari ini adalah unik, maka UMVUE adalah fungsi dari estimator kecukupan. Yaitu

$$\hat{\theta}_{MVUE} = \eta\left(\hat{\theta}_S\right)$$

Dimana η adalah fungsi bernilai real, $\hat{\theta}_{MVUE}$ adalah UMVUE dari θ, dan $\hat{\theta}_S$ adalah estimator kecukupan dari θ.

Definisi 6.7.

Suatu penaksir/estimator dikatakan sebagai estimator yang diperlukan jika dapat ditulis sebagai fungsi dari setiap estimator kecukupan

6.5. Estimator Konsisten

Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak dari populasi X dengan densitas $f(x;\theta)$. Misalkan $\hat{\theta}$ menjadi estimator θ berdasarkan sampel berukuran n. Jelas penduga tergantung pada sampel berukuran n. Untuk mencerminkan bahwa $\hat{\theta}$ bergantung pada n, kita nyatakan $\hat{\theta}$ sebagai $\hat{\theta}_n$.

Definisi 6.8.

Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak dari populasi X dengan densitas $f(x;\theta)$. Barisan estimator $\{\hat{\theta}_n\}$ dari θ dikatakan menjadi konsisten untuk θ jika dan hanya jika barisan $\{\hat{\theta}_n\}$ konvergen pada probabilitas untuk θ, yaitu untuk setiap $\mathcal{E} > 0$

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| \geq \mathcal{E}\right) = 0$$

Perhatikan bahwa konsistensi sebenarnya adalah konsep yang berkaitan dengan barisan penduga $\{\hat{\theta}_n\}_{n=n_0}^{\infty}$ tapi kami biasanya mengatakan "konsistensi $\hat{\theta}_n$" untuk kesederhanaan. Selanjutnya, konsistensi adalah sifat sampel besar dari penduga.

Teorema berikut menyatakan bahwa jika kesalahan kuadrat rata-rata *(Mean Square Eror)* menjadi nol saat n menuju tak terhingga, maka $\{\hat{\theta}_n\}$ konvergen dalam probabilitas θ.

Teorema 6.5.

Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak dari populasi X dengan densitas $f(x;\theta)$ dan $\{\hat{\theta}_n\}$ menjadi barisan dari penaksir θ berdasarkan pada sampel. Jika varians $\hat{\theta}_n$ ada untuk setiap n dan berhingga dan

$$\lim_{n\to\infty} E\left(\left(\hat{\theta}_n - \theta\right)^2\right) = 0$$

Maka, untuk setiap $\mathcal{E} > 0$

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| \geq \mathcal{E}\right) = 0$$

Bukti :

Dengan menggunakan Pertidaksamaan Markov (lihat Teorema 3.8) kita mempunyai

$$P\left(\left(\widehat{\theta}_n - \theta\right)^2 \geq \mathcal{E}^2\right) \leq \frac{E\left(\left(\widehat{\theta}_n - \theta\right)^2\right)}{\mathcal{E}^2}$$

Untuk setiap $\mathcal{E} > 0$. Karena kejadian

$$\left(\widehat{\theta}_n - \theta\right)^2 \geq \mathcal{E}^2 \quad dan \quad \left|\widehat{\theta}_n - \theta\right| \geq \mathcal{E}$$

Adalah sama, kita lihat bahwa

$$P\left(\left(\widehat{\theta}_n - \theta\right)^2 \geq \mathcal{E}^2\right) = P\left(\left|\widehat{\theta}_n - \theta\right| \geq \mathcal{E}\right) \leq \frac{E\left(\left(\widehat{\theta}_n - \theta\right)^2\right)}{\mathcal{E}^2}$$

Untuk setiap $n \in \mathbb{N}$. Oleh karena jika

$$\lim_{n\to\infty} E\left(\left(\widehat{\theta}_n - \theta\right)^2\right) = 0$$

Maka

$$\lim_{n\to\infty} P\left(\left|\widehat{\theta_n} - \theta\right| \geq \mathcal{E}\right) = 0$$

Misalkan

$$B\left(\hat{\theta},\theta\right) = E\left(\hat{\theta}\right) - \theta$$

Adalah bias. Jika estimator tidak bias, maka $B\left(\hat{\theta},\theta\right) = 0$. Selanjutnya kita tunjukkan bahwa

$$E\left(\left(\widehat{\theta_n} - \theta\right)^2\right) = Var\left(\hat{\theta}\right) + \left[B\left(\hat{\theta},\theta\right)\right]^2 \qquad (1)$$

Untuk melihatnya, misalkan

$$E\left(\left(\widehat{\theta_n} - \theta\right)^2\right) = E\left(\left(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2\right)^2\right)$$

$$= E\left(\hat{\theta}^2\right) - 2E\left(\hat{\theta}\right)\theta + \theta^2$$

$$= E\left(\hat{\theta}^2\right) - E\left(\hat{\theta}\right)^2 + E\left(\hat{\theta}\right)^2 - 2E\left(\hat{\theta}\right)\theta + \theta^2$$

$$= Var\left(\hat{\theta}\right) + E\left(\hat{\theta}\right)^2 - 2E\left(\hat{\theta}\right)\theta + \theta^2$$

$$= Var\left(\hat{\theta}\right) + \left[E\left(\hat{\theta}\right) - \theta\right]^2$$

$$= Var\left(\hat{\theta}\right) + \left[B\left(\hat{\theta},\theta\right)\right]^2$$

Dengan melihat persamaan (1), kita dapat katakan bahwa

$$\lim_{n\to\infty} Var\left(\hat{\theta}_n\right) = 0 \qquad (2)$$

Dan

$$\lim_{n\to\infty} B\left(\hat{\theta}_n, \theta\right) = 0 \qquad (3)$$

Maka

$$\lim_{n\to\infty} E\left(\left(\hat{\theta}_n - \theta\right)^2\right) = 0$$

Dengan kata lain, untuk menunjukan barisan dari estimator adalah konsisten kita harus memeriksa limit (2) dan (3).

Contoh 6.22. Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak dari populasi normal X dengan mean μ dan varians $\sigma^2 > 0$. Apakah penaksir likelihood

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Untuk σ^2 merupakan estimator konsisten untuk σ^2 ?

Jawab :

Karena $\widehat{\sigma^2}$ tergantung pada ukuran sampel n, kita notasikan $\widehat{\sigma^2}$ sebagai $\widehat{\sigma_n^2}$. Oleh karena itu

$$\widehat{\sigma_n^2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Varians dari $\widehat{\sigma_n^2}$ adalah diberikan oleh

$$Var\left(\widehat{\sigma_n^2}\right) = Var\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right)$$

$$= \frac{1}{n^2} Var\left(\sigma^2 \frac{(n-1)S^2}{\sigma^2}\right)$$

$$= \frac{\sigma^4}{n^2} Var\left(\frac{(n-1)S^2}{\sigma^2}\right)$$

$$= \frac{\sigma^4}{n^2} Var\left(\chi^2(n-1)\right)$$

$$= \frac{2(n-1)\sigma^4}{n^2}$$

$$= \left[\frac{1}{n} - \frac{1}{n^2}\right] 2\sigma^4$$

Karenanya

$$\lim_{n\to\infty} Var\left(\hat{\theta}_n\right) = \lim_{n\to\infty} \left[\frac{1}{n} - \frac{1}{n^2}\right] 2\sigma^4 = 0$$

Bias $B\left(\hat{\theta}_n, \theta\right)$ diberikan oleh

$$B\left(\hat{\theta}_n, \theta\right) = E\left(\hat{\theta}_n\right) - \sigma^2$$

$$= E\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right) - \sigma^2$$

$$= \frac{1}{n} E\left(\sigma^2 \frac{(n-1)S^2}{\sigma^2}\right) - \sigma^2$$

$$= \frac{\sigma^2}{n} E\left(\chi^2(n-1)\right) - \sigma^2$$

$$= \frac{(n-1)\sigma^2}{n} - \sigma^2$$

$$= -\frac{\sigma^2}{n}$$

Jadi

$$\lim_{n\to\infty} B\left(\hat{\theta}_n,\theta\right) = -\lim_{n\to\infty}\frac{\sigma^2}{n} = 0$$

Oleh karena itu $\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ adalah estimator konsisten untuk σ^2

Dalam contoh terakhir kita melihat bahwa estimator likelihood varians adalah penduga yang konsisten. Secara umum, jika fungsi kepadatan $f\left(x;\theta\right)$ dari suatu populasi memenuhi beberapa kondisi, maka penduga maksimum likelihood dari θ adalah konsisten. Demikian pula, jika fungsi kepadatan $f\left(x;\theta\right)$ dari suatu populasi memenuhi beberapa kondisi, maka estimator yang diperoleh dengan metode momen juga konsisten.

Karena konsistensi adalah sifat sampel yang besar dari penaksir, beberapa ahli statistik menyarankan bahwa konsistensi tidak boleh digunakan sendiri untuk menilai kebaikan estimator, melainkan harus digunakan bersama dengan kriteria lainnya.

LATIHAN SOAL BAB 6

1. Diberikan $X_1,X_2,\ldots,X_n$ adalah sampel acak berdistribusi Poisson, $X_i \sim POI\left(\mu\right)$. Tunjukan bahwa $S=\sum_{i=1}^{n}X_i$ adalah statistik kecukupan untuk μ !

   Penyelesaian :

   Fungsi kepadatan probabilitas dari distribusi Poisson yaitu : $f\left(x\right)=\frac{e^{-\mu}\mu^x}{x!}$

   Sedangkan fungsi kepadatan probabilitas gabungan untuk distribusi Poisson yaitu :

$$f\left(x_1,x_2,\ldots,x_n;\mu\right)=\prod_{i=1}^{n}\frac{\mu^{x_i}}{x_i!}e^{-\mu}=\mu^{\sum_{i=1}^{n}x_i}e^{-\mu}\prod_{i=1}^{n}\frac{1}{x_i!}$$

   Diketahui bahwa $S=\sum_{i=1}^{n}X_i$ maka diperoleh

$$f_s\left(s;\mu\right)=\frac{e^{-n\mu}\mu^s}{n^s s!}$$

   Jika $s=\sum_{i=1}^{n}X_i$ maka

$$\frac{f(x_1,x_2,\dots,x_n;\mu)}{f_s(s;\mu)}=\frac{\mu^{\sum_{i=1}^{n}x_i}e^{-\mu}\prod_{i=1}^{n}\frac{1}{x_i!}}{\frac{e^{-n\mu}\mu^s}{n^s s!}}=\frac{s!}{n^s\prod_{i=1}^{n}x_i!}$$

Sedangkan 0 untuk yang lainnya. Karena PDF tersebut tidak bergantung pada μ maka $S=\sum_{i=1}^{n}X_i$ adalah statistik kecukupan untuk μ.

2. Diberikan sebuah sampel acak berukuran n berdistribusi geometrik, $X_i \sim GEO(p)$. Tunjukan bahwa $S=\sum_{i=1}^{n}X_i$ adalah statistik kecukupan untuk p.

Penyelesaian :

Untuk distribusi Geometrik, fungsi kepadatan probabilitasnya diberikan oleh

$$f(x,p)=\begin{cases} p(1-p)^{x-1}, \; untuk\, x\geq 1 \\ 0, \; untuk\; yang\, lainnya \end{cases}$$

Sedangkan fungsi kepadatan probabilitasnya bersamanya yaitu :

$$f(x_1,x_2,\dots,x_n,p)=p^n(1-p)^{-n+s}=\begin{cases} p^n(1-p)^{ns-n}, & jika\, x_1,x_2,\dots,x_n\geq 1 \\ 0, & untuk\; yang\, lainnya \end{cases}$$

Sehingga dapat dinyatakan dalam bentuk faktorisasi

$$g(s,p)=p^n(1-p)^{ns-n} \quad \text{dan} \quad h(x_1,x_2,\dots,x_n)=\begin{cases} 1, & jika\, x_i\geq 1 \\ 0, & untuk\; yang\, lainnya \end{cases}$$

Karena dapat dinyatakan dalam bentuk faktorisasi maka $S=\sum_{i=1}^{n}X_i$ adalah statistik kecukupan untuk p.

3. Misalkan $Y_1,Y_2,\dots,Y_n$ menjadi sampel random dari distribusi uniform pada interval $(0,\theta)$. Dua penaksir tidak bias dari θ adalah

$$\hat{\theta}_1=2\overline{Y} \quad dan \quad \hat{\theta}_2=\left(\frac{n+1}{n}\right)Y_{(n)},$$

Dimana $Y_{(n)}=\max(Y_1,Y_2,\dots,Y_n)$. Temukan Efisiensi dari $\hat{\theta}_1$ relatif terhadap $\hat{\theta}_2$

Penyelesaian :

Karena setiap Y_i adalah distribusi uniform pada interval $(0,\theta)$, $\mu = E(Y_i) = \dfrac{\theta}{2}$ dan $\sigma^2 = V(Y_i) = \dfrac{\theta^2}{12}$. Oleh karena itu,

$$E\left(\hat{\theta}_1\right) = E\left(2\overline{Y}\right) = 2E\left(\overline{Y}\right) = 2(\mu) = 2\left(\frac{\theta}{2}\right) = \theta$$

Dan $\hat{\theta}_1$ adalah tak bias, seperti yang di klaim. Lebih lanjut,

$$V\left(\hat{\theta}_1\right) = V\left(2\overline{Y}\right) = 4V\left(\overline{Y}\right) = 4\left[\frac{V(Y_i)}{n}\right] = \left(\frac{4}{n}\right)\left(\frac{\theta^2}{12}\right) = \frac{\theta^2}{3n}$$

Untuk menemukan mean dan varians dari $\hat{\theta}_2$, perhatikan bahwa fungsi densitas dari $Y_{(n)}$ adalah diberikan oleh

$$g_{(n)}(y) = n\left[F_Y(y)\right]^{n-1} f_Y(y) = \begin{cases} n\left(\dfrac{y}{\theta}\right)^{n-1}\left(\dfrac{1}{\theta}\right), & untuk\, 0 \le y \le \theta \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Kemudian,

$$E\left(Y_{(n)}\right) = \frac{n}{\theta^n}\int_0^\theta y^n\, dy = \left(\frac{n}{n+1}\right)\theta$$

Dan itu mengikuti $E\left\{\left[\dfrac{(n+1)}{n}\right]Y_{(n)}\right\} = \theta$; itu adalah, $\hat{\theta}_2$ adalah estimator tidak bias untuk θ.

Karena

$$E\left(Y_{(n)}^2\right) = \frac{n}{\theta^n}\int_0^\theta y^{n+1}\, dy = \left(\frac{n}{n+2}\right)\theta^2,$$

Kita peroleh

$$V\left(Y_{(n)}\right) = E\left(Y_{(n)}^2\right) - \left[E\left(Y_{(n)}\right)\right]^2 = \left[\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right]\theta^2$$

Dan

$$V\left(\hat{\theta}^2\right) = V\left[\left(\frac{n}{n+1}\right)Y_{(n)}\right] = \left(\frac{n}{n+1}\right)^2 V\left(Y_{(n)}\right)$$

$$= \left[\frac{(n+1)^2}{n(n+2)} - 1\right]\theta^2 = \frac{\theta^2}{n(n+2)}$$

Oleh karena itu, Efisiensi dari $\hat{\theta}_1$ relatif terhadap $\hat{\theta}_2$ adalah diberikan oleh

$$eff\left(\hat{\theta}_1, \hat{\theta}_2\right) = \frac{V\left(\hat{\theta}_2\right)}{V\left(\hat{\theta}_1\right)} = \frac{\dfrac{\theta^2}{\left[n(n+2)\right]}}{\dfrac{\theta^2}{3n}} = \frac{3}{n+2}$$

Efisiensi ini kurang dari 1 jika $n > 1$. Artinya jika $n > 1$, $\hat{\theta}^2$ memiliki varians yang lebih kecil dari $\hat{\theta}_1$, dan oleh karena itu $\hat{\theta}_2$ umumnya lebih disukai daripada $\hat{\theta}_1$ sebagai estimator dari θ.

4. Misalkan bahwa $Y_1, Y_2, \ldots, Y_n$ merupakan sampel acak seperti pada $E\left(Y_i\right) = \mu, E\left(Y_i^2\right) = \mu_2^{'}$ dan $E\left(Y_i^4\right) = \mu_4^{'}$ adalah semua terbatas. Tunjukkan bahwa

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}_n\right)^2$$

Adalah estimator konsisten dari $\sigma^2 = V\left(Y_i\right)$.

Penyelesaian :

Kita telah melihat di bab sebelumnya bahwa S^2, sekarang ditulis sebagai S_n^2,

$$S_n^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}_n^2\right) = \left(\frac{n}{n-1}\right)\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}_n^2\right)$$

Statistik $\frac{1}{n}\sum_{i=1}^{n} Y_i^2$ adalah rata-rata dari n independen dan distribusi identik variabel acak,

dengan $E\left(Y_i^2\right) = \mu_2^{'}$ dan $V\left(Y_i^2\right) = \mu_4^{'} - \left(\mu_2^{'}\right)^2 < \infty$. Dengan menggunakan aturan bilangan

besar, kita tahu bahwa $\frac{1}{n}\sum_{i=1}^{n} Y_i^2$ konvergen pada probabilitas dari $\mu_2^{'}$.

Kita tahu implikasi bahwa $\overline{Y}_n$ konvergen pada probabilitas dari μ. Karena fungsi $g(x)=x^2$ adalah kontinu untuk semua nilai berhingga dari x. Jika $g(\bullet)$ adalah fungsi bernilai real bahwa kontinu pada θ, sehingga $g(\hat{\theta}_n)$ konvergen pada probabilitas $g(\theta)$ mengimplikasikan bahwa $\overline{Y}_n^2$ konvergen pada probabilitas untuk μ^2. Ini kemudian mengikuti teorema bahwa $\hat{\theta}_n+\hat{\theta}_n^{'}$ konvergen pada probabilitas untuk $\theta+\theta'$. Bahwa

$$\frac{1}{n}\sum_{i=1}^{n}Y_i^2-\overline{Y}_n^2$$

Konvergen pada probabilitas untuk $\mu_2^{'}-\mu^2=\sigma^2$. Karena $\frac{n}{n-1}$ adalah barisan dari konvergen konstan untuk 1 sebagai $n\to\infty$, kita dapat menyimpulkan bahwa S_n^2 konvergen pada probabilitas untuk σ^2. Ekuivalen S_n^2 sampel varians adalah estimator konsisten untuk σ^2 populasi varians

5. Misalkan $Y_1,Y_2,\ldots,Y_n$ menjadi sampel acak dengan $E(Y_i)=\mu$ dan $V(Y_i)=\sigma^2$. Tunjukkan bahwa

$$S'^2=\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2$$

Adalah estimator bias dari σ^2 dan tunjukkan bahwa

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2$$

Adalah estimator tidak bias dari σ^2

Penyelesaian :

Ini dapat ditunjukkan bahwa

$$\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2=\sum_{i=1}^{n}Y_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}Y_i\right)^2=\sum_{i=1}^{n}Y_i^2-n\overline{Y}^2$$

Karenanya

$$E\left[\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2\right]=E\left(\sum_{i=1}^{n}Y_i^2\right)-nE\left(\overline{Y}^2\right)$$

$$= \sum_{i=1}^{n} E(Y_i^2) - n\left(\overline{Y}^2\right)$$

Perhatikan bahwa $E(Y_i^2)$ adalah sama untuk $i = 1, 2, \ldots, n$. Kita menggunakan ini dan fakta bahwa varians dari variabel acak adalah diberikan oleh $V(Y) = E\left(Y^2\right) - \left[E(Y)\right]^2$ untuk menyimpulkan bahwa $E(Y_i^2) = V(Y_i) + \left[E(Y_i)\right]^2 = \sigma^2 + \mu^2$, $E\left(\overline{Y}^2\right) = V\left(\overline{Y}\right) + \left[E\left(\overline{Y}\right)\right]^2 = \dfrac{\sigma^2}{n + \mu^2}$

Dan bahwa

$$E\left[\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2\right] = \sum_{i=1}^{n}\left(\sigma^2 + \mu^2\right) - \left(\frac{\sigma^2}{n} + \mu^2\right)$$

$$= n\left(\sigma^2 + \mu^2\right) - n\left(\frac{\sigma^2}{n} + \mu^2\right)$$

$$= n\sigma^2 - \sigma^2 = (n-1)\sigma^2$$

Ini mengikuti bahwa

$$E\left(S'^2\right) = \frac{1}{n} E\left[\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2\right]$$

$$= \frac{1}{n}(n-1)\sigma^2$$

$$= \left(\frac{n-1}{n}\right)\sigma^2$$

Dan bahwa S'^2 adalah bias karena $E\left(S'^2\right) \neq \sigma^2$. Akan tetapi,

$$E\left(S^2\right) = \frac{1}{n-1} E\left[\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2\right]$$

$$=\frac{1}{n-1}(n-1)\sigma^2$$

$$=\sigma^2$$

Jadi, kita lihat bahwa S^2 adalah estimator tidak bias dari σ^2

6. Misalkan $Y_1, Y_2, \ldots, Y_n$ menjadi sampel acak dengan Y_i memiliki fungsi kepadatan probabilitas

$$f(y_i \mid \theta)=\begin{cases}\left(\frac{1}{\theta}\right)e^{-\frac{y_i}{\theta}}, \text{ untuk } 0 \le y_i < \infty \\ \\ 0, \qquad \text{untuk yang lainnya}\end{cases}$$

Dimana $\theta > 0,\ i = 1, 2, \ldots, n$. Tunjukkan bahwa $\overline{Y}$ adalah statistik kecukupan untuk parameter θ.

Penyelesaian :

Likelihood $L(\theta)$ dari sampel adalah densitas bersama

$$L(y_1, y_2, \ldots, y_n \mid \theta) = f(y_1, y_2, \ldots, y_n \mid \theta)$$

$$= f(y_1 \mid \theta) \times f(y_2 \mid \theta) \times \ldots \times f(y_n \mid \theta)$$

$$=\frac{e^{-\frac{y_1}{\theta}}}{\theta}\times\frac{e^{-\frac{y_2}{\theta}}}{\theta}\times\ldots\times\frac{e^{-\frac{y_n}{\theta}}}{\theta}$$

$$=\frac{e^{-\Sigma\frac{y_1}{\theta}}}{\theta^n}$$

$$=\frac{e^{-\frac{n\overline{y}}{\theta}}}{\theta^n}$$

Perhatikan bahwa $L(\theta)$ adalah fungsi hanya dari θ dan $\overline{y}$ dan bahwa jika

$$g(\bar{y},\theta)=\frac{e^{-\frac{n\bar{y}}{\theta}}}{\theta^n} \quad dan \quad h(y_1,y_2,\ldots,y_n)=1$$

Kemudian

$$L(y_1,y_2,\ldots,y_n\,|\theta)=g(\bar{y},\theta)\times h(y_1,y_2,\ldots,y_n)$$

Oleh karena itu, dengan menggunakan teorema dibawah ini,

Misalkan U menjadi statistik berdasarkan pada sampel acak $Y_1,Y_2,\ldots,Y_n$. Kemudian U adalah statistik kecukupan untuk estimasi dari parameter θ jika dan hanya jika likelihood $L(\theta)=L(y_1,y_2,\ldots,y_n\,|\theta)$ dapat difaktorkan menjadi dua fungsi nonnegatif,

$$L(y_1,y_2,\ldots,y_n\,|\theta)=g(u,\theta)\times h(y_1,y_2,\ldots,y_n)$$

Dimana $g(u,\theta)$ adalah fungsi hanya dari u dan θ dan $h(y_1,y_2,\ldots,y_n)$ bukanlah fungsi dari θ.

Mengimplikasikan bahwa $\bar{Y}$ adalah statistik kecukupan untuk parameter θ.

7. Misalkan bahwa $Y_1,Y_2,\ldots,Y_n$ menunjukan sampel acak dari fungsi densitas Weibull, diberikan oleh

$$f(y\,|\,\theta)=\begin{cases}\left(\dfrac{2y}{\theta}\right)e^{-\frac{y^2}{\theta}}, & untuk\ y>0\\[2ex] 0, & untuk\ yang\ lainnya\end{cases}$$

Temukan MVUE *(Minimum Variance Unbiased Estimation)* untuk θ.

Penyelesaian :

Kita mulai dengan faktorisasi kriteria untuk menemukan statistik kecukupan

$$L(y_1,y_2,\ldots,y_n\,|\theta)=f(y_1,y_2,\ldots,y_n\,|\theta)$$

$$=\left(\frac{2}{\theta}\right)^n(y_1\times y_2\times\ldots\times y_n)\exp\left(-\frac{1}{\theta}\sum_{i=1}^{n}y_i^2\right)$$

$$= \underbrace{\left(\frac{2}{\theta}\right)^n \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} y_i^2\right)}_{g\left(\Sigma y_i^2, \theta\right)} \times \underbrace{\left(y_1 \times y_2 \times \ldots \times y_n\right)}_{h\left(y_1, y_2, \ldots, y_n\right)}$$

Jadi, $U = \sum_{i=1}^{n} Y_i^2$ adalah minimal statistik kecukupan untuk θ.

Sekarang harus menemukan fungsi dari statistik ini bahwa itu adalah tidak bias untuk θ. Misalkan $W = Y_i^2$, kita dapatkan

$$f_w(w) = f\left(\sqrt{w}\right)\frac{d\left(\sqrt{w}\right)}{dw}$$

$$= \left(\frac{2}{\theta}\right)\left(\sqrt{w}\, e^{-\frac{w}{\theta}}\right)\left(\frac{1}{2\sqrt{w}}\right)$$

$$= \left(\frac{1}{\theta}\right)e^{-\frac{w}{\theta}},\ w > 0$$

Dengan demikian Y_i^2 memiliki distribusi eksponensial dengan parameter θ. Karena

$$E\left(Y_i^2\right) = E\left(W\right) = \theta \quad dan \quad E\left(\sum_{i=1}^{n} Y_i^2\right) = n\theta$$

Ini mengikuti bahwa

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

Adalah estimator tidak bias dari θ itu adalah fungsi dari statistik kecukupan $\sum_{i=1}^{n} Y_i^2$. Oleh karena itu, θ adalah MVUE dari parameter θ Weibull.

8. Misalkan $X_1, \ldots, X_n$ sampel acak dari Distribusi Bernoulli, $X_i \sim BIN\left(1; p\right); 0 < p < 1$.
   a. Cari UMVUE dari $Var\left(X\right) = p\left(1 - p\right)$.
   b. Cari UMVUE dari p^2.

Penyelesaian :

$$X_i \sim BIN(1, p)$$

Untuk sampel acak n, $\ell(X_i) = X_i$ dan $S = \sum_{i=1}^{n} X_i$ merupakan statistik kecukupan lengkap untuk p.

a. UMVUE dari $Var(X) = p(1-p)$

   Menggunakan $\bar{X}(1-\bar{X})$

$$E\left[\bar{X}(1-\bar{X})\right] = E(\bar{X}) - E(\bar{X}^2)$$

$$E\left[\bar{X}(1-\bar{X})\right] = E(\bar{X}) - \left[E(\bar{X})^2 + Var(\bar{X})\right]$$

$$E\left[\bar{X}(1-\bar{X})\right] = p - \left(p^2 + Var(\bar{X})\right)$$

$$E\left[\bar{X}(1-\bar{X})\right] = p - p^2 - \frac{p(1-p)}{n}$$

$$E\left[\bar{X}(1-\bar{X})\right] = p(1-p)\left(1 - \frac{1}{n}\right)$$

$$E\left[\bar{X}(1-\bar{X})\right] = p(1-p)\left(\frac{n-1}{n}\right)$$

$$E\left[\bar{X}(1-\bar{X})\left(\frac{n}{n-1}\right)\right] = p(1-p)$$

$$E\left[\frac{n\bar{X}(1-\bar{X})}{n-1}\right] = p(1-p)$$

UMVUE dari $p(1-p)$ adalah $c\bar{X}(1-\bar{X})$ dimana $c = \dfrac{n}{n-1}$

b. UMVUE dari p^2

Menggunakan $\bar{X}\left(1-\bar{X}\right)$

$$E\left[\bar{X}\left(1-\bar{X}\right)\right]=E\left(\bar{X}\right)-E\left(\bar{X}^2\right)$$

$$E\left[\bar{X}\left(1-\bar{X}\right)\right]=E\left(\bar{X}\right)-\left[E\left(\bar{X}\right)^2+Var\left(\bar{X}\right)\right]$$

$$E\left[\bar{X}\left(1-\bar{X}\right)\right]=p-\left(p^2+Var\left(\bar{X}\right)\right)$$

$$E\left[\bar{X}\left(1-\bar{X}\right)\right]=p-p^2-\frac{p\left(1-p\right)}{n}$$

$$E\left[\bar{X}\left(1-\bar{X}\right)\right]=p\left(1-p\right)\left(1-\frac{1}{n}\right)$$

$$E\left[\bar{X}\left(1-\bar{X}\right)\right]=p\left(1-p\right)\left(\frac{n-1}{n}\right)$$

$$E\left[\bar{X}\left(1-\bar{X}\right)\left(\frac{n}{n-1}\right)\right]=p\left(1-p\right)$$

$$E\left[\frac{n\bar{X}\left(1-\bar{X}\right)}{n-1}\right]=p\left(1-p\right)$$

$$E\left[\frac{n\bar{X}\left(1-\bar{X}\right)}{n-1}\right]=p-p^2$$

$$p^2=p-E\left[\frac{n\bar{X}\left(1-\bar{X}\right)}{n-1}\right]$$

UMVUE dari p^2 adalah $p-E\left[\frac{n\bar{X}\left(1-\bar{X}\right)}{n-1}\right]$

9. Misalkan sampel acak n dari Distribusi Poisson, $X_i \sim POI(\mu);\mu > 0$. Cari UMVUE dari $P[X = 0] = e^{-\mu}$.

   Penyelesaian :

   PDF $f(x_1,\ldots,x_n;\mu) = \dfrac{e^{-n\mu}\mu^{\Sigma x}}{\Pi(x_i!)}$

   Dengan kriteria faktorisasi, $S = \sum X_i$ merupakan statistik kecukupan.

$$S \sim POI(n\mu)$$

   Sehingga fungsi $u(s)$

$$E[u(s)] = \sum_{s=0}^{\infty} u(s) f(x_i,\ldots,x_n;\mu)$$

$$= \sum_{s=0}^{\infty} \frac{u(s)e^{-n\mu}n\mu^s}{s!}$$

   Karena $e^{-n\mu} \neq 0$, agar $E[u(s)] = 0$ untuk setiap koefisien $\frac{u(s)}{s!}, n\mu^s$ harus nol. Jika $\frac{u(s)}{s!} = 0$ maka $u(s) = 0$. Dengan kelengkapan, $\bar{X} = \frac{s}{n}$ fungsi tunggal S dan unbias untuk $E(\bar{X}) = \mu$ maka dari teorema yang telah dipelajari jelas bahwa $\bar{X} = \frac{s}{n}$ adalah UMVUE dari μ.

BAB VII

BEBERAPA TEKNIK DALAM MENCARI ESTIMATOR INTERVAL PARAMETER

Dalam estimasi titik kita mencari nilai parameter θ yang diberikan oleh sebuah data sampel. Contoh, jika $X_1, X_2, \ldots, X_n$ adalah sampel acak berukuran n dari sebuah populasi dengan fungsi kepadatan probabilitas

$$f(x;\theta) = \begin{cases} \sqrt{\dfrac{2}{\pi}}\, e^{-\frac{1}{2}(x-\theta)^2} & \textit{untuk } x \geq \theta \\ \\ 0 & \textit{untuk yang lainnya} \end{cases}$$

Kemudian fungsi likelihood dari θ adalah

$$L(\theta) = \prod_{i=1}^{n} \sqrt{\frac{2}{\pi}}\, e^{-\frac{1}{2}(x-\theta)^2}$$

Ketika $x_1 \geq \theta, x_2 \geq \theta, \ldots, x_n \geq \theta$. Fungsi likelihood ini disederhanakan

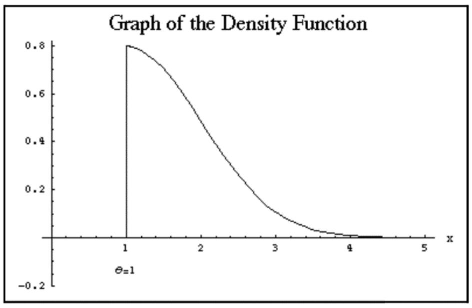
$$L(\theta) = \left[\frac{2}{\pi}\right]^{\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2}$$

Dimana $\min\{x_1, x_2, \ldots, x_n] \geq \theta$. Dengan menggunakan logaritma natural dari $L(\theta)$ dan maksimum, kita memperoleh estimator maksimum likelihood dari θ sebagai statistik terurut pertama dari sampel $X_1, X_2, \ldots, X_n$ yaitu

$$\hat{\theta} = X_{(1)}$$

Dimana $X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}$. Misalkan bilangan asli dari $\theta = 1$. Dengan menggunakan estimator maksimum likelihood dari θ, kita akan mencoba untuk menduga bilangan dari θ berdasarkan sampel acak. Misalkan $X_1 = 1.5,\ X_2 = 1.1,\ X_3 = 1.7,\ X_4 = 2.1,\ X_5 = 3.1$ adalah himpunan data sampel dari populasi diatas. Maka berdasarkan sampel acak kita dapatkan

$$\hat{\theta}_{ML} = X_{(1)} = \min\{1.5, 1.1, 1.7, 2.1, 3.1\} = 1.1$$



Jika kita ambil sampel acak lain, katakanlah

$$X_1 = 1.8,\ X_2 = 2.1,\ X_3 = 2.5,\ X_4 = 3.1,\ X_5 = 2.6$$

maka estimator maksimum likelihood dari θ akan menjadi $\hat{\theta} = 1.8$ berdasarkan sampel tersebut. Grafik dari fungsi kepadatan $f(x;\theta)$ untuk $\theta = 1$ adalah sebagai berikut

Dari grafik diatas, sangat jelas bahwa bilangan yang berdekatan dengan angka 1 mempunyai kesempatan terambil acak yang tinggi dalam proses pengambilan sampel, maka bilangan tersebut jauh lebih besar dari 1. Oleh karena itu, masuk akal jika θ mengestimasikan nilai sampel terkecil. Namun, dari contoh kita lihat bahwa estimasi titik dari θ tidak sama dengan bilangan asli dari θ. Bahkan jika kita mengambil beberapa sampel acak, namun estimasi dari θ akan jarang sama dengan parameter yang sebenarnya. Oleh karena itu, daripada mencari nilai tunggal dari θ, kita harus merencanakan berbagai kemungkinan nilai pada parameter θ dengan interval kepercayaan parameter tertentu.

7.1. Estimator Interval dan Interval Kepercayaan dari Parameter

Masalah estimasi interval dapat dinyatakan sebagai berikut : Diberikan sampel acak $X_1, X_2, \ldots, X_n$ dan nilai probabilitas $1-\alpha$, carilah pasangan statistik $L = L(X_1, X_2, \ldots, X_n)$ dengan $L \le U$ sedemikian sehingga probabilitas dari θ berada di interval $[L,U]$ adalah $1-\alpha$. Sehingga

$$P(L \le \theta \le U) = 1-\alpha$$

Ingat bahwa bagian dari populasi biasanya memilih dengan metode sampling acak dan sama halnya himpunan variabel acak $X_1, X_2, \ldots, X_n$ dengan fungsi kepadatan probabilitas $f(x;\theta)$ yang sama sebagai populasi. Setelah sampling atau pengambilan sampel selesai kita dapatkan

$$X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$$

Dimana $x_1, x_2, \ldots, x_n$ adalah data sampel.

Definisi 7.1.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak berukuran n dari sebuah populasi X dengan kepadatan $f(x;\theta)$ dimana θ adalah parameter yang tidak diketahui. Estimator interval dari θ adalah pasangan dari statistik $L = L(X_1, X_2, \ldots, X_n)$ dan $U = U(X_1, X_2, \ldots, X_n)$ dengan $L \leq U$ sedemikian sehingga jika $x_1, x_2, \ldots, x_n$ adalah himpunan dari data sampel, maka θ termasuk dalam interval $\left[L(x_1, x_2, \ldots, x_n), U(x_1, x_2, \ldots, x_n)\right]$.

Interval $[l, u]$ akan dinotasikan sebagai estimasi interval θ sedangkan interval acak $[L, U]$ akan dinotasikan sebagai estimator interval θ. Perhatikan bahwa, estimator interval θ adalah interval acak $[L, U]$. Selanjutnya, kita definisikan sebagai $100(1-\alpha)\%$ interval kepercayaan untuk parameter θ yang tidak diketahui.

Definisi 7.2.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak berukuran n dari sebuah populasi X dengan kepadatan $f(x;\theta)$ dimana θ adalah parameter yang tidak diketahui. Estimator interval dari θ dinamakan $100(1-\alpha)\%$ interval kepercayaan jika

$$P(L \leq \theta \leq U) = 1 - \alpha$$

Variabel acak L disebut limit kepercayaan terendah dan U dinamakan limit kepercayaan terbesar. Bilangan $(1-\alpha)$ disebut koefisien kepercayaan atau derajat kepercayaan.

Ada beberapa metode dalam menyusun interval kepercayaan untuk parameter yang tidak diketahui seperti :

1) Metode Kuantitas Pivotal
2) Metode Estimator Maksimum Likelihood
3) Metode Bayes
4) Metode Invarian
5) Metode Inversi Uji Statistik
6) Metode Statistik atau Umum

Pada bab ini kita hanya fokus pada metode kuantitas pivot dan metode estimator maksimum likelihood. Kita juga sekilas memeriksa metode statistik atau umum. Metode kuantitas pivot ini mungkin satu dari banyaknya metode elegan dari penyusunan interval kepercayaan dengan parameter yang tidak diketahui.

7.2. Metode Kuantitas Pivotal

Pada bagian ini, kita jelaskan bagaimana kuantitas pivot dapatdigunakan menjadi penyusun interval kepercayaan untuk parameter yang tidak diketahui. Kita akan belajar bagaimana cara menemukan kuantitas pivot terkait dengan fungsi densitas probabilitas tertentu, kita mulai dengan definisi formal dari kuantitas pivot.

Definisi 7.3.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak berukuran n dari populasi X dengan fungsi kepadatan probabilitas $f(x;\theta)$ dimana θ adalah parameter yang tidak diketahui. Kuantitas pivot Q adalah fungsi dari $X_1, X_2, \ldots, X_n$ dan θ merupakan distribusi probabilitas yang independen dari parameter θ

Catatan bahwa kuantitas pivotal $Q(X_1, X_2, \ldots, X_n, \theta)$ biasanya mengandung dua parameter θ dan sebuah estimator dari θ.

Contoh 7.1.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari populasi normal X dengan rata-rata μ dan varians σ^2 yang diketahui. Carilah kuantitas pivotal dengan parameter μ tidak diketahui.

Jawab :

Karena setiap $X_i \sim N(\mu, \sigma^2)$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Standardisasi $\bar{X}$ kita punya

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Statistik Q diberikan oleh

$$Q(X_1, X_2, \ldots, X_n, \mu) = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Adalah kuantitas pivotal yang merupakan fungsi dari $X_1, X_2, \ldots, X_n$. dan μ dan fungsi kepadatan probabilitasnya bebas dari parameter μ.

Tidak ada aturan umum dalam mencari kuantitas pivotal untuk parameter θ dari sebuah fungsi kepadatan $f(x;\theta)$ yang berubah – ubah. Oleh karena itu untuk beberapa tingkatan, mencari kuantitas pivotal mengandalkan tebak menebak. Namun, jika fungsi kepadatan probabilitas $f(x;\theta)$ termasuk kedalam keluarga skala lokasi, lalu ada jalan sistematis untuk mencari pivot.

Definisi 7.4.

Misalkan $g:\mathbb{R}\rightarrow\mathbb{R}$ adalah fungsi kepadatan probabilitas. Untuk setiap $\mu>\theta$ dan $\sigma>\theta$ maka keluarga dari fungsi

$$\mathcal{F}=\left\{f(x;\mu,\sigma)=\frac{1}{\sigma}g\left(\frac{x-\mu}{\sigma}\right)\mid\mu\in(-\infty,\infty),\sigma\in(0,\infty)\right\}$$

Dinamakan keluarga skala lokasi dengan standar kepadatan probabilitas $f(x;\theta)$. Parameter μ disebut parameter lokasi dan parameter σ disebut parameter skala. Jika $\sigma=1$ maka $\mathcal{F}$ disebut keluarga lokasi. Jika $\mu=0$ maka $\mathcal{F}$ disebut keluarga skala.

Ini dinotasikan bahwa setiap keanggotaan $f(x;\mu,\sigma)$ dari keluarga skala lokasi disebut fungsi kepadatan probabilitas. Jika kita ambil

$g(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ maka fungsi kepadatan normal dari

$$f(x;\mu,\sigma)=\frac{1}{\sigma}g\left(\frac{x-\mu}{\sigma}\right)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},\quad -\infty<x<\infty$$

Termasuk keluarga skala lokasi. Fungsi kepadatan dari

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}} & jika\, 0<x<\infty\\ 0 & untuk\ yang\ lainnya\end{cases}$$

rmasuk keluarga skala. Namun, fungsi kepadatan dari

$$f(x;\theta)=\begin{cases}\theta x^{\theta-1} & jika\, 0<x<\infty\\ 0 & untuk\ yang\ lainnya\end{cases}$$

Tidak termasuk keluarga skala lokasi.

Secara keseluruhan sangat mudah dalam mencari lokasi kuantitas pivotal atau parameter skala ketika fungsi kepadatan dari populasi termasuk kedalam keluarga skala lokasi $\mathcal{F}$. Ketika fungsi densitas termasuk keluarga lokasi, pivot dari parameter lokasi μ adalah $\widehat{\mu}-\mu$, dimana $\widehat{\mu}$ adalah estimator maksimum likelihood dari μ. Jika adalah estimator maksimum likelihood dari σ. , maka pivot dari parameter skala σ adalah $\frac{\widehat{\sigma}}{\sigma}$ saat fungsi kepadatannya termasuk dalam keluarga skala. Pivot parameter lokasi μ adalah $\frac{\widehat{\mu}-\mu}{\widehat{\sigma}}$ dan pivot untuk parameter skala σ adalah $\frac{\widehat{\sigma}}{\sigma}$ saat fungsi kepadatannya termasuk dalam keluarga skala lokasi. Terkadang ini sesuai untuk membuat modifikasi minor pada pivot yang diperoleh dengan cara ini, seperti penjumlahan bilangan konstan, sehingga modifikasi pivot dapat diketahui distribusinya.

Catatan 7.1. Kuantitas pivotal bisa juga disusun dengan menggunakan statistik kecukupan untuk parameter. Misalkan, $T = T(X_1, X_2, \ldots, X_n)$ adalah statistik kecukupan berdasarkan sampel acak $X_1, X_2, \ldots, X_n$ dari populasi X dengan fungsi kepadatan probabilitas $f(x;\theta)$. Misalkan fungsi kepadatan probabilitas dari T menjadi $g(t;\theta)$. Jika $g(t;\theta)$ termasuk dalam keluarga lokasi, maka kesesuaian kelipatan konstan dari $T - a(\theta)$ adalah kuantitas pivotal untuk parameter lokasi θ dari beberapa pernyataan yang sesuai $a(\theta)$. Jika $g(t;\theta)$ termasuk dalam keluarga skala lokasi, maka

kesesuaian kelipatan konstan dari $\frac{T}{b(\theta)}$ adalah kuantitas pivotal untuk

parameter skala θ untuk beberapa pernyataan yang sesuai $b(\theta)$. Demikian pula jika $g(t;\theta)$ termasuk dalam keluarga skala lokasi, maka

kesesuaian kelipatan konstan dari $\frac{T-a(\theta)}{b(\theta)}$ adalah kuantitas pivotal dari

parameter lokasi θ untuk beberapa pernyataan yang sesuai $a(\theta)$ dan $b(\theta)$ Manipulasi aljabar dari pivot adalah faktor kunci dalam mencari interval kepercayaan. Jika $Q = Q(X_1, X_2, \ldots, X_n)$ adalah sebuah pivot, maka interval kepercayaan $100(1-\alpha)\%$ bisa disusun dengan menggunakan : Pertama, cari nilai a dan b sehingga

$$P(a \le Q \le b) = 1-\alpha$$

Lalu ubah pertidaksamaan dari $a \le Q \le b$ menjadi bentuk $L \le \theta \le U$

Sebagai contoh, jika X adalah populasi normal dengan rata-rata μ tidak diketahui dan varians σ^2 diketahui, maka PDF termasuk dalam keluarga skala lokasi. Pivot μ adalah $\frac{\overline{X}-\mu}{S}$. Namun, ketika varians σ^2 diketahui, maka tidak perlu mengambil S. Jadi, kita berpendapat bahwa pivot $\frac{\overline{X}-\mu}{S}$ disusun oleh interval kepercayaan $100(1-2\alpha)\%$ untuk μ. Ketika populasi $X \sim N(\mu, \sigma^2)$, rata-rata sampel $\overline{X}$ juga normal dengan rata-rata μ dan varians $\frac{\sigma}{\sqrt{n}}$. Karenanya

$$1-2\alpha = P\left(-z_\alpha \le \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le z_\alpha\right)$$

$$= P\left(\mu - z_\alpha \frac{\sigma}{\sqrt{n}} \le \overline{X} \le \mu + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

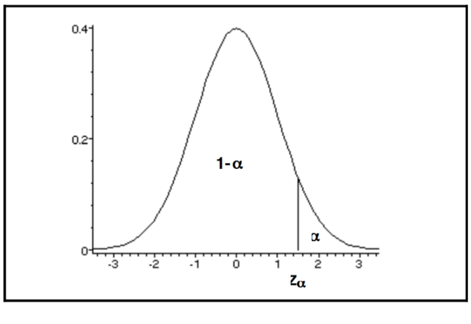
$$= P\left(\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

Oleh karena itu interval kepercayaan $100(1-2\alpha)\%$ untuk μ adalah

$$\left[\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}}, \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}\right]$$

z_α disini dinotasikan sebagai $100(1-\alpha)\%$ persentil atau kuartil $100(1-\alpha)$ dari normal standar variabel acak Z, sehingga

$$P(Z \le z_\alpha) = 1-\alpha$$



Dimana $\alpha \le 0.5$. Perhatikan bahwa $\alpha = P(Z \le -z_\alpha)$ jika $\alpha \le 0.5$

Interval kepercayaan $100(1-\alpha)\%$ untuk parameter θ harus mempunyai interpretasi. Jika $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ adalah sampel berukuran n maka berdasarkan sampel ini, kita susun interval kepercayaan $100(1-\alpha)\%$ $[l,u]$ dengan sub-interval dari garis real $\mathbb{R}$. Misalkan kita ambil bilangan besar dari interval kepercayaan $100(1-\alpha)\%$, maka aproksiamasi interval tersebut termasuk nilai parameter θ yang tidak diketahui.

Pada bagian selanjutnya, kita ilustrasikan bagaimana metode kuantitas pivotal dapat digunakan untuk menentukan interval kepercayaan untuk berbagai parameter

7.3. Interval Kepercayaan untuk Populasi Mean

Pada awalnya, kita menggunakan metode kuantitas pivot untuk membangun interval kepercayaan untuk rata-rata populasi normal. Di sini kita asumsikan terlebih dahulu bahwa populasi varians diketahui dan kemudian varians tidak diketahui. Selanjutnya kita membangun interval kepercayaan untuk rata-rata populasi dengan distribusi probabilitas kontinu, simetris dan unimodal dengan menerapkan teorema limit pusat.

Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak dari populasi $X \sim N\left(\mu, \sigma^2\right)$ dimana μ adalah parameter yang diketahui dan σ^2 adalah parameter yang diketahui. Pertama-tama, kita membutuhkan kuantitas pivot $Q = \left(X_1, X_2, \ldots, X_n, \mu\right)$. Untuk menyusun kuantitas pivot ini, kita definisikan estimator likelihood dari parameter μ. Kita tahu bahwa $\hat{\mu} = \bar{X}$. Karena setiap $X_i \sim N\left(\mu, \sigma^2\right)$, distribusi dari rata-rata sampel adalah diberikan oleh

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Dapat dilihat bahwa distribusi dari estimator μ adalah tidak independen dari parameter μ. Jika kita standardisasi $\bar{X}$, maka kita dapatkan

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Distribusi dari standardisasi $\bar{X}$ adalah independen dari parameter μ. Standardisasi $\bar{X}$ adalah kuantitas pivot karena merupakan fungsi dari sampel $X_1, X_2, \ldots, X_n$ dan parameter μ, dan distribusi probabilitas ini adalah independen dari parameter μ. Dengan menggunakan kuantitas pivot, kita susun interval kepercayaan sebagai berikut :

$$1 - \alpha = P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}\right)$$

$$= P\left(\bar{X} - \left(\frac{\sigma}{\sqrt{n}}\right) z_{\frac{\alpha}{2}} \leq \mu \leq \bar{X} + \left(\frac{\sigma}{\sqrt{n}}\right) z_{\frac{\alpha}{2}}\right)$$

Karenanya, interval kepercayaan $100(1-\alpha)\%$ untuk μ ketika populasi X adalah normal dengan varians σ^2 diketahui adalah diberikan oleh

$$\left[\overline{X}-\left(\frac{\sigma}{\sqrt{n}}\right)z_{\frac{\alpha}{2}} \le \mu \le \overline{X}+\left(\frac{\sigma}{\sqrt{n}}\right)z_{\frac{\alpha}{2}}\right]$$

Ini dikatakan bahwa jika sampel berukuran n diambil dari populasi normal dengan mean μ dan varians σ^2 diketahui dan jika interval

$$\left[\overline{X}-\left(\frac{\sigma}{\sqrt{n}}\right)z_{\frac{\alpha}{2}} \le \mu \le \overline{X}+\left(\frac{\sigma}{\sqrt{n}}\right)z_{\frac{\alpha}{2}}\right]$$

Adalah disusun untuk setiap sampel, maka dalam panjang interval $100(1-\alpha)\%$ akan mencakup parameter yang tidak diketahui dan karenanya dengan kepercayaan $(1-\alpha)100\%$ kita dapat mengatakan bahwa μ terletak pada interval

$$\left[\overline{X}-\left(\frac{\sigma}{\sqrt{n}}\right)z_{\frac{\alpha}{2}},\ \overline{X}+\left(\frac{\sigma}{\sqrt{n}}\right)z_{\frac{\alpha}{2}}\right]$$

Estimasi interval μ ditemukan dengan mengambil estimator terbaik $\overline{X}$ dalam hal ini maksimum likelihod μ dan penjumlahan dan pengurangan $z_{\frac{\alpha}{2}}$ kali standar deviasi $\overline{X}$.

Catatan 7.2. Dengan menggunakan definisi interval kepercayaan $100(1-\alpha)\%$ untuk parameter θ adalah interval $[L,U]$ sedemikian sehingga probabilitas θ berada pada interval $[L,U]$ adalah $1-\alpha$. Artinya

$$1-\alpha = P(L \le \theta \le U)$$

Kita dapat menemukan dapat menemukan banyak pasangan L,U yang tak terhingga, sedemikian sehingga

$$1-\alpha = P(L \le \theta \le U)$$

Oleh karena itu, terdapat banyak interval kepercayaan tak terhingga untuk parameter yang diberikan. Namun, kita hanya memperhitungkan interval kepercayaan dengan panjang terpendek. Jika sebuah interval kepercayaan dibangun dengan menghilangkan area ekor yang sama maka kita peroleh yang disebut interval kepercayaan pusat. Dalam distribusi simetris, dapat ditunjukkan bahwa interval kepercayaan pusat memiliki panjang terpendek.

Contoh 7.2. Misalkan $X_1, X_2, \ldots, X_{11}$ menjadi sampel acak berukuran 11 dari distribusi normal dengan mean μ yang diketahui dan varians $\sigma^2 = 9.9$

Jika $\sum_{i=1}^{11} x_i = 132$, maka berapakah interval kepercayaan 95% untuk μ ?

Jawab :

Karena setiap $X_i \sim N(\mu, 9.9)$, interval kepercayaan untuk μ adalah diberikan dengan

$$\left[\bar{X} - \left(\frac{\sigma}{\sqrt{n}}\right) z_{\frac{\alpha}{2}},\ \bar{X} + \left(\frac{\sigma}{\sqrt{n}}\right) z_{\frac{\alpha}{2}}\right]$$

Karena $\sum_{i=1}^{11} x_i = 132$, rata-rata sampel $\bar{x} = \frac{132}{11} = 12$. Selanjutnya, kita lihat bahwa

$$\sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{9.9}{11}} = \sqrt{0.9}$$

Lebih lanjut, karena $1 - \alpha = 0.95, \alpha = 0.05$. Jadi

$$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96 \qquad (\text{dari tabel normal})$$

Dengan menggunakan keterangan pada pernyataan interval kepercayaan μ kita dapatkan

$$\left[12 - 1.96\sqrt{0.9},\ 12 + 1.96\sqrt{0.9}\right]$$

Sehingga

$$[10.141,\ 13.859]$$

Contoh 7.3. Misalkan $X_1, X_2, \ldots, X_{11}$ menjadi sampel acak berukuran 11 dari distribusi normal dengan mean μ yang diketahui dan varians $\sigma^2 = 9.9$

Jika $\sum_{i=1}^{11} x_i = 132$, maka berapakah nilai konstanta k

$$\left[12 - k\sqrt{0.9},\ 12 + k\sqrt{0.9}\right]$$

Interval kepercayaan 90% untuk μ ?

Jawab :

90% interval kepercayaan untuk μ ketika varians diketahui adalah

$$\left[\overline{x}-\left(\frac{\sigma}{\sqrt{n}}\right)z_{\frac{\alpha}{2}},\ \overline{x}+\left(\frac{\sigma}{\sqrt{n}}\right)z_{\frac{\alpha}{2}}\right]$$

Kemudian kita membutuhkan untuk mencari $\overline{x}$, $\sqrt{\frac{\sigma^2}{n}}$ dan $z_{\frac{\alpha}{2}}$ berkorespondensi dengan $1-\alpha=0.9$. Karenanya

$$\overline{x}=\frac{\sum_{i=1}^{11}x_i}{11}$$

$$=\frac{132}{11}$$

$$=12$$

$$\sqrt{\frac{\sigma^2}{n}}=\sqrt{\frac{9.9}{11}}$$

$$=\sqrt{0.9}$$

$$z_{0.05}=1.64 \qquad (\text{dari tabel normal})$$

Karenanya, interval kepercayaan untuk μ pada 90% tingkat kepercayaan adalah

$$\left[12-(1.96)\sqrt{0.9},\ 12+(1.96)\sqrt{0.9}\right]$$

Bandingkan interval ini dengan interval yang diberikan, kita mendapatkan

$$k=1.64$$

Dan korespondensi interval kepercayaan 90% adalah $[10.444,\ 13.556]$

Catatan 7.3. Perhatikan bahwa panjang dari 90% interval kepercayaan untuk μ adalah 3.112. Akan tetapi, panjang dari 95% interval kepercayaan adalah 3.718. Jadi semakin tinggi tingkat kepercayaan semakin besar panjang interval kepercayaan. Oleh karena itu, tingkat kepercayaan berbanding lurus dengan panjang interval kepercayaan. Mengingat fakta ini, kita melihat bahwa jika tingkat kepercayaan adalah nol, maka panjangnya juga nol. Yaitu ketika tingkat kepercayaan adalah nol, interval kepercayaan dari μ merosot menjadi titik $\overline{X}$.

Sampai sekarang kita telah memberikan kasus ketika populasi normal dengan mean μ yang tidak diketahui dan varians σ^2 yang diketahui. Sekarang kita berikan kasus bahwa ketika populasi tidak normal tetapi fungsi kepadatan probabilitasnya adalah kontinu, simetris, dan unimodal. Jika ukuran sampel besar, maka dengan teorema limit pusat

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \qquad n \to \infty$$

Jadi, pada kasus ini kita dapat mengambil kuantitas pivot untuk menjadi

$$Q(X_1, X_2, \ldots, X_n, \mu) = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

Jika ukuran sampel besar $(n \geq 32)$. Karena kuantitas pivot adalah sama seperti sebelumnya, kita dapatkan sampel pernyataan untuk interval kepercayaan $(1-\alpha)100\%$

$$\left[\overline{X} - \left(\frac{\sigma}{\sqrt{n}}\right) z_{\frac{\alpha}{2}},\ \overline{X} + \left(\frac{\sigma}{\sqrt{n}}\right) z_{\frac{\alpha}{2}}\right]$$

Contoh 7.4. Misalkan $X_1, X_2, \ldots, X_{40}$ menjadi sampel acak berukuran 40 dari distribusi dengan varians yang diketahui dan rata-rata μ tidak diketahui. Jika $\sum_{i=1}^{n} x_i = 286.56$ dan $\sigma^2 = 10$, maka berapakah 90% interval kepercayaan untuk populasi rata-rata μ ?

Jawab :

Karena $1-\alpha = 0.90$, kita dapatkan $\frac{\alpha}{2} = 0.05$. Karenanya, $z_{0.05} = 1.64$ (dari tabel normal standar). Selanjutnya, kita temukan rata-rata sampel

$$\overline{x} = \frac{286.56}{40} = 7.164$$

Oleh karena itu, interval kepercayaan untik μ adalah diberikan dengan

$$\left[7.164 - (1.64)\left(\sqrt{\frac{10}{40}}\right),\ 7.164 + (1.64)\left(\sqrt{\frac{10}{40}}\right)\right]$$

Yaitu

$$[6.344,\ 7.984]$$

Contoh 7.5. Dalam pengambilan sampel dari distribusi tidak normal dengan varians 25, berapa besar ukuran sampel sehingga panjang dari 95% interval kepercayaan untuk rata-rata adalah 1,96 ?

Jawab :

Interval kepercayaan saat sampel diambil dari populasi normal dengan varians 25 adalah

$$\left[\overline{x} - \left(\frac{\sigma}{\sqrt{n}}\right) z_{\frac{\alpha}{2}},\ \overline{x} + \left(\frac{\sigma}{\sqrt{n}}\right) z_{\frac{\alpha}{2}}\right]$$

Jadi panjang interval kepercayaannya adalah

$$\ell = 2 z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}$$

$$= 2 z_{0.025} \sqrt{\frac{25}{n}}$$

$$= 2(1.96)\sqrt{\frac{25}{n}}$$

Tetapi kita diberikan bahwa panjang interval kepercayaan adalah $\ell = 1.96$ Jadi

$$1.96 = 2(1.96)\sqrt{\frac{25}{n}}$$

$$\sqrt{n} = 10$$

$$n = 100$$

Oleh karena itu, ukuran sampel haruslah 100 sedemikian sehingga panjang dari 95% interval kepercayaan akan menjadi 1.96

Sejauh ini, kita telah membahas metode konstruksi interval kepercayaan untuk populasi rata-rata parameter ketika varians diketahui. Ini sangat tidak mungkin bahwa kita akan mengetahui varians tanpa mengetahui rata-rata populasi, dan dengan demikian apa yang telah kita bahas sejauh ini di bagian ini tidak terlalu realistis. Sekarang kita memperlakukan kasus membangun interval kepercayaan untuk rata-rata populasi ketika varians populasi juga tidak diketahui. Pertama-tama, kita mulai dengan membangun interval kepercayaan dengan asumsi populasi X adalah normal.

Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari populasi normal X dengan mean μ dan varians $\sigma^2 > 0$. Misalkan rata-rata sampel dan sampel varians masing-masing menjadi $\overline{X}$ dan S^2. Maka

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Dan

$$\frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

Oleh karena itu, variabel acak didefinisikan sebagai rasio dari $\frac{(n-1)S^2}{\sigma^2}$ terhadap $\frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$ memiliki distribusi-t dengan derajat kebebasan $(n-1)$,

yaitu

$$Q(X_1, X_2, \ldots, X_n, \mu) = \frac{\dfrac{\bar{X}-\mu}{\sqrt{\dfrac{\sigma^2}{n}}}}{\sqrt{\dfrac{(n-1)S^2}{(n-1)\sigma^2}}} = \frac{\bar{X}-\mu}{\sqrt{\dfrac{S^2}{n}}} \sim t(n-1)$$

Dimana Q adalah kuantitas pivot yang akan digunakan untuk membangun interval kepercayaan untuk μ. Dengan menggunakan kuantitas pivot ini, kita membangun interval kepercayaan sebagai berikut :

$$1-\alpha = P\left(-t_{\frac{\alpha}{2}}(n-1) \leq \frac{\bar{X}-\mu}{\dfrac{S}{\sqrt{n}}} \leq t_{\frac{\alpha}{2}}(n-1) \right)$$

$$= P\left(\bar{X} - \left(\frac{S}{\sqrt{n}}\right) t_{\frac{\alpha}{2}}(n-1) \leq \mu \leq \bar{X} + \left(\frac{S}{\sqrt{n}}\right) t_{\frac{\alpha}{2}}(n-1) \right)$$

Oleh karena itu, interval kepercayaan untuk μ ketika populasi X adalah normal dengan varians σ^2 tidak diketahui adalah diberikan oleh

$$\left[\bar{X} - \left(\frac{S}{\sqrt{n}}\right) t_{\frac{\alpha}{2}}(n-1),\ \bar{X} + \left(\frac{S}{\sqrt{n}}\right) t_{\frac{\alpha}{2}}(n-1) \right]$$

Contoh 7.6. Sembilan pengamata sampel acak dari populasi normal pengamatan staistik menghasilkan $\bar{x} = 5$ dan $\frac{1}{8}\sum_{i=1}^{9}(x_i - \bar{x})^2 = 36$. Berapakah 95% interval kepercayaan untuk μ ?

Jawab :

Karena

$$n = 9$$

$$\bar{x} = 5$$

$$s^2 = 36$$

$$1-\alpha = 0.95$$

95% interval kepercayaan untuk μ adalah diberikan oleh

$$\left[\bar{x}-\left(\frac{s}{\sqrt{n}}\right)t_{\frac{\alpha}{2}}(n-1),\ \bar{x}+\left(\frac{s}{\sqrt{n}}\right)t_{\frac{\alpha}{2}}(n-1)\right]$$

Yaitu

$$\left[5-\left(\frac{6}{\sqrt{9}}\right)t_{0.025}(8),\ 5+\left(\frac{6}{\sqrt{9}}\right)t_{0.025}(8)\right]$$

Yang mana

$$\left[5-\left(\frac{6}{\sqrt{9}}\right)(2.306),\ 5+\left(\frac{6}{\sqrt{9}}\right)(2.306)\right]$$

Oleh karena itu, 95% interval kepercayaan untuk μ adalah diberikan oleh $[0.388,\ 9.612]$

Contoh 7.7. Manakah berikut ini yang benar dari interval kepercayaan 95% untuk rata-rata populasi?

a) Interval mencakup 95% dari nilai populasi rata-rata

b) Interval mencakup 95% dari nilai sampel rata-rata

c) Interval memiliki peluang 95% termasuk mean sampel

Jawab :

Tidak ada pernyataan yang benar karena interval kepercayaan 95% untuk rata-rata populasi μ berarti interval tersebut memiliki 95% peluang termasuk rata-rata populasi μ.

Akhirnya, kita berikan kasus ketika populasi tidak normal tetapi fungsi kepadatan probabilitas kontinu, simetris dan unimodal. Jika beberapa kondisi rendah terpenuhi, maka varians sampel S^2 dari sampel acak berukuran $n \geq 2$, konvergen secara stokastik untuk σ^2. Oleh karena itu, pada

$$\frac{\dfrac{\overline{X}-\mu}{\sqrt{\dfrac{\sigma^2}{n}}}}{\sqrt{\dfrac{(n-1)S^2}{(n-1)\sigma^2}}}=\frac{\overline{X}-\mu}{\sqrt{\dfrac{S^2}{n}}}$$

Pembilang bagian kiri konvergen terhadap $N(0,1)$ dan penyebut bagian itu konvergen terhadap 1. Karenanya

$$\frac{\overline{X}-\mu}{\sqrt{\dfrac{S^2}{n}}}\sim N(0,1) \qquad n\to\infty$$

Fakta ini dapat digunakan membangun interval kepercayaan untuk rata-rata populasi ketika varians tidak diketahui dan distribusi populasi adalah tidak normal. Kami berikan kuantitas pivot menjadi

$$Q(X_1,X_2,\ldots,X_n,\mu)=\frac{\overline{X}-\mu}{\sqrt{\dfrac{S^2}{n}}}$$

Dan didapatkan interval kepercayaan sebagai berikut

$$\left[\overline{X}-\left(\frac{S}{\sqrt{n}}\right)z_{\frac{\alpha}{2}},\ \overline{X}+\left(\frac{S}{\sqrt{n}}\right)z_{\frac{\alpha}{2}}\right]$$

Kita merangkum hasil bagian ini pada tabel berikut.

Populasi	Varians σ^2	Ukuran Sampel n	Limit kepercayaan
Normal	Diketahui	$n\geq 2$	$\overline{x}\mp z_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}$
Normal	Tidak diketahui	$n\geq 2$	$\overline{x}\mp t_{\frac{\alpha}{2}}(n-1)\dfrac{s}{\sqrt{n}}$
Tidak normal	Diketahui	$n\geq 32$	$\overline{x}\mp z_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}$

Tidak normal	Diketahui	$n<32$	Tidak ada rumus yang ada
Tidak normal	Tidak diketahui	$n \geq 32$	$\bar{x} \mp t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}}$
Tidak normal	Tidak diketahui	$n<32$	Tidak ada rumus yang ada

7.4. Interval Kepercayaan untuk Populasi Varians

Pada bagian ini, pertama-tama kita akan menjelaskan terkait dengan metode untuk menyusun interval kepercayaan varians saat populasinya normal dengan populasi mean μ diketahui. Sehingga kita mengerjakan kasus ketika populasi mean tidak diketahui.

Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak dari populasi normal X dengan mean μ diketahui dan varians σ^2 tidak diketahui. Kita ingin menyusun interval kepercayaan $100(1-\alpha)\%$ untuk varians σ^2, yaitu kita akan menemukan penaksir dari L dan U sedemikian sehingga

$$P\left(L \leq \sigma^2 \leq U\right) = 1-\alpha$$

Untuk menemukan penaksir dari L dan U, pertama-tama kita susun kuantitas pivot. Sehingga

$$X_i \sim N\left(\mu, \sigma^2\right)$$

$$\left(\frac{X_i-\mu}{\sigma}\right) \sim N(0,1)$$

$$\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(1)$$

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$$

Kita definisikan kuantitas pivot $Q\left(X_1, X_2, \ldots, X_n, \sigma^2\right)$ sebagai

$$Q\left(X_1, X_2, \ldots, X_n, \sigma^2\right) = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2$$

Yang memiliki distribusi chi-square dengan derajat kebebasan n.

Karenanya

$$1-\alpha = P\left(a \le Q \le b\right)$$

$$= P\left(a \le \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \le b\right)$$

$$= P\left(\frac{1}{a} \ge \sum_{i=1}^{n}\frac{\sigma^2}{\left(X_i - \mu\right)^2} \ge \frac{1}{b}\right)$$

$$= P\left(\frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{a} \ge \sigma^2 \ge \frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{b}\right)$$

$$= P\left(\frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{b} \le \sigma^2 \le \frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{a}\right)$$

$$= P\left(\frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)} \le \sigma^2 \le \frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{\chi^2_{\frac{\alpha}{2}}(n)}\right)$$

Oleh karena itu, interval kepercayaan $100\left(1-\alpha\right)\%$ untuk σ^2 ketika mean diketahui adalah diberikan oleh

$$\left[\frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)}, \frac{\sum_{i=1}^{n}\left(X_i - \mu\right)^2}{\chi^2_{\frac{\alpha}{2}}(n)}\right]$$

Contoh 7.8. Sampel acak dari 9 observasi dari populasi normal dengan $\mu = 5$ menghasilkan statistik yang diamati $\frac{1}{8}\sum_{i=1}^{9} x_i^2 = 39.125$ dan $\sum_{i=1}^{9} x_i = 45$. Berapakah interval kepercayaan 95% untuk σ^2

Jawab :

Kita telah diberikan bahwa

$$n = 9 \qquad dan \qquad \mu = 5$$

Lebih lanjut kita tahu bahwa

$$\sum_{i=1}^{9} x_i = 45 \qquad dan \qquad \frac{1}{8}\sum_{i=1}^{9} x_i^2 = 39.125$$

Karenanya

$$\sum_{i=1}^{9} x_i^2 = 313$$

Dan

$$\sum_{i=1}^{9} (x_i - \mu)^2 = \sum_{i=1}^{9} x_i^2 - 2\mu \sum_{i=1}^{9} x_i - 9\mu^2$$

$$= 313 - 450 + 225$$

$$= 88$$

Karena $1 - \alpha = 0.95$, kita dapatkan $\frac{\alpha}{2} = 0.025$ dan $1 - \frac{\alpha}{2} = 0.975$. dengan menggunakan tabel chi-square kita mempunyai

$$\chi^2_{0.025}(9) = 2.700 \qquad dan \qquad \chi^2_{0.975}(9) = 19.02$$

Karenanya, 95% interval kepercayaan untuk σ^2 adalah diberikan oleh

$$\left[\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)}, \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)} \right]$$

Yaitu

$$\left[\frac{88}{19.02}, \frac{88}{2.7}\right]$$

Sehingga

$$\left[4.583,\ 32.59\right]$$

Catatan 7.4. Karena distribusi χ^2 tidaklah simetris, interval kepercayaan diatas seharusnya tidaklah terpendek. Nanti, di bagian selanjutnya, kita menggambarkan bagaimana kita membangun interval kepercayaan dengan panjang terpendek.

Diberikan sampel acak $X_1, X_2, \ldots, X_n$ dari populasi normal $X \sim N\left(\mu, \sigma^2\right)$ dimana populasi mean μ dan populasi varians σ^2 adalah tidak diketahui. Kita akan membangun interval kepercayaan $100\left(1-\alpha\right)\%$ untuk populasi varians. Kita tahu bahwa

$$\frac{\left(n-1\right)S^2}{\sigma^2} \sim \chi^2\left(n-1\right)$$

$$\Rightarrow \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\sigma^2} \sim \chi^2\left(n-1\right)$$

Kita mengambil $\dfrac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\sigma^2}$ sebagai kuantitas pivot Q untuk membangun interval kepercayaan untuk σ^2. Karenanya kita dapatkan

$$1-\alpha = P\left(\frac{1}{\chi^2_{\frac{\alpha}{2}}\left(n-1\right)} \leq Q \leq \frac{1}{\chi^2_{1-\frac{\alpha}{2}}\left(n-1\right)}\right)$$

$$= P\left(\frac{1}{\chi^2_{\frac{\alpha}{2}}\left(n-1\right)} \leq \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\sigma^2} \leq \frac{1}{\chi^2_{1-\frac{\alpha}{2}}\left(n-1\right)}\right)$$

$$= P\left(\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \le \sigma^2 \le \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \right)$$

Karenanya, interval kepercayaan $100(1-\alpha)\%$ untuk varians σ^2 ketika populasi mean adalah diketahui diberikan oleh

$$\left[\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \right]$$

Contoh 7.9. Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak berukuran 13 dari distribusi Normal $N\left(\mu, \sigma^2\right)$. Jika $\sum_{i=1}^{13} x_i = 246.61$ dan $\sum_{i=1}^{13} = 4806.61$. Temukan 90% interval kepercayaan untuk σ^2

Jawab :

$$\overline{x} = 18.97$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{13}\left(x_i - \overline{x}\right)^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{13}\left(x_i^2 - n\overline{x}^2\right)^2$$

$$= \frac{1}{12}\left[4806.61 - 4678.2\right]$$

$$= \frac{1}{12}128.41$$

Karenanya, $12s^2 = 128.41$. lebih lanjut, karena $1-\alpha = 0.90$, kita dapatkan $\frac{\alpha}{2} = 0.05$ dan $1-\frac{\alpha}{2} = 0.95$. Oleh karena itu, dari tabel chi-square kita dapatkan

$$\chi^2_{0.95}(12) = 21.03, \ \chi^2_{0.05} = 5.23$$

Karenanya, 95% interval kepercayaan untuk σ^2 adalah

$$\left[\frac{128.41}{5.23}, \frac{128.41}{21.03}\right]$$

Sehingga

$$[6.11, 24.57]$$

Contoh 7.10. Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak berukuran n dari distribusi $N(\mu, \sigma^2)$, dimana μ dan σ^2 adalah parameter yang tidak diketahui. Berapakah 90% interval kepercayaan terpendek untuk standar deviasi σ?

Jawab :

Misalkan S^2 menjadi sampel varians, sehingga

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Dengan menggunakan variabel acak sebagai pivot, kita dapat menyusun $100(1-\alpha)\%$ untuk σ dari

$$1-\alpha = P\left(a \le \frac{(n-1)S^2}{\sigma^2} \le b\right)$$

Dengan memilih konstanta a dan b yang sesuai. Karenanya, interval kepercayaan untuk σ adalah diberikan oleh

$$\left[\sqrt{\frac{(n-1)S^2}{b}}, \sqrt{\frac{(n-1)S^2}{a}}\right]$$

Panjang dari interval kepercayaan ini adalah diberikan oleh

$$L(a,b) = S\sqrt{n-1}\left[\frac{1}{\sqrt{a}} - \frac{1}{\sqrt{b}}\right]$$

Untuk mencari interval kepercayaan terpendek, kita harus mencari pasangan dari konstanta a dan b sedemikian sehingga $L(a,b)$ adalah minimum. Jadi, kita memiliki kendala masalah minimalisasi. Yaitu

$$\left.\begin{array}{c} Minimize\ L(a,b) \\ \\ Kondisi\,Subyek \\ \\ \int_a^b f(u)\,du = 1-\alpha \end{array}\right\} \qquad (MP)$$

Dimana

$$f(x) = \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}} x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}}$$

Turunkan L terhadap a, kita dapatkan

$$\frac{dL}{da} = S\sqrt{n-1}\left(-\frac{1}{2}a^{-\frac{3}{2}} + \frac{1}{2}b^{-\frac{3}{2}}\frac{db}{da}\right)$$

Dari

$$\int_a^b f(u)\,du = 1-\alpha$$

Kita temukan turunan dari b terhadap a sebagai berikut

$$\frac{d}{da}\int_a^b f(u)\,du = \frac{d}{da}(1-\alpha)$$

Yaitu

$$f(b)\frac{db}{da} - f(a) = 0$$

Karenanya, kita mempunyai

$$\frac{db}{da} = \frac{f(a)}{f(b)}$$

Misalkan ini kedalam penyataan untuk menurunkan L, kita dapatkan

$$\frac{dL}{da} = S\sqrt{n-1}\left(-\frac{1}{2}a^{-\frac{3}{2}} + \frac{1}{2}b^{-\frac{3}{2}}\frac{f(a)}{f(b)}\right)$$

Samakan turunan ini dengan nol, kita dapatkan

$$S\sqrt{n-1}\left(-\frac{1}{2}a^{-\frac{3}{2}} + \frac{1}{2}b^{-\frac{3}{2}}\frac{f(a)}{f(b)}\right) = 0$$

Yang menghasilkan

$$a^{\frac{3}{2}}f(a) = b^{\frac{3}{2}}f(b)$$

Dengan menggunakan pernyataan dari f, kita dapatkan dari pernyataan diatas

$$a^{\frac{3}{2}}a^{\frac{n-3}{2}}e^{-\frac{a}{2}} = b^{\frac{3}{2}}b^{\frac{n-3}{2}}e^{-\frac{b}{2}}$$

Maka dari itu kita dapatkan

$$\ln\left(\frac{a}{b}\right) = \left(\frac{a-b}{n}\right)$$

Karenanya untuk memperoleh pasangan dari konstanta a dan b itu akan menghasilkan interval kepercayaan terpendek untuk σ, kita harus menyelesaikan sistem persamaan nonlinear berikut

$$\left.\begin{aligned} &\int_a^b f(u)\,du = 1-\alpha \\ &\ln\left(\frac{a}{b}\right) = \left(\frac{a-b}{n}\right) \end{aligned}\right\} \qquad (\star)$$

Jika a_0 dan b_0 adalah solusi dari $(\star)$, maka interval kepercayaan terpendek untuk σ adalah diberikan oleh

$$\left[\sqrt{\frac{(n-1)S^2}{b_0}}, \sqrt{\frac{(n-1)S^2}{a_0}}\right]$$

Karena sistem persamaan nonlinier ini sulit dipecahkan secara analitik, solusi numerik diberikan dalam literatur statistik dalam bentuk tabel untuk menemukan interval terpendek untuk varians

7.5. Interval Kepercayaan Parameter dari Beberapa Distribusi bukan milik Keluarga Skala Lokasi

Pada bagian ini, kita ilustrasikan metode kuantitas pivot untuk menemukan interval kepercayaan untuk parameter θ saat fungsi densitas bukan milik keluarga skala lokasi. Berikut ini fungsi densitas bukan milik keluarga skala lokasi :

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & jika\, 0 < x < 1 \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Atau

$$f(x;\theta) = \begin{cases} \dfrac{1}{\theta}, & jika\, 0 < x < 1 \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Kita akan mengkonstruksi penaksir interval parameter pada fungsi densitas ini. Ide yang sama untuk menemukan penaksir interval dapat digunakan untuk menemukan penaksir interval parameter dari fungsi densitas yang termasuk dalam keluarga skala lokasi seperti

$$f(x;\theta) = \begin{cases} \dfrac{1}{\theta} e^{-\frac{x}{\theta}}, & jika\, 0 < x < \infty \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Untuk menemukan kuantitas pivot dari yang disebutkan distribusi diatas dan lainnya kita memerlukan tiga hasil berikut. Hasil pertama adalah teorema 6.2 sedangkan pembuktian hasil yang kedua mudah dan kita serahkan kepada pembaca

Teorema 7.1.

Misalkan $F(X;\theta)$ menjadi fungsi distribusi kumulatif dari variabel acak kontinu X. Maka

$$F(X;\theta) \sim UNIF(0,1)$$

Teorema 7.2.

Jika $X \sim UNIF(0,1)$, maka

$$-\ln X \sim EXP(1)$$

Teorema 7.3.

Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak dari distribusi dengan fungsi densitas

$$f(x;\theta) = \begin{cases} \dfrac{1}{\theta} e^{-\frac{x}{\theta}}, & jika\, 0 < x < \infty \\ \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Dimana $\theta > 0$ adalah parameter. Sehingga variabel acak

$$\frac{2}{\theta}\sum_{i=1}^{n} X_i \sim \chi^2(2n)$$

Bukti :

Misalkan $Y = \frac{2}{\theta}\sum_{i=1}^{n} X_i$. Sekarang kita tunjukkan bahwa distribusi sampling dari Y adalah chi-square dengan derajat kebebasan $2n$. Kita gunakan metode pembangkit momen untuk menunjukkan ini. Fungsi pembangkit momen Y adalah diberikan oleh

$$M_Y(t) = M_{\frac{2}{\theta}\sum_{i=1}^{n} X_i}(t)$$

$$= \prod_{i=1}^{n} MX_i\left(\frac{2}{\theta}t\right)$$

$$=\prod_{i=1}^{n}\left(1-\theta\frac{2}{\theta}t\right)^{-1}$$

$$=(1-2t)^{-n}$$

$$=(1-2t)^{-\frac{2n}{2}}$$

Karena $(1-2t)^{-\frac{2n}{2}}$ berkorespondensi dengan fungsi pembangkit momen dari variabel acak chi-square dengan derajat kebebasan $2n$, kita simpulkan bahwa

$$\frac{2}{\theta}\sum_{i=1}^{n}X_i \sim \chi^2(2n)$$

Teorema 7.4.

Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak dari distribusi dengan fungsi densitas

$$f(x;\theta)=\begin{cases}\theta x^{\theta-1}, & jika\, 0\le x\le 1\\ 0, & untuk\ yang\ lainnya\end{cases}$$

Dimana $\theta>0$ adalah parameter. Sehingga variabel acak $-2\theta\sum_{i=1}^{n}\ln X_i$ memiliki distribusi chi-square dengan derajat kebebasan $2n$.

Bukti :

Kita diberikan bahwa

$$X_i \sim \theta x^{\theta-1},\ 0<x<1$$

Karenanya, fungsi distribusi kumulatif f adalah

$$F(x;\theta)=\int_0^x \theta x^{\theta-1}\,dx = x^{\theta}$$

Jadi dengan Teorema 7.1,

$$F(X_i;\theta) \sim UNIF(0,1)$$

Itu adalah

$$X_i^0 \sim UNIF(0,1)$$

Dengan menggunakan teorema 7.2

$$-\ln X_i^0 \sim EXP(1)$$

Itu adalah

$$-\theta \ln X_i \sim EXP(1)$$

Dengan menggunakan teorema 7.3 (dengan $\theta = 1$), kita peroleh

$$-2\theta \sum_{i=1}^{n} \ln X_i \sim \chi^2(2n)$$

Karenanya, distribusi sampling dari $-2\theta \sum_{i=1}^{n} \ln X_i$ adalah chi-square dengan derajat kebebasan $2n$

Teorema berikut yang pembuktiannya mengikuti dari Teorema 7.1, 7.2 dan 7.3 adalah kunci untuk menemukan kuantitas pivot dari banyak distribusi yang bukan milik keluarga skala lokasi. Selanjutnya, teorema ini juga dapat digunakan untuk menemukan kuantitas pivot untuk parameter dari beberapa distribusi milik keluarga skala lokasi.

Teorema 7.5.

Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak daro populasi kontinu X dengan fungsi distribusi $F(x;\theta)$. Jika $F(x;\theta)$ adalah monoton di θ, maka statistik $Q = -2\sum_{i=1}^{n} \ln F(X_i;\theta)$ adalah kuantitas pivot dan memiliki distribusi chi-square dengan derajat kebebasan $2n$ $\left(Q \sim \chi^2(2n)\right)$

Perlu dicatat bahwa kondisi $F(X_i;\theta)$ adalah monoton di θ diperlukan untuk memastikan interval. Kalau tidak, kita mungkin mendapatkan area kepercayaan daripada interval kepercayaan. Perhatikan juga bahwa jika

$$-2\sum_{i=1}^{n} \ln F(X_i;\theta) \sim \chi^2(2n),$$

maka

$$-2\sum_{i=1}^{n}\ln\left(1-F\left(X_i;\theta\right)\right)\sim\chi^2\left(2n\right)$$

Contoh 7.11. Jika $X_1, X_2, \ldots, X_n$ adalah sampel acak dari populasi dengan densitas

$$f\left(x;\theta\right)=\begin{cases}\theta x^{\theta-1}, & jika\,0<x<1\\ \\ 0, & untuk\ yang\ lainnya\end{cases}$$

Dimana $\theta>0$ adalah parameter yang tidak diketahui, berapakah interval kepercayaan $100(1-\alpha)\%$ untuk θ?

Jawab :

Untuk membangun interval kepercayaan untuk θ, kita memerlukan kuantitas pivot. Maka dari itu, kita memerlukan variabel acak yang merupakan fungsi dari sampel dan parameter, dan yang distribusi probabilitasnya adalah diketahui tetapi tidak melibatkan θ. Kita gunakan variabel acak

$$Q=-2\theta\sum_{i=1}^{n}\ln X_i\sim\chi^2\left(2n\right)$$

Sebagai kuantitas pivot. Interval kepercayaan $100(1-\alpha)\%$ untuk θ dapat disusun dari

$$1-\alpha=P\left(\chi^2_{\frac{\alpha}{2}}\left(2n\right)\le Q\le\chi^2_{1-\frac{\alpha}{2}}\left(2n\right)\right)$$

$$=P\left(\chi^2_{\frac{\alpha}{2}}\left(2n\right)\le-2\theta\sum_{i=1}^{n}\ln X_i\le\chi^2_{1-\frac{\alpha}{2}}\left(2n\right)\right)$$

$$=P\left(\frac{\chi^2_{\frac{\alpha}{2}}\left(2n\right)}{-2\theta\sum_{i=1}^{n}\ln X_i}\le\theta\le\frac{\chi^2_{1-\frac{\alpha}{2}}\left(2n\right)}{-2\theta\sum_{i=1}^{n}\ln X_i}\right)$$

Karenanya, interval kepercayaan $100(1-\alpha)\%$ untuk θ adalah diberikan oleh

$$\left[\frac{\chi^2_{\frac{\alpha}{2}}(2n)}{-2\theta\sum_{i=1}^{n}\ln X_i}, \frac{\chi^2_{1-\frac{\alpha}{2}}(2n)}{-2\theta\sum_{i=1}^{n}\ln X_i}\right]$$

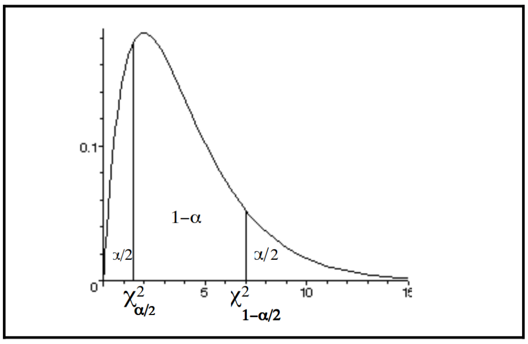
Disini $\chi^2_{1-\frac{\alpha}{2}}(2n)$ menunjukkan $\left(1-\frac{\alpha}{2}\right)$ kuantil dari variabel acak chi-square Y, maka dari itu

$$P\left(Y \le \chi^2_{1-\frac{\alpha}{2}}(2n)\right) = 1-\frac{\alpha}{2}$$

Dan $\chi^2_{\frac{\alpha}{2}}(2n)$ demikian juga menunjukkan $\frac{\alpha}{2}$ kuantil dari Y, sehingga

$$P\left(Y \le \chi^2_{\frac{\alpha}{2}}(2n)\right) = \frac{\alpha}{2}$$

Untuk $\alpha \le 0.5$ (lihat grafik berikut)



■ **Contoh 7.12.** Jika $X_1, X_2, \ldots, X_n$ adalah sampel acak dari distribusi dengan fungsi densitas

$$f(x;\theta) = \begin{cases} \frac{1}{\theta}, & jika\, 0 < x < \theta \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Dimana $\theta > 0$ adalah parameter, maka berapakah interval kepercayaan $100(1-\alpha)\%$ untuk θ?

Jawab :

Fungsi distribusi kumulatif dari $f(x;\theta)$ adalah

$$f(x;\theta)=\begin{cases}\dfrac{x}{\theta}, & jika\, 0<x<\theta \\ \\ 0, & untuk\ yang\ lainnya\end{cases}$$

Karena

$$-2\sum_{i=1}^{n}\ln F(X_i;\theta)=-2\sum_{i=1}^{n}\ln\left(\frac{X_i}{\theta}\right)$$

$$=2n\ln\theta-2\sum_{i=1}^{n}\ln(X_i)$$

Dengan menggunakan teorema 7.5, kuantitas $2n\ln\theta-2\sum_{i=1}^{n}\ln(X_i)\sim\chi^2(2n)$ Karena $2n\ln\theta-2\sum_{i=1}^{n}\ln(X_i)$ adalah fungsi dari sampel dan parameter, dan distribusi ini adalah independen dari θ, itu adalah pivot untuk θ. Karenanya, kita mengambil

$$Q(X_1,X_2,\ldots,X_n,\theta)=2n\ln\theta-2\sum_{i=1}^{n}\ln(X_i)$$

Interval kepercayaan $100(1-\alpha)\%$ untuk θ dapat disusun dari

$$1-\alpha=P\left(\chi^2_{\frac{\alpha}{2}}(2n)\le Q\le\chi^2_{1-\frac{\alpha}{2}}(2n)\right)$$

$$=P\left(\chi^2_{\frac{\alpha}{2}}(2n)\le 2n\ln\theta-2\sum_{i=1}^{n}\ln X_i\le\chi^2_{1-\frac{\alpha}{2}}(2n)\right)$$

$$=P\left(\chi^2_{\frac{\alpha}{2}}(2n)+2\sum_{i=1}^{n}\ln X_i\le 2n\ln\theta\le\chi^2_{1-\frac{\alpha}{2}}(2n)+2\sum_{i=1}^{n}\ln X_i\right)$$

$$= P\left(e^{\frac{1}{2n}\left\{\chi^2_{\frac{\alpha}{2}}(2n)+2\sum_{i=1}^{n}\ln X_i\right\}} \le \theta \le e^{\frac{1}{2n}\left\{\chi^2_{1-\frac{\alpha}{2}}(2n)+2\sum_{i=1}^{n}\ln X_i\right\}} \right)$$

Sehingga, interval kepercayaan $100(1-\alpha)\%$ untuk θ adalah diberikan oleh

$$\left[e^{\frac{1}{2n}\left\{\chi^2_{\frac{\alpha}{2}}(2n)+2\sum_{i=1}^{n}\ln X_i\right\}},\ e^{\frac{1}{2n}\left\{\chi^2_{1-\frac{\alpha}{2}}(2n)+2\sum_{i=1}^{n}\ln X_i\right\}} \right]$$

Fungsi densitas dari contoh berikut termasuk dalam keluarga skala. Akan tetapi, kita dapat menggunakan teorema 7.5 unutk menemukan pivot dari parameter dan menentukan penaksir interval dari parameter.

Contoh 7.13. Jika $X_1, X_2, \ldots, X_n$ adalah variabel acak dari distribusi dengan fungsi densitas

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & jika\, 0 < x < \infty \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Dimana $\theta > 0$ adalah parameter, maka berapakah interval kepercayaan $100(1-\alpha)\%$ untuk θ?

Jawab :

Fungsi distribusi kumulatif $F(x;\theta)$ dari fungsi densitas

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & jika\, 0 < x < \infty \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Adalah diberikan oleh

$$F(x;\theta) = 1 - e^{-\frac{x}{\theta}}$$

Karenanya

$$-2\sum_{i=1}^{n}\ln\left(1-F\left(X_i;\theta\right)\right)=\frac{2}{\theta}\sum_{i=1}^{n}X_i$$

Sehingga

$$\frac{2}{\theta}\sum_{i=1}^{n}X_i \sim \chi^2\left(2n\right)$$

Kita mengambil $Q=\frac{2}{\theta}\sum_{i=1}^{n}X_i$ sebagai kuantitas pivot. Interval kepercayaan $100\left(1-\alpha\right)\%$ untuk θ dapat disusun dengan

$$1-\alpha = P\left(\chi^2_{\frac{\alpha}{2}}\left(2n\right)\le Q \le \chi^2_{1-\frac{\alpha}{2}}\left(2n\right)\right)$$

$$= P\left(\chi^2_{\frac{\alpha}{2}}\left(2n\right)\le \frac{2}{\theta}\sum_{i=1}^{n}X_i \le \chi^2_{1-\frac{\alpha}{2}}\left(2n\right)\right)$$

$$= P\left(\frac{2\sum_{i=1}^{n}X_i}{\chi^2_{1-\frac{\alpha}{2}}\left(2n\right)}\le \theta \le \frac{2\sum_{i=1}^{n}X_i}{\chi^2_{\frac{\alpha}{2}}\left(2n\right)}\right)$$

Karenanya, interval kepercayaan $100\left(1-\alpha\right)\%$ untuk θ adalah diberikan oleh

$$\left[\frac{2\sum_{i=1}^{n}X_i}{\chi^2_{1-\frac{\alpha}{2}}\left(2n\right)},\frac{2\sum_{i=1}^{n}X_i}{\chi^2_{\frac{\alpha}{2}}\left(2n\right)}\right]$$

Pada bagian ini, kita dapat melihat bahwa interval kepercayaan $100\left(1-\alpha\right)\%$ untuk parameter θ dapat disusun dengan mengambil kuantitas pivot Q menjadi salah satu dari

$$Q=-2\sum_{i=1}^{n}\ln F\left(X_i;\theta\right)$$

Atau

$$Q = -2\sum_{i=1}^{n} \ln\left(1 - F\left(X_i;\theta\right)\right)$$

Dalam salah satu kasus, distribusi dari Q adalah chi-squared dengan derajat kebebeasan $2n$, sehingga $Q \sim \chi^2\left(2n\right)$. Karena distribusi chi-squared adalah tidak simetris terhadap sumbu y, interval kepercayaan yang dibangun di bagian ini tidak memiliki panjang terpendek. Untuk memiliki interval kepercayaan terpendek, kita harus menyelesaikan masalah minimasi berikut :

$$\left.\begin{array}{r} Minimize\ \ L\left(a,b\right) \\ \\ Keadaan\,Subyek\ \int_a^b f\left(u\right)du = 1-\alpha \end{array}\right\} \quad \left(MP\right)$$

Dimana

$$f\left(x\right) = \frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{\frac{n-1}{2}}} x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}}$$

Pada kasus contoh 7.13, proses minimalisasi mengarah ke sistem persamaan nonlinier berikut :

$$\left.\begin{array}{l} \int_a^b f\left(u\right)du = 1-\alpha \\ \\ \ln\left(\frac{a}{b}\right) = \frac{a-b}{2\left(n+1\right)} \end{array}\right\} \quad \left(NE\right)$$

Jika a_0 dan b_0 adalah solusi dari $\left(NE\right)$, maka interval kepercayaan terpendek untuk θ adalah diberikan oleh

$$\left[\frac{2\sum_{i=1}^{n} X_i}{b_0}, \frac{2\sum_{i=1}^{n} X_i}{a_0}\right]$$

7.6. Aproksimasi Interval Kepercayaan Parameter dengan MLE

Pada bagian ini, kita akan membahas terkait dengan bagaimana cara membangun aproksimasi interval kepercayaan $100(1-\alpha)\%$ dari populasi parameter θ dengan menggunakan estimator likelihood maksimum $\hat{\theta}$. Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak dari populasi X dengan densitas $f(x;\theta)$. Misalkan $\hat{\theta}$ menjadi estimator likelihood maksimum dari θ. Jika ukuran sampel n adalah besar, maka dengan menggunakan sifat asimtotik estimator likelihood maksimum, kita dapatkan

$$\frac{\hat{\theta}-E(\hat{\theta})}{\sqrt{Var(\hat{\theta})}} \sim N(0,1) \qquad n \to \infty$$

Dimana $Var(\hat{\theta})$ menunjukan varians dari estimator θ. Karena untuk n yang besar, estimator likelihood maksimum θ adalah tidak bias, kita dapatkan

$$\frac{\hat{\theta}-\theta}{\sqrt{Var(\hat{\theta})}} \sim N(0,1) \qquad n \to \infty$$

Varians $Var(\hat{\theta})$ dapat diperhitungkan secara langsung apabila memungkinkan atau dengan menggunakan CRLB *(Cramer-Rao Lower Bound)*

$$Var(\hat{\theta}) \geq \frac{-1}{E\left[\frac{d^2 \ln L(\theta)}{d\theta^2}\right]}$$

Sekarang dengan menggunakan $Q = \dfrac{\hat{\theta}-\theta}{\sqrt{Var(\hat{\theta})}}$ sebagai kuantitas pivot, kita

rancang aproksimasi interval kepercayaan $100(1-\alpha)\%$ dari θ sebagai

$$1-\alpha = P\left(-z_{\frac{\alpha}{2}} \leq Q \leq z_{\frac{\alpha}{2}}\right)$$

$$= P\left(-z_{\frac{\alpha}{2}} \leq \frac{\hat{\theta}-\theta}{\sqrt{Var(\hat{\theta})}} \leq z_{\frac{\alpha}{2}}\right)$$

Jika $Var(\hat{\theta})$ adalah bebas dari θ, maka didapatkan

$$1-\alpha = P\left(\hat{\theta} - z_{\frac{\alpha}{2}}\sqrt{Var\left(\hat{\theta}\right)} \le \theta \le \hat{\theta} + z_{\frac{\alpha}{2}}\sqrt{Var\left(\hat{\theta}\right)}\right)$$

Kemudian aproksimasi interval kepercayaan $100(1-\alpha)\%$ dari θ adalah

$$\left[\hat{\theta} - z_{\frac{\alpha}{2}}\sqrt{Var\left(\hat{\theta}\right)}\ ,\ \hat{\theta} + z_{\frac{\alpha}{2}}\sqrt{Var\left(\hat{\theta}\right)}\right]$$

Asalkan $Var\left(\hat{\theta}\right)$ adalah bebas terhadap θ

Catatan 7.5. Pada beberapa keadaan, $Var\left(\hat{\theta}\right)$ tidak bebas terhadap parameter θ. Pada keadaan ini kita masih menggunakan bentuk diatas interval kepercayaan dengan menggantikan parameter θ dengan $\hat{\theta}$ pada pernyataan dari $Var\left(\hat{\theta}\right)$

Contoh 7.14. Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak dari populasi X dengan fungsi kepadatan probabilitas

$$f(x;p) = \begin{cases} p^x (1-p)^{(1-x)}, & jika\, x = 0,1 \\ \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Berapakah aproksimasi interval kepercayaan $100(1-\alpha)\%$ untuk parameter p ?

Jawab :

Fungsi likelihood dari sampel adalah diberikan oleh

$$L(p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{(1-x_i)}$$

Dapatkan logaritma dari fungsi likelihood, kita dapatkan

$$\ln L(p) = \sum_{i=1}^{n} \left[x_i \ln p + (1-x_i)\ln(1-p)\right]$$

Turunkan pernyataan diatas, kita dapatkan

$$\frac{d \ln L(p)}{dp} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}\sum_{i=1}^{n}(1-x_i)$$

Atur persamaan ini dengan menyamakan dengan 0 dan temukan p, kita dapatkan

$$\frac{n\overline{x}}{p} - \frac{n-n\overline{x}}{1-p} = 0$$

Itu adalah

$$(1-p)n\overline{x} = p(n-n\overline{x})$$

Yang mana

$$n\overline{x} - pn\overline{x} = pn - pn\overline{x}$$

Karenanya

$$p = \overline{x}$$

Oleh karena itu, estimator maksimum likelihood dari p adalah diberikan oleh

$$\hat{p} = \overline{X}$$

Varians dari $\overline{X}$ adalah

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

Karena $X \sim BER(p)$, varians $\sigma^2 = p(1-p)$, dan

$$Var(\hat{p}) = Var(\overline{X}) = \frac{p(1-p)}{n}$$

Karena $Var(\hat{p})$ adalah tidak bebas dari parameter p, kita ganti p dengan $\hat{p}$ pada pernyataan dari $Var(\hat{p})$ untuk mendapatkan

$$Var(\hat{p}) \simeq \frac{\hat{p}(1-\hat{p})}{n}$$

Aproksimasi interval kepercayaan $100(1-\alpha)\%$ dari parameter p adalah diberikan oleh

$$\left[\hat{p}-z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ,\ \hat{p}+z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Yang mana

$$\left[\overline{X}-z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\ ,\ \overline{X}+z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\right]$$

Interval kepercayaan diatas adalah aproksimasi interval kepercayaan $100(1-\alpha)\%$ untuk proporsi

Contoh 7.15. Sebuah jajak pendapat diambil dari mahasiswa Universitas sebelum mahasiswa pemilihan. Dari 78 siswa yang dihubungi, 33 mengatakan mereka akan memilih Dr. Smith. Populasi diambil 2200. Dapatkan batas kepercayaan 95% untuk proporsi pemilih dalam populasi mendukung Dr. Smith.

Jawab :

Sample proporsi $\hat{p}$ adalah diberikan oleh

$$\hat{p}=\frac{33}{78}=0.4231$$

Karenanya

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}=\sqrt{\frac{(0.4231)(0.5769)}{78}}=0.0559$$

Persentil ke-2.5 dari distribusi normal adalah diberikan oleh

$$z_{0.025}=1.96,\ (dari\,tabel)$$

Karenanya, batas kepercayaan bawah dari 95% interval kepercayaan adalah

$$\hat{p}-z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}=0.4231-(1.96)(0.0559)$$

$$= 0.4231 - 0.1096$$

$$= 0.3135$$

Demikian pula, batas kepercayaan atas dari 95% interval kepercayaan adalah

$$\hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.4231 + (1.96)(0.0559)$$

$$= 0.4231 + 0.1096$$

$$= 0.5327$$

Karenanya, 95% batas kepercayaan untuk proporsi pemilih dari populasi yang mendukung Dr. Smith adalah 0.3135 dan 0.5327.

Catatan 7.6. Pada contoh 7.15, 95% aproksimasi interval kepercayaan untuk parameter p adalah $[0.3135, 0.5327]$. Interval kepercayaan ini dapat ditingkatkan menjadi interval singkat melalui pertidaksamaan kuadrat. Sekarang kita jelaskan bagaimana interval dapat ditingkatkan. Pertama, perhatikan bahwa pada contoh 7.14, yang mana kita menggunakan dari contoh 7.15, nilai aproksimasi varians dari estimator 3maksimum likelihood $\hat{p}$ diperoleh menjadi $\sqrt{\frac{p(1-p)}{n}}$.

Oleh karena itu, ini adalah varian yang tepat dari $\hat{p}$. Sekarang kuantitas pivot $Q = \frac{\hat{p}-p}{\sqrt{Var(\hat{p})}}$ menjadi

$$Q = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$$

Dengan menggunakan kuantitas pivot, kita dapat mengatur 95% interval kepercayaan

$$0.05 = P\left(-z_{0.025} \leq \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{0.025}\right)$$

$$= P\left(\left|\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}\right| \leq 1.96\right)$$

Dengan menggunakan $\hat{p} = 0.4231$ dan $n = 78$, kita selesaikan dengan pertidaksamaan

$$\left|\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}\right| \leq 1.96$$

Yang mana

$$\left|\frac{0.4231-p}{\sqrt{\frac{p(1-p)}{78}}}\right| \leq 1.96$$

Kuadratkan kedua sisi dari pertidaksamaan diatas dan selesaikan, kita dapatkan

$$78(0.4231-p)^2 \leq (1.96)^2 \left(p-p^2\right)$$

Pertidaksamaan terakhir adalah ekuivalen

$$13.96306158 - 69.84520000p + 81.84160000p^2 \leq 0$$

Selesaikan persamaan kuadrat ini, kita memperoleh $[0.3196\ ,0.5338]$ sebagai interval kepercayaan untuk p. Interval ini adalah meningkat sejak panjangnya adalah 0.2142 dimana panjang interval $[0.3135\ ,0.5327]$ adalah 0.2192.

Contoh 7.16. Jika $X_1, X_2, \ldots, X_n$ adalah sampel acak dari populasi dengan densitas

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & jika\, 0 < x < 1 \\ \\ 0, & untuk\ yang\ lainnya \end{cases}$$

Dimana $\theta > 0$ adalah parameter yang tidak diketahui, berapakah $100(1-\alpha)\%$ aproksimasi interval kepercayaan untuk θ jika ukuran sampel besar?

Jawab :

Fungsi likelihood $L(\theta)$ dari sampel adalah

$$L(\theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1}$$

Karenanya

$$\ln L(\theta) = n \ln \theta + (\theta - 1) \sum_{i=1}^{n} \ln x_i$$

Turunan pertama logaritma dari fungsi likelihood adalah

$$\frac{d}{d\theta} \ln L(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \ln x_i$$

Atur turunan ini dengan menyamakan dengan 0 dan temukan θ, kita peroleh

$$\theta = -\frac{n}{\sum_{i=1}^{n} \ln x_i}$$

Karenanya, estimator lokelihood maksimum dari θ adalah diberikan oleh

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \ln X_i}$$

Untuk menemukan varians dari estimator ini adalah sangat sulit. Kita hitung varians dengan menghitung CRLB *(Cramer-Rao Lower Bound)* untuk estimator ini. Turunan kedua logaritma dari fungsi likelihood adalah diberikan oleh

$$\frac{d^2}{d\theta^2}\ln L(\theta)=\frac{d}{d\theta}\left(\frac{n}{\theta}+\sum_{i=1}^{n}\ln x_i\right)$$

$$=-\frac{n}{\theta^2}$$

Karenanya

$$E\left(\frac{d^2}{d\theta^2}\ln L(\theta)\right)=-\frac{n}{\theta^2}$$

Oleh karena itu

$$Var\left(\hat{\theta}\right)\geq\frac{\theta}{n}$$

Kemudian kita ambil

$$Var\left(\hat{\theta}\right)\simeq\frac{\theta}{n}$$

Saat $Var\left(\hat{\theta}\right)$ memiliki θ pada pernyataan, kita ganti θ yang tidak diketahui dengan estimasi $\hat{\theta}$ sehingga

$$Var\left(\hat{\theta}\right)\simeq\frac{\hat{\theta}^2}{n}$$

Aproksimasi interval kepercayaan $100(1-\alpha)\%$ dari θ adalah diberikan oleh

$$\left[\hat{\theta}-z_{\frac{\alpha}{2}}\frac{\hat{\theta}}{\sqrt{n}}\quad,\ \hat{\theta}+z_{\frac{\alpha}{2}}\frac{\hat{\theta}}{\sqrt{n}}\right]$$

Yang mana

$$\left[-\frac{n}{\sum_{i=1}^{n}\ln X_i}+z_{\frac{\alpha}{2}}\left(\frac{\sqrt{n}}{\sum_{i=1}^{n}\ln X_i}\right)\quad,\ -\frac{n}{\sum_{i=1}^{n}\ln X_i}-z_{\frac{\alpha}{2}}\left(\frac{\sqrt{n}}{\sum_{i=1}^{n}\ln X_i}\right)\right]$$

Catatan 7.7. Pada bagian selanjutnya, kita turunkan interval kepercayaan yang tepat untuk θ saat distribusi populasi eksponensial. Interval kepercayaan $100(1-\alpha)\%$ yang tepat untuk θ adalah diberikan oleh

$$\left[-\frac{\chi^2_{\frac{\alpha}{2}}(2n)}{2\sum_{i=1}^{n}\ln X_i}\ ,\ -\frac{\chi^2_{1-\frac{\alpha}{2}}(2n)}{2\sum_{i=1}^{n}\ln X_i}\right]$$

Perhatikan bahwa interval kepercayaan yang tepat ini bukanlah interval kepercayaan pendek untuk parameter θ

Contoh 7.17. Jika $X_1, X_2, \ldots, X_{49}$ adalah sampel acak dari populasi dengan densitas

$$f(x;\theta)=\begin{cases}\theta x^{\theta-1}, & jika\, 0<x<1\\ 0, & untuk\ yang\, lainnya\end{cases}$$

Dimana $\theta > 0$ adalah parameter yang tidak diketahui, berapakah 90% aproksimasi dan interval kepercayaan yang tepat untuk θ jika $\sum_{i=1}^{49}\ln X_i = -0.7567$?

Jawab :

$$n = 49$$

$$\sum_{i=1}^{49}\ln X_i = -0.7567$$

$$1-\alpha = 0.90$$

Karenanya kita dapatkan

$$z_{0.05} = 1.64$$

$$\frac{n}{\sum_{i=1}^{n}\ln X_i} = \frac{49}{-0.7567} = -64.75$$

Dan

$$\frac{\sqrt{n}}{\sum_{i=1}^{n} \ln X_i} = \frac{7}{-0.7567} = -9.25$$

Karenanya, aproksimasi interval kepercayaan adalah diberikan oleh

$$\left[64.75-(1.64)(9.25),\ 64.75+(1.64)(9.25)\right]$$

Itu adalah $[49.58,\ 79.92]$

Selanjutnya, kita hitung 90% interval kepercayaan yang tepat untuk θ dengan menggunakan formula

$$\left[-\frac{\chi^2_{\frac{\alpha}{2}}(2n)}{2\sum_{i=1}^{n} \ln X_i}\ ,\ -\frac{\chi^2_{1-\frac{\alpha}{2}}(2n)}{2\sum_{i=1}^{n} \ln X_i}\right]$$

Dari tabel chi-square, kita dapatkan

$$\chi^2_{0.05}(98) = 77.93 \quad dan \quad \chi^2_{0.95}(98) = 124.34$$

Karenanya 90% interval kepercayaan yang tepat adalah

$$\left[\frac{77.93}{(2)(0.7567)}, \frac{124.34}{(2)(0.7567)}\right]$$

Itu adalah $[51.49\ ,82.16]$

Contoh 7.18. Jika $X_1, X_2, \ldots, X_n$ adalah sampel acak dari populasi dengan densitas

$$f(x;\theta) = \begin{cases} (1-\theta)\theta^2, \ jika\, x = 0,1,2,\ldots,\infty \\ \\ 0, \qquad\qquad untuk\ yang\ lainnya \end{cases}$$

Dimana $0 < \theta < 1$ adalah parameter yang tidak diketahui, berapakah aproksimasi interval kepercayaan $100(1-\alpha)\%$ untuk θ jika ukuran sampel besar?

Jawab :

Logaritma dari fungsi likelihood dari sampel adalah

$$\ln L(\theta) = \ln\theta \sum_{i=1}^{n} x_i + n\ln(1-\theta)$$

Dengan menggunakan turunan kita peroleh

$$\frac{d}{d\theta}\ln L(\theta) = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{n}{1-\theta}$$

Menyamakan turunan ini dengan nol dan menyelesaikannya untuk θ, kita dapatkan $\theta = \frac{\overline{x}}{1+\overline{x}}$. Kemudian, estimator maksimum likelihood dari θ adalah diberikan oleh

$$\hat{\theta} = \frac{\overline{X}}{1+\overline{X}}$$

Selanjutnya, kita temukan varians dari estimator ini dengan menggunakan CRLB. Untuk itu, kita memerlukan turunan kedua dari $\ln L(\theta)$. Karenanya

$$\frac{d^2}{d\theta^2}\ln L(\theta) = -\frac{n\overline{x}}{\theta^2} - \frac{n}{(1-\theta)^2}$$

Oleh karena itu

$$E\left(\frac{d^2}{d\theta^2}\ln L(\theta)\right) = E\left(-\frac{n\overline{X}}{\theta^2} - \frac{n}{(1-\theta)^2}\right)$$

$$= \frac{n}{\theta^2}E(\overline{X}) - \frac{n}{(1-\theta)^2}$$

$$= \frac{n}{\theta^2}\frac{1}{(1-\theta)} - \frac{n}{(1-\theta)^2} \quad (X_i \sim GEO(1-\theta))$$

$$= -\frac{n}{\theta(1-\theta)}\left[\frac{1}{\theta}+\frac{\theta}{1-\theta}\right]$$

$$= -\frac{n\left(1-\theta+\theta^2\right)}{\theta^2\left(1-\theta\right)^2}$$

Oleh karena itu

$$Var\left(\hat{\theta}\right) \simeq \frac{\hat{\theta}^2\left(1-\hat{\theta}\right)^2}{n\left(1-\hat{\theta}+\hat{\theta}^2\right)}$$

Aproksimasi interval kepercayaan $100(1-\alpha)\%$ untuk θ adalah diberikan oleh

$$\left[\hat{\theta} - z_{-\frac{\alpha}{2}}\frac{\hat{\theta}\left(1-\hat{\theta}\right)}{\sqrt{n\left(1-\hat{\theta}+\hat{\theta}^2\right)}}\right], \hat{\theta} + z_{\frac{\alpha}{2}}\frac{\hat{\theta}\left(1-\hat{\theta}\right)}{\sqrt{n\left(1-\hat{\theta}+\hat{\theta}^2\right)}}$$

Dimana

$$\hat{\theta} = \frac{\overline{X}}{1+\overline{X}}$$

7.7. Metode Statistik atau Umum

Sekarang kita deskripsikan secara singkat Metode Statistik dan Umum untuk mengkonstruksi sebuah interval kepercayaan. Misalkan $X_1, X_2, \ldots, X_n$ menjadi sampel acak dari populasi dengan densitas $f(x;\theta)$, dimana θ adalah parameter yang tidak diketahui. Kita akan menentukan estimator interval untuk θ. Misalkan $T(X_1, X_2, \ldots, X_n)$ menjadi beberapa statistik memiliki fungsi densitas $g(t;\theta)$. Misalkan p_1 dan p_2 menjadi dua bilangan positif tetap pada interval terbuka $(0,1)$ dengan $p_1 + p_2 < 1$. Sekarang kita definisikan dua fungsi $h_1(\theta)$ dan $h_2(\theta)$ sebagai berikut :

$$p_1 = \int_{-\infty}^{h_1(\theta)} g(t;\theta)\,dt \qquad dan \qquad p_2 = \int_{-\infty}^{h_2(\theta)} g(t;\theta)\,dt$$

Sedemikian sehingga

$$P\left(h_1(\theta) < T(X_1, X_2, \ldots, X_n) < h_2(\theta)\right) = 1 - p_1 - p_2$$

Jika $h_1(\theta)$ dan $h_2(\theta)$ adalah fungsi monoton pada θ, maka kita dapat menemukan interval kepercayaan

$$P(u_1 < \theta < u_2) = 1 - p_1 - p_2$$

Dimana $u_1 = u_1(t)$ dan $u_2 = u_2(t)$. Statistik $T(X_1, X_2, \ldots, X_n)$ mungkin menjadi statistik kecukupan atau estimator maksimum likelihood. Jika kita meminimalkan panjang $u_2 - u_1$ dari interval kepercayaan, keadaan subyek $1 - p_1 - p_2 = 1 - \alpha$ untuk $0 < \alpha < 1$, kita peroleh interval kepercayaan yang pendek berdasarkan pada statistik T.

7.8. Kriteria untuk Mengevaluasi Interval Kepercayaan

Dalam banyak situasi, kita dapat memiliki lebih dari satu interval kepercayaan untuk parameter yang sama. Oleh karena itu diperlukan beberapa kriteria untuk memutuskan apakah interval tertentu lebih baik dari interval lainnya. Beberapa kriteria baik yang diketahui adalah :

a) Panjang Terpendek

b) Tidak bias

Kriteria panjang terpendek menuntut bahwa interval kepercayaan $[L, U]$ $100(1-\alpha)\%$ yang baik dari parameter harus memiliki panjang terpendek $\ell = U - L$. Dalam metode kuantitas pivot, kita menemukan pivot Q untuk sebuah parameter dan kemudian mengubah pernyataan probabilitas

$$P(a < Q < b) = 1 - \alpha$$

Menjadi

$$P(L < Q < U) = 1 - \alpha$$

Diperoleh interval kepercayaan $100(1-\alpha)\%$ untuk θ. Jika a dan b konstanta dapat ditemukan perbedaan $U - L$ tergantung pada sampel $X_1, X_2, \ldots, X_n$ adalah minimum untuk setiap realisasi sampel, maka interval acak $[L, U]$ dikatakan sebagai interval kepercayaan terpendek berdasarkan Q. Jika kuantitas pivot Q memiliki jenis fungsi densitas tertentu, maka kita dapat dengan mudah membangun interval kepercayaan dengan panjang terpendek.

Teorema 7.6.

Misalkan fungsi densitas dari pivot $Q \sim h(q;\theta)$ menjadi kontinu dan unimodal. Jika dalam beberapa interval $[a,b]$ fungsi densitas h memiliki mode, dan kondisi memenuhi

$$\int_a^b h(q;\theta)\,dq = 1-\alpha$$

$$h(a) = h(b) > 0$$

kemudian interval $[a,b]$ adalah dari panjang terpendek antara seluruh interval yang memenuhi kondisi a

Jika fungsi densitas bukanlah unimodal, maka meminimalkan ℓ adalah diperlukan untuk membangun interval kepercayaan terpendek. Salah satu kelemahan dari kriteria panjang terpendek adalah dalam beberapa kasus, ℓ bisa menjadi variabel acak. Seringkali, panjang interval yang diharapkan $E(\ell) = E(U-L)$ juga digunakan sebagai kriteria untuk mengevaluasi kebaikan suatu interval. Namun, ini juga memiliki kelemahan. Kelemahan dari kriteria ini adalah minimisasi $E(\ell)$ tergantung pada nilai sebenarnya yang tidak diketahui dari parameter . Jika ukuran sampel sangat besar, maka setiap perkiraan interval kepercayaan dibangun menggunakan metode MLE memiliki panjang minimum yang diharapkan.

Interval kepercayaan hanya terpendek berdasarkan pivot Q tertentu. Ini adalah mungkin untuk menemukan pivot $Q^\star$ lain yang mungkin menghasilkan interval yang lebih pendek dari interval terpendek yang ditemukan berdasarkan Q. Pertanyaan yang secara alami muncul adalah bagaimana untuk menemukan pivot yang memberikan interval kepercayaan terpendek di antara pivot yang lainnya. Telah ditunjukkan bahwa suatu kuantitas pivot Q yang merupakan suatu fungsi dari kelengkapan dan statistik kecukupan memberikan interval kepercayaan terpendek.

Ketidakbiasan, adalah kriteria lain untuk menilai kebaikan suatu penaksir interval. Ketidakbiasan didefinisikan sebagai berikut. Interval kepercayaan $100(1-\alpha)\%$ $[L,U]$ dari parameter θ dikatakan tidak bias jika

$$P\left(L \le \theta^\star \le U\right)\begin{cases} \ge 1-\alpha & jika\, \theta^\star = \theta \\ \le 1-\alpha & jika\, \theta^\star \ne \theta \end{cases}$$

LATIHAN SOAL BAB 7

1. Diberikan sampel acak berukuran n dari distribusi normal, $X_i \sim N\left(\mu,\sigma^2\right)$.

   a. Jika diketahui $\sigma^2 = 9$, tentukan 90% interval kepercayaan untuk μ berdasarkan estimasi $\bar{x} = 19.3$ dengan $n = 16$

   Penyelesaian :

$$\sigma^2 = 9$$

$$\sigma = 3, n = 16, \bar{x} = 19.3$$

Untuk 90% interval kepercayaan, $1 - \alpha / 2 = 0,95$ maka

$$1 - \alpha / 2 = P\left[-z_{1-\alpha/2} < \frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma} < z_{1-\alpha/2} \right]$$

$$= P\left[\bar{X} - \frac{z_{0.95}\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{z_{0.95}\sigma}{\sqrt{n}} \right]$$

$$= P\left[19.3 - 1.65\left(\frac{3}{\sqrt{16}} \right) < \mu < 19.3 + 1.65\left(\frac{3}{\sqrt{16}} \right) \right]$$

$$= P\left[19.3 - 1.65\left(\frac{3}{4} \right) < \mu < 19.3 + 1.65\left(\frac{3}{4} \right) \right]$$

$$= P\left[19.3 - 1.2375 < \mu < 19.3 + 1.2375 \right]$$

Sehingga diperoleh interval kepercayaan untuk $\alpha = 90\%$

$$\left(19.3 - 1.2375 \ , 19.3 + 1.2375\right)$$

$$\left(18.1 \ , 20.5\right)$$

b. Berdasarkan pada informasi yang sudah diperoleh dari huruf (a), tentukan batas bawah satu arah 90% limit kepercayaan untuk μ. Juga tentukan batas atas satu arah 90% limit kepercayaan untuk μ

Penyelesaian :

Batas bawah satu arah 90% limit kepercayaan yakni

$$P\left[L\left(X_1,\ldots,X_n\right)<\theta\right]=\gamma$$

$$L\left(x\right)=L\left(x_1,\ldots,x_n\right)$$

$$L\left(x\right)=18.1$$

Batas atas satu arah 90% limit kepercayaan yakni

$$P\left[\theta<U\left(X_1,\ldots,X_n\right)\right]=\gamma$$

$$U\left(x\right)=U\left(x_1,\ldots,x_n\right)$$

$$U\left(x\right)=20.5$$

c. Untuk interval kepercayaan yang diberikan oleh contoh $\left(\overline{x}-\frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}},\overline{x}+\frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}\right)$, diperoleh rumus ukuran sampel yang diperlukan untuk menghasilkan panjang spesifik λ dari sebuah interval. Jika $\sigma^2=9$, maka berapakah ukuran sampel yang dibutuhkan untuk mencapai 90% interval kepercayaan dengan panjang 2

Penyelesaian :

Contoh adalah sebagai berikut

$$\left(\overline{x}-\frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}},\overline{x}+\frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}\right)$$

$$\sigma^2 = 9$$

$$\sigma = 3$$

Untuk 90% interval kepercayaan, $1-\alpha/2 = 0.95$ maka

$$z_{1-\alpha/2} = \sqrt{n}\left(\bar{X}-\mu\right)/\sigma$$

$$z_{0.95} = \sqrt{n}\left(\bar{X}-\mu\right)/3$$

$$1.65 = \sqrt{n}\left(\bar{X}-\mu\right)/3$$

$$(1.65)(3) = \sqrt{n}\left(\bar{X}-\mu\right)$$

$$4.95 = \sqrt{n}\left(\bar{X}-\mu\right)$$

$$n = (4.95)^2$$

$$n = 24.5$$

$$n \approx 25$$

Dengan demikian diperoleh sampel berukuran 25 dengan panjang 2

d. Misalkan σ^2 tidak diketahui. Tentukan 90% interval kepercayaan untuk μ jika $\bar{x} = 19.3$ dan $s^2 = 10.24$ dengan $n = 16$

Penyelesaian :

$$s^2 = 10.24$$

$$s = 3.2, n = 16, \bar{x} = 19.3, v = n-1 = 16-1 = 15$$

Untuk 90% interval kepercayaan, $1-\alpha/2=0.95$ maka

$$1-\alpha=P\left[-t_{1-\alpha/2}<\frac{\left(\overline{X}-\mu\right)}{s/\sqrt{n}}<t_{1-\alpha/2}\right]$$

$$=P\left[\overline{X}-t_{1-\alpha/2}\ s/\sqrt{n}<\mu<\overline{X}+t_{1-\alpha/2}\ s/\sqrt{n}\right]$$

$$=P\left[19.3-t_{0.95}\,3.2/\sqrt{16}<\mu<19.3+t_{0.95}\,3.2/\sqrt{16}\right]$$

$$=P\left[19.3-1.753\left(\frac{3.2}{\sqrt{16}}\right)<\mu<19.3+1.753\left(\frac{3.2}{\sqrt{16}}\right)\right]$$

$$=P\left[19.3-1.753\left(0.8\right)<\mu<19.3+1.753\left(0.8\right)\right]$$

$$=\left[19.3-1.4024<\mu<19.3+1.4024\right]$$

Sehingga diperoleh interval kepercayaan untuk $\alpha=90\%$

$$\left(19.3-1.4024\ ,19.3+1.4024\right)$$

$$\left(17.9\ ,20.7\right)$$

e. Berdasarkan pada data huruf (d), tentukan 99% interval kepercayaan untuk σ^2

Penyelesaian :

$$s=3.2, n=16, \overline{x}=19.3, v=n-1=16-1=15$$

Untuk 99% interval kepercayaan, $\frac{\alpha}{2}=0.005$ dan $1-\frac{\alpha}{2}=0.995$ maka

$$1-\alpha = P\left[\chi^2_{\alpha/2} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{1-\alpha/2}\right]$$

$$= P\left[\chi^2_{0.005} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{0.995}\right]$$

$$= P\left[\frac{(n-1)s^2}{\chi^2_{0.995}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{0.005}}\right]$$

$$= P\left[\frac{(15)(10.24)}{\chi^2_{0.995}} < \sigma^2 < \frac{(15)(10.24)}{\chi^2_{0.005}}\right]$$

$$= P\left[\frac{153.6}{32.8} < \sigma^2 < \frac{153.6}{4.6}\right]$$

Sehingga diperoleh interval kepercayaan untuk $\alpha = 99\%$

$$\left(\frac{153.6}{32.8}, \frac{153.6}{4.6}\right) = (4.68\ , 33.39)$$

2. Misalkan bahwa $X_1, X_2, \ldots, X_n$ merupakan sampel acak dari distribusi eksponensial $X_i \sim EXP(\theta)$

   a. Jika $\overline{x} = 17.9$ dengan $n = 50$, tentukan batas bawah satu arah 95% limit kepercayaan untuk θ

   Penyelesaian :

   Untuk 95% interval kepercayaan dengan $\gamma = 0.95$

$$\gamma = P\left[2n\overline{X} / \chi^2_\gamma(2n) < \theta\right]$$

$$L(x) = 2n\overline{x} / \chi^2_\gamma(2n)$$

$$= 2(50)(17.9) / \chi^2_{0.95}(2(50))$$

$$= 1790 / \chi^2_{0.95}(100)$$

$$= 1790 / 124.34$$

$$= 14.4$$

b. Tentukan batas bahwa satu arah 95% limit kepercayaan untuk $P(X > t) = e^{-t/\theta}$ dimana t adalah sebuah nilai yang berubah-ubah

Penyelesaian :

$$\gamma = P\left[2n\overline{X} / \chi^2_{\gamma}(2n) < \theta\right]$$

$$L(x) = 2n\overline{X} / \chi^2_{\gamma}(2n)$$

$$= 2(50)(17.9) / \chi^2_{0.95}(2(50))$$

$$= 1790 / \chi^2_{0.95}(100)$$

$$= \frac{1790}{124.34}$$

$$= 14.4$$

Untuk $P(X > t) = e^{-t/\theta}$ maka dapat diperoleh batas bawah satu arah 95% limit kepercayaan adalah $e^{-t/14.4}$

3. Data berikut adalah waktu (per jam) gangguan peralatan AC di bagian pesawat terbang :

$$74, 57, 48, 29, 502, 12, 70, 21, 29, 386, 59, 27, 153, 26, 326$$

Asumsikan bahwa data adalah nilai yang diamati dari sampel acak berdistribusi eksponensial $X_i \sim EXP(\theta)$

a. Tentukan 90% interval kepercayaan untuk rata-rata waktu gangguan, θ

Penyelesaian :

Dari data yang diberikan, dapat diperoleh

$$n = 15$$

$$\bar{x} = \frac{74+57+48+29+502+12+70+21+29+386+59+27+153+26+326}{15}$$

$$= \frac{1819}{15}$$

$$= 121.3$$

Untuk 90% interval kepercayaan, $1-\alpha/2 = 0.95$ maka

$$1-\alpha = P\left[\chi^2_{\alpha/2}(2n) < 2n\bar{X}/\theta < \chi^2_{1-\alpha/2}(2n)\right]$$

$$= P\left[2n\bar{X}/\chi^2_{1-\alpha/2}(2n) < \theta < 2n\bar{X}/\chi^2_{\alpha/2}(2n)\right]$$

$$= P\left[2(15)(121.3)/\chi^2_{0.95}(2(15)) < \theta < 2(15)(121.3)/\chi^2_{0.05}(2(15))\right]$$

$$= P\left[3639/\chi^2_{0.95}(30) < \theta < 3639/\chi^2_{0.05}(30)\right]$$

$$= P\left[3639/43.77 < \theta < 3639/18.49\right]$$

Sehingga diperoleh interval kepercayaan untuk $\alpha = 90\%$

$$\left(\frac{3639}{43.77}, \frac{3639}{18.49}\right) = (83.1, 196.8)$$

b. Tentukan batas bawah satu arah 95% limit kepercayaan untuk persentil ke-10 dari distribusi waktu gangguan

Penyelesaian :

Untuk 95% interval kepercayaan dengan persentil ke-10 maka

$$\gamma = P\left[2n\overline{X} / \chi^2_\gamma(2n) < \theta\right]$$

$$L(x) = 2n\overline{X} / \chi^2_\gamma(2n)$$

$$= 2(15)(121.3) / \chi^2_{0.1}(2(15))$$

$$= 3639 / \chi^2_{0.1}(30)$$

$$= \frac{3639}{20.6}$$

$$= 176.65$$

4. Misalkan $X_1, \ldots, X_n$ adalah sampel acak dari distribusi eksponensial dua parameter, $X_i \sim EXP(\theta, \eta)$. Misalkan diketahui $\eta = 150$, tentukan kuantitas pivot untuk parameter θ berdasarkan statistik kecukupan

Penyelesaian :

Fungsi kepadatan probabilitas : $f(x;\theta,\eta) = \frac{1}{\theta} e^{-\frac{(x-\eta)}{\theta}}$, $X_i \sim EXP(\theta,\eta)$

$\eta = 150$ sehingga $f(x;\theta) = \frac{1}{\theta^n} e^{-\frac{1}{\theta}(\Sigma X_i - 150)}$

$\delta = \sum X_i - 150$ adalah statistik kecukupan untuk θ

$$X_i - 150 \sim EXP(\theta, \eta)$$

$$\delta = \sum X_i - 150 \sim GAM(\theta, \eta)$$

$$\frac{2\delta}{\theta} \sim \chi^2(2n)$$

Jadi, $\dfrac{2\sum X_i - 150}{\theta}$ adalah kuantitas pivot

5. Misalkan $X_1, X_2, \dots, X_n$ adalah contoh acak dari distribusi Weibull $X \sim WEI(\theta, 2)$,

   a. Tunjukkan bahwa $Q = 2\sum_{i=1}^{n} \dfrac{X_i^2}{\theta^2} \sim \chi^2(2n)$
   b. Gunakan Q untuk mendapatkan interval kepercayaan $100\gamma\%$ terhadap θ
   c. Dapatkan batas bawah limit kepercayaan $100\gamma\%$ untuk $P(x > t) = \exp\left[-\left(\dfrac{t}{\theta}\right)^2\right]$
   d. Dapatkan batas atas limit kepercayaan untuk persentil ke- p dari distribusi ini

Penyelesaian :

Diketahui :

$X_1, X_2, \dots, X_n$ adalah contoh acak

distribusi Weibull $X \sim WEI(\theta, 2)$

Fungsi kepadatan probabilitas : $f(x;\theta,\beta) = \begin{cases} \dfrac{\beta}{\theta^\beta} x^{\beta-1} e^{-\left(\frac{x}{\theta}\right)^\beta}, & untuk\ x > 0 \\ 0, & untuk\ yang\ lainnya \end{cases}$

Maka $f(x;\theta,2) = \dfrac{2}{\theta^2} x^{2-1} e^{-\left(\frac{x}{\theta}\right)^2}$

Untuk $f(x_1, x_2, \dots, x_n; \theta, 2) = \dfrac{2}{\theta^2} \sum_{i=1}^{n} x_i e^{-\left(\frac{x}{\theta}\right)^2}$

a. Akan ditunjukkan bahwa $Q = 2\sum_{i=1}^{n} \dfrac{X_i^2}{\theta^2} \sim \chi^2(2n)$

   Jika $\overline{X}$ adalah kecukupan untuk θ dan juga $2n\overline{X}/\theta \sim \chi^2(2n)$. Telah jelas bahwa persentil ke-γ, $\chi^2_\gamma(2n)$ dapat dilihat pada Tabel 4 (*Introduction to Probability and Mathematical Statistics*) halaman 604. Dengan demikian,

$$\gamma = P\left[2n\overline{X}/\theta < \chi^2_\gamma(2n)\right]$$

$$\gamma = P\left[2n\overline{X}/\chi^2_\gamma(2n) < \theta\right]$$

Jika $\overline{x}$ dapat dihitung, maka batas bawah satu arah $100\gamma\%$ limit kepercayaan diberikan oleh

$$L(x) = 2n\overline{X} / \chi_{\gamma}^{2}(2n)$$

Sama halnya, batas atas satu arah $100\gamma\%$ limit kepercayaan diberikan oleh

$$U(x) = 2n\overline{X} / \chi_{1-\gamma}^{2}(2n)$$

Sehingga dari definisi kuantitas pivot diperoleh

$$Q = \frac{2}{\theta^2}\sum_{i=1}^{n} X_i^2 \sim \chi^2(2n)$$

$$Q = 2\sum_{i=1}^{n} \frac{X_i^2}{\theta^2} \sim \chi^2(2n)$$

b. Menentukan interval kepercayaan $100\gamma\%$ terhadap θ dengan menggunakan Q

$$Q = 2\sum_{i=1}^{n} \frac{X_i^2}{\theta^2} \sim \chi^2(2n)$$

Dalam hal ini, θ adalah parameter skala karena Q juga $Q = \frac{2}{\theta}\sum_{i=1}^{n} \frac{X_i^2}{\theta} \sim \chi^2(2n)$ sehingga $Q = \frac{\hat{\theta}}{\theta}$ adalah kuantitas pivot

$$2\sum_{i=1}^{n} \frac{X_i^2}{\theta^2} \sim \chi^2(2n)$$

Anggaplah bahwa kita menginginkan $100\gamma\%$ limit kepercayaan untuk θ. Jika dipilih nilai $\gamma_1 > 0$ dan $\gamma_2 > 0$ seperti $\gamma_1 + \gamma_2 = \gamma_0 = 1 - \gamma$, maka

$$P\left[\chi_{\gamma_1}^2(2n) < 2\sum_{i=1}^{n} \frac{X_i^2}{\theta^2} \sim \chi^2(2n) < \chi_{1-\gamma_2}^2(2n)\right] = 1 - \gamma_2 - \gamma_1$$

Dengan demikian,

$$P\left[2\sum_{i=1}^{n}X_i^2 / \chi^2_{1-\gamma_2}(2n) < \theta^2 < 2\sum_{i=1}^{n}X_i^2 / \chi^2_{\gamma_1}(2n)\right] = \gamma$$

$$P\left[\sqrt{2\sum_{i=1}^{n}X_i^2 / \chi^2_{1-\gamma_2}(2n)} < \theta < \sqrt{2\sum_{i=1}^{n}X_i^2 / \chi^2_{\gamma_1}(2n)}\right] = \gamma$$

Sehingga interval kepercayaan $100\gamma\%$ terhadap θ adalah

$$\theta = \left(\sqrt{\frac{\left(2\sum_{i=1}^{n}X_i^2\right)}{\chi^2_{1-\gamma_2}(2n)}} \sqrt{\frac{\left(2\sum_{i=1}^{n}X_i^2\right)}{\chi^2_{\gamma_1}(2n)}}\right)$$

c. Batas bawah limit kepercayaan $100\gamma\%$ untuk $P(x > t) = \exp\left[-\left(\frac{t}{\theta}\right)^2\right]$

$$P\left(\chi^2_\gamma > t\right) = \gamma$$

$$P\left(\chi^2_\gamma > \exp\left[-\left(\frac{t}{\theta}\right)^2\right]\right) = 1 - P\left(\chi^2_\gamma < \exp\left[-\left(\frac{t}{\theta}\right)^2\right]\right) = \gamma$$

$$= 1 - P\left(\chi^2_\gamma < \exp\left[-\left(\frac{t}{\theta}\right)^2\right]\right) = \gamma$$

Diperoleh batas bawah limit kepercayaan $100\gamma\%$ untuk $P(x > t) = \exp\left[-\left(\frac{t}{\theta}\right)^2\right]$

adalah $1 - P\left(\chi^2_\gamma < \exp\left[-\left(\frac{t}{\theta}\right)^2\right]\right)$

d. Batas atas limit kepercayaan untuk persentil ke- p

Jika $\chi^2_{1-\gamma/2} = \chi^2_{1-\frac{\gamma}{2}}(2n)$ adalah persentil ke- p dari distribusi chi-square dengan derajat kebebasan $(2n)$, maka

p dengan daerah $1 - \gamma = P\left[\chi^2_{\gamma/2} < 2\sum_{i=1}^{n}\frac{X_i^2}{\theta^2} < \chi^2_{1-\gamma/2}\right]$

$$= P\left[\frac{\chi^2_{\gamma/2}}{2\sum_{i=1}^{n} X_i^2} < \frac{1}{\theta^2} < \frac{\chi^2_{1-\gamma/2}}{2\sum_{i=1}^{n} X_i^2}\right]$$

$$= P\left[\frac{\chi^2_{1-\gamma/2}}{2\sum_{i=1}^{n} X_i^2} < \theta^2 < \frac{\chi^2_{\gamma/2}}{2\sum_{i=1}^{n} X_i^2}\right]$$

$$= P\left[\sqrt{\frac{\chi^2_{1-\gamma/2}}{2\sum_{i=1}^{n} X_i^2}} < \theta < \sqrt{\frac{\chi^2_{\gamma/2}}{2\sum_{i=1}^{n} X_i^2}}\right]$$

Diperoleh limit kepercayaan

$$\theta = \left(\sqrt{\frac{\chi^2_{1-\gamma/2}}{2\sum_{i=1}^{n} X_i^2}}\ ,\sqrt{\frac{\chi^2_{\gamma/2}}{2\sum_{i=1}^{n} X_i^2}}\right)$$

Dengan batas atas $\theta = \sqrt{\frac{\chi^2_{\gamma/2}}{2\sum_{i=1}^{n} X_i^2}}$

6. Misalkan p adalah proporsi orang Yogyakarta yang berambut keriting. Dalam suatu sampel berukuran 40, ditemukan 5 orang berambut keriting. Temukan pendekatan 90% dari interval kepercayaan terhadap p

Penyelesaian :

Diketahui bahwa p adalah proporsi orang Yogyakarta yang berambut keriting

$x = 5$ adalah orang yang berambut keriting

$n = 40$ adalah ukuran sampel

$$\hat{p} = \frac{x}{n} = \frac{5}{40} = \frac{1}{8} = 0.125$$

$1-\alpha = 90\% = 0.9$ sehingga $\alpha = 0.1$ (koefisien kepercayaan)

$$1-\frac{\alpha}{2} = 1-\frac{0.1}{2} = 1-0.05 = 0.95$$

$$-Z_{1-\frac{\alpha}{2}} < Z < Z_{1-\frac{\alpha}{2}}$$

$$-Z_{1-\frac{\alpha}{2}} < \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}} < Z_{1-\frac{\alpha}{2}}$$

$$-Z_{0.95} < \frac{0.125-p}{\sqrt{\frac{0.125\left(1-0.125\right)}{40}}} < Z_{0.95}$$

$$-Z_{0.95} < \frac{0.125-p}{\sqrt{\frac{0.125\left(0.875\right)}{40}}} < Z_{0.95}$$

$$-Z_{0.95} < \frac{0.125-p}{\sqrt{\frac{0.109375}{40}}} < Z_{0.95}$$

$$-Z_{0.95} < \frac{0.125-p}{\sqrt{0.0027344}} < Z_{0.95}$$

$$-Z_{0.95} < \frac{0.125-p}{0.0523} < Z_{0.95}$$

$$p = 0.125 \pm Z_{0.95} \bullet 0.0523$$

$$p = 0.125 \pm 1.645 \bullet 0.0523$$

$$p = 0.125 \pm 0.086$$

Diperoleh $p = \left(0.211; 0.039\right)$

Dengan demikian, pendekatan 90% dari interval kepercayaan terhadap p adalah $\left(0.211; 0.039\right)$

7. Terdapat 45 pekerja bangunan yang dipilih secara acak untuk mempelajari tingkat kecelakaan kerja yang terjadi. Angka kecelakaan tiap pekerja diasumsikan berdistribusi Poisson dengan mean μ, sedangkan nilai rata-rata kecelakaan tiap pekerja adalah $\bar{x} = 1.7$

   a. Temukan pendekatan dari batas bawah sebesar 90% terhadap limit kepercayaan untuk μ menggunakan persamaan $\frac{\bar{X}-\mu}{\sqrt{\mu/n}} \to Z \sim N(0,1)$

   b. Gunakan perintah pada soal huruf (a), jika diselesaikan dengan persamaan $\frac{\bar{X}-\mu}{\sqrt{\bar{X}/n}} \to Z \sim N(0,1)$

   Penyelesaian :

   Diketahui bahwa :

   $n = 45$

   $\bar{x} = 1.7$

   $1-\alpha = 90\% = 0.9$ sehingga $\alpha = 0.1$ (koefisien kepercayaan)

$$1-\frac{\alpha}{2} = 1-\frac{0.1}{2} = 1-0.05 = 0.96$$

   Angka kecelakaan tiap pekerja diasumsikan berdistribusi Poisson dengan mean μ

   Fungsi kepadatan probabilitas distribusi Poisson

$$f(x;\mu) = \frac{e^{-\mu} \bullet \mu^{x}}{x!}$$

   Akan didapatkan dengan pendekatan sesuai persamaan

$$\frac{\bar{X}-\mu}{\sqrt{\mu/n}} \to Z \sim N(0,1)$$

   Dan juga persamaan

$$\frac{\bar{X}-\mu}{\sqrt{\bar{X}/n}} \to Z \sim N(0,1)$$

   a. Dengan persamaan $\frac{\bar{X}-\mu}{\sqrt{\mu/n}} \to Z \sim N(0,1)$

$$-Z_{1-\frac{\alpha}{2}} < Z < Z_{1-\frac{\alpha}{2}}$$

$$-Z_{1-\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{\sqrt{\mu / n}} < Z_{1-\frac{\alpha}{2}}$$

$$-Z_{1-\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{\sigma} < Z_{1-\frac{\alpha}{2}}$$

$$-Z_{0.95} < \frac{1.7-\mu}{1} < Z_{0.95}$$

$$-1.645-1.7 < -\mu < 1.645-1.7$$

$$-1.645+1.7 < \mu < 1.645+1.7$$

$$0.055 < \mu < 3.345$$

Jadi, $\mu = (0.055; 3.345)$

b. Dengan persamaan $\frac{\overline{X}-\mu}{\sqrt{\overline{X} / n}} \rightarrow Z \sim N(0,1)$

$$-Z_{1-\frac{\alpha}{2}} < Z < Z_{1-\frac{\alpha}{2}}$$

$$-Z_{1-\frac{\alpha}{2}} < \frac{\overline{X}-\mu}{\sqrt{\overline{X} / n}} < Z_{1-\frac{\alpha}{2}}$$

$$-Z_{0.95} < \frac{\overline{X}-\mu}{\sqrt{\overline{X} / n}} < Z_{0.95}$$

$$-Z_{0.95} < \frac{1.7-\mu}{\sqrt{1.7 / 45}} < Z_{0.95}$$

$$-Z_{0.95} < \frac{1.7-\mu}{\sqrt{0.038}} < Z_{0.95}$$

$$-1.645 < \frac{1.7-\mu}{0.1944} < 1.645$$

$$-1.645 \bullet 0.1944 < 1.7-\mu < 1.645 \bullet 0.1944$$

$$-0.3198 < 1.7-\mu < 0.3198$$

$$-0.3198-1.7 < -\mu < 0.3198-1.7$$

$$-0.3198+1.7 < -\mu < 0.3198+1.7$$

$$1.380 < \mu < 2.020$$

Dengan demikian, $\mu = (1.380; 2.020)$

8. Diberikan sampel acak independen dari dua distribusi normal, $X_i \sim N\left(\mu_1, \sigma_1^2\right)$ dan $Y_j \sim N\left(\mu_2, \sigma_2^2\right);\ i=1,\ldots,n_1, j=1,\ldots,n_2$. Jika diasumsikan bahwa μ_1 dan μ_2 diketahui, dapatkan interval kepercayaan $100(1-\alpha)\%$ untuk σ_2^2/σ_1^2 berdasarkan statistik kecukupan

Penyelesaian :

Diketahui : ada dua sampel acak independen distribusi normal

$X_i \sim N\left(\mu_1, \sigma_1^2\right)$ dan $Y_j \sim N\left(\mu_2, \sigma_2^2\right);\ i=1,\ldots,n_1, j=1,\ldots,n_2$

Diasumsikan bahwa μ_1 dan μ_2 diketahui

Akan didapatkan interval kepercayaan $100(1-\alpha)\%$ untuk σ_2^2/σ_1^2 berdasarkan statistik kecukupan

Karena $\overline{Y}-\overline{X} \sim N\left(\mu_2-\mu_1, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right), \mu_1$ dan μ_2 diketahui

$$-Z_{1-\frac{\alpha}{2}} < Z < Z_{1-\frac{\alpha}{2}}$$

$$-Z_{1-\frac{\alpha}{2}} < \frac{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < Z_{1-\frac{\alpha}{2}}$$

$$\frac{-Z_{1-\frac{\alpha}{2}}}{\bar{Y}-\bar{X}-(\mu_2-\mu_1)} < \frac{1}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < \frac{Z_{1-\frac{\alpha}{2}}}{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}$$

$$\left(\frac{-Z_{1-\frac{\alpha}{2}}}{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}\right)^2 < \frac{n_1}{\sigma_1^2}+\frac{n_2}{\sigma_2^2} < \left(\frac{Z_{1-\frac{\alpha}{2}}}{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}\right)^2$$

$$\left(\frac{-Z_{1-\frac{\alpha}{2}}}{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}\right)^2 < \frac{n_1\sigma_2^2+n_2\sigma_1^2}{\sigma_1^2\sigma_2^2} < \left(\frac{Z_{1-\frac{\alpha}{2}}}{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}\right)^2$$

Dengan syarat statisik kecukupan pertidaksamaan diatas dapat diubah menjadi

$$\frac{1}{n_1}\left(\sigma_2^2\left(\frac{-Z_{1-\frac{\alpha}{2}}}{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}\right)^2 - n_2\right) < \frac{\sigma_2^2}{\sigma_1^2} < \frac{1}{n_1}\left(\left(\frac{Z_{1-\frac{\alpha}{2}}}{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}\right)^2 - n_2\right)$$

Sehingga interval kepercayaan $100(1-\alpha)\%$ untuk σ_2^2/σ_1^2 berdasarkan statistik kecukupan adalah

$$\frac{\sigma_2^2}{\sigma_1^2} = \left(\frac{1}{n_1}\left(\sigma_2^2\left(\frac{-Z_{1-\frac{\alpha}{2}}}{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}\right)^2 - n_2\right), \frac{1}{n_1}\left(\left(\frac{Z_{1-\frac{\alpha}{2}}}{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}\right)^2 - n_2\right)\right)$$

Namun yang sering dijumpai dalam persoalan, interval kepercayaan untuk $\frac{\sigma_2^2}{\sigma_1^2}$ didapatkan dengan menggunakan distribusi F Snedecor.

9. Misalkan $X_1,\ldots,X_n$ adalah sampel acak dari distribusi Normal, $X \sim N\left(\mu,\sigma^2\right)$. Jika t adalah bilangan asli, temukan statistik yang merupakan statistik kecukupan dan distribusi yang bergantung pada t,μ,σ^2 hanya jika $F\left(t;\mu,\sigma^2\right)=P\left(X \le t\right)$

Penyelesaian :

Parameter skala lokasi $f\left(x;\mu,\sigma^2\right)=F_0\left(\frac{x-\mu}{\sigma^2}\right)$

Misalkan :

$$c=\left(\frac{t-\mu}{\sigma^2}\right)$$

$$c\sigma^2=t-\mu$$

$$t=c\sigma^2+\mu$$

Jika $\hat{\mu}$ dan $\hat{\sigma}^2$ adalah MLE dari μ dan σ^2 yang mana bergantung pada t,μ,σ^2

$$F\left(t;\mu,\sigma^2\right)=P\left(X \le t\right)$$

$$F_0=\left(\frac{x-\hat{\mu}}{\hat{\sigma}^2}\right)$$

$$F_0=\left(\frac{c\sigma^2+\mu-\hat{\mu}}{\hat{\sigma}^2}\right)$$

$$F_0=\left(\frac{c\sigma^2}{\hat{\sigma}^2}-\frac{\left(\mu-\hat{\mu}\right)}{\hat{\sigma}^2}\right)$$

Yang mana adalah fungsi dari c dan kuantitas pivot untuk μ dan untuk σ^2 yang bergantung pada $F\left(t;\mu,\sigma^2\right)=P\left(X \le t\right)$

Daftar Pustaka

Sahoo, Prasanna. 2006. *Probability and Mathematical Statistics*. Louisville: Departement of Mathematics, University of Louisville

Walpole, Ronald E dkk. 2012. *Probability & Statistics for Engineers & Scientists* (Ninth Edition), United States of America: Prentice Hall

Wibisono, Yusuf. 2015. *Metode Statistik*. Yogyakarta: Gadjah Mada University Press

Ross, Sheldon M. 2010. *Introduction to Probability Models* (Tenth Edition), California: Elsevier

Wackerly, Dennis D, William Mendenhall III, Richard L Scheaffer. 2008. *Mathematical Statistics with Applications* (Seventh Edition). Florida: Duxbury Press

Soong, T.T. 1988. *Fundamentals of Probability and Statistics for Engineers.* New York:John Wiley and Sons

Ott R Lyman, Michael Longnecker. 2015. *An Introduction to Statistical Methods & Data Analysis* (Seventh Edition). Texas:Cengage Learning

Devore, Jay L, Kenneth N Berk. 2012. *Modern Mathematical Statistics with Applications* (Second Edition), New York:Springer

Roussas, G. 2003. *An Introduction to Probability and Statistical Inference*. San Diego: Academic Press.

Ross, Sheldon M. 2000. *Introduction to Probability and Statistics for Engineers and Scientists.* San Diego: Harcourt Aca

Taylor, L D. 1974. *Probability and Mathematical Statistics.* New York: Harper & Row

Ross, S. 1988. *A First Course in Probability.* New York: Macmillan.

Rinaman, W. C. 1993. *Foundations of Probability and Statistics.* New York: Saunders College Publishing.

Meyer, P. L. 1970. *Introductory Probability and Statistical Applications.* Reading: Addison-W

Casella, G., and R. L. Berger. 2002. *Statistical Inference, (Second Edition).* Pacific Grove,California: Duxbury.

Cramer, H. 1973. *The Elements of Probability Theory and Some of Its Applications, (Second Edition).* Huntington, New York: Krieger.

Hogg, R. V., A. T. Craig, and J. W. McKean. 2005. *Introduction to Mathematical Statistics, (Sixth Edition).* Upper Saddle River, New Jersey: Pearson Prentice Hall.

Lindgren, B. W. 1993. *Statistical Theory, (Fourth Edition).* Boca Raton, Florida: Chapman and Hall/CRC.

Miller, I., and M. Miller. 2003. *John E. Freud's Mathematical Statistics with Applications, (Seventh Edition).* Upper Saddle River, New Jersey: Pearson Prentice Hall.

Mood, A. M., F. A. Graybill, and D.Boes. 1974. *Introduction to the Theory of Statistics, (Third Edition).* New York: McGraw-Hill.

Serfling, R. J. 2002. *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

Berger, J. O. 1985. Statistical Decision Theory and Bayesian Analysis, (Second Edition). NewYork: Springer-Verlag.

Box, G. E. P., and G. C. Tiao. 1992. Bayesian Inference in Statistical Analysis. New York: WileyClassics.

Rice, J. A. 1995. Mathematical Statistics and Data Analysis, (Second Edition). Belmont,California: Duxbury.

Mosteller, F., R. E. K. Rourke, and G. B. Thomas. 1970. Probability with Statistical Applications,(Second Edition). Reading, Mass. Addison-Wesley