③ Sampling System に就いて

水　野

§0 前書

Sampling Estimation をする時の考へ方を分析してみると次の様になると思はれる。

我々が調べたい対象の集りがある。之は我々が merkmal と呼んでゐるものを Element とする集合である。之を Universe と云ふ事にする。我々が知りたいのは此の Universe の Element に依って決定されるある量 Θ である。之を Universe Parameter と呼ばう。

我々は此処で Universe の任意の Element に夫々等しい Probability を與へる。此の Probability Field を Ω で表はさう。

註 始めの Ω を Population と呼んだが、Population なる語の不明確性を失くす為関係諸氏との discussion の結果此の名前を使はない事にした。

Ω で Expectation が量としては Θ に等しくなる様な函数 f に着目する。

我々は f の実現値の大きさを以て、θ の大きさと見做すのである。

註 biand な estimate には以下の論を簡単にする為、ふれないが全く平行的に行はれる

之が Sampling に依つて、或る Universe Parameter θ を推定する時の構造である。

此処で、我々の見做しが有効である様にと幾つかの提案がなされてきた。其は種々の Estimate の提唱とか Stratification の使用とかの形をとつたが、本質的なものとしては f の変動の範囲が人間的な尺度に於ける θ を除いて、充分 θ に近接してゐるといふ事、換言すれば Unbiased で Variame が小さいといふ事か、negligible な Rias を持ち、その mean Square Error が小さいと云ふ事であつた。そして此処に於て Optimum な Estimate とか、Optimum な Allocation といふ事が問題とされてきた。

斯う考へてくると問題が生じてくる。推定を行ふ為に確率論的構造を持つた Ω を想定して考へるのが一番穏当な方法であつたといふ事を認めるとしても、何故 Effisiency を考へる時に、此の特殊な Ω で考察する必要があつたらうかと。

Probability field を作る時、その各 Element に対応する Probability を等しくする必要はないではないか。考へられる種

種の Probability field で Effisiecy を考へた方がより合理的であらう。斯くして Sampling System の話は始る。

§ 1 Probability field Ω_p の構成

Universe の size を N とする。その Element を $X_1, \cdots\cdots, X_N$ とする。簡単の為にその量も同一の記号で表はす事にする。

註 X は一般に Vector であってよい。

我々が知りたいのは Universe parameter θ だとする。

(1) $\theta \equiv F(X_1, \cdots\cdots, X_N)$

$X_{(i)}$; $i = 1, \cdots\cdots, n$ を $X_1, \cdots\cdots, X_N$ の何れか一つの値をとる様な変量をとる。

今 $(i_1, \cdots, i_n)$ を $1, \cdots, N$ からとられた size n の任意の組合せとする。

註. $(i_1, \cdots\cdots, i_n)$ を任意の size n なる順列とする事も可能である。此の場合以下の議論も全く平行的に行へる。

今任意の $(X_{i_1}, \cdots\cdots, X_{i_n})$ なる組に対して定義された負ならざる函数 $P(X_{(1)}, \cdots\cdots, X_{(n)})$ を考へる。そして $(X_{i_1}, \cdots\cdots, X_{i_n})$ なる Element の組に対して、$P(X_{i_1}, \cdots\cdots, X_{i_n})$ に比例する Probability を与へる。即ち $\dfrac{P(X_{i_1}, \cdots\cdots, X_{i_n})}{\sum P(X_{i_1}, \cdots\cdots, X_{i_n})}$ なる Probability で

$(X_{i_1}, \cdots, X_{i_n})$ なる組が抽出される様な Probability field を構成する。此処に $\sum$ は総ての $(i_1, \cdots\cdots i_n)$ に亘るものとする。

此の Probability field を P に Proportionate な Probability field と呼び Ω_p で表はさう。之は Equally likely な Probability P として常数をとった時即1に proportionate な probabulity field を以下 Ω で表はす。

field と呼ばれるものである。

§.2 *Estimate* f_p の構成

今 Ω で或函数 $f(X_{(1)},----,X_{(n)})$ が θ の *Unbiased estimate* だとする。即ち

$$(2)\qquad E_\Omega=\left[f(X_{(1)},----,X_{(n)})\right]=\sum f(X_{i1},---,X_{in})\frac{1}{\binom{N}{n}}$$
$$=\theta$$

こゝに E_Ω は Ω での *Expectation* を表はす。

此処で組 $X_{i1},-----,X_{in}$ に対して

$$(3)\qquad f_P(X_{i1},-----,X_{in})=\frac{f(X_{i1},---X_{in})}{\dfrac{P(X_{i1},----,X_{in})}{\sum P(X_{i1},----,X_{in})}}\cdot\frac{1}{\binom{N}{n}}$$

で定義される函数 $f_P(X_{(1)},----,X_{(n)})$ を考へる。

然らば明らかに $f_P(X_{(1)},----,X_{(n)})$ は Ω_P で θ の *unbiased estimate* となる。

$$(4)\qquad E_{\Omega P}\left[f_P(X_{(1)},-----,X_{(n)})\right]$$

$$=\sum\frac{f(X_{i1},---,X_{in})}{\dfrac{P(X_{i1},----,X_{in})}{\sum P(X_{i1},---,X_{in})}}\cdot\frac{1}{\binom{N}{n}}\cdot\frac{P(X_{i1},---,X_{in})}{\sum P(X_{i1},---,X_{in})}$$

$$=\sum f(X_{i1},----,X_{in})\cdot\frac{1}{\binom{N}{n}}$$

$$=\theta$$

§.3. *Sampling System*

或る函数 f が θ の Estimate であるのは Probability field Ω が与へられてるからであつて f は Ω に relative なものである。

依つて、今 f が Ω で θ の estimate とされる時に Ω, f の組を以て θ に対する Sampling System と称し (Ω, f) で表はす。

(5) $E_\Omega(f)=\theta$

が成立する時、Sampling System (Ω, f) は unbiased なりといふ。

又 Ω に於ける f の variance を以て Sampling System (Ω, f) の Variance と称し $V_\Omega(f)$ で表はす。

(6) $$V_\Omega(f)=E_\Omega[f-E_\Omega(f)]^2$$

然らば我々としては、Unbiased で minimum な Variance を持つ Sampling System を求めるのが妥当であらう。

前に考察した Ω_p, f_p を以て Sampling System (Ω_p, f_p) を考へれば、之は θ の unbiased な Sampling System となり Variance は次式で与へられる。

(7) $$V_{\Omega_p}(f_p)=E_{\Omega_p}(f_p^2)-\{E_{\Omega_p}(f_p)\}^2$$

$$=\sum \frac{\{f(X_{i_1},\cdots,X_{i_n})\}^2}{\left\{\frac{P(X_{i_1},\cdots,X_{i_n})}{\sum P(X_{i_1},\cdots,X_{i_n})}\right\}^2}\cdot\frac{1}{\binom{N}{n}^2}\cdot\frac{P(X_{i_1},\cdots,X_{i_n})}{\sum P(X_{i_1},\cdots,X_{i_n})}-\theta^2$$

$$=\sum \frac{\{f(X_{i_1},\cdots,X_{i_n})\}^2}{P(X_{i_1},\cdots,X_{i_n})}\cdot\frac{1}{\binom{N}{n}}\cdot\sum\frac{P(X_{i_1},\cdots,X_{i_n})}{\binom{N}{n}}-\theta^2$$

$$=E_\Omega\left[\frac{f^2(X_{(1)},\cdots,X_{(n)})}{P(X_{(1)},\cdots,X_{(n)})}\right]\cdot E_\Omega\left[P(X_{(1)},\cdots,X_{(n)})\right]-\theta^2$$

§. 4. Efficiency

Sampling System (Ω_p, f_p) の Effisiency $\alpha(\Omega_p, f_p)$ を次の量で測る。

(8) $\alpha(\Omega_p, f_p)=\dfrac{V_\Omega(f)}{V_{\Omega_p}(f_p)}$

(9) $\alpha(\Omega_p, f_p)>\alpha(\Omega_{p'}, f_{p'})$

なる時 (Ω_p, f_p) は $(\Omega_{p'}, f_{p'})$ より efficient である。といふ。

θ の為の Sampling System (Ω_p, f_p) をきめる為には出来るだけ efficient な Sampling System をとればよい。

明らかに Equally likely な Sampling System の efficiency は 1 であり、(Ω_f, f_f) の Efficiency は ∞ である。

此の事は以下の論に示唆を与へる。

我々の採用し得る P が任意の函数であるなら問題はないのだが、我々の知識は、即我々の利用し得る函数族 $\{P\}$ は限定されたものであるから、我々は此の函数族の中で最も Efficiency の高いものを採用するより仕方がない。

与へられた函数族 $\{P\}$ に対して $\alpha(\Omega_p, f_p)$ が maximum である函数 P の決定する Sampling System を optimum といふ。

§. 5. $\alpha(\Omega_p, f_p)$ の近似式

次の式が成立する。

$$(9)\quad E\left(\frac{Y^2}{X}\right)=E\left(\frac{Y^2}{X^2}\right)\cdot E(X)+\varphi_1(X.Y)$$

こゝに

$$(10)\quad \varphi_1(X.Y)=\frac{E\left(\dfrac{E(\frac{1}{X})-\frac{1}{X}}{\frac{1}{X}}\cdot\dfrac{Y^2}{X^2}\right)}{E\left(\frac{1}{X}\right)}-E(X)E\left(\frac{Y^2}{X^2}\right)\frac{E\left(\frac{E(X)-X}{X}\right)}{1+E\left(\frac{E(X)-X}{X}\right)}$$

又

$$(11)\quad E\left(\frac{Y}{X}\right)=\frac{E(Y)}{E(X)}+\varphi_2(X,Y)$$

こゝに

$$(12)\quad \varphi_2(X.Y)=\frac{E\left(\frac{E(X)-X}{X}\cdot Y\right)}{E(X)}$$

又

$$(13)\quad V\left(\frac{Y}{X}\right)=\frac{E^2(Y)}{E^2(X)}\left\{C_X^2-2\rho C_X C_Y+C_Y^2\right\}+\varphi_3(X.Y)$$

こゝに

$$(14)\quad \varphi_3(X.Y)=\frac{1}{E^2(X)}\left\{\frac{E(Y)}{E(X)}-E\left(\frac{Y}{X}\right)\right\}^2+2\rho C_X C_Y\frac{E(Y)}{E(X)}\left\{\frac{E(Y)}{E(X)}-E\left(\frac{Y}{X}\right)\right\}$$
$$-C_X^2\left\{\frac{E^2(Y)}{E^2(X)}-E^2\left(\frac{Y}{X}\right)\right\}-\frac{1}{E(X)}E\left\{\frac{\{X^2-E^2(X)\}\left\{Y^2-X\frac{E(Y)}{E(X)}\right\}^2}{X^2}\right\}$$

P_1 は *Correlation Coefficient* を C は *Coefficient of variation* を表はす。

依つて

$$(15)\quad \alpha(\Omega_p, f_p) = \frac{C_f^2}{C_f^2(1-\rho)+(C_p-\rho C_f)^2+\varphi'(f.P)}$$

$$= \frac{1}{(1-\rho^2)+\left(\frac{C_p}{C_f}-\rho\right)^2}+\varphi(f.p)$$

となる。こゝにρは$f.P.$のCovelation Coefficient を $\varphi(f.P)$は $\varphi_1, \varphi_2, \varphi_3$, を含む函数である。

§. 6. Optimum Sampling System の近似的決定

前§のφが第一項に対して省略できるなら、$\alpha(\Omega_p, f_p)$は近似的に

$$(16)\quad \frac{1}{(1-\rho)^2+\left(\frac{C_p}{C_f}-\rho\right)^2}$$

で与へられる。

今我々が一つの函数Pの知識を持ってるとするなら次の様にする事が可能である。

λを常数として

$$(17)\quad P^* = P + \lambda E_\Omega(P)$$

とおく

λが与へられるなら　当然P^*の知識は完全である。

然る時

$$(18)\quad E_\Omega(P^*) = (1+\lambda)E_\Omega(P)$$

$$(19)\quad V_\Omega(P^*) = V_\Omega(P)$$

(20) $\rho_{f \cdot p} = \rho_{f \cdot p}$

であるから

(21) $\alpha(\Omega_{p^*}, f_{p^*}) = \dfrac{1}{(1-\rho^2)+(\frac{C_p}{C_f}\frac{1}{1+\lambda}-\rho)} + \varphi$

となる。

依って

(22) $\lambda = \dfrac{1}{\rho}\cdot\dfrac{C_p}{C_f} - 1$

なる如く λ を決定すれば

(23) $\alpha(\Omega_{p^*}, f_{p^*}) = \dfrac{1}{1-\rho^2} + \varphi$

註　$p^* \geqq 0$ でなければならないので(22)で与へるλを用い得ない事がある。

之は Correlation Coefficient の大きさが ρ なる函数族に対して、Optimum Sampling System を近似的に与へるものである。

之を graph で説明するなら、次の様である。(graph は終頁へ)
横軸は ρ、縦軸は C_p/C_f である。然らば任意の函数、依って、Sampling System は多対一に此の平面上の一点で表現される。曲線群は Equi-efficiency α を与へる。Sampling System を表はす。

我々の知識は此の平面上の discrete な点の集りであるが、p^* の導入により、知識の範囲を斯る点を通り縦軸に平行な直線上に

近拡大する事が可能となる。

こゝで此の直線に切する Equi-efficiency 曲線 $\alpha=\frac{1}{1-\rho^2}$ との切点が近似的に Optimum Sampling System を与へるのである。

図に於て横軸と原点と(±1.2,)を通る直線上の System は Equally likely な Sampling System と同様な Efficiency を持つ System を表はしておる。

註　函数族{P}の correlation coefficient ρ の例へば linear form で correlation が maximum となる様にして P^* を導入すれば更に Efficient な System を与へ得る事は明らかである。

§.7. 後書

以上で Sampling System の考へ方を説明したのであるが最後に一つだけ附言しておきたい。

所論から明かな様に(8)からPを決定する事が山である。此の効果的な決定法に就いての研究が行はれなければならない。此の問題に関する御協力を期待する。

註　実例は Hansen - Hurwitz の考へを拡張した水野に依る

Sampling with probability proportionate to sizes が その一つである。之は統計数理研究所談話会。日本数学会に於て昨年報告した。何れ原稿にして発表する予定である。

統計数理研究所

麹町研究室にて

graph