Journal of the Arkansas Academy of Science

Volume 21 Article 17

1967

Characterization of the Bivariate Negative Binomial Distribution

James E. Dunn
University of Arkansas, Fayetteville

CHARACTERIZATION OF THE BIVARIATE NEGATIVE BINOMIAL DISTRIBUTION

James E. Dunn

INTRODUCTION

The univariate negative binomial distribution (also known as Pascal's distribution and the Polya-Eggenberger distribution under various reparameterizations) has recently been characterized by Bartko (1962). Its broad acceptance and applicability in such diverse areas as medicine, ecology, and engineering is evident from the references listed there. On the other hand, the corresponding joint distribution, the bivariate negative binomial, seems to have received only negligible attention with the principal applications having been made in studying accident proneness, c.f. Arbous and Kerrich (1951), Bates and Neyman (1952).

In trying to trace the history of this distribution, one becomes aware that no comprehensive study of the distribution apparently exists; in the context where it occurs, it appears as an intermediate step to some other result. Hence, the purpose of this paper is two-fold. First, it will be desirable to compile a list of properties which characterize the distribution. It is hoped that this availability will make it easier for applied scientists to examine their research for new applications of the bivariate negative binomial distribution. Second, since the estimation problem must certainly arise in applications, some new results, in particular the maximum likelihood (ML) solution, will appear here.

BIVARIATE PROBABILITY GENERATING FUNCTIONS

It is well known that use of probability generating functions provides a powerful tool in revealing properties of probability distributions. For example, one may consult Bailey (1964) or Feller (1957) for excellent discussions of their characteristics in the univariate case. However, it seems worthwhile to review some of their basic properties in the bivariate situation.

Definition: Let X, Y be jointly distributed, non-negative, integral valued random variables. If $Pr(X=x, Y=y) = p_{xy}$, then the associated bivariate p.g.f. is defined to be the power series transformation.

$$G(z_1, z_2) = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} p_{xy} z_1^x z_2^y. \tag{1}$$

Theorem 1: The univariate p.g.f. for the marginal distribution of X is given by $G_x(z) = G(z, 1)$; the univariate p.g.f. for the marginal distribution of Y is given by $G_y(z) = G(1, z)$.

Note that either z_1 or z_2, corresponding to X, or Y, is set equal to 1 and the subscript is dropped from the remaining Z.

Theorem 2: Let W = X + Y define a new random variable. The univariate p.g.f. for the marginal distribution of the sum W is $G_w(Z) = G(Z, Z)$.

Theorem 3: Given a bivariate p.g.f. $G(Z_1, Z_2)$, the terms of the corresponding distribution $\{p_{xy}\}$ may be obtained as

$$p_{00} = G(0, 0) \qquad \text{if } x = y = 0$$

$$p_{xy} = \frac{\left.\dfrac{\partial^{x+y} G(Z_1, Z_2)}{\partial Z_1^x \, \partial Z_2^y}\right|_{Z_1 = Z_2 = 0}}{x!\, y!} \qquad \text{if } x>0 \text{ or } y>0 \quad .$$

Theorem 4: We define the joint r,s factorial moment by

$$\overset{*}{\mu}_{[r,s]} = E\{X(X-1) \ldots (X-r+1)Y(Y-1) \ldots (Y-x+1)\}$$

where E denotes the expected value operator. Then

$$\overset{*}{\mu}_{[r,s]} = \left.\frac{\partial^{r+s} G(Z_1, Z_2)}{\partial Z_1^{r\cdot} \, \partial Z_2^s}\right|_{Z_1 = Z_2 = 1}$$

Proofs of these results follow by direct application of the indicated operations to the definition of the bivariate p.g.f.

BIVARIATE NEGATIVE BINOMIAL DISTRIBUTION

Since we shall need to make frequent references to the univariate negative binomial in what follows, we state the following:

Definition: We say that a non-negative, integral valued random variable V has a (univariate) negative binomial distribution if for parameters

$A>0, \ 0<P<1,$

$$Pr\{V=v\} = p_v = (1-P)^A \qquad \text{if } v=0$$

$$(1-P)^A \ \frac{A(A+1)\ldots(A+v-1)P^v}{v!} \qquad \text{if } v>0 \tag{2}$$

It is easily verified that the mean and variance in this case are respectively

$$E\{V\} = \frac{AP}{1-P} \ , \ var\{V\} = \frac{AP}{(1-P)^2} \quad . \tag{3}$$

The associated univariate p.g.f. is, by definition,

$$G_v(Z) = \sum_{v=0}^{\infty} p_v \, Z^v = \left(\frac{1-P}{1-PZ}\right)^A \quad . \tag{4}$$

Feller (op. cit.) gives the following

Definition: We say that non-negative, integral valued random variables X,Y have a bivariate negative binomial distribution if their joint p.g.f. is

$$G(Z_1, Z_2) = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} p_{xy} \, Z_1^x \, Z_2^y = \left(\frac{p_0}{1-p_1Z_1-p_2Z_2}\right)^a \tag{5}$$

where

$$a, \ p_0, \ p_1, \ p_2>0; \ p_0+p_1+p_2 = 1. \tag{6}$$

Applying theorem 3, we obtain terms of the distribution directly

$$p_{00} = p_0^a$$

$$p_{xy} = \frac{p_0^a}{x!y!} \left. \frac{\overbrace{(-a)(-a-1)\dots(-a-x+1)(-p_1)^x}^{x \text{ terms}}\overbrace{(-a-x)\dots(-a-x-y+1)(-p_2)^y}^{y \text{ terms}}}{(1-p_1Z_1-p_2Z_2)^{a+x+y}} \right|_{Z_1=Z_2=0}$$

$$= p_0^a \frac{a(a+1)\dots(a+x+y-1)}{x!y!} p_1^x p_2^y \text{ if } x \neq 0 \text{ or } y \neq 0 \tag{7}$$

Since the general expression for $\dfrac{\partial^{r+s} G(Z_1, Z_2)}{\partial Z_1^r \, \partial Z_2^s}$ appears as an intermediate step in obtaining (7), we can apply theorem 4 directly to obtain the joint r,s factorial moment, viz.

$$\mu'_{[r,s]} = a(a+1)\dots(a+r+s-1)(p_1/p_0)^r(p_2/p_0)^s \tag{8}$$

Marginal Distributions

Applying theorem 1 and recalling that $p_0 = 1-p_1-p_2$, we obtain the p.g.f. for the marginal distribution of X as

$$G_x(Z) = G(Z,1) = \left\{\frac{(1-p_2)-p_1}{(1-p_2)-p_1Z}\right\}^a = \left\{\frac{1-\left(\frac{p_1}{1-p_2}\right)}{1-\left(\frac{p_1}{1-p_2}\right)Z}\right\}^a$$

which, by identifying $p = \frac{p_1}{1-p_2}$ and A = a, is identical with (4). It follows from the uniqueness of the p.g.f. that the marginal distribution of X is negative binomial with respective mean and variance,

$$E(X) = \frac{ap_1}{p_0}, \quad var(X) = \frac{ap_1(p_0+p_1)}{p_0^2} \tag{9}$$

and probability distribution given by

$$Pr(X=0) = \left(1-\frac{p_1}{1-p_2}\right)^a = \left(\frac{p_0}{1-p_2}\right)^a \tag{10}$$

$$Pr(X=x) = \left(\frac{p_0}{1-p_2}\right)^a \frac{a(a+1)\dots(a+x-1)}{x!} \left(\frac{p_1}{1-p_2}\right)^x \quad \text{if } x>0 \ .$$

By symmetry of p_1Z_1 and p_2Z_2 in (5), it follows immediately that the marginal distribution of Y is also negative binomial with

$$E(Y) = \frac{ap_2}{p_0}, \quad var(Y) = \frac{ap_2(p_0+p_2)}{p_0^2} \tag{11}$$

and probability distribution given by

$$Pr(Y=0) = \left(1-\frac{p_2}{1-p_1}\right)^a = \left(\frac{p_0}{1-p_1}\right)^a$$

$$Pr(Y=y) = \left(\frac{p_0}{1-p_1}\right)^a \frac{a(a+1)\dots(a+y-1)}{y!} \left(\frac{p_2}{1-p_1}\right)^y \quad \text{if } y>0 \ . \tag{12}$$

The marginal distribution of the sum W = X + Y is equally simple since from theorem 1,

$$G_w(Z) = G(Z, Z) = \left\{\frac{1-(p_1+p_2)}{1-(p_1+p_2)Z}\right\}^a$$

which, by identifying $p = p_1+p_2$ and $A = a$, is again identical with (4). Hence, this marginal distribution is also negative binomial with respective mean and variance

$$E(W) = \frac{a(p_1+p_2)}{p_0} \quad , \quad var(W) = \frac{a(p_1+p_2)}{p_0^{\,2}}$$

and probability distribution

$$Pr(W=0) = (1-p_1-p_2)^a = p_0^{\,a}$$

$$Pr(W=w) = p_0^{\,2} \frac{a(a+1)\ldots(a+w-1)}{w!} (p_1+p_2)^k \quad \text{if } w>0$$

The existence of these three derived distributions is well-known, cf. Feller (op. cit.) but it is interesting to see their explicit functional forms.

Covariance and Correlation

Bivariate and multivariate distributions possess an additional interest over their univariate analogues inasmuch as they allow characterization of the association between random variables. Setting $r = s = 1$ in (8), we easily obtain

$$E(XY) = \frac{a(a+1)p_1p_2}{p_0^{\,2}}$$

from which, by definition, the covariance and correlation between X and Y are respectively

$$cov(X,Y) = E(XY) - E(X)E(Y) = \frac{ap_1p_2}{p_0} \tag{13}$$

$$corr(X,Y) = \frac{cov(X,Y)}{(var(X) \cdot var(Y))^{1/2}} = \left(\frac{p_1p_2}{(p_0+p_1)(p_0+p_2)}\right)^{1/2}$$

where means and variances of X and Y are given in (9) and (11). Obviously $0 \leq corr(X,Y) \leq 1$ where the lower bound is attained if $p_1=0$ or $p_2=0$ and the upper bound by setting $p_0=0$.

Conditional Distributions

Since conditional distributions form the theoretical basis of all regression analyses, it is informative to examine these properties in the special context of the bivariate negative binomial distribution. By definition, the conditional probability that $Y = y$ given that $X = x$, written $Pr(Y=y|X=x)$, is

$$Pr(Y=y|X=x) = \frac{Pr(Y=y,\ X=x)}{Pr(X=x)}$$

Hence, by taking the ratio of (7) and (10) and simplifying, one obtains the familiar expressions

$$Pr(Y=0|X=x) = (1-p_2)^{a+x} \tag{14}$$

$$Pr(Y=y|X=x) = (1-p_2)^{a+x} \frac{(a+x)(a+x+1)\ldots(a+x+y-1)}{y!} p_2^{\,y} \quad \text{if } y>0$$

i.e. the conditional distribution of Y is negative binomial. By identifying $P=p_2$ and $A=a+x$ in (2), it follows immediately that the conditional mean and variance of Y are respectively

$$E(Y|X=x) = \frac{p_2(a+x)}{1-p_2} = \frac{ap_2}{1-p_2} + \left(\frac{p_2}{1-p_2}\right)x \tag{15}$$

$$var(Y|X=x) = \frac{p_2(a+x)}{(1-p_2)^2} \tag{16}$$

with p.g.f. given by

$$G_{y|x}(Z) = \left(\frac{1-p_2}{1-p_2 Z}\right)^{a+x} \qquad (17)$$

By symmetry of X with Y and p_1 with p_2, it follows that the conditional distribution of X given Y = y is also negative binomial with

$$Pr\{X=0|Y=y\} = (1-p_1)^{a+y} \qquad (18)$$

$$Pr\{X=x|Y=y\} = (1-p_1)^{a+y} \frac{(a+y)(a+y+1)\ldots(a+y+x-1)p_1^{\,x}}{x!} \qquad x>0$$

with

$$E(X|Y=y) = \frac{p_1(a+y)}{1-p_1} = \frac{ap_1}{1-p_1} + \left(\frac{p_1}{1-p_1}\right) y \qquad (19)$$

$$var(X|Y=y) = \frac{p_1(a+y)}{(1-p_1)^2} \qquad (20)$$

$$G_{x|y}(Z) = \left(\frac{1-p_1}{1-p_1 Z}\right)^{a+y} \qquad (21)$$

Arbous and Kerrich (op. cit.) were aware of the nature of these conditional distributions but their rather cumbersome notation involving several exponential terms gives a foreboding appearance to inherently simple formulae.

Expressions (15) and (19) are fundamental to regression analysis inasmuch as they provide the classical prediction models, i.e. regression of the mean for y given x and for x given y respectively. In each case, they are written in two forms to emphasize the linearity of the conditional mean on the conditional variable. Under these conditions, as Meyer (1965) points out, we might equivalently have written

$$E(Y|X=x) = E(Y) + \frac{cov(X,Y)}{var(X)} \cdot (x-E(X)) \qquad (22)$$

and

$$E(X|Y=y) = E(X) + \frac{cov(X,Y)}{var(Y)} \cdot (y-E(Y)), \qquad (23)$$

these relations being easily verified by substituting the results of (9), (11), and (13) in the above expressions and comparing the results to (15) and (19).

Intuitively, one might suppose that since expressions (22) and (23) hold for the bivariate normal distribution, c.f. Fraser (1958), and also, as we have shown, for the bivariate negative binomial, that they hold for all bivariate distributions. Meyer (op. cit.) gives several counter examples. In actual fact, one must verify in each case, as we have done, that these regression models are linear. Of course, in the process, we have revealed the salient feature that each of the conditional distributions is univariate negative binomial.

ESTIMATION

The existence of the regression models shown in (15) and (19) immediately suggests the practical question of trying to estimate the unknown parameters of the models from experimental data. In what follows, let us suppose that $x_1, y_1; x_2, y_2; \ldots; x_n, y_n$ is a random sample of n pairs drawn from a common bivariate negative binomial distribution. The immediate problem is to estimate p_0, p_1, p_2, and a defined in (5), (6), and (7).

Graphical Solution

(a) Plot y as a function of x. Draw a "sight line" (or fit by least squares). Call it $y = G + Hx$.

(b) Plot x as a function of y. Draw a "sight line" and call it $x = K + Ly$.

(c) Equating the constants of these fitted lines to their equivalents in (15) and (19), i.e. estimated intercepts to theoretical intercepts, etc., and solving for the unknowns yields estimates

$$\hat{p}_1 = \frac{L}{1+L} \qquad \hat{p}_2 = \frac{H}{1+H} \tag{24}$$

$$\hat{a} = \frac{(1-\hat{p}_1)K}{\hat{p}_1} \quad \text{or} \quad \hat{a} = \frac{(1-\hat{p}_2)G}{\hat{p}_2} \tag{25}$$

$$\hat{p}_0 = 1-\hat{p}_1-\hat{p}_2$$

Moment Estimation

Let

$$m_x = \frac{\sum_{i=1}^{n} x_i}{n}, \quad m_y = \frac{\sum_{i=1}^{n} y_i}{n} \qquad m_{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n}. \tag{26}$$

Equating these sample moments to their population equivalents $E(X)$, $E(Y)$, $E(XY)$ given previously and solving for the unknowns, we obtain the moment estimators

$$\hat{a} = \frac{m_x m_y}{m_{xy}-m_x m_y} \qquad \hat{p}_1 = \frac{m_x}{\hat{a}+m_x+m_y} \tag{27}$$

$$\hat{p}_2 = \frac{m_y}{\hat{a}+m_x+m_y} \qquad \hat{p}_0 = 1-\hat{p}_1-\hat{p}_2$$

Arbous and Kerrich (op. cit.) used moment estimates in their example. However, a profusion of exponential terms makes their expressions look formidable compared with those given above.

Maximum Likelihood Estimation

Suppose $X=0$ and $Y=0$ for n_0 pairs of observations and that either $X\neq 0$ or $Y\neq 0$ for the remaining $n-n_0=r$ pairs of observations. Using (7), the joint likelihood function is

$$L(a_1,p_1,p_2)=(1-p_1-p_2)^{na} \prod_{i=1}^{r} \frac{p_1^{x_i} p_2^{y_i} a(a+1)\ldots(a+x_i+y_i-1)}{x_i!\,y_i!}$$

where the product involves only those pairs where either $x_i\neq 0$ or $y_i\neq 0$. The maximum likelihood (ML) estimators are those values of a, p_1, p_2 which maximize $L(a, p_1, p_2)$ or equivalently $\ln L(a,p_1,p_2)$. We obtain

Journal of the Arkansas Academy of Science, Vol. 21 [1967], Art. 17

$$\frac{\partial \ln L}{\partial p_1} = \frac{-na}{1-p_1-p_2} + \frac{nm_x}{p_1}$$

$$\frac{\partial \ln L}{\partial p_2} = \frac{-na}{1-p_1-p_2} + \frac{nm_y}{p_2}$$

$$\frac{\partial \ln L}{\partial a} = n \ln(1-p_1-p_2) + \sum_{i=1}^{r} \sum_{j=1}^{x_i+y_i} (a+j-1)^{-1}$$

Equating each of these to zero, we obtain the ML estimators as

$$\hat{p}_1 = \frac{m_x}{\hat{a}+m_x+m_y}, \quad \hat{p}_2 = \frac{m_y}{\hat{a}+m_x+m_y} \tag{28}$$

where $\hat{a}$ is the positive solution of

$$f(a) = n \ln\left(\frac{a}{a+m_x+m_y}\right) + \sum_{i=1}^{r} \sum_{j=1}^{x_i+y_i} (a+j-1)^{-1} = 0. \tag{29}$$

Newton's iteration $\hat{a}_{k+1} = \hat{a}_k - \frac{f(\hat{a}_k)}{f'(\hat{a}_k)}$ is useful in solving (29) where $f'(a) = \frac{n(m_x+m_y)}{a(a+m_x+m_y)} - \sum_{i=1}^{r} \sum_{j=1}^{x_i+y_i} (a+j-1)^{-2}$ and either moment estimates or the graphical solution may be used as a starting point of the iteration.

If

$$R = \begin{bmatrix} \frac{a(np_1+rp_0)}{p_0^2 p_1} & \frac{an}{p_0^2} & \frac{n}{p_0} \\ \text{sym.} & \frac{a(np_2+rp_0)}{p_0^2} & \frac{n}{p_0} \\ & & J \end{bmatrix}$$

where $J = rp_0^a \sum_{k=1}^{\infty} \sum_{j=1}^{k} \frac{\Gamma(a+k)(p_1+p_2)^k}{\Gamma(a)\Gamma(k+1)(a+j-1)^2}$, the asymptotic variance — covariance matrix for the ML estimators is given by

$$\text{var}\left\{\begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \\ \hat{a} \end{bmatrix}\right\} = R^{-1} \quad .$$

Having shown the feasibility of the ML solution, we are then in a position to suggest tests of hypothesis based on Wilks' (1962) likelihood ratio criterion. For example, suppose we wish to know if the means of X and Y are identical. Formally, H_0: $E(X) = E(Y)$ where the logical alternative is H_1: $E(X) \neq E(Y)$. But from (9) and (11), we see that H_0 is true only if $p_1 = p_2$. Hence, the hypotheses may be restated as H_0: $p_1 = p_2$, H_1: $p_1 \neq p_2$. We have already indicated in (28) and (29) the ML solution in the unconstrained parameter space under H_1. Let $L(\hat{a}, \hat{p}_1, \hat{p}_2)$ denote the value of the likelihood function at this solution.

In the constrained parameter space where H_0 is true, i.e. where $p_1 = p_2 = p$ say, after equating derivatives to zero we see that the likelihood function is maximized by choosing

$$\hat{p} = \frac{m_x+m_y}{2(\hat{a}+m_y+m_y)} \tag{30}$$

where $\hat{\hat{a}}$ again is the positive solution of (29), i.e. $\hat{\hat{a}}$ and $\hat{a}$ are identical. From (30), we see that

$$\hat{p} = \frac{\hat{p}_1 + \hat{p}_2}{2}$$

i.e. simply take the average of the estimates of $\hat{p}_1$ and $\hat{p}_2$. Let $L(\hat{a},\hat{p},\hat{p})$ be the value of the likelihood function at this solution. Wilks' likelihood ratio procedure suggests that when H_0 is true, test statistic

$$T = -2 \ln[L(\hat{a},\hat{p},\hat{p})/L(a,\hat{p}_1,\hat{p}_2)] \quad (31)$$

has an asymptotic chi-square distribution with one degree of freedom. Hence, reject H_0 at the a-level of significance of $T > X^2_{(1)1-a}$.

Example

Table 1 gives data reproduced from Arbous and Kerrich (1951) involving accident proneness of 122 experienced railroad shunters. The columns (y) refer to the number of accidents suffered by an individual in the 6 year period 1937-42; the rows (x) refer to the number suffered in the following 5 years 1943-47. Table entries indicate the number of individuals suffering a particular combination of accident rates.

Table 1: Accidents among 122 experienced railroad shunters (25 years experience), from Adelstein (1951).

1937-42 (6 years)

		0	1	2	3	4	5	6	
	0	21	18	8	2	1	0	0	50
	1	13	14	10	1	4	1	0	43
	2	4	5	4	2	1	0	1	17
1943-47	3	2	1	3	2	0	1	0	9
	4	0	0	1	1	0	0	0	2
	5	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0
	7	0	1	0	0	0	0	0	1
		40	39	26	8	6	2	1	122

Table 2 gives estimates of a, p_0, p_1 and p_2 obtained from the procedures just obtained. In the graphical solution, we actually fitted the regression of x on y and y on x by least squares as an intermediate step and obtained

$$y = 0.98574 + 0.29193\,x \quad (32)$$

$$x = 0.67986 + 0.23263\,y$$

Table 2: Comparison of estimates by three methods.

	p_0	p_1	p_2	a
Graphical	0.58531	0.18873	0.22596	2.92242 3.37671
Moment	0.59505	0.17587	0.22908	3.30023
Maximum Likelihood	0.60361	0.17215	0.22423	3.42002

Two estimates of a result from the graphical solution as indicated in (25). No clear guide for choosing one over the other is evident. Moment and ML estimates are in close agreement, though from our knowledge of estimation problems in the univariate case, we suspect that this may be a coincidence, that in fact the two estimation procedures may lead to quite different estimates depending on the actual values of the parameters. Usually the ML estimates are more efficient; i.e. smaller variances, and hence are preferred where available.

Since the invariance principle holds for our ML estimates of a, p_0, p_1, p_2, we obtain ML estimates of various functions of these parameters directly by simply replacing parameters with the corresponding estimates. Estimates of some of the more interesting functions are shown in table 3.

Table 3. Maximum Likelihood Estimates of Various Functions

Function	Estimate
$E(X)$	0.97539
$var(X)$	1.25357
$E(Y)$	1.27047
$var(Y)$	1.74243
$cov(X,Y)$	0.36234
$corr(X,Y)$	0.77529
$E(Y\|X=x)$	0.98853 + 0.28904x
$E(X\|y=y)$	0.71119 + 0.20795y
$E(Y\|X=2)$	1.56666
$E(X\|Y=2)$	1.1271

The ML estimates of the regression lines may be compared with the corresponding least squares estimates given in (32). The discrepancy does not look too serious for most purposes.

To test H_0: $E(X) = E(Y)$, we calculate the following quantities:

$-\ln L(\hat{a},\hat{p}_1,\hat{p}_2) = -341.77495$

$\hat{p} = 0.19819$

$-\ln L(\hat{a},\hat{p},\hat{p}) = -13245.082$

$T = 25807.$

Since $Pr(\chi^2_{(1)} > 25807) < 10^{-5}$, we conclude that there is strong evidence of a location shift in accident frequency from Y (1937) to X (1943-47). Examination of the estimates in table 3 suggests the shift is $E(Y) > E(X)$, i.e. the over-all rate of accidents has been reduced. One might suspect that new safety innovations may be responsible for this change.

SUMMARY

So far as the author is aware, the bivariate negative binomial is the only joint distribution other than the bivariate normal in which both marginal random variables, both conditional random variables, and the sum of the random variables all have a common probability distribution, namely the univariate analogue of the bivariate distribution. We have shown that the correct models for regression of the mean are linear in both cases and have indicated a choice of estimation procedures. In addition to the indicated test for shift in means, one would still like to formulate a likelihood ratio test for independence. Though the test seems plausible, the method of approach is not evident.

LITERATURE CITED

Adelstein, A. M. 1951. Unpublished doctoral dissertation on accident rates of railway operatives. Univ. of the Witwatersrand S. A.

Arbous, A. G. and J. E. Kerrich. 1951. Accident statistics and the concept of accident-proneness. Biometrics 7:340-432.

Bailey, N. T. J. 1964. The elements of stochastic processes. John Wiley & Sons, Inc. New York. Pp. 5-15.

Bartko, J. J. 1962. A note on the negative binomial distribution. Technometrics 4: 609-610.

Bates, G. E. and J. Neyman. 1952. University of California Publications in Statistics, Vol. I.

Feller, W. 1957. An Introduction to Probability Theory and Its Applications. John Wiley & Sons, Inc.: New York. Page 267.

Meyer, P. L. 1965. Introductory Probability and Statistical Applications. Addison-Wesley Reading, Massachusetts. Pp. 137-140.

Wilks, S. S. 1962. Mathematical Statistics. John Wiley & Sons: New York. Pp. 419-422.