# SKEWNESS FOR MAXIMUM LIKELIHOOD ESTIMATORS OF THE NEGATIVE BINOMIAL DISTRIBUTION

K.O. BOWMAN[1] and L.R. SHENTON[2]

[1]Computational Sciences and Engineering Division
Oak Ridge National Laboratory, P.O.Box 2008, 4500N, MS-6191
Oak Ridge, TN 37831-6191, U.S.A.
e-mail: bowmanko@ornl.gov
[2]Department of Statistics, University of Georgia
Athens, Georgia 30602, U.S.A.

July 13, 2007

## Abstract

The probability generating function of one version of the negative binomial distribution being $(p + 1 - pt)^{-k}$, we study elements of the Hessian and in particular Fisher's discovery of a series form for the variance of $\hat{k}$, the maximum likelihood estimator, and also for the determinant of the Hessian. There is a link with the Psi function and its derivatives. Basic algebra is excessively complicated and a Maple code implementation is an important task in the solution process. Low order maximum likelihood moments are given and also Fisher's examples relating to data associated with ticks on sheep. Efficiency of moment estimators is mentioned, including the concept of joint efficiency. In an Addendum we give an interesting formula for the difference of two Psi functions.

2000 Mathematics Subject Classification: 62E20.
*Key words and phrases:* asymptotic skewness, asymptotic variance, Hessian, iterative schemes, Maple code, Psi function.

# 1 Introduction

In Fisher's *Contributions to Mathematical Statistics* (1950), the 38th paper, a short one, considers certain aspects of the negative binomial distribution, moment and maximum likelihood methods. The paper was written some sixty or so years ago (1941), so that in the ensuring years there have been notation modifications. For example, the Hessian matrix was referred to as the information matrix; also Fisher preferred to use factorial notation rather than the gamma function, for example, $(\frac{1}{2})!$

We shall use $i_{kk}$, $i_{kp}$, $i_{pp}$ for elements in the information matrix, the probability function for the negative binomial random variable being

$$Pr(X = x) = (q)^{-k} \left(\frac{p}{q}\right)^x \frac{\Gamma(k+x)}{\Gamma(k)x!} \quad (k > 0, p > 0, q = 1 + p)$$

and $x = 0, 1, \cdots$. We recall the basic moments

$$\text{Mean}: \ \mu_1' = kp, \quad \text{Variance}: \ \mu_2 = kpq.$$

In this paper we have remarks to make on Fisher's expression for the variance of the maximum likelihood estimator $\hat{k}$; but more importantly for the mathematical structure of the asymptotic form of the skewness of the distribution of $\hat{k}$; this takes the form $\sqrt{\beta_{11}(\hat{k})}$, the coefficient of $1/\sqrt{N}$ in the skewness, $N$ the sample size.

Maximum likelihood estimators of the parameters $k$, $p$ for the negative binomial distribution are fundamentally based on the Psi function and its derivatives. Actually $\psi(k + x) - \psi(k)$ and its derivatives turn up frequently.

We shall also mention Fisher's idea of efficiency for a single estimator and jointly. Data based examples are included.

# 2 Fisher's expression for $i_{kk}$

## 2.1 The formulas

The third central moment of the estimator $\hat{\theta}_a$ is (Bowman and Shenton, 1998, 1999)

$$\mu_{32}(\hat{\theta}_a) = L^{a\alpha} L^{a\beta} L^{a\gamma} \{[\theta_\alpha, \theta_\beta, \theta_\gamma] + 3[\theta_\alpha \theta_\beta \theta_\gamma] + 6[\theta_\alpha \theta_\beta, \theta_\gamma]\} \tag{1}$$

where subscript 32 indicate third moment, coefficient of $N^{-2}$, $[\alpha\beta\gamma] = E\left(\frac{\partial^3 \ln(Pr)}{\partial\alpha\partial\beta\partial\gamma}\right)$, $[\alpha\beta, \gamma] = E\left(\frac{\partial^2 \ln(Pr)}{\partial\alpha\partial\beta} \frac{\partial \ln(Pr)}{\partial\gamma}\right)$, $[\alpha, \beta, \gamma] = E\left(\frac{\partial \ln(Pr)}{\partial\alpha} \frac{\partial \ln(Pr)}{\partial\beta} \frac{\partial \ln(Pr)}{\partial\gamma}\right)$, summed for $\alpha$, $\beta$,

$\gamma$ having the values $1, 2, \cdots, s$ in the $s$-parameter case. We shall define the square bracket terms as $T_r(\underline{\theta})$, $\underline{\theta}$ referring to the parameters space.

For the Hessian, we have

$$H = \begin{bmatrix} i_{kk} & i_{kp} \\ i_{pk} & i_{pp} \end{bmatrix}.$$

The second derivatives of $\log(Pr)$ with respect to $k$ and $p$ are

$$\frac{\partial^2 \ln(Pr)}{\partial k^2} = \psi_1(k+x) - \psi_1(k),$$
$$\frac{\partial^2 \ln(Pr)}{\partial k \partial p} = -\frac{1}{q} = \frac{\partial^2 \ln(Pr)}{\partial p \partial k}, \qquad \frac{\partial^2 \ln(Pr)}{\partial p^2} = \frac{k}{q^2} - \frac{x(2p+1)}{p^2 q^2}.$$

For the element in the Hessian matrix $i_{kk}$

$$i_{kk} = \sum_0^\infty \frac{1}{q^k} \left(\frac{p}{q}\right)^x \frac{\Gamma(k+x)}{\Gamma(k)x!} [\psi_1(k) - \psi_1(k+x)]$$

$\psi_1$ being the derivative of the Psi function. Fisher defined $r = \frac{p}{p+1} = \frac{p}{q}$, and inserted $\frac{1}{k^2} + \frac{1}{(k+1)^2} + \cdots + \frac{1}{(k+x-1)^2}$ for $\psi_1(k) - \psi_1(k+x)$.

Thus

$$i_{kk} = (1-r)^k \left\{ rk\left(\frac{1}{k^2}\right) + \frac{r^2}{2!}k(k+1)\left(\frac{1}{k^2} + \frac{1}{(k+1)^2}\right) \right.$$
$$\left. + \frac{r^3}{3!}k(k+1)(k+2)\left(\frac{1}{k^2} + \frac{1}{(k+1)^2} + \frac{1}{(k+2)^2}\right) + \cdots \right\}$$

and Fisher goes on to expand this in powers of $r$ $(0 < r < 1)$, giving

$$i_{kk} = \frac{r}{k} + \frac{r^2}{2k(k+1)} + \frac{4r^3}{6k(k+1)(k+2)} + \cdots \tag{2}$$
$$= \sum_{x=1}^\infty \frac{r^x}{x} \frac{(x-1)!\Gamma(k)}{\Gamma(k+x)}. \tag{3}$$

## 2.2 Validity of $i_{kk}$

We can only guess that Fisher used the pattern of terms displayed in (2) to extend to the series in (3). We have been unable to produce a mathematical proof. However the Maple code has polygamma functions as standard mathematical functions so we have checked out (3) for $i_{kk}$ up to the coefficient of $r^{10}$. For the term in $r^6$, Maple gives the following output:

$$
\begin{aligned}
\text{Coeff. of } r^6 = {} & \frac{1}{6k} + \frac{5k}{16} - \frac{17k^2}{144} + \frac{k^3}{48} - \frac{k^4}{720} + \frac{k}{16\,(k+1)^2} - \frac{13k^2}{72\,(k+1)^2} \\
& - \frac{k^2}{18\,(k+2)^2} + \frac{k}{6\,(k+2)^2} - \frac{5k^3}{24\,(k+2)^2} - \frac{5k^3}{48\,(k+1)^2} - \frac{5k^3}{24\,(k+3)^2} \\
& + \frac{k^2}{18\,(k+3)^2} + \frac{k}{6\,(k+3)^2} + \frac{25k^4}{144\,(k+1)^2} - \frac{5k^4}{72\,(k+2)^2} - \frac{5k^4}{72\,(k+3)^2} \\
& - \frac{k^5}{16\,(k+1)^2} + \frac{k^5}{24\,(k+2)^2} + \frac{k^5}{24\,(k+3)^2} - \frac{25k^4}{144\,(k+4)^2} - \frac{k^5}{48\,(k+4)^2} \\
& + \frac{13k^2}{72\,(k+4)^2} + \frac{k}{6\,(k+4)^2} - \frac{k^5}{16\,(k+4)^2} - \frac{k^5}{48\,(k+5)^2} - \frac{17k^4}{144(k+5)^2} \\
& + \frac{5k^4}{(k+2)^2} + \frac{k^5}{48\,(k+3)^2} + \frac{17k^4}{144\,(k+5)^2} + \frac{5k^3}{16\,(k+5)^2} + \frac{137k^2}{360\,(k+5)^2} \\
& + \frac{k}{6\,(k+5)^2} + \frac{k^6}{144\,(k+1)^2} - \frac{k^6}{72\,(k+2)^2} + \frac{k^6}{72\,(k+3)^2} - \frac{k^6}{144\,(k+4)^2} \\
& + \frac{k^6}{720\,(k+5)^2} - \frac{137}{360} \\
= {} & \frac{20}{(k+5)(k+4)(k+3)(k+2)(k+1)k}.
\end{aligned}
$$

Fisher also stated the results

$$i_{pp} = \frac{k}{pq}, \quad i_{kp} = -\frac{1}{q} \quad \text{and} \quad |H| = \frac{k}{pq} \sum_{x=2}^{\infty} \frac{r^x}{x} \frac{(x-1)!\Gamma(k)}{\Gamma(k+x)}.$$

It is gratifying to see that our low order moments of maximum likelihood estimators found using Maple code (Bowman and Shenton, 2005) check out against Fisher's values.

Risking criticism because of repetition, we think interesting to high light the relation

$$
\begin{aligned}
i_{kk} &= \sum_{x=0}^{\infty} \frac{1}{q^k} \left(\frac{p}{q}\right)^x \frac{\Gamma(k+x)}{\Gamma(k)} \{\psi_1(k) - \psi_1(k+x)\} \\
&= \sum_{x=1}^{\infty} \frac{r^x (x-1)!\Gamma(k+x)}{x\Gamma(k)}, \qquad (r = p/q, 0 < r < 1, k > 0) \qquad (4)
\end{aligned}
$$

a type of generating function for the Psi function derivative. The reader will note the inversion of $\Gamma(k+x)/\Gamma(k)$. **A strict mathematical proof of (4) has not been discovered**

# 3 Terms in skewness

## 3.1 Square bracket terms

The formulas for the third central moment of the estimator $\hat{\theta}_a$ is given in (1).

A complete listing of square bracket terms, third, forth, and fifth is given in Bowman and Shenton (1965), where the symbol 1 refers to $\hat{k}$, the symbol 2 to $\hat{p}$. For examples,

$$[kkk] = 2E[S_3(x)], \quad [kkp] = 0, \quad [kpp] = 1/q^2, \quad [ppp] = 2k(1+2p)/(p^2q^2),$$
$$[kk,k] = E[(x-kp)S_2(x)]/(pq), \qquad [kk,p] = -E[(x-kp)S_2(x)]/(pq),$$
$$[kp,k] = [kp,p] = 0, \quad [pp,k] = -(1+2p)/(pq^2), \quad [pp,p] = -k(1+2p)/(p^2q^2),$$
$$[k,k,k] = E[S_1(x) - \ln(q)]^3, \qquad [k,k,p] = E[(x-kp)S_2(x)]/(pq),$$
$$[k,p,p] = 1/(pq), \qquad [p,p,p] = k(1+2p)/(p^2q^2).$$

Here

$$S_\lambda(x) = \frac{1}{k^\lambda} + \frac{1}{(k+1)^\lambda} + \cdots + \frac{1}{(k+x-1)^\lambda}, \qquad (x = 1, 2, \cdots, \lambda = 1, 2, \cdots).$$

Using the Maple code we have set up expressions in terms of $r = p/q$, $0 < r < 1$, for the expectations of $S_1(x)$, $S_2(x)$, $S_1(x)^2$, $S_1(x)^3$, $S_1(x)S_2(x)$, $xS_1(x)$, $xS_2(x)$, $S_3(x)$. We shall discuss these components in the sequel.

For the present, note that

$$E[S_1(x)] = r + \frac{r^2}{2} + \frac{r^3}{3} + \cdots = \ln(q) = L, \qquad (0 < r < 1, k > 0)$$
$$E[S_2(x)] = \frac{r}{k} + \frac{r^2}{2k(k+1)} + \frac{2r^3}{3k(k+1)(k+2)} + \cdots = \sum_{x=1}^{\infty} \frac{r^x (x-1)!\Gamma(k)}{x\Gamma(k+x)},$$
$$E[S_1(x)^2] = E[S_2(x)] + L^2,$$
$$E[S_1(x)^3] = 3E[S_1(x)S_2(x)] - 2E[S_3(x)] + L^3.$$

## 3.2 The asymptotic skewness for $\hat{k}$

### 3.2.1 Basic forms

Using the Maple code implementation for the asymptotic skewness ($\sqrt{\beta_{11}(\hat{k})} = \mu_3/\mu_2^{3/2}$) of a maximum likelihood estimator we have found approximations to $\sqrt{\beta_{11}(\hat{k})}$.

$$\sqrt{\beta_{11}(\hat{k})} = [Var_1(\hat{k})]^{3/2}\left\{-E(S_1^3+3S_2L+L^3+2S_3)+\frac{3E(x-pk)S_2}{kq}+\frac{2p(p+2)}{k^2q^2}\right\} \quad (5)$$

where

$$\begin{aligned} E(S_1) &= \ln(q) = L, \\ E(xS_1) &= p + pkL, \\ E(S_1^2 - S_2) &= L^2, \\ 2E(S_3) &= E(S_1S_2 - S_1^3 + L^3), \\ E(x-kp)S_1^2 &= E[(x-kp)S_2 + 2pL]. \end{aligned} \quad (6)$$

These expressions may be derived from

$$\sum_{x=1}^{\infty}\frac{r^x}{x!}\Gamma(k+x) = (1-r)^{-k}\Gamma(k)$$

by differentiations. For examples

$$(1-r)^k\sum_{1}^{\infty}\frac{r^x}{x!}S_2\Gamma(k+x) = \Gamma(k)E(S_2)$$

Differentiate with respect to $k$, so that

$$\begin{aligned} -L\Gamma(k)E[S_2(x)] + (1-r)^k\sum_{x=1}^{\infty}\frac{r^x}{x!}\Gamma(k+x)\{-2E[S_3(x)]\} \\ + (1-r)^k\sum_{x=1}^{\infty}\frac{r^x}{x!}\Gamma(k+x)\{E[S_1(x)S_2(x)]\} \\ = \Gamma'(k)E[S_2(x)] + \Gamma(k)\frac{\partial E[S_2(x)]}{\partial k}. \end{aligned}$$

By adjustment we deduce the identity

$$E[S_1(x)S_2(x)] = LE[S_2(x)] + 2E[S_3(x)] + \frac{\partial E[S_2(x)]}{\partial k}, \quad (7)$$

where

$$E[S_2(x)] = \sum_{x=1}^{\infty}\frac{r^x(x-1)!\Gamma(k)}{x\Gamma(k+x)}, \quad (0 < r < 1, k > 0).$$

### 3.2.2 Identities for special cases of the parameter $k$

Example 1: The case $k = 1$.

Here the probability function is simple in form, but keep in mind we are still involved with sampling from the 2 parameter negative binomial distribution. For $0 < r < 1$, and $k = 1$,

$$E[S_1(x)] = r + \frac{r^2}{2} + \frac{r^3}{3} + \cdots = \ln\left(\frac{1}{1-r}\right) = \ln(q) = L,$$
$$E[S_2(x)] = \frac{r}{1^2} + \frac{r^2}{2^2} + \frac{r^3}{3^2} + \cdots = \sum_1^\infty \frac{r^x}{x^2}\left(< \frac{\pi^2}{6}\right) = \text{Fisher's } i_{kk},$$
$$E[S_3(x)] = \frac{r}{1^3} + \frac{r^2}{2^3} + \frac{r^3}{3^3} + \cdots = \sum_1^\infty \frac{r^x}{x^3} = \frac{(2\pi)^3}{12}\int_0^1 B(x)\cot(\pi x)dx.$$

The example above was conjectured from the pattern exhibited in $E[S_3(x)]$ expanded in power of $r$ to $r^5$.

Information matrix $H(k, p)$

$$|H(k,p)| = \frac{1}{pq}\sum_{x=2}^\infty \frac{r^x}{x^2}, \qquad \frac{\partial E[S_2(x)]}{\partial r} = \frac{1}{r}\sum_{x=1}^\infty \frac{r^x}{x^2}.$$

For $E[S_s(x)S_t(x)]$, $s \in |N$, $t \in |N$.

Coefficient $r = 1/k^{s+t-1}$

Coefficient $r^2 = \dfrac{k(k+1)}{2}\left(\dfrac{1}{k^{s+t}} + \dfrac{1}{k^s(k+1)^t} + \dfrac{1}{k^t(k+1)^s} + \dfrac{1}{(k+1)^{t+s}}\right) - \dfrac{1}{k^{s+t-1}}$ $(k > 0)$

### 3.2.3 Numerical examples in the general case

Table 1 Expectation of $S_2$ and $S_3$

| $k$ | $p$ | $E[S_2(x)]$ | $E[S_3(x)]$ | $k$ | $p$ | $E[S_2(x)]$ | $E[S_3(x)]$ |
|---|---|---|---|---|---|---|---|
| 1.0 | 0.1 | 0.093063 | 0.091971 | 10.0 | 0.1 | 0.009129 | 0.000878 |
| | 1.0 | 0.582241 | 0.537213 | | 1.0 | 0.051206 | 0.003994 |
| | 5.0 | 1.144978 | 0.955870 | | 5.0 | 0.086833 | 0.005340 |
| 5.0 | 0.1 | 0.018322 | 0.003547 | | | | |
| | 1.0 | 0.104632 | 0.016884 | | | | |
| | 5.0 | 0.180712 | 0.023384 | | | | |

Values of $E[S_2(x)]$ and $E[S_3(x)]$ arise in the next section.

# 4 Skewness $\sqrt{\beta_{11}(\hat{k})}$

## 4.1 Series in powers of $r$ for some terms

An examination of the power series for $E[S_1(x)]^3$, $E[S_1(x)S_2(x)]$, and $E[S_3(x)]$, only a few coefficients being available, indicates that only $E[S_2(x)]$ has coefficients of $k$ in descending order; thus

$$\text{coeff. of } r^2 \sim 1/(2!k^2), \quad \text{coeff. of } r^3 \sim 4/(3!k^2), \quad \text{coeff. of } r^4 \sim 36/(4!k^2).$$

This strongly suggests that an expression for $\sqrt{\beta_{11}(\hat{k})}$ would be more satisfactory if it could be expressed in terms of $E[S_3(x)]$, only and not involving $E[S_1(x)]^3$, $E[S_1(x)S_2(x)]$.

Now from §(3.1)

$$\begin{aligned}[kkk] &= 2E[S_3(x)], \quad [kk,k] = E[(x-kp)S_2(x)]/(pq), \\ [k,k,k] &= E[S_1(x)-L]^3.\end{aligned}$$

Now $\mu_3(\hat{k})$, from (1) involves these square bracket terms with symbols $\alpha$, $\beta$, $\gamma$. We define the sum of these as the triplet $T_r(\alpha\beta\gamma)$.

For example

$$\begin{aligned}T_r(kkk) &= [k,k,k] + 3\,[kkk] + 6\,[kk,k]\,, \\ T_r(kkp) &= [k,k,p] + 3\,[kkp] + 6\,[kk,p]\,, \\ T_r(kpp) &= [k,p,p] + 3\,[kpp] + 6\,[kp,p]\,, \\ T_r(ppp) &= [p,p,p] + 3\,[ppp] + 6\,[pp,p]\,.\end{aligned} \tag{8}$$

Each triplet $T_r$ is multiplied by triples of $L^{ij}$'s which we call $A$.

$$\begin{aligned}A(kkk) &= Var_1(\hat{k})^3, \quad A(kkp) = -Var_1(\hat{k})^3\frac{p}{k}, \\ A(kpp) &= Var_1(\hat{k})^3\frac{p^2}{k^2}, \quad A(ppp) = -Var_1(\hat{k})^3\frac{p^3}{k^3}.\end{aligned}$$

## 4.2 The term $T_r(kkk)$

Let

$$\begin{aligned}E[S_2(x)] &= (1-r)^k \sum_{x=1}^{\infty} \frac{r^x\Gamma(k+x)}{x!\Gamma(k)} S_2(x) \\ &= F_1(k,p).\end{aligned}$$

Then $[kk, k] = E[xS_2(x)] - kpF_1(k, p)$, and differentiating $F_1(k, p)$ with respect to $r$ yields

$$E[xS_2(x)] = rqkF_1(k, p) + r\frac{\partial F_1(k, p)}{\partial r}.$$

But $rq = r/(1 - r) = p$. Hence

$$[kk, k] = r\frac{\partial F_1(k, p)}{\partial r}$$

Again from §(3.1)

$$[k, k, k] = E[S_1(x) - L]^3 = E[S_1(x)]^3 - 3LE[S_1(x)]^2 + 2L^3$$

Now we can use the identities of (6) and (7). Namely

$$E[S_1(x)]^3 = -2E[S_3(x)] + 3E[S_1(x)S_2(x)] + L^3$$

(See Bowman and Shenton, 1965, p30).

Taking into account these terms we find

$$T_r(kkk) = -2E[S_3(x)] + \frac{\partial F_1(k, p)}{\partial k}.$$

with contribution $T_r(kkk)[Var_1(\hat{k})]^3$.

## 4.3 The remaining triplets in $\mu_{32}(\hat{k})$

$$\begin{aligned} T_r(kkp) &= \frac{3}{pq}E(x - kp)S_2(x) - \frac{6}{pq}E(x - kp)S_2(x) \\ &= -\frac{3}{pq}E(x - kp)S_2(x) = -\frac{3}{pq}r\frac{\partial F_1(k, p)}{\partial r}. \end{aligned}$$

Contribution is $(-p/k)T_r(kkp)[Var_1(\hat{k})]^3$.

Next

$$T_r(kpp) = \frac{3}{pq} + \frac{3}{q^2} - \frac{6(1 + 2p)}{pq^2} = \frac{3}{pq^2}(-1 - 2p)$$

Contribution is $(p^2/k^2)T_r(kpp)[Var_1(\hat{k})]^3$.

Lastly,

$$\begin{aligned} T_r(ppp) &= \frac{3k(1 + 2p)}{p^2q^2} + \frac{6k(1 + 2p)}{p^2q^2} - \frac{6k(1 + 2p)}{p^2q^2} \\ &= \frac{3k(1 + 2p)}{p^2q^2} \end{aligned}$$

Contribution is $(-p^3/k^3)T_r(ppp)[Var_1(\hat{k})]^3$.

### 4.4 Final form for skewness of $\hat{k}$

Adding the terms in the previous section for $\mu_{32}(\hat{k})$ we have

$$\sqrt{\beta_{11}(\hat{k})} = [Var_1(\hat{k})]^{3/2}\left\{-2E[S_3(x)] - 3\frac{\partial F_1(k,p)}{\partial k} + \frac{3r}{kp}\frac{\partial F_1(k,p)}{\partial r} - \frac{4r(1+r)}{k^2}\right\}$$

where

$$\frac{\partial F_1(k,p)}{\partial k} = -\sum_{x=1}^{\infty}\frac{r^x}{x}\frac{(x-1)!\Gamma(k)}{\Gamma(k+x)}[\psi(k+x)-\psi(k)],$$

and

$$r\frac{\partial F_1(k,p)}{\partial r} = \sum_{x=1}^{\infty}\frac{r^x(x-1)!\Gamma(k)}{\Gamma(k+x)} \quad (0<r<1, k>0)$$

and $\psi(k+x)-\psi(x) = \frac{1}{k}+\frac{1}{k+1}+\cdots+\frac{1}{k+x-1},\ \ x=1,2,\cdots$.

## 5 Skewness $\sqrt{\beta_{11}(\hat{p})}$

$\mu_{32}(\hat{p})$ ($N^{-2}$ terms in $\mu_3(\hat{p})$) has triplet $T_r(\alpha\beta\gamma)$ terms identical to $\hat{k}$ displayed in equation (8). Only multiplier triples $L^{ij}$ will be different.

$$B(kkk) = -Var_1(\hat{k})^3\frac{p^3}{k^3}, \quad B(kkp) = Var_1(\hat{k})^2\frac{p^2}{k^2}\left(\frac{pq}{k}+\frac{p^2}{k^2}Var_1(\hat{k})\right),$$

$$B(kpp) = -Var_1(\hat{k})\frac{p}{k}\left(\frac{pq}{k}+\frac{p^2}{k^2}Var_1(\hat{k})\right)^2, \quad B(ppp) = \left(\frac{pq}{k}+\frac{p^2}{k^2}Var_1(\hat{k})\right)^3.$$

The figure 1 display skewness of $\hat{k}$ and $\hat{p}$.

## 6 Efficiency of an estimator and joint efficiency

Fisher (1941) introduced the concept of efficient estimator in the form

$$E_f = \lim_{N\to\infty}\frac{Var\theta_1}{Var\theta_2}$$

Other comparisons may be considered, but keep in mind that estimators must be consistent, i.e. $E(\hat{\theta}) = \theta + \frac{\theta_1}{N} + \cdots$.

Fisher (1922) established the fact that maximum likelihood estimators have the highest efficiency; and that moment estimators, in comparison, always inefficient. It should however be remembered that high efficiency for maximum likelihood estimators

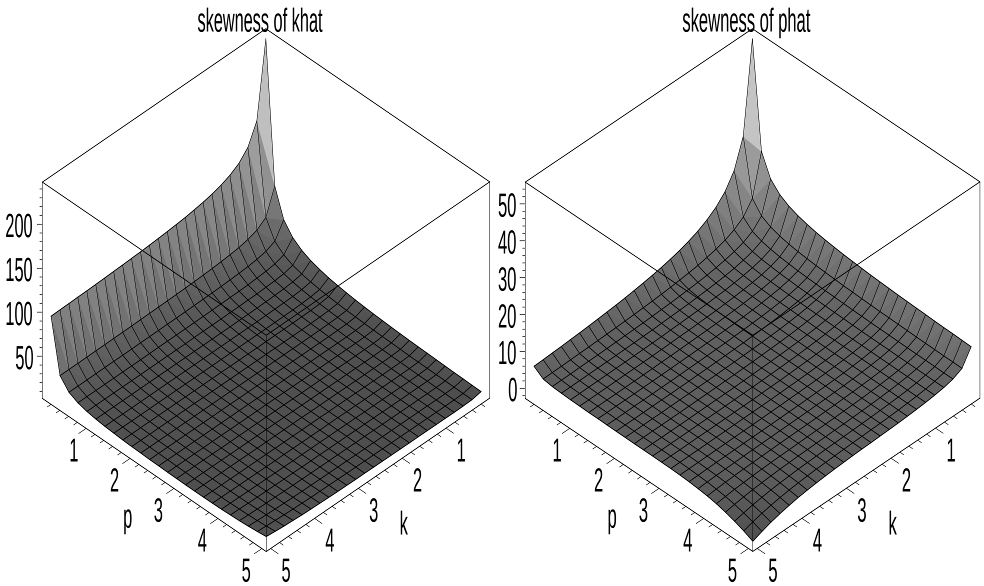


Figure 1: $\sqrt{\beta_{11}(\hat{k})}$ and $\sqrt{\beta_{11}(\hat{k})}$

does not in general imply other advantages. For example, there is no guarantee that high efficiency implies smaller value of the skewness.

Joint efficiency is described in Fisher (1941). For two parameters, for example,

$$\frac{1}{E_f^*} = \frac{\begin{vmatrix} Var_1(m_1) & Cov_1(m_1, m_2) \\ Cov_1(m_2, m_1) & Var_1(m_2^*) \end{vmatrix}}{\begin{vmatrix} Var_1(\hat{\theta}_1) & Cov_1(\hat{\theta}_1, \hat{\theta}_2) \\ Cov_1(\hat{\theta}_2, \hat{\theta}_1) & Var_1(\hat{\theta}_2) \end{vmatrix}}$$

where in the numeration the $m$'s refer to sample values. For the negative binomial model $(k, p)$,

$$\frac{1}{E_f^*} = \frac{\begin{vmatrix} Var_1(p^*) & Cov_1(p^*, k^*) \\ Cov_1(k^*, p^*) & Var_1(k^*) \end{vmatrix}}{\begin{vmatrix} Var_1(\hat{p}) & Cov_1(\hat{p}, \hat{k}) \\ Cov_1(\hat{k}, \hat{p}) & Var_1(\hat{k}) \end{vmatrix}}$$

Using moment estimators $k^*$ and $p^*$, the numerator is

$$\begin{vmatrix} q\left(2q + \frac{3p+2}{k}\right) & -\frac{2(k+1)q^2}{p} \\ -\frac{2(k+1)q^2}{p} & \frac{2k(k+1)q^2}{p^2} \end{vmatrix} = \frac{2(k+1)q^3}{p}$$

and the members of denominator are

$$Var_1(\hat{k}) = \frac{1}{\sum_{x=2}^{\infty} \frac{r^x}{x} \frac{(x-1)!\Gamma(k)}{\Gamma(k+x)}} \qquad Var_1(k^*) = \frac{2k(k+1)q^2}{p^2}$$

$$Var_1(\hat{p}) = \frac{pq}{k}\left(1 + \frac{r}{k\sum_{x=2}^{\infty} \frac{r^x}{x} \frac{(x-1)!\Gamma(k)}{\Gamma(k+x)}}\right) \qquad Var_1(p^*) = q\left(2q + \frac{3p+2}{k}\right)$$

$$Cov_1(\hat{k}, \hat{p}) = -\frac{p/k}{\sum_{x=2}^{\infty} \frac{r^x}{x} \frac{(x-1)!\Gamma(k)}{\Gamma(k+x)}} \qquad Cov_1(k^*, p^*) = -\frac{2(k+1)q^2}{p}$$

Hence the Fisher joint efficiency of maximum likelihood estimator against moment estimators is

$$\begin{aligned}\frac{1}{E_f} &= \left(\frac{2(k+1)q^3}{p}\right)\left(\frac{k}{qp}\right)\sum_{x=0}^{\infty} \frac{r^x}{x} \frac{(x-1)!\Gamma k}{\Gamma(k+x)} \\ &= \sum_{x=0}^{\infty} \frac{2r^x(x+1)!\Gamma(k+2)}{(x+2)\Gamma(k+x+2)}.\end{aligned}$$

Hence

$$E_f^* = \frac{1}{\sum_{x=2}^{\infty} \frac{2r^x(x+1)!\Gamma(k+2)}{(x+2)\Gamma(k+x+2)}}, \qquad (k > 0, 0 < r < 1)$$

the denominator a convergent power series in the ratio with positive terms, the first term being unity.

The reader may refer to Fisher's brief account in his paper (Fisher, 1941). Also see Mikulski (1982) and Shenton (1950).

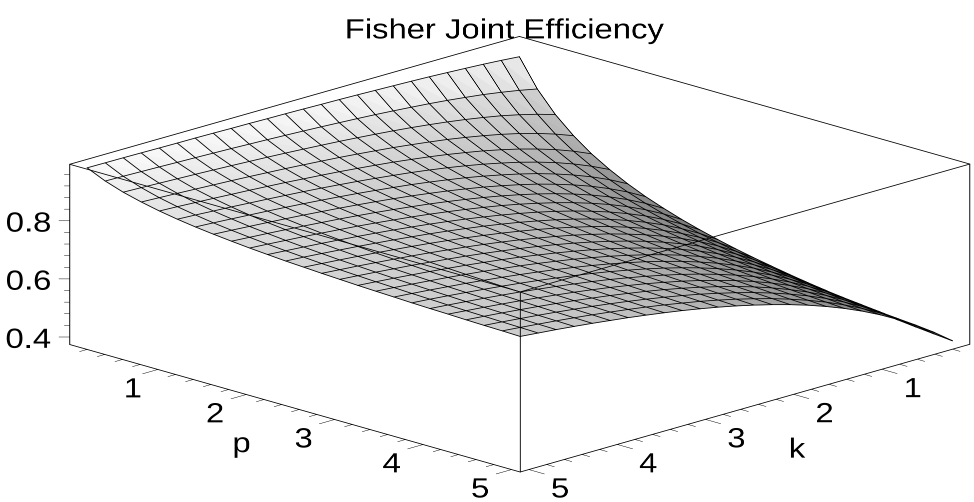


Figure 2: Joint efficiency of $k$ and $p$

# 7 Fisher's data sets

## 7.1 Data

These are given by Fisher (1941) and relate to two samples of sheep classified according to the number of ticks found on each sheep.

Table 2 The first example with sample size $N = 60$

| Number of ticks | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of sheep | 7 | 9 | 8 | 13 | 8 | 5 | 4 | 3 | 0 | 1 | 2 |

Fisher uses the method of moments to fit a negative binomial distribution. For the two moment estimators $k^*$, $p^*$ he uses the moments $\bar{x} = 3.25$, $m_2 = 349.25/59 = 5.9194915$ to lead to the estimates $p^* = 0.821382$, $k^* = 3.956740$. Note Fisher has used the divisor 59 in $m_2$ to avoid bias. There is little discussion since a criterion of efficiency sets this at about 90%.

Table 3 The second example with sample size $N = 82$

| Number of ticks | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 and more |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of sheep | 4 | 5 | 11 | 10 | 9 | 11 | 3 | 5 | 3 | 2 | 19 |

(More details are given in Fisher, Table 2).

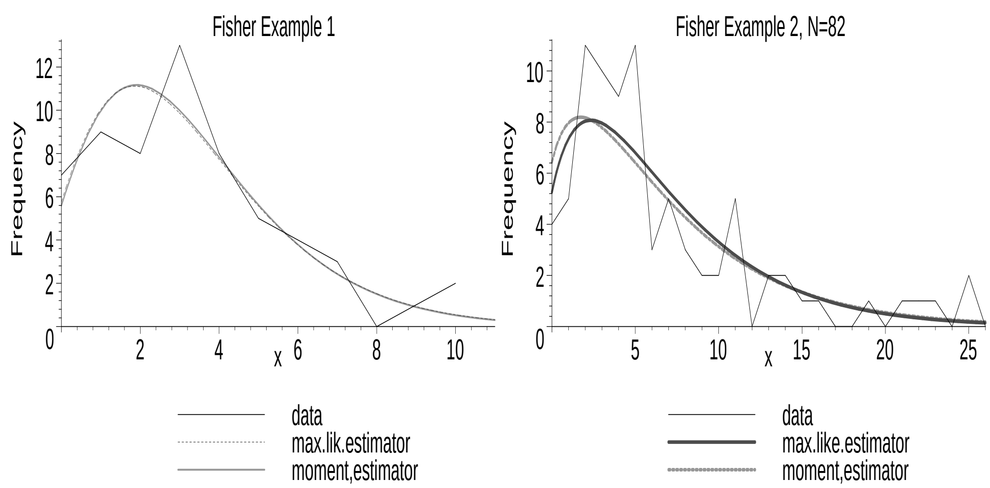


Figure 3: Fisher's examples of two cases

Maximum likelihood estimators $\hat{p}$ and $\hat{k}$ are

$$\frac{\partial \ln Pr}{\partial p} = 0 \quad \text{leads to} \quad \hat{p}\hat{k} = \bar{x} \tag{9}$$

where $\bar{x}$ is the sample mean. Using this

$$\frac{\partial \ln Pr}{\partial k} = -k\ln(p-1) + \psi(x+k) - \psi(k)$$

where $\psi(\cdot)$ is the Psi function. This leads to, using (9) the transcendental equation

$$\psi(\hat{k}) = \sum_{x=0}^{n} \frac{n_x}{N}\psi(\hat{k}+x) - \ln\left(1+\frac{\bar{x}}{\hat{k}}\right)$$

For a solution of $\hat{k}$ we use use

$$\begin{cases} f(k) = \psi(k) - \sum_{x=0}^{n} \frac{n_x}{N}\psi(k+x) + \ln\left(1+\frac{\bar{x}}{k}\right) \\ f'(k) = \psi_1(k) - \sum_{x=0}^{n} \frac{n_x}{N}\psi_1(k+x) - \frac{\bar{x}}{k(\bar{x}+k)} \end{cases}$$

and

$$_{new}\hat{k} = k - \frac{f(k)}{f'(k)} \quad \text{and} \quad _{new}\hat{p} = \frac{\bar{x}}{_{new}\hat{k}}.$$

The starting values for the iteration cycles, we use moment estimators

$$p^* = \frac{m_2}{\bar{x}} - 1, \qquad k^* = \frac{\bar{x}}{p^*}.$$

We have used the maximum likelihood estimator $\hat{k}$ and the mean to set up $\hat{p}$. It turns out the $\hat{k} = 1.77476$, $\hat{p} = 3.69175$, in exact agreement with Fisher (1950, page 38.186).

Table 4. Low order moments of $\hat{k}$ and $\hat{p}$

| | $\mu'_{11}$ | $\mu'_{12}/\mu'_{11}$ | $\mu'_{21}$ | $\mu'_{22}/\mu'_{21}$ | $\sigma$ | $\sqrt{\beta_1}$ | $\beta_2$ |
|---|---|---|---|---|---|---|---|
| Example 1 $N = 60$ | | | | | | | |
| $\hat{k} = 3.7513$ | 0.8664 | 0.3775 | 2.4552 | 1.1323 | 1.5669 | 2.0306 | 14.8144 |
| $\hat{p} = 0.8651$ | -0.2357 | -0.0029 | 0.1374 | 0.0109 | 0.3706 | 0.4920 | 6.1987 |
| Example 2 $N = 82$ | | | | | | | |
| $\hat{k} = 1.7775$ | 0.0923 | 0.0734 | 0.1326 | 0.2291 | 0.3642 | 0.8376 | 5.0285 |
| $\hat{p} = 3.6912$ | -0.0358 | -0.0109 | 0.6931 | 0.0108 | 0.8325 | 0.4547 | 3.8903 |

We consider the "goodness of fit" by considering the moment ratios. For example if for a statistic $t$,

$$E(t) \sim \tau_0 + \frac{\tau_1}{N} + \frac{\tau_2}{N^2} + \cdots \qquad (n \to \infty)$$
$$Var(t) \sim \frac{\mu_{21}}{N} + \frac{\mu_{22}}{N^2} + \cdots.$$

Then we look at $\tau_2/\tau_1$, and $\mu_{22}/\mu_{21}$. Terms omitted in these expressions are modified by introducing by taking $N$, the sample size, to be large enough.

For the estimator $\hat{k}$:

$$\tau_2/\tau_1 = 0.677, \quad \mu_{22}/\mu_{21} = 0.229.$$

In addition the asymptotic skewness is $\sqrt{\beta_1} = 0.83$ and asymptotic kurtosis, $\beta_2 = 5.0$. Looking at the four criteria, bias, variance, skewness, and kurtosis suggests a reasonable fit. It is not surprising the criteria for the estimator $\hat{p}$ is more promising. We have for $\hat{p}$:

$$\tau_2/\tau_1 = -0.010, \quad \mu_{22}/\mu_{21} = 0.011.$$
$$\sqrt{\beta_1} = 0.49, \quad \beta_2 = 3.8.$$

With 8 degrees of freedom, the $\chi^2$ value of 8.40 given by Fisher would in random sample only be less in about 40% cases.

# 8 Conclusion

This paper brings out a relation between maximum likelihood estimators for parameters of the negative binomial distribution and the Psi function and it derivatives. Basically, it depends on R.A. Fisher's discovery for a power series expression for $i_{kk}$, this being derived from the expectation of a second derivative of the logarithm of the probability function. Fisher's proof, as far as we can detect, depends on a deduction from a pattern in the first three terms. The underlying algebra is exceptionally complicated and far out of reach of pen and paper approaches, relying on the implementation of a Maple Code. The random variable $\frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{k+x-1}$, its derivatives with respect to $k$, and powers plays an essential role.

The asymptotic skewness of the estimator $\hat{k}$ (the negative binomial generating function being $(p + 1 - pt)^{-k}$) using the maximum likelihood approach, is stated in terms of $Var_1(\hat{k})$; similarly for the estimator $\hat{p}$.

It is interesting to note that in Fisher's 1941 paper he still adheres to the importance of sample efficiency, directed to moment estimators (Karl Pearson) against maximum likelihood estimators. Now let us recall Fisher's note on his "Foundations" paper (1922) in the Wiley (1950) publication. We quote Fisher.

*He did not clearly see, for example, that the variance of an estimate does not, in the theory of small samples, supply a satisfactory basis for comparison*

In our approach we consider low oder moments up to the fourth instead of efficiency and joint efficiency.

## Reference

Bowman, K. O. and Shenton, L. R. (1965), *Asymptotic Covariances for the Maximum Likelihood Estimators of the Parameters of a Negative Binomial Distribution*, K-1643.

Bowman, K.O. and Shenton, L.R. (1998), Asymptotic skewness and the distribution of maximum likelihood estimators, *Commun.Statist.Theory and Method*, 27(11), 2743-2760.

Bowman, K.O. and Shenton, L.R. (1999), The asymptotic moment profile and maximum likelihood, *Commun.Statist.Theory and Method*, 28(10), 2497-2508.

Bowman, K.O. and Shenton, L.R. (2005). Bias, Variance, Skewness and Kurtosis of Maximum Likelihood Estimators using Maple, *Far East Journal of Theoretical Statistics*, 17(2), 137-195.

Bowman, K.O. and Shenton, L.R. (2006). Moments of Maximum Likelihood Estimators in the Discrete Case, *Far East Journal of Theoretical Statistics*, 18(1), 61-91.

Fisher, R.A. (1922), On the mathematical foundations of theoretical statistics, *Philosophical Transactions of the Royal Society of London, Series A*, 222,309-368.

Fisher, R.A. (1941), The negative binomial distribution, *Annals of Eugenics*, XI(II), 182-187.

Fisher. R.A. (1950), *Contributions to Mathematical Statistics*, John Wiley & Sons, Inc, New York.

Mikulski, W. (1982), Efficiency, Second order, Encyclopedia of Statistical Sciences, 2, 468-472.

Shenton, L.R. (1950), Maximum Likelihood and the efficiency of the method of moments, *Biometrika*, 37, 112-116.

Shenton, L.R. and Bowman, K.O. (1971), Continued Fractions for the PSI Function and Its Derivatives, *SIAM Jour.Appl.Math.*, 20(4), 547-554.

# Addendum

## Validity of Fisher's formula for $i_{kk}$ using an identity given in Whittaker and Watson

### A.1 A surprise result for Psi functions

We have noted in the paper, that the validity of the expression for Fisher's $i_{kk}$ is not clearly proved. By good fortune, glancing through the chapter on the gamma function in Whittaker and Watson (1915, p257), we noticed the expression,

$$\psi(k+x)-\psi(k)=\frac{x}{k}-\frac{x(x-1)}{2k(k+1)}+\frac{x(x-1)(x-2)}{3k(k+1)(k+2)}+\cdots \quad (x=1,2,\cdots,k>0). \quad (1)$$

This result appears in the form

$$\frac{d}{dz}\frac{\ln\Gamma(z+x)}{\Gamma(z)}=\psi(z+x)-\psi(z),$$

for $x+z>0$.

An equivalent formula, namely

$$\psi(x)-\psi(x-\alpha)=\sum_{s=0}^{\infty}\frac{1}{s+1}\frac{\alpha(\alpha+1)\cdots(\alpha+s)}{x(x+1)\cdots(x+s)} \quad (\Re(x-\alpha)>0)$$

appears in Nielsen (1906). In particular see page 83, equation 7 of the contribution due to General Major V.H.O. Madsen.

To us, this result on the psi function is quite remarkable, especially when considered from a statistical point of view, and especially the negative binomial distribution and factorial moments. For

$$E[x]=kp, \quad E[x(x-1)]=k(k+1)p^2, \quad E[x(x-1)(x-2)]=k(k+1)(k+2)p^3,$$

and in general

$$E[x(x-1)\cdots(x-s+1)]=p^s k(k+1)\cdots(k+s-1).$$

Taking expectations in (1) for a negative binomial variate, we have

$$\begin{aligned} E\left(\psi(k+x)-\psi(k)\right) &= E\left(\frac{1}{k}+\frac{1}{k+1}+\cdots+\frac{1}{k+x-1}\right) \\ &= p-\frac{p^2}{2}+\frac{p^3}{3}-\cdots=\ln(p+1) \quad (0<p<1) \end{aligned}$$

as it should as shown in the main part of the paper.

## A.2 Further examples

We can write (1) in the form

$$S_1(x,k)=\frac{x}{k}-\frac{x(x-1)}{2k(k+1)}+\frac{x(x-1)(x-2)}{3k(k+1)(k+2)}-\cdots \tag{2}$$

using the notation

$$\begin{aligned} S_\lambda(x,k) &= \frac{1}{k^\lambda}+\frac{1}{(k+1)^\lambda}+\cdots+\frac{1}{(k+x-1)^\lambda} \quad (\lambda=1,2,\cdots,x=1,2,\cdots,k>0) \\ &= x, \quad \text{if } \ \lambda=0. \end{aligned}$$

If we differentiate (2) with respect to $k$, then derivatives of $(k(k+1)\cdots(k+x-1))^{-1}$ are required. Set $z=\frac{1}{k(k+1)\cdots(k+x-1)}$, then

$$\begin{aligned} \frac{\partial z}{\partial k} &= -[S_1(x,k)]z, \\ \frac{\partial^2 z}{\partial k^2} &= [S_2(x,k)+S_1(x,k)^2]z, \\ \frac{\partial^3 z}{\partial k^3} &= -\left\{2S_3(x,k)+3S_1(x,k)S_2(x,k)+S_1(x,k)^3\right\}z, \end{aligned} \tag{3}$$

and so on.

From (1) then

$$\begin{aligned} E(\psi_1(k+x)-\psi_1(k)) &= E\left(\frac{1}{k^2}+\frac{1}{(k+1)^2}+\cdots+\frac{1}{(k+x-1)^2}\right) \\ &= \sum_{x=1}^{\infty}\frac{(-1)^{x-1}}{x}p^x\left(\frac{1}{k}+\frac{1}{(k+1)}+\cdots+\frac{1}{(k+x-1)}\right) \\ &= \frac{p}{k}-\frac{p^2}{2}\left(\frac{1}{k}+\frac{1}{k+1}\right)+\frac{p^3}{3}\left(\frac{1}{k}+\frac{1}{k+1}+\frac{1}{k+2}\right)+\cdots \end{aligned}$$

where

$$\text{Coeff. } r = \frac{1}{k},$$
$$\text{Coeff. } r^2 = \frac{1}{k} - \frac{1}{2}\left(\frac{1}{k} + \frac{1}{k+1}\right) = \frac{1}{2k(k+1)},$$
$$\text{Coeff. } r^3 = \frac{1}{k} - \frac{1}{2}\left(\frac{1}{k} + \frac{1}{k+1}\right) + \frac{1}{3}\left(\frac{1}{k} + \frac{1}{k+1} + \frac{1}{k+2}\right) = \frac{2}{3k(k+1)(k+2)}.$$

Further coefficient have been checked and agree with Fisher's $i_{kk}$ series.

## A.3 A new expression for $E[S_3(x,k)]$

From the derivatives in (3)

$$\begin{aligned} E(\psi_2(k+x) - \psi_2(k)) &= 2E\left(\frac{1}{k^3} + \frac{1}{(k+1)^3} + \cdots + \frac{1}{(k+x-1)^3}\right) \\ &= \sum_{x=1}^{\infty} \frac{(-1)^{x-1}}{x} p^x \left(S_2(x,k) + [S_1(x,k)]^2\right). \end{aligned}$$

Thus

$$\begin{aligned} 2E\left(\frac{1}{k^3}+\frac{1}{(k+1)^3}+\cdots+\frac{1}{(k+x-1)^3}\right) &= \left(\frac{1}{k^2}+\frac{1}{k^2}\right)p - \frac{1}{2}\left\{\frac{1}{k^2}+\frac{1}{(k+1)^2}+\left(\frac{1}{k}+\frac{1}{k+1}\right)^2\right\}p^2 \\ &+ \frac{1}{3}\left\{\frac{1}{k^2}+\frac{1}{(k+1)^2}+\frac{1}{(k+2)^2}+\left(\frac{1}{k}+\frac{1}{k+1}+\frac{1}{k+2}\right)^2\right\}p^3 \\ &- \frac{1}{4}\left\{\frac{1}{k^2}+\frac{1}{(k+1)^2}+\frac{1}{(k+2)^2}+\frac{1}{(k+3)^2}+\left(\frac{1}{k}+\frac{1}{k+1}+\frac{1}{k+2}+\frac{1}{k+3}\right)^2\right\}p^4 \\ &+ \frac{p^5}{5}\left\{\sum_0^4 \frac{1}{(k+s)^2} + \left(\sum_0^4 \frac{1}{k+s}\right)^2\right\} + \cdots. \end{aligned} \tag{4}$$

If we express this in terms of $r$, where $r = p/(p+1)$, $p = r/(1-r)$, then the coefficient of $r^s$ in $2E[S_3(x,p)]$ is

$$(-1)^{s-1}\left\{\frac{\alpha_s(k)}{s} - \frac{\alpha_{s-1}(k)}{s-1}\frac{\Gamma(s)}{1!\Gamma(s-1)} + \frac{\alpha_{s-2}(k)}{s-2}\frac{\Gamma(s)}{2!\Gamma(s-2)}\right.$$
$$\left. -\frac{\alpha_{s-3}(k)}{s-3}\frac{\Gamma(s)}{3!\Gamma(s-3)} + \cdots + \frac{(-1)^{s-1}\alpha_1(k)}{s-(s-1)}\frac{\Gamma(s)}{(s-1)!\Gamma(1)}\right\}$$

where

$$\alpha_t(k) = \left(\frac{1}{k^2} + \frac{1}{(k+1)^2} + \cdots + \frac{1}{(k+t-1)^2}\right) + \left(\frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{k+t-1}\right)^2$$
$$\alpha_1(k) = \frac{1}{k^2} + \left(\frac{1}{k}\right)^2 = \frac{2}{k^2}, \qquad (t = 1, 2, \cdots).$$

Concerning the extended expression in (4), we have for $2E\left(\frac{1}{k^3} + \frac{1}{(k+1)^3} + \cdots + \frac{1}{(k+x-1)^3}\right)$,

$$\text{Coeff. } r = \frac{1}{k^2},$$
$$\text{Coeff. } r^2 = -\frac{k^2-k-1}{2k^2(k+1)^2}$$

Previous algebraic results using Maple code for $E(S_3)$

$$\text{Coeff. } r = \frac{1}{k^2}$$
$$\text{Coeff. } r^2 = -\frac{k^2-k-1}{2k^2(k+1)^2}$$
$$\text{Coeff. } r^3 = -\frac{3k^3+3k^2-6k-4}{3k^2(k+1)^2(k+2)^2}$$
$$\text{Coeff. } r^4 = \frac{11k^4-42k^3+13k^2-66k-36}{4k^2(k+1)^2(k+2)^2(k+3)^2}$$
$$\text{Coeff. } r^5 = -\frac{2(25k^5+190k^4+395k^3-10k^2-600k-288}{4k^2(k+1)^2(k+2)^2(k+3)^2(k+4)^2}.$$

For $S_4$ we have

$$6E[S_4(x,k)] = \sum_{x=1}^{\infty} \frac{(-1)^{x-1}}{x} p^x \left(2S_3(x,k) + 3S_2(x,k)S_1(x,k) + S_1^3(x,k)\right).$$

Differentiating with respect to $k$

$$\sum_{x=1}^{\infty} \frac{r^x}{x!} \frac{\Gamma(k+x)}{\Gamma(x)} S_m(x,k),$$

we can set up an expression for $E[S_1(x,k)S_m(x,k)]$. But expressions such as $E[S_2(x,k)S_m(x,k)]$, $m = 2, 3, \cdots$ have not been found.

The reader may like to check out the validity of the identity

$$E[S_1(x,k)]^3 = 4E[S_3(x,k)] + 3LE[S_2(x,k)] + \frac{3\partial}{\partial k}E[S_2(x,k)] + L^3,$$

where the first term on the right is given in the text, and

$$(i) \qquad L = \ln(1+p),$$
$$(ii) \qquad E[S_2(x,k)] = \sum_{x=1}^{\infty} \frac{r^x (x-1)!\Gamma(k)}{x\Gamma(k+x)} \qquad \text{(Fisher)}$$
$$(iii) \qquad p > 0, k > 0, r = \frac{p}{p+1}, 0 < r < 1.$$

## A.4 Conclusion

The mathematical identity in equation (1) is stated in a book by Whittaker and Watson (1915, p 257). The book concerns so-called "mathematical Analysis" and has little to do with statistics or probability.

Now in the identity (1) the denominators are $k$, $k(k+1)$, $k(k+1)(k+2)$ and so on; but these occur in the factorial moments of the negative binomial variate. For example, the factorial moments are

$$\mu_{[s]} = p^s k(k+1)\cdots(k+s-1) = p^s\Gamma(k+s)/\Gamma(k). \quad (s = 0, 1, \cdots)$$

Stieltjes moment problem which refers to the case of distributions on $(0, \infty)$ are involved. For a set of moments, Stieltjes (1918,1874-5) considered the Stieltjes integral transform

$$\int_0^\infty \frac{d\sigma(x)}{x+1} = \frac{1}{\alpha_0 x+} \frac{1}{\alpha_1+} \frac{1}{\alpha_2 x+} \frac{1}{\alpha_3+} \cdots$$

referring to a continued fraction form. The moment problem is determinate if and only if $\alpha$'s are positive and also if $\sum \alpha_s = \infty$. Now from Bowman and Shenton (1989, p48), for the negative binomial distribution $B(k,t;p)$, there is the Stieltjes continued fraction

$$\sum_{t=0}^{\infty} \frac{B(k,t;1)}{z+t} = \frac{1}{z+} \frac{kp}{1+} \frac{q}{z+} \frac{(k+1)p}{1+} \frac{2q}{z+} \frac{(k+2)p}{1+} \cdots$$
$$= \frac{1}{\alpha_1 z+} \frac{1}{\alpha_1+} \frac{1}{\alpha_3 z+} \cdots.$$

If

$$\frac{q_0}{x+} \frac{p_1}{1+} \frac{q_1}{x+} \frac{p_2}{1+} \frac{q_2}{z+} \cdots = \frac{1}{\alpha_0+} \frac{1}{\alpha_1+} \frac{1}{\alpha_2+} \frac{1}{\alpha_3+} \cdots$$

then

$$\alpha_{2p-1} = \frac{q_0 q_1 \cdots q_{s-1}}{p_1 p_2 \cdots p_s}.$$

Examples

$$\alpha_1 = \frac{1}{kp}$$
$$\alpha_3 = \frac{1}{kp}\frac{q}{(k+1)p}$$
$$\alpha_5 = \frac{1}{kp}\frac{q}{(k+1)p}\frac{2q}{(k+2)p}$$
$$\alpha_7 = \frac{1}{kp}\frac{q}{(k+1)p}\frac{2q}{(k+2)p}\frac{3q}{(k+3)p}$$

and in general

$$\alpha_{2s-1} = \frac{1}{p}\left(\frac{q}{p}\right)^s \frac{\Gamma(s)\Gamma(k)}{\Gamma(k+s)} \sim \frac{\Gamma(k)}{p}\left(\frac{q}{p}\right)^{s-1} s^{-k} \quad (s \to \infty)$$
$$= \frac{\Gamma(k)}{p} \exp\left\{(s-1)\ln\left(\frac{q}{p}\right) - k\ln(s)\right\}. \quad (q = 1+p, 0 < p)$$

Hence $\sum \alpha_s$ diverges and $\sum \alpha_s = \infty$. No other distribution has the same moments as the negative binomial random variable.

# Reference


Bowman, K.O. and Shenton, L.R. (1989). *Continued Fractions in Statistical Applications*, Marcel Dekker, Inc., New York.

Nielsen, N. (1906). *Handbuch der Theorie der Gammafunktion*, Chelsea Publiching Co., Bronx, New York.

Stieltjes, T.J. (1918). *Oeuvres Complétes*, Tome 2, P. Noordoff, Gronigen.

Whittaker, E.T. and Watson, G.N. (1915). *A Course of Modern Analysis*, 2nd Edition, Cambridge University Press.