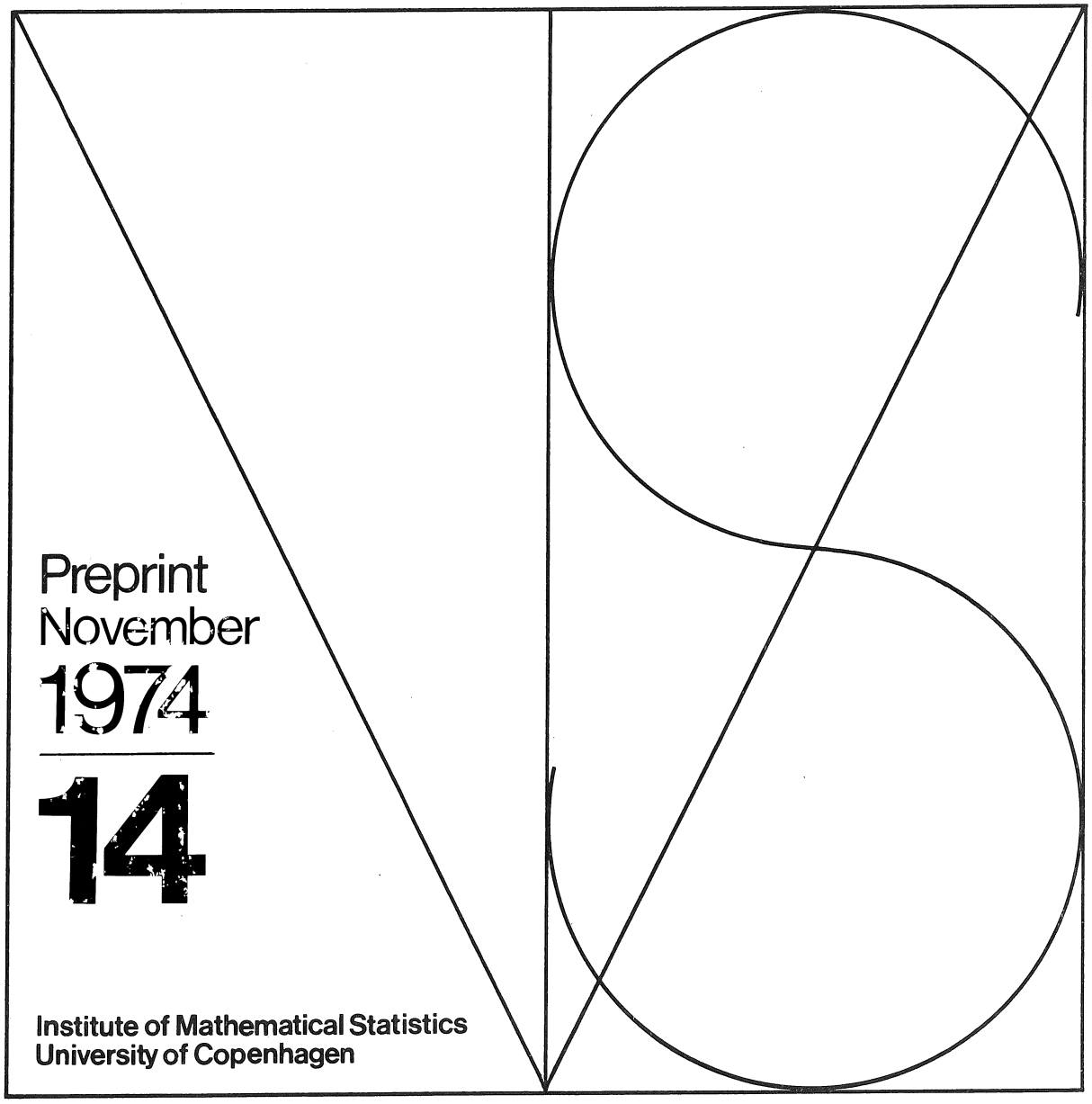
J. E. Beyer N. Keiding W. Simonsen

The Exact Behaviour of the Maximum Likelihood Estimator in the Pure Birth Process and the Pure Death Process



Preprint
November
1974

14

Institute of Mathematical Statistics
University of Copenhagen

Jan E. Beyer* Niels Keiding W. Simonsen**

THE EXACT BEHAVIOUR OF THE MAXIMUM LIKELIHOOD ESTIMATOR IN THE PURE BIRTH PROCESS AND THE PURE DEATH PROCESS

Preprint 1974 No. 14

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

November 1974

* Institute of Mathematical Statistics and Operations Research, The Technical University of Denmark.

** Laboratory of Insurance Mathematics, University of Copenhagen.

Summary.

This is a small-sample study of the maximum likelihood estimator of the parameters in the pure (linear) birth process and the pure (linear) death process, observed in a fixed time interval. The expectation, variance, and coefficient of skewness are tabulated and compared with various approximations. Some new asymptotic results on mean convergence of the estimators are also derived, and particular attention is called to the properties for large observation times (or large birth or death intensities) and small numbers of initial individuals.

Key Words: Estimation in Markov processes, Life testing, Birth-and-death processes, Small-sample theory, Maximum likelihood estimation, Maximum likelihood approximations, Truncated exponential distributions.

1. Introduction.

This paper is a small-sample study of the maximum likelihood estimators of the parameter in the pure birth process and in the pure death process, derived under the assumption that the process is observed continuously in a fixed time interval $[0,t]$ and conditional on the value at 0.

Let $\{X_t,\ 0 \leq t < \infty\}$ be a simple (linear) birth-and-death process given by

$$P\{X_{t+h} = j \mid X_t = i\} = \begin{cases} i\lambda h + o(h), & j = i + 1, \\ 1 - i(\lambda+\mu)h + o(h), & j = i, \\ i\mu h + o(h), & j = i - 1, \\ o(h), & \text{otherwise}, \end{cases}$$

$i = 0,1,2,\ldots, \lambda \geq 0,\ \mu \geq 0$ and assume throughout that X_0 is degenerate at some $q > 0$. Maximum likelihood estimation of λ and μ in this process, assuming continuous observation over some fixed time interval $[0,t]$ was studied by Keiding (1975). The maximum likelihood estimators are the occurrence-exposure rates $\hat{\lambda} = B_t/S_t$ and $\hat{\mu} = D_t/S_t$ where B_t and D_t are the number of births and deaths in $[0,t]$, respectively, and $S_t = \int_0^t X_u\,du$ is the total time lived by the population in $[0,t]$. The exact distribution of these estimators, being ratios between a discrete and a continuous random variable, is not very attractive.

In this paper we shall study the moments of the estimators in the two extreme cases: that of the pure birth process, where $\mu = 0$, and that of the pure death process, where $\lambda = 0$. Accordingly, the paper is divided into two parts, corresponding to these two cases.

Part I, concerning the pure birth process, gives in Section 2 some exact and asymptotic properties of the moments, not previously stated in the literature, as well as the basic representation of the moments $E(B_t^{(m)}/S_t^r)$ as an integral. In Section 3 the expectation $E(1/\hat{\lambda})$ of the maximum likelihood estimator of $1/\lambda$ is derived and used to provide a lower bound on $E(\hat{\lambda})$. Section 4 gives a derivation of approximations to the moments based on asymptotic maximum likelihood theory. Finally, we report in Section 5 on the numerical results. We study the properties of the exact moments and discuss the approximations mentioned above, thus also illustrating the asymptotic distribution results given by Keiding (1974).

In Part II, a similar discussion is devoted to the pure death process.

For estimation in the pure death process, a large literature is available, cf. the bibliographies by Mendenhall (1958) and Govindarajulu (1964). The form of the estimator is latent in Grenander (1950, p. 253) and was derived by Littell (1952), Bartlett (1953a) and Deemer and Votaw (1955). Earlier studies of small-sample properties and validity of approximate distributions are by Deemer and Votaw (loc. cit.), Mendenhall and Hader (1958) and Bartholomew (1963). A survey of tests for the assumption of exponential life time distributions was given by Epstein (1960), and studies on the robustness of the methods based on exponential life

time distributions when this assumption is violated are by Zelen and Dannemiller (1961) and Barlow and Proschan (1967). Some further references were given by Keiding (1975).

Part I: The pure birth process.

2. The moments of the maximum likelihood estimator.

Keiding (1974) stated the distribution of the minimal sufficient statistic (B_t,S_t) in the following way. With probability $e^{-q\lambda t}$, $(B_t,S_t) = (0,qt)$. Otherwise the density with respect to counting measures on the integers and Lebesgue measure on the Borel field is given as

$$\binom{b+q-1}{q-1} (\lambda t)^b e^{-\lambda s} g_b(\frac{s}{t} - q) \tag{2.1}$$

$b = 1,2,\ldots,$ $qt \leq s < \infty$, where g_b is the probability density of a sum of b independent, uniformly distributed random variables on $[0,1]$.

Remark. From this expression the distribution of the maximum likelihood estimator $\hat{\lambda} = B_t/S_t$ is readily derived as given by an atom at 0 with probability $e^{-q\lambda t}$ and otherwise with the density

$$v^{-2} \sum_{b=1}^{\infty} \binom{b+q-1}{q-1} b\lambda^b t^{b-1} e^{-\lambda b/v} g_b(\frac{b}{tv} - q) ,$$

$v > 0$. It may be seen that this density is zero on the intervals

$$(\frac{i-1}{qt}, \frac{i}{(q+i)t}), \qquad i=1,\ldots,m$$

where m is the greatest integer such that $m(m-1) < q$. We shall not consider this rather irregular distribution further.

Let now

$$\varphi(x) = \int_0^1 e^{-xt}dt = (1-e^{-x})/x ,$$

$x \geq 0$, be the Laplace transform of the uniform distribution on $[0,1]$ and let $\gamma = \lambda t$. The basic formula in the study of the moments of $\hat{\lambda}$ is given by the following Theorem.

Theorem 2.1. Define $\psi_q(\gamma,m,r) = \lambda^{-r}E(B_t^{(m)}/S_t^r)$, where m and r are non-negative integers and $a^{(x)} = a(a-1)\cdots(a-x+1)$. Then for $m = 0,1,2,\ldots$ and $r = 1,2,\ldots$

$$\psi_q(\gamma,m,r) = (q+m-1)^{(m)}\gamma^{m-r}\int_\gamma^\infty \frac{(x-\gamma)^{r-1}}{(r-1)!} e^{-qx}\varphi^m(x)[1-\gamma\varphi(x)]^{-q-m}dx \qquad (2.2)$$

and we have the recurrence relation

$$\psi_q(\gamma,m+1,r+1) = \frac{1}{r}[(q+m)\psi_q(\gamma,m,r) - q\psi_{q+1}(\gamma,m,r)] . \qquad (2.3)$$

Proof. By (2.1),

$$\psi_q(\lambda t,m,r) = \sum_{b=m}^\infty \frac{(q+b-1)^{(b)}}{(b-m)!}\lambda^{b-r}t^{b-1}\int_{qt}^\infty s^{-r}e^{-\lambda s}g_b(\frac{s}{t} - q)ds . \qquad (2.4)$$

To evaluate the integral, we use the fact that the Laplace transform of g_b is φ^b, that is,

$$\int_0^b e^{-xu}g_b(u)du = \varphi^b(x)$$

so that by multiplying both sides by $e^{-qx}(x-\gamma)^{r-1}/(r-1)!$ and subsequent integration, we obtain

$$\int_0^b \left[\int_\gamma^\infty \frac{(x-\gamma)^{r-1}}{(r-1)!} e^{-x(q+u)} dx \right] g_b(u)du = \int_\gamma^\infty \frac{(x-\gamma)^{r-1}}{(r-1)!} e^{-qx} \varphi^b(x)dx \ .$$

The inner integral on the left-hand side is the Laplace transform of a gamma distribution on $[\gamma,\infty)$ and equals $e^{-\gamma(q+u)}/(q+u)^r$. Inserting this result into (2.4), we obtain

$$\psi_q(\gamma,m,r) =$$

$$(q+m-1)^{(m)}\gamma^{m-r} \int_\gamma^\infty \frac{(x-\gamma)^{r-1}}{(r-1)!} e^{-qx}\varphi^m(x) \sum_{b=0}^{\infty} \binom{q+m+b-1}{b} [\gamma\varphi(x)]^b dx \ .$$

Since $\varphi(x) = \int_0^1 e^{-xt}dt$, where e^{-xt} is decreasing in x for each fixed $t \in (0,1]$, it is itself decreasing, so that $0 \leq \gamma\varphi(x) \leq \gamma\varphi(\gamma) = 1-e^{-\gamma} < 1$ for $x \geq \gamma > 0$. We may therefore complete the sum under the integral sign which ends the proof of the representation of $\psi_q(\gamma,m,r)$.

The proof of the recurrence relation is obtained by noting that

$$(x-\gamma)\varphi(x) = [1-\gamma\varphi(x)] - e^{-x} \ .$$

Corollary. Define $\mu_r = E[(\hat{\lambda}/\lambda)^r]$. Then

$$\mu_1 = \psi_q(\gamma,1,1)$$
$$\mu_2 = \psi_q(\gamma,1,2) + \psi_q(\gamma,2,2)$$
$$\mu_3 = \psi_q(\gamma,1,3) + 3\psi_q(\gamma,2,3) + \psi_q(\gamma,3,3)$$

and in general

$$\mu_r = \sum_{i=1}^{r} T_r^{(i)} \psi_q(\gamma,i,r)$$

where the $T_r^{(i)}$ are the Stirling's numbers of the second kind defined by

$$t^r = \sum_{k=0}^{r} T_r^{(k)} t^{(k)}$$

and satisfying the recurrence relation

$$T_r^{(i)} = T_{r-1}^{(i-1)} + iT_{r-1}^{(i)} \qquad 2 \leq i + 1 \leq r ,$$

$T_r^{(1)} = T_r^{(r)} = 1$. In particular,

$$E(\hat{\lambda}) = \lambda\mu_1, \quad Var(\hat{\lambda}) = \lambda^2(\mu_2-\mu_1^2)$$

and the coefficient of skewness

$$\beta = \frac{E[\{\hat{\lambda}-E(\hat{\lambda})\}^3]}{[Var(\hat{\lambda})]^{3/2}} = \frac{\mu_3-3\mu_1\mu_2+2\mu_1^3}{(\mu_2-\mu_1^2)^{3/2}} .$$

Keiding (1974) proved strong consistency and results on asymptotic distribution of $\hat{\lambda}$ as $t \to \infty$ for fixed q or as $q \to \infty$ for fixed t. The following Theorem gives results on the behaviour of $E(\hat{\lambda})$ under these limits as well as for small observation times or small parameter values.

Theorem 2.2.

(a) $E(\hat{\lambda}) \leq \lambda$.

(b) For fixed q, $\lim_{\lambda t \to 0} E(\hat{\lambda}/\lambda) = q \log(\frac{q+1}{q})$ and $\lim_{\lambda t \to \infty} E(\hat{\lambda}/\lambda) = 1$.

(c) For fixed $t > 0$, (and fixed λ), $\lim_{q \to \infty} E(\hat{\lambda}) = \lambda$.

Proof. The argument x is omitted from $\varphi(x)$ in this proof. To prove (a), we write

$$D_x[e^{-qx}(1-\gamma\varphi)^{-q}] = -\,qe^{-qx}(1-\gamma\varphi)^{-q-1}(1-\gamma\varphi-\gamma\varphi')$$

and since $\varphi + \varphi' = (1-\varphi)/x$, the last factor is $1-\gamma(1-\varphi)/x$ $= \varphi + (1-\varphi)(1-\gamma/x)$, so that

$$qe^{-qx}\varphi(1-\gamma\varphi)^{-q-1} = -D_x[e^{-qx}(1-\gamma\varphi)^{-q}] - qe^{-qx}(1-\varphi)(1-\gamma\varphi)^{-q-1}(1-\gamma/x) .$$

Now when $x > \gamma$, all factors in the last term are positive, and consequently by integration from γ to ∞

$$E(\hat{\lambda}/\lambda) = \psi_q(\gamma,1,1) = \int_\gamma^\infty qe^{-qx}\varphi(x)[1-\gamma\varphi(x)]^{-q-1}dx$$

$$\leq -\,[e^{-qx}\{1-\gamma\varphi(x)\}^{-q}]_\gamma^\infty = 1 .$$

We next prove $\lim E(\hat{\lambda}) = \lambda$ as $t \to \infty$ or $q \to \infty$. Since, as mentioned above, we know that $\hat{\lambda} \to \lambda$ a.s. in these cases, we obtain by Fatou's lemma $\liminf E(\hat{\lambda}) \geq E(\lim \hat{\lambda}) = \lambda$ and the result follows by (a). Since $E(\hat{\lambda}/\lambda)$ depends only on λ and t through $\gamma = \lambda t$, it is now also obvious that $E(\hat{\lambda}/\lambda) \to 1$ as $\gamma \to \infty$.

Finally, the first part of (b) is proved directly.

By (2.4), we may write $E(\hat{\lambda}/\lambda)$ as

$$\psi_q(\lambda t,1,1) = \sum_{b=1}^\infty \frac{(q+b-1)^{(b)}}{(b-1)!}(\lambda t)^{b-1}\int_q^\infty u^{-1}e^{-\lambda tu}g_b(u-q)du .$$

Since $u^{-1} \leq 1$ for $u \geq q \geq 1$, the integral is, for each b, bounded by

$$e^{-q\lambda t}\int_0^\infty e^{-\lambda tu}g_b(u)du = e^{-q\lambda t}\varphi^b(\lambda t) \leq 1 ,$$

and for $\lambda t_0 < 1$

$$\sum_{b=1}^{\infty} \frac{(q+b-1)^{(b)}}{(b-1)!} (\lambda t_0)^{b-1} = \frac{q}{(1-\lambda t_0)^q} .$$

We may therefore use dominated convergence for $t_0 > t \to 0$, obtaining

$$\lim_{t \to 0} \psi(\lambda t,1,1) = q \int_q^{\infty} u^{-1} g_1(u-q)du = q \int_q^{q+1} u^{-1} du = q \log[(q+1)/q]$$

as claimed.

We now turn to the following L_2-convergence results.

Theorem 2.3. For fixed λ and t as $q \to \infty$ and for fixed q as $\lambda t \to \infty$ $E[\{(\hat{\lambda}/\lambda) - 1\}^2] \to 0$ and therefore $\mathrm{Var}(\hat{\lambda}/\lambda) \to 0$.

Proof. Let first $\lambda > 0$ and $t > 0$ be fixed. Then, since $S_t \geq qt$,

$$E\{(\hat{\lambda}-\lambda)^2\} = E\left(\frac{B_t-\lambda S_t}{S_t}\right)^2 \leq E(B_t-\lambda S_t)^2/(qt)^2 ,$$

where $E(B_t-\lambda S_t) = 0$. The birth process with $X_0 = q$ may be interpreted as the sum of q stochastically independent birth processes with $X_0 = 1$ and the same parameter λ, and similarly $B_t-\lambda S_t$ may be interpreted as the sum of independent contributions from each of these processes. Denoting $\mathrm{Var}(B_t-\lambda S_t)$ in a process with $X_0 = 1$ by σ_1^2, we therefore have

$$E\{(\hat{\lambda}-\lambda)^2\} \leq q\sigma_1^2/(qt)^2 \to 0$$

as $q \to \infty$.

Next, consider a fixed positive integer q. According to the Corollary of Theorem 2.1

$$\mathrm{Var}(\hat{\lambda}/\lambda) = \psi_q(\gamma,1,2) + \psi_q(\gamma,2,2) - \psi_q^2(\gamma,1,1)$$

and it is known from Theorem 2.2(b) that the last term tends to 1 as $t \to \infty$. Furthermore, since $S_t \geq qt$

$$\psi_q(\gamma,1,2) = \lambda^{-2}E(B_t/S_t^2) \leqq (qt)^{-1}\lambda^{-2}E(\hat{\lambda}) \to 0$$

as $t \to \infty$ by Theorem 2.2(b). And by Theorem 2.1

$$\psi_q(\gamma,2,2) = (q+1)\psi_q(\gamma,1,1) - q\psi_{q+1}(\gamma,1,1) \to q + 1 - q = 1$$

as $t \to \infty$ and hence $\gamma \to \infty$ by Theorem 2.2(b). This proves $\lim_{t\to\infty} \mathrm{Var}(\hat{\lambda}) = 0$, and $E\{(\hat{\lambda}-\lambda)^2\} \to 0$ then follows from the mean convergence proved in Theorem 2.2(b).

Finally, the behaviour for small observation periods or small parameter values of the variance and the coefficient of skewness may be obtained by passing to the limit in the relevant integrals, which may be justified as in the proof of Theorem 2.2(b). Both quantities become infinite in this limit, and the speed of divergence is described below. Notice that $\beta(\hat{\lambda}) = \beta(\hat{\lambda}/\lambda)$.

Theorem 2.4. For fixed q and $\lambda t \to 0$,

$$(q + 1)\lambda t + \mathrm{Var}(\hat{\lambda}/\lambda) \to 1$$

and

$$t^{1/2}\lambda^{1/2}\,\frac{q(q+1)^{1/2}}{q+\frac{1}{2}}\,\beta(\hat{\lambda}) \to 1 .$$

Remark. Monotonicity properties of the moments as functions of q for fixed λ and t and as functions of $\gamma = \lambda t$ for fixed q have been further studied by Simonsen (unpublished). Among the results obtained are that

$$(-1)^n D_q^{\,n}\,\psi_q(\gamma,1,1) \leqq 0 \qquad \text{for } n = 1,2,\ldots$$

so that, in particular, $E(\hat{\lambda})$ is increasing and concave as function of q, and

$$D_\gamma \psi_q(\gamma,1,1) \geqq 0$$

so that $E(\hat{\lambda}/\lambda)$ is increasing as function of λt. These properties are illustrated by the numerical results in Section 5 below.

There are also general results on the asymptotic behaviour of the moments as $q \to \infty$ or $\lambda t \to \infty$.

3. The expectation of the inverse estimator and an inequality.

It may in several situations be of interest to inquire into the possible bias of the maximum likelihood estimator $\hat{\lambda}^{-1}$ of λ^{-1}. We give in this Section results on this problem as well as an inequality for $E(\hat{\lambda})$ obtained via Jensen's inequality.

Since $\hat{\lambda} = 0$ with positive probability, $E(\hat{\lambda}^{-1}) = \infty$. But it is still possible to get the following results concerning $E(\hat{\lambda}^{-1}|B_t > 0)$.

Theorem 3.1. Let $\gamma = \lambda t$. Then, interpreting an empty sum as 0,

$$E(\lambda/\hat{\lambda}|B_t>0) = 1 - \frac{\gamma}{e^{\gamma}-1} + \frac{q\gamma}{e^{q\gamma}-1}\left\{\gamma + \sum_{k=1}^{q-1}\binom{q-1}{k}\frac{(e^{\gamma}-1)^k}{k}\right\}.$$

Proof. We compute $E(\lambda/\hat{\lambda})$ by conditioning on B_t. First,

$$E(\lambda/\hat{\lambda}|B_t>0) = E[E(\lambda S_t/B_t|B_t)|B_t>0] \qquad (3.1)$$

and we use the representation $S_t = tX_t - \sum_{i=q+1}^{X_t} T_i$ where $T_{q+1},\ldots,T_{X_t}$ are the epochs at which the births take place (Keiding 1974).

Given X_t, or equivalently $B_t = X_t - q$, these epochs are distributed like an ordered sample of B_t independent random variables, all with density $\lambda e^{\lambda u}/(e^{\lambda t}-1)$, $0 < u < t$ (Puri 1968). Since a sum of ordered variables has the same distribution as a sum of the same number of corresponding independent variables, and since the expectation of the above distribution is computed as

$$(e^{\lambda t}-1)^{-1}\int_0^t \lambda u e^{\lambda u}du = \frac{t}{1-e^{-\lambda t}} - \frac{1}{\lambda},$$

we conclude that

$$E(S_t|B_t) = tB_t + qt - B_t\left(\frac{t}{1-e^{-\lambda t}} - \frac{1}{\lambda}\right) = B_t\left(\frac{1}{\lambda} - \frac{t}{e^{\lambda t}-1}\right) + qt \ .$$

Next, by inserting this into (3.1), we obtain

$$E(\lambda/\hat{\lambda}|B_t>0) = 1 - \frac{\lambda t}{e^{\lambda t}-1} + \lambda q t \, E(B_t^{-1}|B_t>0) \ .$$

But B_t is negative binomially distributed with parameters $(q,e^{-\lambda t})$, and the result is therefore obtained from Rider (1962), who not only gave the formula used above, but also numerical tables. Further formulas and approximations were provided by Govindarajulu (1962).

Corollary. For fixed q as $t \to \infty$

$$E(\lambda/\hat{\lambda}|B_t>0) \to 1 \ .$$

Theorem 3.1 may be used to provide a lower bound on $E(\hat{\lambda}/\lambda)$, which by Theorem 2.2(a) is bounded above by 1.

Theorem 3.2. Let $\gamma = \lambda t$. Then

$$E(\hat{\lambda}/\lambda) \geq \frac{1 - e^{-q\gamma}}{1 - \frac{\gamma}{e^{\gamma}-1} + \frac{q\gamma}{e^{q\gamma}-1}\left\{\gamma + \sum_{k=1}^{q-1}\binom{q-1}{k}\frac{(e^{\gamma}-1)^k}{k}\right\}} \ . \tag{3.2}$$

Proof. First notice that

$$E(\hat{\lambda}/\lambda) = E\{B_t/(\lambda S_t)|B_t=0\}P\{B_t=0\} + E\{B_t/(\lambda S_t)|B_t>0\}P\{B_t>0\} \ ,$$

where $E\{B_t/(\lambda S_t) | B_t = 0\} = 0$ and $P\{B_t > 0\} = 1 - e^{-q\lambda t}$. By Jensen's inequality

$$E(\hat{\lambda}/\lambda | B_t > 0) \geq \frac{1}{E(\lambda/\hat{\lambda} | B_t > 0)}$$

and the result is then immediate from Theorem 3.1.

Remark. Denote the right-hand side of (3.2) by $a_q(\gamma)$. Then it is seen that for fixed q as $\gamma \to \infty$

$$a_q(\gamma) = 1 - \gamma(q-1)e^{-\gamma} + o(\gamma e^{-\gamma})$$

which in conjunction with Theorem 2.2(a) in particular proves the mean convergence $\lim_{\gamma \to \infty} E(\hat{\lambda}/\lambda) = 1$ of Theorem 2.2 once more and also yields a bound on the rate of convergence. (Notice the special case $q = 1$.) As $\gamma \to 0$, we have $a_q(\gamma) \to 2q/(2q+1) \sim 1 - (2q)^{-1} + (4q^2)^{-1}$ whereas the exact result by Theorem 2.2(b) is $\lim_{\gamma \to 0} E(\hat{\lambda}/\lambda) = q \log(1+q^{-1}) \sim 1 - (2q)^{-1} + (3q^2)^{-1}$.

We finally call attention to the fact that $a_q(\gamma)$ is not monotone. As an example, $a_1(\gamma)$ has a minimum at $\gamma = 1$, at which point it attains the value $1 - e^{-1} = .632$. The possible numerical usefulness of this inequality is discussed further in Section 5 below.

4. Approximations from maximum likelihood theory

A straightforward though of course not in general warranted application of asymptotic distribution results such as those given by Keiding (1974) is to approximate the moments of the estimator by the moments in the limiting distribution. The asymptotic results are that $\{q(e^{\lambda t} - 1)\}^{\frac{1}{2}}(\hat{\lambda}/\lambda - 1)$ is asymptotically normal $(0,1)$ for fixed t and $q \to \infty$ and asymptotically Student with $2q$ d.f. for fixed q and $t \to \infty$.

Consider first the normal approximation. This invites the hypothesis of $E(\hat{\lambda}) = \lambda$ and $\beta(\hat{\lambda}) = 0$ and the usual inverse information approximation for the variance

$$\text{Var}\ (\hat{\lambda}/\lambda) \sim \frac{1}{q(e^{\lambda t} - 1)} . \qquad (4.1)$$

The Student approximation again invites the hypothesis of no bias or skewness and since the variance of the Student distribution with $2q$ d.f. for $q \geq 2$ is $q/(q-1)$, we would expect

$$\text{Var}\ (\hat{\lambda}/\lambda) \sim \frac{1}{(q-1)(e^{\lambda t} - 1)} \qquad (4.2)$$

Comparing (4.1) and (4.2), it is seen that the price paid for the limiting randomness as $t \to \infty$, discussed at length by Keiding (1974), is that the variance should be considered as if the process had started with one individual less.

Returning to the case of independent replications as $q \to \infty$, more refined approximations to the moments of the maximum likelihood estimators were given by Haldane and Smith (1956). In a general framework of "curved exponential families" Efron (1974) showed in particular how the second approximation for the variance may be interpreted in terms of what he defines as the "statistical curvature". We give below the approximations to the expectation, the variance, and the coefficient of skewness in the case of the birth process.

Theorem 4.1. Let $\gamma = \lambda t$. Then as $q \to \infty$ for fixed γ,

$$E\left(\frac{\hat{\lambda}}{\lambda}\right) = 1 - \frac{\gamma e^{\gamma} - e^{\gamma} + 1}{q(e^{\gamma} - 1)^2} + O\left(\frac{1}{q^2}\right),$$

$$\mathrm{Var}\left(\frac{\hat{\lambda}}{\lambda}\right) = \frac{1}{q(e^{\gamma}-1)}\left[1 + \frac{1}{q(e^{\gamma}-1)^3}\{e^{3\gamma} + (3\gamma^2 - 10\gamma + 2)e^{2\gamma} + (2\gamma^2+10\gamma-7)e^{\gamma}+4\}\right] + O\left(\frac{1}{q^3}\right),$$

and the coefficient of skewness

$$\beta = \frac{E[\{\hat{\lambda}-E(\hat{\lambda})\}^3]}{\mathrm{Var}(\hat{\lambda})^{3/2}} = \frac{4(e^{\gamma} - 1) - 3\gamma e^{\gamma}}{q^{1/2}(e^{\gamma} - 1)^{3/2}} + O\left(\frac{1}{q}\right). \qquad (4.3)$$

Proof. Let $\ell(\lambda) = \log L(\lambda)$, and define the information as

$$i = -E\{D^2\ell(\lambda)\} = E[\{D\ell(\lambda)\}^2] = \frac{q}{\lambda^2}(e^{\lambda t} - 1).$$

The covariance matrix Σ of the sufficient statistic (B_t, S_t) is given by (Puri 1966)

$$\Sigma_{11} = \mathrm{Var}(B_t) = qe^{\lambda t}(e^{\lambda t} - 1),$$

$$\Sigma_{22} = \mathrm{Var}(S_t) = q\{\frac{1}{\lambda^2}(e^{2\lambda t} - 1) - \frac{2t}{\lambda}e^{\lambda t}\},$$

$$\Sigma_{12} = \Sigma_{21} = \mathrm{Cov}(B_t, S_t) = qe^{\lambda t}\{\frac{1}{\lambda}(e^{\lambda t} - 1) - t\}.$$

Following Efron (1974), let now

$$\eta(\lambda) = \begin{pmatrix} \log\lambda \\ -\lambda \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} B_t \\ S_t \end{pmatrix}$$

be the vectors of canonical parameters and canonical statistics, respectively. Then $\ell(\lambda) = \eta(\lambda)'X$, and therefore moments of ℓ, $D\ell$, etc. are easily obtained directly from Σ as given above. In particular, define $a = E\{D\ell(\lambda)\ D^2\ell(\lambda)\}$. Then since $E\{D\ell(\lambda)\} = 0$,

$$a = E\{D\eta(\lambda)'X\ \ D^2\eta(\lambda)'X\} = Cov\{D\eta(\lambda)'X, D^2\eta(\lambda)'X\}$$

$$= D\eta(\lambda)'\Sigma\ D^2\eta(\lambda) = -qte^{\lambda t}/\lambda^2 \ .$$

Let furthermore

$$b = E\left\{\frac{DL(\lambda)}{L(\lambda)}\ \ \frac{D^2L(\lambda)}{L(\lambda)}\right\} = E\{D\ell(\lambda)\ D^2\ell(\lambda)\} + E[\{D\ell(\lambda)\}^3]$$

Then since (see Bartlett (1953b))

$$DE\{D^2\ell(\lambda)\} + 2E\{D\ell(\lambda)\ D^2\ell(\lambda)\} + E[\{D\ell(\lambda)\}^3] = 0 \ , \qquad (4.4)$$

we obtain

$$b = -DE\{D^2\ell(\lambda)\} - a = \frac{2q}{\lambda^3}\left[\lambda te^{\lambda t} - e^{\lambda t} + 1\right] \ .$$

Haldane and Smith (1956) gave the approximate expectation as

$$E(\hat{\lambda}) = \lambda - \frac{b}{2(qi)^2} + O(q^{-2})$$

and the first result is now easily proved; similarly, they gave the

approximation of the skewness β as

$$\beta = \frac{E[\{D\ell(\lambda)\}^3] - 3b}{q^{1/2} i^{3/2}} + O(q^{-3/2})$$

and the result concerning β is proved using (4.4) once more.

For the variance, we want one extra term, which is quite complicated in Haldane and Smith's original formulation. We shall therefore quote Efron's (1974) version, which is

$$\mathrm{Var}(\hat{\lambda}) = \frac{1}{qi}\left[1 + \frac{1}{q}\left(c^2 + \frac{b^2}{2i^3}\right) + 2DE(\hat{\lambda} - \lambda)\right] + O(q^{-3}) .$$

All quantities have been defined above except the "statistical curvature" c given by $c^2 = |M| i^{-3}$, M being the symmetric matrix with elements $M_{ij} = D^i\eta(\lambda)' \, \Sigma \, D^j\eta(\lambda)$. It may be seen that

$$c^2 = \frac{1}{q}\left[\frac{1}{1 - e^{-\lambda t}} - \frac{(\lambda t)^2 e^{2\lambda t}}{(e^{\lambda t} - 1)^3}\right] .$$

We refer to Efron's paper for further discussion of this concept and a rigorous derivation of the approximation. The variance approximation is now derived by carrying through the necessary algebra and using the approximation to $E(\hat{\lambda}-\lambda)$ derived above.

5. Numerical results

By direct numerical integration we have tabulated the expectation, variance and coefficient of skewness of $\hat{\lambda}$, and a detailed table of the results is available (Beyer 1974). We give in this Section first a report on the integration procedures used and then a discussion of the exact behaviour of the first three moments along with a numerical evaluation of the approximations and inequalities proposed in the earlier Sections.

a. The numerical integration

The integrals $\psi_q(\gamma,m,r)$ given by (2.2) had to be computed for $1 \leq m \leq r \leq 3$ and a representative set of parameter values γ and q. Since $\lim_{\gamma\to\infty} \psi(\gamma,m,r) = 0$ if $m < r$ and 1 if $m = r$, it turned out to be advantageous to compute $1 - \psi_q(\gamma,m,m)$ instead of $\psi_q(\gamma,m,m)$ itself. The integrand of $1 - \psi_q(\gamma,m,m)$ has a peak approximately located at $me^{-\gamma/q}$ and since we did not succeed in transforming the integrand to avoid this, we first truncated and then split the actual integration interval at the point $4\,me^{-\gamma/q}$, thus obtaining two intervals, one around the peak which was densely tabulated and another with more sparse evaluations.

Using an adaptive Simpson quadrature procedure by Lyness (1969,1970) in double precision on the IBM 370/165 at "Northern Europe University Computer Center" in Lyngby, Denmark, the computations were done with at least 8 correct significant digits in the ψ's, giving the bias and the

variance with a similar accuracy and the skewness with 4 correct significant digits for a large range of parameter values in the region $0.0001 \leq \gamma \leq 10$ and $q = 1, 2, \ldots, 15$. Each $\psi_q(\gamma,m,r)$ was computed separately and the recurrence relation (2.3) might then be used to check the accuracy.

A detailed report on the numerical integration is available separately (Beyer 1974).

b. The expectation.

It is stated in Section 2 that $E(\hat{\lambda}/\lambda) \leq 1$ and that $E(\hat{\lambda}/\lambda) \longrightarrow 1$ as λt or $q \longrightarrow \infty$. Table 1 gives a sample of values, we use $\gamma = \lambda t$ throughout.

It appears that $E(\hat{\lambda}/\lambda)$ is increasing and concave in q and increasing in γ, cf. the Remark at the end of Section 2 above.

We derived in Theorem 3.2 the inequality $E(\hat{\lambda}/\lambda) \geq a_q(\gamma)$, where $a_q(\gamma)$ was defined in the Theorem. $a_q(\gamma)$ is very accurate for $\gamma = 0$, but then it decreases and only for large γ is it of any numerical significance.

Theorem 4.1 states the approximation

$$E(\frac{\hat{\lambda}}{\lambda}) = 1 - \frac{\gamma e^{\gamma} - e^{\gamma} + 1}{q(e^{\gamma}-1)^2} + O(\frac{1}{q^2}) .$$

This approximation of the bias is remarkably good even for q as small as 5.

In Figs.1-3 we show the exact expectations together with the lower limit (called "Jensen inequality") and the approximation for $q = 1$, 2 and 10.

TABLE 1

Expectation of $\hat{\lambda}/\lambda$

r \ q	1	2	5	10	15	100
0	.6931	.8109	.9116	.9531	.9681	.9950
.5	.7173	.8309	.9235	.9600	.9730	
1	.7431	.8517	.9354	.9669	.9777	.9966
2	.7976	.8931	.9575	.9790	.9861	
4	.8991	.9587	.9866	.9938		
8	.9894	.9977	.9994	.9997		

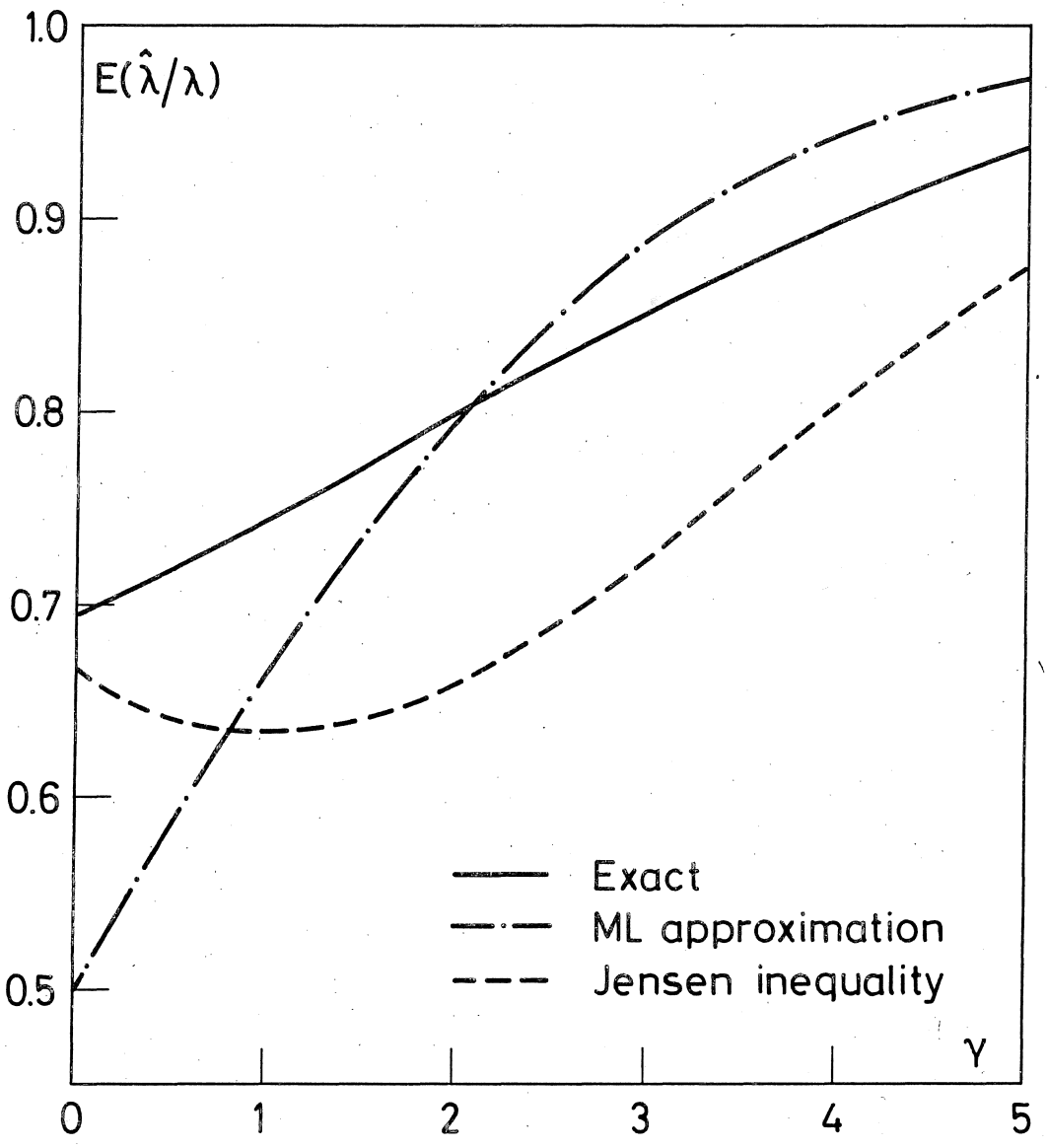


Figure 1. $E(\hat{\lambda}/\lambda)$ and approximations for q=1.

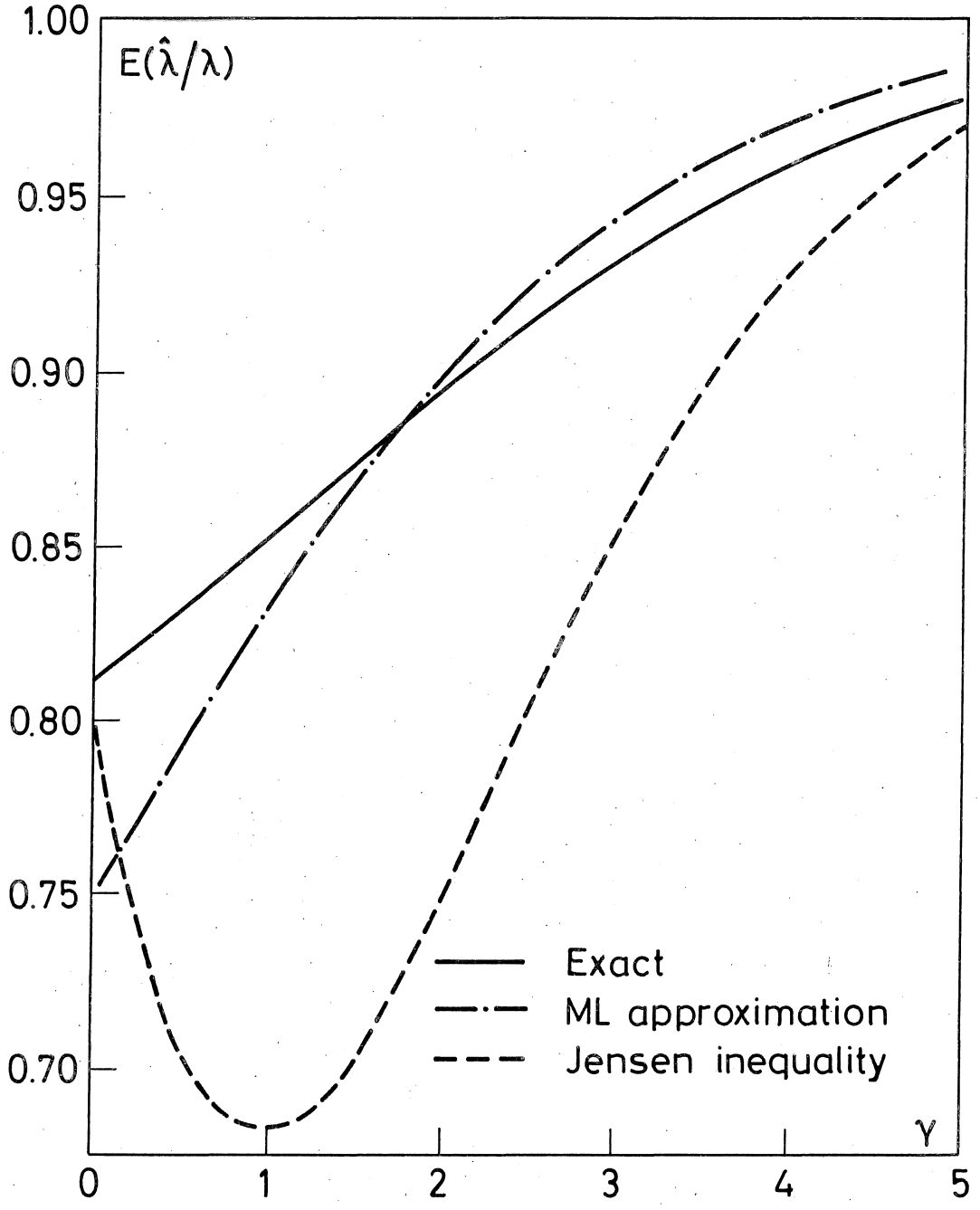


Figure 2. $E(\hat{\lambda}/\lambda)$ and approximations for q=2.

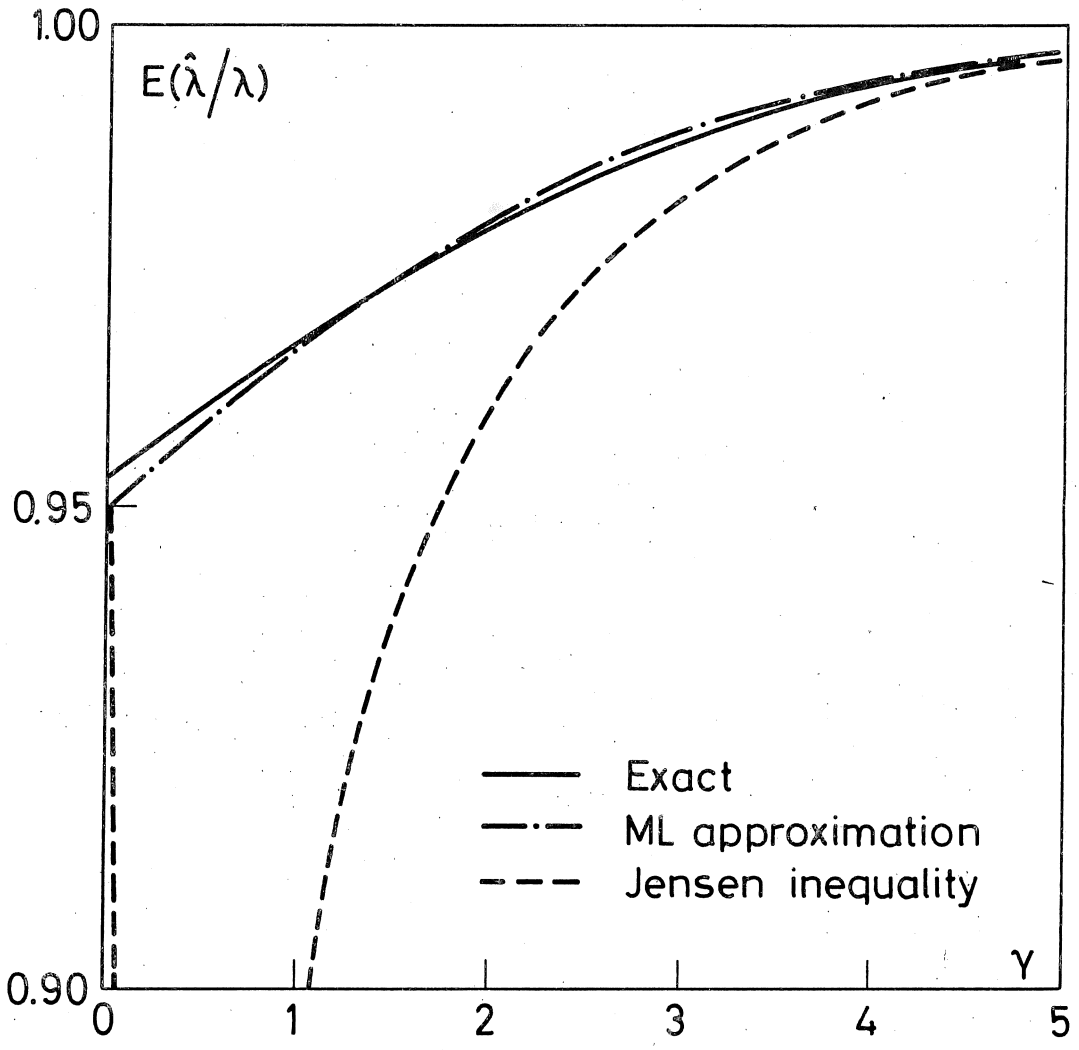


Figure 3. $E(\hat{\lambda}/\lambda)$ and approximations for q=10.

c. The variance

The asymptotic results of Section 2 concerning the variance are that $\mathrm{Var}(\hat{\lambda}) \longrightarrow 0$ as $q \longrightarrow \infty$ or $t \longrightarrow \infty$ and that $\mathrm{Var}(\hat{\lambda}) \longrightarrow \infty$ as t^{-1} as $t \longrightarrow 0$ for fixed q. Approximations to the variance were discussed in Section 4. We give in Table 2 the exact variances, the Student approximation (4.2), the inverse information approximation (4.1) and the refined approximation from Theorem 4.1 for $\gamma = 0.1, 1, 10, 20$ and selected values of q.

It is seen that for small γ and not too small values of q, the traditional inverse information approximation ($q \longrightarrow \infty$) is quite good. The second normal approximation is a definite improvement over the first for $\gamma = 0.1$ and 10 and $q \geq 2$ but for $\gamma = 1$, the second approximation only improves the first for large q. The Student approximation ($\gamma \longrightarrow \infty$) behaves better than the normal for $\gamma = 10$ and 20 but worse for $\gamma = 0.1$ and 1. We notice the rather irregular behaviour of the exact variance for $q = 1$ and large γ. In this case the limiting Student distribution has infinite variance.

d. The skewness

The coefficient of skewness $\beta(\gamma,q)$ is positive for small γ, (in fact, $\beta(\gamma,q) \longrightarrow \infty$ as $\gamma \longrightarrow 0$ by Theorem 2.3), then becomes negative with larger γ and reaches a minimum from which it approaches 0. This is, however, not true when $q = 1$ in which case $\beta \longrightarrow -\infty$ as $\gamma \longrightarrow \infty$. See Figs. 4 and 5. The root $\gamma = \gamma_q$ of $\beta(\gamma,q) = 0$ is decreasing in q and we conjecture that as $q \longrightarrow \infty$, it approaches 0.606, being the

TABLE 2

Variance of $\hat{\lambda}/\lambda$ and approximations.

γ	q	Student approx. (4.2) $t \to \infty$	Exact	First normal approx. (4.1) $q \to \infty$	Second normal approx. Theorem 4.1 $q \to \infty$
0.1	1	∞	4.976	9.508	1.184
	2	9.508	3.278	4.754	2.673
	5	2.378	1.614	1.902	1.569
	10	1.0565	0.8737	0.9508	0.8676
	15	0.6792	0.5987	0.6339	0.5969
1	1	∞	0.4689	0.5820	0.6660
	2	0.5820	0.2737	0.2910	0.3120
	5	0.1455	0.1165	0.1164	0.1198
	10	0.06466	0.05860	0.05820	0.05904
	15	0.04157	0.03904	0.03880	0.03917
	100	0.005879	0.005828	0.005820	0.005828
10*	1	∞	68.13	4.54	9.12
	2	4.540	4.786	2.270	3.416
	5	1.135	1.139	0.9080	1.091
	10	0.5045	0.5051	0.4540	0.4998
20**	1	∞	71.06	2.06	4.12
	2	2.061	2.061	1.031	1.546
	3	1.031	1.031	0.687	0.916

*All values have been multiplied by 10^5.

**All values have been multiplied by 10^9.

root of the approximation (4.3). This limiting value is marked on Fig. 5. Similarly, the minimum point γ_q^* of $\beta(\gamma,q)$ decreases in q with the conjectured limit of 2.78, which is the minimum of (4.3).

The approximation depends only on q through the factor $q^{-1/2}$ and therefore reflects neither these details in the sign pattern and minimum point nor the particular behaviour of $\beta(1,\gamma)$ as $\gamma \longrightarrow \infty$. Table 3 gives some information on the validity of the approximation. The general conclusion from the data presented there and our other investigations is that the approximation is only able to give a very rough guide, and is particularly misleading for small q.

TABLE 3

Exact and approximate coefficients of skewness

Exact values with 4 significant digits, approximate values to the same number of decimal places.

r \ q	1		5		10	
	Exact	Approx.	Exact	Approx.	Exact	Approx.
.0001	106.1	100.0	44.90	44.71	31.65	31.62
.1	3.190	2.613	1.249	1.169	.8582	.8264
.5	1.115	0.233	.2464	.1042	.1290	.0737
.6	.9417	.0116	.1523	.0052	.06018	.00367
.7	.7997	- .1703	.07322	- .07618	.002671	- .053867
.8	.6782	- .3237	.004713	- .144751	- .04676	- .10235
1	.4768	- .5690	- .1108	- .2545	- .1287	- .1800
1.5	.09583	- .96064	- .3280	- .4296	- .2740	- .3038
2	- .2118	-1.1628	- .4877	- .5200	- .3676	- .3677
5	-2.175	-0.914	- .7039	-. 4089	- .3735	- .2892
10	-14.83	-0.18	- .1474	- .0784	- .07329	- .05540

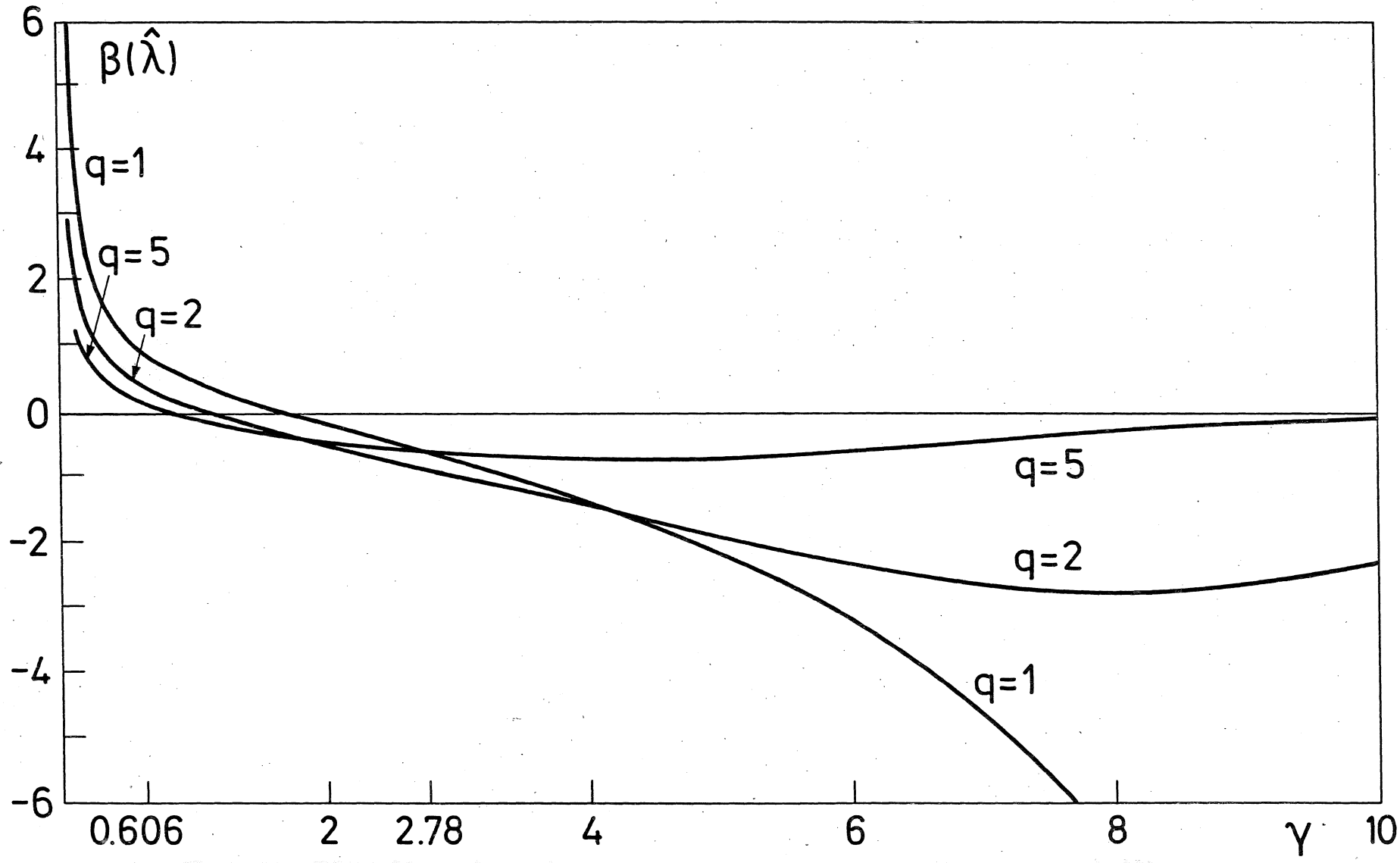


Figure 4. The coefficient of skewness $\beta(\hat{\lambda})$ for q=1, 2, and 5 and $0 \leq \gamma \leq 10$.

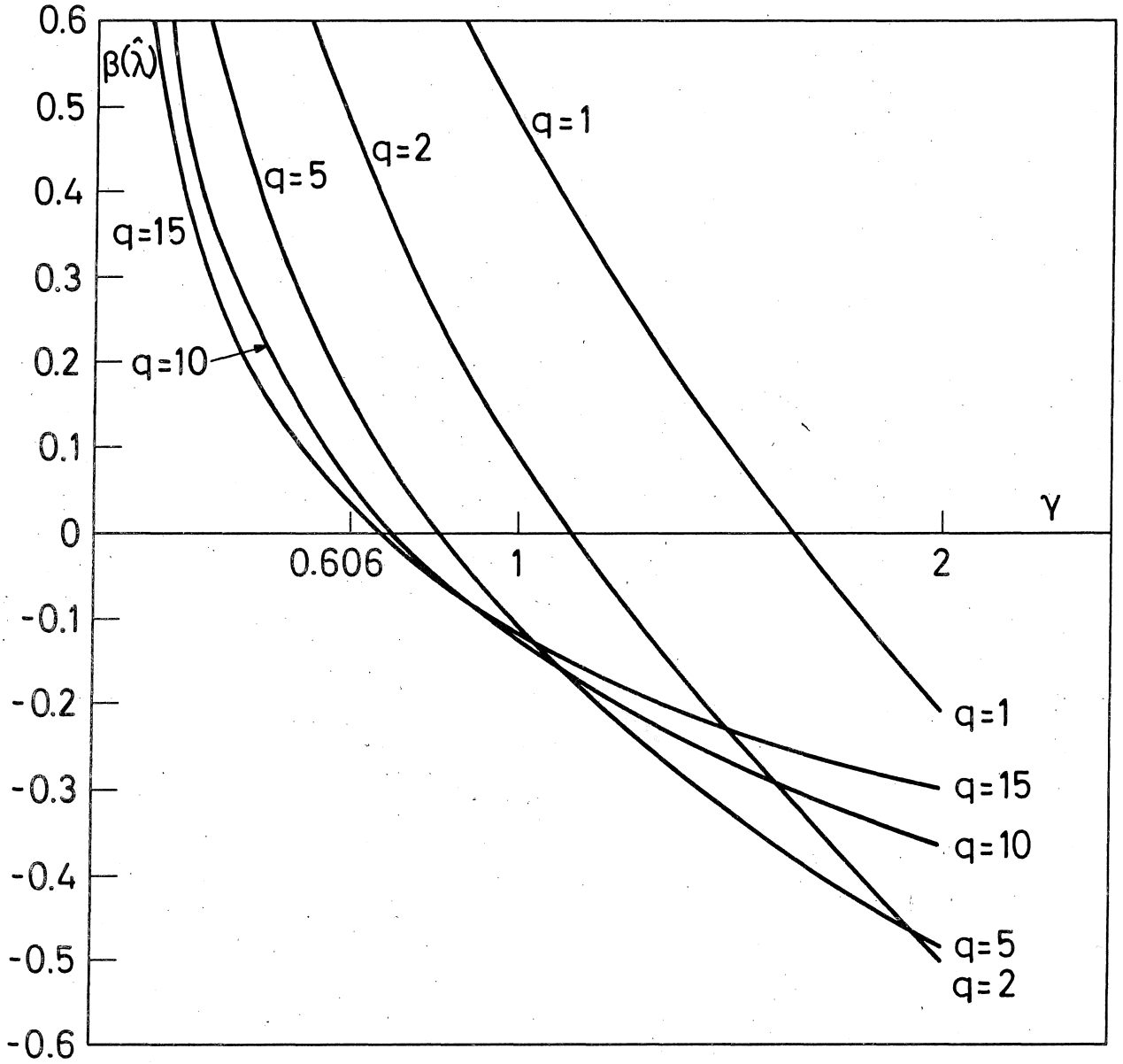


Figure 5. The coefficient of skewness $\beta(\hat{\lambda})$ for q=1,2,5,10,15 and $0 \leq \gamma \leq 2$.

Part II: The pure death process

6. The moments of the maximum likelihood estimator

The distribution of the minimal sufficient statistic (D_t, S_t) is given in the following way. With probability $e^{-q\mu t}$, $(D_t, S_t) = (0, qt)$. Otherwise the density, with a similar notation as in Section 2, is

$$\binom{q}{d} \mu^d t^{d-1} e^{-\mu s} g_d\left(\frac{s}{t} - q + d\right), \qquad d = 1,2,\ldots,q;\ 0 \leq s \leq qt. \tag{6.1}$$

Accordingly, the distribution of the maximum likelihood estimator D_t/S_t is given by being 0 with probability $e^{-q\mu t}$ and otherwise has the density

$$u^{-2} \sum_{d=1}^{q} \binom{q}{d} d\mu^d t^{d-1} e^{-\mu d/u} g_d\left(\frac{d}{tu} - q + d\right), \qquad a > 0 .$$

This density has a similar irregular shape as that described in Section 2 above concerning the birth process. These results and references to previous work leading up to them were given by Hoem (1969), see, however, also Puri (1968).

Let $\eta = \mu t$ and, as in Section 2,

$$\varphi(x) = \int_0^1 e^{-xt}\, dt = (1-e^{-x})/x, \qquad x \geq 0 .$$

Theorem 6.1. *Define* $\zeta_q(\eta,m,r) = \mu^{-r} E(D_t^{(m)}/S_t^r)$, *where* m *and* r *are nonnegative integers and* $a^{(x)} = a(a-1)\cdots(a-x+1)$. *Then* $\zeta_q(\eta,m,r) = 0$ *for* $m > q$ *and* $\zeta_q(\eta,m,r) = \infty$ *for* $r \geq q$. *Otherwise* $0 < \zeta_q(\eta,m,r) < \infty$ *and may be computed as*

$$\zeta_q(\eta,m,r) = q^{(m)}\eta^{m-r}\int_{\eta}^{\infty}\frac{(x-\eta)^{r-1}}{(r-1)!}\varphi^m(x)[e^{-x}+\eta\varphi(x)]^{q-m}\,dx\ . \qquad (6.2)$$

Proof. The proof is rather similar to the proof of Theorem 2.1 above and we shall only outline the steps.

Notice first that for $m > q$, both sides of the equation are 0. By (6.1) we get for $m \leq q$

$$\zeta_q(\eta,m,r) = \sum_{d=m}^{q}\frac{q^{(d)}}{(d-m)!}\eta^{d-r}e^{-\eta(q-d)}\int_0^{\infty}\frac{e^{-\eta x}}{(x+q-d)^r}g_d(x)\,dx\ .$$

For all values of $d < q$, the integrand in the above integral will be bounded and nonnegative over the support $[0,d]$. Since $g_q(x) = cx^{q-1}$ for $x \in (0,1)$, where $c > 0$ is a constant, the integral for $d = q$ becomes

$$c\int_0^1 e^{-\eta x}x^{q-1-r}\,dx + \text{a finite contribution}\ ,$$

so that the integral is finite if and only if $r \leq q-1$.

In that case, we now proceed as for the pure birth process to get

$$q^{(m)}\eta^{m-r}\int_{\eta}^{\infty}\frac{(x-\eta)^{r-1}}{(r-1)!}\varphi^m(x)\sum_{d=m}^{q}\frac{(q-m)^{(d-m)}}{(d-m)!}\{\eta\varphi(x)\}^{d-m}e^{-(q-d)x}\,dx$$

and the main result is obtained by completing the sum.

The moments may now be obtained as in the Corollary of Theorem 2.1 above.

As $\eta \longrightarrow \infty$, it is obvious (see e.g., Keiding (1975)), that $D_t \longrightarrow q$ so that $\hat{\mu} \longrightarrow q/S$ a.s. where S, being the sum of the exponentially distributed life-lengths of the q individuals that were alive at $t = 0$, is gamma-distributed (q, μ^{-1}). It follows from well-known properties of the gamma distribution that $\mu^{-r} E(q^{(m)}/S^r) = q^{(m)}/(q-1)^{(r)}$, which is in particular 0 for $m > q$ and ∞ for $r \geq q$. It is a consequence of the following Theorem that the moments derived for finite t converge to the moments of the limiting distribution.

Theorem 6.2. As $\mu t \longrightarrow \infty$,

$$\zeta_q(\mu t, m, r) = \mu^{-r} E(D_t^{(m)}/S_t^r) \longrightarrow q^{(m)}/(q-1)^{(r)} \quad .$$

Proof. This result is obvious by the remarks above for $m \geq q$ and for $r \geq q$. If $m \leq q$ and $r < q$ we may by a change of variable rewrite (6.2) as

$$\zeta_q(\eta, m, r) = q^{(m)} \int_1^\infty \frac{(x-1)^{r-1}}{(r-1)!} \{\eta\varphi(\eta x)\}^m [e^{-\eta x} + \eta\varphi(\eta x)]^{q-m} \, dx \ .$$

Choose t_0 so large that $e^{-\eta x} < \eta\varphi(\eta x) = (1-e^{-\eta x})/x$ for all $\eta > \eta_0 = \mu t_0$. The integrand is then bounded above by

$$\frac{(x-1)^{r-1}}{(r-1)!}\{\eta\varphi(\eta x)\}^q\, 2^{q-m} \leq \frac{2^{q-m}}{(r-1)!}\,\frac{(x-1)^{r-1}}{x^q}$$

for $\eta > \eta_0$, which is integrable (since $r < q$), and the result may be obtained by dominated convergence.

Corollary. As $\mu t \longrightarrow \infty$

$$E\{(\hat{\mu}/\mu)^r\} \longrightarrow q^r/\{(q-1)^{(r)}\}, \qquad r = 1,\ldots,q-1,$$

and in particular

$$E(\hat{\mu}/\mu) \longrightarrow q/(q-1)\,, \qquad q = 2,3,\ \ldots,$$

$$\mathrm{Var}(\hat{\mu}/\mu) \longrightarrow q^2/\{(q-1)^2\,(q-2)\}\,, \qquad q = 3,4,\ \ldots,$$

and the coefficient of skewness

$$\beta(\hat{\mu}) \longrightarrow 4(q-2)^{1/2}/(q-3)\,, \qquad q = 4,5,\ \ldots\ .$$

Theorem 6.3.

(a) $E(\hat{\mu}) \geq \mu$

(b) For fixed $q \geq 2$, $\lim_{\mu t \to 0} E(\hat{\mu}/\mu) = q\log(\frac{q}{q-1})$ and $\lim_{\mu t \to \infty} E(\hat{\mu}/\mu) = \frac{q}{q-1}$.

(c) For fixed $\lambda > 0$ and $t > 0$ $\lim_{q \to \infty} E(\hat{\mu}) = \mu$.

Proof. The second part of (b) was proved above and the rest of the proof is similar to the proof of Theorem 2.2. We omit the details.

7. The expectation of the inverse estimator

In a similar vein as the study in Section 3, one may investigate $\hat{\mu}^{-1}$ which is the maximum likelihood estimator of μ^{-1}. A careful study of small sample properties of $\hat{\mu}^{-1}$ was in fact performed by Bartholomew (1963), cf. also his references. Also, an inequality similar to the result of Theorem 3.2 above may be derived. However, the resulting lower limit is less than 1 and the result, thus being weaker than Theorem 6.3(a), is omitted.

8. Approximations from maximum likelihood theory

For fixed t and large q we may derive similar approximations to the moments of $\hat{\mu}$ as was done in Section 4 for $\hat{\lambda}$. The results are given in Theorem 8.1 below, the proof of which is very similar to that of Theorem 4.1 and hence is omitted.

Theorem 8.1. Let $\eta = \mu t$. Then as $q \to \infty$ for fixed η,

$$E(\hat{\mu}/\mu) = 1 + \frac{1 - e^{-\eta} - \eta e^{-\eta}}{q(1 - e^{-\eta})^2} + O(\frac{1}{q^2}) , \tag{8.1}$$

$$\begin{aligned} &\mathrm{Var}(\hat{\mu}/\mu) \\ &= \frac{1}{q(1-e^{-\eta})} \left[1 + \frac{1}{q(1-e^{-\eta})^3} \{4+(2\eta^2-10\eta-7)e^{-\eta}+(3\eta^2+10\eta+2)e^{-2\eta}+e^{-3\eta}\}\right] \\ &\quad + O(\frac{1}{q^3}) , \end{aligned} \tag{8.2}$$

and the coefficient of skewness

$$\beta(\hat{\mu}) = \frac{E[\{\hat{\mu} - E(\hat{\mu})\}^3]}{\{\mathrm{Var}(\hat{\mu})\}^{3/2}} = \frac{4(1-e^{-\eta}) - 3\eta e^{-\eta}}{q^{1/2}(1-e^{-\eta})^{3/2}} + O(\frac{1}{q}) . \tag{8.3}$$

Remark. It is easily seen that the approximation to the bias is greater than 1, cf. Theorem 6.3, and that the approximation to the skewness is positive for all η.

As $\eta \to \infty$, it is interesting to verify that the approximations are in accordance (to the order of approximation considered) with the exact results of the Corollary of Theorem 6.2.

Thus as $\eta \to \infty$, the approximate expectation tends to

$1 + 1/q + O(1/q^2) = q/(q-1) + O(1/q^2)$, the variance has the limit $(1/q)(1 + 4/q) + O(1/q^3) = q^2/\{(q-1)^2\ (q-2)\} + O(1/q^3)$ and for the skewness we obtain $4/q^{1/2} + O(1/q) = 4(q-2)^{1/2}/(q-3) + O(1/q)$. However, we shall discuss in Section 9 below that in some cases it is advantageous to amend the approximations by making them exact in the limit $\eta \to \infty$.

9. Numerical results

As for the birth process, we have tabulated the first three moments of $\hat{\mu}$, and a detailed table is available (Beyer 1974).

a. The numerical integration

The same adaptive Simpson quadrature procedure as described in Section 5a above was used. Some care had to be taken in order to assure that the cutoff point in approximating the integral was chosen properly. For details, see Beyer (1974).

b. The expectation

The expectation $E(\hat{\mu}/\mu)$ is an increasing function of η and a decreasing function of q; its limiting behavior is stated in Section 6. An improved version of the approximation (8.1) is given by

$$E\left(\frac{\hat{\mu}}{\mu}\right) \sim 1 + \frac{1 - e^{-\eta} - \eta e^{-\eta}}{(q-1)(1 - e^{-\eta})^2}$$

and is quite precise even for quite small values of q and all but very small values of η. Table 4 gives some values of the exact and approximate expectations.

c. The variance

The variance appears to decrease in q and η. The limit behaviour as $\eta \longrightarrow \infty$ is given in Section 6 and as $q \longrightarrow \infty$, $\mathrm{Var}(\hat{\mu})$ seems to go to 0. (This could of course be proved rigorously by going through the details of the derivations of Section 8.) Finally, $\mathrm{Var}(\hat{\mu}/\mu) = \infty$ for $q = 1$ and 2 and $\longrightarrow \infty$ as $\eta \longrightarrow 0$ for fixed $q \geq 3$.

TABLE 4

Exact and approximate expectations of $\hat{\mu}/\mu$ in the death process

η \ q	5		10		50	
	Exact	Approx.	Exact	Approx.	Exact	Approx.
0	1.116	1.125	1.054	1.056	1.0101	1.0102
.1	1.121	1.129	1.056	1.057	1.0105	1.0105
.5	1.144	1.146	1.064	1.065	1.0119	1.0119
1	1.172	1.165	1.074	1.073	1.0135	1.0135
2	1.214	1.197	1.091	1.088	1.0163	1.0162
5	1.248	1.243	1.109	1.108	1.0199	1.0198
∞	1.250	1.250	1.111	1.111	1.0204	1.0204

Two approximations to $\mathrm{Var}(\hat{\mu}/\mu)$ are given in (8.2). We present in Table 5 some values of the exact variance, the first approximation given by $[q(1-e^{-\eta})]^{-1}$, the amended first approximation $q^2/[(q-1)^2\,(q-2)\,(1-e^{-\eta})]$, the second approximation given by (8.2) and an amended version of that obtained by multiplication by $q^4/[(q-1)^2\,(q-2)\,(q+4)]$. The amendments all have the purpose of yielding exact results as $\eta \to \infty$.

It is seen that the amendments are worthwhile only for larger values of η and that the second approximation, thus amended, is quite good.

d. The skewness

The skewness $\beta(\eta,q)$ is infinitely large as $\eta \to 0$. For increasing η, $\beta(\eta,q)$ decreases until it reaches a positive minimum, then increases toward the asymptotic value given in Section 6. For $q \leq 3$, $\beta = \infty$. The value η_q where $\beta = 0$ is increasing in q and we conjecture that the limiting value is .528, being the minimum of the approximation (8.3).

The approximation is of course not able to reflect the moving minima. We show in Fig. 6 the exact coefficients of skewness for $q = 5$ and 10 and $0 \leq \eta \leq 2$. The asymptotes as $\eta \to \infty$ and the limiting minimum point are also indicated. The dashed curves represent an amended approximation obtained from (8.3) by replacing $q^{-1/2}$ by the limiting value $\beta(\infty,q) = (q-2)^{1/2}/(q-3)$. This approximation is quite good for $q = 10$ and not too small η.

TABLE 5

The Variance of $\hat{\mu}/\mu$ and approximations

η	q	Exact	First approx.	Am. First appr.	Second appr.	Am. Second appr.
.1	5	2.733	2.102	5.474	2.575	3.725
	10	1.186	1.051	1.622	1.169	1.289
	20	.5569	.5254	.6468	.5550	.5694
	50	.2150	.2102	.2280	.2149	.2158
1	5	.6065	.3164	.8240	.4571	.6613
	10	.2052	.1582	.2441	.1934	.2132
	20	.08914	.07910	.09738	.08789	.09017
	50	.03312	.03164	.03432	.03305	.03319
2	5	.5540	.2313	.6024	.3783	.5473
	10	.1679	.1157	.1785	.1524	.1680
	20	.06865	.05783	.07119	.06701	.06875
	50	.02470	.02313	.02509	.02460	.02471
∞	5	.5208	.2000	.5208	.3600	.5208
	10	.1543	.1000	.1543	.1400	.1543
	20	.06156	.0500	.06156	.06000	.06156
	50	.02169	.0200	.02169	.02160	.02169

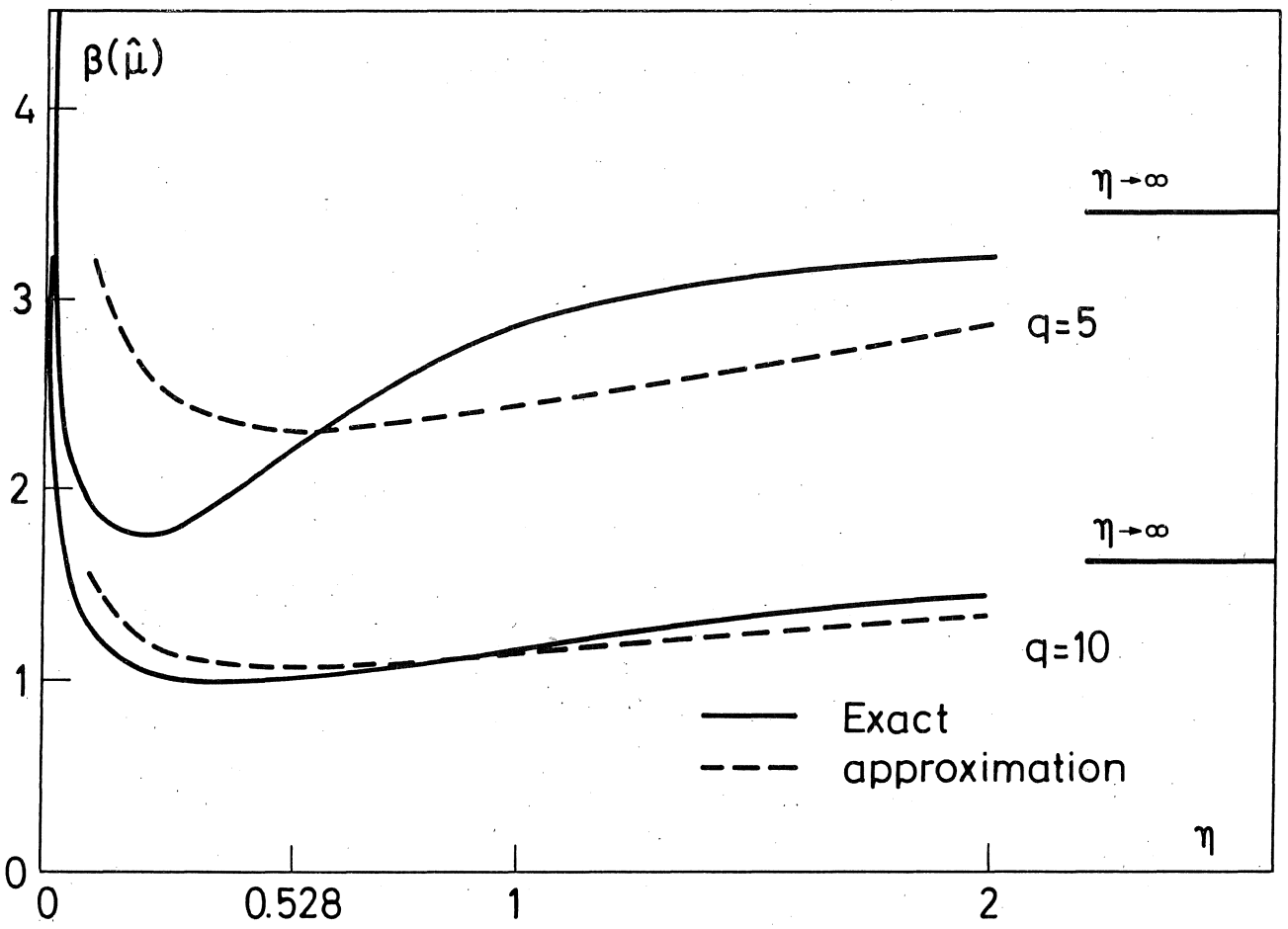


Figure 6. The coefficient of skewness $\beta(\hat{\mu})$ and approximations for q=5 and 10.

Acknowledgements.

We are grateful to Hanne Dalgas Christiansen and Niels Herman Hansen for discussions, in particular, concerning the numerical problems, and to Bradley Efron for providing us with early versions of his work on statistical curvature. NEUCC provided computing facilities.

Most of Niels Keiding's work was done at Stanford University with partial support from Danish Natural Sciences Research Council and U.S. National Sciences Foundation Grants GP-30711X-2 and GK-34363.

REFERENCES

Barlow, R. E. and Proschan, F. (1967). Exponential life test procedures when the distribution has monotone failure rate. J. Amer. Statist. Assoc. 62, 548-560.

Bartholomew, D. J. (1963). The sampling distribution of an estimate arising in life-testing. Technometrics 5, 361-372.

Bartlett, M. S. (1953a). On the statistical estimation of mean life-times. Phil. Mag. Ser. 7, 44, 249-262.

Bartlett, M. S. (1953b). Approximate confidence intervals. Biometrika 40, 12-19.

Beyer, J. E. (1974). The exact behavior of the maximum likelihood estimator in the pure birth process and the pure death process -- tables and graphs. Technical Report, IMSOR, Technical University of Denmark.

Deemer, W. L., Jr. and Votaw, D. F., Jr. (1955). Estimation of parameters of truncated or censored exponential distributions. Ann. Math. Statist. 26, 498-504.

Efron, B. (1974). Defining the curvature of a statistical problem (with applications to second order efficiency). Technical Report, Department of Statistics, Stanford University.

Epstein, B. (1960). Test for the validity of the assumption that the underlying distribution of life is exponential. I. Technometrics 2, 83-101.

Govindarajulu, Z. (1962). The reciprocal of the decapitated negative binomial variable. J. Amer. Statist. Assoc. 57, 906-913. Correction note: J. Amer. Statist. Assoc. 58, 1162.

Govindarajulu, Z. (1964). A supplement to Mendenhall's bibliography on life testing and related topics. J. Amer. Statist. Assoc. 59, 1231-1291. Correction note: J. Amer. Statist. Assoc. 60, 1965, 1249-50.

Grenander, U. (1950). Stochastic processes and statistical inference. Arkiv för Matematik 1, 195-277.

Haldane, J.B.S. and Smith, S.M. (1956). The sampling distribution of a maximum likelihood estimate. Biometrika 43, 96-103.

Hoem, J. M. (1969). The sampling distribution of an estimator arising in connection with the truncated exponential distribution. Ann. Math. Statist. 40, 702-703.

Keiding, N. (1974). Estimation in the birth process. Biometrika 61, 71-80.

Keiding, N. (1975). Maximum likelihood estimation in the birth-and-death process. Ann. Statist. 3.

Littell, A. S. (1952). Estimation of the T-year survival rate from follow-up studies over a limited period of time. Human Biology 24, 87-116.

Lyness, J. N. (1969). Notes on the adaptive Simpson quadrature routine. J. Assoc. Comp. Mach. 16, 483-495.

Lyness, J. N. (1970). Algorithm 379, SQUANK (Simpson quadrature used adaptively-noise killed). Comm. Assoc. Comp. Mach. 13, 260-263.

Mendenhall, W. (1958). A bibliography on life testing and related topics. Biometrika 45, 521-543.

Mendenhall, W. and Hader, R.J. (1958). Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data. Biometrika 45, 504-520.

Puri, P. S. (1966). On the homogeneous birth-and-death process and its integral. Biometrika 53, 61-71.

Puri, P. S. (1968). Some further results on the birth-and-death process and its integral. Proc. Camb. Phil. Soc. 64, 141-154.

Rider, P. R. (1962). Expected values and standard deviations of the reciprocal of a variable from a decapitated negative binomial distribution. J. Amer. Statist. Assoc. 57, 439-445.

Zelen, M. and Dannemiller, M. C. (1961). The robustness of life testing procedures derived from the exponential distribution. Technometrics 3, 29-49.

Niels Keiding

Institute of Mathematical Statistics

Universitetsparken 5

DK - 2100 Copenhagen Ø

Denmark