Simple and accurate inference for the mean of the gamma model

D.A.S. Fraser
N. Reid
Department of Statistics, University of Toronto
Toronto, Canada M5S 1A1

A. Wong
Department of Mathematics and Statistics, York University
North York, Canada M3J 1P3

Abstract

The two parameter gamma model is widely used in reliability, environmental, medical, and other areas of statistics. It has a two dimensional sufficient statistic, and a two dimensional parameter which can be taken to describe shape and mean. This makes it closely comparable to the normal model, but it differs substantially in that the exact distribution for the minimal sufficient statistic is not available. Some recent and simple asymptotics are used to derive a new approximation for observed levels of significance and confidence intervals for the mean parameter of the model; the method extends to cover the shape parameter. The approximation is accessible and easy to apply, often easier than first order methods. Simulations show that the approximation has high accuracy, typically better than other third order methods. Comparisons are given using real data.

Keywords: Asymptotic approximation; Averaging; Canonical parameter; Conditional likelihood; Exponential model; Location model; Saddlepoint approximation; Significance function; Third order approximation.

1. Introduction

Consider a sample $(y_1, \ldots, y_n)$ from the two parameter gamma model with shape β and mean μ. The joint density is

$$f(y_1, \ldots, y_n; \beta, \mu) = \Gamma^{-n}(\beta)(\beta/\mu)^{n\beta} \exp(\beta t_1 - \beta t_2/\mu) \prod (1/y_i) \tag{1}$$

where $(t_1, t_2) = (\sum \log y_i, \sum y_i)$ is the minimal sufficient statistic for (β, μ). An alternate version of the minimal sufficient statistic for (β, μ) is given by (d, t_2) where d is the log off-set of the arithmetic mean from the geometric mean,

$$d = \log(\sum y_i/n) - \log(\prod y_i)^{1/n} = \log(t_2/n) - t_1/n \ .$$

For fixed β, the density of $\log y_i$ has location model form and it follows that the conditional density for t_2 given the residuals and thus given d is readily available. We then have that the joint density factors as $f(d, t_2; \beta, \mu) = f(d; \beta) f(t_2|d; \beta, \mu)$ into marginal and conditional components,

$$\begin{aligned} f(d; \beta) &= \Gamma(n\beta)\Gamma^{-n}(\beta) n^{-(n\beta - 1/2)} \exp(-n\beta d) h_n(d) \\ f(t_2|d; \beta, \mu) &= \Gamma^{-1}(n\beta)(\beta/\mu)^{n\beta} \exp(n\beta \log t_2 - \beta t_2/\mu) t_2^{-1} \end{aligned} \tag{2}$$

where $h_n(d)$ requires $(n-2)$- dimensional integration and is available only for very small values of n (Jensen, 1986). Note that the distribution of t_2 given d is free of d and thus t_2 and d are statistically independent. Also note that the joint distribution of (t_1, t_2) then has the form

$$f(t_1, t_2; \beta, \mu) = \Gamma^{-n}(\beta)(\beta/\mu)^{n\beta} \exp(\beta t_1 - \beta t_2/\mu) \frac{1}{t_2} \frac{1}{\sqrt{n}} h_n \Big\{ \log\left(\frac{t_2}{n}\right) - \frac{t_1}{n} \Big\} \ , \tag{3}$$

which is an exponential family model with canonical variable (t_1, t_2) and canonical parameter $(\beta, -\beta/\mu)$.

The log likelihood function in terms of $\theta = (\beta, \mu)$ is

$$l(\theta) = l(\beta, \mu) = -n \log \Gamma(\beta) + n\beta \log \beta - n\beta \log \mu + \beta t_1 - \beta t_2/\mu \ . \qquad (4)$$

For later reference, the first and negative second derivatives of $l(\theta)$ are

$$l_\theta(\theta) = \begin{pmatrix} l_\beta(\theta) \\ l_\mu(\theta) \end{pmatrix} = \begin{pmatrix} -ng(\beta) + n + n\log\beta - n\log\mu + t_1 - t_2/\mu, \\ -n\beta/\mu + \beta t_2/\mu^2 \end{pmatrix} \qquad (5)$$

$$j_{\theta\theta}(\theta) = \begin{pmatrix} -l_{\beta\beta}(\theta) & -l_{\beta\mu}(\theta) \\ -l_{\mu\beta}(\theta) & -l_{\mu\mu}(\theta) \end{pmatrix} = \begin{pmatrix} ng'(\beta) - n/\beta & n/\mu - t_2/\mu^2 \\ n/\mu - t_2/\mu^2 & -n\beta/\mu^2 + 2\beta t_2/\mu^3 \end{pmatrix} \qquad (6)$$

where $g(t) = d\log\Gamma(t)/dt$ is the digamma function; the subscripts θ and $\theta\theta$ denote the first and second order differentiation with respect to θ.

Perhaps the most straightforward method for testing μ is based on the first order normal approximation for the maximum likelihood estimate $\hat{\theta} = (\hat{\beta}, \hat{\mu})$ obtained from the equation $l_\theta(\hat{\theta}) = 0$: this is discussed in Gross & Clark (1975). The full maximum likelihood estimate $\hat{\theta}$ is asymptotically normal with mean θ and variance $j^{\theta\theta}(\hat{\theta}) = j_{\theta\theta}^{-1}(\hat{\theta})$, so $\hat{\mu}$ is asymptotically normal with mean μ and variance $\hat{\mu}^2/n\hat{\beta}$. This method has first order accuracy, meaning the error in the approximation is of order $O(n^{-1/2})$.

Several methods of inference can be obtained from the location model factorization (2) in which the simple gamma distribution $f(t_2|d; \beta, \mu) = f(t_2; \beta, \mu)$ can be viewed as containing all the information concerning μ; however, to reliably test a value μ, the effect of the nuisance parameter β needs to be accommodated. Grice & Bain (1980) approached this by first substituting the maximum likelihood estimate $\hat{\beta}$ into (2), thus treating $\hat{\beta}\bar{y}/\mu$ as gamma $(n, \hat{\beta})$. They then used Monte Carlo simulations to adjust for the effect of this substitution. Shiue, Bain & Engelhardt (1988) reviewed this approximation and extended the method to obtain confidence intervals for the difference in means for two independent gamma models. Shiue & Bain (1990) discussed further the Grice & Bain (1980) method and fine tuned it using a Weibull approximation for the negative log of the probability integral transform of $\hat{\beta}\bar{y}/\mu$.

In addition to its use in applications, the two parameter gamma is often taken as a test case for higher order approximations. It is a relatively simple example of an exponential family model in which the parameter of interest is a ratio of canonical parameters, and for which exact calculations are not possible.

Wong (1993) used a parameter averaging method to eliminate β from $f(t_2|d;\beta,\mu) = f(t_2;\beta,\mu)$: it is shown in Fraser & Wong (1995) that the averaging method yields second order accuracy; i.e. with relative error $O(n^{-1})$. The averaging method has some similarities to Grice & Bain's (1980) method and also to the partially Bayes method of Cox (1975). Another second order method based on Bartlett adjustment of the likelihood ratio statistic is given in Jensen & Kristensen (1991).

Jensen (1986) obtained a third order approximation, with relative error $O(n^{-3/2})$, by examining the conditional density of t_2 given $(t_1 - t_2/\mu_0)$, which is exponential with canonical parameter $\gamma = \beta(1/\mu_0 - 1/\mu)$ for fixed $\mu = \mu_0$. The explicit form of this conditional density is typically unavailable except for small n. By using saddlepoint methods for h_n and numerical integration, Jensen derived tables for constructing 95% and 98% confidence intervals for μ with sample sizes $n = 10, 20, 40$ and ∞. A different evaluation of the conditional density is needed for each value of μ_0, and confidence intervals are constructed iteratively. The method is of third order accuracy but the calculations for other than tabulated values are substantial. Jensen (1991) discussed another saddlepoint approximation which simplifies the calculation of the observed level of significance. This method is accurate only to second order, but has the advantage that the error is uniformly bounded in a large deviation region. Jensen (1992) examined large deviation properties of some third order tests. Barndorff-Nielsen (1986) derived a third order approximation based on an adjusted version of the signed root of the log-likelihood ratio statistic that is similar to the one outlined below.

Fraser & Reid (1995) used tangent exponential models to derive a fairly simple third order approximation to the significance level for testing an interest parameter. An advantage of the method is that it needs only maximum likelihood estimates and observed information for a recalibrated parameter. The development in Fraser & Reid (1995) is substantially more general than is required here, and this obscures to some extent the simplicity of the method. An outline of the method and its application to the gamma mean problem are given in Section 2, and some numerical comparisons are provided in Section 3.

2. Third order inference for a component parameter

Consider a model with parameter $\theta = (\lambda, \psi)$ of dimension p, where ψ is a scalar parameter of interest. Many recently developed third order methods for inference on ψ can be presented in terms of one or the other of the formulas

$$\Phi^1(R,Q) = \Phi(R) + \phi(R)\{R^{-1} - Q^{-1}\} \quad , \quad \Phi^2(R,Q) = \Phi\{R - R^{-1}\log(R/Q)\} \qquad (8)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions, and R and Q are statistics whose first order distribution is the standard normal: their exact form depends on the type of problem. The two formulas were introduced respectively by Lugannani & Rice (1980) for a one parameter exponential model and by Barndorff-Nielsen (1986) for a reduced model after a sufficiency or ancillary reduction. As $\Phi(R,R)$, $\Phi_Q(R,R)$, $\Phi_{QQ}(R,R)$ are the same for the two definitions in (8), it follows easily that the approximations are third order equivalent. Various expressions have been developed for R and Q appropriate to problems of varying generality; in most cases R is the signed square root of a log-likelihood ratio statistic. Use of the approximations have generally been hampered by the difficulty in calculating Q, which is typically some version of a standardized maximum likelihood departure.

A reduction by approximate ancillarity is usually needed to compute Q, although it is not needed for the gamma mean problem, since the sufficient statistic and the full parameter are of the same dimension. The third order procedure developed by Fraser & Reid (1995, Sec.5) gives a method for obtaining the conditionality reduction without explicit computation of the approximate ancillary statistic. A further reduction of dimension to a scalar pivotal quantity is necessary for testing the interest parameter ψ: this is outlined in Fraser & Reid (1995, Sec.6). The method is relatively simple in requiring only the observed likelihood function $l^0(\theta) = l(\theta; y^0)$ and the likelihood gradient $l^0_V(\theta) = l_V(\theta; y^0)$. In general problems the directions V of differentiation record tangents to a first derivative ancillary, but in the gamma mean problem it suffices to consider the likelihood gradient relative to the minimum sufficient statistic.

The computation of Q is obtained from a reparameterization that is determined by the gradient of the log-likelihood in the direction of the minimal sufficient statistic, evaluated at the data point:

$$\varphi = l_{;y}(\theta; y^0) = \frac{\partial}{\partial y} l(\theta; y)\Big|_{y^0} . \tag{9}$$

We define the related Jacobians

$$J_\theta(\theta) = \frac{\partial \varphi}{\partial \theta'} = l_{\theta;y}(\theta; y^0) \ , \ J^\theta(\theta) = \frac{\partial \theta}{\partial \varphi'} = J_\theta^{-1}(\theta) \tag{10}$$

so that the observed information can be recalibrated in the new parameterization:

$$|j_{(\theta\theta)}(\hat\theta)| = |j_{\theta\theta}(\hat\theta)||J_\theta(\hat\theta)|^{-2} \tag{11}$$

$$|j_{(\lambda\lambda)}(\hat\theta_\psi)| = |j_{\lambda\lambda}(\hat\theta_\psi)||J'_\lambda(\hat\theta_\psi) J_\lambda(\hat\theta_\psi)|^{-2} \tag{12}$$

where J_λ records the columns of J_θ associated with λ. We also need the signed square root of the profile log-likelihood ratio statistic

$$R = \mathrm{sgn}\,(\hat{\psi} - \psi_0)\{2[l^0(\hat{\theta}) - l^0(\hat{\theta}_{\psi_0})]\}^{1/2} \tag{13}$$

and the standardized maximum likelihood departure

$$Q = \{\hat{\chi} - \hat{\chi}_{\psi_0}\}\,\frac{|j_{(\theta\theta)}(\hat{\theta})|^{1/2}}{|j_{(\lambda\lambda)}(\hat{\theta}_{\psi_0})|^{1/2}} \tag{14}$$

for a scalar parameter χ that measures departure from the value ψ_0. The parameter χ is a scaled linear version of $\varphi(\theta)$:

$$\chi = J^{\psi}(\hat{\theta}_{\psi_0})\varphi(\theta)/|J^{\psi}(\hat{\theta}_{\psi_0})|\ , \tag{15}$$

that corresponds to $d\psi$ at $\theta = \hat{\theta}_{\psi_0}$. The vector J^{ψ} is the row of J^{θ} that corresponds to ψ. The significance obtained from (8) is viewed as probability left of the data point, appropriately defined; for some discussion, see Fraser (1991). Significance would correspond to values near 0 or 1.

For testing $\mu = \mu_0$ in the gamma model (1), (2) or (3), the parameter of interest ψ is replaced by μ, and the nuisance parameter λ is replaced by β in the formulas given above. The maximum likelihood estimates $\hat{\theta} = (\hat{\beta}, \hat{\mu})$, $\hat{\theta}_{\mu_0} = (\hat{\beta}_{\mu_0}, \mu_0)$ are routinely obtained by solving $l_\theta(\hat{\theta}) = 0$ and $l_\beta(\hat{\beta}_{\mu_0}, \mu_0) = 0$ from (5). The observed information matrices $j_{\theta\theta}(\hat{\theta})$ and $j_{\beta\beta}(\hat{\theta}_{\mu_0})$ are obtained by substituting $\hat{\theta}$ in (6) and $\hat{\theta}_{\mu_0}$ in $j_{\beta\beta}(\theta)$. The new parameterization is the canonical parameterization $\varphi = (\beta, -\beta/\mu)'$ with Jacobians

$$J_\theta(\theta) = \begin{pmatrix} 1 & 0 \\ -1/\mu & \beta/\mu^2 \end{pmatrix}\ ,\ J^{\theta}(\theta) = \begin{pmatrix} 1 & 0 \\ \mu/\beta & \mu^2/\beta \end{pmatrix}\ . \tag{16}$$

and the scalar version of φ specific to $\mu = \mu_0$ is

$$\chi = \left(\frac{\mu_0}{\hat{\beta}_{\mu_0}}\ ,\ \frac{\mu_0^2}{\hat{\beta}_{\mu_0}}\right)\begin{pmatrix} \beta \\ -\beta/\mu \end{pmatrix}\Big/\left\{\left(\frac{\mu_0}{\hat{\beta}_{\mu_0}}\right)^2 + \left(\frac{\mu_0^2}{\hat{\beta}_{\mu_0}}\right)^2\right\}^{1/2} = \frac{\beta - \beta\mu_0/\mu}{(1+\mu_0^2)^{1/2}}. \tag{17}$$

After some simplification the expression for Q is given by

$$Q = \sqrt{n\hat{\beta}}(\hat{\mu}/\mu_0 - 1)\{g'(\hat{\beta}) - \hat{\beta}^{-1}\}^{1/2}\{g'(\hat{\beta}_{\mu_0}) - \hat{\beta}_{\mu_0}^{-1}\}^{-1/2} \tag{18}$$

which is the same as the Q used by Barndorff-Nielsen (1986, eq. 3.28). This equivalence provides an alternate derivation of the third order accuracy of (8), as well as an alternate route to Barndorff-Nielsen's r^* approximation in k, k exponential families. The p-value or significance for testing μ_0 is then given by (8) using the signed log profile likelihood ratio (13) and the standardized maximum likelihood departure (18).

3. Examples

Example 1

We first examine four of the approximations in the small sample case with $n = 2$, where exact tail probabilities are readily computable; we do not include Jensen's third order procedure due to the complexity of the associated calculations. As an observed data point we use $(y_1^0, y_2^0) = (1, 4)$.

For the present sample size, the function $h_2(d)$ is available by marginalizing from (y_1, y_2) to the scalar variable $d = \log\{(y_1 + y_2)/2\} - \log(y_1 y_2)^{1/2}$. This gives

$$h_2(d) = 2\sqrt{2}(1 - e^{-2d})^{-1/2} \ .$$

The conditional relative density of t_2 given $t_1 - t_2/\mu_0$, is available exactly:

$$f_c(t_2|(t_1 - t_2/\mu_0); \beta, \gamma) = C \exp(\gamma t_2)(t_2^2 - 4e^{t_1})^{-1/2}$$

.

The sample space for (t_2, t_1) is the region beneath and to the right of the curve $d = 0$ or $t_2^2 = 4e^{t_1}$ in Figure 1; the observed data point $(t_2^0, t_1^0) = (5, \log 4)$ is also plotted. The conditional distributions for testing $\mu = 1, 3, 5, 7$ and 9 are distributions along the lines correspondingly marked through the observed data point. These conditional distributions are recorded on the right side, in each case with the location of the data marked *; the left tail probabilities $p(1)$, $p(3)$, $p(5)$, $p(7)$ and $p(9)$ were obtained by numerical integration. It is clear from Figure 1 that the conditional distributions are very far from normal.

Table 1 records the significance $p(\mu)$ for $\mu = 1, 3, 5, 7$ and 9 obtained from the exact distribution and from six approximations: the first order method, Shiue & Bain's Weibull approximation, the parameter averaging method, Jensen's third order method, and the new third order methods using (13) and (14) with Φ^1 or Φ^2 in (8). The third order method using Φ^1 gives the best approximations. The averaging method also gives satisfactory results, while the first order method and Shiue & Bain's (1990) Weibull approximation are not very good.

Table 1: Approximate and exact p-values for a sample of size 2

	μ				
Method	1	3	5	7	9
First order mle	0.905	0.331	0.014	4.09×10^{-5}	6.37×10^{-9}
Shiue & Bain	0.944	0.429	0.257	0.184	0.145
Averaging	0.892	0.512	0.330	0.280	0.257
Jensen	0.908	0.464	0.292	0.231	0.200
From (13),(14) with Φ^1	0.910	0.466	0.291	0.230	0.200
From (13),(14) with Φ^2	0.911	0.464	0.280	0.215	0.182
Exact	0.901	0.464	0.318	0.256	0.225

Example 2

In order to examine smaller significance levels than those attained in Example 1, it seems appropriate to increase the sample size. Table 2 summarizes the results for 10 observations from a gamma with mean 1 and shape parameter $\beta = 2$. For 10,000 simulations, the observed p-values for testing $\mu = 1$ were computed by the various approximate methods, and the percent of 2-sided p-values less than the normal 1%, 2.5%, and 5.0% values are recorded. The averaging method is satisfactory; the third order methods are better and within two sigma limits. Again the first order method and Shiue & Bain's (1990) Weibulll approximation are not good.

Table 2: Results of 10,000 simulations with sample size 10

	% of p-values					
Method	$< 0.5\%$	$< 1.25\%$	$< 2.5\%$	$> 97.5\%$	$> 98.75\%$	$> 99.5\%$
First order mle	5.00	5.73	8.53	1.52	0.74	0.35
Shiue & Bain	0.28	0.70	1.19	1.20	0.67	0.30
Averaging	0.29	0.69	2.18	2.05	0.95	0.41
From (13),(14) with Φ^1	0.37	0.90	2.30	2.41	1.13	0.47
From (13),(14) with Φ^2	0.37	0.91	2.30	2.41	1.13	0.47

Example 3

Grice & Bain (1980) discuss the Gross & Clark (1975) data on survival times on 20 mice exposed to 240 rads of gamma radiation:

152	152	115	109	137	88	94	77	160	165
125	40	128	123	136	101	62	153	83	69.

The 95% confidence intervals are recorded in Table 3 for six different approximation methods: the first order method, Shiue & Bain's (1990) Weibull approximation, the method of parameter averaging, Jensen's third order method, and the third order methods from equations (13) and (14) with Φ^1 and Φ^2 in (8). The three third order methods are very close. In terms of ease of calculation, the two proposed methods using maximum likelihood estimates and information are computationally simple.

Table 3: Confidence intervals for survival time data

	95% confidence interval for μ
First order mle	(96.7 , 130.9)
Shiue & Bain	(96.7 , 130.8)
Averaging	(97.0 , 134.7)
Jensen	(96.8 , 133.5)
From (13),(14) with Φ^1	(97.0 , 134.7)
From (13),(14) with Φ^2	(97.2 , 134.2)

References

Barndorff-Nielsen, O.E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307-322.

Cox, D.R. (1975). A note on partially Bayes inference and linear models. *Biometrika* **62**, 651-654.

Fraser, D.A.S. (1991). Statistical inference: Likelihood to significance, *J. Amer. Statist. Assoc.* **86**, 258-265.

Fraser, D.A.S. and Reid, N. (1995). Ancillaries and third order significance. *Utilitas Mathematica* **47**, 33-53.

Fraser, D.A.S. and Wong, A. (1995). On accuracy of approximate Studentization. Submitted for publication.

Grice, J.V. and Bain, L.J. (1980). Inference concerning the mean of the gamma distribution. *J. Amer. Statist. Assoc.* **75**, 929-933.

Gross, A.J. and Clark, V.A. (1975). *Survival Distributions: Reliability Applications in the Biomedical Sciences.* New York: John Wiley and Sons.

Jensen, J.L. (1986). Inference for the mean of a gamma distribution with unknown shape parameter. *Scand. J. Statist.* **13**, 135-151.

Jensen, J.L. (1991). Uniform saddlepoint approximations and log-concave densities. *J.R. Statist. Soc. B* **53**, 157-172.

Jensen, J.L. (1992). The modified signed likelihood statistic and saddlepoint approximation. *Biometrika* **79**, 693-703.

Jensen, J.L. and Kristensen, L.B. (1991). Saddlepoint approximations to exact test and improved likelihood ratio tests for the gamma distribution. *Comm. Statist. – Theory and Methods* **20**, 1515-1532.

Lugannani, R. and Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. Appl. Probab.* **12**, 479-490.

Shiue, W.K., and Bain, L.J. (1990). Simple approximate inference procedures for the mean of the gamma model. *J. Statist. Comput. Simul.* **34** 67-73.

Shiue, W.K., Bain, L.J. and Engelhardt, M. (1988). Test of equal gamma-distribution means with unknown and unequal shape parameters. *Technometrics* **30**, 169-174.

Wong, A. (1993). A note on inference for the mean parameter of the gamma distribution. *Statist. & Probab. Letters* **17**, 61-66.