TECHNOMETRICS ©, VOL. 22, NO. 1, FEBRUARY 1980

Bayesian Analysis of the Two-Parameter Gamma Distribution

Robert B. Miller
Department of Statistics and
Graduate School of Business
University of Wisconsin—Madison
Madison, WI 53706

This paper presents a Bayesian analysis of shape, scale, and mean of the two-parameter gamma distribution. Attention is given to conjugate and "non-informative" priors, to simplifications of the numerical analysis of posterior distributions, and to comparison of Bayesian and classical inferences.

KEY WORDS

Gamma distribution
Bayesian analysis

1. INTRODUCTION

The gamma distribution has found extensive application in reliability and life testing (see Engelhardt and Bain, 1977, Glaser, 1976, and Gross and Clark, 1975, for example) and in insurance (see Ammeter, 1970, and Seal, 1969). Maximum likelihood estimation of the parameters of the gamma distribution is discussed by Choi and Wette (1969), Gross and Clark (1975), and Johnson and Kotz (1970), among others. Hypothesis testing has been considered recently by Cox and Hinkley (1974, p. 125), Engelhardt and Bain (1977), Bain and Engelhardt (1975), and Glaser (1976*a*, *b*). The testing results can be used to construct confidence intervals for the parameters of the gamma distribution.

Damsleth (1975) considers a Bayesian analysis of the two-parameter gamma family, but the numerical calculations involved in his approach are prohibitive. In this paper I present an approach that requires relatively little computing involving subroutines that can be easily written and that in fact are already available on most computers.

Increasingly, statistical techniques are being made accessible to users via interaction with computer systems. Much of what formerly had to be looked up in tables (and often obtained by interpolation) is now available at the touch of a button. The methods presented here can be easily incorporated into an interactive computer system. What is more, even though the approach is Bayesian, classical inferences can be obtained by choosing suitable "non-informative" priors.

Received November 1977; revised March 1979

2. THEORY

Let the population density have the gamma form

$$g(x|\alpha,\theta) = \begin{cases} \theta^\alpha x^{\alpha-1}\exp(-x\theta)/\Gamma(\alpha), & x>0, \alpha>0 \\ 0 \quad , & \text{otherwise,} \end{cases} \tag{1}$$

where α is a shape parameter and θ is the reciprocal of a scale parameter. Denote the population mean by $\mu = \alpha/\theta$. If $\tilde{x}_1, \cdots \tilde{x}_n$ denotes a random sample of fixed size n from the population, then given $\tilde{x}_i = x_i$, $i = 1, \cdots, n$, any likelihood function of α and θ is proportional to the kernel

$$\prod_{i=1}^{n} g(x_i|\alpha, \theta) = \theta^{n\alpha}p^{\alpha-1}\exp(-s\theta)/[\Gamma(\alpha)]^n, \tag{2}$$

where $s = \sum_{i=1}^n x_i$ and $p = \pi_{i=1}^n x_i$.

Damsleth (1975) exhibits a conjugate class of distributions for α and θ which, however, can be extended and enriched according to the prescriptions in sections 3.2.3 and 3.2.4 of Raiffa and Schlaifer (1961). A very general conjugate class is defined by the joint density

$$f'(\alpha,\theta) \propto \theta^{\nu'\alpha-1}(p')^{\alpha-1}\exp(-s'\theta)/[\Gamma(\alpha)]^{n'} \tag{3}$$

where $\alpha > 0$, $\theta > 0$, $n' > 0$, $\nu' > 0$, $s' > 0$, and $p' > 0$, such that $n' \sqrt[n']{p'}/s' < 1$. The posterior joint density is proportional to the product of the expressions in (2) and (3), namely

$$\theta^{\nu''\alpha-1}(p'')^{\alpha-1}\exp(-s''\theta)/[\Gamma(\alpha)]^{n''} \quad (4)$$

where $\nu'' = \nu' + n$, $p'' = p'p$, $s'' = s' + s$, and $n'' = n' + n$. Thus the posterior conditional density of $\tilde{\theta}$, given $\tilde{\alpha} = \alpha$, has the gamma form $g(\theta|\nu''\alpha,s'')$, and the marginal posterior density of $\tilde{\alpha}$ is proportional to

$$\Gamma(\nu''\alpha)(r''/n'')^{\nu''\alpha}/[\Gamma(\alpha)]^{n''} \quad (5)$$

where

$$r''/n'' = \sqrt[\nu'']{p''}/s'' = (p')^{1/(\nu'+n)}\,(r/n)^{n/(\nu'+n)}s^{n/(\nu'+n)}/(s'+s).$$

The last expression shows that the marginal distribution of $\tilde{\alpha}$ depends on both r and s (unless $n' = 0$ and $s' = 0$). Thus Bayesian inferences about α are not based on the same statistics as are optimum unbiased and optimum invariant hypothesis tests about α, which are functions of r and n alone (Glaser, 1976*a*, p. 480).

The posterior marginal density of $\tilde{\theta}$ can be found by numerically integrating on α in expression (4) for each desired value of θ (as was done by Damsleth, 1975). But in practice such extensive calculation will seldom be warranted, so a convenient approximation must be sought. I will suggest an approximation in Section 3.

The Jeffreys (1961) invariant prior "density," which is proportional to the square root of the determinant of the information matrix of the population density (1), is proportional to $[\alpha\psi'(\alpha) - 1]^{1/2}/\theta$, where $\psi(\alpha)$ is the digamma function $\psi(\alpha) = d\ \ln\Gamma(\alpha)/d\alpha$. This "density" implies the a priori "independence" of $\tilde{\alpha}$ and $\tilde{\theta}$ (in the sense that it factors). Jeffreys suggested that when the parameters are assumed to be "independent" and when they are restricted to $(0,\infty)$, the logarithms of the parameters should be assumed to have uniform "densities". Thus for gamma parameters the prior "density" should be proportional to $1/\alpha\theta$. The two Jeffreys priors imply different marginal "densities" for $\tilde{\alpha}$. The disagreement is substantial for small values of α and persists for large α since $[\alpha\psi'(\alpha) - 1]^{1/2} = 0(1/\sqrt{\alpha})$ as $\alpha \to \infty$ (Abramowitz and Stegun, 1964, p. 260). Presumably, the "density" $1/\alpha\theta$ would be preferred in such circumstances.

An heuristic approach to a "non-informative" prior is to ask what form of the density in (3) will have minimal impact on the posterior density. If we choose $\nu' = n' = 0$, $s' = 0$, and $p' = 1$, the posterior parameters are functions of the data only, so in this sense the impact of the prior is negligible. With the above choice of prior parameters, the density in (3) becomes the improper "density" $f'(\alpha,\theta) \propto 1/\theta$.

3. NUMERICAL ANALYSIS OF POSTERIOR DISTRIBUTIONS

The normalizing constant and moments of the marginal density of $\tilde{\alpha}$, whose kernel is in expression (5), can be found easily by numerical integration. I constructed simple computer programs on both IBM 360 and UNIVAC 1110 machines using canned Gaussian integration and gamma function subroutines. (A brief discussion of the numerical techniques involved appears in the Appendix). The calculations are certainly no worse than those required by the classical analysis described by Glaser (1976*a*). Once the moments have been computed, trivial manipulations yield the cumulants of $\tilde{\alpha}$.

The method of deriving the marginal density of $\tilde{\theta}$ suggested in Section 2 is rather expensive and results in a tabular representation of the density. I developed an approximation technique that requires nothing more than the first four cumulants of $\tilde{\alpha}$ and yields an approximating curve. The idea is to compute the first four moments of $\tilde{\theta}$ and then fit a member of the Pearson family of curves by the method of moments. Methods for accomplishing the fit are discussed by Elderton and Johnson (1969) and Pearson and Hartley (1970).

Since the conditional distribution of θ, given $\tilde{\alpha} = \alpha$, is gamma, it follows that the moments of the marginal distribution of $\tilde{\theta}$ are

$$E''(\tilde{\theta}^k) = E''E''(\tilde{\theta}^k|\tilde{\alpha}) = E''\prod_{j=1}^{k}(\nu''\tilde{\alpha} + j - 1)/(s'')^k, \quad (6)$$

and these moments are easily-found functions of the moments of $\tilde{\alpha}$. I found the formulas relating the mean, variance, skewness, and kurtosis of $\tilde{\theta}$ to the cumulants of $\tilde{\alpha}$ a bit easier to work with, but this is a matter of taste. The formulas I used are

$$E''(\tilde{\theta}) = (\nu''/s'')\kappa_1(\alpha) \quad (7)$$

$$\mathrm{Var}''(\tilde{\theta}) = (\nu''/s'')^2[\nu''\kappa_2(\alpha) + \kappa_1(\alpha)]/\nu'' \quad (8)$$

$$\sqrt{\beta_1(\theta)} = \frac{[(\nu'')^2\kappa_3(\alpha) + 3\nu''\kappa_2(\alpha) + 2\kappa_1(\alpha)]}{\sqrt{\nu''}\,[\nu''\kappa_2(\alpha) + \kappa_1(\alpha)]^{3/2}} \quad (9)$$

$$\beta_2(\theta) - 3 = [(\nu'')^3\kappa_4(\alpha) + 6(\nu'')^2\kappa_3(\alpha) + 11\nu''\kappa_2(\alpha) + 6\kappa_1(\alpha)]/\{\nu''[\nu''\kappa_2(\alpha) + \kappa_1(\alpha)]^2\}, \quad (10)$$

where $\kappa_j(\alpha)$, $j = 1,2,3,4$, are cumulants of $\tilde{\alpha}$, and $\beta_1(\theta)$ and $\beta_2(\theta)$ are the usual measures of skewness and kurtosis of $\tilde{\theta}$. It is easy to see that when either ν'' or $\kappa_2(\alpha)$ is large, then $\sqrt{\beta_1(\theta)} \approx \sqrt{\beta_1(\alpha)}$ and $\beta_2(\theta) \approx \beta_2(\alpha)$.

The same sort of approximation procedure could in principle be used to find the moments of the population mean $\tilde{\mu} = \tilde{\alpha}/\tilde{\theta}$. This would involve calculating expectations of ratios of polynomials in $\tilde{\alpha}$ by numerical integration. However such calculations would rarely be worth the effort since a simple approximation is at hand. The conditional mean, variance, skewness, and kurtosis of $\tilde{\mu}$, given $\tilde{\alpha} = \alpha$, are

$$E''(\tilde{\mu}|\alpha) = s''\alpha/(\nu''\alpha - 1)$$
$$= (s''/\nu'')(\nu''\alpha)(\nu''\alpha - 1)^{-1} \approx s''/\nu'', \quad (11)$$
$$\text{Var}''(\tilde{\mu}|\alpha) = (s''\alpha)^2/(\nu''\alpha - 1)^2(\nu''\alpha - 2)$$
$$= (\nu''\alpha - 2)^{-1}[E''(\tilde{\mu}|\alpha)]^2, \quad (12)$$
$$\sqrt{\beta_1(\mu)} = 4\sqrt{\nu''\alpha - 2}/(\nu''\alpha - 3)(\to 0 \text{ as } \nu'' \to \infty), \quad (13)$$

and

$$\beta_2(\mu) = 3\,\frac{(\nu''\alpha - 2)(\nu''\alpha + 5)}{(\nu''\alpha - 3)(\nu''\alpha - 4)}\,(\to 3 \text{ as } \nu'' \to \infty). \quad (14)$$

The mode of the conditional density is $s''\alpha/(\nu''\alpha + 1) \approx s''/\nu'' = E''(\tilde{\mu}|\alpha)$. Thus unless α is small its influence on these quantities will be slight, and it appears that the marginal density of $\tilde{\mu}$ can usually be approximated by fitting a normal curve whose mean and variance are the conditional mean and variance of $\tilde{\mu}$, given an appropriate value of α, say the mode of $\tilde{\alpha}$. Miller and Hickman (1975) give some numerical illustrations of the efficacy of this approximation.

4. NUMERICAL ILLUSTRATION

Gross and Clark (1975, p. 104) report $n = 20$ randomly selected survival times (in weeks) of male mice exposed to 240 rads of gamma radiation. Gross and Clark obtained maximum likelihood estimates $\hat{\alpha} = 8.53$ (st. err. = 2.72) and $\hat{\theta} = 0.075$ (st. err. = 0.024). Their approximate 98% confidence intervals for α and θ are [2.2, 14.86] and [0.018, 0.132] for θ. The approximate method of Bain and Engelhardt (1975) yields [3.46, 15.76] as a 98% confidence interval for α. The estimated correlation coefficient between $\hat{\alpha}$ and $\hat{\theta}$ is 0.98.

Table 1 shows some of the Bayesian posterior analysis of the marginal distributions of $\tilde{\alpha}$ and $\tilde{\theta}$ assuming the following three priors:

(1) improper $\propto 1/\alpha\theta$
(2) improper $\propto 1/\theta$
(3) conjugate $\nu' = n' = 3$, $s' = 300$, $\ln p' = 13.50$.

The conjugate prior might have arisen as follows. Interpret the prior information as having come from a hypothetical experiment involving a sample of size $n' = 3$, a mean $s'/n' = 100$, and a ratio of geometric to arithmetic means $r' = n' \sqrt[n']{p'}/s' = .9$. These specifications yield the parameters of prior (3). This prior implies prior mean and standard deviation of 6.49 4.41 for $\tilde{\alpha}$ and 0.065 and 0.047 for $\tilde{\theta}$. In practice, if an interactive computer program were available, one could study the impact of variations in the prior parameters on tables like Table 1 (with double primes replaced by single primes, of course) in order to find a specification consistent with prior knowledge. All the suggestions, caveats, and limitations of Raiffa and Schlaifer (1961, pp. 58–69) apply to the above discussion.

For all three priors in Table 1 the point $(\sqrt{\beta_1(\cdot)}, \beta_2(\cdot))$ appears to fall near the Pearson Type III (gamma) line in Table 43 of Pearson and Hartley (1970), and the fact that $2\beta_2(\cdot) - 3\beta_1(\cdot) - 6 \approx 0$ signals a Pearson Type III according to Elderton and Johnson (1969). (In another context Lindley (1969) tried a gamma approximation to the posterior distribution for $\tilde{\alpha}$ that is slightly different from mine.) Table 42 of Pearson and Hartley (1970) or the more complete table of Johnson, Nixon and Amos (1963) can be used to obtain selected percentage points of the posterior distributions. These tables were used to calculate the 98% credible intervals in Table 1. They are not highest density intervals since they place equal areas in each tail of the posterior distribution.

Notice that the Jeffreys prior (prior (1)) leads to 98% credible intervals for α and θ that are practically the same as the intervals obtained in Bain and Engelhardt (1975) and (1977), and these intervals differ considerably from those obtained by Gross and Clark (1975) using asymptotic likelihood theory. That the Jeffreys Bayesian and efficient classical inferences agree is to be expected.

A feature of Bayesian analysis is its ability to accommodate a variety of expressions of prior belief. (Whether this be boon or bane is a matter of opinion.) Prior (2) is an improper alternative to Jeffreys' prior, and prior (3) is a relatively mild proper prior. Of course the credible intervals based on the proper prior are shorter than those based on the improper priors.

It is interesting to note that in all cases in Table 1 the skewness and kurtosis measures of $\tilde{\alpha}$ and $\tilde{\theta}$ are equal to two decimal places, signaling that the conditions stated below equation (14) are satisfied.

Table 2 displays the sensitivity to α of my suggested normal approximation to the posterior distribution of $\tilde{\mu}$ for priors (1) and (3). The MLE's of μ given by Gross and Clark and by Engelhardt and Bain are 113.45 and 116.53, respectively.

5. CONCLUSIONS

If an investigator feels that his prior opinion will have little effect on the information in the data, then he may as well use the Jeffreys prior, which is essentially equivalent to accepting the classical inference. Once the relevant computer program has been written, the Bayesian procedure can then be looked at as a convenient way to get efficient classical inferences about α, θ, and μ.

The Bayesian framework allows us to study the sensitivity of our inferences to variations in the specification of prior parameters. If an investigator is willing to express prior knowledge in terms of the con-

TABLE 1—*Bayesian posterior analysis of $\tilde{\alpha}$ and $\tilde{\theta}$ based on Gross and Clark data.*

	Analysis Based on		
	Prior (1)	Prior (2)	Prior (3)
$E''(\tilde{\alpha})$	8.39	9.24	8.19
$\sqrt{V''(\tilde{\alpha})}$	2.66	2.80	2.32
$\sqrt{\beta_1(\alpha)}$	0.65	0.62	0.58
$\beta_2(\alpha)$	3.63	3.57	3.50
$2\beta_2(\alpha)-3\beta_1(\alpha)-6$	0.92×10^{-3}	0.56×10^{-3}	-0.92×10^{-2}
98% credible interval for $\tilde{\alpha}$	(3.50,15.81)	(4.03,17.00)	(3.80,14.55)
$E''(\tilde{\theta})$	0.074	0.081	0.073
$\sqrt{V''(\tilde{\theta})}$	0.024	0.025	0.021
$\sqrt{\beta_1(\theta)}$	0.65	0.62	0.58
$\beta_2(\theta)$	3.63	3.57	3.50
$2\beta_2(\theta)-3\beta_1(\theta)-6$	0.63×10^{-3}	0.12×10^{-3}	-0.92×10^{-2}
98% credible interval for $\tilde{\theta}$	(0.030.0.141)	(0.035,0.150)	(0.033,0.131)

TABLE 2—*Sensitivity of a normal approximation to the posterior distribution of $\tilde{\mu}$ (Gross and Clark's data).*

Jeffreys' Prior

α	$E''(\tilde{\mu}\|\alpha)$	Mode	$V''(\tilde{\mu}\|\alpha)$	$\pm 2.326\sqrt{V''(\tilde{\mu}\|\alpha)}$
4	114.89	112.05	169.22	± 30.26
8	114.16	112.75	82.49	± 21.13
12	113.92	112.98	54.53	± 17.18

Conjugate Prior

α	$E''(\tilde{\mu}\|\alpha)$	Mode	$V'(\tilde{\mu}\|\alpha)$	$\pm 2.326\sqrt{V''(\tilde{\mu}\|\alpha)}$
4	112.92	110.49	141.68	± 27.69
8	112.31	111.09	69.30	± 19.36
12	112.10	111.29	45.86	± 15.75

jugate family of distributions, and to use the approximations I have suggested, then the amount of numerical integration is quite small. Thus even an extensive sensitivity analysis can be both quick and inexpensive, given the appropriate computer software.

6. APPENDIX

The main numerical problem in my approach is the integration of the kernel in (5). Let us denote it by $h(\alpha)$. To get the normalizing constant and the first four moments of the marginal posterior distribution of $\tilde{\alpha}$ we must evaluate $\int_0^\infty \alpha^i h(\alpha) d\alpha$ for $i = 0,1,2,3,4$. The integrands are "well-behaved" so that straightforward Laguerre-Gauss quadrature can be used (see Hildebrand, 1956, p. 325, and Abramowitz and Stegun, 1964, p. 890). I used a 48-point formula. To use the numerical integration subroutine, one must supply another subroutine to calculate the integrand. I found it convenient first to compute

$$\ln[\alpha^i h(\alpha)] = i \ln\alpha + \ln\Gamma(\nu''\alpha) + \nu''\alpha \ln(r''/n'') - n''\ln\Gamma(\alpha),$$

and then exponentiate. This calculation involved calling a canned subroutine that returned values of $\ln\Gamma(x)$. The routine used double precision arithmetic. For $x \geq 8$, it used the asymptotic expansion

$$\ln\Gamma(x) \sim (x - 1/2)\ln x - x + (1/2)\ln(2\pi) + 1/12x - 1/360x^3 + 1/1260x^5 - 1/1680x^7 + 1/1188x^9. \qquad \text{(A1)}$$

(See Abramowitz and Stegun, 1964, p. 257.) For values $x < 8$, the recursion $\Gamma(x + 1) = x\Gamma(x)$ was applied repeatedly until (A1) could be used.

7. ACKNOWLEDGEMENTS

This research was partially supported by grants from the Graduate School of Business, University of Wisconsin, Madison, and the Office of Water Research and Technology. The work was completed while I was visiting the Department of Statistics, University of Iowa, during the 1977–78 academic year. I used the computer facilities of both the University of Wisconsin and the University of Iowa. I am grateful for the comments of the Associate Editor and two referees.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* U.S. National Bureau of Standards Applied Mathematics Series, 55.

AMMETER, H. (1970). The solvency problem in risk life insurance. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmatematiker,* 70(2), 250–258.

BAIN, L. and ENGELHARDT, M. (1975). A two-moment chi-square approximation for the statistic log $(\bar{x}/\tilde{x})$. *J. Amer. Statist. Assoc. 70,* 948–950.

CHOI, S. and WETTE, R. (1969). Maximum likelihood estimation of the parameters of the gamma distribution and their bias. *Technometrics, 11,* 683–690.

COX, C. and HINKLEY, D. (1974). *Theoretical Statistics.* London: Chapman and Hall.

DAMSLETH, E. (1975). Conjugate classes for gamma distributions. *Scandinavian Journal of Statistics, 2,* 80–84.

ELDERTON, W. and JOHNSON, N. (1969). *Systems of Frequency Curves.* Cambridge: Cambridge University Press.

ENGELHARDT, M. and BAIN, L. (1977). Uniformly most powerful unbiased tests on the scale parameter of a gamma distribution with a nuisance shape parameter. *Technometrics, 19,* 77–81.

GLASER, R. (1976a). The ratio of the geometric mean to the arithmetic mean for a random sample from a gamma distribution. *J. Amer. Statist. Assoc., 71,* 480–487.

GLASER, R. (1976b). Exact critical values for Bartlett's test for homogeneity of variances. *J. Amer. Statist. Assoc., 71,* 488–490.

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

HILDEBRAND, F. (1956). *Introduction to Numerical Analysis.* New York: McGraw-Hill Book Co.

JEFFREYS, H. (1961). *Theory of Probability, 3rd. ed.* Oxford: Clarendon Press.

JOHNSON, N. NIXON, E., and AMOS, D. (1963). Table of percentage points of Pearson curves, for given $\sqrt{\beta_1}$ and β_2, expressed in standard measure. *Biometrika, 50,* 459–498.

JOHNSON, N. and KOTZ, S. (1970). *Continuous Univariate Distributions.* Boston: Houghton Mifflin Company.

LINDLEY, D. (1969). A Bayesian solution for some education prediction problems: II. *Research Bulletin,* 69–91. Princeton: Educational Testing Service.

MILLER, R. and HICKMAN, J. (1975). Insurance credibility theory and Bayesian estimation. In *Credibility: Theory and Applications,* P. M. Kahn, pp. 249–270. New York: Academic Press.

PEARSON, E. and HARTLEY, H. (1970). *Biometrika Tables for Statisticians, 3rd. ed.* Cambridge: Cambridge University Press.

RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory.* Boston: Division of Research, Harvard Business School.

SEAL, H. (1969). *Stochastic Theory of a Risk Business.* New York: John Wiley and Sons.