Maximum Likelihood Estimation of Parameters in the Inverse Gaussian Distribution, With Unknown Origin

R. C. H. Cheng and N. A. K. Amin

Institute of Science and Technology
University of Wales
Cardiff CF1 3NU Wales

Maximum likelihood estimation is applied to the three-parameter Inverse Gaussian distribution, which includes an unknown shifted origin parameter. It is well known that for similar distributions in which the origin is unknown, such as the lognormal, gamma, and Weibull distributions, maximum likelihood estimation can break down. In these latter cases, the likelihood function is unbounded and this leads to inconsistent estimators or estimators not asymptotically normal. It is shown that in the case of the Inverse Gaussian distribution this difficulty does not arise. The likelihood remains bounded and maximum likelihood estimation yields a consistent estimator with the usual asymptotic normality properties. A simple iterative method is suggested for the estimation procedure. Numerical examples are given in which the estimates in the Inverse Gaussian model are compared with those of the lognormal and Weibull distributions.

KEY WORDS: Maximum likelihood estimation; Inverse Gaussian distribution; Lognormal distribution; Weibull distribution.

1. INTRODUCTION

Suppose a parametric model is to be fitted to a random sample of observations $x_1, x_2, \ldots, x_n$ that is positively skewed, where moreover the x's are bounded below by an unknown bound. For convenience, we shall assume that x's are ordered with x_1 smallest. If the physical situation does not suggest an obvious distribution, one can be chosen on the grounds of goodness of fit and convenience. There are a number of three-parameter generalizations of well-known two-parameter distributions that can be used, the best known probably being the lognormal, gamma, and Weibull distributions. These three are similar in that the two-parameter versions have positive density only if $x > 0$, while the three-parameter versions allow for a shifted origin so that the density is positive only for $x > \alpha$. In their three-parameter versions, the lognormal model was investigated by Cohen (1951), Hill (1963), and Harter and Moore (1966), among others; the gamma model by Harter and Moore (1965) and Cohen and Norgaard (1977); the Weibull model by Harter and Moore (1965), Dubey (1966), and Rockette, Antle, and Klimko (1974).

It is well known that maximum likelihood estimation in these three-parameter versions does not satisfy the usual regularity conditions and conse-

quently can give rise to estimators not having the usual asymptotic normality properties, and even to inconsistent estimators for these three models. The critical difficulty when using the three-parameter lognormal, gamma, and Weibull models is that in all these cases there are paths in the parameter space, with α (the shifted parameter) tending to the smallest observation x_1, along which the likelihood becomes infinite. This yields $\hat{\alpha} = x_1$ as the ML estimator of α. Unfortunately, this does not then yield sensible estimates of the remaining parameters. In certain cases, such as the lognormal model, the problem can be circumvented by searching for a local maximum of the likelihood which yields consistent estimators for all parameters. Unfortunately, for a given sample there may be no local maximum. Thus the technique is not foolproof and it has to be augmented by additional ad hoc methods (see Harter and Moore 1966).

To avoid these difficulties, we suggest an alternative distribution, the three-parameter Inverse Gaussian (IG) distribution, with density function

$$f(x; \alpha, \lambda, \mu)$$

$$= [\lambda/\{2\pi(x - \alpha)^3\}]^{1/2}$$

$$\exp[-\lambda(x - \alpha - \mu)^2/\{2\mu^2(x - \alpha)\}]$$

$$x > \alpha; \mu, \lambda > 0. \tag{1.1}$$

The well-known two-parameter model, of which (1.1) is a generalization, was reviewed recently by Folks and Chhikara (1978) and no consideration of the three-parameter case appears to have been made. Our main aim is to show that ML estimation for this distribution is not only numerically straightforward, but more importantly cannot give rise to inconsistent estimators. It is suggested that if the only criteria for using a particular model are those of convenience and goodness of fit, then the IG distribution is an especially appropriate model.

In the next section we describe the properties of the likelihood function of the IG distribution and suggest a method for obtaining ML estimates of the parameters. In Section 3, some properties of the estimators are discussed. Numerical examples are given in the final section and the ML estimates in the IG model are compared with those of the lognormal and Weibull models.

2. ML ESTIMATION IN THE INVERSE GAUSSIAN MODEL

We first consider the likelihood function of the IG distribution. The log-likelihood corresponding to the density (1.1) is

$$L(\alpha, \lambda, \mu) = (n/2)\log\{\lambda/(2\pi)\} - (3/2)\sum_{i=1}^{n} \log(x_i - \alpha)$$

$$- \lambda\sum_{i=1}^{n} (x_i - \alpha - \mu)^2/\{2\mu^2(x_i - \alpha)\}. \quad (2.1)$$

For fixed α, the maximum of L can be written as

$$L^*(\alpha) = \max_{\mu, \lambda} L(\alpha, \mu, \lambda) = L(\alpha, \hat{\mu}(\alpha), \hat{\lambda}(\alpha)),$$

where

$$\hat{\mu}(\alpha) = \bar{x} - \alpha, \quad \hat{\lambda}(\alpha) = \left\{n^{-1}\sum_{i=1}^{n} (x_i - \alpha)^{-1} - \hat{\mu}^{-1}(\alpha)\right\}^{-1}.$$

$$(2.2)$$

The overall ML estimators $\hat{\alpha}$, $\hat{\mu}$, and $\hat{\lambda}$ can now be found by maximizing $L^*(\alpha)$ with respect to α.

Unlike the lognormal, gamma, and Weibull cases, $L^*(\alpha)$ has a finite maximum. Straightforward series expansion of $L^*(\alpha)$ in terms of $\delta = x_1 - \alpha$ gives

$$L^*(\alpha) = \tfrac{1}{2}(n - 3)\log \delta + 0(1) \quad \text{as} \quad \delta \to 0, \quad (2.3)$$

so that $L^*(\alpha) \to -\infty$ as $\alpha \to x_1$, provided $n > 3$, which we assume from now on. Similarly,

$$L^*(\alpha) = -\tfrac{1}{2}n\{1 + \log(2\pi s^2)\} - ng_1 s(2\alpha)^{-1}$$

$$+ 0(\alpha^{-2}) \quad \text{as} \quad \alpha \to -\infty, \quad (2.4)$$

where s^2 and g_1 are respectively the sample variance and skewness given by $s^2 = \sum x_i^2/n - \bar{x}^2$, $g_1 = s^{-3}$ $(\sum x_i^3/n - 3\sum x_i^2 \sum x_i/n^2 + 2\bar{x}^3)$, so that $L^*(\alpha)$ tends to a constant as $\alpha \to -\infty$.

It follows from (2.3) and (2.4) that $L^*(\alpha)$ is bounded above for all $-\infty < \alpha < x_1$. Equation (2.4) also shows that, when the sample skewness is positive, $L^*(\alpha)$ ultimately decreases as α tends to $-\infty$. Consequently, $L^*(\alpha)$ must have a global maximum at a stationary point where $\partial L^*(\alpha)/\partial\alpha = 0$.

When the sample skewness is negative, $L^*(\alpha)$ can achieve its overall maximum at $\alpha = -\infty$. This might appear unusual, but is actually easily explained. The point is that the IG distribution encompasses the normal distribution as a special case. Letting $\alpha \to -\infty$ is equivalent to fitting the normal distribution, with the ML estimates of the mean $\hat{\theta}(\alpha) = \alpha + \hat{\mu}(\alpha)$ and the variance $\hat{\sigma}^2(\alpha) = \hat{\mu}^3(\alpha)/\hat{\lambda}(\alpha)$ becoming the ML estimates of the mean and variance of a normal sample. We have

$$\hat{\theta}(\alpha) = \bar{x} \text{ (independent of } \alpha), \quad \lim_{\alpha \to -\infty} \hat{\sigma}^2(\alpha)$$

$$= \sum x_i^2/n - \bar{x}^2. \quad (2.5)$$

Thus, when the sample is negatively skew, ML estimation is equivalent to fitting the normal model, this being obtained by letting $\alpha \to -\infty$.

With the above points in mind, the following procedure is suggested for deciding when to fit the IG distribution and when to fit the normal one. First, calculate the sample skewness g_1 as previously defined. Then, fit the three-parameter IG model if $g_1 > k(6/n)^{1/2}$, with k a positive constant to be chosen; otherwise, fit the normal model. (Recall that the sample is assumed to be positively skewed. The case in which the normal model is not appropriate because the sample is too negatively skewed does not therefore occur.) The reason for using this test is that it is known (see for example Kendall and Stuart 1969) that g_1 is approximately $N(0, 6n^{-1})$ if the underlying distribution is nearly normal. The constant k can therefore be chosen to make small the probability of fitting the IG model when the true distribution is actually nearly normal. For example, with $k = 1.64$ this probability is approximately 5 percent. It should be stressed that it is not desirable to make this probability too small, as this would give an over-conservative procedure with the three-parameter IG model being fitted only when the distribution is very skew.

When the IG model is to be fitted, α has to be found from the implicit equation $\partial L^*(\alpha)/\partial\alpha = 0$. The suggested method for doing this is to use the iteration

$$\alpha_0 = x_1 - (2s^2 \log n)^{-1}(\bar{x} - x_1)^3,$$

$$\alpha_{m+1} = \alpha_m + \left[3\sum_i (x_i - \alpha_m)^{-1}/n\right.$$

$$\left. + \lambda_m\left\{\mu_m^{-2} - \sum_i (x_i - \alpha_m)^{-2}/n\right\}\right]$$

$$\times \{3\mu_m^{-1}\lambda_m^{-1} + 12\lambda_m^{-2}\}(m = 0, 1, 2, \ldots), \quad (2.6)$$

where μ_m, λ_m are the estimates (2.2) evaluated at $\alpha = \alpha_m$. The recursion is based on the Newton-Raphson iteration

$$\alpha_{m+1} = \alpha_m - \{\partial L^*/\partial\alpha)/(\partial^2 L^*/\partial\alpha^2)\}_{\alpha_m} \quad (2.7)$$

It replaces $\partial^2 L^*/\partial\alpha^2$, however, by an estimate of its expected value, this being approximately

$$E\{n^{-1}(\partial^2 L^*/\partial\alpha^2)_{\alpha_{true}}\} = -1.5\lambda^{-1}\mu^{-1} - 6\lambda^{-2}$$

for large n.

Apart from simplifying calculations, the approximation to $\partial^2 L^*/\partial\alpha^2$ has the important advantage of being always negative. This ensures that the modified iterations converge only to a maximum and never to a minimum of L^*. The suggested starting value α_0 has been obtained by using x_1 adjusted by a bias correction term. Finally, if the initial choice of α_0 happens to be poor, it is possible to get at some stage, $\alpha_{m+1} > x_1$, which is clearly invalid. If this occurs, then it is suggested that iterations are continued with $\alpha_{m+1} = (x_1 + \alpha_m)/2$.

3. DISTRIBUTIONAL PROPERTIES OF ESTIMATORS

We consider both large and small sample properties of the ML estimators in the IG distribution. We have used the parametrization (α, μ, λ) as being a natural generalization of the two-parameter (μ, λ) model. This is not always the most informative set of parameters and we shall therefore consider also the parameters $\theta = \alpha + \mu$, $\sigma^2 = \mu^3/\lambda$, and $\phi = \lambda/\mu$, these being respectively the mean, variance, and shape parameters of the distribution.

For large samples, the usual normality and efficiency properties still hold, even though the model does not satisfy the usual regularity conditions.

Theorem. There is in probability as $n \to \infty$, a stationary point $(\hat{\alpha}, \hat{\mu}, \hat{\lambda})$ of L such that $\sqrt{n}(\hat{\alpha} - \alpha, \hat{\mu} - \mu, \hat{\lambda} - \lambda)$ is asymptotically normally distributed with mean zero and covariance matrix $\sum$ with elements $d^{-1}\sigma_{ij}$, where

$$\sigma_{11} = \tfrac{1}{2}\lambda^{-1}\mu^{-3},$$

$$\sigma_{12} = -\tfrac{1}{2}\lambda^{-1}\mu^{-3}$$

$$\sigma_{13} = -\tfrac{3}{2}(\mu^{-4} + \lambda^{-1}\mu^{-3}),$$

$$\sigma_{23} = \tfrac{3}{2}(\mu^{-4} + \lambda^{-1}\mu^{-3})$$

$$\sigma_{22} = \tfrac{1}{2}\lambda^{-1}\mu^3 + \tfrac{3}{4}\lambda^{-3}\mu^{-1} + 3\lambda^{-4},$$

$$\sigma_{33} = \tfrac{9}{2}\lambda\mu^{-5} + \tfrac{21}{2}\mu^{-4} + \tfrac{21}{2}\lambda^{-1}\mu^{-3}$$

and $d = \tfrac{3}{4}\lambda^{-2}\mu^{-4} + 3\lambda^{-3}\mu^{-3}$.

The proof is given in the Appendix. The asymptotic distribution of $\hat{\theta}$, $\hat{\sigma}$, $\hat{\phi}$ can be calculated directly from this result using well-known "standard errors" formu-

las. For example, this yields the asymptotic distribution of $\hat{\theta}$ as $\sqrt{n}[\hat{\theta} - (\alpha + \mu)] \sim N(0, \mu^3\lambda^{-1})$.

The extent to which the asymptotic results hold for small samples depends very much on the shape of the parent distribution. The discussion is simplified because the distributional properties do not depend on α, and because μ is only a scale parameter. Thus, with no loss of generality we can assume that $\alpha = 0$, and that $\mu = 1$. The shape of the distribution then depends only on ϕ, being progressively more skew as $\phi \to 0$, and tending to normality as $\phi \to \infty$. Table 1 gives the values of the biases to order n^{-1} for selected values of ϕ. The biases of $\hat{\alpha}$, $\hat{\mu}$, and $\hat{\lambda}$ have been computed from the formula given by Shenton and Bowman (1977, Eq. (3.12b)); the biases of the other estimates have been calculated from these using

$$\text{bias}\{f(\hat{\xi})\} = \sum_i \frac{\partial f}{\partial \xi_i} \text{bias}(\hat{\xi}_i)$$

$$+ \frac{1}{2}\sum_i \sum_j \frac{\partial^2 f}{\partial \xi_i \, \partial \xi_j} \text{cov}(\hat{\xi}_i, \hat{\xi}_j),$$

where the derivatives are evaluated at $E(\hat{\xi})$.

The most notable feature is that the estimators $\hat{\lambda}$, $\hat{\phi}$, and to a lesser extent $\hat{\mu}$ and $\hat{\alpha}$ all show increasingly substantial bias as the underlying distribution tends to normality (i.e., as ϕ increases). In contrast, the estimates of mean and variance, $\hat{\theta}$ and $\hat{\sigma}^2 = \hat{\mu}^3/\hat{\lambda}$ are relatively unbiased.

This feature shows there is little point in trying to fit the three-parameter model when the data look normal irrespective of the true shape of the population distribution. In essence, the only reliable information is being supplied by the sample mean and variance. All the other estimators, $\hat{\mu}$, $\hat{\lambda}$, $\hat{\phi}$, and $\hat{\alpha}$, precisely those connected directly with the IG model, are heavily biased.

There is another feature of note in Table 1, which occurs in the converse situation where the parent distribution is very skew. Here the estimates $\hat{\alpha}$, $\hat{\mu}$, $\hat{\lambda}$, $\hat{\phi}$, and $\hat{\theta}$ remain reasonably unbiased. However, $\hat{\sigma}^2$ now becomes increasingly positively biased as skewness increases. We do not pursue in detail how this bias

Table 1. *First Order Biases (with $\mu = 1$)*

ϕ	$n$bias($\hat{\alpha}$)	$n$bias($\hat{\lambda}$)	$n$bias($\hat{\phi}$)	$n$bias($\hat{\sigma}^2$)
0.1	.1433	-.0652	.9440	338.4
0.5	.6008	-.4815	.6708	18.54
2.0	1.296	1.000	1.482	1.722
10.0	-1.667	259.3	184.1	.0372
100.0	-60.49	38250.	19190.	-.008411

Notes : (i) The bias of $\hat{\mu}$ is bias ($\hat{\mu}$) = - bias ($\hat{\alpha}$).
(ii) The bias of $\hat{\theta}$ is zero.

might be removed, but note that $v = ns^2/(n - 1)$ is unbiased for μ^3/λ. Now, as ϕ increases, so $\hat{\mu}^3/\hat{\lambda}$ becomes approximately equal to s^2 in probability. This implies that v is nearly fully efficient when ϕ is large. Interestingly, v retains a fair proportion of its efficiency when ϕ is small. To order n^{-1},

$$\text{var}(v) = n^{-1}\mu^4(2\phi^{-2} + 15\phi^{-3}),$$

which tends to $15n^{-1}\mu^4\phi^{-3}$ as $\phi \to 0$. To the same order,

$$\text{var}(\hat{\mu}^3/\hat{\lambda})$$
$$= n^{-1}\mu^4(2\phi^{-2} + 23\phi^{-3} + 36\phi^{-4})(1 + 4\phi^{-1})^{-1},$$

and this tends to $9n^{-1}\mu^4\phi^{-3}$ as $\phi \to 0$. As $\hat{\mu}^3/\hat{\lambda}$ has a large bias when ϕ is small, this suggests that v is to be preferred when either n is small or ϕ is large.

4. NUMERICAL EXAMPLES

To illustrate this discussion we fit the IG model to three sets of data, selected to illustrate different types of behavior of the ML estimates. All three sets of data are sufficiently skewed to warrant, under the suggested criterion, fitting the IG rather than the normal model. Also, in each case there are grounds for believing a shifted origin as being possibly more appropriate than an origin with $\alpha = 0$, so that it is reasonable to try to fit the three-parameter IG model.

The fitting of the IG model of (1.1) will be compared with the three-parameter lognormal model with density

$$f(x) = (2\pi)^{-1/2}\sigma^{-1}(x - \alpha)^{-1}$$
$$\times \exp\{-(2\sigma^2)^{-1}[\log(x - \alpha) - \beta]^2\}$$

and the three-parameter Weibull model with density

$$f(x) = \gamma\delta^{-\gamma}(x - \alpha)^{\gamma-1}\exp\{-[(x - \alpha)/\delta]^\gamma\}.$$

In these two models *local* maximum likelihood estimation has been used to obtain estimates.

This method has been used to illustrate the advantage of the IG model over the lognormal and Weibull models in always allowing ML estimates to be obtained. Though we have not done so here, when ML estimation breaks down, alternative methods (such as the modifications suggested by Cohen 1975 and 1976 in which an alternative equation is substituted for $\partial L/\partial \alpha = 0$) might be used that enable the lognormal or Weibull model to be successfully fitted. The problem, of course, is that it is not clear at the outset when such an alternative procedure is required. It should be said, however, that though ML estimates can always be obtained with the IG model, this does not mean they will always be completely satisfactory. As discussed at the end of Section 3, there is the possibility in small samples of bias adversely affecting estimates.

There does not seem to be a precise goodness-of-fit test suitable for these three-parameter distributions when the sample size is small. For example, the Kolmogorov-Smirnov test cannot be applied where all three parameters have to be estimated from the sample. The chi squared goodness-of-fit test might be used, but the smallness of the sample size renders its application impractical. We therefore give only a rough indication of the goodness of fit of each of the models by plotting the empirical cdf and cdf of the fitted distributions.

Additional comparison of the three models can be made using the ML estimates themselves. A direct comparison of parameter estimates is not possible as they are not simply related except for the parameter α, which is the shifted origin in all three models—and μ in the IG model and δ in the Weibull model, which are both scale parameters. They can, however, be used to give ML estimates of the mean, variance, and skewness that give a more direct comparison of the three models. For the IG model, the mean, variance, and skewness are $\alpha + \mu$, μ^3/λ, and $3(\mu/\lambda)^{1/2}$; for the lognormal model they are $\alpha + \exp(\mu + \frac{1}{2}\sigma^2)$, $w(w - 1)e^{2\mu}$, and $(w - 1)^{1/2}(w + 2)$, where $w = e^{\sigma^2}$; and for the Weibull model they are $\alpha + \delta\Gamma(1 + \gamma^{-1})$, $v\delta^2$, and

$$v^{-3/2}\{\Gamma(1 + 3\gamma^{-1}) - 3\Gamma(1 + 2\gamma^{-1})\Gamma(1 + \gamma^{-1})$$
$$+ 2\Gamma^3(1 + \gamma^{-1})\},$$

where $v = \Gamma(1 + 2\gamma^{-1}) - \Gamma^2(1 + \gamma^{-1})$.

Example 1. Dumonceaux and Antle (1973) cite data, obtained in a civil engineering context, of the maximum flood level (in millions of cubic feet per second) for the Susquehanna River at Harrisburg, Pennsylvania over 20 four-year periods as

.654	.613	.315	.449	.297
.402	.379	.423	.379	.3235
.269	.740	.418	.412	.494
.416	.338	.392	.484	.265

This is an example in which all three models can be successfully fitted. The ML estimates for the IG model are $\hat{\alpha} = .178$, $\hat{\mu} = .245$, and $\hat{\lambda} = .914$; for the lognormal they are $\hat{\alpha} = .185$, $\hat{\beta} = -1.56$, and $\hat{\sigma}^2 = .257$, and for the Weibull model they are $\hat{\alpha} = .261$, $\hat{\gamma} = 1.25$, and $\hat{\delta} = .173$. We give some idea of the goodness of fit by comparing the empirical cdf and the cdf's of the fitted models, shown in Figure 1. The three models give a similar fit to the data with the cdf's for the lognormal and IG appearing almost identical. Estimates of mean, variance, and skewness listed in Table 2 compare reasonably well and with the corresponding sample statistics. In the Weibull model, though the fit is satisfactory, ML estimation is actually close to breaking down. The α value giving the local maximum of the
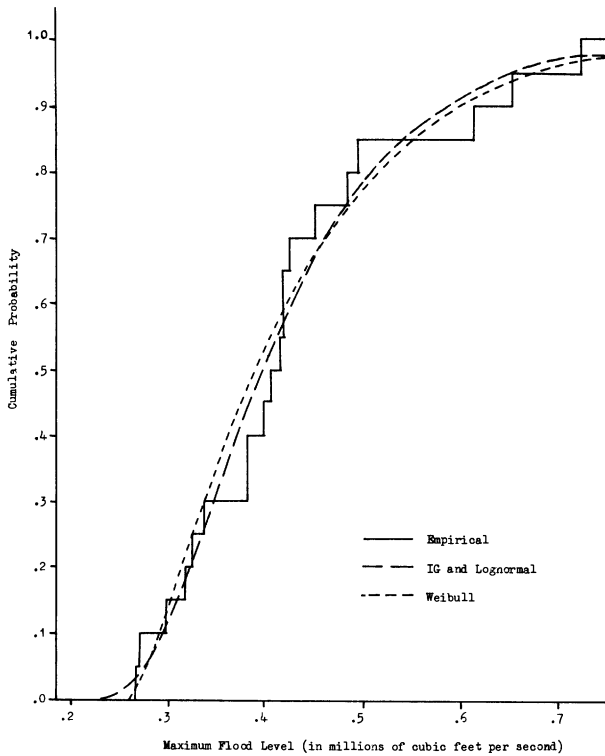
Maximum Flood Level (in millions of cubic feet per second)

Figure 1. *Empirical and Fitted Distributions: Dumonceaux and Antle Data*

Fatigue Lives (in hours)

Figure 2. *Empirical and Fitted Distributions: Engelhardt and Bain Data*

Weibull likelihood is very close to the smallest observation and the data would only need to be slightly more skewed before the local maximum disappears altogether. The next example illustrates this.

Example 2. Engelhardt and Bain (1979) give the fatigue lives (in hours) for 10 bearings tested on a certain tester as follows.

152.7	172.0	172.5	173.3	193.0
204.7	216.5	234.9	262.6	422.6

The ML estimates for the IG model are $\hat{\alpha} = 139.7$, $\hat{\mu} = 80.8$, and $\hat{\lambda} = 88.2$. The lognormal model can also be fitted giving estimates $\hat{\alpha} = 144.1$, $\hat{\beta} = 3.94$, and $\hat{\sigma}^2 = .827$. ML estimation breaks down for the Weibull model as no local maximum of the likelihood exists.

The cdf's of the fitted lognormal and IG models are

again very similar as seen in Figure 2 and they provide a reasonable fit compared with the sample cdf. However, a comparison of the estimates of mean, variance, and skewness in Table 3 suggests better agreement between the sample and IG estimates, the lognormal estimates being somewhat larger that the sample values.

Example 3. Steen and Stickler (1976), in a survey of beach pollution in South Wales, give the pollution (measured in number of coliform per 100 m) on 20 days over a five-week period at Cold Knap Beach as follows.

200	6091	336	327	154	109	111	282	2120	1082
918	718	482	1345	53600	5900	1918	900	1045	1454

The ML estimates in the IG model are $\hat{\alpha} = 44.4$, $\hat{\mu} = 3910$, and $\hat{\lambda} = 318$; in the lognormal model they are $\hat{\alpha} = 108.5$, $\hat{\beta} = 6.08$, and $\hat{\sigma}^2 = 6.36$. ML estimation again breaks down in the Weibull model.

Table 2. *Comparison of ML Estimates in the IG, Lognormal, and Weibull Models: Dumonceaux and Antle Data*

Estimate		Sample Value	Inverse Gaussian	Lognormal	Weibull
α	$(x_1 = .265)$		.178	.185	.261
Mean		.423	.423	.424	.422
Variance		.0149	.0161	.0167	.0169
Skewness		1.07	1.55	1.78	1.43

Table 3. *Comparison of ML Estimates in the IG, Lognormal, and Weibull Models: Engelhardt and Bain Data*

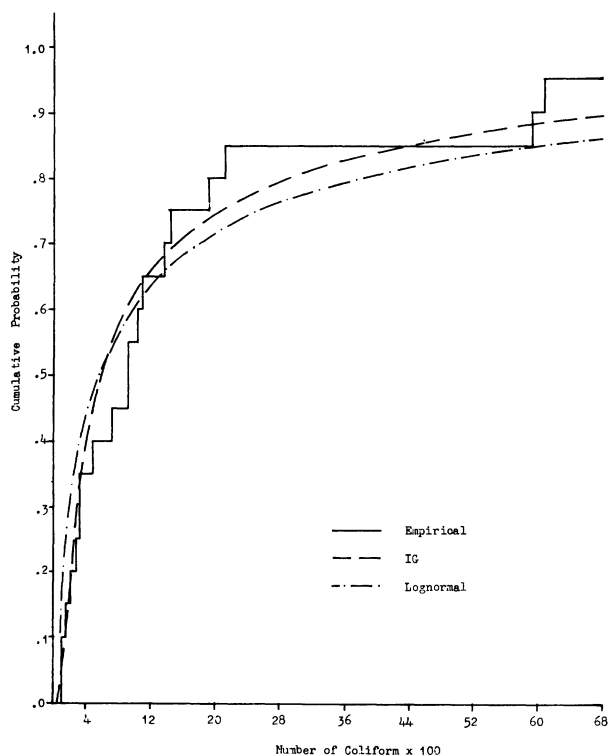| Estimate		Sample Value	Inverse Gaussian	Lognormal	Weibull
α	$(x_1 = 152.6)$		139.6	144.1	ML estimation
Mean		220.5	220.5	221.8	Fails
Variance		5533.	5984.	7751.	
Skewness		1.86	2.87	4.86	

Figure 3. *Empirical and Fitted Distributions: Steen and Stickler Data*

In Figure 3, the plots of the cdf of the fitted lognormal and IG models are seen to give similar fits to the empirical cdf, with the IG model giving a somewhat better fit to the middle percentiles and upper tail of the empirical distribution. Comparison of the estimates of mean, variance, and skewness in Table 4, however, indicates the lognormal estimates $\hat{\beta}$ and $\hat{\sigma}^2$ as being highly suspect. In fact, the data are close to the borderline where ML estimation in the lognormal model cannot be applied; for slightly more skewed data, the lognormal likelihood would have no local maximum at all.

5. CONCLUSION

When the physical situation does not suggest an obvious distribution to be fitted to (positively) skewed data, one can be chosen on the grounds of goodness of fit and convenience. It is suggested that the three-parameter IG distribution may provide an alternative

Table 4. *Comparison of ML Estimates in the IG, Lognormal and Weibull Models: Steen and Stickler Data*

Estimate	Sample Value	Inverse Gaussian	Lognormal	Weibull
α	$(x_1 = 109)$	44.4	108.5	ML estimation
Mean	3955	3955	10659	Fails
Variance	1.32×10^8	1.88×10^8	6.45×10^{10}	
Skewness	3.99	10.53	14028	

skewed model in such situations. Unlike other three-parameter models, such as the lognormal and the Weibull cases, the IG model has a bounded likelihood and ML estimators for the parameters can always be obtained. The estimators possess the usual asymptotic properties of normality and minimum-variance. A method of fitting the distribution has been given in Section 2. Computation of the estimates is relatively simple and requires no greater effort than in the other models.

APPENDIX

Proof of theorem. Only an outline is given, as the proof is similar to standard proofs of asymptotic normality of ML estimates. A more detailed proof has been given by Cheng, Amin, and Feast (1979).

In standard proofs (see for example Kendall and Stuart 1969), a consistent root of the likelihood equation is first exhibited. In the present case, because of the unknown shifted origin α, it is necessary to establish a sharper initial result than just that of consistency.

We write $X = O_p(n^{-q})$ for a variable X of order n^{-q} in probability. We first show that a solution $\hat{\alpha}$ can be found for which $\hat{\alpha} - \alpha_0 = O_p(n^{-1/2})$, where α_0 is the true parameter value. Expanding $\partial L^*/\partial \alpha$, or more simply $h(\alpha) = \hat{\lambda}(\alpha) \, \partial L^*/\partial \alpha$, as a series in α about α_0, we find that

$$n^{-2}h(\alpha) = L_0 + L_1(\alpha_0 - \alpha) + R,$$

where $L_0 = O_p(n^{-1/2})$, $L_1 = O_p(1)$, and $R = O_p\{(\alpha_0 - \alpha)^2/(x_1 - \alpha_0)^2\}$. The expansion is valid provided $x_1 - \alpha_0 > \alpha - \alpha_0$. From the distribution of x_1 we find that $(x_1 - \alpha_0)^{-1} = O_p(\log n)$, so that if we restrict $(\alpha - \alpha_0)$ to being $O(n^{-1/2})$ then the inequality will hold in probability. Under this condition we find that R can be neglected and there is a root of $h(\alpha) = 0$ given approximately by $\hat{\alpha} - \alpha_0 = L_0/L_1$ for which $\hat{\alpha} - \alpha_0 = O_p(n^{-1/2})$.

A similar argument applied to $\hat{\mu} = \hat{\mu}(\hat{\alpha})$, $\hat{\lambda} = \hat{\lambda}(\hat{\alpha})$ as defined in (2.2) now shows $\hat{\mu} - \mu_0 = O_p(n^{-1/2})$ and $\hat{\lambda} - \lambda_0 = O_p(n^{-1/2})$.

The asymptotic distribution of $\hat{\theta} = (\hat{\alpha}, \hat{\mu}, \hat{\lambda})$ can now be established along standard lines. For example, the asymptotic distribution of $\hat{\alpha} - \alpha_0$ is the same as that of L_0/L_1. A straightforward evaluation shows that the asymptotic distribution of $\hat{\theta}$ is normal with mean $\theta^0 (=$ true parameter values) and variance, which can be calculated in the usual way as $E\{-n^{-1}(\partial^2 L/\partial\theta_i \, \partial\theta_j)^{-1}|_{\theta^0}\}$. A direct computation of this quantity gives the result of the theorem.

REFERENCES

CHENG, R. C. H., AMIN, N., and FEAST, G. M. (1979), "Maximum Likelihood Estimation of Parameters in the Three-

Parameter Inverse Gaussian Distribution," MATH Report 79-2, UWIST, Department of Mathematics, Cardiff.

COHEN, A. C. (1951), "Estimating Parameters of Logarithmic-Normal Distributions by Maximum Likelihood," *J. Amer. Statist. Assoc.*, 46, 206–212.

———— (1975), "Multi-Censored Sampling in the Three Parameter Weibull Distribution," *Technometrics*, 17, 347–351.

———— (1976), "Progressively Censored Sampling in the Three Parameter Log-Normal Distribution," *Technometrics*, 18, 99–103.

COHEN, A. C., and NORGAARD, N. J. (1977), "Progressively Censored Sampling in the Three-Parameter Gamma Distribution," *Technometrics*, 19, 333–340.

DUBEY, S. D. (1966), "Hyper-Efficient Estimator of the Location Parameter of the Weibull Laws," *Naval Research Logistic Quarterly*, 13, 253–263.

DUMONCEAUX, R., and ANTLE, C. E. (1973), "Discrimination Between the Lognormal and Weibull Distributions," *Technometrics*, 15, 923–926.

ENGELHARDT, M., and BAIN, L. J. (1979), "Prediction Limits and Two Sample Problems With Complete or Censored Weibull Data," *Technometrics*, 21, 233–237.

FOLKS, J. L., and CHHIKARA, R. S. (1978), "The Inverse Gauss-ian Distribution and its Statistical Application—A Review (with Discussion). *J. Roy. Statist. Soc.*, Ser. B, 40, 263–289.

HARTER, L. H., and MOORE, A. H. (1965), "Maximum Likelihood Estimation of the Parameters of Gamma and Weibull Populations From Complete and Censored Samples," *Technometrics*, 7, 639–643.

———— (1966), "Local Maximum-Likelihood Estimation of the Three-Parameter Lognormal Population From Complete and Censored Samples, *J. Amer. Statist. Assoc.*, 61, 842–851.

HILL, B. M. (1963), "The Three-Parameter Lognormal Distribution and Bayesian Analysis of a Point-Source Epidemic," *J. Amer. Statist. Assoc.*, 58, 72–84.

KENDALL, M. G., and STUART, A. (1969), *The Advanced Theory of Statistics*, Vol. 2, London: Griffin.

ROCKETTE, H., ANTLE, C., and KLIMKO, L. A. (1974), "Maximum Likelihood Estimation With the Weibull Model," *J. Amer. Statist. Assoc.*, 69, 246–249.

SHENTON, L. R., and BOWMAN, K. O. (1977), "Maximum Likelihood Estimation in Small Samples, London: Griffin.

STEEN, P. J., and STICKLER, D. J. (1976), "A Sewage Pollution Study of Beaches From Cardiff to Ogmore," UWIST, Dept. of Applied Biology Report, January, Cardiff.