4.1.8. Suppose we have observations **Z** on the variance component model in which

$$Z_{ij} = \mu + a_i + u_{ij}, \qquad j = 1, \ldots, m, i = 1, \ldots, g,$$

where $\{a_i\}$ are unobserved independent $N(0, \sigma_a^2)$ random variables which are independent of $\{u_{ij}\}$ which are independent $N(0, \sigma_u^2)$ random variables, and μ is an unknown parameter. Find the exact distribution of $\hat{\tau}_a/m = \sum_{i=1}^{g} (\bar{Z}_i - \bar{Z})^2/g$, where $\bar{Z}_i = m^{-1} \sum_{j=1}^{m} Z_{ij}$, and then show that it is a consistent estimator of $\tau_a/m = \sigma_a^2 + \sigma_u^2/m$ when $g \to \infty$ with m fixed, but not when $m \to \infty$ with g fixed. What happens when $m, g \to \infty$? Interpret the results.

4.2 MULTIPARAMETER PROBLEMS

In Section 4.1, we treated the pressure vessel failure time data as though the observations were actually generated by the exponential model (4.1). As we noted in Section 1.2.2, the gamma model is often used to explore the appropriateness of the exponential model. That is, we treat the data as a realization of **Z** generated by the model

$$\mathscr{F} = \left\{ f(\mathbf{y}, \lambda, \kappa) = \prod_{i=1}^{n} \frac{1}{\Gamma(\kappa)} \lambda(\lambda y_i)^{\kappa-1} \exp(-\lambda y_i), y_i > 0: \lambda > 0 \right\} \qquad (4.12)$$

and explore whether the nonexponentiality parameter κ is close to 1.

4.2.1 Maximum Likelihood Estimation Under the Gamma Model

The log-likelihood under the gamma model is

$$\ell(\lambda, \kappa) \propto n\kappa \log(\lambda) + \kappa \sum_{i=1}^{n} \log(z_i) - \sum_{i=1}^{n} \lambda z_i - n \log\{\Gamma(\kappa)\}$$

which is maximized at (λ, κ) satisfying

$$0 = \frac{n\kappa}{\lambda} - \sum_{i=1}^{n} z_i$$

$$0 = n \log(\lambda) + \sum_{i=1}^{n} \log(z_i) - n\psi(\kappa),$$

where $\psi(x) = \partial \log\{\Gamma(x)\}/\partial x$ is the digamma function. Using the first equation to eliminate λ from the second, we obtain after some manipulation

$$\lambda = \frac{\kappa}{\bar{z}}$$

and

$$0 = n^{-1} \sum_{i=1}^{n} \log(z_i) - \log(\bar{z}) - \psi(\kappa) + \log(\kappa). \tag{4.13}$$

Noting that the right-hand side of (4.13) is continuous in $\kappa > 0$, that

$$n^{-1} \sum_{i=1}^{n} \log(z_i) - \log(\bar{z}) < 0$$

by Jensen's inequality (see 1 in the Appendix), and that $\psi(\kappa) - \log(\kappa) < 0$,

$$\begin{aligned} \psi(\kappa) - \log(\kappa) &\to 0 \qquad \text{as } \kappa \to \infty \\ &\to -\infty \qquad \text{as } \kappa \to 0, \end{aligned}$$

we see that (4.13) always has at least one solution. Since the derivative of the right-hand side of (4.13) is

$$-\psi'(\kappa) + \frac{1}{\kappa} < 0,$$

(4.13) has precisely one solution and there is a unique $(\hat{\lambda}, \hat{\kappa})$ which maximizes the likelihood.

To find the maximum likelihood estimates $(\hat{\lambda}, \hat{\kappa})$, we need to solve (4.13) for κ and then set $\lambda = \kappa/\bar{z}$. Greenwood and Durand (1960) suggested an approximation to the solution of this equation: let $y = |n^{-1} \sum_{i=1}^{n} \log(z_i) - \log(\bar{z})|$ and then set

$$\hat{\kappa} = \begin{cases} \dfrac{0.5000876 + 0.1648852y - 0.0544274y^2}{y} & 0 < y < 0.5772 \\[2ex] \dfrac{8.898919 + 9.059950y + 0.9775373y^2}{y(17.79728 + 11.968477y + y^2)} & 0.5772 \le y \le 17. \end{cases}$$

We find that $y = 1.072$ so we should use the second approximation. Solving the equations for our data, we find that $(\hat{\lambda}, \hat{\kappa}) = (0.001, 0.579)$. The right-hand side of the estimating equation (4.13) evaluated at this value equals 0.0002.

The next step is to find an approximation to the sampling distribution of the maximum likelihood estimator $(\hat{\lambda}, \hat{\kappa})$. This is complicated by the fact that these estimators are only implicitly defined. However, we can obtain an asymptotic expansion for $(\hat{\lambda}, \hat{\kappa})$ and then approximate the asymptotic distribution of the terms in this expansion. The approach we use is much more transparent if, instead of restricting attention to the maximum likelihood estimators under the gamma model, we work with a general class of estimators and a general model.

4.2.2 Estimating Equations

Suppose that we treat the data as a realization of **Z** generated by the model

$$\mathscr{F} = \left\{ f(\mathbf{y}; \theta) = \prod_{i=1}^{n} f(y_i; \theta) \colon \theta \in \Omega \right\}. \tag{4.14}$$

Consider the class of estimators of θ which are solutions $\hat{\theta}$ of a general *estimating equation* of the form

$$\sum_{i=1}^{n} \eta(Z_i, \theta) = 0. \tag{4.15}$$

We call $\sum_{i=1}^{n} \eta(Z_i, \theta)$ an *estimating function* (Edgeworth, 1908–9; Godambe, 1960; 1991) and $\hat{\theta}$ a *maximum likelihood type* or *M-estimator* (Huber, 1964).

Maximum likelihood estimators $\hat{\theta}$ for the model $\mathscr{F}$ correspond to setting $\eta(x, \theta) = \partial \log \{f(x; \theta)\}/\partial\theta$ but we obtain a useful generalization of maximum likelihood estimation if we allow flexibility in the choice of η in (4.15).

If the expected value of the estimating function is 0 under $\mathscr{F}$ so

$$\int_{-\infty}^{\infty} \eta(z, \theta) f(z; \theta)\, dz = 0,$$

we say that the estimating equation is *unbiased* for θ under $\mathscr{F}$. Unbiasedness of the estimating function implies that when the estimating equation procedure is applied to the population represented by $\mathscr{F}$, the estimator is the parameter we are trying to estimate. This property is called *Fisher consistency*.

More formally, define a function $\theta(\cdot)$ from the set of all distribution functions to the parameter space Ω as a solution of the equation

$$\int_{-\infty}^{\infty} \eta(z, \theta)\, dF(z) = 0, \tag{4.16}$$

where F is an arbitrary distribution function. Replacing F in (4.16) by the empirical distribution function F_n (Section 1.5.2) produces (4.15) so we can write $\hat{\theta} = \theta(F_n)$. Let F_0 denote the distribution which generated **Z** so that $F_0(x) = F(x; \theta_0)$ for some θ_0 (called the *true parameter value*) whenever $F_0 \in \mathscr{F}$. Solving (4.16) at F_0 produces $\theta(F_0)$ and $\hat{\theta}$ is Fisher consistent for θ_0 if $\theta(F_0) = \theta_0$.

The formalization in terms of the function $\theta(\cdot)$ is useful because it tells us that $\hat{\theta}$ is estimating $\theta(F_0)$ when $Z_i, \ldots, Z_n$ are independent and identically distributed random variables with common distribution function ($F_0 \notin \mathscr{F}$).

Fisher consistency is not the same as consistency defined in Section 4.1.6 but it is closely related because F_n is consistent for F_0 and, when $\theta(\cdot)$ is continuous at F_0, it follows that $\hat{\theta}$ is consistent for $\theta(F_0)$. We will show in Section 4.2.4 that, under further conditions, $\theta(F_0)$ is the asymptotic mean of $\hat{\theta}$.

We derive the properties of M-estimators in Sections 4.2.4–4.2.6, specialize the results to maximum likelihood estimators in Sections 4.2.7–4.2.10, and then apply them to make inferences about the parameters in the gamma model (4.12) in Section 4.2.11.

4.2.3 Establishing Convergence for Random Vectors

In multiparameter problems, we have to establish approximations to the sampling distribution of a vector estimator. Although the calculations become more complicated, the manipulation of vectors raises no substantive difficulties. The techniques we use are very similar to those we would use in the case $p = 1$ and the results include $p = 1$ as a special case.

A simple extension of the central limit theorem can be established using the *Cramer–Wold device.*

Theorem 4.4 (Cramer and Wold, 1936) *The random p-vector X_n converges in distribution to X if and only if for each fixed p-vector a, $a^{\mathrm{T}}X_n$ converges in distribution to $a^{\mathrm{T}}X$.*

For our purposes we require only the multivariate version of the Lindeberg–Levy central limit theorem which is readily established from Corollary 4.1 and Theorem 4.4.

Theorem 4.5 *Let $Z_1, \ldots, Z_n$ be independent and identically distributed random p-vectors with mean p-vector μ and $p \times p$ variance matrix Σ. Then*

$$n^{-1/2} \sum_{i=1}^{n} (Z_i - \mu) \xrightarrow{\mathscr{D}} \mathrm{N}_p(0, \Sigma) \qquad \text{as } n \to \infty,$$

where N_p denotes the p-dimensional multivariate Gaussian distribution.

Convergence in probability is even easier to extend to the multiparameter case because a vector or matrix converges in probability if and only if its components do so.

4.2.4 The Approximate Sampling Distribution of an M-Estimator

The basic procedure for approximating the sampling distribution of an M-estimator is to expand the estimating equation (4.15) in a Taylor series (see 2 in the Appendix) about the point $\theta(F_0)$ to produce an expansion in increasing powers of $\hat{\theta} - \theta(F_0)$ which is of the form

$$0 = n^{-1}\sum_{i=1}^{n}\eta(Z_i,\hat{\theta}) = n^{-1}\sum_{i=1}^{n}\eta(Z_i,\theta(F_0)) + n^{-1}\sum_{i=1}^{n}\eta'(Z_i,\theta(F_0))(\hat{\theta}-\theta(F_0)) + \cdots,$$

where $\eta'(z,t)$ denotes the matrix with (i,j)th component $\partial\eta_i(z,t)/\partial t_j$. We then invert this expansion to obtain an expansion for $\hat{\theta} - \theta(F_0)$ in increasing powers of $n^{-1}\sum_{i=1}^{n}\eta(Z_i,\theta(F_0))$ which is typically of order $n^{-1/2}$ in probability and can be used to obtain approximations to the sampling distribution of $\hat{\theta} - \theta(F_0)$.

The procedure is particularly straightforward if we are only interested in the leading term. In this case, we have the expansion

$$n^{1/2}(\hat{\theta} - \theta(F_0)) = -\left\{n^{-1}\sum_{i=1}^{n}\eta'(Z_i,\theta^*)\right\}^{-1} n^{-1/2}\sum_{i=1}^{n}\eta(Z_i,\theta(F_0)), \quad (4.17)$$

where θ^* is between $\hat{\theta}$ and $\theta(F_0)$ in the sense that $|\theta^* - \theta(F_0)| \leq |\hat{\theta} - \theta(F_0)|$.

Provided $Z_1, \ldots, Z_n$ are independent and identically distributed random variables, $n^{-1/2}\sum_{i=1}^{n}\eta(Z_i,\theta(F_0))$ is a sum of independent and identically distributed random variables. If in addition

$$\mathrm{E}_{F_0}\eta(Z,\theta(F_0)) = 0,$$

and

$$\mathrm{E}_{F_0}\eta(Z,\theta(F_0))\eta(Z,\theta(F_0))^{\mathrm{T}} = A_{F_0}(\theta(F_0)) < \infty,$$

it follows from the central limit theorem that

$$n^{-1/2}\sum_{i=1}^{n}\eta(Z_i,\theta(F_0)) \xrightarrow{\mathcal{D}} \mathrm{N}(0, A_{F_0}(\theta(F_0))). \quad (4.18)$$

Next, write

$$n^{-1}\sum_{i=1}^{n}\eta'(Z_i,\theta^*) = n^{-1}\sum_{i=1}^{n}\eta'(Z_i,\theta(F_0)) + n^{-1}\sum_{i=1}^{n}\{\eta'(Z_i,\theta^*) - \eta'(Z_i,\theta(F_0))\}. \quad (4.19)$$

The first term is the mean of independent and identically distributed random variables so, provided $-\mathrm{E}_{F_0}\eta'(Z,\theta(F_0)) = B_{F_0}(\theta(F_0)) < \infty$, the weak law of large numbers ensures that

$$n^{-1}\sum_{i=1}^{n}\eta'(Z_i,\theta(F_0)) = -B_{F_0}((\theta(F_0)) + o_p(1). \quad (4.20)$$

Also, provided $\hat{\theta} = \theta(F_0) + o_p(1)$ and $\eta'(y, t)$ is continuous at $t = \theta(F_0)$ uniformly in y, we can show that

$$n^{-1} \sum_{i=1}^{n} \{\eta'(Z_i, \theta^*) - \eta'(Z_i, \theta(F_0))\} = o_p(1). \tag{4.21}$$

Finally, if $B_{F_0}(\theta(F_0))$ is nonsingular, we can apply Theorem 4.2 to (4.17)–(4.21) to obtain

$$n^{1/2}(\hat{\theta} - \theta(F_0)) = B_{F_0}(\theta(F_0))^{-1} n^{-1/2} \sum_{i=1}^{n} \eta(Z_i, \theta(F_0)) + o_p(1)$$

from which it follows that

$$n^{1/2}(\hat{\theta} - \theta(F_0)) \xrightarrow{\mathscr{D}} \mathrm{N}(0, B_{F_0}(\theta(F_0))^{-1} A_{F_0}(\theta(F_0)) B_{F_0}(\theta(F_0))^{-\mathrm{T}}),$$

where $M^{-\mathrm{T}} = (M^{-1})^{\mathrm{T}}$.

Collecting all the conditions, we have proved the following result.

Theorem 4.6 *Let $Z_1, \ldots, Z_n$ be independent and identically distributed random variables with common distribution function F_0. Suppose that*

1. *$\theta(F_0)$ is an interior point of the parameter space Ω and an isolated root of the equation $\mathrm{E}_{F_0}\eta(Z_i, \theta) = 0$,*
2. *$\eta'(y, t)$ is continuous at $t = \theta(F_0)$ uniformly in y,*
3. *$\mathrm{E}_{F_0}\eta(Z_i, \theta(F_0))\eta(Z_i, \theta(F_0))^{\mathrm{T}} = A_{F_0}(\theta(F_0)) < \infty$*

and

4. *$\mathrm{E}_{F_0}\eta'(Z_i, \theta(F_0)) = B_{F_0}(\theta(F_0)) < \infty$ and nonsingular.*

Then if $\hat{\theta} = \theta(F_0) + o_p(1)$ and $n^{-1/2} \sum_{i=1}^{n} \eta(Z_i, \hat{\theta}) = o_p(1)$,

$$n^{1/2}(\hat{\theta} - \theta(F_0)) = -B_{F_0}(\theta(F_0))^{-1} n^{-1/2} \sum_{i=1}^{n} \eta(Z_i, \theta(F_0)) + o_p(1)$$

which implies that

$$n^{1/2}(\hat{\theta} - \theta(F_0)) \xrightarrow{\mathscr{D}} \mathrm{N}(0, B_{F_0}(\theta(F_0))^{-1} A_{F_0}(\theta(F_0)) B_{F_0}(\theta(F_0))^{-\mathrm{T}}).$$

The last two (unnumbered) conditions ensure that $\hat{\theta}$ is a consistent estimator which satisfies the estimating equations. Condition 1 ensures that $\theta(F_0)$ is not on the boundary of Ω so that a Gaussian approximation to the sampling distribution is plausible (see Moran, 1971). Condition 2 justifies the Taylor series expansion of the estimating equations and ensures that the remainder vanishes. Conditions 3 and 4 justify the application of the weak law of large numbers to the denominator and then the central limit theorem to the numerator in the rearranged Taylor expansion expressing $n^{1/2}(\hat{\theta} - \theta(F_0))$ as the ratio of two terms.

A more general version of Theorem 4.6 has been given by Huber (1967).

4.2.5 Approximate Standard Errors for Estimating *M*-Estimators

Theorem 4.6 establishes that the asymptotic variance of an M-estimator $\hat{\theta}$ is

$$V_{F_0}(\theta(F_0)) = n^{-1}B_{F_0}(\theta(F_0))^{-1}A_{F_0}(\theta(F_0))B_{F_0}(\theta(F_0))^{-\mathrm{T}}.$$

A natural estimator of $V_{F_0}(\theta(F_0))$ is

$$\hat{V}(\hat{\theta}) = n^{-1}\hat{B}(\hat{\theta})^{-1}\hat{A}(\hat{\theta})\hat{B}(\hat{\theta})^{-\mathrm{T}}, \tag{4.22}$$

where

$$\hat{A}(\theta) = n^{-1}\sum_{i=1}^{n} \eta(Z_i; \theta)\eta(Z_i; \theta)^{\mathrm{T}} \quad \text{and} \quad \hat{B}(\theta) = n^{-1}\sum_{i=1}^{n} \eta'(Z_i; \theta).$$

The estimator (4.22) is consistent for $V_{F_0}(\theta(F_0))$ under the conditions of Theorem 4.6 so $\{\text{diag } \hat{V}(\hat{\theta})\}^{1/2}$ gives approximate standard errors for $\hat{\theta}$.

4.2.6 Solving Estimating Equations

The Newton–Raphson method described in Section 2.7.7 as an algorithm for obtaining the maximum likelihood estimates can be applied to solve the more general (4.15). The algorithm is based on a linear expansion of the estimating function in (4.15) instead of a quadratic expansion of the log-likelihood function but the end result is the same.

We can modify the Newton–Raphson method by replacing the normalized Hessian matrix $n^{-1}\sum_{i=1}^{n} \eta'(Z_i, \theta_{(m)})$ by the estimate $-B_{F(\cdot;\theta_{(m)})}(\theta_{(m)})$ of its limit under the model $\mathscr{F}$. At least when the estimating function is the derivative of the log-likelihood, the resulting algorithm is known as *Fisher's method of scoring*. As we saw in Section 4.2.1, the form of η may suggest additional alternative algorithms.

4.2.7 Why Maximum Likelihood Estimates the True Parameter

Suppose that the model $\mathscr{F} = \{f(\mathbf{y}; \theta) = \prod_{i=1}^{n} f(y_i; \theta): \theta \in \Omega\}$ holds so that $F_0 \in \mathscr{F}$. Let θ_0 denote the true parameter value which identifies the distribution in the model which actually generated the data so F_0 denotes the distribution with density $f(y; \theta_0)$. By Jensen's inequality (see 1 in the Appendix)

$$\mathrm{E}_{F_0} \log \left\{\frac{f(Z;\theta)}{f(Z;\theta_0)}\right\} < \log \mathrm{E}_{F_0} \left\{\frac{f(Z;\theta)}{f(Z;\theta_0)}\right\}$$

$$= \log \left\{\int_{\text{support}\{f(\cdot;\theta_0)\}} f(y;\theta)\,dy\right\}$$

If the densities in $\mathscr{F}$ have the same support and the densities in the model are distinct in the sense that $f(y; \theta_0) \neq f(y; \theta)$ whenever $\theta \neq \theta_0$,

$$\mathrm{E}_{F_0} \log \left\{\frac{f(Z;\theta)}{f(Z;\theta_0)}\right\} < 0 \tag{4.23}$$

and θ_0 maximizes $\mathrm{E}_{F_0} \log \{f(Z; \theta)\}$.

Since $\mathrm{E}_{F_0} \log \{f(Z_i; \theta)\}$ is maximized at $\theta = \theta_0$, the true value θ_0 satisfies the equation

$$\left.\frac{\partial \mathrm{E}_{F_0} \log f(Z;\theta)}{\partial \theta}\right|_{\theta=\theta_0} = 0.$$

If we can interchange the order of expectation (i.e., integration) and differentiation we have that

$$\mathrm{E}_{F_0}\eta(Z,\theta_0) = \mathrm{E}_{F_0} \left.\frac{\partial \log f(Z;\theta)}{\partial \theta}\right|_{\theta=\theta_0} = 0$$

so the estimating equation is unbiased for θ_0 when $\mathscr{F}$ holds and $\hat{\theta}$ is Fisher consistent for θ_0.

4.2.8 The Approximate Sampling Distribution of Maximum Likelihood Estimators

A simplification to Theorem 4.6 is often available for maximum likelihood estimators. If we can interchange the order of integration and differentiation twice,

$$\begin{aligned} B_{F_0}(\theta_0) &= -\mathrm{E}_{F_0}\eta'(Z,\theta_0) \\ &= -\int \frac{\partial^2 \log f(x,\theta_0)}{\partial \theta_0\, \partial \theta_0^{\mathrm{T}}} f(x,\theta_0)\,dx \\ &= -\int \left[\frac{1}{f(x,\theta_0)} \frac{\partial^2 f(x,\theta_0)}{\partial \theta_0\, \partial \theta_0^{\mathrm{T}}} - \frac{1}{f(x,\theta_0)^2} \frac{\partial f(x,\theta_0)}{\partial \theta_0} \frac{\partial f(x,\theta_0)^{\mathrm{T}}}{\partial \theta_0}\right] f(x,\theta_0)\,dx \end{aligned}$$

$$= -\int \frac{\partial^2 f(x,\theta_0)}{\partial\theta_0\,\partial\theta_0^{\mathrm{T}}}\,dx + \int \frac{\partial \log\{f(x,\theta_0)\}}{\partial\theta_0}\frac{\partial \log\{f(x,\theta_0)\}^{\mathrm{T}}}{\partial\theta_0} f(x,\theta_0)\,dx$$

$$= -\frac{\partial^2}{\partial\theta_0\,\partial\theta_0^{\mathrm{T}}}\int f(x,\theta_0)\,dx + \mathrm{E}_{F_0}\eta(Z,\theta_0)\eta(Z,\theta_0)^{\mathrm{T}}$$

$$= A_{F_0}(\theta_0).$$

The common value of $A_{F_0}(\theta) = B_{F_0}(\theta)$ is denoted by $I(\theta)$ and is the Fisher information matrix defined in Section 2.3.2.

Specializing the statement of Theorem 4.6, we have the following result.

Corollary 4.6 *Let $Z_1, \ldots, Z_n$ be observations on a model*

$$\mathscr{F} = \left\{ f(\mathbf{y};\theta) = \prod_{i=1}^{n} f(y_i;\theta) \colon \theta \in \Omega \right\}$$

which satisfies

1. *θ_0 is an interior point of the parameter space Ω*
2. *the support of $f(\cdot,\theta)$ does not depend on θ and $f(y;\theta_0) \neq f(y;\theta)$ for $\theta_0 \neq \theta$*
3. *$\eta(x,\theta) = \partial \log\{f(x,\theta)\}/\partial\theta$ and the second derivative of $f(x,\theta)$ with respect to θ is finite for each x in the support of $f(x,\theta)$ and continuous at θ_0 uniformly in x*
4. *the integral $\int f(x,\theta)\,dx$ can be differentiated twice under the integral sign*
5. *the Fisher information*

$$I(\theta) = \int_{-\infty}^{\infty} \frac{\partial \log f(x,\theta)}{\partial\theta}\frac{\partial \log f(x,\theta)^{\mathrm{T}}}{\partial\theta} f(x,\theta)\,dx$$

is finite and nonsingular at $\theta = \theta_0$.

Then if $\hat\theta = \theta_0 + o_p(1)$ and $n^{-1/2}\sum_{i=1}^{n}\eta(Z_i,\hat\theta) = o_p(1)$,

$$n^{1/2}(\hat\theta - \theta_0) = -I(\theta_0)^{-1} n^{-1/2} \sum_{i=1}^{n} \eta(Z_i,\theta_0) + o_p(1)$$

which implies that

$$n^{1/2}(\hat\theta - \theta_0) \xrightarrow{\mathscr{D}} \mathrm{N}(0, I(\theta_0)^{-1}).$$

4.2.9 Approximate Standard Errors for Maximum Likelihood Estimators

To use Corollary 4.6 to make inferences about θ_0, we need to construct an estimator of the Fisher information matrix $I(\theta_0)$. We could use the estimator $\hat{V}(\hat{\theta})$ defined in (4.22) but, under the assumed model, we usually use the *observed information*

$$\mathscr{I} = -n^{-1} \sum_{i=1}^{n} \left. \frac{\partial^2 \log f(Z_i, \theta)}{\partial \theta^2} \right|_{\theta=\hat{\theta}}$$

or the *expected information*

$$I(\hat{\theta}) = -\left. \left\{ \mathrm{E}_\theta \frac{\partial^2 \log f(Z, \theta)}{\partial \theta^2} \right\} \right|_{\theta=\hat{\theta}}.$$

These estimators are generally different except in the case of distributions in the exponential family (Section 1.3.1) for which they are identical. In particular, they are identical for the gamma model (4.12). Both of these estimators are consistent for $I(\theta_0)$ under the conditions of Corollary 4.6 so either estimator can be used to make approximate inferences about θ_0. Both estimators require $\hat{\theta}$ but $\mathscr{I}$ is typically simpler to obtain than $I(\hat{\theta})$ because it does not require the expectation of the second derivative matrix.

Using the observed Fisher information, an approximate $100(1-\alpha)\%$ confidence interval for θ_{01} is given by

$$[\hat{\theta}_1 - n^{-1/2}(\mathscr{I}^{11})^{1/2}\Phi^{-1}(1-\alpha/2), \hat{\theta}_1 + n^{-1/2}(\mathscr{I}^{11})^{1/2}\Phi^{-1}(1-\alpha/2)], \quad (4.24)$$

where $\mathscr{I}^{11}$ denotes the (1, 1)th element of $\mathscr{I}^{-1}$. Confidence intervals for the other components of θ_0 are easily obtained.

4.2.10 Consistency of Maximum Likelihood Estimators

To apply Corollary 4.6, we still have to show that $\hat{\theta}$ is a consistent estimator of θ_0. This is surprisingly difficult to do and requires rather technical conditions.

Since we are trying to maximize the log-likelihood $\sum_{i=1}^{n} \log \{f(Z_i, \theta)\}$ to estimate θ_0, we can base a consistency proof on the likelihood function. This has been done very elegantly by Wald (1949) who first proved consistency for the case that the parameter space Ω contains only a finite number of points and then extended the result to more general sets Ω satisfying compactness conditions which enable them to be approximated by finite sets.

An alternative approach due to Cramer (1946, pp. 500–4) is to show that a root of the likelihood equation is consistent. The difficulty with this approach is that if the equtions have multiple roots, it is impossible to tell which of these are consistent. This difficulty can be overcome by specifying an algorithm for choosing a single root and then showing that this root is consistent.

4.2.11 Maximum Likelihood Inference Under the Gamma Model

Technical arguments can be used to show that distributions in the exponential family satisfy the conditions of Corollary 4.6; see for example Lehmann (1959/1991, pp. 57–60; 1983, pp. 438–42). For our problem with the gamma model (4.12), $\theta = (\lambda, \kappa)^{\mathrm{T}}$ and

$$\eta(x, \theta) = \left(\frac{\kappa}{\lambda} - x, \log(\lambda) + \log(x) - \psi(\kappa)\right)^{\mathrm{T}}.$$

It follows that

$$\begin{aligned} I(\theta) &= -\mathrm{E}\eta'(Z, \theta) \\ &= \begin{pmatrix} \kappa/\lambda^2 & -\lambda^{-1} \\ -\lambda^{-1} & \psi'(\kappa) \end{pmatrix} \end{aligned}$$

and inverting this matrix, we obtain

$$I(\theta)^{-1} = \frac{\lambda^2}{\kappa\psi'(\kappa) - 1}\begin{pmatrix} \psi'(\kappa) & \lambda^{-1} \\ \lambda^{-1} & \kappa/\lambda^2 \end{pmatrix}.$$

Solving the estimating equations for our data, we find that $\hat{\theta} = (0.001, 0.579)$ and

$$\left(\frac{\mathscr{I}}{20}\right)^{-1} = \begin{pmatrix} 1.58 \times 10^{-7} & 4.08 \times 10^{-5} \\ 4.08 \times 10^{-5} & 0.023 \end{pmatrix}$$

An approximate 95% confidence interval for κ is obtained from (4.24) as

$$(0.27, 0.88).$$

This interval does not contain $\kappa = 1$ so provides evidence against the adoption of an exponential model for the pressure vessel failure data.

The fact that κ is a non-negative parameter suggests that we should consider a log-normal approximation to the sampling distribution of $\hat{\kappa}$. From Theorem 4.3, we obtain

$$\mathrm{Var}(\log(\hat{\kappa})) \sim \frac{1}{n\kappa(\kappa\psi'(\kappa) - 1)}$$

which produces the standard error 0.265 and hence the 95% confidence interval

$$(0.34, 0.97).$$

Although both approximations lead to similar conclusions in this case, we may still seek to investigate which scale provides the better approximation. Ideally, we would like to compare the approximations to the exact results but we do not know the exact distribution of $\hat{\kappa}$. Nonetheless, we can make some progress through a computer simulation (see Section 3.10). We can generate 1000 data sets of size $n = 20$ from a gamma distribution with $\lambda = 0.001$ and $\kappa_0 = 0.5$, compute a nominal 95% confidence interval for κ from each data set using the two approximations, and then compare the estimated coverage probabilities of the intervals (the proportion of intervals containing the actual value of κ_0) and the distributions of the lengths of the intervals.

Using the Gaussian approximation to the binomial distribution to set confidence intervals for the actual coverage probabilities (see Section 3.7.2), we find that the approximation on the raw scale produces nominal 95% confidence intervals for κ_0 which have an estimated coverage probability of $0.97 \pm 1.96\sqrt{0.97 \times 0.03/1000} = 0.97 \pm 0.01$ whereas that on the log scale produces nominal 95% confidence intervals for κ_0 which have an estimated coverage probability of $0.92 \pm 1.96\sqrt{0.92 \times 0.08/1000} = 0.92 \pm 0.01$. The distributions of the lengths of the confidence intervals are shown in Figure 4.3. Using the Gaussian approximation to the distribution of the mean lengths, we see

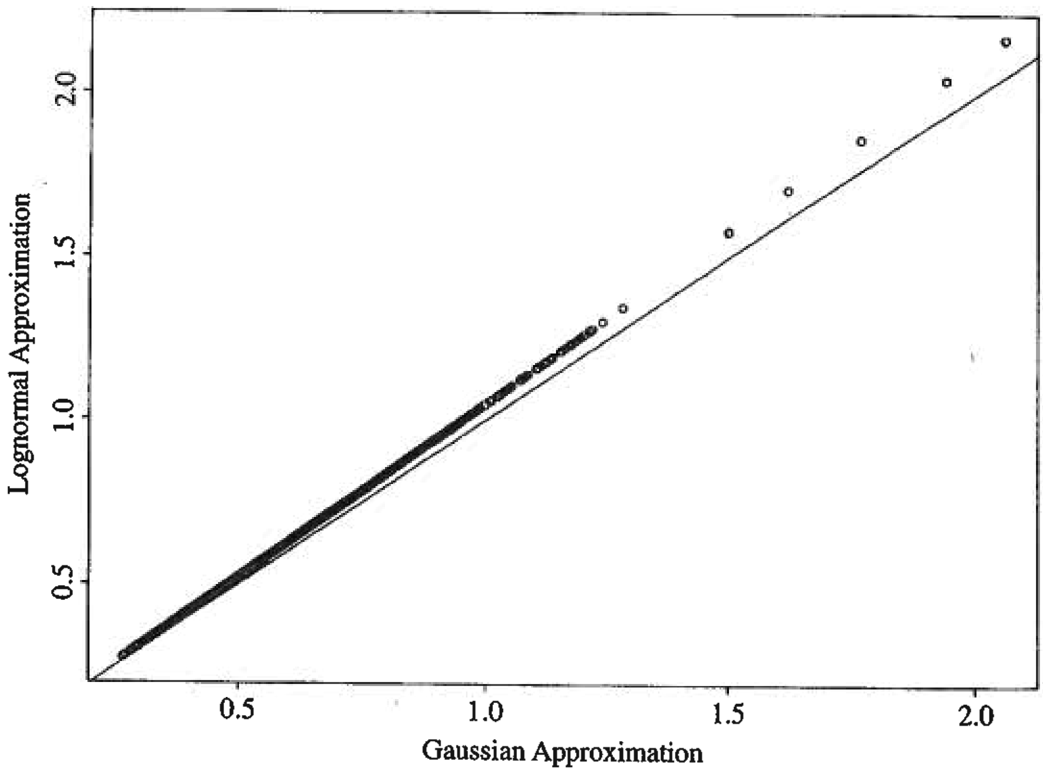

Figure 4.3. A qq-plot of the lengths of simulated 95% confidence intervals using the Gaussian and lognormal approximations to the sampling distribution of the maximum likelihood estimator of the gamma shape parameter κ under the $\Gamma(0, 5, 0.001)$ model.

that the approximation on the raw scale produces nominal 95% confidence intervals for κ_0 with mean length $0.60 \pm 1.96 \times 0.006 = 0.60 \pm 0.01$ whereas that on the log scale produces nominal 95% confidence intervals for κ_0 with mean length $0.63 \pm 1.96 \times 0.007 = 0.63 \pm 0.01$. There is not a great difference between the results produced by the two methods but the confidence intervals produced on the raw scale have slightly better coverage and are typically slightly shorter than those produced on the log scale. More detailed comparisons can be made by extending the range of n, λ, κ, and nominal levels considered.

PROBLEMS

4.2.1. Suppose that we have observations **Z** on the multivariate Student t model

$$\mathscr{F} = \left\{ f(\mathbf{y}; \mu, \sigma) = \frac{\Gamma\left(\dfrac{\nu + n}{2}\right)}{(\nu\pi\sigma^2)^{n/2}\Gamma\left(\dfrac{\nu}{2}\right)} \frac{1}{\left\{1 + \sum_{i=1}^{n} \dfrac{(y_i - \mu)^2}{2\nu}\right\}^{(\nu+n)/2}} \right.$$
$$\left. -\infty \le y_i \le \infty : \mu \in \mathbb{R}, \sigma > 0 \right\},$$

where $\nu > 0$ is known. Use the representation $Z_i = \mu + \sigma Y_i/h^{1/2}$, where Y_i are independent standard Gaussian random variables which are independent of $h \sim \chi_\nu^2/\nu$ to find the sampling distribution of the maximum likelihood estimator $\hat{\mu}$ of μ as $n \to \infty$. Show that $\hat{\sigma}^2$ is inconsistent by showing that it converges in probability to the random variable σ^2/h.

4.2.2. Suppose that we observe **Z** on the model

$$\mathscr{F} = \left\{ f(\mathbf{z}; \theta) = \prod_{i=1}^{n} \theta z_i^{\theta - 1} \exp(-z_i^\theta), z_i > 0 : \theta > 0 \right\}.$$

The maximum likelihood estimator of θ cannot be written down explicitly. Nonetheless, show that the likelihood equations have a unique root which equals the maximum likelihood estimator. (Hint: show that the likelihood equation is a continuous function of θ, takes positive and negative values and crosses the zero axis once.) Find the asymptotic sampling distribution of the maximum likelihood estimator and show how to use it to set an approximate $100(1 - \alpha)\%$ confidence interval for θ.

4.2.3. *Pareto's distribution* is sometimes used to represent the distribution of incomes over a population. Suppose that we observe **Z** on the Pareto model

$$\mathscr{F} = \left\{ f(\mathbf{y}; \lambda) = \prod_{i=1}^{n} \frac{\kappa\lambda^{\kappa}}{y_i^{\kappa+1}},\ y_i > \lambda : \kappa, \lambda > 0 \right\}.$$

Suppose initially that λ is known. Obtain the maximum likelihood estimator $\hat{\kappa}$ of κ. Find the asymptotic distribution of $\hat{\kappa}$ and hence of the maximum likelihood estimator of the median $\lambda 2^{1/\kappa}$. Construct a $100(1-\alpha)\%$ large sample confidence interval for the median income in the population.

4.2.4. In the context of Problem 4.2.3, suppose now that both κ and λ are unknown. Obtain the maximum likelihood estimators of κ and λ. Show that $\hat{\lambda} - \lambda = O_p(n^{-1})$. Hence or otherwise show that the asymptotic distribution of the maximum likelihood estimator of the median when λ is unknown is the same as when λ is known. (Malik, 1970, has obtained the exact distribution theory for this problem.)

4.2.5. We noted in Section 1.3.2 that the Weibull model is often used in place of the gamma model to explore the applicability of the exponential model. In the Weibull model we treat the data as a realization of **Z** generated by

$$\mathscr{F} = \left\{ f(\mathbf{y}, \lambda, \kappa) = \prod_{i=1}^{n} \kappa\lambda(\lambda y_i)^{\kappa-1} \exp\{-(\lambda y_i)^{\kappa}\},\ y_i > 0 : \lambda > 0 \right\}.$$

Show that the log-likelihood is maximized at (λ, κ) satisfying

$$\lambda = \frac{1}{\{n^{-1}\sum_{i=1}^{n} z_i^{\kappa}\}^{1/\kappa}}$$

$$0 = \frac{1}{\kappa} + n^{-1}\sum_{i=1}^{n} \log(z_i) - \frac{\sum_{i=1}^{n} z_i^{\kappa} \log(z_i)}{\sum_{i=1}^{n} z_i^{\kappa}}$$

and show that there is a unique $(\hat{\lambda}, \hat{\kappa})$ which maximizes the likelihood. Write down an approximation to the sampling distribution of the maximum likelihood estimator.

4.2.6. Fit the Weibull model of Problem 4.2.5 to the pressure vessel failure time data presented in Table 1.2 and use it to make inferences about κ and then the median of the failure time distribution. Carry out a simulation to explore the quality of the Gausian approximation and the repeated sampling properties of the inferences about κ at the estimated parameter values.

4.2.7. Suppose that the conditions of Theorem 4.6 hold and that $\theta^* - \theta(F_0) = O_p(n^{-1/2})$. Show that the one-step estimator

$$\hat{\theta} = \theta^* - \left\{\sum_{i=1}^{n} \eta'(Z_i, \theta^*)\right\}^{-1} \sum_{i=1}^{n} \eta(Z_i, \theta^*)$$

is asymptotically equivalent to a root of the estimating equations (4.15).

4.3 THE CHOICE OF INFERENCE PROCEDURE

We showed in Sections 4.2.6–4.2.11 that maximizing the likelihood for the gamma model (4.12) is a reasonable method of estimating the parameters of the model. However, the implementation of the procedure requires us to solve an implicit equation and to approximate the sampling distribution of the implicitly defined estimator so other approaches may be simpler to implement.

4.3.1 Method of Moments Estimation for the Gamma Model

If we compute the sample moments $m_k = n^{-1} \sum_{j=1}^{n} Z_j^k$, $k = 1, 2$, their expectations under the gamma model (4.12), which are $\mathrm{E}m_1 = \kappa/\lambda$ and $\mathrm{E}m_2 = \kappa(1 + \kappa)/\lambda^2$, and then solve the system of equations

$$m_1 = \frac{\hat{\kappa}_{\mathrm{m}}}{\hat{\lambda}_{\mathrm{m}}}$$

$$m_2 = \frac{\hat{\kappa}_{\mathrm{m}}(1 + \hat{\kappa}_{\mathrm{m}})}{\hat{\lambda}_{\mathrm{m}}^2},$$

we obtain the explicit method of moments estimators (Section 3.1.1)

$$\hat{\lambda}_{\mathrm{m}} = \frac{\hat{\kappa}_{\mathrm{m}}}{m_1}$$

$$\hat{\kappa}_{\mathrm{m}} = \frac{m_1^2}{m_2 - m_1^2}$$

of λ and κ.

4.3.2 The Sampling Distribution of the Method of Moments Estimators

We can apply Theorem 4.6 with $\theta = (\lambda, \kappa)^{\mathrm{T}}$ and

$$\eta(y, \theta) = (y - \kappa/\lambda,\ y^2 - \kappa(1 + \kappa)/\lambda^2)^{\mathrm{T}}$$