CHAPTER 17

Gamma Distributions

1 DEFINITION

A random variable X has a gamma distribution if its probability density function is of form

$$p_X(x) = \frac{(x-\gamma)^{\alpha-1}\exp[-(x-\gamma)/\beta]}{\beta^\alpha\Gamma(\alpha)}, \qquad \alpha > 0, \beta > 0; x > \gamma. \tag{17.1}$$

This distribution–denoted gamma (a, β, γ)—is Type **III** of Pearson's system (Chapter 12, Section 4). It depends on three parameters a, β, and y. If $\gamma = 0$, the distribution is termed a two-parameter gamma distribution, denoted gamma (a, β); see equation (17.23).

The standard form of distribution is obtained by setting $\beta = 1$ and $\gamma = 0$. This gives

$$p_X(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}, \qquad x \geq 0. \tag{17.2}$$

If a = 1, we have an exponential distribution (see Chapter 19). If a is a positive integer, we have an Erlang distribution.

The distributions of Y = –X, namely

$$p_Y(y) = \frac{(-y-\gamma)^{\alpha-1}\exp[(y+\gamma)/\beta]}{\beta^\alpha\Gamma(\alpha)}, \qquad y \leq -\gamma, \qquad (17.1)'$$

and

$$p_Y(y) = \frac{(-y)^{\alpha-1}e^{y}}{\Gamma(\alpha)}, \qquad y \leq 0, \qquad (17.2)'$$

are also gamma distributions. But such distributions rarely need to be considered, and we will not discuss them further here.

The probability integral of distribution (17.2) is

$$\Pr[X \leq x] = [\Gamma(\alpha)]^{-1} \int_0^x t^{\alpha-1} e^{-t}\, dt. \tag{17.3}$$

This is an incomplete gamma function ratio. The quantity

$$\Gamma_x(\alpha) = \int_0^x t^{\alpha-1} e^{-t}\, dt \tag{17.4}$$

is sometimes called an incomplete gamma *function*, but this name is also quite commonly applied to the ratio (17.3) (Chapter 1, Section 3).

This ratio depends on x and a, and it would be natural to use a notation representing it as a function of these variables. However, Pearson (1922) found it more convenient to use $u = x\alpha^{-1/2}$ in place of x for tabulation purposes, and he defined the incomplete gamma function as

$$I(u, \alpha - 1) = \frac{1}{\Gamma(\alpha)} \int_0^{u\sqrt{\alpha}} t^{\alpha-t} e^{-t}\, dt. \tag{17.5}$$

The main importance of the (standard) gamma distribution in statistical theory is the fact that if $U_1, U_2, \ldots, U_\nu$ are independent unit normal variables, the distribution of $\Sigma_{j=1}^{\nu} U_j^2$ is of form (17.1) with a $= \nu/2$, $\beta = 2$, and $\gamma = 0$. This particular form of gamma distribution is called a chi-square distribution with ν degrees of freedom. The corresponding random variable is often denoted by χ_ν^2, and we will follow this practice. It is clear that $\frac{1}{2}\Sigma_{j=1}^{\nu} U_j^2$ has a standard gamma distribution with a $= \nu/2$. Expressed symbolically:

$$p_{\chi_\nu^2}(x^2) = \left\{2^{\nu/2}\Gamma\left(\tfrac{1}{2}\nu\right)\right\}^{-1} (x^2)^{(\nu/2)-1} \exp\left(-\tfrac{1}{2}x^2\right), \qquad x^2 \geq 0. \tag{17.6}$$

Although in the definition above ν must be an integer, the distribution (17.6) is also called a "χ^2 distribution with ν degrees of freedom" if ν is any positive number. This distribution is discussed in detail in Chapter 18.

2 MOMENTS AND OTHER PROPERTIES

The moment generating function of the standard gamma distribution (17.2) is

$$E[e^{tX}] = \{\Gamma(\alpha)\}^{-1} \int_0^{m} x^{\alpha-1} \exp[-(1-t)x]\, dx = (1-t)^{-\alpha}, \qquad t < 1. \tag{17.7}$$

The characteristic function is $(1 - it)^{-\alpha}$.

Since distributions of form (17.1) can be obtained from those of form (17.2) by the linear transformation $X = (X' - \gamma)/\beta$, there is no difficulty in deriving formulas for moments, generating functions, and so on, for (17.1) from those for (17.2).

The formula for the rth moment about zero of distribution (17.2) is

$$\mu'_r = \{\Gamma(\alpha)\}^{-1} \int_0^\infty x^{\alpha+r-1} e^{-x}\, dx = \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)}. \tag{17.8}$$

From (17.8) cumulants can be obtained. These are very simple:

$$\kappa_r = (r - 1)!\alpha. \tag{17.9}$$

Hence for distribution (17.2)

$$\begin{aligned} E[X] &= \operatorname{Var}(X) = \alpha, \\ \mu_3 &= 2\alpha, \\ \mu_4 &= 3\alpha^2 + 6\alpha, \end{aligned} \tag{17.10}$$

so

$$\begin{aligned} \alpha_3 &= \sqrt{\beta_1} = 2\alpha^{-1/2}, \\ \alpha_4 &= \beta_2 = 3 + 6\alpha^{-1}. \end{aligned} \tag{17.11}$$

The mean deviation of distribution (17.2) is

$$\frac{2\alpha^{\alpha} e^{-\alpha}}{\Gamma(\alpha)}. \tag{17.12}$$

The standard distribution (17.2) has a single mode at $x = a - 1$ if $a \geq 1$. [Distribution (17.1) has a mode at $x = y + \beta(\alpha - 1)$.] If $a < 1$, $p_X(x)$ tends to infinity as x tends to zero; if $a = 1$ (the standard exponential distribution), $\lim_{x \to 0} p_X(x) = 1$.

There are points of inflexion, equidistant from the mode, at

$$x = \alpha - 1 \pm \sqrt{\alpha - 1} \tag{17.13}$$

(provided that the values are real and positive). The standardized variable

$$W = \frac{X - \alpha}{\sqrt{\alpha}} \tag{17.14}$$

is referred to as the frequency factor by hydrologists in flood frequency analysis [see, e.g., Phien (1991) or Chow (1969)].

Vodă (1974) studies a reparametrized version of the gamma distribution (17.1), with $\gamma = 0$, whose pdf is

$$p_X(x;\alpha,\theta) = \left(\frac{\alpha}{\theta}\right)^{\alpha} \times \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp\left(\frac{-\alpha x}{\theta}\right), \qquad x > 0;\ \alpha,\theta > 0. \tag{17.15}$$

Some typical probability density functions are shown in Figures 17.1 and 17.2. Figure 17.2 shows three different gamma distributions (17.1), each having the same expected value (zero) and standard deviation (unity).

$\alpha = 1$; $p(x) = \exp[-(x+1)]$ $(x > -1)$ Mode at -1

$\alpha = 4$; $p(x) = \frac{8}{3}(x+2)^3 \exp[-2(x+2)]$ $(x > -2)$ Mode at $-\frac{1}{2}$

$\alpha = 9$; $p(x) = \frac{2187}{4480}(x+3)^8 \exp[-3(x+3)]$ $(x > -3)$ Mode at $-\frac{1}{3}$

It can be seen from Figure 17.1 that, as α increases, the shape of the curve becomes similar to the normal probability density curve. In fact the standardized gamma distribution tends to the unit normal distribution as the value of the parameter α tends to infinity:

$$\lim_{\alpha\to\infty} \Pr\left[(X-\alpha)\alpha^{-1/2} \le u\right] = \Phi(u) \tag{17.16}$$

for all real values of u, where $\Phi(u) = (2\pi)^{-1/2}\int_{-\infty}^{u} \exp(-\frac{1}{2}t^2)\,dt$.

A similar result holds for the general distribution (17.1), namely

$$\lim_{\alpha\to\infty} \Pr\left[\left\{\frac{X-\gamma}{\beta} - \alpha\right\}\alpha^{-1/2} \le u\right] = \Phi(u). \tag{17.17}$$

For χ^2_ν,

$$\lim_{\nu\to\infty} \Pr\left[(\chi^2_\nu - \nu)(2\nu)^{-1/2} \le u\right] = \Phi(u). \tag{17.18}$$

It can be checked from (17.10) and (17.11) that $\alpha_3 \to 0$, $\alpha_4 \to 3$ (the values for the normal distribution) as α, ν, respectively, tend to infinity.

One of the most important properties of the distribution is the reproductive property: If X_1, X_2 are independent random variables each having a distribution of form (17.1), with possibly different values α', α'' of α, but with common values of β and γ, then $(X_1 + X_2)$ also has a distribution of this form, with the same value of β, double the value of γ, and with $\alpha = \alpha' + \alpha''$.

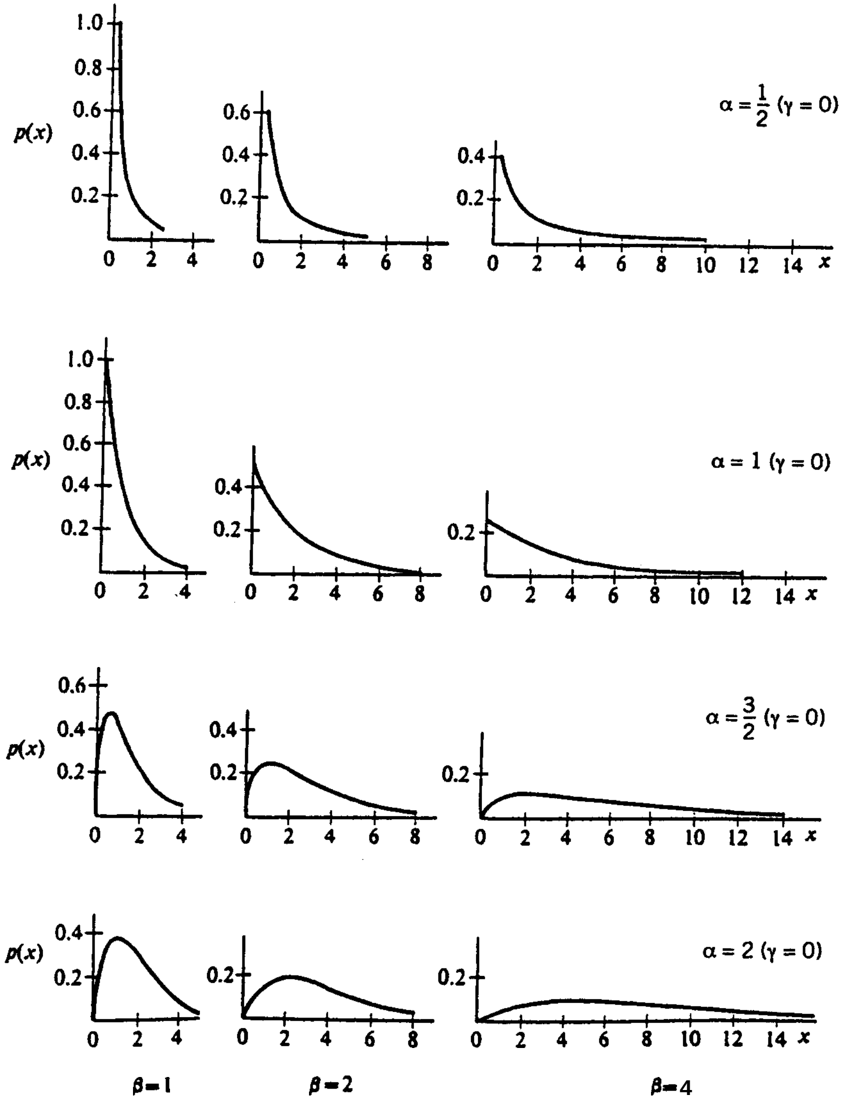

Figure 17.1 Gamma Density Functions

For distribution (17.2) Gini's concentration ratio is

$$G = \frac{\Gamma(\alpha + \frac{1}{2})}{\sqrt{\pi}\,\Gamma(\alpha + 1)}. \tag{17.19a}$$

The Lorenz concentration ratio is

$$L = 2B_{0.5}(\alpha, a + 1), \tag{17.19b}$$

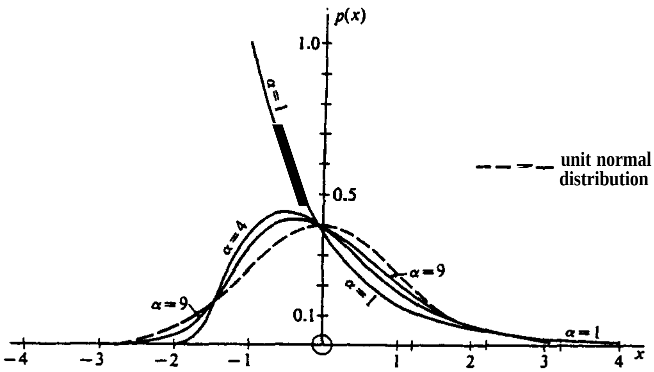

Figure 17.2 *Standardized Type III Density Functions*

$\alpha = 1$;	$p(x) = \exp[-(x+1)]$	$(x > -1)$	*Mode at* -1
$\alpha = 4$;	$p(x) = (8/3)(x+2)^3 \exp[-2(x+2)]$	$(x > -2)$	*Mode at* $-1/2$
$\alpha = 9$;	$p(x) = (2187/4480)(x+3)^8 \exp[-3(x+3)]$	$(x > -3)$	Mode at $-1/3$

where $B_p(a,b) = \int_0^p y^{a-1}(1-y)^{b-1}\,dy$ is the incomplete beta function. The Pietra ratio is

$$\rho = \frac{\alpha^{\alpha} e^{-\alpha}}{\Gamma(\alpha+2)}\,{}_1F_1(2, \alpha+2; \alpha). \tag{17.20}$$

The Theil entropy measure of inequality is

$$T = \frac{1}{\alpha} + \Psi(\alpha) - \log \alpha \tag{17.21}$$

$[\Psi(\alpha) = d \log \Gamma(\alpha)/d\alpha]$ [Salem and Mount (1974); McDonald and Jensen (1979)]. Saunders and Moran (1978) show that the ε-quantile of the gamma distribution (17.1), denoted by $y_{\varepsilon|\alpha}$, where

$$\int_0^{y_{\varepsilon|\alpha}} e^{-y} y^{\alpha-1}\,dy/\Gamma(\alpha) = \varepsilon \tag{17.22}$$

has the property that when $1 > \varepsilon_2 > \varepsilon_1 > 0$, the ratio $\gamma_\alpha = y_{\varepsilon_2|\alpha}/y_{\varepsilon_1|\alpha}$ is a decreasing function of α and thus the equation

$$\gamma_\alpha = b$$

has a unique solution, $\phi(b)$ for any $b \in (1,\infty)$. Moreover $\gamma_\alpha \to 1$ as $\alpha \to \infty$.

The reproductive property is utilized, among other things, for determination of gamma priors in Bayesian reliability analysis [see Waller and Waterman (1978)]. Also $y_{\varepsilon_2|\alpha} - y_{\varepsilon_1|\alpha}$ increases with α, implying that the gamma distributions are "ordered in dispersion" (see Chapter 33).

3 GENESIS AND APPLICATIONS

Lancaster (1966) quotes from Laplace (1836) in which the latter obtains a gamma distribution as the posterior distribution of the "precision constant" $(h = \frac{1}{2}\sigma^{-2})$, (Chapter 13, Section 1) given the values of n independent normal variables with zero mean and standard deviation σ (assuming a "uniform" prior distribution for h). Lancaster (1966) also states that Bienaymé (1838) obtained the (continuous) χ^2 distribution as the limiting distribution of the (discrete) random variable $\sum_{i=1}^{k}(N_i - np_i)^2(np_i)^{-1}$, where $(N_1, \ldots, N_k)$ have a joint multinomial distribution with parameters $n, p_1, p_2, \ldots, p_k$.

The gamma distribution appears naturally in the theory associated with normally distributed random variables, as the distribution of the sum of squares of independent unit normal variables. (See Chapter 18.) The use of the gamma distribution to approximate the distribution of quadratic forms (particularly positive definite quadratic forms) in multinormally distributed variables is well established and widespread. One of the earliest examples was its use, in 1938, to approximate the distribution of the denominator in a test criterion for difference between expected values of two normal populations with possibly different variances [Welch (1938)]. It has been used in this way many times since. The use of gamma distributions to represent distributions of range and quasi-ranges in random samples for a normal population has been discussed in Chapter 13. In most applications the two-parameter form $(\gamma = 0)$,

$$p_X(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, \qquad x > 0; \alpha > 0, \beta > 0, \tag{17.23}$$

is used (this is equivalent to approximating by the distribution of $\frac{1}{2}\beta\chi^2_{2\alpha}$). However, the three-parameter form has also been used with good effect [e.g., see Pearson (1963)].

The gamma distribution may be used in place of the normal distribution as "parent" distribution in expansions of Gram-Charlier type (Chapter 12, Section 4.2). Series with Laguerre polynomial multipliers rather than Hermite polynomial multipliers are obtained in this situation. Formulas for use with such expansions and their properties have been described by Khamis (1960). These *Laguerre series* have been used by Barton (1953) and Tiku (1964a, b) to approximate the distributions of "smooth test" (for goodness-of-fit) statistics and noncentral F (Chapter 30).

In applied work, gamma distributions give useful representations of many physical situations. They have been used to make realistic adjustments to exponential distributions in representing lifetimes. The "reproductive" property (mentioned in Section 1) leads to the appearance of gamma distributions in the theory of random counters and other topics associated with random processes in time, in particular in meteorological precipitation processes [Kotz and Neumann (1963); Das (1955)]. Some other applications from diverse fields are described in papers cited in references. Among the latter papers Salem and Mount (1974) provide a comparison of the gamma and lognormal distributions and demonstrate that the gamma distributions provides a better fit for personal income data in the United States for the years 1960 to 1969.

Dennis and Patil (1984) discuss applications of gamma distribution in statistical ecology (standard model for sampling dispersion). Among more recent applications we note Costantino and Desharnais's (1981) empirical fit of the steady-state abundances of laboratory flour beetle (Tribolium) populations. Dennis and Patil (1984) generalize this result and show that a gamma distribution is an approximate stationary distribution for the abundance of a population fluctuating around a stable equilibrium. Starting from the stochastic model of population growth

$$\frac{dn}{dt} = n[g(u) + h(n)z(t)],$$

where n is population density at time t, $g(u)$ is the specific growth rate, $z(t)$ is a Gaussian process (noise) with variability a^2, and $h(n)$ is a function specifying the density dependence on the effects of the noise, Dennis and Patil (1984) approximate Wright's formula for the equilibrium pdf by a gamma distribution emphasizing its right-skewness in the distribution of single-species abundances at equilibrium and its positive range. Then, modifying their deterministic model from $du/dt = ng(u)$ to $du/dt = n[g(u) - p(u)]$, where $p(u)$ is a specific rate describing the effects of predation, harvesting and other forces, Dennis and Patil (1984) arrive at a weighted gamma distribution for the equilibrium pdf.

Gamma distributions share with lognormal distributions (Chapter 14) the ability to mimic closely a normal distribution (by choosing α large enough) while representing an essentially positive random variable (by choosing $\gamma \geq 0$).

4 TABLES AND COMPUTATIONAL ALGORITHMS

In 1922 there appeared a comprehensive Tables of the Incomplete *Γ-Function*, edited by Pearson (1922). This contains values of $I(u,p)$ [see (17.5) with $\alpha = p + 1$] to seven decimal places for p = $-1(0.05)0(0.1)5(0.2)50$ and u at intervals of 0.1. These are supplemented by a table of values of

$$\log I(u,p) - (p+1)\log u$$

for p = $-1(0.05)0(0.1)10$ and $u = 0.1(0.1)1.5$. This function was chosen to make interpolation easier, particularly for low values of p (Section 5).

Harter (1964) published tables of $I(u, p)$ to nine decimal places for $p = -0.5(0.5)74(1)164$ and u at intervals of 0.1. In this work he covers a greater range of values of p and has two extra decimal places, although Pearson (1922) has p at finer intervals. Harter (1969) published further tables of Type III distributions, giving the 0.01, 0.05, 0.1, 0.5, 1, 2, 2.5, 4, 5, 10(10)90, 95, 96, 97.5, 98, 99, 99.5, 99.9, 99.95, and 99.99 percentage points to 5 decimal places for $\sqrt{\beta_1} = 0.0(0.1)4.8(0.2)9.0$. Harter (1971) extends his 1969 tables and provides percentage points of the one-parameter gamma distribution (Pearson Type III) corresponding to cumulative probabilities 0.002 and 0.998 as well as 0.429624 and 0.570376. The first pair is used for determination of the magnitude of an event (flood) corresponding to 500-year return period, while the second corresponds to a return period of 32,762 years (the so-called mean annual flood according to the guidelines of the U.S. Department of Housing and Urban Development). These are the most important direct tables of $I(u, p)$.

Pearson (1922) gave the more general formula for distribution (17.1):

$$\Pr[X \le x] = e^{-y} \sum_{j=0}^{\infty} \left\{ \frac{y^{\alpha+j}}{\Gamma(\alpha + j + 1)} \right\} \quad \text{for } y = (x - \gamma)/\beta > 0. \quad (17.24)$$

Salvosa (1929, 1930) published tables of the probability integral, probability density function and its first six derivatives for distribution (17.1), with β and γ so chosen that X is standardized (i.e., $\beta = \alpha_3/2$; $\gamma = -2/\alpha_3$). Values are given to six decimal places for a, $(= 2\alpha^{-1/2}) = 0.1(0.1)1.1$ at intervals of 0.01 for x. Cohen, Helm, and Sugg (1969) have calculated tables of the probability integral to nine decimal places for a, = 0.1(0.1)2.0(0.2)3.0(0.5)6.0 at intervals of 0.01 for x. Bobée and Morin (1973) provide tables of percentage points of order statistics for gamma distributions.

Thom (1968) has given tables to four decimal places of the distribution function $\Gamma_x(\alpha)/\Gamma(\alpha)$:

1. for a = 0.5(0.5)15.0(1)36 and x = 0.0001, 0.001, 0.004(0.002)0.020(0.02)0.80(0.1)2.0(0.2)3.0(0.5)—the tabulation is continued for increasing x until the value of the tabulated function exceeds 0.9900;
2. values of x satisfying the equation

$$\frac{\Gamma_x(\alpha)}{\Gamma(\alpha)} = \varepsilon$$

for a = 0.5(0.5)15.0(1)36 and $\varepsilon = 0.01, 0.05(0.05)0.95, 0.99$. Burgin (1975) provides some interesting numerical computations of the gamma and associated functions. He cites the following representation for the incomplete gamma function ratio:

$$I(u,p) = \frac{e^{-u}}{\Gamma(p+1)} u^p \sum_{n=0}^{\infty} \left(\frac{u^{n+1}}{\prod_{i=1}^{n}(p+i+1)} \right). \qquad (17.25)$$

Lau (1980) uses a series expansion similar to Wilk, Gnanadesikan, and Huyett (1962a) to calculate the incomplete gamma function ratio. Moore (1982) finds this series expansion not always satisfactory. Bhattacharjee (1970) provides an alternative algorithm. Phien (1991) presents an algorithm for computing the quantile $y_{\varepsilon|\alpha}$. Newton's method seems to be appropriate to solving the equation

$$I(y_{\varepsilon|\alpha}, \alpha + 1) = \varepsilon.$$

Efficient computation requires a powerful algorithm for computing the incomplete gamma function ratio and a good initial value y_0. Phien used Moore's algorithm (1982) for calculation of the incomplete gamma function ratio and the approximate value provided by Hoshi and Burges (1981) is used for the initial value y_0.

Moore's algorithm is highly accurate. In the experiment conducted by Phien (1991) Moore algorithm's value agreed with those values tabulated by E. S. Pearson (1963) up to the sixth decimal place.

5 APPROXIMATION AND GENERATION OF GAMMA RANDOM VARIABLES

The best-known approximations for probability integrals of gamma distributions have been developed in connection with the χ^2 distribution. Modifications to apply to general gamma distributions are direct, using the linear transformation $y = 2(x - \gamma)/\beta$. The reader should consult Section 5 of Chapter 18.

From (17.25) it may be observed that for u small, $u^{-(p+1)}I(u,p)$ is very approximately, a linear function of u. It is for this reason that the values

$$\log I(u,p) - (p+1)\log u$$

tabulated by Pearson (1922) lead to relatively easy interpolation. Gray, Thompson, and McWilliams (1969) have obtained the relatively simple approximation:

$$\frac{1}{x^{\alpha-1}e^{-x}}\int_x^{\infty} t^{\alpha-1}e^{-t}\,dt \doteqdot \frac{x}{x-\alpha+1}\left[1-\frac{\alpha-1}{(x-\alpha+1)^2+2x}\right], \quad (17.26)$$

which gives good results when x is sufficiently large.

If Y has the standard uniform distribution (Chapter 26, Section 1)

$$p_Y(y) = 1, \qquad 0 < y < 1, \quad (17.27)$$

then $(-2\log Y)$ is distributed as χ^2 with 2 degrees of freedom. If $Y_1, Y_2, \dots, Y_s$ each have distribution (17.27) and are independent then $\sum_{j=1}^{s}(-2\log Y_j)$ is distributed as χ^2_{2s}; that is, it has a gamma distribution with $\alpha = s$, $\beta = 2$, and $\gamma = 0$. Using this relation, it is possible to generate gamma distributed variables from tables of random numbers. Extension to cases when α is not an integer can be effected by methods of the kind described by Bánkövi (1964).

If X has distribution (17.2), then the moment generating function of $\log X$ is

$$E[e^{t\log X}] = E[X^t] = \frac{\Gamma(\alpha+t)}{\Gamma(\alpha)}. \quad (17.28)$$

Hence the rth cumulant of $\log X$ is

$$\kappa_r(\log X) = \psi^{(r-1)}(\alpha). \quad (17.29)$$

Note that for α large

$$\begin{aligned}\beta_1(\log X) &\doteqdot \alpha^{-1},\\ \beta_2(\log X) &\doteqdot 3 + 2\alpha^{-1},\end{aligned} \quad (17.30)$$

which may be compared with

$$\begin{aligned}\beta_1(X) &= 4\alpha^{-1},\\ \beta_2(X) &= 3 + 6\alpha^{-1}.\end{aligned}$$

The distribution of $\log X$ is more nearly normal than the distribution of X. Although this approximation is not generally used directly, it is often very useful when approximating the distributions of functions of independent gamma variables.

For example, suppose that $X_1, X_2, \dots, X_k$ are independent variables, each distributed as χ^2 with ν degrees of freedom. Then the distribution of

$$R_k = \frac{\max(X_1, X_2, \dots, X_k)}{\min(X_1, X_2, \dots, X_k)} \quad (17.31)$$

may be approximated by noting that

$$\log R_k = \max(\log X_1, \dots, \log X_k) - \min(\log X_1, \dots, \log X_k)$$

is approximately distributed as the range of k independent normal variables each having the same expected value, and standard deviation

$$\sqrt{\psi^{(1)}\left(\frac{\nu}{2}\right)} \doteqdot \sqrt{\frac{2}{\nu - 1}} \,. \tag{17.32}$$

[See also (17.107).]

Gray and Schucany (1968) and Gray and Lewis (1971) apply the method of H- and B_n-transforms (see Chapter 12, Section 4) to approximate the tail probabilities of chi-squared (and hence of gamma) distributions. Alfers and Dinges (1984) sought an approximate normalizing transformation. They obtained a polynomial expression.

A recent survey of approximations for gamma distribution quantiles is contained in a paper by Phien (1991). He notes that for a variable X with pdf given by (17.2), the standardized variable is $W = (X - \alpha)/\sqrt{\alpha}$ (for which $E[W] = 0$, $\operatorname{Var}(W) = 1$, and $\sqrt{\beta_1(W)} = 2/\sqrt{\alpha}$). The Wilson-Hilferty chi-squared approximation (see Chapter 18, Section 5) gives

$$W_\varepsilon \doteqdot \sqrt{\alpha}\left[\left\{1 - \tfrac{1}{9}\alpha^{-1} + \tfrac{1}{3}\alpha^{-1/2}z_\varepsilon\right\}^3 - 1\right], \tag{17.33}$$

where $\Phi(z_\varepsilon) = \varepsilon$. When α is small (skewness is large), the values of W_ε given by (17.33) are too low.

Kirby (1972) modifies (17.33) and obtains the following approximation:

$$W_\varepsilon \doteqdot A(U - B), \tag{17.34}$$

where $A = \max(\sqrt{\alpha}, 0.40)$, $B = 1 + 0.0144\,[\max(0, 2\alpha^{-1/2} - 2.25)]^2$, and

$$U = \max\left[B - \sqrt{\alpha}/A, \left\{1 - \left(\tfrac{1}{6}D\right)^2 + \left(\tfrac{1}{6}D\right)z_\varepsilon\right\}^3\right]$$

with

$$D = 2\alpha^{-1/2} - 0.063\left\{\max(0, 2\alpha^{-1/2} - 1)\right\}^{1.85}.$$

Kirby (1972) provides tables to assist in the calculation of A, B, and D. Kirby's approximation was in turn modified by Hoshi and Burges (1981) who express A, B, $B - A\sqrt{\alpha}$, and D as polynomials of degree 5 in $\alpha^{-1/2}$. Phien (1991) reproduces the values given by Hoshi and Burges (1981). Harter (1969) provided tables of exact values of W_ε, to which Bobée (1979) fitted polynomials in $\alpha^{-1/2}$ of degree four. For $\sqrt{\beta_1} = 2/\sqrt{\alpha} < 4$ (i.e., $\alpha > \frac{1}{4}$) Bobée's approximation is superior to those of Hoshi and Burges (1981) and Kirby (1979), but it is not satisfactory for smaller values of a.

Tadikamalla and Ramberg (1975), Wheeler (1975), and Tadikamalla (1977) approximate the gamma distribution (17.2) by a four-parameter Burr distribution (see Chapter 12, Section 4.5) by equating the first four moments—expected value, variance, skewness, and kurtosis.

For the Burr distribution with cdf,

$$F_X(x) = 1 - \left\{1 + \left(\frac{x-a}{b}\right)^c\right\}^{-k},$$

the $100p\%$ points of the largest and smallest order statistics from a random sample of size n are

$$a + b\left\{(1-p)^{-1/(nk)} - 1\right\}^{1/c}, \tag{17.35a}$$

$$a + b\left\{(1-p^{1/n})^{-1/k} - 1\right\}^{1/c}, \tag{17.35b}$$

respectively. These equations give good approximations to the corresponding value for the appropriate gamma distributions. Tables are available from Tadikamalla (1977) to facilitate calculation of appropriate values for a, b, k, and c.

Values of c and k for given values of a are given in Tadikamalla and Ramberg (1975) and Wheeler (1975). These two papers were published in the same issue of the same journal, and yet they do not cross-reference each other. (Fortunately the values of c and k given in the two papers agree.)

Many papers on generation of gamma random variables have been written in the years 1964 to 1990. It is impossible to survey them in detail. We draw attention to Ahrens and Dieter (1974, "GO algorithm"), Fishman (1976), Johnk (1964), Odell and Newman (1972), Wallace (1974), and Whittaker (1974)—all are based on the general von Neumann rejection method. Cheng and Feast (1979, 1980) use the ratio of uniform random variables on the lines suggested by Kinderman and Monahan (1977).

Bowman and Beauchamp (1975) warn of pitfalls with some gamma distribution simulation routines. They note that an algorithm given by Phillips and Beightler (1972) does not actually generate random variables with gamma distributions, but rather with Weibull distributions (see Chapter 21).

6 CHARACTERIZATIONS

If X_1 and X_2 are independent standard gamma random variables [i.e., having distributions of form (17.2), possibly with different values of a; α_1, a,, say], then the random variables

$$X_1 + X_2 \quad \text{and} \quad \frac{X_1}{X_1 + X_2}$$

are mutually independent. [Their distributions are, respectively, a standard gamma with $\alpha = \alpha_1 + \alpha_2$ and a standard beta (Chapter 25) with parameters α_1, α_2.]

Lukacs (1965) showed that this property characterizes the gamma distribution in that, if X_1 and X_2 are independent positive random variables, and $X_1 + X_2$ and $X_1/(X_1 + X_2)$ are mutually independent, then X_1 and X_2 must each have gamma distributions of form (17.1) with $\gamma = 0$, common β, but possibly different values of α. If it be assumed that X_1 and X_2 have finite second moments and identical distributions, it is sufficient to require that the regression function

$$E\left[\frac{a_{11}X_1^2 + 2a_{12}X_1X_2 + a_{22}X_2^2}{(X_1 + X_2)^2}\,\middle|\, X_1 + X_2\right], \qquad a_{11} + a_{22} \neq 2a_{12},$$

be independent of $X_1 + X_2$ to ensure that the common distribution is a gamma distribution [Laha (1964)].

Marsaglia (1974) extended Lukacs's result by removing the condition that X_1 and X_2 should be positive. He shows that "If X_1 and X_2 are independent nondegenerate random variables, then $X_1 + X_2$ is independent of X_1/X_2 if and only if there is a constant c such that cX_1 and cX_2 have standard gamma distributions." Marsaglia (1989) provides a simpler proof of this result. He uses a method of deriving Lukacs's (1965) result, which was developed by Findeisen (1978), without use of characteristic functions (although a disclaimer, suggested by the referees of the Findeison paper, claims that characteristic functions are implicit in the argument). Marsaglia (1989) also remarks that the "$X_1 + X_2, X_1/X_2$" characterization has been used in developing computer methods for generating random points on surfaces by projections of points with independent (not necessarily positive) components.

Earlier Marsaglia (1974) had obtained the following result "Let $X_1, X_2, \ldots, X_n$ $(n \geq 2)$ be independent random variables. Then the vector

$$\left(\frac{X_1}{S_n}, \frac{X_2}{S_n}, \ldots, \frac{X_n}{S_n}\right),$$

where $S_n = \sum_{j=1}^n X_j$, is independent of S_n if and only if there is constant c such that $cX_1, cX_2, \ldots, cX_n$ are gamma." Wang and Chang (1977) used this result to develop several sensitive nonparametric tests of exponentiality. Many multivariate generalizations of these results are surveyed in Wang (1981).

On the one hand, the distribution of X_1/X_2 is not sufficient to establish that each $|X_j|$ has a gamma distribution. If X_j is distributed as standard gamma $(a,)$, $(j = 1, 2)$ and X_1 and X_2 are mutually independent then the probability density function of $G = X_1/X_2$ is

$$p_G(g) = [B(\alpha_1, \alpha_2)]^{-1} g^{\alpha_1 - 1}(1 + g)^{-(\alpha_1+\alpha_2)}, \qquad g > 0, \quad (17.36)$$

which is a Pearson Type VI distribution (see Chapter 12 and also Chapter 25). However, it is possible for X_1 and X_2 to be independent, identically distributed positive random variables, and for $G = X_1/X_2$ to have distribution (17.36), without each X_j having a gamma distribution [Laha (1954); Mauldon (1956); Kotlarski (1962, 1965)]. However, Kotlarski (1967) showed that the joint distribution of ratios $X_1/X_3, X_2/X_3$ (in a similar situation) does characterize the distribution (up to a constant multiplier).

It follows that any result depending only on the distribution (17.36) of the ratio X_1/X_2 cannot characterize the distribution of each X_j. In particular, it can be shown that if X_1 and X_2 are independent and identically distributed as in (17.2), then

$$\sqrt{\frac{a}{2}}\left(\sqrt{\frac{X_1}{X_2}} - \sqrt{\frac{X_2}{X_1}}\right)$$

has a t_{2a} distribution (as defined in Chapter 28), although this property is not sufficient to establish the form of the common distribution of X_1 and X_2 (given they are positive, independent, and identically distributed).

However, if X_3 is a third random variable (with the same properties relative to X_1 and X_2), then the joint distribution of

$$\sqrt{\frac{a}{2}}\left(\sqrt{\frac{X_1}{X_2}} - \sqrt{\frac{X_2}{X_1}}\right) \quad \text{and} \quad \sqrt{\frac{a}{2}}\left(\sqrt{\frac{X_1}{X_3}} - \sqrt{\frac{X_3}{X_1}}\right)$$

is sufficient to establish that common distribution is a gamma distribution with $\gamma = 0$. [Kotlarski (1967).]

Khatri and Rao (1968) have obtained the following characterizations of the gamma distribution, based on constancy of various regression functions:

1. If $X_1, X_2, \ldots, X_n$ $(n \geq 3)$ are independent positive random variables and

$$Y_1 = \sum_{i=1}^{n} b_{1i} X_i, \qquad b_{1i} \neq 0,\ i = 1, 2, \ldots, n,$$

$$Y_j = \prod_{i=1}^{n} X_i^{b_{ji}}, \qquad j = 2, \ldots, n,$$

with the $(n-1) \times n$ matrix (b_{ji}) $(j = 2, \ldots, n,\ i = 1, 2, \ldots, n)$ nonsingular, then the constancy of

$$E[Y_1 | Y_2, \ldots, Y_n]$$

ensures that the X's must have a common gamma distribution (unless they have zero variances).

Setting $b_{11} = b_{12} = \cdots = b_{1n} = 1$ and $b_{j,j-1} = -1$, $b_{j,j} = 1$, with all other b's zero, the condition becomes the constancy of

$$E\left[\sum_{j=1}^{n} X_j \middle| \frac{X_2}{X_1}, \frac{X_3}{X_1}, \ldots, \frac{X_n}{X_1}\right].$$

2. In the conditions of 1, if $E[X_j^{-1}] \neq 0$ $(j = 1, 2, \ldots, n)$ and

$$Z_1 = \sum_{i=1}^{n} b_{1i} X_i^{-1},$$

$$Z_j = \sum_{i=1}^{n} b_{ji} X_i, \qquad j = 2, \ldots, n,$$

with the b's satisfying the same conditions as in 1, then the constancy of

$$E[Z_1 | Z_2, Z_3, \ldots, Z_n]$$

ensures that each X_j has a gamma distribution (not necessarily the same for all j), unless they have zero variances. Choosing special values of b's as in 1, we obtain the condition that $E[\sum_{j=1}^{n} X_j^{-1} | X_2 - X_1, \ldots, X_n - X_1]$ should be constant.

3. Under the same conditions as in 1, if $E[X_1 \log X_1]$ is finite, then the constancy of

$$E\left[\sum_{j=1}^{n} a_j X_j \middle| \prod_{i=1}^{n} X_i^{b_i}\right],$$

with $\sum_{j=1}^{n} a_j b_j = 0$, $|b_n| > \max(|b_1|, |b_2|, \ldots, |b_{n-1}|)$, and $a_j b_j / a_n b_n < 0$ for all $j = 1, 2, \ldots, n-1$ ensures that X_1 has a gamma distribution (unless it has zero variance).

As a special case, setting $a_1 = a_2 = \cdots = a_n = 1$, $b_n = n-1$, $b_1 = b_2 = \cdots = b_{n-1} = -1$, we obtain the condition as the constancy of

$$E\left[\sum_{j=1}^{n} X_j \middle| X_n^{n-1} \left\{\prod_{j=1}^{n-1} X_j\right\}^{-1}\right].$$

4. If $X_1, \ldots, X_n$ are independent, positive, and identically distributed random variables, and if $E[X_i^{-1}] \neq 0$ $(i = 1, 2, \ldots, n)$ and

$$E\left[\sum_{j=1}^{n} a_j X_j^{-1} \middle| \sum_{j=1}^{n} b_j X_j\right]$$

is constant with the same conditions on the a's and *b*'s as in 3, the common distribution of the X's is a gamma distribution (unless it has a zero variance).

Giving the a's and *b*'s the same special values as in 3, we obtain the condition of constancy of

$$E\left[\sum_{j=1}^{n} X_j^{-1} \middle| X_n - \bar{X}\right],$$

where $\bar{X} = n^{-1}\sum_{j=1}^{n} X_j$.

Khatri and Rao (1968) have also obtained a number of further conditions characterizing gamma distributions. Hall and Simons (1969) have shown that if X and Y are mutually independent and nondegenerate, and if

$$E\left[X^2(X+Y)^{-2} | X+Y\right] \quad \text{and} \quad E\left[Y^2(X+Y)^{-2} | X+Y\right]$$

do not depend on $X + Y$, then either X and Y or $-X$ and $-Y$ have two-parameter gamma distributions with a common value of the scale parameter β.

The following characterization based on conditional expectation was suggested by Wang (1972): "If $\beta = (b_{jk})$ be a $n \times n$ real matrix satisfying

$$\sum_{j=1}^{n} b_{jj} = 1 \quad \text{and} \quad \sum_{j,k} b_{jk} = c \neq 0,$$

$X = (X_1, X_2, \ldots, X_n)'$ is a $n \times 1$ random vector, $Q^* = \mathbf{X}'\boldsymbol{\beta}\mathbf{X}$, and $L = C'X$, where $C = (c_1, c_2, \ldots, c_n)'$ is an $n \times 1$ real vector such that $\sum_{j=1}^{n} c_j = 1$ and $c_j = c_k \neq 0$ for some j, k, then provided the distribution F of X_i $(i = 1, 2, \ldots, n)$ is nondegenerate, the conditional expectation $E[Q^*|L]$ is zero almost everywhere if and only if each X_i has a gamma distribution." If N is a random variable defined by

$$N = 0 \quad \text{if } X_1 > x$$

and

$$N = n \quad \text{if } X_1 + \cdots + X_n \leq x < X_1 + \cdots + X_{n+1},$$

where $X_1, X_2, \ldots, X_n$ are independent random variables each having distribution (17.1) with α an integer and with $\gamma = 0$ (**Erlang** distribution), then N has the generalized Poisson distribution (Chapter 9, Section 3)

$$\Pr[N \leq n] = e^{-x/\beta} \sum_{j=0}^{(n+1)\alpha - 1} \left\{ \frac{(x/\beta)^j}{j!} \right\}.$$

Nabeya (1950) showed that for $\alpha = 1$, the converse is true. That is to say, it is a characterization of the common distribution of $X_1, \ldots, X_n$ (exponential in this case) given that they are positive, independent, identically distributed, and continuous. Goodman (1952) extended the result to apply to any positive integer value of $\alpha \geq 2$, thus providing a characterization of a gamma distribution.

Under conditions (4), Linnik, Rukhin, and Stzelic (1970) characterize the gamma distribution by the property

$$E\left\{ P_k\left(\frac{X_1}{S_n}, \ldots, \frac{X_n}{S_n} \right) \middle| S_n \right\} \text{ does not depend on } S_n = \sum_{j=1}^{n} X_j,$$

where P_k is a polynomial of degree k, $n \geq 2$, and $n > k$. Some conditions on the behavior of the cdf of the positive i.i.d. X_i and its derivative in an interval $[0, \varepsilon]$ were required to prove the validity of this result.

If $X_1, X_2, \ldots, X_n$ are independent gamma variables with the same scale parameter β_j, and $S(X_1, X_2, \ldots, X_n)$ is a statistic invariant under the scale transformation $X \to cX$ (for all $c \neq 0$), then $U = \sum_{j=1}^{n} X_j$ and $S = S(X_1, X_2, \ldots, X_n)$, are independent [Wang (1981)]. However, whether the gamma distribution can be characterized by the independence of U and S seems still to be an open problem.

The following characterization has been proved by Wang (1981): Let $X_1, \ldots, X_n$ $(n > 2)$ be nondegenerate i.i.d. positive random variables and I_1 and I_2 be arbitrary nonempty subsets of $\{1, 2, \ldots, n\}$ of size $k \leq n/2$. Define $T_1 = \prod_{j \in I_1} X_j$ and $T_2 = \prod_{j \in I_2} X_j$. If

$$E[T_1^2 | U] = \theta E(T_1 + T_2)^2,$$

then the X_i's have a (two-parameter) gamma distribution (17.23).

Characterization of gamma distributions by the negative binomial has been provided by Engel and Zijlstra (1980):

> Let events A occur independently and at random so that the number of events occurring in a given interval is a Poisson random variable with rate θ. The waiting time between events has a negative exponential distribution with mean θ^{-1}, and the total time T between r + 1 events has a gamma (r, θ^{-1}) distribution.

Consider now a second process independent of the first in which events B occur at an average rate β. If we start from a specified instant and count the number N_B of B events occurring before the rth A event, the distribution of N_B is negative binomial with parameters r and $p = \alpha/(\alpha + \beta)$.

Engel and Zijlstra (1980) have shown that T has a gamma (r, α) distribution if and only if N_B has a negative binomial distribution.

Letac (1985) characterizes the gamma distributions as follows: Given two positive independent random variables X and Y, if the distribution of Y is defined by its moments

$$E[Y^s] = \left(1 + \frac{s}{a}\right)^{a+s} \quad \text{for } s > 0$$

and a fixed given positive a, then $X\exp\{(-X/\alpha)Y\}$ and X have the same distribution if and only if the distribution of X is gamma with shape parameter a [as in (17.2)]. Motivation for this characterization is that $(UV)^{uv}$ and U^u have the same distribution provided U and V are independent uniform [0, 1] variables.

She (1988) has proved several theorems dealing with characterizations of the gamma distribution based on regression properties; see also Wesolowski (1990) for a characterization result based on constant regression, and Yeo and Milne (1991) for a characterization based on a mixture-type distribution.

7 ESTIMATION

Estimation of parameters of gamma distribution has also received extensive attention in the literature in the last two decades. The contributions of Bowman and Shenton and of A. C. Cohen and his coworkers should be particularly mentioned. Bowman and Shenton's (1988) monograph provides detailed analysis of maximum likelihood estimators for two-parameter gamma distributions with emphasis on the shape parameter and presents valuable information on distributions and moments of these estimators, including their joint distributions. Careful discussion of estimation problems associated with the three-parameter gamma density is also presented. The major emphasis of the monograph is on the distribution of sample standard deviation, skewness and kurtosis in random samples from a Gaussian density. The authors also deal with the moments of the moment estimators. The list of references in the monograph covers the development of the authors' work on this area from 1968 onwards.

A. C. Cohen's contributions to estimation of parameters of the gamma distribution are covered in the monographs by Cohen and Whitten (1988) and Balakrishnan and Cohen (1991) with an emphasis on modified moment estimators and censored samples. As in Chapters 14 and 15 we will concentrate on results not available in monographic literature and especially those discussed in less easily available sources.

7.1 Three Parameters Unknown

We will first consider estimation for the three parameter distribution (17.1), although in many cases it is possible to assume γ is zero, and estimate only a and β in (17.1). Given values of n independent random variables $X_1, X_2, \dots, X_n$, each distributed as in (17.1), the equations satisfied by the maximum likelihood estimators $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ of a, β, and y, respectively, are

$$\sum_{j=1}^{n} \log(X_j - \hat{\gamma}) - n \log \hat{\beta} - n\psi(\hat{\alpha}) = 0, \qquad (17.37a)$$

$$\sum_{j=1}^{n} (X_j - \hat{\gamma}) - n\hat{\alpha}\hat{\beta} = 0, \qquad (17.37b)$$

$$-\sum_{j=1}^{n} (X_j - \hat{\gamma})^{-1} + n\{\hat{\beta}(\hat{\alpha} - 1)\}^{-1} = 0. \qquad (17.37c)$$

From (17.37c) it can be seen that if $\hat{\alpha}$ is less than 1, then some X_j's must be less than $\hat{\gamma}$. This is anomalous, since for $x < \gamma$ the probability density function (17.1) is zero. It is also clear that equations (17.37) will give rather unstable results if $\hat{\alpha}$ is near to 1, even though it exceeds 1. It is best, therefore, not to use these equations unless it is expected that $\hat{\alpha}$ is at least 2.5, say.

It is possible to solve equations (17.37) by iterative methods. A convenient (but not the only) method is to use (17.37a) to determine a new value for $\hat{\beta}$, given $\hat{\alpha}$ and $\hat{\gamma}$. Then (17.37b) for a new $\hat{\gamma}$, given $\hat{\alpha}$ and $\hat{\beta}$, and (17.37c) for a new 2, given $\hat{\beta}$ and $\hat{\gamma}$.

The asymptotic variance-covariance matrix of 6 6 , $\sqrt{n}\,\hat{\beta}$, and $\sqrt{n}\,\hat{\gamma}$ is the inverse of the matrix

$$\begin{pmatrix} \psi'(\alpha) & \beta^{-1} & \beta^{-1}(\alpha - 1)^{-1} \\ \beta^{-1} & \alpha\beta^{-2} & \beta^{-2} \\ \beta^{-1}(\alpha - 1)^{-1} & \beta^{-2} & \beta^{-2}(\alpha - 2)^{-1} \end{pmatrix}.$$

The determinant of this matrix is

$$\beta^4 \left[\frac{2\psi'(\alpha)}{\alpha - 2} - \frac{2\alpha - 3}{(\alpha - 1)^2(\alpha - 2)} \right].$$

Hence

$$\operatorname{Var}(\hat{\alpha}) \doteqdot 2n^{-1}\left[2\psi'(\alpha) - (2\alpha - 3)(\alpha - 1)^{-2}\right]^{-1}, \tag{17.38a}$$

$$\operatorname{Var}(\hat{\beta}) \doteqdot n^{-1}\beta^2\left[(\alpha - 1)^2\psi'(\alpha) - \alpha + 2\right]\left[2(\alpha - 1)^2\psi'(\alpha) - 2\alpha + 3\right]^{-1}, \tag{17.38b}$$

$$\operatorname{Var}(\hat{\gamma}) \doteqdot n^{-1}\beta^2(\alpha - 2)\{\alpha\psi'(\alpha) - 1\}\left[2\psi'(\alpha) - (2\alpha - 3)(\alpha - 1)^{-2}\right]^{-1}. \tag{17.38c}$$

Using the approximation

$$\psi'(\alpha) \doteqdot \alpha^{-1} + \tfrac{1}{2}\alpha^{-2} + \tfrac{1}{6}\alpha^{-3}, \tag{17.39}$$

we obtain the simple formulas,

$$\operatorname{Var}(\hat{\alpha}) \doteqdot 6n^{-1}\alpha^3, \tag{17.38a$'$}$$

$$\operatorname{Var}(\hat{\beta}) \doteqdot 3n^{-1}\beta^2\alpha, \tag{17.38b$'$}$$

$$\operatorname{Var}(\hat{\gamma}) \doteqdot \tfrac{3}{2}n^{-1}\beta^2\alpha^3, \tag{17.38c$'$}$$

giving the orders of magnitude of the variances when a is large. Fisher *(1922)* obtained the more precise approximation:

$$\operatorname{Var}(\hat{\alpha}) = 6n^{-1}\left[(\alpha - 1)^3 + \tfrac{1}{5}(\alpha - 1)\right] \tag{17.40}$$

by using more terms in the expansion *(17.39).*

If the method of moments is used to estimate a, β, and y, the following simple formulas are obtained:

$$\tilde{\gamma} + \tilde{\alpha}\tilde{\beta} = \bar{X}, \tag{17.41a}$$

$$\tilde{\alpha}\tilde{\beta}^2 = m,, \tag{17.41b}$$

$$2\tilde{\alpha}\tilde{\beta}^3 = m_3, \tag{17.41c}$$

where

$$\bar{X} = n^{-1}\sum_{j=1}^{n} X_j,$$

$$m_2 = n^{-1}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2,$$

$$m_3 = n^{-1}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^3$$

are the sample mean, second, and third central moments. (Since this method would be used only when n is rather large, there is no need to attempt to make the estimators unbiased; it is also not clear whether this would improve the accuracy of estimation.) Note that (17.41a) and (17.37b)′ are identical. From equations (17.41) the following formulas for the ***moment estimators*** $\tilde{\alpha}$, $\tilde{\beta}$, and $\tilde{\gamma}$ are obtained:

$$\tilde{\alpha} = \frac{4m_2^3}{m_3^2} = \frac{4}{b_1} \qquad \left(\text{where } \sqrt{b_1} = \frac{m_3}{m_2^{3/2}}\right) \tag{17.42a}$$

$$\tilde{\beta} = \frac{\frac{1}{2}m_3}{m_2}, \tag{17.42b}$$

$$\tilde{\gamma} = \bar{X} - \frac{2m_2^2}{m_3}. \tag{17.42c}$$

Although these are simple formulas, the estimators are often, unfortunately, considerably less accurate than the maximum likelihood estimators $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$.

It can be shown that if n and α are large

$$\text{Var}(\tilde{\alpha}) \doteqdot 6\alpha(\alpha + 1)(\alpha + 5)n^{-1} \tag{17.43}$$

[Fisher (1922)]. Comparing (17.40) and (17.43), it can be seen that the ratio of approximate values $\text{Var}(\hat{\alpha})/\text{Var}(\tilde{\alpha})$ is substantially less than 1 unless α is rather large. The ratio

$$\frac{(\alpha - 1)^3 + \frac{1}{5}(\alpha - 1)}{\alpha(\alpha + 1)(\alpha + 5)}$$

increases with α and reaches the value 0.8 at $\alpha = 39.1$.

On the other hand, we have already noted that when α is less than 2.5, the maximum likelihood estimators are of doubtful utility. It then becomes necessary to consider yet other methods of estimation. When α is less than 1, the distribution is shaped like a reversed J, with the probability density function tending to infinity as x tends to γ (see Figure 17.1). If n is large (as it usually is if a three-parameter distribution is being fitted), it is reasonable to estimate γ as the smallest observed value among $X_1, X_2, \ldots, X_n$, or a value slightly smaller than this. Estimation of α and β then proceeds as for the two-parameter case, to be described later. Using the value of α so estimated, a new value for γ can be estimated, and so on.

As in the case of the lognormal and inverse Gaussian distributions (Chapters 14 and 15), Cohen and Whitten (1988) advocate the use of modified moment estimators

$$E[X] = \bar{X} = \tilde{\gamma} + \tilde{\alpha}\tilde{\beta}, \tag{17.44a}$$

$$\operatorname{Var}(X) = s^2 = \tilde{\alpha}\tilde{\beta}^2, \tag{17.44b}$$

$$E[F(X'_{1:n})] = \frac{1}{n+1} = F\left(\frac{X'_{1:n} - \bar{X}}{s}; 0, 1, \tilde{\alpha}\right). \tag{17.44c}$$

Tables and graphs to facilitate solution of (17.44c) are given by Cohen and Whitten (1986, 1988), Bai, Jakeman, and Taylor (1990). Note that (17.37b) and (17.37c) can be written as

$$\hat{\alpha} = \sum_{i=1}^{n} \left(\frac{X_i - \hat{\gamma}}{n\hat{\beta}}\right), \tag{17.45a}$$

$$\hat{\beta} = \sum_{i=1}^{n} \left(\frac{X_i - \hat{\gamma}}{n}\right) - \frac{n}{\sum_{i=1}^{n}(X_i - \hat{\gamma})^{-1}}, \tag{17.45b}$$

respectively. Assuming a value for $\hat{\gamma}$, $\hat{\alpha}(\hat{\gamma})$, and $\hat{\beta}(\hat{\gamma})$ can be computed, the corresponding likelihood $L(\hat{\gamma})$ calculated. The value of $\hat{\gamma}$ maximizing $L(\hat{\gamma})$ is then found numerically. [See Bai, Jakeman, and Taylor (1990).]

Cheng and Amin (1983) applied their maximum product of spacings (MPS) estimator method (see Chapters 12, 14, and 15) to provide consistent estimators of a, β, and y. This method yields the following first-order equations:

$$\frac{\partial \log G}{\partial \gamma} \equiv \sum_{i=1}^{n+1} \frac{\int_{X'_{i-1}}^{X'_i} [-(\alpha - 1) + \beta^{-1}(x - \gamma)](x - \gamma)^{\alpha-2} e^{-(x-\gamma)/\beta}\, dx}{(n+1)\int_{X'_{i-1}}^{X'_i} (x - \gamma)^{\alpha-1} e^{-(x-\gamma)/\beta}\, dx}, \tag{17.46a}$$

$$\frac{\partial \log G}{\partial \beta} \equiv -\frac{\alpha}{\beta} + \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{\int_{X'_{i-1}}^{X'_i} (x - \gamma)^{\alpha} \beta^{-2} e^{-(x-\gamma)/\beta}\, dx}{\int_{X'_{i-1}}^{X'_i} (x - \gamma)^{\alpha-1} e^{-(x-\gamma)/\beta}\, dx} = 0, \tag{17.46b}$$

$$\frac{\partial \log G}{\partial \alpha} \equiv -\log \beta - \log \psi(\alpha)$$

$$+ \sum_{i=1}^{n+1} \frac{\int_{X'_{i-1}}^{X'_i} (x - \gamma)^{\alpha-1} e^{-(x-\gamma)/\beta} \log(x - y)\, dx}{(n+1)\int_{X'_{i-1}}^{X'_i} (x - \gamma)^{\alpha-1} e^{-(x-\gamma)/\beta}\, dx} = 0, \tag{17.46c}$$

where $X'_1 \le X'_2 \le \dots \le X'_n$ are the order statistics corresponding to $X_1, \dots, X_n$.

7.2 Some Parameters Unknown

We now consider estimation when the value of one of the three parameters a, β, and γ is known. The commonest situation is when the value of γ is known (usually it is zero). Occasionally a is known (at least approximately) but not β or γ. Inadmissibility of standard estimators of gamma parameters in the case when $\gamma = 0$ has received attention in the literature.

Let $X_1, X_2, \dots, X_k$ be independent random variables with two-parameter gamma distributions with parameters α_i, β_i $(i = 1, \dots, k)$ where the values of the α_i's are known but the β_i's (> 0) are unknown. Berger (1980) considered weighted quadratic losses $\sum_{i=1}^{k} \beta_i^{-m}(\delta_i \beta_i^{-1} - 1)^2$ for m = $0, 2, 1, -1$, and showed that the standard estimator of $(\beta_1, \beta_2, \dots, \beta_k)$, namely $(X_1/(\alpha_1 + 1), \dots, X_k/(\alpha_k + 1))$ is inadmissible for k $\geq$ 2 except when m = 0, in which case it is inadmissible for k $\geq$ 3. Ghosh and Parsian (1980) also discussed this problem for the same weighted quadratic losses. Gupta (1984) has shown that inadmissibility also holds for the loss function

$$L(\underset{\sim}{\beta}', \underset{\sim}{\delta}) = \sum_{i=1}^{k} \delta_i \beta_i - \sum_{i=1}^{k} \log \delta_i \beta_i - k.$$

The vector of natural estimators $(X_1/\alpha_1, \dots, X_k/\alpha_k)$ is an inadmissible estimator of $(\beta_1, \dots, \beta_k)$ for k $\geq$ 3. It **would** seem, however, that the critical dimension for inadmissibility is typically 2; three dimensions are required only in special cases. The problem whether the natural estimator is admissible for k = 2 is an open question. Zubrzycki (1966) has considered the case when β is **known** to exceed some positive number β_0. He has shown that with a loss function $(\beta^* - \beta)^2/\beta^2$, where β^* denotes an estimator of β, and given a single observed value of X, estimators

$$\beta^* = (\alpha + 1)^{-1} X + b, \tag{17.47}$$

with

$$\beta_0(\alpha + 1)^{-1} \leq b \leq 2\beta_0(\alpha + 1)^{-1},$$

have minimax risk [equal to $(\alpha + 1)^{-1}$] and are admissible in the class of estimators linear in X.

If γ is known, the maximum likelihood estimators of a and β might be denoted $\hat{\alpha}(\gamma), \hat{\beta}(\gamma)$ to indicate their dependence on y. We will, however, simply use $\hat{\alpha}$ and $\hat{\beta}$; no confusion between this use and that in Section 7.1 should arise.

If γ is known to be zero, the probability density function is of form (17.23). If $X_1, X_2, \dots, X_n$ are independent random variables each having distribution (17.23), then equations for the maximum likelihood estimators $\hat{\alpha}, \hat{\beta}$ are

$$n^{-1} \sum_{j=1}^{n} \log X_j = \log \hat{\beta} + \psi(\hat{\alpha}), \tag{17.48a}$$

$$\bar{X} = \hat{\alpha}\hat{\beta}. \tag{17.48b}$$

From (17.48b), $\hat{\beta} = \bar{X}/\hat{\alpha}$. Inserting this in (17.48a), we obtain the following equation for $\hat{\alpha}$:

$$n^{-1} \sum_{j=1}^{n} \log X_j - \log \bar{X} = \psi(\hat{\alpha}) - \log \hat{\alpha}; \tag{17.48c}$$

that is,

$$R_n = \log\left[\frac{\text{Arithmetic mean } (X_1, X_2, \dots, X_n)}{\text{Geometric mean } (X_1, X_2, \dots, X_n)}\right] = \log \hat{\alpha} - \psi(\hat{\alpha}).$$

(Note that $R_n \geq 0$.)

It is readily seen that the estimator $\hat{\alpha}$, of a the shape parameter, and the ratio $\hat{\beta}/\beta$ are distributed independently of β. In particular, the variance of $\hat{\beta}/\beta$ does not depend on the parent population value of β. The value of $\hat{\alpha}$ can be determined by inverse interpolation in a table of the function [log a $- \psi(\alpha)$]. Such a table has been published by Masuyama and Kuroiwa (1952). Chapman (1956) has published a table giving the results of such inverse interpolation (i.e., values of $\hat{\alpha}$) corresponding to a few values of the ratio of arithmetic to geometric mean. (He reported that a more complete table was available from the Laboratory of Statistical Research, University of Washington.)

Greenwood and Durand (1960) pointed out that the function α[log a $- \psi(\alpha)$] progresses much more smoothly than does [log a $- \psi(\alpha)$] and so is more convenient for interpolation. They gave a table of values of α[log a $- \psi(\alpha)$] as a function of α to eight decimal places for argument values 0.00(0.01)1.40, and to seven decimal places for argument values 1.4(0.2)18.0. This method eliminates the necessity of inverse interpolation and assures high accuracy using linear interpolation. Bain and Engelhardt (1975) show that 2na $\log R_n$ is approximately distributed as $c\chi^2_\nu$ for appropriate values of c and ν (depending on n and a). For a ≥ 2 we have $2n\alpha \log R_n$ approximately distributed as χ^2_{n-1}. [See also (17.107).]

$$H = \frac{\text{Arithmetic mean}}{\text{Arithmetic mean} - \text{Geometric mean}},$$

except when $\hat{\alpha}$ is less than about 2. They give a table of solutions of equation (17.48c) to five decimal places for

$$H = 1.000(0.001)1.010(0.002)1.030(0.005)1.080(0.01)1.16(0.02)1.40(0.05)$$
$$2.00(0.1)3.0(0.2)5.0(0.5)7.0(1)10(2)20(10)50.$$

For $H > 1.001$, linear interpolation gives four decimal place accuracy for 6.

If $\hat{\alpha}$ is large enough, the approximation $\psi(\alpha) = \log(\alpha - \frac{1}{2})$ may be used. Then from (17.48c) we have

$$\frac{\text{Arithmetic mean}}{\text{Geometric mean}} \doteqdot \frac{\hat{\alpha}}{\hat{\alpha} - \frac{1}{2}};$$

that is,

$$\hat{\alpha} \doteqdot \frac{\text{Arithmetic mean}}{2(\text{Arithmetic mean} - \text{Geometric mean})} = \tfrac{1}{2}H. \qquad (17.49)$$

For a better approximation 1/12 (= 0.083) should be subtracted from the right-hand side.

Thom (1968) suggests the approximation

$$\hat{\alpha}_T \doteqdot \tfrac{1}{4}R_n^{-1}\left(1 + \sqrt{1 + \tfrac{4}{3}R_n}\right). \qquad (17.50)$$

Thom further suggests adding the correction $[(\hat{\alpha}_T - 1)(24 - 96\hat{\alpha}_T)^{-1} + 0.00921$ if $\hat{\alpha}_T > 0.9$, and gives a table of corrections for $\hat{\alpha}_T < 0.9$. It is stated that with these corrections the value of $\hat{\alpha}_T$ should be correct to three decimal places.

Asymptotic formulas (as $n \to \infty$) for the variances of $\sqrt{n}\,\hat{\alpha}$ and $\sqrt{n}\,\hat{\beta}$, and the correlation between these statistics, are

$$\begin{cases} \operatorname{Var}(\sqrt{n}\,\hat{\alpha}) \doteqdot \alpha\{\alpha\psi'(\alpha) - 1\}^{-1}, \\ \operatorname{Var}(\sqrt{n}\,\hat{\beta}) \doteqdot \beta^2\psi'(\alpha)\{\alpha\psi'(\alpha) - 1\}^{-1}, \\ \operatorname{Corr}(\hat{\alpha}, \hat{\beta}) \doteqdot -\{\alpha\psi'(\alpha)\}^{-1/2}. \end{cases} \qquad (17.51)$$

Masuyama and Kuroiwa (1952) give tables with values of $\alpha\{\alpha\psi'(\alpha) - 1\}^{-1}$ and $\psi'(\alpha)\{\alpha\psi'(\alpha) - 1\}^{-1}$. If the approximation $\psi'(\alpha) = (\alpha - \frac{1}{2})^{-1}$, useful for α large, is used, we have

$$\begin{cases} \operatorname{Var}(\sqrt{n}\,\hat{\alpha}) \doteqdot 2\alpha(\alpha - \frac{1}{2}), \\ \operatorname{Var}(\sqrt{n}\,\hat{\beta}) \doteqdot \beta^2\alpha, \\ \operatorname{Corr}(\hat{\alpha}, \hat{\beta}) \doteqdot -\sqrt{1 - \frac{1}{2}\alpha^{-1}}\,. \end{cases} \tag{17.52}$$

Bowman and Shenton (1968) investigated the approximate solutions [due to Greenwood and Durand (1960)]

$$\hat{\alpha} \doteqdot R_n^{-1}\left(0.500876 + 0.1648852R_n - 0.0544274R_n^2\right), \qquad 0 < R_n \le 0.5772, \tag{17.53a}$$

$$\hat{\alpha} \doteqdot R_n^{-1}\left(17.79728 + 11.968477R_n + R_n^2\right)^{-1}\left(8.898919 + 9.059950R_n + 0.9775373R_n^2\right), \qquad 0.5772 \le R_n \le 17. \tag{17.53b}$$

The error of (17.53a) does not exceed 0.0088% and that of (17.53b) does not exceed 0.0054%.

If α is known but not β or γ, maximum likelihood estimators $\hat{\beta} = \hat{\beta}(\alpha)$, $\hat{\gamma} = \hat{\gamma}(\alpha)$ satisfy equations (17.37b)′ and (17.37c)′ with $\hat{\alpha}$ replaced by α. From (17.37b)′,

$$\hat{\gamma} = \bar{X} - \alpha\hat{\beta},$$

and hence (17.37b)′ can be written as an equation for β,

$$(\alpha - 1)\hat{\beta} = \left[n^{-1}\sum_{j=1}^{n}\left(X_j - \bar{X} + \alpha\hat{\beta}\right)^{-1}\right]^{-1}. \tag{17.37c)″}$$

Alternatively, using the first two sample moments, we have for moment estimators $\tilde{\beta} = \tilde{\beta}(\alpha)$ and $\tilde{\gamma} = \tilde{\gamma}(\alpha)$,

$$\tilde{\gamma} = \bar{X} - \alpha\tilde{\beta},$$

$$\alpha\tilde{\beta}^2 = m_2 \qquad \text{[cf. (17.41a) and (17.41b)]},$$

whence

$$\tilde{\beta} = \sqrt{\frac{m_2}{\alpha}}, \qquad \tilde{\gamma} = \bar{X} - \sqrt{\alpha m_2}\,. \tag{17.54}$$

In this case (a known) for n large,

$$\begin{cases} \mathrm{Var}(\hat{\beta}) = \beta^2 n^{-1}, \\ \mathrm{Var}(\tilde{\gamma}) = \frac{1}{2}\beta^2\alpha(\alpha - 2)n^{-1}, \\ \mathrm{Corr}(\hat{\beta}, \hat{\gamma}) \doteqdot -\dfrac{(\alpha + 1)}{a + 3} \end{cases} \tag{17.55}$$

while

$$\begin{cases} \mathrm{Var}(\tilde{\beta}) \doteqdot \frac{1}{2}\beta^2(1 + 3\alpha^{-1})n^{-1}, \\ \mathrm{Var}(\tilde{\gamma}) \doteqdot \frac{1}{2}\beta^2\alpha(\alpha + 3)n^{-1}, \\ \mathrm{Corr}(\tilde{\beta}, \hat{\gamma}) \doteqdot -\dfrac{(\alpha + 1)}{\alpha + 3}. \end{cases} \tag{17.56}$$

The advantage of the maximum likelihood estimators is not so great in this case as when all three parameters have to be estimated.

Glaser **(1976a)** observes that the distribution of R, is the same as that of $\prod_{i=1}^{n-1} V_i$, where the V_i's are independently distributed beta random variables with parameters a and i/n $(i = 1,\dots,$ n $- 1)$. Various methods for calculation of the distribution of R, and its lower critical values are available in the literature.

Provost (1988) provides expression for the jth moment of R, and provides an expression for its probability density function by inverting the Mellin transform. In their seminal paper and book Bowman and Shenton (1983, 1988) provide an approximation to the distribution of R,, along with a new approximation to the inverse function $\hat{\alpha} = \phi^{-1}(R_n)$, namely

$$\hat{\alpha} = \frac{1}{2R_n} + \frac{1}{6} - \frac{R_n}{18} - \frac{4R_n^2}{135} + \frac{47R_n^3}{810} \cdots. \tag{17.57a}$$

This can be used when R, is not too small; if R, is small, the formula

$$\hat{\alpha} \sim (\text{R, } + \log R_n)^{-1} \tag{17.57b}$$

is suggested. They also suggest using Thom's formula **(17.50)** as a starting value for an iterative procedure, calculating the mth iterate, $\hat{\alpha}_m$ from the formula

$$\hat{\alpha}_m = \frac{\hat{\alpha}_{m-1}\{\log \hat{\alpha}_{m-1} - \psi(\hat{\alpha}_{m-1})\}}{R_n} \tag{17.58}$$

They observe that about ten iterations suffices for reasonable accuracy.

Bowman and Shenton (1983) also obtain the formula

$$\kappa_s(R_n) = (-1)^s\{n^{1-s}\psi^{(s-1)}(\alpha) - \psi^{(s-1)}(n\alpha)\} \tag{17.59}$$

for the sth cumulant of R_n. As $n \to \infty$,

$$\mu_1'(R_n) \quad \frac{n-1}{2n\alpha}, \tag{17.60a}$$

$$\mu_2(R_n) \sim \frac{n-1}{2n^2\alpha^2}, \tag{17.60b}$$

$$\mu_3(R_n) \sim -2\sqrt{\frac{2}{n-1}}, \tag{17.60c}$$

$$\mu_4(R_n) \sim 3 + \frac{12}{n-1}. \tag{17.60d}$$

These values suggest that for n large R_n is approximately distributed as $\chi^2_{n-1}/(2n\alpha)$. For n large

$$E[\hat{\alpha}] \sim \frac{n\alpha}{n-3} - \frac{2}{3(n-3)} + \frac{n-1}{9(n-3)n\alpha} + \frac{7(n^2-9)}{54(n^2-9)n^2\alpha^2} + \frac{(n^2-1)(26n^3+33n^2-324n-707)}{810(n^2-9)(n+5)n^3\alpha^3} + \dots, \tag{17.61a}$$

$$\operatorname{Var}(\hat{\alpha}) \sim \frac{2n^2\alpha^2}{(n-3)^2(n-5)} - \frac{2n(n+1)\alpha}{3(n-3)^2(n-5)} + \frac{2(n^2-3n+\delta)}{9(n-3)^2(n-5)} + \cdots \tag{17.61b}$$

$$\sqrt{\beta_1(\hat{\alpha})} \sim \frac{4\sqrt{(2n-10)}}{n-7}\left\{1 - \frac{(n-3)^2}{24n^2\alpha^2} + \cdots\right\}, \tag{17.61c}$$

$$\beta_2(\hat{\alpha}) \sim \frac{3(n+9)(n-5)}{(n-7)(n-9)}\left\{1 - \frac{4(n-3)^2}{3(n+9)n^2\alpha^2} - \cdots\right\}$$

[Bowman and Shenton (1988)]. (17.61d)

Bowman and Shenton (1982) give tables of coefficients of series expansions for the expected value, variance and third and fourth central moments of $\hat{\alpha}$ and $\hat{\beta}$ (terms up to n^{-6}) for $\alpha = 0.2(0.1)3(0.2)5(0.5)15$ and $n = 6(1)50(5)100(10)150$.

Choi and Wette (1969) present results of sampling investigations. As an example, with $\alpha = 2$ and $\beta = 1$ they obtain the average values (arithmetic means of 100 repetitions) shown below:

n	$\hat{\alpha}$	$\hat{\beta}$
40	2.10	1.03
120	2.04	1.02
200	2.03	1.01

The positive bias in $\hat{\alpha}$ is to be expected from the expansion (17.61a). Note that the next term in the expansion is

$$+\frac{(n^2-1)\pi(n)}{17010(n^2-9)(n+5)(n+7)n^4\alpha^4},$$

where

$$\pi(n) = 1004n^5 + 9363n^4 + 13358n^3 - 82019n^2 - 296760n - 288472.$$

The expansion is asymptotic in nature, so only limited accuracy can be expected from it's use. It is more accurate than an expansion simply in descending powers of n.

Anderson and Ray (1975), noting that the bias in $\hat{\alpha}$ can be considerable when n is small, suggest using the following, less biased estimator based on (17.61a):

$$\hat{\alpha}^* = \frac{n-3}{n}\hat{\alpha} + \frac{2}{3n}. \tag{17.62}$$

For estimating $\theta = \beta^{-1}$, these authors suggest $\hat{\beta}^{-1}f(\hat{\alpha}^*)$, where

$$f(a) = \left\{1 + \frac{3na}{(n-3)(na-1)}\left(1 + \frac{1}{9a} - \frac{1}{na} + \frac{n-1}{27na^2}\right)\right\}. \tag{17.63}$$

Shenton and Bowman (1973) introduce the "almost unbiased" estimators

$$\tilde{\alpha} = \bar{X}\left\{\frac{2nR_n}{n-1} - \frac{2nR_n^2}{2(n-1)} + \frac{4n(n+1)R_n^3}{9(n-1)(n+3)} - \frac{2n(7n^2+60n+7)R_n^4}{135(n-1)(n+3)(n+5)}\right\}, \tag{17.64a}$$

$$\tilde{\beta} = \frac{n-3}{2nR_n} + \frac{n+1}{6n} - \frac{(n+1)R_n}{18n} - \frac{(4n^2-10n+4)R_n^2}{135n(n+3)}. \tag{17.64b}$$

Dahiya and Gurland (1978) develop generalized minimum chi-squared estimators of α and β. Stacy (1973) proposed "quasimaximum likelihood" estimators for $1/\alpha$ and β,

$$\left(\frac{1}{\alpha}\right)' = n(n-1)^{-1}\sum(Z_i - n^{-1})\log Z_i, \qquad (17.65a)$$

$$\beta' = \bar{X}\left(\frac{1}{\alpha}\right)', \qquad (17.65b)$$

where $Z_i = X_i/n\bar{X}$ $(i = 1,\dots,n)$.

These estimators arise from the maximum likelihood estimators for Stacy and Mihram's (1965) generalized gamma distribution (see Section 8). The estimators are unbiased and

$$\operatorname{Var}\left(\left(\frac{1}{\alpha}\right)'\right) = \frac{1}{(n-1)\alpha^2}\left\{1 + \frac{n\alpha^2 + \psi'(\alpha+1)}{n\alpha+1}\right\}, \qquad (17.66a)$$

$$\operatorname{Var}(\tilde{\beta}) = \frac{\beta^2}{(n-1)\alpha}\{1 + \alpha + \alpha^2\psi'(\alpha+1)\}. \qquad (17.66b)$$

The asymptotic efficiencies are for $(1/\alpha)'$

$$\frac{n-1}{n}\left[\{1 + \alpha^2\psi'(\alpha+1)\}\left\{1 + \frac{n\alpha^2\psi'(\alpha+1)}{n\alpha+1}\right\}\right]^{-1}, \qquad (17.67a)$$

and for $\tilde{\beta}$

$$\frac{n-1}{n}\{1 + \alpha + \alpha^2\psi'(\alpha+1)\}^{-1}. \qquad (17.67b)$$

Grouped data can be tested by standard methods. Rosaiah, Kantam, and Narasimham (1991) give a detailed analysis for fitting a two-parameter gamma distribution. They provide tables for optimum group length d, in the sense of (1) minimum asymptotic generalized variance or (2) minimum sum of asymptotic variances, for $\alpha = 2(1)5$ and number of groups, $k = 2(1)10$ for the cases when neither α nor β are known or when only α or only β is known. The group boundaries are $0, d\beta, 2d\beta, \dots, (k-1)d\beta, \infty$. Since it is necessary to know both α and β to use the tables, they can be regarded as only providing a basis for intelligent selection of group length. Tables are also provided for use when group lengths may vary.

7.3 Estimation of Shape Parameter (β and γ Known)

Bowman and Shenton (1970) carried out an extensive study of the distributional properties of two different estimators of the shape parameters, a, of the gamma distribution (17.2). These were as follows:

1. The maximum likelihood estimator (MLE) $\hat{\alpha}$.
2. Thom's (1968) estimator $\hat{\alpha}_T$ [see (17.50)].

We recall that $\hat{\alpha}$ is the solution of the equation

$$R_n = \log \hat{\alpha} - \psi(\hat{\alpha}), \tag{17.48c$'$}$$

while

$$\hat{\alpha}_T = \tfrac{1}{4}R_n^{-1}\left(1 + \sqrt{1 + \tfrac{4}{3}R_n}\right), \tag{17.50}$$

which is applicable when $\tilde{\alpha}_T$ is not too small.

3. The moment estimator

$$\tilde{\alpha} = \frac{m_1'^2}{m_2} = \frac{\bar{X}^2}{S^2}. \tag{17.68}$$

The estimator (17.68) is much easier to compute than either $\hat{\alpha}$ or $\hat{\alpha}_T$. All three estimators are unaffected by the value of the scale parameter (β). Dusenberry and Bowman (1977) compare these three estimators. In regard to $\tilde{\alpha}$, they apply the techniques of David and Johnson (1951) to calculate the cdf of $\tilde{\alpha}$ by noting that

$$\Pr[\tilde{\alpha} < a] = \Pr\left[\frac{\bar{X}^2}{S^2} < a\right] = \Pr[\bar{X}^2 - aS^2 < 0].$$

It is not difficult to determine the moments of $(\bar{X}^2 - aS^2)$. These are then used to approximate the required probability. Dusenberry and Bowman (1977) use the Cornish-Fisher expansion to evaluate percentage points of the distribution of $\tilde{\alpha}$. They plot values of $\Pr[\tilde{\alpha} < a]$ for

$$\alpha = 0.5(0.5)3.0(1)5(5)20,$$

$$n = 50, 75, 100, 150, 200, 300, 750, 1000.$$

Figure 17.3 is the plot for a = 2.0.

The estimator $\hat{\alpha}_T$ is quite close to $\hat{\alpha}$. Although $\hat{\alpha}_T$ has a slight systematic bias, this is offset by a somewhat lower variance and greater ease of calculation. The moment estimator $\tilde{\alpha}$ is much easier to compute than either

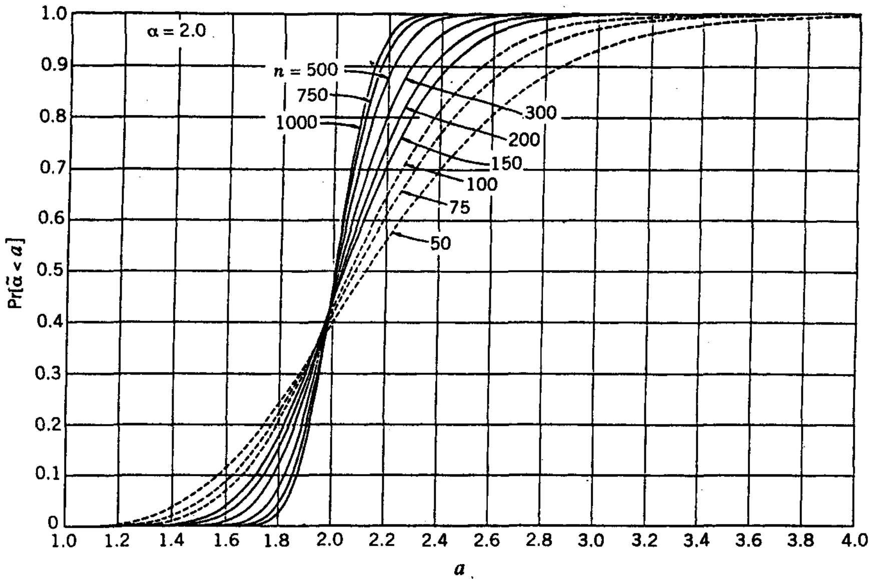

Figure 17.3 $\Pr[\tilde{\alpha} < a]$ **for** $\alpha = 2.0$.

a or $\hat{\alpha}_T$ but has a greater bias and variance than either of these two estimators. As measured by $\sqrt{\beta_1}$ and β_2, the distribution of $\tilde{\alpha}$ is closer to normality than that of either $\hat{\alpha}$ or $\hat{\alpha}_T$ for nearly all the values of ***a*** and n considered by Dusenberry and Bowman (1977).

Blischke (1971) constructed a BAN (best asymptotically normal) estimator of a, when β is known, as

$$\tilde{\alpha}' = \tilde{\alpha} + \frac{n^{-1}\sum_{i=1}^{n} \log(X_i/\beta) - \psi(\tilde{\alpha})}{\psi'(\tilde{\alpha})} \tag{17.69}$$

Since $\tilde{\alpha}$ is a consistent estimator of α and

$$\mathrm{E}\left[\log\left(\frac{X}{\beta}\right)\right] = \psi(\alpha),$$

the estimator $\hat{\alpha}'$ is asymptotically unbiased.

Results from 50 simulated samples of sizes $n = 11$, 31, and 51 are set out in Table 17.1, comparing values of $\hat{\alpha}$ and $\tilde{\alpha}'$. Note that $\hat{\alpha}$ has a positive bias, except for the case where ***a*** = 2.00 and $n = 31$. Huque and Katti (1976) and Bar-Lev and Reiser (1983) discuss maximum ***conditional*** likelihood estimation of ***a***. There is little to choose between the two estimators. Note the generally positive bias of the maximum likelihood estimator.

Table 17.1 Results of Simulation for $\tilde{\alpha}$ and $\hat{\alpha}$

		BAN Estimator ($\hat{\alpha}'$)		ML Estimator ($\hat{\alpha}$)	
α	n	Arithmetic Mean	Sample Variance	Arithmetic Mean	Sample Variance
0.25	11	0.234[a]	0.00565[a]	0.269	0.0067
	31	0.243	0.00214	0.258	0.0020
	51	0.238	0.00088	0.252	0.0010
0.50	11	0.530	0.0271	0.543	0.0219
	31	0.476	0.0059	0.498	0.0052
	51	0.494	0.0025	0.506	0.0045
0.75	11	0.760	0.0480	0.773	0.0339
	31	0.743	0.0101	0.765	0.0135
	51	0.745	0.0063	0.747	0.0069
1.25	11	1.228	0.0678	1.301	0.0879
	31	1.195	0.0204	1.257	0.0221
	51	1.253	0.0171	1.260	0.0172
1.50	11	1.533	0.0958	1.537	0.1010
	31	1.487	0.0362	1.508	0.0355
	51	1.529	0.0233	1.525	0.0199
2.00	11	2.063	0.121	2.036	0.1270
	31	1.965	0.067	1.996	0.0514
	51	1.979	0.031	2.004	0.0321

[a] Based on 48 values; in two cases negative estimates were obtained.

7.4 Order Statistics and Estimators Based on Order Statistics

A considerable amount of work has been done in evaluating the lower moments of order statistics $X_1' \le X_2' \le \cdots \le X_n'$ corresponding to sets of independent random variables $X_1, \dots, X_n$ having a common standard gamma distribution of form (17.2). Since γ and β are purely location and scale parameters, the results are easily extended to the general form (17.1). [Moments of order statistics of the exponential distribution ($\alpha = 1$) are discussed in some detail in Chapter 19, Section 6.] Tables of moments of order statistics from random samples of size n from the standard gamma distribution (17.2) are summarized in Table 17.2.

We next note that Kabe (1966) has obtained a convenient formula for the characteristic function of any linear function $\sum_{j=1}^n a_j X_j'$ of the order statistics. The characteristic function is

$$E\left[\exp\left\{it\sum_{j=1}^n a_j X_j'\right\}\right] = n![\Gamma(\alpha)]^{-n}\int_0^\infty \cdots \int_0^{x_3}\int_0^{x_2}\left\{\prod_{j=1}^n x_j\right\}^{\alpha-1}$$

$$\times \exp\left\{-\sum_{j=1}^n (1 - ia_j t)x_j\right\} dx_1\, dx_2 \dots dx_n. \quad (17.70)$$

Applying the transformation

$$x_r = \prod_{j=r}^{n} w_j$$

so that $0 < w_j < 1$ for $j = 1, 2, \dots, (n-1)$, and $w_n > 0$, we obtain the formula

$$n![\Gamma(\alpha)]^{-n} \int_0^1 \int_0^1 \cdots \int_0^1 \int_1^{\infty} \left\{ \prod_{j=1}^{n} w_j^{j\alpha - 1} \right\} \exp\{-w_n D(w)\}\, dw_n\, dw_{n-1} \dots dw_1$$

$$= \frac{n!\Gamma(n\alpha)}{[\Gamma(\alpha)]^n} \int_0^1 \int_0^1 \cdots \int_0^1 \left\{ \prod_{j=1}^{n-1} w_j^{j\alpha - 1} \right\} [D(w)]^{-n\alpha}\, dw_{n-1} \dots dw_1, \tag{17.71}$$

where

$$D(w) = (1 - ia_n t) + w_{n-1}(1 - ia_{n-1}t) + w_{n-2}w_{n-1}(1 - ia_{n-2}t) + \cdots + w_1 w_2 \dots w_{n-1}(1 - ia_1 t).$$

Table 17.2 Details of Available Tables on Moments of Gamma Order Statistics

Reference	Values of n	Shape Parameter α	Order of Moments	Serial Number (j) of Order Statistics	Number Figures[c]
Gupta (1961)[a]	1(1)10	1(1)5	1(1)4	1(1)n	6 s.f.
	11(1)15	1(1)5	7(1)4	1	
Breiter and Krishnaiah (1968)	1(1)16	0.5(1)5.5 (0.5)10.5	1(1)4	1(1)n	5 s.f.
Harter (1970)	1(1)40	0.5(0.5)4.0	1	1(1)n	5 d.p.
Prescott (1974)	2(1)10	2(1)5	1	1(1)n	4 d.p.
Walter and Stitt (1988)	1(1)25	1(1)10(5)20	1	1, n	4 d.p.
	1(1)5	1, 5(5)20	1	1(1)n	
	10(5)25	1, 5(5)20	1	1, 5(5)n	
Balasooriya and Hapuarachchi (1991)[b]	1(1)10 (5)40	5(1)8	1(1)5	1(1)2	5 d.p.

[a]The values of BS (1992) occasionally differ from those of Gupta (1961) by more than 0.00001.
[b]The last of these references also includes covariances for $n = 15(5)25$, $a = 2(1)5$, extending Prescott (1974) which includes covariances for $n = 2(1)10$ and the same values of a. More extensive tables of expected values of order statistics for $n = 15(1)40$, $\alpha = 5(1)8$, and covariances for $n = 2(1)25$ and $a = 2(1)8$ are available on request from Balasooriya and Hapuarachchi (Memorial University of Newfoundland, St. John's).
[c]s.f. = significant figures; d.p. = decimal places.

The multiple integral can be expanded as a series of beta functions. Although we will not use it here directly, equation (17.71) is very convenient as a starting point for studying the distributions of linear functions of order statistics from gamma distributions.

The distribution of X_r', the rth smallest among n independent random variables each having distribution (17.2), has the probability density function

$$p_{X_r'}(x) = \frac{n!}{(r-1)!(n-r)!}\left[\frac{\Gamma_x(\alpha)}{\Gamma(\alpha)}\right]^{r-1}\left[1-\frac{\Gamma_x(\alpha)}{\Gamma(\alpha)}\right]^{n-r}\frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}, \qquad x>0. \quad (17.72)$$

In general, this expression does not lend itself to simple analytic treatment. However, if α is a positive integer,

$$\frac{\Gamma_y(\alpha)}{\Gamma(\alpha)} = \frac{1}{\Gamma(\alpha)}\int_0^y t^{\alpha-1}e^{-t}\,dt = 1 - e^{-y}\sum_{j=0}^{\alpha-1}\left\{\frac{y^j}{j!}\right\} \qquad \text{[cf. (17.24a)]}$$

and so (17.72) becomes

$$p_{X_r'}(x) = \frac{n!}{(r-1)!(n-r)!}\left[1-e^{-x}\sum_{j=0}^{\alpha-1}\left\{\frac{x^j}{j!}\right\}\right]^{r-1} \times\left[\sum_{j=0}^{\alpha-1}\left\{\frac{x^j}{j!}\right\}\right]^{n-r}\frac{x^{\alpha-1}e^{-(n-r+1)x}}{(\alpha-1)!}, \qquad x>0. \quad (17.73)$$

In this case it is possible to express all moments (of integer order) of X_r' and all product moments (of integer orders) of order statistics as finite sums of terms involving factorials, although these expression will usually be cumbersome.

Johnson (1952) has obtained an approximation to the distribution of range $(X_n' - X_1')$ for random samples from (17.2). Tiku and Malik (1972) provide β_1 and β_2 values of the rth-order statistic X_r' for the gamma distribution and compare them with approximate values obtained by equating the first three moments of χ_f^2 with those of $(X_r' + a)/g$. Numerical comparisons are presented in Table 17.3. The agreement is excellent except for a = 0.01. It is, however, doubtful whether any method of approximation based on moments can provide accurate values for the extreme lower tail, as the classical paper of E. S. Pearson (1963) indicated.

Lingappaiah (1974) obtained the joint density of X_r' and X_s' $(s > r)$, from the standard gamma distribution (17.1), and he derives from it that of the difference $U_{r,s} = X_s' - X_r'$. More recently Lingappaiah (1990) has obtained the distribution of $U_{r,s}$ in the presence of a single outlier—when n − 1 observations are from a standard gamma α distribution but one observation is from a gamma $\alpha + \delta$ distribution. Typical values of $\Pr[U_{r,s} > u]$ under these conditions are shown in Table 17.4.

Table 17.3 Comparison of the Percentage Points of rth-Order Statistics from Standard Gamma Distribution with values from Tiku and Malik's (1972) Approximation (with a = 2, sample size n = 6)

		Upper Percentage Points				
		50	75	90	95	99
$r = 2$	Exact	1.135	1.536	1.963	2.250	2.854
	Approximate	1.135	1.535	1.962	2.249	2.854
$r = 4$	Exact	2.376	3.036	3.744	4.225	5.254
	Approximate	2.370	3.040	3.754	4.234	5.245
		Lower Percentage Points				
		25	10	05	01	
$r = 2$	Exact	0.807	0.570	0.453	0.280	
	Approximate	0.808	0.571	0.453	0.276	
$r = 4$	Exact	1.829	1.423	1.214	0.885	
	(Approximate	1.821	1.423	1.224	0.924	

As is to be expected, in Table 17.4 $\Pr[U_{r,s} > u]$ increases with 6. Note that the values depend only on αu, so a single table (with a = 1) might suffice. Lingappaiah (1991) also provides tables (to 5 d.p.) with negative moments $E[X_r'^{-i}]$ of the standard gamma a distribution for n = 2(1)5 with i = $-(\alpha -$ 1), – (a – 2),..., –1. As far as we know, the most general recurrence relationship in the literature for moments of order statistics from gamma a samples are those in Thomas and Moothathu (1991).

These formulas relate values of (descending) factorial moments

$$\mu_{n:n}^{(i)} = E\left[X_n'^{(i)}\right] = E[X_n'(X_n' - 1) \cdots (X_n' - i + 1)]$$

of the greatest X in a random sample of size n from the standard gamma distribution (17.1). The recurrence formula is

$$\frac{\Gamma(k + \alpha n - n + 1)}{\{\Gamma(\alpha)\}^n} \times \frac{1}{n^{k+n(\alpha-1)}} = \sum_{j=1}^{n} A_{n-1,j}\mu_{n,n}^{(k-n+j)},$$

$$n \geq 2;\ k \geq \max[1 - \alpha, n(1 - a)], \qquad (17.74)$$

Table 17.4 Values of $\Pr[U_{r,s} > u]$ for n = 5, r = 1, s = 2

u	$\alpha = 0.5$ 0.25	0.5	0.75	1.0	a = 1.0 0.25	0.5	0.75	1.0
δ								
0	0.6065	0.3679	0.2231	0.1353	0.3679	0.1353	0.0498	0.0183
1	0.6672	0.4414	0.2901	0.1895	0.4414	0.1895	0.0796	0.0329
2	0.6831	0.4654	0.3161	0.2138	0.4654	0.2138	0.0968	0.0432

Note: Included are four observations from gamma a, one from gamma $\alpha + \delta$.

where

$$A_{0,0} = 1,$$

$$A_{r,1} = -r^{-1}\{k - r + 1 + (n - r)(\alpha - 1)\}A_{r-1,1},$$

$$A_{r,j} = r^{-1}(n - r)A_{r-1,j-1} - r^{-1}\{k - r + j + (n - r)(\alpha - 1)\}A_{r-1,j} \qquad (j = 2,3,\dots,r - 1)$$

$$A_{r,r} = r^{-1}(n - r)A_{r-1,r}.$$

If α is an integer, expected values and moments of the other order statistics can be evaluated using Joshi's (1979) formula:

$$\mu'^{(i)}_{k:n} = \mu'^{(i)}_{k-1:n-1} + in^{-1}\Gamma(\alpha)\sum_{j=0}^{\alpha-1}\left\{\frac{\mu'^{(i+j-\alpha)}_{k:n}}{j!}\right\}, \tag{17.75}$$

with $\mu'^{(i)}_{k:n} = E[X'^{i}_{k:n}]$; $\mu'^{(0)}_{k:n} = 1$; $\mu'^{(i)}_{0:n} = 0$ for $i \ge 1$. Prescott (1974) showed that

$$E[X'_r X'_s] = \frac{C}{\{\Gamma(\alpha)\}^2}\sum_b\sum_c\sum_u\sum_v\sum_j(-1)^{b+c}\binom{r-1}{b}\binom{s-r-1}{c}$$
$$\times a_u(\alpha,t-1)a_v(\alpha,q-1)\frac{\Gamma(u+\alpha+1)\Gamma(u+\alpha+j+1)}{t^{u+\alpha-j+1}(t+q)^{v+\alpha+j+1}j!}, \tag{17.76}$$

where $t = n - s + c + 1$, $q = b + s - r - c$, and summation is over $0 \le b \le r - 1$, $0 \le c \le r - s - 1$, $0 \le u \le (\alpha - 1)(t - 1)$, $0 \le v \le (\alpha - 1)(q - 1)$, $0 \le j \le u + \alpha$, and

$$C = \frac{n!}{(r-1)!(s-r-1)!(n-s)!},$$

and where $a_g(h,i)$ is the coefficient of z^g in the expansion of $\{\sum_{j=1}^{h-1}(z^j/j!)\}^i$.

Gupta (1962) obtained recurrence formulas for the crude moments $\mu'^{(i)}_{k} = E[X'^{i}_{k:n}]$ of the kth-order statistic $(X'_{k:n})$. These are

$$\mu'^{(i)}_{1:n} = \frac{n}{\Gamma(\alpha)}\sum_{j=0}^{(n-1)(\alpha-1)} a_j(\alpha,n-1)\frac{\Gamma(\alpha+i+j)}{n^{\alpha+i+j}}, \tag{17.77a}$$

$$\mu'^{(i)}_{k:n} = \frac{n!}{(k-1)!(n-k)!}\sum_{j=0}^{k-1}(-1)^j\binom{k-1}{j}\frac{\mu'^{(i)}_{1:n-k+j+1}}{n-k+j+1}, \tag{17.77b}$$

where $a_j(\alpha,p)$ is as defined in (17.76); see Balakrishnan and Cohen (199

Prescott (1974) provided a table of variances and covariances of all ord statistics for n = 2(1)10 with a = 2(1)5 to four decimal places. (He used log-gamma computer routine, accurate to about ten decimal places, to avc errors occurring in tables prepared by Gupta (1962)—which have be ascribed to the way in which ratios of gamma functions were calculated.)

Using Gupta's tables (1962), it is possible to construct best linear unbiased estimators of the parameter β if α and γ are known. Coefficients of such estimators have been given by Musson (1965). Coefficients for best linear unbiased estimators, not using all of the sample values have been given by

1. Karns (1963) using only one-order statistic,
2. Bruce (1964) using the least M values out of n,
3. Hill (1965) using only the least number of order statistics (from a complete or censored sample) to give a specified efficiency relative to the best linear unbiased estimator using all available order statistics,
4. Sarndal (1964) using the best k-order statistics. (Sarndal also considers estimation of β and γ, α being known.)

Returning now to situations where it is necessary to estimate all three parameters α, β, and γ, we consider maximum likelihood estimation when the least r_1 and greatest r_2 of the X's have been censored. The maximum likelihood equations are [introducing $\hat{Z}_j = (X_j' - \hat{\gamma})/\hat{\beta}$ for convenience]

$$\sum_{j=r_1+1}^{n-r_2} \log \hat{Z}_j - n\psi'(\hat{\alpha}) + \frac{\Gamma'(\hat{\alpha}) - \Gamma'_{\hat{Z}_{n-r_2}}(\hat{\alpha})}{\Gamma(\hat{\alpha}) - \Gamma_{\hat{Z}_{n-r_2}}(\hat{\alpha})} r_2 + \frac{\Gamma'_{\hat{Z}_{r_1+1}}(\hat{\alpha})}{\Gamma_{\hat{Z}_{r_1+1}}(\hat{\alpha})} r_1 = 0, \tag{17.78a}$$

$$-(n - r_1 - r_2)\hat{\alpha} + \sum_{j=r_1+1}^{n-r_2} \hat{Z}_j + \frac{\hat{Z}_{n-r_2}^{\hat{\alpha}} e^{-\hat{Z}_{n-r_2}}}{\Gamma(\hat{\alpha}) - \Gamma_{\hat{Z}_{n-r_2}}(\hat{\alpha})} r_2 - \frac{\hat{Z}_{r_1+1}^{\hat{\alpha}} e^{-\hat{Z}_{r_1+1}}}{\Gamma_{\hat{Z}_{r_1+1}}(\hat{\alpha})} r_1 = 0, \tag{17.78b}$$

$$-(\hat{\alpha} - 1) \sum_{j=r_1+1}^{n-r_2} \hat{Z}_j^{-1} + (n - r_1 - r_2) + \frac{\hat{Z}_{n-r_2}^{\hat{\alpha}-1} e^{-\hat{Z}_{n-r_2}}}{\Gamma(\hat{\alpha}) - \Gamma_{\hat{Z}_{n-r_2}}(\hat{\alpha})} r_2 - \frac{\hat{Z}_{r_1+1}^{\hat{\alpha}-1} e^{-\hat{Z}_{r_1+1}}}{\Gamma_{\hat{Z}_{r_1+1}}(\hat{\alpha})} r_1 = 0. \tag{17.78c}$$

The equations simplify if either $r_1 = 0$ or $r_2 = 0$. For the case $r_1 = 0$ (censoring from above) a method of solving the equations is given by Harter and Moore (1965); also see Balakrishnan and Cohen (1991).

Estimation is simplified if the value of γ is known. Without loss of generality it may be arranged (if γ is known) to make $\gamma = 0$ (by adding, if necessary, a suitable constant to each observed value). For this case, with data censored from above ($r_1 = 0$), Wilk, Gnanadesikan, and Huyett (1962a, b) have provided tables which considerably facilitate solution of the

maximum likelihood equations. They express these equations in terms of

$$P = \frac{\left(\prod_{j=1}^{n-r_2} X'_j\right)^{1/(n-r_2)}}{X'_{n-r_2}}, \tag{17.79}$$

$$S = \frac{\left(\sum_{j=1}^{n-r_2} X'_j\right)}{(n-r_2)X'_{n-r_2}}; \tag{17.80}$$

that is, the ratios of the geometric and arithmetic means of the available observed values to their maximum. The maximum likelihood equations for $\hat{\alpha}$ and $\hat{\beta}$ are

$$(n-r_2)\log P = n\left[\psi'(\hat{\alpha}) - \log\left(\frac{X'_{n-r_2}}{\hat{\beta}}\right)\right] - r_2\frac{\partial \log J(\hat{\alpha})}{\partial\hat{\alpha}}, \tag{17.81a}$$

$$\frac{SX'_{n-r_2}}{\hat{\beta}} = \hat{\alpha} - \left\{\frac{r_2}{n-r_2}\right\}\frac{e^{-X'_{n-r_2}/\hat{\beta}}}{J(\hat{\alpha})}, \tag{17.81b}$$

where

$$J(\hat{\alpha}) = \int_1^\infty t^{\hat{\alpha}-1}e^{-X'_{n-r_2}t/\hat{\beta}}\,dt. \tag{17.82}$$

Note that r_2 and n enter the equations only in terms of the ratio r_2/n, and X'_{n-r_2} and $\hat{\beta}$ only as the ratio X'_{n-r_2}/β. Wilk, Gnanadesikan, and Huyett (1962b) provide tables, giving $\hat{\alpha}$ and $\hat{\alpha}\hat{\beta}/X'_{n-r_2}$ to three decimal places for

$$\frac{n}{r_2} = 1.0, 1.1, 1.2(0.2)2.0, 2.3, 2.6, 3.0,$$

$$P = 0.04(0.04)1.00, \quad \text{and} \quad S = 0.08(0.04)1.00.$$

The values for $n/r_2 = 1$ of course correspond to uncensored samples. A special table, which we have already mentioned in Section 7.2, is provided for this case. Wilk, Gnanadesikan, and Lauh (1966) discuss generalizations and modifications of these techniques for estimation of an unknown common scale parameter based on order statistics from a sample of gamma random variables with known shape parameters not necessarily all equal.

If a is known, it is possible to use "gamma probability paper," as described by Wilk, Gnanadesikan, and Huyett (1962a) to estimate β and γ graphically. This entails plotting the observed order statistics against the corresponding expected values for the standard distribution (17.2) (which of course depends on a) or, if these are not available, the values ξ_j satisfying the equations

$$\frac{j}{n+1} = [\Gamma(\alpha)]^{-1} \int_0^{\xi_j} x^{\alpha-1} e^{-x}\, dx. \tag{17.83}$$

In the case of progressively censored sampling, Cohen and Norgaard (1977) and Cohen and Whitten (1988) suggest the following procedure for solving the maximum likelihood equations (in the case when a > 1): Let n denote the total sample size and D, the number of failing items, for which there are completely determined life spans. Suppose that censoring occurs in k stages at times $T_1 < T_2 < \cdots < T_k$ and that C_j surviving items are selected randomly and withdrawn (censored) from further observation at time I;.. Then

$$n = D + \sum_{j=1}^{k} C_j. \tag{17.84}$$

The sample data consist of the ordered life span observations $\{X_i\}$ $(i = 1, 2, \dots, D)$, the censoring times {I;.}and the numbers of censored items $\{C_j\}$ $(j = 1, 2, \dots, k)$. The likelihood function is

$$L = K \prod_{i=1}^{D} p_X(X_i) \prod_{j=1}^{k} \{1 - F_X(T_j)\}^{C_j}, \tag{17.85}$$

where K is a constant, and $p_X(\cdot)$ and $F_X(\cdot)$ are the pdf and cdf of the lifetime distribution, respectively.

For the three-parameter gamma lifetime distribution (17.1), we have

$$\log L = -D \log \Gamma(\alpha) - n\alpha \log \beta - \beta^{-1} \sum_{i=1}^{D} (X_i - y)$$

$$+ (\alpha - 1) \sum_{i=1}^{D} \log(X_i - \gamma) + \sum_{j=1}^{k} C_j \log(1 - F_j) + \log K, \tag{17.86}$$

where

$$F_j = \{\beta^\alpha \Gamma(\alpha)\}^{-1} \int_0^{T_j - \gamma} y^{\alpha-1} \exp\left(\frac{-y}{\beta}\right) dy. \tag{17.87}$$

The corresponding maximum likelihood equations are

$$\frac{\partial \log L}{\partial \alpha} = -D\psi(\alpha) - D\log\beta + \sum_{i=1}^{D}\log(X_i-\gamma) - \sum_{j=1}^{k}\frac{C_j}{1-F_j}\frac{\partial F_j}{\partial \alpha} = 0, \tag{17.88a}$$

$$\frac{\partial \log L}{\partial \beta} = -\frac{D\alpha}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{D}(X_i-\gamma) - \sum_{j=1}^{k}\frac{C_j}{1-F_j}\frac{\partial F_j}{\partial \beta} = 0, \tag{17.88b}$$

$$\frac{\partial \log L}{\partial \gamma} = \frac{D}{\beta} - (\alpha-1)\sum_{i=1}^{D}(X_i-\gamma)^{-1} - \sum_{j=1}^{k}\frac{C_j}{1-F_j}\frac{\partial F_j}{\partial \gamma} = 0 \tag{17.88c}$$

[cf. **(17.37a)–(17.37c)** for the case of uncensored samples].

Evaluation of the partial derivatives of F_j with respect to α, β, and γ and computational details are given in Cohen and **Whitten** (1988). They caution that convergence problems may arise in iterative solution of equations (17.88) unless $\alpha \gg 1$ (in our opinion, $\alpha \geq 2.5$). Cohen and Norgaard (1977) assert that for $\alpha \geq 4$, the formulas can be used "without any hesitation."

When α is less than 1, the likelihood function tends to infinity as $\gamma \to X_1'$. Cohen and Norgaard (1977) suggest setting an initial value $\hat{\gamma} = X_1' - \frac{1}{2}\eta$, where η is "the precision with which observations are made," and then proceeding iteratively. They also provide **computational details** for calculation of the asymptotic variance-covariance matrix of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$.

The maximum likelihood estimator $\hat{\alpha}$ is the solution of

$$n^{-1}\sum_{i=1}^{n}\log\left(\frac{X_i}{\beta}\right) = \psi(\hat{\alpha}). \tag{17.90}$$

A median estimator α^* is the solution of

$$\log\left(\frac{\text{Median}(X)}{\beta}\right) = \psi(\alpha^*). \tag{17.91}$$

Harner, **Shyu,** and **Trutzer (1991)** carried out simulation studies of robustness of these estimators with respect to contamination of a gamma $(\alpha, 1)$ distribution by a gamma $(\alpha_c, 1)$ distribution. They took sample sizes $n = 25, 75$, with $\alpha = 1, 2, 5$ and $\alpha_c = 0.1, 1, 5, 10$; the proportions of contaminant [gamma $(\alpha_c, 1)$] were $p = 0, 0.01, 0.05$, and 0.1. They concluded the following:

1. The moment estimator $(\tilde{\alpha})$ "greatly overestimates" the value of α.
2. The maximum likelihood estimator $(\hat{\alpha})$ "is competitive except when $\alpha_c = 0.1$."
3. The median estimator (α^*) is "fairly stable over all combinations of the simulation parameters," with positive bias for small α, decreasing to "negligible amounts" as α or n increase.

Harner, Shyu, and Trutzer (1991) also consider trimmed mean estimators $(\hat{\alpha}_{\theta_1,\theta_2})$ satisfying the equation

$$\frac{\sum_{i=k_1}^{k_2} \log(X_i'/\beta)}{n - [\theta_1 n + \frac{1}{2}] - [\theta_2 n] - 1} = \psi(\alpha), \tag{17.92}$$

where X_i' is the ith order statistic among $X_1, \ldots, X_n$, $k_1 = [\theta_1 n + \frac{1}{2}] + 1$ and $k_2 = n - [\theta_2 n] - 1$. Simulations were carried out for the same values of the parameters as for $\tilde{\alpha}$, $\hat{\alpha}$, and a^*, with

$$(\theta_1, \theta_2) = (0.025, 0.075) \quad \text{and} \quad (0.0, 0.1).$$

The trimmed means gave good results with $(\theta_1, \theta_2) = (0.025, 0.075)$, and it was suggested that a trimmed mean, omitting the first few order statistics, might be used in place of the maximum likelihood estimator if a is thought to be small (giving rise to observations near zero).

8 RELATED DISTRIBUTIONS

If Y has the standard uniform (rectangular) distribution

$$p_Y(y) = 1, \qquad 0 \le y \le 1, \tag{17.93}$$

then Z = $-\log$ Y has the exponential distribution

$$p_Z(z) = e^{-z}, \qquad 0 \le z, \tag{17.94}$$

which is a special form of gamma distribution. If $Y_1, Y_2, \ldots, Y_k$ are independent random variables, each distributed as Y, and $Z_j = -\log Y_j$ $(j = 1, \ldots, k)$, then $Z_{(k)} = \sum_{j=1}^{k} Z_j$ has a gamma distribution with parameters a = k, $\beta = 1$, $\gamma = 0$. [$2Z_{(k)}$ is distributed as χ_{2k}^2; see Chapter 18.1 Relationships between gamma and beta distributions are described in Chapter 25 (see also Section 6 of this chapter).

Apart from noting these interesting relationships, we will devote this section to an account of classes of distributions that are related to gamma distributions, in particular

1. truncated gamma distributions,
2. compound gamma distributions,
3. transformed gamma distributions especially the generalized gamma distributions (which are assigned a special section of their own),
4. distributions of mixtures, sums, and products of gamma variables.

8.1 Truncated Gamma Distributions

The most common form of truncation of gamma distributions, when used in life-testing situations, is truncation from above. This is omission of values exceeding a **fixed** number τ, which is usually (though not always) known. If τ is not known, and the distribution before truncation is of the general form **(17.1)**, there are four parameters (a, β, y, τ) to estimate, and technical problems become formidable. However, it is not difficult to construct fairly simple (but quite likely not very accurate) formulas for estimating these parameters.

Fortunately it is often possible to assume that γ is zero in these situations [see Parr and Webster **(1965)** for examples], and we will restrict ourselves to this case. We will suppose that we have observations that can be regarded as observed values of independent random variables X,, $X_2, \ldots,$ X,,, each having the probability density function

$$\frac{x^{\alpha-1}e^{-x/\beta}}{\int_0^{\tau}t^{\alpha-1}e^{-t/\beta}\,dt}, \qquad 0 \le x \le \tau. \tag{17.95}$$

This may be denoted as a gamma (a, $\beta|\tau$) distribution. Estimation of the parameters a and β has been discussed by Chapman **(1956)**, Cohen (1950, **1951)**, Das **(1955)**, Des **Raj (1953)**, and Iyer and Singh (1963).

The moments of distribution (17.95) are conveniently expressed in terms of incomplete gamma functions:

$$\mu'_r(X) = \beta^r\Gamma_{\tau/\beta}(\alpha + r)/\Gamma_{\tau/\beta}(\alpha). \tag{17.96}$$

Gross (1971) notes that this is an increasing function of both a and β and, further, that μ'_{r+1}/μ'_r **is** an increasing function of a for r > 0.

The preceding results imply that

$$0 \le \mu'_1 \le \frac{\beta\tau}{\beta + 1} \qquad \text{for all a} > 0, \tag{17.97}$$

$$\frac{\mu'_{r+s}}{\mu'_r} \le (\beta + \text{r})\frac{\tau^s}{\beta + r + s} \qquad \text{for all a} > 0. \tag{17.98}$$

Nath (1975) obtained the minimum variance unbiased estimator of the reliability function $(R(t) = \Pr[X > t])$ for the gamma (a, $\beta|\tau$) distribution with integer a. In his analysis he showed that the sum of n independent

gamma$(\alpha, \beta|\tau)$ variables, $Y_n = X_1 + \cdots + X_n$ has cdf

$$p_{Y_n}(y|\alpha,\beta;\tau) = \frac{C^n\{\Gamma(\alpha)\}^n}{\beta^{n\alpha}\Gamma(n\alpha)}e^{-y/\beta}$$
$$\times \sum_{r=0}^{k_y}(-1)^r\binom{n}{r}(y-r\tau)^{n\alpha-1}\theta\left(n\alpha-1, \frac{\tau}{y-r\tau}\right)$$
$$\text{for } k_y\tau < y < (k_y+1)\tau;\ k = 0,1,\ldots,n-1, \quad (17.99)$$

with

$$\theta\left(n\alpha-1, \frac{\tau}{y-r\tau}\right) = \sum^{*} \frac{r!s(\mathbf{r})!}{\prod_{j=0}^{\alpha-1}(r_j)!\prod_{j=0}^{\alpha-1}(j!)^{r_j}}\binom{n\alpha-1}{s(\mathbf{r})}\left(\frac{\tau}{y-r\tau}\right)^{s(\mathbf{r})}, \quad (17.100)$$

where $s(\mathbf{r}) = \Sigma_{j=0}^{\alpha-1} jr_j$; the multinomial summation Σ^* is over all nonnegative $\mathbf{r} = (r_1,\ldots,r_{\alpha-1})$ satisfying $\Sigma_{j=0}^{\alpha-1} r_j = r$.

The formula for the MVU estimator of $R(t)$ appears to be extremely cumbersome, although Nath (1975) claims that "it is not so in practical application, particularly when the sample size is small."

As $\tau \to \infty$, the distribution of Y_n tends to gamma $(n\alpha, \beta)$, as is to be expected. The MVU estimator of $R(t)$ tends to the incomplete beta function ratio $I_{t/Y_n}(\alpha, (n-1)\alpha)$, corresponding to Basu's (1964) MVU estimator of $R(t)$, with corrected factorial term. [See also Wani and Kabe (1971), and for exponential distributions, Pugh (1963).]

8.2 Compound Gamma Distributions

Starting from (17.1), compound gamma distributions can be constructed by assigning joint distributions to α, β, and γ. The great majority of such distributions used in applied work start from (17.2) (i.e., with $\gamma = 0$) and assign a distribution to one of α and β (usually β).

If β^{-1} itself be supposed to have a gamma (δ, b^{-1}) distribution with

$$p_{\beta^{-1}}(x) = \frac{b^\delta x^{\delta-1}e^{-xb}}{\Gamma(\delta)}, \qquad 0 \le x, \qquad (17.101)$$

the resulting compound distribution has probability density function

$$p_X(x) = \frac{\Gamma(\alpha+\delta)}{\Gamma(\alpha)\Gamma(\delta)}x^{\alpha-1}(x+1)^{-(\alpha+\delta)}, \qquad 0 \le x. \qquad (17.102)$$