CHAPTER 25

Beta Distributions

1 DEFINITION

The family of beta distributions is composed of all distributions with probability density functions of form:

$$p_Y(y) = \frac{1}{B(p,q)} \frac{(y-a)^{p-1}(b-y)^{q-1}}{(b-a)^{p+q-1}}, \qquad a \le y \le b, \quad (25.1)$$

with $p > 0$, $q > 0$. It is denoted beta (p, q). This will be recognized as a Pearson Type I or II distribution (see Chapter 12, Section 4.1). If $q = 1$, the distribution is sometimes called a ***power-function*** distribution.

If we make the transformation

$$X = \frac{Y-a}{b-a},$$

we obtain the probability density function

$$p_X(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}, \qquad 0 \le x \le 1. \qquad (25.2)$$

This is the ***standard form*** of the beta distribution with ***parameters*** p, q. It is the form that will be used in most of this chapter. The ***standard power-function*** density is

$$p_X(x) = px^{p-1}, \qquad 0 \le x \le 1. \qquad (25.2)'$$

Harter (1978) introduced the family of symmetric $(p = q)$ ***standardized*** beta variables with the density function

$$p_X(x) = \left[\frac{\Gamma(2p)}{\Gamma^2(p)\left(2\sqrt{2p+1}\right)^{2p-1}}\right](2p+1-x^2)^{p-1}, \qquad (25.3)$$

$$-\sqrt{2p+1} \le x \le \sqrt{2p+1}.$$

Of course $E[X] = 0$ and $\operatorname{var}(X) = 1$. For $p = 1.5(0.5)4.0$ he provides explicit formulas for the cdfs. The simplest, for $p = 2$, is

$$F_X(x) = \left(\frac{3\sqrt{5}}{100}\right)\left(5x - \frac{1}{3}x^3\right) + \frac{1}{2}, \qquad -\sqrt{5} \le x \le \sqrt{5}. \qquad (25.4)$$

The probability density function of a symmetric beta distribution with parameter p, mean μ, and standard deviation σ is

$$p_X(x) = \frac{\Gamma(2p)}{\sigma\{\Gamma(p)\}^2\left(2\sqrt{2p+1}\right)^{2p-1}}\left[2p+1-\left(\frac{x-\mu}{\sigma}\right)^2\right]^{p-1}, \qquad (25.5)$$

$$\mu - \sigma\sqrt{2p+1} \le x \le \mu + \sigma\sqrt{2p+1}.$$

The probability integral of the distribution (25.2) up to x is the ***incomplete beta function ratio,*** and it is denoted by $I_x(p,q)$ so that

$$I_x(p,q) = \frac{1}{B(p,q)}\int_0^x t^{p-1}(1-t)^{q-1}\,dt \qquad (25.6)$$

The word "ratio," which distinguishes *(25.6)* from the ***incomplete beta function***,

$$B_x(p,q) = \int_0^x t^{p-1}(1-t)^{q-1}\,dt, \qquad (25.7a)$$

is often omitted. A description of the properties of $I_x(p,q)$ is contained in Chapter 1 (Section A5) and in Chapter 3 (Section *6*).

Dutka *(1981)* provides a detailed account of the history of $B_x(p,q)$ and $I_x(p,q)$, tracing it back to *1676,* in a letter from Isaac Newton to Henry Oldenberg. The formula given is the special case of

$$B_x(p,q) = p^{-1}x^p(1-x)^q{}_2F_1(p+q,1;p+1;x), \qquad (25.7b)$$

where ${}_2F_1(\cdot)$ denotes the Gaussian hypergeometric function defined in *Eq. (1.104)* of Chapter *1.*

2 GENESIS AND RANDOM NUMBER GENERATION

in "normal theory" the beta distribution arises naturally as the distribution of $V^2 = X_1^2/(X_1^2 + X_2^2)$, where X_1^2, X_2^2 are independent random variables, and X_j^2 is distributed as χ^2 with ν_j degrees of freedom $(j = 1, 2)$ (see Chapter 18). The distribution of V^2 is then a standard beta distribution, as in (25.2), with $p = \frac{1}{2}\nu_1$, $q = \frac{1}{2}\nu_2$. Generally $Y = W_1/(W_1 + W_2)$ has a standard beta distribution with parameters p_1 and p_2 if W_j has the gamma density (see Chapter 17) with parameters (p_j, β) $(j = 1, 2)$ (for any $\beta > 0$).

Notice that V^2 and $(X_1^2 + X_2^2)$ are mutually independent. An extension of this result is that if $X_1^2, X_2^2, \cdots, X_k^2$ are mutually independent with X_j^2 distributed as χ^2 with ν_j degrees of freedom $(j = 1, 2, \dots, k)$ (see Chapter 18), then

$$V_1^2 = \frac{X_1^2}{X_1^2 + X_2^2}$$

$$V_2^2 = \frac{X_1^2 + X_2^2}{X_1^2 + X_2^2 + X_3^2}$$

$$\vdots$$

$$V_{k-1}^2 = \frac{X_1^2 + \cdots + X_{k-1}^2}{X_1^2 + \cdots + X_k^2}$$

are mutually independent random variables, each with a beta distribution, the values of p, q for V_j^2 being $\frac{1}{2}\sum_{i=1}^{j}\nu_i$, $\frac{1}{2}\nu_{j+1}$, respectively. Under these conditions the product of any consecutive set of V_j^2's also has a beta distribution [see Jambunathan (1954) and Section 8]. This property also holds when the v's are any positive numbers (not necessarily integers). Kotlarski (1962) has investigated general conditions under which products of independent variables have a beta distribution.

The special standard beta distribution with $p = q = \frac{1}{2}$ [known as the arc-sine distribution because $\Pr[X \le x] = (2/\pi)\sin^{-1}\sqrt{x}$ for $0 \le x \le 1$] arises in an interesting way in the theory of "random walks." Suppose that a particle moves along the real line by steps of unit length, starting from zero, it being equally likely that a step will be to the left (decreasing) or right (increasing). Let the random variable T_{2n} denote the number of times in the first 2n steps for which the point is in the interval 0 to 2n inclusive at the conclusion of a step. Then

$$\Pr[T_{2n} = 2k] = \binom{2k}{k}\binom{2n-2k}{n-k}2^{-2n}, \qquad k = 0, 1, \dots, n.$$

The ratio $T_{2n}/(2n)$ can be regarded as the fraction of time spent on the positive part of the real line. As n tends to infinity, the limiting distribution of $T_{2n}/(2n)$ is the arc-sine distribution:

$$\lim_{n\to\infty}\left\{\sum_{k\le nx}\Pr[T_{2n}=2k]\right\}=\frac{1}{\pi}\int_0^x t^{-1/2}(1-t)^{-1/2}\,dt=\left(\frac{2}{\pi}\right)\sin^{-1}\sqrt{x}. \tag{25.8}$$

Standard beta distributions with $p+q=1$, but $p \neq \frac{1}{2}$, are sometimes called generalized arc-sine distributions. For more details on the arc-sine distribution, see Section 7.

A beta distribution can also be obtained as the limiting distribution of eigenvalues in a sequence of random matrices. Suppose that $\mathbf{A}_n$ to be a symmetric $n \times n$ matrix whose elements a_{ij} $(i \le j)$ are independent random variables, all a_{ij}'s with $i \neq j$ having a common distribution, and all a_{ii}'s another common distribution, both distributions being symmetrical about zero with variance σ^2 and with all absolute moments finite. Under these conditions Wigner (1958) has shown that the proportion of eigenvalues of the "normalized" matrix $(2\sigma\sqrt{n})^{-1}\mathbf{A}_n$, which are less than x, tends to the limit

$$2\pi^{-1}\int_{-1}^{x}\sqrt{1-t^2}\,dt$$

as $n \to \infty$. This is of form (25.1) with $a=-1$, $b=1$, $p=q=3/2$. Arnold (1967) has shown that this result holds under much weaker conditions on the distributions of the a_{ij}'s.

A class of distributions that includes the beta $(\frac{1}{2},\frac{1}{2})$ and beta $(2,2)$ distributions, can be generated by the following procedure: Starting with the interval $(0,1)$, observe the value of a random variable X_1 distributed uniformly over $(0,1)$ [i.e., as beta $(1,1)$]. Then choose one of the two subintervals $(0,X_1),(X_1,1)$, with probabilities p, $1-p$ of choosing the longer or shorter one, respectively. Denoting the chosen interval by (L_1,U_1), then observe the value of a random variable X_2, uniformly distributed over (L_1,U_1), and choose as (L_2,U_2), the longer or shorter of the intervals $(L_1,X_2),(X_2,U_1)$ with probabilities p, $1-p$, respectively. Continue in this way, choosing (L_{n+1},U_{n+1}) as the longer or shorter of $(L_n,X_{n+1}),(X_{n+1},U_n)$ with probabilities p, $1-p$, respectively. It is easy to see that as $n\to\infty$, the interval length (U_n-L_n) tends to zero with probability one, and there is a limiting value Y_p, say, to which L_n and U_n tend.

The distribution of $Y_{1/2}$ is beta $(\frac{1}{2},\frac{1}{2})$ [Chen, Lin, and Zame (1981)], and the distribution of Y_1 is beta $(2,2)$ [Chen, Goodman, and Zame (1984)]. It is natural to conjecture that Y_p has an approximate (but not exact) beta distribution for values of p other than $\frac{1}{2}$ or 1. Johnson and Kotz (1994) show that

$$\operatorname{var}(Y_p) = \frac{2(7-6p)}{4(11-6p)} \tag{25.9}$$

and that

$$\beta_2(Y_p) = \frac{3(11-6p)(151-204p+60p^2)}{(7-6p)^2(79-30p)}. \tag{25.10}$$

If Y_p had a beta (α, α) distribution, the value of α giving the correct value for $\operatorname{var}(Y_p)$ would be $2(7-6p)^{-1}$. This would result in a "nominal" value

$$\beta_2 = 3(11-6p)(25-18p)^{-1}. \tag{25.11}$$

Table 25.1 compares values of β_2 from (25.10) (actual) and (25.11) (nominal), for selected values of p.

The agreement between actual and nominal values supports the conjecture that beta $[2(7-6p)^{-1}, 2(7-6p)^{-1}]$ would be a good approximation to the distribution of Y_p. O'Connor, Hook, and O'Connor (1985) came to the same conclusion on the basis of simulations.

Another procedure leading to limiting beta distributions has been described by Kennedy (1988). Values of k independent variables $Z_{n1}, \cdots, Z_{nk}$ each uniformly distributed over (L_n, U_n) are observed. The interval (L_{n+1}, U_{n+1}) is then chosen as $(L_n, \max(Z_{n1}, \cdots, Z_{nk}))$,

$$(\min(Z_{n1}, \cdots, Z_{nk}), U_n), \quad \text{or} \quad (\min(Z_{n1}, \cdots, Z_{nk}), \max(Z_{n1}, \cdots, Z_{nk}))$$

with probabilities p, q, r, respectively $(p+q+r=1)$. Kennedy (1988) showed that if the initial interval is $(0,1)$, the limit to which both L_n and U_n converge (with probability 1) is distributed as beta $(k(p+r), k(q+r))$ over $(0,1)$. [Of course, if the initial interval is (A, B), the limit distribution is beta $(k(p+r), k(q+r))$ over (A, B).] There is an alternative proof, based on moment calculations, in Johnson and Kotz (1993).

Yet another way in which a beta distribution arises is as the distribution of an ordered variable from a rectangular distribution (Chapter 26). If $Y_1, Y_2, \cdots, Y_n$ are independent random variables each having the standard

Table 25.1 Actual and Nominal Values of β_2

p	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Actual β_2	1.287	1.315	1.348	1.388	1.438	1.500	1.580	1.687	1.831	2.019	2.143
Nominal β_2	1.320	1.345	1.374	1.408	1.449	1.500	1.563	1.645	1.754	1.909	2.143

rectangular distribution so that

$$p_{Y_j}(y) = 1, \qquad 0 \le y \le 1, \tag{25.12}$$

and the corresponding order statistics are $Y_1' \le Y_2' \le \cdots \le Y_n'$, the sth-order statistic Y_s' has the beta distribution

$$p_{Y_s'}(y) = [B(s, n-s+1)]^{-1} y^{s-1}(1-y)^{n-s}, \qquad 0 \le y \le 1. \tag{25.13}$$

Fox (1963) suggested that this result may be used to generate beta-distributed random variables from standard rectangularly distributed variables. By this method only integer values can be obtained for n and n − s. A method applicable for fractional values of n and $n - s$ was constructed by Johnk (1964). He showed that if X and Y are independent standard rectangular variables, then the conditional distribution of $X^{1/n}$, given that $X^{1/n} + Y^{1/r} \le 1$, is a standard beta distribution with parameters n, r + 1, and the conditional distribution of $Y^{1/r}$ is beta with parameters n + 1 and r.

This process involves the calculation of $X^{1/n}$ and $Y^{1/r}$, which may be awkward. If n and/or r are large, then a large number of pairs of values (X, Y) is likely to be needed to ensure $X^{1/n} + Y^{1/r} \le 1$, as pointed out by Pekh and Marchenko (1992). [In fact $\Pr[X^{1/n} + Y^{1/r} \le 1] < 1 - \Pr[X^{1/n} > \frac{1}{2}]\Pr[Y^{1/r} > \frac{1}{2}] < 2^{-n} + 2^{-r}$. Hence, if $\min(n, r) \ge 11$, $\Pr[X^{1/n} + Y^{1/r} \le 1]$ is less than 0.001.] Bánkövi (1964) has suggested a method whereby these calculations may be avoided if n and r are both rational. This consists of selecting integers $a_1, a_2, \dots, a_M$, $b_1, b_2, \dots, b_N$ such that

$$n = \sum_{j=1}^{M} a_j^{-1}, \qquad r = \sum_{j=1}^{N} b_j^{-1}.$$

Then using the fact that if $X_1, X_2, \dots, X_M, Y_1, \dots, Y_N$ are independent standard rectangular variables, $\max(X_1^{a_1}, X_2^{a_2}, \dots, X_M^{a_M})$ and $\max(Y_1^{b_1}, Y_2^{b_2}, \dots, Y_N^{b_N})$ are distributed as $X^{1/n}$, $Y^{1/r}$, respectively.

If n (or r) is not a rational fraction, it may be approximated as closely as desired by such a fraction. Bánkövi has investigated the effects of such approximation on the desired beta variates. The GR method is based on the property that $X = Y/(Y+Z)$ has a beta (p, q) distribution if Y and Z are independent gamma variables with shape parameters p and q, respectively (see the beginning of this section).

Generation of beta random variables based on acceptance/rejection algorithms was studied by Ahrens and Dieter (1974) and Atkinson and Pearce (1976), among others. The latter authors recommend the Forsythe (1972) method, which was originally applied to generate random normal deviates. Chen (1978) proposed a modified algorithm BA: $(p, q > 0)$.

Initialization: Set $\alpha = p + q$. If $\min(p, q) \leq 1$, set $\beta = \max(p^{-1}, q^{-1})$; otherwise set $\beta = \sqrt{\{(\alpha - 2)/(2pq - \alpha)\}}$. Set $\gamma = p + \beta^{-1}$.

1. Generate uniform (0, 1) random numbers U_1, U_2, and set $V = \beta \log\{U_1/(1 - U_1)\}$, $W = pe^V$.
2. If $\alpha \log\{\alpha/(q + W)\} + \gamma V - 1.3862944 < \log(U_1^2 U_2)$, go to 1.
3. Deliver $X = W/(q + W)$.

This algorithm is reasonably fast for values of p and q down to about 0.5. More complicated versions (*BB*)(*BC*), also described by Chen (1978), cover all a, b > 0 and offer quicker variate generation speed. Here is

Algorithm BB ($\min(p_0, q_0) > 1$)
Initialization: Set $p = \min(p_0, q_0)$, $q = \max(p_0, q_0)$, $\alpha = p + q$, $\beta = \sqrt{[(\alpha - 2)/(2pq - \alpha)]}$; $\gamma = p + \beta^{-1}$.

1. Generate uniform (0, 1) random numbers U_1, U_2, and set $V = \beta \log\{U_1/(1 - U_1)\}$, $W = pe^V$, $Z = U_1^2 U_2$, $R = \gamma^V - 1.3862944$, $S = p + R - W$.
2. If $S + 2.609438 \geq 5Z$, go to 5.
3. Set $T = \log Z$. If $S \geq T$, go to 5.
4. If $R + \alpha \log\{\alpha/(q + W)\} < T$, go to 1.
5. If $p = p_0$, deliver $X = W/(q + W)$; otherwise deliver $X = q/(q + W)$.

Schmeiser and Shalaby (1980) developed three exact methods applicable for $\min(p, q) > 1$ (corresponding to Chen's BB algorithm). One of the methods is a minor modification of the Ahrens and Dieter (1974) algorithm: BNM. All the methods use the property that points of inflexion of the beta density are at

$$x = \frac{(p-1) \pm [(p-1)(q-1)]/(p+q-3)^{1/2}}{p+q-2}$$

if these values lie between zero and one, and are real.

A detailed comparison of the various methods, carried out by Schmeiser and Shalaby (1980), shows that BB is the fastest for heavily skewed distributions but yields to BNM for heavy-tailed symmetric distributions. No algorithm does better than BB for the following values of the parameters:

$p = 1.01$	$q = 1.01, 1.50, 2.00, 5.00, 10.00, 100.00$
$p = 1.50$	$q = 1.50, 2.00, 5.00, 10.00, 100.00$
$p = 2.00$	$q = 2.00$
$p = 5.00$	$q = 5.00$
$p = 10.00$	$q = 10.00$
$p = 100.00$	$q = 100.00$

Devroye (1986) contains summaries of methods of generating random variables with beta distributions.

3 PROPERTIES

If X has the standard beta distribution (25.2), its rth moment about zero is

$$\mu'_r = \frac{B(p+r,q)}{B(p,q)} = \frac{\Gamma(p+r)\Gamma(p+q)}{\Gamma(p)\Gamma(p+q+r)}$$

$$= \frac{}{(p+q)^{[r]}} \qquad \text{(if } r \text{ is an integer)}, \tag{25.14}$$

where $y^{[r]} = y(y+1)\cdots(y+r-1)$ is the ascending factorial. In particular

$$E[X] = \frac{p}{p+q}, \tag{25.15a}$$

$$\operatorname{var}(X) = pq(p+q)^{-2}(p+q+1)^{-1}, \tag{25.15b}$$

$$\alpha_3(X) = \sqrt{\beta_1(X)}$$

$$= 2(q-p)\sqrt{p^{-1}+q^{-1}+(pq)^{-1}} \cdot (p+q+2)^{-1}, \tag{25.15c}$$

$$\alpha_4(X) = \beta_2(X) = 3(p+q+1)\{2(p+q)^2 + pq(p+q-6)\}$$
$$\times [pq(p+q+2)(p+q+3)]^{-1}, \tag{25.15d}$$

$$E[X^{-1}] = (p+q-1)(p-1)^{-1}, \tag{25.15e}$$

$$E\left[(1-X)^{-1}\right] = (p+q-1)(q-1)^{-1}. \tag{25.15f}$$

Pham-Gia (1994) has recently established some simple bounds for $\operatorname{var}(X)$. Specifically, he has shown that $\operatorname{var}(X) < 1/4$, and if the density of X is unimodal (i.e., $p > 1$ and $q > 1$) then $\operatorname{var}(X) < 1/12$; further, if the density of X is U-shaped (i.e., $p < 1$ and $q < 1$), then he has proved that $\operatorname{var}(X) > 1/12$.

Writing $\lambda = (p+q)^{-1}$ and $\theta = p(p+q)^{-1}$, we have the following recurrence relation among the central moments of the standard beta distribution:

$$\mu_{s+1} = -\frac{s\lambda}{1+s\lambda}\mu_s + \theta\sum_{j=1}^{s}\binom{s}{j}\frac{\lambda^j(1-\theta)^j j!}{(1+s\lambda)\cdots(1+[s-j]\lambda)}\mu_{s-j}, \tag{25.16}$$

with $\mu_0 = 1$, μ_1 $(= E[X - E[X]]) = 0$, $\mu_2 = \lambda\theta(1-\theta)/(1+\lambda)$,

$$\mu_3 = \frac{2\lambda^2\theta(1-\theta)}{(1+\lambda)(1+2\lambda)}(1-2\theta) \qquad \text{[Mühlbach (1972)]}.$$

The moment-generating function can be expressed as a confluent **hypergeo**metric function [Eq. (1.1211, Chapter 1]:

$$E[e^{tX}] = M(p; p+q; t) \tag{25.17}$$

and, of course, the characteristic function is $M(p; p+q; it)$.

The moment-generating function of $(-\log X)$, where X is a standard beta, is

$$M(t) = E[\exp(-t \log X)] = \frac{B(p-t,q)}{B(p,q)}, \tag{25.17'}$$

and the corresponding cumulant-generating function is

$$K(t) = \log\left[\frac{\Gamma(p+q)}{\Gamma(p)}\right] - \log\left[\frac{\Gamma(p+q-t)}{\Gamma(p-t)}\right].$$

The cumulants are

$$\kappa_r = (r-1)! \sum_{j=0}^{q-1} (p+j)^{-r}, \qquad r = 1,2,\ldots, \tag{25.17''}$$

if q is an integer. In the general case

$$\kappa_r = (-1)^r\left[\psi^{(r-1)}(p) - \psi^{(r-1)}(p+q)\right], \tag{25.17'''}$$

where $\psi^{(r-1)}(x) = (d^r/dx^r)\log\Gamma(x)$ is the $(r+1)$-gamma function (see Chapter 1, Section A2).

The mean deviation of X is

$$\delta_1(X) = E(|X - E[X]|) = \frac{2}{B(p,q)} \frac{p^p q^q}{(p+q)^{p+q+1}}. \tag{25.18a}$$

If p = q, the expression reduces to

$$\tilde{\delta}_1(X) = \left[B(p,p)p2^{2p}\right]^{-1}. \tag{25.18b}$$

The authors thank Dr. T. Pham-Gia for pointing out an error in the expression for $\delta_1(X)$ which appeared in the first edition of this volume. [See also Pham-Gia and Turkkan (1992).]

For p and q large, using Stirling's approximation to the gamma function, the mean deviation is approximately

$$\sqrt{\frac{2pq}{\pi(p+q)}}\cdot\frac{1}{p+q}\left\{1+\frac{1}{12}(p+q)^{-1}-\frac{1}{12}p^{-1}-\frac{1}{12}q^{-1}\right\}, \quad (25.19)$$

and

$$\frac{\text{Mean deviation}}{\text{Standard deviation}} \doteqdot \sqrt{\frac{2}{\pi}}\left\{1+\frac{7}{12}(p+q)^{-1}-\frac{1}{12}p^{-1}-\frac{1}{12}q^{-1}\right\}.$$

The mean deviation about the median (m) is

$$\frac{2m^p(1-m)^q}{(p+q)B(p,q)} = 2\operatorname{var}(X)\left\{\frac{1}{B(p+1,q+1)}m^p(1-m)^q\right\}. \quad (25.20)$$

If $p>1$ and $q>1$, then $p_X(x)\to 0$ as $x\to 0$ or $x\to 1$; if $0<p<1$, $p_X(x)\to\infty$ as $x\to 0$; and if $0<q<1$, $p_X(x)\to\infty$ as $x\to 1$. If $p=1$ $(q=1)$, $p_X(x)$ tends to a finite nonzero value as $x\to 0$ (1).

If $p>1$ and $q>1$, the density function has a single mode at $x=(p-1)/(p+q-2)$. If $p<1$ and $q<1$, there is an **antimode** (minimum value) of $p_X(x)$ at this value of x. Such distributions are called U-shaped beta (or Type I or II) distributions. If $(p-1)(q-1)$ is not positive, the probability density function does not have a mode or an **antimode** for $0<x<1$. Such distributions are called J-shaped or reverse J-shaped beta (or Type I) distributions. [Peleg and Normand (1986) advocate using the reparametrization $am = p-1$, $m = q-1$ so that the mode is at $a/(a+1)$ and does not depend on m. Although they call this a *modified* beta distribution, it is in fact just a regular beta distribution that is differently parametrized.] If $p=q$, the distribution is symmetrical about $x=\frac{1}{2}$.

For all positive values of p and q, there are points of inflexion at

$$\frac{p-1}{p+q-2}\pm\frac{1}{p+q-2}\sqrt{\frac{(p-1)(q-1)}{p+q-3}} \quad (25.21)$$

provided these values are real and lie between 0 and 1. Note that as for all Pearson curves, the points of inflexion are equidistant from the modes.

The expected value $p/(p+q)$ depends on the ratio p/q. If this ratio is kept constant, but p and q are both increased, the variance decreases, and the (standardized) distribution tends to the unit normal distribution. Some of the properties of beta distributions described in this section are shown in Figures 25.1*a*, *b*. Note that if the values of p and q are interchanged, the distribution is "reflected" about $x=\frac{1}{2}$.

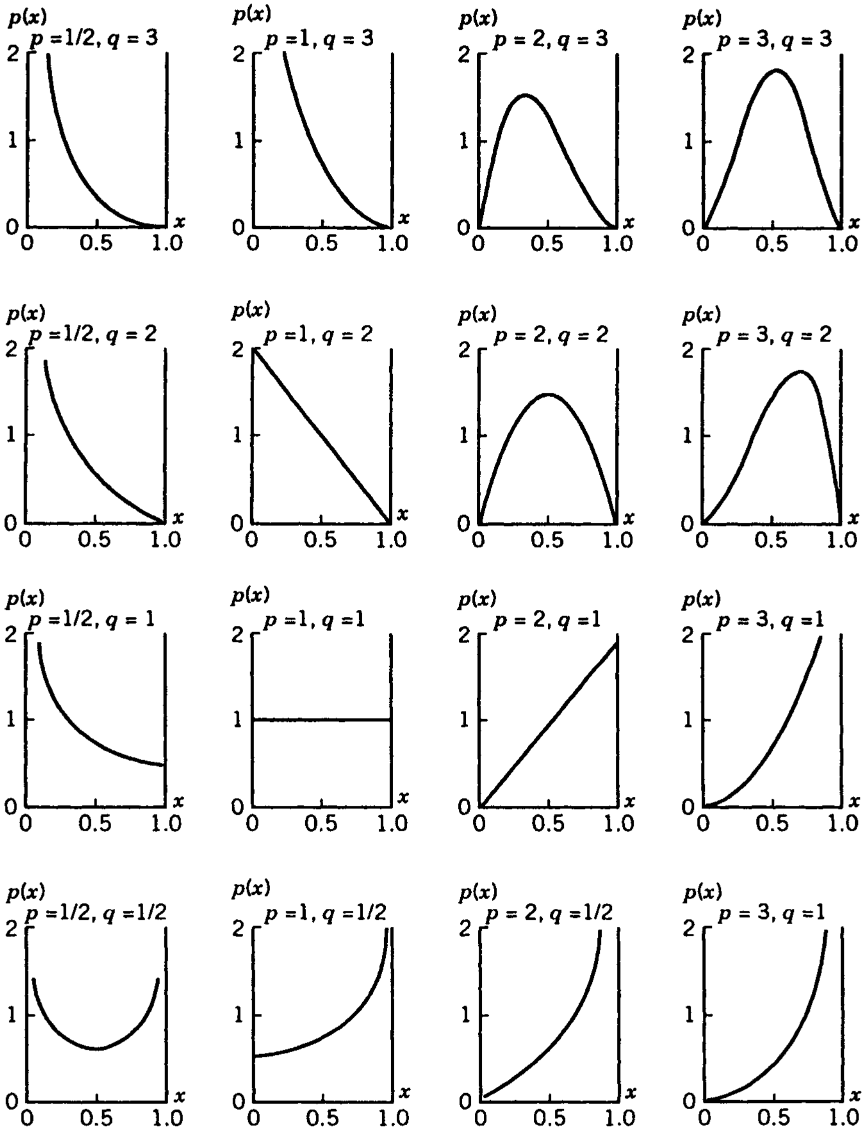

Figure 25.1a Beta density functions

The Lorenz curve [see Chapter 12, Eq. (12.1611 has coordinates

$$[I_x(p,q), I_x(p+1,q)],$$

and the Gini index [Chapter 12, Eq. (12.911 is

$$\frac{2B(2p,2q)}{p[B(p,q)]^2}. \tag{25.22}$$

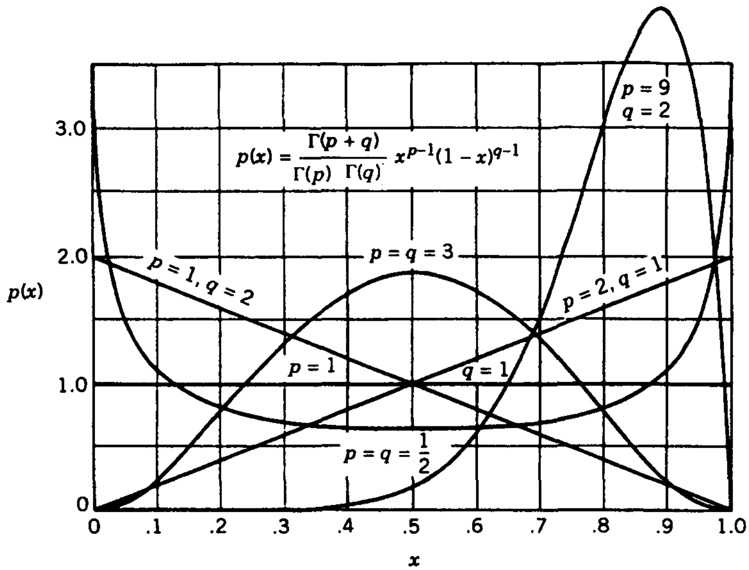

Figure 25.1b Beta density functions

4 ESTIMATION

Discussion of parameter estimation for beta distributions goes back to Pearson's classical paper of 1895 where the method of moments was introduced. Direct algebraic solution of the ML equations cannot be obtained for beta distributions. Koshal (1933, 1935) tackled the ML estimation of four-parameter beta distributions, approximating the actual ML parameter estimates by an interactive method using estimates derived by the method of moments as initial values.

Estimation of all four parameters in distribution (25.1) can be effected by equating sample and population values of the first four moments. Calculation of a, b, p, and q from the mean μ'_1 and central moments μ_2, μ_3, μ_4 uses the following formulas [Elderton and Johnson (1969)]. Putting

$$r = \frac{6(\beta_2 - \beta_1 - 1)}{6 + 3/31 - 2/32},$$

then

$$p, q = \tfrac{1}{2} r \left\{ 1 \pm (r+2) \sqrt{\beta_1 \{ (r+2)^2 \beta_1 + 16(r+1) \}^{-1}} \right\} \qquad (25.23)$$

with $p \lessgtr q$ according as $\alpha_3 = \sqrt{\beta_1} \gtrless 0$. Also

$$\frac{p-1}{q-1} = \frac{\text{mode}(Y) - a}{b - \text{mode}(Y)} \tag{25.24}$$

[where mode $(Y) = a + (b-a)(p-1)/(p+q-2)$] and

$$b - a = \tfrac{1}{2}\sqrt{\mu_2}\sqrt{(r+2)^2\beta_2 + 16(r+1)}\,. \tag{25.25}$$

If the values of a and b are known, then only the first and second moments need to be used, giving

$$\mu'_1 = a + (b-a)p/(p+q),$$

$$\mu_2 = (b-a)^2 pq(p+q)^{-2}(p+q+1)^{-1},$$

whence

$$\frac{\mu'_1 - a}{b - a} = \frac{p}{p+q}, \tag{25.26}$$

$$\frac{\mu_2}{(b-a)^2} = \frac{p}{p+q}\left(1 - \frac{p}{p+q}\right)\frac{1}{p+q+1}. \tag{25.27}$$

Thus

$$p + q = \frac{[(\mu'_1 - a)/(b-a)][1 - (\mu'_1 - a)/(b-a)]}{\left(\mu_2/(b-a)^2\right)} - 1, \tag{25.28}$$

$$p = \left(\frac{\mu'_1 - a}{b-a}\right)^2\left(1 - \frac{\mu'_1 - a}{b-a}\right)\left(\frac{\mu_2}{(b-a)^2}\right)^{-1} - \frac{\mu'_1 - a}{b-a}. \tag{25.29}$$

The existence, consistency, and asymptotic normality and efficiency of a root of the likelihood equations are usually proved under conditions similar to those given by Cramér (1946) or Kulldorff (1957), which, among other things, allow Taylor expansion of the derivative of the log-likelihood function in a fixed neighborhood of the true parameter value.

When it is necessary to estimate at least one of the end points (a or b) of the four-parameter beta distribution, no such fixed neighborhood of Taylor expansion validity exists. But if the true shape parameters (p and q) are large enough (> 2, regular case), Whitby (1971) has shown that the conditions can be weakened to allow Taylor expansion in a sequence of shrinking neighborhoods, and the usual asymptotic results, with $n^{1/2}$ normalization, can be proved.

If a and b are known and $Y_1, Y_2, \cdots, Y_n$ are independent random variables each having distribution (25.1), the maximum likelihood equations for estimators $\hat{p}, \hat{q}$ of p, q, respectively are

$$\psi(\hat{p}) - \psi(\hat{p} + \hat{q}) = n^{-1} \sum_{j=1}^{n} \log\left(\frac{Y_j - a}{b - a}\right), \qquad (25.30a)$$

$$\psi(\hat{q}) - \psi(\hat{p} + 4) = n^{-1} \sum_{j=1}^{n} \log\left(\frac{b - Y_j}{b - a}\right), \qquad (25.30b)$$

where $\psi(\cdot)$ is the digamma function [Eq. (1.371, Chapter 1]. The Cramér and Kulldorff conditions cover this case and Eqs. (25.30a) and (25.30b) have to be solved by trial and error. If $\hat{p}$ and $\hat{q}$ are not too small, the approximation

$$\psi(t) \doteqdot \log(t - \tfrac{1}{2})$$

may be used. Then approximate values of $(\hat{p} - \frac{1}{2})/(\hat{p} + \hat{q} - \frac{1}{2})$ and $(\hat{q} - \frac{1}{2})/(\hat{p} + \hat{q} -$ \$1 can be obtained from (25.30a) and (25.30b), whence follow, as first approximations to p and q,

$$\hat{p} \doteqdot \frac{\frac{1}{2}\left\{1 - \Pi_{j=1}^{n}\left((b - Y_j)/(b - a)\right)^{1/n}\right\}}{1 - \Pi_{j=1}^{n}\left((Y_j - a)/(b - a)\right)^{\prime\prime\prime\prime} - \Pi_{j=1}^{n}\left((b - Y_j)/(b - a)\right)^{1/n}}, \qquad (25.31a)$$

$$\hat{q} \doteqdot \frac{\frac{1}{2}\left\{1 - \Pi_{j=1}^{n}\left((Y_j - a)/(b - a)\right)^{1/n}\right\}}{1 - \Pi_{j=1}^{n}\left((Y_j - a)/(b - a)\right)^{1/n} - \Pi_{j=1}^{n}\left((b - Y_j)/(b - a)\right)^{1/n}}. \qquad (25.31b)$$

Starting from these values, solutions of (25.30a) and (25.30b) can be obtained by an iterative process. Gnanadesikan, Pinkham, and Hughes (1967) give exact numerical solutions for a few cases.

The asymptotic covariance matrix of $\sqrt{n}\,\hat{p}$ and $\sqrt{n}\,\hat{q}$ (as $n \to \infty$) is

$$[\psi'(p)\psi'(q) - \psi'(p+q)\{\psi'(p) + \psi'(q)\}]^{-1}$$
$$\times \begin{pmatrix} \psi'(q) - \psi'(p+q) & \psi'(p+q) \\ \psi'(p+q) & \psi'(p) - \psi'(p+q) \end{pmatrix}. \qquad (25.32)$$

Introducing approximations for $\psi'(\cdot)$, we have for large values of p and q,

$$\operatorname{var}(\hat{p}) \doteqdot p(2p-1)n^{-1},$$

$$\operatorname{var}(\hat{q}) \doteqdot q(2q-1)n^{-1},$$

$$\operatorname{corr}(\hat{p},\hat{q}) \doteqdot \sqrt{(1-2p^{-1})(1-2q^{-1})}\,. \tag{25.33}$$

Fielitz and Myers (1975, 1976) and Romesburg (1976), in brief communications, discuss the comparative advantages and disadvantages of the method of moments versus the maximum likelihood method for estimating parameters p and q. The difficulties involved in the maximum likelihood method are related to employing efficient search procedures to maximize the likelihood function. The Newton-Raphson method is extremely sensitive to the initial values of $\tilde{p}$ and $\tilde{q}$, and there is no guarantee that convergence will be achieved. Fielitz and Myers (1976) point out that for the sample problem considered by Gnanadesikan, Pinkham, and Hughes (1967), the method of moments yield more accurate estimates of p and q than does the method of maximum likelihood, possibly due to bias introduced by the computational method used in determining the ML estimators.

Beckman and Tietjen (1978) have shown that the equations (25.30a) and (25.30b) can be reduced to a single equation for $\hat{q}$ alone:

$$\psi(\hat{q}) - \psi\{[\psi^{-1}[\log G_1 - \log G_2 + \psi(\hat{q})]] + \hat{q}\} - \log G_2 = 0, \tag{25.34a}$$

where

$$G_1 = \prod_{j=1}^{n}\left(\frac{Y_j - a}{b - a}\right)^{1/n},$$

$$G_2 = \prod_{j=1}^{n}\left(\frac{b - Y_j}{b - a}\right)^{1/n}.$$

Having solved (25.34a) for $\hat{q}$, the estimator, $\hat{p}$, of p is calculated from

$$\hat{p} = \psi^{-1}[\log G_1 - \log G_2 + \psi(\hat{q})]\,. \tag{25.34b}$$

Lau and Lau (1991) provide a detailed investigation of methods of calculating good *initial* estimators p,, q, of p and q, respectively.

For $G_1 + G_2 = G_T \le 0.95$ they recommend

$$\log p_e = -3.929 + 10.523G_2 - 3.026G_1^3 + 1.757\exp\left(G_2\sqrt{G_1}\right) \tag{25.35a}$$

and

$$\log q_e = -3.895 + 1.222\sqrt{G_2} - 6.9056G_1^3 + 39.057G_1^2G_1^3 + 1.5318\exp(G_T). \tag{25.35b}$$

But for $0.95 \le G_T \le 0.999$ they suggest

$$\log p_e = 110706.79 + 3.0842\sqrt{G_1} + 110934.01\log G_T + 6.3908\exp(G_1G_2^2) - 233851.3G_T + 45300.7\exp(G_T) \tag{25.35c}$$

and

$$\log q_e = 113753.4 - 2.1G_1^2 + 113979.94\log G_T + 2.154G_1G_2^6 - 240149.9G_T + 46500.7\exp(G_T). \tag{25.35d}$$

They also study the sampling distribution of the ML estimators $\hat{p}$ and $\hat{q}$ and provide a table of the sample values of percentage bias $d = 100 \times (m - p)/p$, where $m = \Sigma p_e/K$ [computed for $K = 1000$ values of n, p, q ($n = 30, 60, 100$, $p(= q) = 2, 6, 10, 20$, and 40)], skewness $K^{-1}\Sigma(p_e - m)^3/S^3 = a$, and kurtosis $b_2 = K^{-1}\Sigma(p_e - m)^4/S^4$, where $S^2 = \Sigma(p_e - m)^2/K$.

For $p = q = 10$ representative values are

$n =$	30	100
d	11.1%	3.1%
a_1	1.17	0.59
b_2	5.6	3.7

The same authors also provide a procedure for estimating a confidence interval for p,, using Bowman and Shenton's (1979a, 1979b) method for calculating fractiles of distributions belonging to Pearson's system.

If a and b are unknown, and maximum likelihood estimators of a, b, p, and q are required, the procedure based on (25.31a) and (25.31b) can be repeated using a succession of trial values of a and b, until the pair (a, b), for which the maximized likelihood (given a and b) is as great as possible, is attained.

Carnahan (1989) investigated in detail maximum likelihood estimation for four-parameter beta distributions. He adds to (25.31a) and (25.31b) the additional ML equations

$$\frac{1}{n(p-1)} \cdot \frac{\partial \log L}{\partial a} = \frac{p+q-1}{p-1} - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{b-a}{Y_i-a}\right) = 0, \quad (25.31c)$$

and

$$-\frac{1}{n(q-1)} \cdot \frac{\partial \log L}{\partial b} = \frac{p+q-1}{q-1} - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{b-a}{b-Y_i}\right) = 0. \quad (25.31d)$$

(Note that these are essentially "method of moments" expressions, relating the sample values of the harmonic means $E[(Y-a)^{-1}]^{-1}, E[(b-Y)^{-1}]^{-1}$ to the corresponding theoretical values; see Section 2.) Unfortunately, the likelihood function for the distribution is unbounded and has a global maximum that is infinite, so values of a "near" to Y_1', and of b "near" to Y_n' must be excluded. There is also a possibility of local maxima, which may not be well defined for small sample sizes and which plague various numerical schemes for maximizing likelihood. The ML estimators are asymptotically normal and unbiased with variances asymptotically equal to the Cramér-Rao lower bounds provided that $\min(p,q) > 2$. However, a numerical study by Carnahan indicates that only for very large sample sizes ($n \geq 500$) does the bias become small and the Cramér-Rao bound become a good approximation to the variance. The author recommends employing the least and greatest order statistics to improve the estimates of the end points.

The information matrix, from which the asymptotic variances and covariances of ML estimates can be obtained, (in the regular case of $p, q > 2$) is

$$\mathbf{I} = n\begin{pmatrix} \dfrac{q(p+q-1)}{(p-2)(b-a)^2} & \dfrac{(p+q-1)}{(b-a)^2} & \dfrac{q}{(p-1)(b-a)} & -\dfrac{1}{b-a} \\ \dfrac{p+q-1}{(b-a)^2} & \dfrac{p(p+q-1)}{(q-2)(b-a)^2} & \dfrac{1}{b-a} & -\dfrac{p}{(q-1)(b-a)} \\ \dfrac{q}{(p-1)(b-a)} & \dfrac{1}{b-a} & -\psi'(p+q)+\psi'(p) & -\psi'(p+q) \\ -\dfrac{1}{b-a} & \dfrac{-p}{(q-1)(b-a)} & -\psi'(p+q) & \psi'(p+q)+\psi'(q) \end{pmatrix}. \quad (25.36)$$

The diagonal elements of $\mathbf{I}^{-1}$ are the asymptotic variances of the parameter estimates. Explicit inversion has not been attempted. Carnahan (1989) provides numerical results.

AbouRizk, Halpin, and Wilson (1993) [see also AbouRizk, Halpin, and Wilson (1991)], using their program "Beta Fit," compare several methods for estimating the parameters of four-parameter beta distributions as in (25.1) (which they term "generalized beta distributions"). Among these were the following:

1. Moment methods. Equating first four sample and population moments; next taking a $= Y_1'$ and b $= Y_n'$ and using only the first two moments [e.g., as suggested by Riggs (1989)].
2. "Feasibility moment matching" method. Minimizing the (unweighted) sum of squares of differences between sample and population means, variances, skewness ($\sqrt{b_1}$ and $\sqrt{\beta_1}$) and kurtosis (b_2 and β_2), subject to a $< Y_1'$ and b $> Y_n'$, and possibly other restrictions (e.g., a > 0, b > 0).
3. Maximium likelihood method. Maximizing with arbitrary values of a and b [as described formally in (25.30)]; any variation of a and b is not considered.
4. "Regression-based" methods [see Swain, Venkataraman, and Wilson (1988)]. Using order statistics and the relationships [see Chapter 12, Eq. (12.20)]

$$E[F_Y(Y_j')] = \frac{j}{n+1},$$

$$\operatorname{var}(F_Y(Y_j')) = \frac{j(n-j+1)}{(n+1)^2(n+2)},$$

$$\operatorname{cov}(F_Y(Y_j'), F_Y(Y_k')) = \frac{j(n-k+1)}{(n+1)^2(n+2)}, \qquad j<k.$$

Two variants of a least-squares method are used in minimizing $\sum_{j=1}^n w_j\{F_Y(Y_j') - j/(n+1)\}^2$ with respect to a, b, p, and q.

Case 1. $w_j = 1$ for all j ("ordinary least squares"),
Case 2. $w_j = \{\operatorname{var}(F_Y(Y_j'))\}^{-1}$, $j = 1,\dots,n$ ("diagonally weighted least squares").
In each case minimization is subject to the restrictions a $< Y_1'$, b $> Y_n'$ (a > 0, b > 0), as in the third method mentioned above.

Dishon and Weiss (1980) provide small sample comparisons of maximum likelihood and moment estimators for standard beta distributions (25.2) ($a = 0$ and b $= 1$): The maximum likelihood estimators $\hat{p}$ and $\hat{q}$, which in this case are solutions of the equations,

$$\psi(\hat{p} + \hat{q} + 2) - \psi(\hat{p} + 1) = \frac{1}{n} \sum \log\left(\frac{1}{X_i}\right) \tag{25.37a}$$

and

$$\psi(\hat{p} + 4 + 2) - \psi(\hat{q} + 1) = \frac{1}{n} \sum \log\left(\frac{1}{1 - X_i}\right) \tag{25.37b}$$

are compared with the moment estimators

$$\tilde{p} = \frac{\tilde{\mu}'_1(\tilde{\mu}'_1 - \tilde{\mu}'_2)}{\tilde{\mu}'_2 - (\tilde{\mu}'_1)^2} - 1 \tag{25.38a}$$

and

$$\tilde{q} = \frac{(1 - \tilde{\mu}'_1)(\tilde{\mu}'_1 - \tilde{\mu}'_2)}{\tilde{\mu}'_2 - (\tilde{\mu}'_1)^2} - 1, \tag{25.38b}$$

where $\tilde{\mu}'_1$ and $\tilde{\mu}'_2$ are estimators of the first and second moments, respectively.

The results are summarized in Table 25.2. The authors generated a number of beta variables with known values of p and q and calculated, for different values of sample size n with 100 replications, $\hat{p}$, $\hat{q}$, $\tilde{p}$, and $\tilde{q}$ by means of the equations given above. (Both estimators tend to produce errors of the same sign.) They defined

$$R_p = \frac{\sum_{j=1}^{100}(\hat{p}_j - p)^2}{\sum_{j=1}^{100}(\tilde{p}_j - p)^2}$$

(with an analogous definition for R_q), where $\hat{p}_j$ is the ML estimator of p and $\tilde{p}_j$ is the moment estimator of p in the jth replication. [The authors also develop efficient procedures for computing $\psi(z)$ using the expansion

$$\psi(1 + z) = -\gamma + \sum_{n=1}^{\infty} \frac{z}{n(n + z)}, \qquad z \neq -1, -2, -3, \ldots,$$

where $\gamma = 0.57722\ldots$ is Euler's constant, as defined in Chapter 1, Eq. (1.19), and the Euler-Maclaurin summation formula.] The data in the table show that when n is low, the ML estimator is usually more accurate than the moment estimator (with notable exception when $p = q$).

Table 25.2 Comparison of estimates obtained from ML and moment estimates for a univariate beta distribution. Each row is for 100 replications

Parameter Values		Sample Size	N^a			
p	q	n	p	q	R_p	R_q
$-\frac{1}{2}$	$-\frac{1}{2}$	25	58	56	0.935	0.888
		50	58	64	0.911	0.799
		100	53	57	0.805	0.847
$-\frac{1}{2}$	1	25	64	61	0.793	0.802
		50	70	57	0.765	0.953
		100	62	56	0.646	0.829
1	1	25	42	44	1.020	1.020
		50	48	51	1.004	0.977
		100	51	50	0.962	0.975
$-\frac{1}{2}$	5	25	75	66	0.663	0.778
		50	66	63	0.564	0.706
		100	61	59	0.728	0.758
5	1	25	57	56	0.984	0.984
		50	57	59	0.932	0.912
		100	55	51	0.961	0.940
5	5	25	44	46	1.007	1.000
		50	41	42	1.017	1.021
		100	58	63	0.980	0.970
10	5	25	54	58	1.000	0.996
		50	57	58	0.996	0.989
		100	51	59	0.984	0.981
$-\frac{1}{2}$	100	25	64	67	0.806	0.852
		50	70	67	0.777	0.840
		100	76	68	0.693	0.801
1	100	25	56	61	0.915	0.889
		50	70	70	0.837	0.833
		100	62	64	0.914	0.896
50	100	25	53	54	0.996	0.996
		50	55	56	0.992	0.993
		100	50	50	1.000	1.000
100	100	25	57	55	0.999	0.999
		50	43	47	1.000	1.000
		100	57	57	1.000	1.000

[a] N = number of cases in which the MLE is closer to the true value of p, q, than the moment estimator.

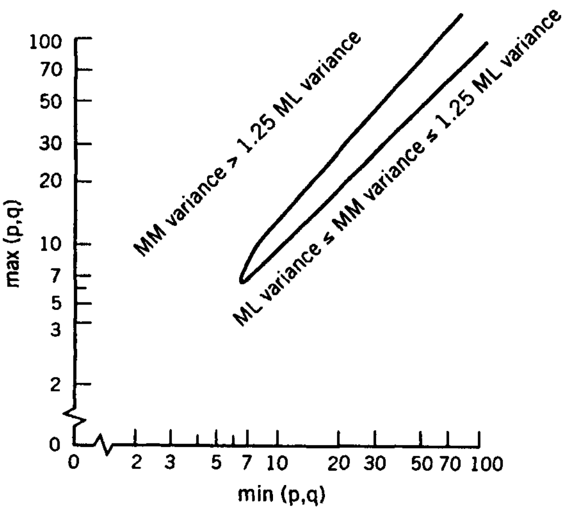

Figure 25.2 Comparison of the variances associated with MM and ML estimators

As Figure *25.2* from Kottes and Lau (1978) indicates, when p and q are small or their difference is large, the (asymptotic) method of moments variance exceeds the (asymptotic) maximum likelihood variance by at least *25%.* These are the situations when the need to fit beta distributions is the greatest. Fortunately, in many cases a and b, or at least one of these parameters, can be assigned known values.

If only the r smallest values $X_1', X_2', \dots, X_r'$ are available, the maximum likelihood equations are

$$\frac{r}{n}\log\left[\left(\prod_{j=1}^{r} X_j'\right)^{1/r}\right] = \psi(\hat{p}) - \psi(\hat{p}+\hat{q}) - \left(1-\frac{r}{n}\right)\frac{\partial}{\partial\hat{p}}\log\left[\int_{X_r'}^{1} t^{\hat{p}-1}(1-t)^{\hat{q}-1}\,dt\right] \tag{25.39a}$$

and

$$\frac{r}{n}\log\left[\left(\prod_{j=1}^{r}(1-X_j')\right)^{1/r}\right] = \psi(\hat{q}) - \psi(\hat{p}+\hat{q}) - \left(1-\frac{r}{n}\right)\frac{\partial}{\partial\hat{q}}\log\left[\int_{X_r'}^{1} t^{\hat{p}-1}(1-t)^{\hat{q}-1}\,dt\right] \tag{25.39b}$$

[Gnanadesikan, **Pinkham,** and Hughes **(1967)].**

Fang and Yuan (1990) apply the sequential algorithm for optimization by number theoretic methods (SNTO) proposed by Fang and Wang (1989) to obtain ML estimators of parameters of standard beta distributions. The method is superior to the Newton-Raphson method. It does not require unimodality or existence of derivatives (only continuity of the likelihood) and is not sensitive to the initial values. For the data provided by Gnanadesikan, Pinkham, and Hughes (1967), this method yields more accurate values than those of moment estimators or Gnanadesikan, Pinkham, and Hughes's (1967) estimators.

If one of the values p and q is known, the equations are much simpler to solve. In particular, for the standard power-function distribution ($q = 1$), the maximum likelihood estimator of p is

$$\hat{p} = \left[n^{-1} \sum_{j=1}^{n} \log X_j \right]^{-1}, \tag{25.40}$$

and we have

$$n \operatorname{var} \hat{p} \doteqdot p^2. \tag{25.41}$$

A moment estimator of p in this case is

$$\tilde{p} = \bar{X}(1 - \bar{X})^{-1}, \tag{25.42}$$

for which

$$n \operatorname{var} \tilde{p} \doteqdot p(p+1)^2(p+2)^{-1}. \tag{25.43}$$

Note that $(\operatorname{var} \hat{p})/(\operatorname{var} \tilde{p}) = p(p+2)(p+1)^{-2}$. The asymptotic relative efficiency of $\tilde{p}$ increases with p; it is as high as 75% for $p = 1$, tends to 100% as $p \to \infty$, but tends to zero as $p \to 0$. There is further discussion of power-function distributions in Chapter 20, Section 8.

Interestingly Guenther (1967) has shown that for the special case of the standard power-function distribution, with the pdf

$$p_{X_j}(x) = px^{p-1}, \qquad 0 < x < 1, j = 1, \dots, n, \tag{25.44}$$

the minimum variance unbiased estimator of p is $-(n-1)[\sum_{j=1}^{n} \log X_j]^{-1}$. Its variance is $p^2(n-2)^{-1}$, while the Cramér-Rao lower bound (Chapter 1, Section B15) is p^2n^{-1}.

In operations research applications (especially in connection with PERT) it is often *assumed* that the standard deviation *must* be one-sixth of the range of variation (!). Thus, for a standard beta (p, q) distribution (range 0 to 1), it is assumed that

$$\sigma(X) = \tfrac{1}{6}, \tag{25.45}$$

while for the more general distribution (25.1),

$$\sigma(X) = \tfrac{1}{6}(b - a). \tag{25.46}$$

This assumption is used in fitting a beta distribution on the basis of "least possible" (a^*), "greatest possible" (b^*), and "most likely" (m^*) values as estimated from engineers' experience of a process. These are used as estimates of a, b and the modal value

$$m = a + \frac{p-1}{p+q-2}(b-a) \qquad [\min(p,q) > 1], \tag{25.47}$$

respectively [Hillier and Lieberman (1980)].

Values of estimates p^*, q^* of p, q, respectively, can be obtained from the simultaneous equations

$$\frac{p^*q^*}{(p^* + q^*)^2(p^* + q^* + 1)} = \frac{1}{36} \qquad \text{(cf. (25.46))} \tag{25.48a}$$

and

$$\frac{p^* - 1}{p^* + q^* - 2} = \frac{m^* - a^*}{b^* - a^*}. \qquad \text{(cf. (25.47))} \tag{25.48b}$$

It would be, perhaps, more natural to use an estimated expected value, $\bar{X}$, say, and equate that to the population value, leading to

$$\frac{p^*}{p^* + q^*} = \frac{\bar{X} - a^*}{b^* - a^*} \tag{25.48c}$$

in place of (25.48b). In fact it appears to be customary to use an equation like (25.48c) but with $\bar{X}$ replaced by an estimate of the expected value

$$a^* + \frac{1}{6}\left\{\frac{4(m^* - a^*)}{b^* - a^*} + 1\right\}(b^* - a^*) = a^* + \frac{1}{6}(4m^* + b^* - 5a^*),$$

leading to

$$\frac{p^*}{p^* + q^*} = \frac{1}{6}\left\{4\left(\frac{m^* - a^*}{b^* - a^*}\right) + 1\right\}. \tag{25.48d}$$

From (25.48a), (25.48c), or (25.48d), $(p^* + q^*)$ can be expressed in terms of p*, a*, and b*. Inserting this expression for $(p^* + q^*)$ in (25.48a), we obtain an equation in p*. For example, using (28.48c),

$$p^* + q^* = \left(\frac{b^* - a^*}{\bar{X} - a^*}\right)p^*$$

and

$$\frac{p^* q^*}{(p^* + q^*)^2} = \left(\frac{\bar{X} - a^*}{b^* - a^*}\right)\left(1 - \frac{\bar{X} - a^*}{b^* - a^*}\right),$$

whence (25.48a) becomes

$$\left(\frac{\bar{X} - a^*}{b^* - a^*}\right)\left(1 - \frac{\bar{X} - a^*}{b^* - a^*}\right) = \frac{1}{36}\left\{\left(\frac{b^* - a^*}{\bar{X} - a^*}\right)p^* + 1\right\},$$

that is,

$$p^* = \frac{\bar{X} - a^*}{b^* - a^*}\left\{36\left(\frac{\bar{X} - a^*}{b^* - a^*}\right)\left(1 - \frac{\bar{X} - a^*}{b^* - a^*}\right) - 1\right\}. \tag{25.49}$$

Using (28.48d) would also lead to a simple explicit value for p*, but (28.48b) would lead to a cubic equation for p*.

Farnum and Stanton (1987) carried out a critical investigation of the accuracy of the assumption that for a standard beta (p, q) variable

$$\text{Expected value} = \tfrac{1}{6}\{4(\text{mode}) + 1\},$$

that is,

$$\frac{p}{p + q} \doteqdot \frac{1}{6}\left(\frac{4(p - 1)}{p + q - 2} + 1\right) \tag{25.50}$$

[presumably when (25.48a) is satisfied]. They found the approximation is correct to within 0.02 when the mode is between 0.13 and 0.87, and they suggest using improved approximations

$$\frac{2}{2 + (\text{mode})^{-1}} \qquad \text{for mode} < 0.13, \qquad (25.51a)$$

and

$$(3 - 2(\text{mode}))^{-1} \qquad \text{for mode} > 0.87. \qquad (25.51b)$$

Moitra (1990) suggested that some allowance should be made for "skewness" (which he measures by $E[(X - E[X])^3]$, rather than the shape factor $\sqrt{\beta_1}$, which would be $6\sqrt{6}E[(X - E[X])^3]$ if $\sigma(X) = 1/6$). Moitra noted that the "traditional" assumptions can be expressed as

$$E[X] = \frac{a + b + k(\text{mode})}{k + 2} \qquad (25.52a)$$

and

$$\sigma(X) = c^{-1}(b - a), \qquad (25.52b)$$

with $k = 4$, $c = 6$. He found that $c = 6$ is not "optimal for values of k other than 4 or 5, and $k = 4$ is not optimal for values of c other than 6."

Moitra made the following recommendations: "If the skewness is judged or known to be high, p would be between 2 and 3, and since we are estimating subjective distributions, we can set $p = 2.5$." But "if the skewness is judged to be moderate, then we can see from the graphs that p is very likely to be between 3 and 4, and so we can similarly set $p = 3.5$. Finally, if the skewness is considered to be only a little, we set $p = 4.5$." He also provided the "best" combinations of values for k and c, which are given in Table 25.3, and an analysis appropriate to triangular distributions (see Chapter 26, Section 9) for which

$$E[X] = \tfrac{1}{3}(a + b + m). \qquad (25.53)$$

Table 25.3 Best combinations of k and c

	k					
c	1	2	3	4	5	6
3	Best					
4		Best	Good			
5			Good	Best		
6				Good	Best	
7						Best
8						Best