# II. Statistics

1. **Univariate distributions**
    - a. **Models for positive random variables**
    - b. **Models for real-valued random variables**
    - c. **Hazard functions**
    - d. **Moments and characteristics functions**
    - e. **Transformations of random variables**

2. **Multivariate distributions**
    - a. **Expectations**
    - b. **Multivariate normal**
    - c. **Distribution results**
        - (1) **Chi-square**
        - (2) **Some useful theorems**
        - (3) **Example**
        - (4) **t-distribution**
        - (5) **F-distribution**
    - d. **Some generalized multivariate distributions**

3. **Estimation**
    - a. **Maximum likelihood**
    - b. **Method of moments**
    - c. **Kernel estimation**

4. **Asymptotic distribution theory**
    - a. **Modes of convergence**
    - b. **Law of large numbers**

5. **Problem sets**

James B. McDonald
Brigham Young University
12//2012

# II. STATISTICS

Statistical distributions play an important role in many areas of economics. These include descriptive models for the distribution of such economic variables as income, firm size, prices, and stock returns to mention but a few. Assumptions about underlying distributions also provide the theoretical basis for such common estimation techniques as least squares. It can be shown that many of the common distributions are closely related to each other and are special cases of a few generalized density functions. These general univariate distributions will be considered in the first section. Some multivariate distributions will be considered in the second section. Section three will discuss principles of maximum likelihood estimation and the distribution of these estimators under correct model specification. If the density function is misspecified, these (maximum likelihood) estimators will be referred to as quasi maximum likelihood (QMLE) and in some cases can still be consistent and asymptotically normal; however, in some important cases misspecification of the density can lead to inconsistent estimators. Section four reviews different statistical concepts of convergence and important theorems which arise in asymptotic distribution theory.

## 1. Some Univariate Distributions

A few statistical distributions have played a dominant role in empirical and theoretical work. Some of the most widely used models in economics include the normal, t, lognormal, Pareto, gamma and beta. Often these models have been adopted without testing their validity or considering the consequences of model misspecification. In selecting a statistical distribution it is important to consider distributional characteristics such as the mean, variance, skewness and kurtosis. For example, the selection of an exponential distribution implies that the associated variance equals the mean squared. We first consider some probability density functions (pdf) for positive random variables and then some pdf's for random variables which can assume positive and negative values. Hazard functions are then discussed. It is important to be able to determine the distribution of transformations of random variables with known distribution functions. The methods of change of variable and using moment generating or cumulant generating functions will be discussed and illustrated with examples.

**a.** **Models for Positive Random Variables:**

The generalized beta (GB), generalized beta of the first and second kind (GB1, GB2), and generalized gamma (GG) are density functions for positive random variables and are defined by

$$(1.1)\quad GB(y;a,b,c,p,q) = \frac{|a|y^{ap-1}(1-(1-c)(y/b)^{a})^{q-1}}{b^{ap}B(p,q)(1+c(y/b)^{a})^{p+q}} \text{ for } 0<Y^{a}<\frac{b^{a}}{1-c}$$

$$(1.2)\quad GB1(y;a,b,p,q) = \frac{|a|y^{ap-1}(1-(y/b^{a}))^{q-1}}{b^{ap}B(p,q)} \quad 0<y<b,$$

$$(1.3)\quad GB2(y;a,b,p,q) = \frac{|a|y^{ap-1}}{b^{ap}B(p,q)(1+(y/b)^{a})^{p+q}} \quad 0< y,$$

$$(1.4)\quad GG(y;a,\beta,p) = \frac{|a|y^{ap-1}e^{-(y/\beta)^{a}}}{\beta^{ap}\Gamma(p)} \quad 0<y$$

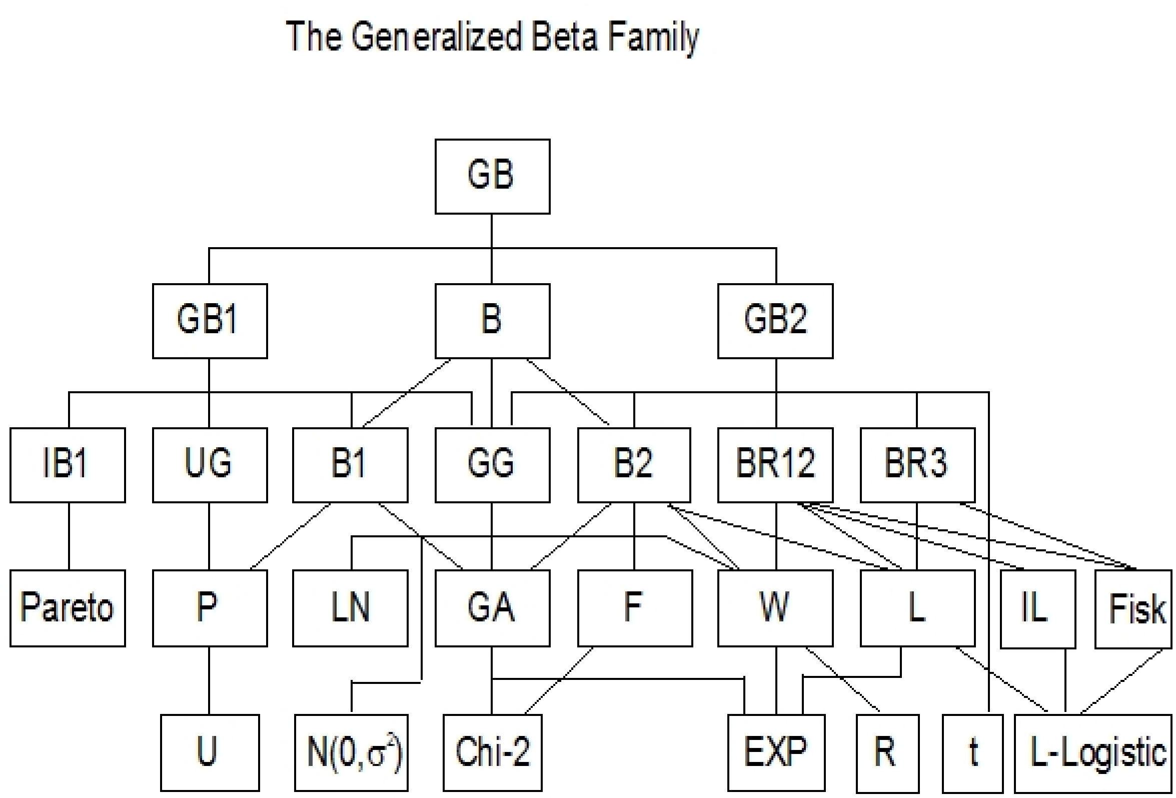
$$= 0 \qquad \text{otherwise.}$$

These distributions can be shown to include the beta of the first kind (B1), the beta of the second kind (B2), or Burr type 12 (BR12), Burr type 3 (BR3), Power (P), Lognormal (LN), Weibull (W), gamma (GA), F, Lomax (L), Fisk (Fisk), Uniform (U), Rayleigh (R), exponential (EXP), Chi Square ($\chi^2$) and half Normal and half t distributions as special or limiting cases. These relationships are depicted in the form of a distribution tree in Figure 1 and are developed in detail in McDonald [1984], McDonald and Butler [1987], McDonald and Richards [1987], and McDonald and Xu [1991]. The GB2 distribution is referred to as a generalized F by Kalbfleisch and Prentice [1980] and a modified version (with a non-zero threshold) as a Feller-Pareto distribution, Arnold [1983].

The parameters in these distributions generally determine the shape and location of the density in a complex manner. A few comments may be helpful. The parameters b and β are merely scale parameters and depend upon the units of measurement. The generalized beta type I and generalized gamma have defined moments ($E(y^h)$) of all integer order for a > 0 or more generally as long as p+h/a > 0; whereas, the generalized beta type II has integer moments of order up to "h" where -p < h/a < q. Consequently, the GB2 density permits the analysis of situations characterized by infinite variance. Generally speaking, the larger the value of "a" or "q," the "thinner" the tails of the density function. In fact, for "large" values of the parameter "a" the corresponding GB2 density function is characterized by the probability mass being concentrated near the value of the parameter "b." This can be verified by noting that for "large" values of "a" the mean is approximately "b" and the variance is near zero. The relative values of the parameters "p" and "q" play an important role in determining the value of skewness and permit positive or negative skewness. This is in contrast to such distributions as the lognormal which is always positively skewed.

The interrelationships between many of these distributions are summarized in Figure 1, where the notation is formalized in Table 1. From Figure 1, we observe that the GB (eqn. 1.1) includes the GB1 (eqn. 1.2) and GB2 (eqn. 1.3) corresponding to c = 0, 1, respectively.

**Figure 1**




see McDonald and Xu, "A Generalization of the Beta Distribution with Applications," *Journal of Econometrics* 66 (1995) pp.133-152.

Table 1. Key to notation used in the distribution tree in figure 1.

| Symbol | Distribution |
|---|---|
| GB = Beta(y;a,b,c,p,q) | Generalized Beta |
| GB1 = GB1(y;a,b,p,q) | Generalized beta type I |
| GB2 = GB2(y;a,b,p,q) | Generalized beta type II |
| IB1 = IB1(y;b,p,q) | Inverse beta type I |
| UG = UG(y;b,$\delta$,q) | Unit gamma |
| B1 = B1(y;b,p,q) | Beta of first kind |
| LT = LT(y;$\mu$,$\sigma^2$,q) | Log t |
| GG = GG(y;a,$\beta$,p) | Generalized gamma |
| B2 = B2(y;b,p,q) | Beta of second kind |
| BR12 = BR12(y;a,b,q) | Burr(type 12), Beta-P or Singh Maddala |
| BR3 = BR3(y;a,b,p) | Kappa (K-3), Beta-K or Burr (type 3) |
| Pareto = Pareto(y;b,p) | Pareto |
| P = P(y;b,p) | Power |
| LN = LN(y;$\mu$,$\sigma^2$) | Lognormal |
| W = W(y;a,$\beta$ | Weibull, Rosin-Rammler |
| GA = GA(y;$\beta$,p) | Gamma |
| F = F(y;u,$\nu$) | Snedecor F |
| L = L(y;b,q) | Lomax |
| IL = IL(y;b,q) | Inverse Lomax |
| Fisk = FISK(y;a,b) | Fisk |
| U = U(y;b) | Uniform |
| ½(N) = ½(N(O,$\sigma^2$)) | Half normal |
| $\chi^2$ = $\chi^2$(y;u) | Chi square |
| Exp = E(y;$\beta$) | Exponential |
| R = R(y;$\beta$) | Rayleigh |
| ½(t) = ½(t(y;$\nu$)) | Half t |
| Logistic | Logistics |

The generalized beta distributions (GB1 and GB2) includes the generalized gamma as a limiting case,

$$GG(y;a,\beta,p) = \lim_{q\to\infty} GB1(y;a,\beta q^{1/a},p,q)$$

$$= \lim_{q\to\infty} GB2(y;a,\beta q^{1/a},p,q),$$

and hence include all the special cases of the generalized gamma as limiting cases. See McDonald [1984] and McDonald and Richards [1987] and McDonald and Xu [1992] for details. Additional relationships can be more formally written as

$$LN(y;\mu,\sigma^2) = \operatorname*{Limit}_{a\to 0} GG\left(y;a,\beta = (\sigma^2 a^2)^{1/a},\ p = \frac{a\mu+1}{\sigma^2 a^2}\right)$$

$$GA(y;\beta,p) = GG(y;a = 1,\beta,p)$$

$$W(y;a,\beta) = GG(y;a,\beta,p = 1)$$

$$BR12(y;a,b,q) = GB2(y;a,b,p = 1,q)$$

$$W(y;a,\beta) = \lim_{q\to\infty} BR12(y;a,b = \beta q^{1/a},q).$$

We see from figure 1 and the previous discussion that both the generalized gamma and Burr type 12 are more general models than the widely used Weibull (W).

The generalized beta of the second kind involves four parameters and is a particularly useful family of distributions. It includes the generalized gamma, beta of the second kind, the Burr type 12 or Beta-P and the Burr type 3, the three parameter Kappa or Beta-K distribution, and all the previously mentioned associated special cases as members. The distribution of the Snedecor F statistic (variance ratio) is also a special case of the generalized beta of the second kind, as are the half normal and half t. The hierarchical relationships among the distributions can be used in testing the nested hypotheses.

Table 2 is included for completeness and includes expressions for the density, cumulative distribution and moments. The notation is defined in McDonald [1984] and McDonald and Richards [1987].

**Table 2. Some Probability Density Functions**

| Distribution | Domain of Y | pdf | Restrictions | Mean | Variance* | MGF |
|---|---|---|---|---|---|---|
| Poisson$(y;\lambda)$ | $\{0,1,2,..\}$ | $e^{-\lambda}\lambda^{y}/y!$ | 0< $\lambda$ | $\lambda$ | $\lambda$ | $e^{\mu(e^{t}-1)}$ |
| Uniform$(y;b)$ | $\{0<y<b\}$ | $1/b$ | 0<b | b/2 | $b^2/12$ | $(e^{bt}-1)/(bt)$ |
| $EXP(y;\beta)$ | $\{0<y\}$ | $e^{-y/\beta}/\beta$ | $0<\beta$ | $\beta$ | $\beta^2$ | $(1-\beta t)^{-1}$ |
| $\chi^2(y;d)$ | $\{0<y\}$ | $y^{(d/2)-1}e^{-y/2}/\left(\Gamma(d/2)2^{d/2}\right)$ | $d=1,2,..$ | d | 2d | $(1-2t)^{-d/2}$ |
| $GA(y;\beta,p)$ | $\{0<y\}$ | $y^{p-1}e^{-y/\beta}/\left(\Gamma(p)\beta^{p}\right)$ | 0< $\beta$, p | $\beta p$ | $\beta^2 p$ | $(1-\beta t)^{-p}$ |
| $B1(y;b,p,q)$ | $\{0<y<b\}$ | $y^{p-1}(1-y/b)^{q-1}/\left(b^{p}B(p,q)\right)$ | 0<b, p, q | $\dfrac{bp}{p+q}$ | $\dfrac{b^2pq}{(p+q)^2(p+q+1)}$ | Involves an infinite series |
| $B2(y;b,p,q)$ | $\{0<y\}$ | $y^{p-1}/\left(b^{p}B(p,q)(1+y/b)^{p+q}\right)$ | 0<b, p, q | $\dfrac{bp}{q-1}$ | $\dfrac{bB(p+h,q-h)^{*}}{B(p,q)}$ | “ |
| $F(y;p,q)$ | $\{0<y\}$ | $y^{p-1}/\left(b^{p}B(p,q)(1+y/b)^{p+q}\right)$ | $p=\dfrac{d_1}{2}, q=\dfrac{d_2}{2}$ <br> $b=d_2/d_1$ | $\dfrac{d_2}{d_2-2}$ | B2 with appropriate substitutions | “ |
| $GG(y;a,\beta,p)$ | $\{0<y\}$ | $\lvert a\rvert y^{ap-1}e^{-(y/\beta)^{a}}/\left(\Gamma(p)\beta^{ap}\right)$ | $0<\beta,p$ | $\dfrac{\beta\Gamma(p+1/a)}{\Gamma(p)}$ | $\dfrac{\beta^{h}\Gamma(p+h/a)^{*}}{\Gamma(p)}$ | ” |

| | | | | | | |
|---|---|---|---|---|---|---|
| $GB1(y;a,b,p,q)$ | $\{0<y<b\}$ | $\lvert a\rvert y^{ap-1}\left(1-(y/b)^a\right)^{q-1}/\left(b^{ap}B(p,q)\right)$ | 0<b, p, q | $\frac{bB(p+q,1/a)}{B(p,1/a)}$ | $\frac{b^h B(p+q,h/a)^*}{B(p,h/a)}$ | “ |
| $GB2(y;a,b,p,q)$ | $\{0<y\}$ | $\lvert a\rvert y^{ap-1}/\left(b^{ap}B(p,q)\left(1+(y/b)^a\right)^{p+q}\right)$ | 0<b, p, q | $\frac{bB\left(p+\frac{1}{a},q-\frac{1}{a}\right)}{B(p,q)}$ | $\frac{b^h B(p+h/a,q-h/a)}{B(p,q)}*$ | “ |
| $Laplace(y;s)$ | $\{-\infty<y<\infty\}$ | $\frac{e^{-\lvert y/s\rvert}}{2s}$ | 0<s | 0 | $2s^2$ | $\frac{1}{(1-s^2t^2)}$ |
| | $\{-\infty<y<\infty\}$ | $\frac{e^{-(y-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma}$ | $0<\sigma^2$ | $\mu$ | $\sigma^2$ | $e^{\mu t+(t\sigma)^2/2}$ |
| $GED(y;\mu,s,p)$ | $\{-\infty<y<\infty\}$ | $\frac{pe^{-\lvert(y-\mu)/s\rvert^p}}{2s\Gamma(1/p)}$ | 0<p,s | $\mu$ | $\frac{s^2\Gamma(3/p)}{\Gamma(1/p)}$ | $N(y;\mu,\sigma^2)$ |
| $SGT(y;\lambda,s,p,q)$ ** | $\{-\infty<y<\infty\}$ | $\frac{2}{2sq^{1/p}B\left(\frac{1}{p},q\right)\left(1+\frac{\lvert y\rvert^p}{(1+\lambda sign(y))^p qs^p}\right)^{q+1/p}}$ | 0<s,p,q $0\le\lvert\lambda\rvert<1$ | $2\lambda sq^{1/p}\frac{B\left(\frac{2}{p},q-\frac{1}{p}\right)}{B\left(\frac{1}{p},q\right)}$ | ALC | ALC |
| $EGB2(y;m,s,p,q)$ | $\{-\infty<y<\infty\}$ | $\frac{e^{p(y-m)/s}}{sB(p,q)\left(1+e^{(y-m)/s}\right)^{p+q}}$ | $0<s,p,q$ | $m+s[\psi(p)-\psi(q)]$ | $s^2[\psi'(p)+\psi'(q)]$ | $\frac{B(p+u,q-u)}{e^{-u}B(p,q)}$ |

* when an asterisk appears in the variance column, the entry denotes the h[th] order moment about the origin

$\psi(s)=d\ell n\Gamma(s)/ds$ is called ı$\psi'(s)$mma function. is the trigamma function.

ALC = a little complicated, ** Results for many special cases of the SGT can be obtained by referring to the SGT figure in the class notes. Also, *y* can be replaced by *y-m*.

**b. <u>Models for Real Valued (Positive and Negative) Random Variables</u>:**

Two other important families of distributions for random variables which can be positive or negative are the skewed generalized T (SGT) and exponential generalized beta (EGB) distributions.

(1) **The skewed generalized T (SGT) is defined by**

(1.5) $$\mathrm{SGT}(\varepsilon;\mu,\sigma,n,k,\lambda)=\frac{C}{\sigma}\left(1+\frac{|\varepsilon|^k}{((n-2)/k)(1+\mathrm{sign}(\varepsilon)\lambda)^k\theta^k\sigma^k}\right)^{-(n+1)/k}$$

for $-\infty < \varepsilon < \infty$ and for $n > 2$ where C and θ are rather complicated expressions involving beta functions and other parameters, see Theodossiou (1998). An alternative less restrictive formulation is given in (2.5)'

(1.5)' $$\mathrm{SGT}(\varepsilon;\lambda,s,p,q)=\frac{p}{2sq^{1/p}B\left(\frac{1}{p},q\right)\left(1+\frac{|\varepsilon|^p}{qs^p(1+\lambda\,\mathrm{sign}(\varepsilon))^p}\right)^{q+1/p}}$$

The moments for the SGT given in (1.5)' can be expressed as

$$E\left(\varepsilon_{SGT}^h\right)=\left(\frac{s^h}{2}\right)\left\{\frac{q^{h/p}B\left(\frac{h+1}{p},q-\frac{h}{p}\right)}{B\left(\frac{1}{p},q\right)}\right\}\left((1+\lambda)^{h+1}+(-1)^h(1-\lambda)^{h+1}\right).$$

The product of the parameters p and q is the degrees of freedom of the SGT. An important limiting case of the SGT is the SGED (skewed generalized error distribution) which corresponds to the limit of (1.5)' as the parameter q grows indefinitely large. The pdf of the SGED, sometimes referred to as the skewed Box-Taio generalized error distribution, or skewed generalized power distribution is given by

$$\mathrm{SGED}(\varepsilon;s,\lambda,p)=\frac{pe^{-\left(\frac{|\varepsilon|^p}{(1+\lambda\,\mathrm{sign}(\varepsilon))^p s^p}\right)}}{2s\Gamma\left(\frac{1}{p}\right)}$$

For large values of the parameter q, the second bracketed expression for the moments of the SGT approaches ($\Gamma((h+1)/p)/\Gamma(1/p)$) to yield expressions for the moments of the SGED..

The relationship between the SGT, GED, skewed t, the t, double exponential or Laplace, normal and Cauchy variates is given in Figure 2. The $\lambda$ parameter is the skewness parameter where the probability of $\varepsilon$ being positive is given by $\left(\frac{1+\lambda}{2}\right)$ with $\lambda=0$ corresponding to a symmetric generalized t distribution (GT).

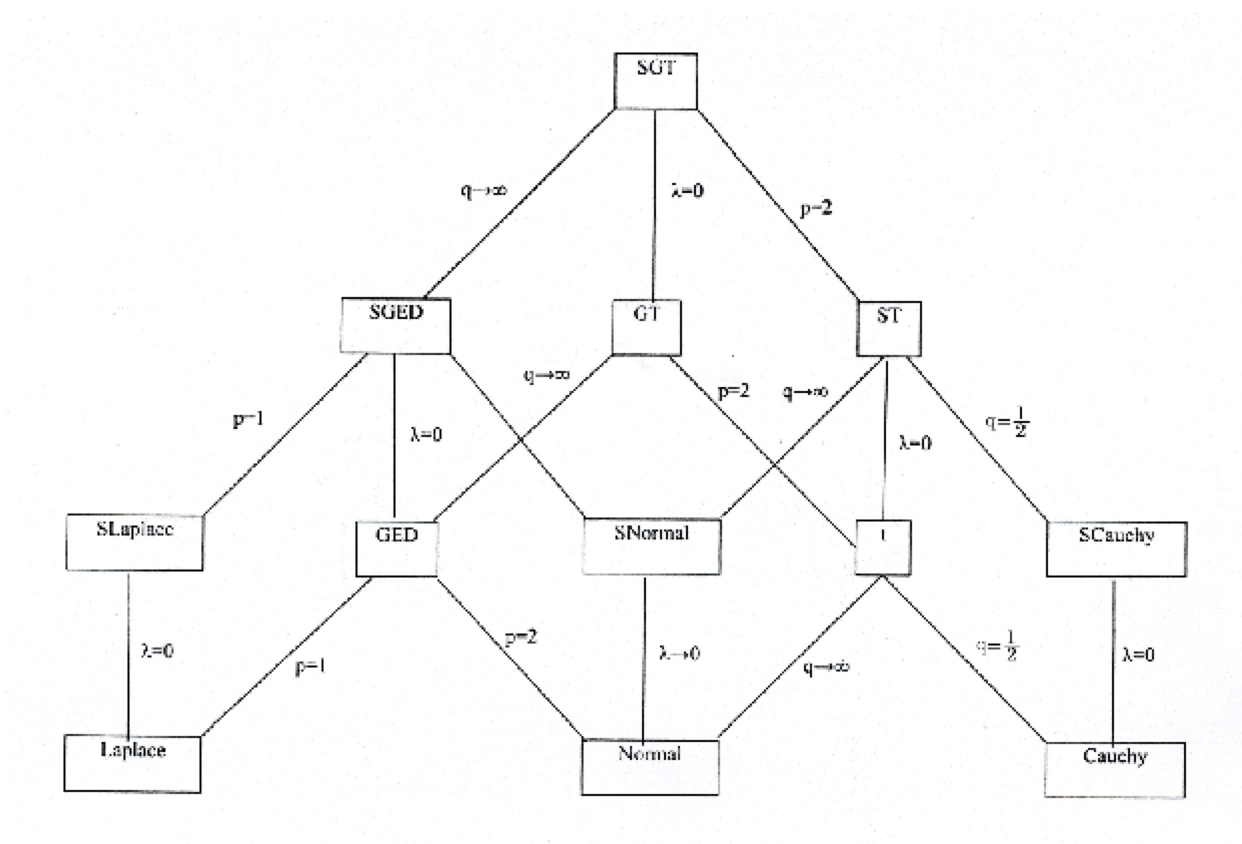



**Figure 2**

**SGT Distribution Family**

The GT distribution was introduced in McDonald and Newey [1988]. The GT and its special cases are all symmetric, but can accommodate tails which are thicker or thinner than the normal. They provide the basis for "robust" or partially adaptive estimation of regression and time series models.

(2) **The pdf for the exponential generalized beta** is given by

$$(1.6) \qquad EGB(\varepsilon;\delta,\sigma,c,p,q) = \frac{e^{p(\varepsilon-\delta)/\sigma}\left(1-(1-c)e^{(\varepsilon-\delta)/\sigma}\right)^{q-1}}{|\sigma|B(p,q)\left(1+ce^{(\varepsilon-\delta)/\sigma}\right)^{p+q}}$$

The domain for ε is $-\infty < \frac{\varepsilon-\delta}{\sigma} < \ln\left(\frac{1}{1-c}\right)$ for the EGB.

If Y ~ GB(y; a, b, c, p, q), then the random variable $\varepsilon = \ln(Y)$, $Y = e^{\varepsilon}$, will be said to be distributed as an exponential generalized beta (EGB). A case could be made for saying that ε is distributed as a log - GB2(LGB2). However, the convention introduced by the relationship between the lognormal and normal distribution will be adopted. Since the EGB and GB distributions are related by the logarithmic transformation, similar "exponential" distributions can be obtained by transforming each of the distributions shown in figure 1. Several of these distributions are of special interest. Consider, for example, the exponential generalized beta of the first and second kind (EGB1 and EGB2) and the exponential generalized gamma (EGG) defined by the probability density functions:

$$(1.7) \qquad EGB1(Z;\ \delta,\ \sigma,\ p,\ q) = EGB(Z;\ \delta,\ \sigma,\ c=0,\ p,\ q) = \frac{e^{p(Z-\delta)/\sigma}(1-e^{(Z-\delta)/\sigma})^{q-1}}{|\sigma|\ B(p,q)};$$

$$(1.8) \qquad EGB2(Z;\ \delta,\ \sigma,\ p,\ q) = EGB(Z;\ \delta,\ \sigma,\ c=1,\ p,\ q) = \frac{e^{p(Z-\delta)/\sigma}}{|\sigma|\ B(p,q)\ (1+e^{(Z-\delta)/\sigma})^{p+q}};$$

and

$$(1.9)\quad EGG(z;\ \delta,\ \sigma,\ p) = \operatorname*{Limit}_{q\to\infty} EGB(Z;\delta* = \sigma\ln q+\delta, c, p, q)$$

$$= \frac{e^{p(Z-\delta)/\sigma}\ e^{-e^{(Z-\delta)/\sigma}}}{|\sigma|\ \Gamma(p)}.$$

The EGB1, EGB2 and EGG are merely alternative representations of the generalized exponential, logistic and Gompertz distributions reviewed in Johnson and Kotz [1970, vol. 2] and Patil et al. [1984]. If (δ, σ, p, q) in the EGB1, EGB2 and EGG is replaced with (-σlnρ, -σ, δ, θ), the notation found in Johnson and Kotz [p. 271] is obtained. The terminology for these and closely related distribution differs in the literature. For example, Patil et al. [1984] refer to generalized exponential and logistics distributions which are different from those referred to by Johnson and Kotz. The notation used here may help clarify some of the interrelationships. The generalized Gumbell corresponds to the EGB2 with p = q. The EBR3 is merely the Burr type 2 distribution. The exponential Weibull (EW) is the extreme value type I distribution. These distributions and interrelationships could be graphically summarized as in figure 1. In fact a figure similar to figure 1 could be constructed for the EGB distribution with each distributional type in figure 1 being preceded with an "E" to denote "exponential," and the parameter a replaced with 1/σ and b with $e^{\delta}$. Thus the EGB would appear in place of GB. In the case of the lognormal, $LN(y;\mu,\sigma^2)$, ELN would correspond to the normal, $N(z;\mu,\sigma^2)$. The structure of this distributional tree would be the same as figure 1, but would include generalized forms of the exponential, logistics, Gompertz, Gumbell and extreme value distributions.

Rather than replicating figure 1 for the "exponential" distributions, we present an abbreviated form which includes some of the most well known members of the EGB family. A comparison of figures 1 and 3 suggests many other distributions which could have been included.

The EGB density with c = 1, the EGB2, can be skewed to the right or left or symmetric depending on the relative values of p and q and has tails which are thicker than the normal. This density also includes the normal as a special case as well as many models used in the reliability literature. See McDonald and Xu for additional details.

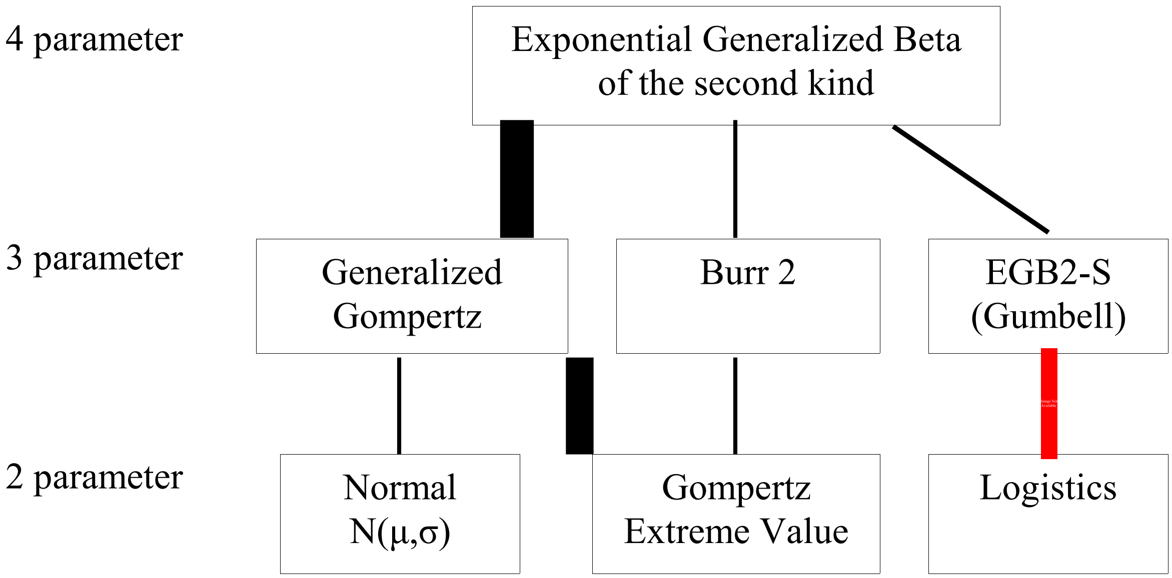



**Figure 3**

**EGB Distribution Tree**

$$*BR2(Z;\ \delta,\sigma,q) = EGB2(Z;\ \delta,\sigma,p=1,q) = EBR12(Z;\delta,\sigma,q)$$

$$EGB2(Z;\delta,-\sigma,p=1,\ q) = EBR3(Z;\delta,-\sigma,q)$$

**** EGB2-S represents a S(symmetric) EGB2**

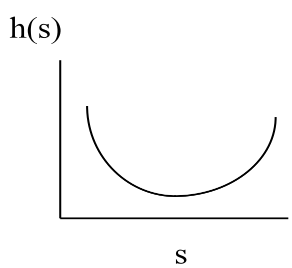
The density functions summarized in Figures 1, 2, and 3 include nearly all of the Pearson family and additional distributions as well. They have the advantage of starting from a few relatively tractable functional forms with the remaining distributions being special cases of the general forms. They are very flexible and can be "thick" or "thin" tailed. In fact, the GB2 and SGT define a finite number of integer moments. The EGB has finite integer moments of all order. b(β) or σ are merely scale parameters. a, p, q are shape parameters. δ is a location parameter. The densities in Figure 2 (SGT and special cases) can be skewed or symmetric about the origin. The generalized beta density functions depicted in Figure 1 and the EGB (figure 3) can be skewed to the left or right depending on the relative values of p and q; whereas, the lognormal is positively skewed for all parameter values.

**c. <u>Hazard Functions</u>**

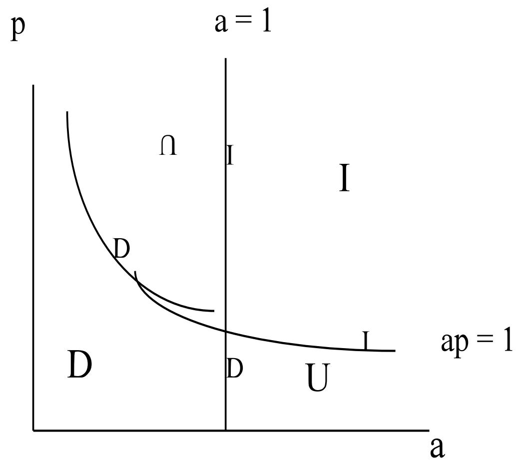
The hazard function is a very useful characterization of density functions and is defined by

(1.10) $$h(s) = \frac{f(s)}{1 - F(s)}$$

and is important in many applications. For example, if s denotes the length of life, then h(s) is the rate of death at age s, given that an individual has lived up to s. The hazard function for mortality data might be expected to appear as follows:




h(s) exhibits decreasing mortality in the first few months of life, then a period of relatively constant mortality and finally an increasing probability of death at older ages. The notion of hazard functions is useful in modeling unemployment duration, length of time of first purchase of a new product or failure time of products. Many models in current use have very limited flexibility for hazard functions. However, many applications may require a "∪" or "∩" shaped or strictly increasing (I) or decreasing (D) hazard functions. The generalized gamma is "∪" shaped for $a > 1$ and $p < 1/a$; "∩" shaped for $a < 1$ and $p > 1/a$; I shaped for $a > 1$ and $p > 1/a$; and D shaped for $a < 1$ and $p < 1/a$, Glaser [1980]. This is compactly summarized in Figure 4, see McDonald and Richards [1987b]. Exercise: What are possible shapes of the hazard function for the gamma, Weibull, or exponential density? (Hint: refer to figure 1 and figure 4)




**Hazard Functions for the Generalized Gamma**
**Figure 4**

## d. Moments, Moment Generating Functions, and Cumulant Generating Functions:

A knowledge of $M_X(t)$ can be used to obtain defined moments about the origin as follows (where the ith moment is defined)

$$E(x^i) = \mu_i' = \int x^i f(x)\,dx$$

$$= \frac{d^i M_x(t)}{dx^i}\Big|_{t=0}.$$

The proof of this results follows from the moment generating function

$$M_X(t) = E\,(e^{tx}) = \int e^{tx} f(x)\,dx$$

$$= \int \sum_{i=0}^{\infty} \frac{(tx)^i}{i} f(x)\,dx$$

$$= \sum_{i=0}^{\infty} \frac{t^i}{i!} E(x^i)$$

which requires the assumption that the integral and summation can be interchanged.

Exercise: (1) Demonstrate that the moment generating function for the normal, $N(\mu,\sigma^2)$, is given by

$$M_X(t) = e^{\mu t + (\sigma^2 t^2/2)}$$

(2) Given this result derive the mean and variance for $N(\mu,\sigma^2)$

$$E(X) = (dM_X(t)/dt)|_{t=0}$$

$$Var(X) = E(X^2) - E^2(X)$$

$$= \frac{d^2M_x(t)}{dt^2}\Big|_{t=0} - E^2(X)$$

The cumulant generating function

$$\Psi_x(t) = \ln M_x(t)$$

is very useful for obtaining the first four defined moments about the mean. These results are obtained from

$$\frac{d\psi_x(t)}{dt}\Big|_{t=0} \qquad \mu = \text{mean} = E(X) =$$

$$\frac{d^2\psi_x(t)}{dt^2}\Big|_{t=0} \qquad = \text{Var}\ \sigma^2 = E(X - E(X))^2 =$$

$$\frac{d^3\psi_X(t)}{dt^3}\Big|_{t=0} \qquad = \text{Skewness}(X) = E(X - E(X))^3$$

$$\frac{d^4\psi_x(t)}{dt^4}\Big|_{t=0} \qquad = \text{Excess Kurtosis} = E(X - E(X))^4 - 3\ (\text{Variance})^2$$

Dividing the expressions for skewness and excess kurtosis by $\sigma^3$ or $\sigma^4$, respectively, yields standardized or normalized measures of skewness and excess kurtosis which are independent of the units of measurement.

Exercises:

(1) Use this result to derive the variance, skewness $\left(\frac{E(X-E(X))^3}{\text{variance}^{3/2}}\right)$ and kurtosis $\left(\frac{E(X-E(X))^4}{\text{variance}^2}\right)$ coefficients for the normal.

(2) The moment generating function for the EGB2 can be shown to be

$$\frac{e^{\delta t}\Gamma(p+t\sigma)\Gamma(q-t\sigma)}{\Gamma(p)\Gamma(q)}$$

Show that the mean and variance of the EGB2 are given

$$\delta + \sigma[\psi(p) - \psi(q)]$$

$$\sigma^2[\psi'(p) + \psi'(q)]$$

where $\psi(a) = d \ln \Gamma(a)/da$ is the digamma function. Demonstrate that the EGB2 is symmetric if $p = q$.

**e. Transformations of Random Variables**

The relationships between the random variables considered in figures 1, 2 and 3 can also be considered as arising from transformations of random variables (moving from special cases to more general forms) rather than as special cases of the general forms. The techniques for determining the distribution of transformations are important to understand. The methods of change of variables and those based on moment generating functions are two of the most common.

(1) Change of Variable Approach:

The change of variable approach for determining the distribution can be summarized as follows. Let f(x) denote the known probability density of the random variable X. Let a new random variable be defined by Y = g(X). Then the density of Y can be written as

$$\boxed{h(y) = f(g^{-1}(y) = x)\left|\frac{dx}{dy}\right|} \tag{1.11}$$

where $\left|\frac{dx}{dy}\right|$ denotes the Jacobian of the transformation Y = g(X). The result given by (1.11) can be easily extended to the case of multivariate distributions.

Consider the following applications of (1.11).

(1.12)

If $X \sim N(\mu, \sigma^2)$, then

- $Y = e^X \sim LN(\mu, \sigma^2)$
- $Y^2 = (e^X)^2 \sim LN(2\mu, 4\sigma^2)$

If X ~ N(μ, σ²), then

(1.13) $$Z = \left[\frac{X-\mu}{\sigma}\right]^2 \sim \chi^2(1)$$

The proofs of (1.12) and (1.13) follow directly from (1.11) and have important implications in economics. A variablye Y is lognormally distributed LN(μ, $\sigma^2$), if the natural logarithm of that variable is N(μ, $\sigma^2$). The lognormal distribution has been used as a model for the size distribution of income. It is also important to know that the square of N(0,1) is $\chi^2(1)$.

Two other probability density functions which have recently been applied in the finance literature are the IHS and the g-h distributions. The IHS (inverse hyperbolic sine distribution) was introduced in the literature by Johnson (1949) and is defined by the transformation:

$$Y_{IHS} = a + b\sinh(\lambda + z/k) = a + b\left(\frac{e^{\lambda+z/k} - e^{-\lambda-z/k}}{2}\right)$$

and the g- and h- random variable $\left(Y_{g,h}\right)$ ,defined by John Tukey (1977) as follows:

$$Y_{g,h} = a + b\left(e^{gZ} - 1\right)\left(\frac{e^{hZ^2/2}}{g}\right)$$

where, in each case, Z denotes a standard normal or z statistic. The IHSand g-and-h distributions are both four parameter distributions. The EGB2 and g- and h-distribution have been used as generalizations of the famous Black Scholes Option Pricing formula. Each of these four parameter distributions allows for skewness and kurtosis common in many data series, particularly in financial data. See the papers by Bookstaber and McDonald (1991), Dutta and Babble (2005), McDonald and Bookstaber (1987), and Mauler and McDonald (2012) for more details.

(2) Moment Generating Function Approach:

A distribution function is uniquely characterized by its moment generating function (where defined). This one-to-one relationship provides the basis for using moment generating functions to determine the distributions of some transformations of random variables. Recall that the moment generating function of the random variable X is defined by

$$M_X(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx \tag{1.14}$$

$$= \sum_{i=0}^{\infty} t^i E(x^i)/i!.$$

If X and Y denote two independent random variables, then

$$M_{X+Y}(t) = E(e^{t(x+y)}) = E(e^{tx+ty}) \tag{1.15}$$

$$= E(e^{tx} \bullet e^{ty})$$

$$= E(e^{tx})E(e^{ty}) \text{ or}$$

$$\boxed{M_{X+Y}(t) = M_X(t)M_Y(t)}$$

(1.15) demonstrates that the moment generating function of the sum of two independent random variables is the product of the individual moment generating functions. The chi-square density provides an application of the moment generating function approach to determining the density of a transformation of random variables. If X ~ $\chi^2(\nu)$, then

$$M_X(t) = E(e^{tx}) = \int_0^{\infty} e^{tx} \left\{ \frac{e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}}{2^{\nu/2}\Gamma(\nu/2)} \right\} dx$$

$$= \int_0^{\infty} \frac{e^{-\frac{x}{2}(1-2t)} x^{\frac{\nu}{2}-1}}{2^{\nu/2}\Gamma(\nu/2)}$$

$$= \frac{1}{(1-2t)^{\nu/2}} \int_0^{\infty} \frac{e^{\frac{-x(1-2t)}{2}} (x(1-2t))^{\frac{\nu}{2}-1}}{2^{\nu/2}\Gamma(\nu/2)} (1-2t)\,dx$$

$$= \frac{1}{(1-2t)^{\nu/2}} \int_0^{\infty} \frac{e^{-s/2} s^{\nu/2-1}}{2^{\nu/2}\Gamma(\nu/2)} ds$$

$$= \frac{1}{(1-2t)^{\nu/2}} \tag{1.16}$$

which is the moment generating function for a $\chi^2(\nu)$.

(1.15) and (1.16) can be used to show that the sum of two independent chi-square variables is distributed as a chi-square with degrees of freedom equal to the sum of the degrees of freedom of the original independent chi-square variables. Let Z = X + Y where X and Y are independently distributed as $\chi^2(\nu_x)$ and $\chi^2(\nu_y)$ respectively. The moment generating function of Z is given by

$$M_Z(t) = E(e^{t(x+y)}) = M_X(t)M_Y(t)$$

$$= \left\{\frac{1}{(1-2t)^{\nu_x/2}}\right\}\left\{\frac{1}{(1-2t)^{\nu_y/2}}\right\}$$

$$= \frac{1}{(1-2t)^{(\nu_x+\nu_y)/2}};$$

which is the moment generating function of a chi-square with degrees of freedom $\nu_x+\nu_y$. Hence, the sum of two independently distributed chi-square variables is distributed as a chi-square with degrees of freedom equal to the sum of the degrees of freedom of the individual random variables. These are important results which will be used in the next section and link univariate results to some important relationships involving multivariate distributions.

## 2. Multivariate Distributions

The distribution of a vector of random variables $Y = (y_1, y_2, \ldots, y_n)'$ is referred to as a multivariate distribution. Many multivariate distributions have been considered and include the multivariate gamma, beta, Wishart, t, generalized t, and normal distributions. The multivariate normal is probably the most widely used and has some very useful properties. We first review the notion of the expected value and variance of a vector of random variables and then summarize some important properties of the multivariate normal. We conclude by defining some other multivariate distributions.

**a. Expectations.** Let Y denote an nxl vector of random variables and Z be an nxm matrix of random variables

The expected value of Y, mean vector, is defined by

$$E(Y) = (E(y_1), \ldots, E(y_n))' = \mu$$

The expected value of a matrix Z is defined by

$$\begin{pmatrix} E(z_{11}) & \cdots & E(z_{1m}) \\ \vdots & & \vdots \\ E(z_{n1}) & \cdots & E(z_{nm}) \end{pmatrix}$$

The variance covariance matrix associated with Y is defined by

$$\sum = E(Y - \mu)(Y - \mu)'$$

$$= \begin{pmatrix} E(y_1-\mu_1)^2 & \cdots & E(y_1-\mu_1)(y_n-\mu_n) \\ E(y_2-\mu_2)(y_1-\mu_1) & \cdots & E(y_2-\mu_2)(y_n-\mu_n) \\ \vdots & & \vdots \\ E(y_n-\mu_n)(y_1-\mu_1) & \cdots & E(y_n-\mu_n)^2 \end{pmatrix}$$

$$= \begin{pmatrix} Var(y_1) & Cov(y_1,y_2) & \cdots & Cov(y_1,y_n) \\ Cov(y_2,y_1) & Var(y_2) & \cdots & Cov(y_2,y_n) \\ \vdots & \vdots & & \vdots \\ Cov(y_n,y_1) & Cov(y_n,y_2) & \cdots & Var(y_n) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$$

$= \mathrm{Var}(Y)$

Note that the variance-covariance matrix includes the variances of the individual variables on the main diagonal and covariance between individual random variables in the corresponding off-diagonal elements.

**b.** **Multivariate Normal:** The vector $Y_{nxl}$ is said to be distributed as the multivariate normal with mean vector $\mu$ and variance-covariance matrix $\Sigma$, denoted $Y \sim N(\mu,\Sigma)$, if the density of Y is given by

$$N(Y;\mu,\Sigma) = \frac{e^{-(Y-\mu)'\Sigma^{-1}(Y-\mu)/2}}{(2\pi)^{n/2}|\Sigma|^{1/2}}.$$

The multivariate normal includes the univariate (n=1) and bivariate (n=2) normal as special cases. For example, if n = 1,

$Y = (y_1)$, $\mu = (\mu)$, $\Sigma = (\sigma^2)$ and

$$N(y_1;\mu,\sigma^2) = \frac{e^{-1/2(y_1-\mu)(1/\sigma^2)(y_1-\mu)}}{\sqrt{2\pi}\ \sqrt{\sigma^2}}.$$

$$= \frac{e^{-(y_1-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\ \sqrt{\sigma^2}}$$

The reader may want to consider the case n = 2.

**c.** **Distribution Theory for the normal** (See Appendix A for proofs of the results contained in this section)

(1) Chi Square

(a) Let $y_1, \ldots, y_n$ be independently distributed as $N(0,\sigma^2)$, i.e.,

$Y = (y_1, \ldots, y_n)' \sim N(0,\sigma^2 I)$, then

$$Y(\sigma^2 I)^{-1}Y = \frac{Y'Y}{\sigma^2} = \sum_{i=1}^{n} \frac{y_i^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{y_i-0}{\sigma}\right)^2 \sim \chi^2(n).$$

The sum of squares of n independent standard normal variables is distributed as a chi-squared variable with n degrees of freedom.

(b) If $Y \sim N(\mu,\sigma^2 I)$, then

$$(Y-\mu)'(\sigma^2 I)^{-1}(Y-\mu) = \frac{(Y-\mu)'(Y-\mu)}{\sigma^2} \sim \chi^2(n).$$

Note this is the quadratic form which appears in the density function for Y.

(c) **More generally, If $Y \sim N(\mu,\Sigma)$, then**

$$(Y - \mu)' \Sigma^{-1}(Y - \mu) \sim \chi^2(n).$$

**Note: This is the quadratic form which appears in the exponent of the multivariate normal. Hint: take the "square root" of the matrix $\Sigma$, Cholesky decomposition.**

(2) Some Useful Theorems

(a) **If $Y \sim N(\mu_y, \Sigma_y)$, then $Z = AY \sim N(\mu_Z = A\mu_Y; \Sigma_Z = A\Sigma_Y A')$**

**where A is a matrix of constants.**

(b) **If $Y \sim N(O,I)$ and A is a symmetric idempotent matrix, then $Y'AY \sim \chi^2(m)$, where the degrees of freedom (m) = trace(A).**

Proof: Diagonalize A. Recall that the characteristic roots of an idempotent matrix are zero or one.

(c) **If $Y \sim N(O,I)$, then the idempotent quadratic forms $Y'AY$ and $Y'BY$ are independently distributed $\chi^2$ variables if $AB = 0$.**

Proof: $Y'AY = Y'A'AY$ and $Y'BY = Y'B'BY$. The covariance matrix of AY and BY is AB.

(d) If $Y \sim N(O,I)$ and L is a kxn matrix of rank k, then LY and the idempotent quadratic form $Y'AY$ are independently distributed if $LA = 0$.

Proof: Similar to (c). $Y'AY = Y'A'AY$. The covariance matrix of LY and AY is LA.

See Appendix A for more details.

(3) A simple example (see exercise 9, 10)

If $Y \sim N(\mu,\sigma^2 I)$, then

(a) $$\frac{(Y-\mu)'(Y-\mu)}{\sigma^2} \sim \chi^2(n);$$

(b) $$\frac{(n-1)s^2}{\sigma^2} = \Sigma\frac{(y_t-\overline{Y})^2}{\sigma^2} = Y'(I-(1/n)\mathbf{ii'})(I-(1/n)\mathbf{ii'})Y/\sigma^2$$

$$= Y'(I - 1/n\mathbf{ii'})Y/\sigma^2 \sim \chi^2(n-1)$$

where **i** denotes a nx1column of 1's, thus **ii'** is an nxn matrix of 1's. It is also important to show that $(I - (1/n)\mathbf{ii'})$ is symmetric and idempotent and has trace = n-1

Recall $\cdot E(\chi^2(d)) = d$ (degrees of freedom)

$\cdot s^2$ is an unbiased estimator of $\sigma^2$

$$\cdot\ E(n-1)\frac{S^2}{\sigma^2} = (n-1)\frac{E(s^2)}{\sigma^2} = n-1$$

(4) t distribution

The ratio of a standard normal variable divided by the square root of an independent chi-Square variable divided by its degrees of freedom is distributed as a t statistic with the same degrees of freedom as the chi-square. This might be represented:

$$\boxed{\frac{N(0,1)}{\sqrt{\chi^2(d)/d}} \sim t(d)}$$

The t-density is symmetric about zero, has variance [d/(d-2)] and normalized kurtos 3 + 6/(d-4); hence, the tails of the t are thicker than the normal, but approach the normal N(0,1) as $d \to \infty$.

(5) F distribution

The ratio of two <u>independent</u> chi-square variables divided by their respective degrees of freedom is distributed as an F statistic with degrees of freedom equal to those associated with the chi-square variables. This can be represented as

$$\boxed{\frac{\chi^2(d_1)/d_1}{\chi^2(d_2)/d_2} \sim F(d_1,d_2)}$$

<u>Exercise</u>: Show that the mean of the F distribution is given by $d_2/(d_2 - 2)$. Can you provide any intuition why this is close to one for large $d_2$? Hint: $F = \left(\frac{d_2}{d_1}\right)(\chi^2(d_1))(\chi^2(d_2))^{-1}$. order moment of a chi-square variable is $E(\chi^2(d))^h = 2^h\Gamma(d/2+h)/\Gamma(d/2)$. Let h=1 for the first chi-square and h=-1 for the second. The same "trick" can be used to evaluate the moments of the t(d). For example,

$$Var(t(d)) = E(t(d))^2 = E[(N(0,1))^2(\chi^2(d))^{-1}]d$$

$$= (1)(d)E(\chi^2(d))^{-1} = d(2)^{-1}\Gamma(d/2 -1)/\Gamma(d/2) = d/ (d-2).$$

**d. Multivariate generalizied error and multivariate generalized t distributions**

(1) **Multivariate generalized error distribution (MGED), multivariate power exponential (MPEXP), or multivariate Box-Tiao distribution (MBT)** of an nx1 vector Y is defined by the pdf

$$MGED(y,\mu,\Sigma,p)=\left(\frac{n\Gamma(n/2)e^{-\left((y-\mu)'\Sigma^{-1}(y-\mu)\right)^{p}/2}}{|\Sigma|^{1/2}\pi^{n/2}\Gamma\left(1+\frac{n}{2p}\right)2^{1+n/2p}}\right)$$

If n=1, the univariate GED results. The multivariate normal corresponds to p=2. For additional details, see [Gomez, E., M.A. Gomez-Villegas, and J.M. Marin, 1998. "A Multivariate generalization of the power exponential family of distributions," Communications in Statistics: Theory and Methods, 27(3), 589-600.

(2) **Multivariate generalized t (MGT)**. Arslan (2001) extended the method used by McDonald and Newey in obtaining the generalized t to derive a multivariate generalized t. The pdf of the MGT is given by

$$MGT(y;\mu,\Sigma,\sigma,p,q)=\left(\frac{p\Gamma\left(\frac{n}{2}\right)q^{q}}{\pi^{n/2}B(q,n/2p)\left(q+\left((y-\mu)'\Sigma^{-1}(y-\mu)/\sigma\right)^{p}\right)^{q+n/2p}}\right)$$

The MGED and MGT are members of elliptically contoured random variables, see (Fang, K. and Y. Zhang, 1990. Generalized Multivariate Analysis, Springer Verlag. Beijing) for additional details on elliptical distributions. When $\sigma = p=1, \Sigma=2\Sigma_1, \text{ and } 2q=\nu$

standard multivariate t-distribution with location and *scatter* parameters $\mu \text{ and } \Sigma_1$ and degrees

of freedom $\nu$ . The multivariate normal is obtained from the MGT by letti $p=\sigma=1$ and

$q\rightarrow\infty$ . The MGED is obtained from the MGT by letting q grow indefinitely large. Finally, as

$\sigma \to \infty$ the MGT approaches a multivariate generalization of the uniform distribution.

Arskan(2001) reports expressions for the first four moments of MGT. We report the first two moments in these notes:

- $E(Y) = \mu$

- $Var(Y) = \left( \dfrac{\sigma q^{1/p} \Gamma\left(\dfrac{n+2}{2p}\right) \Gamma(q - 1/p)}{n \Gamma\left(\dfrac{n}{2p}\right) \Gamma(q)} \right) \Sigma$

Reference: Arslan, Olcay, 2001. "Family of Multivariate Generalized t-Distributions," working paper, University of Cukurova, Turkey. Submitted to Journal of Multivariate Analysis.

**(3) Multivariate generealized beta distributions (MGB1 and MGB2)**. There are often several ways multivariate generalizations of univariate distributions can be formed which yield the same marginal distributions. An extension of the GB1 and GB2 are the MGB1 and MGB2 given as follows

$$MGB1(y; a, b, p, q) = \frac{\left(\Pi_{i=1}^{n} |a_i| y_i^{a_i p_i - 1}\right)\left(1 - \sum_{i=1}^{n} (y_i / b_i)^{a_i}\right)^{q-1}}{\left(\Pi_{i=1}^{n} b_i^{a_i p_i}\right) B(p_1, p_2, \ldots, p_n, q)}$$

$$for\ 0 < \sum_{i=1}^{n} (y_i / b_i)^{a_i} < 1$$

$$MGB2(y; a, b, p, q) = \frac{\left(\Pi_{i=1}^{n} |a_i| y_i^{a_i p_i - 1}\right)}{\left(\Pi_{i=1}^{n} b_i^{a_i p_i}\right) B(p_1, p_2, \ldots, p_n, q)\left(1 + \sum_{i=1}^{n} (y_i / b_i)^{a_i}\right)^{\overline{p}+q}}$$

$$for\ 0 < y_i$$

where $B(p_1, p_2, \ldots, p_n, q) = \dfrac{\Gamma(p_1)\Gamma(p_2)\ldots\Gamma(p_n)\Gamma(q)}{\Gamma(\overline{p} + q)}$ with $\overline{p} = \sum_{i=1}^{n} p_i$.

Multivariate GG, EGB1, and EGB2 distributions can be defined in a similar manner. The pdf for the MEGB2 is given by

$$MEGB2\left(z;\delta,\sigma,p,q\right)=\frac{\left(\Pi_{i=1}^{n}e^{p_i\left(z_i-\delta_i\right)/\sigma_i}\right)}{\left(\Pi_{i=1}^{n}\left|\sigma_i\right|\right)B\left(p_1,p_2,\ldots,p_n,q\right)\left(1+\sum_{i=1}^{n}e^{\left(z_i-\delta_i\right)/\sigma_i}\right)^{p+q}}$$

$$for\ -\infty<y_i<\infty$$

Expressions for the moments are reported in McDonald (Some Multivariate Generalized Beta Distributions, Working Paper, February 1993). Other forms of multivariate beta distributions are summarized in several books by Johnson and Kotz dealing with continuous multivariate distributions.

## 3. Estimation

### a. **Maximum likelihood estimation: basic theory and results**

Assume that the random variables $y_t$, t = 1, 2, . . ., n are independently and identically distributed with the probability density function $f(y_t;\theta)$ where $\theta' = (\theta_1, \theta_2, \ldots, \theta_k)$ denotes a vector of unknown parameters. We will review two important methods of estimating the unknown parameters. In comparing alternative estimators of the same parameter we will want to compare their statistical properties. **Three important criteria** to consider in such a comparison are **bias**, **consistency**, and **relative efficiency**. Rcall that $\hat{\theta}$ is an *unbiase*d estimator if

$$E\left(\hat{\theta}\right) = \theta$$

.

*Consistency* requires that $p\lim_{n\to\infty} \hat{\theta} = \theta \text{ or } \lim_{n\to\infty} \Pr\left(\left|\hat{\theta}-\theta\right| \geq \varepsilon\right) = 0 \text{ for any } \varepsilon > 0$

If the variance and bias of $\hat{\theta}$ approach ze $\hat{\theta}$ will be consistent; however, there are cases in which the variance/bias may not be defined and the estimator is still consistent. An unbiased estimator is said to be efficient if its variance is less than the variance of any other unbiased estimator. In the case of a vector of estimators, one would investigate whether the difference of the variance covariance matrices are positive definite. In the case of biased estimators, comparisons can be made using the mean squared error which is equal to sum of the variance and square of the bias, i.e. $MSE\left(\hat{\theta}\right) = VAR\left(\hat{\theta}\right) + BIAS^2\left(\hat{\theta}\right)$

We now turn to the development of maximum likelihood estimation and then we briefly summarize the notion of extremum estimators.

The likelihood function associated with the random sample $\{y_t t = 1, 2, \ldots, n\}$ is given by

$$(3.1) \quad L(Y;\theta) = \prod_{t=1}^{n} f(y_t;\theta)$$

with corresponding <u>log likelihood function</u>

$$(3.2)\qquad \ell(Y;\theta) = \ln(L(Y;\theta))$$

$$= \sum_{t=1}^{n} \ln f(y_t;\theta)$$

$$= \sum_{t=1}^{n} \ell(y_t;\theta),$$

where $\ell(y_t;\theta) = \ln f(y_t;\theta)$.

The maximum likelihood estimators (MLE) of the parameters $\theta$ are implicitly defined by the equation

$$(3.3)\qquad \frac{d\ell(Y;\theta)}{d\theta} = \begin{pmatrix} \frac{\partial \ell}{\partial \theta_1} \\ \vdots \\ \frac{\partial \ell}{\partial \theta_k} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = 0.$$

These estimators will optimize the log likelihood function $\ell(y;\theta)$. It is possible that multiple solutions to (3.3) will exist; however, a unique maximum will exist if $\ell(Y;\theta)$ is concave in $\theta$. The concavity can be investigated by determining whether the matrix

$$\frac{d^2\ell(Y;\theta)}{d\theta^2} = \begin{pmatrix} \ell_{11} & \cdots & \ell_{1k} \\ \vdots & & \\ \ell_{k1} & \cdots & \ell_{kk} \end{pmatrix}$$

is negative definite where $\ell_{ij} = \dfrac{\partial^2 \ell(Y;\theta)}{\partial\theta_i \partial\theta_j}$.

Under rather general regularity conditions the MLE of θ, $\hat{\theta}$ , will be asymptotically normally distributed.

The regularity conditions will be given and the main results cited. The literature contains a number of alternative formulations of "regularity conditions."

Regularity Conditions:

Let Θ denote the set of permissible values of θ

(R.1) For almost all $y_t$

$$\frac{d\ell(y_t;\theta)}{d\theta}, \quad \frac{d^2\ell(y_t;\theta)}{d\theta^2}, \quad \frac{d^3\ell(y_t;\theta)}{d\theta^3}$$

exist for all θ ε Θ

(R.2) There exist integrable functions $F_1(y)$, $F_2(y)$ and $H(y)$ such that

$$\frac{d\ell(y_t;\theta)}{d\theta} < F_1(y_t)$$

$$\frac{d^2\ell(y_t;\theta)}{d\theta^2} < F_2(y_t)$$

$$\frac{d^3\ell(y_t;\theta)}{d\theta^3} < H(y_t)$$

for all θ ε Θ

and

$$E(H(y)) = \int_{-\infty}^{\infty} H(y) f(y;\theta)\, dy < M$$

where M is independent of θ.

$$\text{(R.3)} \quad E\left[\frac{\partial\ell(y_t;\theta)}{\partial\theta}\right]^2 = \int_{-\infty}^{\infty}\left[\frac{\partial \ln f(y;\theta)}{\partial\theta}\right]^2 f(y;\theta)\, dy < +\infty.$$

These assumptions are important in demonstrating the asymptotic normality of MLE of θ. R.1 insures the existence of necessary Taylor Series expansions while R.2 permits differentiation under an integral and (R.3) implies that the random variable(s) $\left\{\frac{\partial \ln f(y_t;\theta)}{\partial\theta}\right\}$ will have a finite variance.

Theorem. Under the regularity conditions (R.1), (R.2) and (R.3) the MLE of θ will be consistent and asymptotically normal.

$$\boxed{\theta_{MLE} \overset{a}{\sim} N[\theta, \Sigma_{\theta_{MLE}}]} \tag{3.4}$$

where

$$\Sigma_{\theta_{MLE}} = -\left(\frac{Ed^2\ell}{d\theta d\theta'}\right)^{-1}$$

$$= \left(E\left(\frac{d\ell}{d\theta}\right)\left(\frac{d\ell}{d\theta}\right)\right)^{-1}.$$

This distributional result can be alternatively written

$$\boxed{\sqrt{n}\left(\theta_{MLE} - \theta\right) \overset{a}{\sim} N[0, n\Sigma_{\theta_{MLE}}].}$$

Appendix B contains a proof of this important result:

The proof follows from the basic condition for probability density functions $\int L(Y;\theta)dY = 1$ and leads to useful intermediate results that

$$E\left(\frac{d\ell(Y;\theta)}{d\theta}\right) = 0$$

$$\mathrm{Var}\left(\frac{d\ell(Y;\theta)}{d\theta}\right) = E\left(\frac{d\ell(Y;\theta)}{d\theta}\frac{d\ell(Y;\theta)}{d\theta'}\right) = \Sigma_{MLE}^{-1}$$

Two excellent references to this material are Rao [1973] and Theil [1971].

**In another section , we will explore how computer programs such as STATA can be used to obtain MLE.**

**Examples of MLE estimators and their asymptotic distributions**

We now consider two examples. The first example is based upon the **power density** function

$f(x) = px^{p-1}$ for $0 < x < 1$

$= 0$ otherwise.

The log likelihood function is

$$\ell = \ln\left\{\prod_{t=1}^{n} f(x_t)\right\} = \ln\left\{p^n\left(\prod_{t=1}^{n} x_t\right)^{p-1}\right\}$$

$$= n \ln p + (p-1)\sum_{t=1}^{n} \ln X_t .$$

The necessary condition for a maximum of $\ell$ is

$$\frac{d\ell}{dp} = \frac{n}{p} + \sum_{t=1}^{n} \ln X_t = 0$$

which yields

$$\hat{p} = -[n / (\Sigma \ln X_t)] .$$

The asymptotic distribution of $\hat{p}$ depends on

$$-E\frac{d^2\ell}{dp^2} = -E\left\{\frac{-n}{p^2}\right\} = \frac{n}{p^2};$$

hence

$$\hat{p} \overset{a}{\sim} N\left[p, p^2/n\right]$$ or

$$\sqrt{n}\ (\hat{p} - p) \overset{a}{\sim} N\left[0, p^2\right].$$

(2) **The normal.** The second example involves the derivation of MLE of the two parameters in $N(X;\mu,\sigma^2) = \dfrac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\Pi}\sigma}$. The loglikelihood function is given by

$$\ell = \frac{-n}{2}\ln(2\Pi) - \frac{n}{2}(\ln \sigma^2) - \frac{1}{2}\sum_t\left[\frac{(X_t-\mu)^2}{\sigma^2}\right].$$

Differentiating $\ell$ with respect to the parameters to be estimated yields:

$$\frac{\partial\ell}{\partial\mu} = \frac{1}{\sigma^2}\sum_t (X_t-\mu) = 0$$

$$\frac{\partial\ell}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_t (X_t-\mu)^2 = 0.$$

Solving these equations yields

$$\hat{\mu} = \overline{X} \qquad \text{and}$$

$$\hat{\sigma}^2 = \sum_{t=1}^{n} (X_t - \overline{X})^2/n$$

.

The asymptotic-variance covariance matrix of $\hat{\mu}$ and $\hat{\sigma}^2$ can be obtained from the Hessian matrix associated with $\ell$.

$$H = \begin{pmatrix} \dfrac{d^2\ell}{d\mu^2} & \dfrac{d^2\ell}{d\sigma^2 d\mu} \\ \dfrac{d^2\ell}{d\mu d\sigma} & \dfrac{d^2\ell}{d^2\sigma^2} \end{pmatrix} = \begin{pmatrix} -n/\sigma^2 & \dfrac{-\sum_{t=1}^{n}(X_t-\mu)}{\sigma^4} \\ \dfrac{-\sum_{t=1}^{n}(X_t-\mu)}{\sigma^4} & \dfrac{n}{2\sigma^4} - \dfrac{2\sum_{t=1}^{n}(X_t-\mu)^2}{2\sigma^6} \end{pmatrix};$$

hence

$$E(H) = \begin{pmatrix} -n/\sigma^2 & 0 \\ 0 & -n/2\sigma^4 \end{pmatrix}.$$

The asymptotic variance matrix of ($\hat{u}$, $\hat{\sigma}^2$) is

$$(-E(H))^{-1} = \begin{pmatrix} \sigma^2/n & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix};$$

hence,

$$\sqrt{n}\left[\begin{pmatrix} \hat{\mu}-\mu \\ \hat{\sigma}^2-\sigma^2 \end{pmatrix}\right] \underset{\sim}{a} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix};\begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}\right]$$

or

$$\begin{bmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{bmatrix} \underset{\sim}{a} N\left[\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix};\begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{pmatrix}\right].$$

**b. Method of moments** (MOM)

Let the data generating process (DGP) be defined by the pdf f(y, $\theta_i$ i = 1, 2, .. p). Method of moments estimators of the unknown parameters are obtained by selecting parameter estimators which equate sample moments and theoretical moments. We generally use the same number of moments as parameters. If more moments are used than parameters, generalized methods of moments (GMM) can be used.

To formalize the method, let

$$\hat{m}_h = \frac{1}{n}\sum_{i=1}^{n} Y_i^h \text{ and } m_h(\theta) = E\left(Y^h\right)$$

The MOM estimators are then obtained by solving:

$\Sigma\, Y^h_i/n = E(Y^h) = m_h(\hat{\theta})$, h = 1, 2, .. p, for the vector of parameter estimates $\hat{\theta}$.

Examples:

Exponential: (one parameter)

Sample mean ($\bar{Y}$)= E(Y) = β and solve

Power distribution:

The method of moments estimator of p ($\tilde{p}_{mom}$) in the power distributioin is found from

$$E(x) = \int_0^1 x\{px^{p-1}\}dx = \frac{\tilde{p}}{\tilde{p}+1} = \bar{x};$$

hence

$$\tilde{p}_{mom} = \frac{\bar{x}}{1-\bar{x}}.$$

The expected value and variance of $\bar{X}$ are given by $\left(\frac{p}{p+1}\right)$ and $p/((p+1)^2(p+2)n)$. Using the Slutsky Theorem: $\tilde{p}_{mom}$ can be shown to be consistent

$$\text{plim}\left(\tilde{p}_{mom}\right) = \frac{\text{plim}(\bar{X})}{1-\text{plim}(\bar{X})}$$

$$= \frac{\frac{p}{p+1}}{1-(p/p+1)} = p.$$

We will consider ways of determining its asymptotic distribution in another section.

Normal: two parameters

Set the sample mean = $\mu$ and the sample $E(Y^2) = \sigma^2 + \mu^2$ or Sample Var = $\sigma^2$

and solving for $\hat{\mu}$ and $\hat{\sigma}^2$ .

Gamma: two parameters

$E(Y) = \boldsymbol{\beta\, \Gamma(p+1)/\Gamma(p) = \beta\, p}$

$E(Y^2) = \boldsymbol{\beta^2\, \Gamma(p+2)/\Gamma(p) = \beta^2(p+1)(p)}$

**In summary**, method of moments estimators are obtained by solving solve

$$\hat{m}_h = m_h\left(\hat{\theta}\right) \text{ for h=1, 2, ...., p}$$

for $\hat{\theta}$ where there are p parameters. Method of Moment estimators are generally:

- Consistent
- Not always defined (equations may not have a solution)
- May involve nonlinear equations

**Alternatively**, let

$$g(\theta) = \left(m(\theta) - \hat{m}\right)$$

where $\hat{m} = \left(\hat{m}_1, \hat{m}_1, \ldots \hat{m}_p\right)'$ & $m(\theta) = \left(m_1(\theta), m_2(\theta), \ldots, m_p(\theta)\right)'$

and, as before, $\hat{m}_h = \dfrac{1}{n}\sum_{i=1}^{n} Y_i^h$ and $m_h(\theta) = E\left(Y^h\right)$

can be obtained by solving any of the following problems:

$$(1) \quad g\left(\hat{\theta}\right) = 0$$

$$(2) \quad \min_\theta \left(g(\theta)\right)' g(\theta)$$

$$(3) \quad \min_\theta \left(g(\theta)\right)' W g(\theta)$$

where W is a positive definite weighting matrix. If there is a method of moments estimator (with the same number of moment restrictions as parameters) it is given as the solution to (1) and will be the same as the solution to problems (2) or (3) which will yiled a value of zero for the corresponding quadratic forms. If there are more moment conditions (m) than parameters (p), there will generally not be a solution to (1) and the solution to (2) and (3) will be referred to as the **generalized method of moments** (GMM) estimator.

For example, the mean and variance of the exponential pdf can be shown to be

$\beta$ and $\beta^2$ $\quad g(\beta) = \begin{pmatrix} \overline{Y} - \beta \\ s^2 - \beta^2 \end{pmatrix}$ and there is not a method o

unless the sample variance is equal to the square of the sample mean. The GMM estimator would be obtained by solving equation (2) or (3) above for $\beta$ which seeks to make the entries in g( ) as small as possible, with small being measured by the value of the quadratic form.

**c. Kernel Estimators**

Kernel Estimators are referred to as non-parametric because no "distributional parameters" need be estimated. The kernel estimator might be thought of as "high-tech" histogram. However; the user still needs to make a decision about the *kernel* to use and about a *window width* to use. We will merely present some basic building blocks here and in class will show an example of the use of a kernel to describe stock returns.

Denote n observations on the random variable Z as $z_i\ i = 1, 2, \ldots, n.$ The empirical cdf can then be defined as :

$$\hat{F}(z) = \Pr(Z<z)\left(\frac{\#\text{observations} \le z}{\text{Total number of observations}}\right)$$

The corresponding empircal pdf (histogram) with "window width" or "band width" h is defined by

$$\hat{f}_h(z) = \frac{\hat{F}(z+h) - \hat{F}(z-h)}{2h}$$

- the empirical pdf gives the fraction of observations which lie within a neighborhood of width ***h*** of z. *h* is referred to as the bandwidth.
- Increasing h makes the pdf smoother and decreasing h makes the pdf choppier

An alternative way of expressing the histogram is as follows:

$$\hat{f}_h(z) = \left(\frac{1}{nh}\right)\sum_{i=1}^{n} I\left(-\frac{1}{2} \le \frac{z - z_i}{h} \le \frac{1}{2}\right)$$

where *I( )* denotes an indicator function which equals 1 if the inequality is valid and zero otherwise, ( Pagan and Ullah).

If the indicator function is replaced by a pdf, say *K( )*, the corresponding function is referred to as a kernel estimator

$$\hat{f}_h(z) = \left(\frac{1}{nh}\right)\sum_{i=1}^{n} K\left(\frac{z - z_i}{h}\right)$$

with bandwidth *h* and kernel *K( )*.

An **alternative development of the kernel density** estimator can be structured in terms of a new variable W defined by $W = Z + hU$ where

- U is continuous with pdf K(u)
- h is the bandwith
  - Small h, W behaves like the data
  - Large h, W behaves like U

The CDF of W (you can skip this and go straight to the kernel pdf) is given by

$$F_W(w) = Pr(W \le w)$$

$$= Pr(Z + hU \le w) = Pr\left(U \le \frac{w - Z}{h}\right)$$

$$\frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\frac{w-z_i}{h}} K(u)du$$

How would we estimate the pdf of Z? (Use Leibnitz Rule).

The kernel estimator of the pdf is given by the following equation:

$$\hat{f}_h(z) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{z - z_i}{h}\right)$$

- K(u) is the Kernel (often the standard normal, but there are other choices)
- It is easily verified by integrating over z. that the empirical pdf is in fact a pdf.
- h is the bandwidth. As h increases, the pdf gets smoother, variance decreases, but the bias increases. Silverman suggests: the following rule of thumb:

  h=$1.06\hat{\sigma}n^{-.2}$ as the optimal bandwidth.

**STATA** facilitates kernel estimation of density functions with a number of different kernels and window widths. The default kernel density uses a Epanechnikov kernel and optimal window width.

The command is

**kdensity " variable name"**

**kdensity "variable name", normal** overlays the kernel density and a fitted normal

**histogram "variable name", bin(b) normal kdensity** overlays the normal and kernel on the histogram with *b* bins

# 4. Large Sample Theory

Reference: Amemiya [1985].

Let $\{X_n\}$ denote a sequence of random variables. An example of a sequence of random variables is the sample mean based upon a random sample of size n. Another example is the estimator of a regression coefficient based on a sample of size n. We will discuss (a) four different types of convergence of a sequence of random variables and the relationships between them and then consider some (b) central limit theorems which help us identify approximate distribution functions for the sample mean drawn from different populations.

**a. Modes of convergence**

(1) Convergence in probability.

A sequence of random variables $\{X_n\}$ is said to converge to a random variable X (could be a constant) <u>in probability</u> if

$$\lim_{n \to \infty} P_r(|X_n - X| > \varepsilon) = 0$$

for any $\varepsilon > 0$. This is frequently written as

$$\boxed{X_n \xrightarrow{P} X}$$

or

$$\boxed{\operatorname*{plim}_{n \to \infty} X_n = X}$$

(2) Convergence in mean square.

$\{X_n\}$ is said to converge to X in <u>mean square</u> if $\lim_{n \to \infty} E(X_n - x)^2 = 0$ which is frequently denoted

$$\boxed{X_n \xrightarrow{M} X.}$$

Recall that $E(X_n - X)^2 = E(X_n - E(X)^2 + (E(X_n) - X)^2 = Var(X_n) + Bias^2(X_n)$. Hence, convergence in mean square requires that the variance and bias approach zero as n grows indefinitely larger.

(3) Convergence in distribution.

$\{X_n\}$ is said to converge to X <u>in distribution</u>, $X_n \to X$, if

$\lim_{n\to\infty} F_n(X) = F(X)$ for every continuity point of $F(X)$ where

$F_n(\ )$ and $F(\ )$, respectively, denote the distributions of $X_n$ and $X$.

(4) Convergence almost everywhere.

$\{X_n\}$ is said to converge to X almost everywhere,

$$\boxed{X_n \xrightarrow{a.e.} X,}$$

if $\Pr\{w \mid \lim_{n\to\infty} X_n(w) = X(w)\} = 1$

where $X_n(w) = X(w)$ denotes the events in the sample space which are associated with both random variables having the same value.

(5) Relationship between the modes of convergence.

There is a hierarchical relationship between the modes of convergence which Amemiya depicts as

$$\begin{array}{ccccc} & & a.e & & \\ & & \downarrow & & \\ M & \longrightarrow & P & \longrightarrow & d \end{array}$$

Consequently, convergence in mean square or almost everywhere will imply convergence in probability and in distribution.

(6) Some Theorems

1. If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} Y$, then

$$X_n + Y_n \xrightarrow{D} X + Y$$

$$X_n Y_n \xrightarrow{D} X \bullet Y$$

$$\frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{Y}$$

provided $\Pr(Y = 0) \neq 0$.

2. Slutsky's Theorem. Let $X_n$ denote a vector of random variables and $g(\ )$ a real valued continuous function, then $X_n \xrightarrow{P} \alpha$ implies $g(X_n) \xrightarrow{P} g(\alpha)$, where $\alpha$ is a constant vector.

$$\boxed{\operatorname*{plim}_{n\to\infty} g(X_n) = g(\operatorname*{plim}_{n\to\infty} X_n)}$$

D P

3. If $X_n \rightarrow X$ and $Y_n \rightarrow \alpha$, then

$$X_n + Y_n \xrightarrow{D} X + \alpha$$

$$X_n \bullet Y_n \xrightarrow{D} \alpha X$$

$$\frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{\alpha} \text{ if } \alpha \neq 0.$$

**b. Law of Large Numbers (LLN) and Central Limit Theorems (CLT)**

(1) Law of Large Numbers (LLN)

A Law of Large Numbers specifies conditions under which difference between the sample mean ($\bar{X}$) and its expected value will converge to zero.

a. Kolmoqorov LLN1

Let $\{X_t\}$ be independent with finite variance, $Var(X_t) = \sigma_t^2$. If

$$\sum_{t=1}^{\infty} \sigma_t^2/t^2 < \infty,$$

then $\bar{X}_n - E\bar{X}_n \xrightarrow{a.e.} 0.$

b. Kolmoqorov LLN2

Let $\{X_t\}$ be independently distributed. Then a necessary <u>and</u> sufficient condition that $\bar{X}_n \xrightarrow{a.e.} \mu$ is

that $E(X_t)$ exists and is equal to $\mu$.

(2) Central Limit Theorems

Suppose $\bar{X}_n - E(\bar{X}_n) \xrightarrow{P} 0$, then we can conclude that
D

It is more interesting to investigate distributions which approximate that of

$$Z_n = \frac{\bar{X}_n - E(X_n)}{\sqrt{Var(\bar{X}_n)}}$$

Central limit theorems (CLT) specifies conditions under which $Z_n$ will converge to N(0,1), a standard normal.

Three central limit theorems will be stated and the reader is referred to Amemiya for additional discussion and references.

a. Lindeberg-Levy CLT

Let $\{X_t\}$ be independently and identically distributed

with $E(X_t) = \mu$ and $Var(X_t) = \sigma^2$, then

$$Z_n \xrightarrow{D} N(0,1).$$

b. Liapounov CLT

Let $\{X_t\}$ be independent with $E(X_t) = \mu_t$, $Var(X_t) = \sigma_t^2$

and $E|X_t - \mu_t|^3 = \rho_t^3$.

If

$$\lim_{n\to\infty} \frac{\left(\sum_{t=1}^{n} \rho_t^3\right)^{1/3}}{\left(\sum_{t=1}^{n} \sigma_t^2\right)^{1/2}} = 0,$$

then

$$Z_n \xrightarrow{P} N(0,1).$$

c. Lindeberg-Feller CLT

Let $\{X_t\}$ be independent with distribution $F_t$, $E(X_t) = \mu_t$, $Var(X_t) = \sigma_t^2$. Define

$$C_n = \left(\sum_{t=1}^{n} \sigma_t^2\right)^{1/2}.$$

$$\text{If } \underset{n\to\infty}{\text{Limit}} \left(\frac{1}{C_n^2}\right) \sum_{t=1}^{n} \int_{A_n} (x-\mu_t)^2 dF_t(X) = 0.$$

with $A_n = \{X \mid |X - \mu_t| > \varepsilon C_n\}$

for all $\varepsilon > 0$,

then

$$Z_n \xrightarrow{D} N(0,1).$$

**References**

Amemiya, T. Advanced Econometrics, Cambridge: Harvard University, 1985.

Arnold, B. C. (1983). Pareto Distributions, Bartonsville: International Cooperative Publishing House.

Bookstaber, R. M. and J. B. McDonald, A general Distribution for Describing Security Price Returns," Journal of Business, 60(1987), 401-424.

Dutta, Kabir K. and D. F. Babbel, "Extracting Probabilitistic Information from the Prices of Interest Rate Options: Tests of Distribuitonal Assumptions," Journal of Business, 78(2005), 841-870.

Glaser, Ronald E. "Bathtub and Related Failure Rate Characterizations," Journal of the American Statistical Association, 75(1980), 667-672.

Goldberger, A. S. (1964). Econometric Theory, New York: Wiley.

Graybill, F. A. (1961). An Introduction to Linear Statistical Models, New York: McGraw-Hill.

Huber, P. J. (1981). Robust Statistics, New York: Wiley.

Intriligator, M. (1977). Econometric Models, Techniques, and Applications, Englewood Cliffs, N.J.: Prentice-Hall.

Johnson, N. L. (1949). "Systems of frequency curves generated by methods of translation," Biometrica 36, 149-176.

Johnson, N. L. and S. Kotz. (1970). Continuous Univariate Distributions, Vols. 1, 2, New York: John Wiley & Sons.

Johnston, J. (1972). Econometric Methods, New York: McGraw-Hill.

Kalbfleisch, J. D. and R. L. Prentice. (1980). The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons.

Kmenta, J. (1971). Elements of Econometrics, New York: Macmillan.

Lindgren, B. W. (1976). Statistical Theory, 3rd ed., New York: Macmillan.

Maddala, G. S. (1977). Econometrics, New York: McGraw-Hill.

Mauler,D. and J. B. McDonald, "Option Pricing and Distribution Characteristics," working paper, 2012.

McDonald, J. B. "Some Generalized Functions for the Size Distribution of Income," Econometrica, 52(1984), 647-663.

McDonald, J. B. and R. M. Bookstaber, "Optin Pricing for Generalized Distributions," Communications in Statistics--Theory and Method, 20(1991), 4053-4068.

McDonald, J. B. and R. J. Butler. "Some Generalized Mixture Distributions with an Application to Unemployment Duration," Review of Economics and Statistics, 69(1987), 232-240.

McDonald, J. B. and D. O. Richards. "Hazard Functions and Generalized Beta Distributions," IEEE Transactions on Reliability, 36(1987b), 463-466.

McDonald, J. B. and D. O. Richards. "Model Selection: Some Generalized Distributions," Communications in Statistics, 16(1987), 1049-1074.

McDonald, J. B. and W. K. Newey. "Partially Adaptive Estimation of Regression Models Via the Generalized T Distribution," Econometric Theory, 4(1988), 428-457.

McDonald, J.B. and Y. Xu, "A Generalization of the Beta Distribution with Applications," Journal of Econometrics, 66(1995), 133-152.

Patil, G. P., M. T. Boswell and M. V. Ratnaparkhi. Dictionary and Classified Biography of Statistical Distributions in Scientific Work: Continuous Univariate Models, Bartonsville: International Cooperative Clearing House, 1984.

Rao, C. R. (1973). Linear Statistical Inference and Its Applications, New York: Wiley, 2nd edition.

Theil, H. (1971). Principles of Econometrics, New York: Wiley.

Tukey, J. W. (1977). Exploratory Data Analysis, Reading. MA: Addison-Wesley.

White, H. "A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroskedasticity," Econometrica, 48(1980), 817-838.

White, H. "Maximum Likelihood Estimation of Misspecified Models," Econometrica, 50(1982), 1-25.

Zellner, A. (1962). "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association, 57, 348-368.

# APPENDIX A

Proofs of Some Theorems Associated with the Distribution of Quadratic Forms of Normally Distributed Variables

1. If $\underset{n\times 1}{Y} \sim N(\mu,\Sigma)$ where $\Sigma$ is positive definite, then

$$Q = (Y-\mu)'\Sigma^{-1}(Y-\mu) \sim \chi^2(n).$$

Proof:

Since $\Sigma$ is positive definite, it can be factored

$$\Sigma = \Sigma^{1/2}(\Sigma^{1/2})' \text{ (e.g., Cholesky Decomposition).}$$

This implies that

$$\Sigma^{-1} = ((\Sigma^{1/2})')^{-1}(\Sigma^{1/2})^{-1}.$$

Making this substitution into the quadratic form yields

$$\begin{aligned} Q &= (Y-\mu)'\Sigma^{-1}(Y-\mu) = (Y-\mu)'(\Sigma^{-1/2})'\Sigma^{1/2}(Y-\mu) \\ &= [\Sigma^{-1/2}(Y-\mu)]'[\Sigma^{-1/2}(Y-\mu)] \\ &= Z'Z \qquad \text{where } Z = \Sigma^{-1/2}(Y-\mu) \sim N[0,I_n] \end{aligned}$$

because

$$E(Z) = \Sigma^{-1/2}(E(Y)-\mu) = 0$$

$$Var(Z) = \Sigma^{-1/2}\Sigma(\Sigma^{-1/2})' = I;$$

therefore

$$Q \sim \chi^2(n).$$

2. If $Y \sim N[0,I]$ and A is a real symmetric idempotent matrix, then $Q = Y'AY \sim \chi^2(\text{trace}(A))$.

Proof:

The characteristic roots of an idempotent matrix are zero or one and the trace of the idempotent matrix will equal the number of unitary characteristic roots.

Let C denote the orthogonal matrix of characteristic vectors associated with A. In the previous section on matrix results it was shown that

$$C'AC = \Lambda$$

where $\Lambda$ is a diagonal matrix with the characteristic roots on the diagonal.

The quadratic form Q can be rewritten as

$$\begin{aligned} Q &= Y'AY = Y'CC'ACC'Y \\ &= (C'Y)'\,\Lambda(C'Y) \quad \text{(C is orthogonal)} \end{aligned}$$

$$= Z' \Lambda Z \text{ where } Z = C'Y \sim N[C'0, C'IC] = N[0, I]$$

Therefore,

$$Q = Z' \Lambda Z \sim \chi^2(\text{Rank}(A) = \text{trace}(A)).$$

This follows from Z being a vector of "standard normal" variables and $\Lambda$ having 0's and one's on the main diagonal. The number of ones on the diagonal of $\Lambda$ is equal to the trace of A.

3. Let $Y_{nx1} \sim N(0,I)$ and let A denote an nxn real symmetric idempotent matrix. The vector of random variables LY and the quadratic form Q = Y'AY will be independent if LA = 0.

<u>Proof</u>: The symmetry and idempotent property of A permits rewriting Q as

$$Q = Y'AY = Y'A'AY = (AY)'(AY)$$

The covariance between LY and AY is given by

$$\begin{aligned} COV(LY, AY) &= E(LY-E(LY))(AY - E(AY))' \\ &= E(LY)(AY)' = E(LYY'A') \\ &= L(E(YY'))A' \\ &= LIA' \qquad \text{Because } Y \sim N[0,I] \\ &= LA \qquad \text{Because A is symmetric} \end{aligned}$$

It follows that LA = 0 implies that LY and AY are stochastically independent; hence Q and LY will be stochastically independent.

# APPENDIX B

Cramer-Rao Inequality--Information Matrix
Theil, pp. 384-86

1. The equivalence of two important matrices

Let $L(Y_1,...,Y_n, \theta) = \prod_t f(Y_t;\theta) = L(Y;\theta)$ and let

$\ell = \ln(L) = \sum_t \ln f(Y_t;\theta) = \ell(Y;\theta)$

Integrating the likelihood function over all values of the Y's is unity,

i.e.,

$$(B.1) \qquad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} L(Y_1,...,Y_n;\theta)dY_1 \ldots dY_n = \int L(Y;\theta)dY = 1.$$

Differentiating both sides of (B.1) with respect to θ yields

$$(B.2) \qquad \int \frac{dL(Y;\theta)}{d\theta}\, dY = 0,$$

but this is the same as

$$(B.2)' \qquad \int \frac{\frac{dL(Y;\theta)}{d\theta}}{L(Y;\theta)} L(Y;\theta)dY = \int \frac{d\ell(Y;\theta)}{d\theta} L(Y;\theta)dY = 0$$

or

$$\boxed{E\left(\frac{d\ell(Y;\theta)}{d\theta}\right) = 0}$$

Differentiating the second expression in (B.2)' with respect to θ yields

$$(B.3) \qquad \int \frac{d^2\ell}{d\theta^2} L(Y;\theta) + \frac{d\ell}{d\theta}\frac{dL(Y;\theta)}{d\theta}\, dY$$

$$= \int \left\{\frac{d^2\ell(Y;\theta)}{d\theta^2} + \left(\frac{d\ell}{d\theta}\right)\left(\frac{d\ell}{d\theta}\right)\right\} L(Y;\theta)dY = 0.$$

This implies that

$$(B.4) \qquad \boxed{E\left\{\left(\frac{d\ell}{d\theta}\right)\left(\frac{d\ell}{d\theta}\right)\right\} = -E\left\{\frac{d^2\ell(Y;\theta)}{d\theta^2}\right\} = \mathrm{var}\left(\frac{d\ell}{d\theta}\right)}$$

This is a very important theoretical and practical result which can be used in estimating variances of MLE of θ as well as in testing for correct specifications.

2. <u>The asymptotic distribution of the MLE of θ</u>

Let $\theta = \theta(Y)$ denote an estimator of θ. The expected value of θ(Y) is given by

(B.5) $E(\theta(Y)) = \int\theta(Y)L(Y;\theta)dY.$

Differentiating (B.5) with respect to θ yields

$$\text{(B.6)} \quad \frac{\partial E(\theta(Y))}{\partial\theta} = \int\theta(Y)\frac{dL(Y;\theta)}{d\theta}dY = \int\theta(Y)\frac{d\ell(Y;\theta)}{d\theta}L(Y;\theta)dY$$

$$= \text{cov}(\theta(Y), \partial\ell(Y;\theta)/\partial\theta).$$

It will be useful to make a couple of observations at this point. <u>First</u>, for unbiased estimators $E(\theta(Y)) = \theta$ and $\frac{dE(\theta(Y))}{d\theta} = \frac{d\theta}{d\theta} = 1$ and then equation (B.6) implies $\text{Cov}(\theta(Y), d\ell/d\theta) = 1$. <u>Second</u>, recal correlation$^2(Z_1, Z_2) \le 1$ implies that

$$\text{COVAR}^2(Z_1, Z_2) \le \text{Var}(Z_1)\text{Var}(Z_2).$$

Combining these results yields

$$\text{(B.7)} \quad 1 = \left(\frac{\partial E(\theta(Y))}{\partial\theta}\right)^2 = \text{Cov}^2(\;) \le \text{Var}(\theta(Y))\,\text{Var}\left(\frac{\partial\ell(Y;\theta)}{\partial\theta}\right);$$

i.e.,

$$1 \le \text{Var}(\theta(Y))\left\{-E\left(\frac{\partial^2\ell}{\partial\theta^2}\right)\right\}$$

Therefore

$$\text{(B.8)} \quad \boxed{\begin{aligned} \text{var}(\theta(Y)) &\ge \left\{-E\left(\frac{\partial^2\ell}{\partial\theta^2}\right)\right\}^{-1} \\ &\ge \left\{E\left(\frac{d\ell}{d\theta}\frac{d\ell}{d\theta}\right)\right\}^{-1} \end{aligned}}$$

which is the Cramer-Rao lower bound for unbiased estimators.

## 5. Problem sets

Consider the exponential distribution defined by the probability density function

$$f(X;\beta) = \frac{e^{-X/\beta}}{\beta} \qquad \text{for } X \geq 0$$

$= 0$ otherwise.

1. Show that the cumulative distribution function for the exponential is given by

   $F(X;\theta) = 1 - e^{-X/\beta}$

2. Derive an expression for the corresponding hazard function.

3. Show that the moment generating function for the exponential distribution is given by $\left(\frac{1}{1 - \beta t}\right)$.

4. Write an expression for the cumulant generating function for the exponential distribution.

5. Using the cumulant generating function for the exponential evaluate

   $E(X) = \mu =$

   $Var(X) = \sigma^2 =$

   $E(X - \mu)^3 =$

   $E(X - \mu)^4 =$

Note that these results imply that

$$\text{Skewness (normalized)} \quad = \quad \frac{E(X - \mu)^3}{\sigma^3} = 2 > 0$$

$$\text{Kurtosis (normalized)} \quad = \quad \frac{E(X - \mu)^4}{\sigma^4} = 9$$

6. Show that the mode and median for the exponential distribution are given by 0 and (β ln 2), respectively.

   Hint: $f(\text{mode}) \geq f(x)$ $\qquad\qquad$ $\int_{-\infty}^{median} f(s)ds = .5$ .

7. Demonstrate that the MLE of β in the exponential distribution is given by $\tilde{\beta} = \bar{X} = \sum_{i=1}^{N}\left(\frac{X_i}{N}\right)$ .

8. The mgf for the gamma probability density,

$$GA(x;p,b) = \frac{X^{p-1}e^{-X/\beta}}{\beta^p\Gamma(p)},$$

   is given by

$$\frac{1}{(1-\beta t)^p}$$

   Show that the sum of n independent and identically distributed exponential variables is distributed as GA( ; β, p=n).

9. Using your results in question (8), demonstrate that the exact distribution of the sample mean ($\bar{Y}$) is given by:

   $\bar{X}$ ~ GA( ; β/n, p=n).

10. Consider a random sample

$\{y_1, y_2, . . ., y_n\}$ where the density of $y_i$ is given by

$GG(y_i;a,\beta,p)$.

(a) Form the log likelihood function.

(b) Obtain the MLE of β for the case where a = p = 1, i.e., the exponential. (same as # 7)

(c) Evaluate the second derivative of the log likelihood function with respect to β for the exponential?

(d) What is the asymptotic distribution of the MLE of β for the exponential?

11. The exact distribution of the sample mean of exponentially distributed variables is a gamma , GA(z; β/n, p=n), (see problem 9) and the approximating "asymptotic" normal distribution is $N(z; \beta, \beta^2/n)$. The corresponding moment generating functions are given by:

Gamma: $$M_{GA}(t) = \frac{1}{(1-\beta t/n)^n}$$

Normal: $$M_N(t) = e^{\beta t+(t\beta)^2/2n}$$

Compare the mean, variance, skewness, and kurtosis of the *exact* distribution and the approximating or asymptotic normal distribution.

| | GA( ; $\beta / n$ , p=n) | $N\left(\mu=\beta,\ \sigma^2=\beta^2/n\right)$ |
|---|---|---|
| Mean | | |
| Variance | | |
| Skewness* | | |
| Kurtosis* | | |

Hint: Use the cumulant generating functions. * Note you are only asked to "normalize" the skewness and kurtosis coefficients. This is done by dividing the skewness and kurtosis by $\sigma^3$ and $\sigma^4$ , respectively.

12. The Burr distribution and hazard functions

    a. Demonstrate that

$$GB2(y;a,b,p = 1,q) = \frac{|a|q(y/b)^{a-1}}{b(1 + (y/b)^{a})^{q+1}}$$

Hint:

$$B(1,q) = \frac{\Gamma(1)\Gamma(q)}{\Gamma(q + 1)} = \frac{\Gamma(1)\Gamma(q)}{q\Gamma(q)} = \left\{\frac{1}{q}\right\}$$

$$F_{Burr12}(y;a,b,p) = 1 - \left(\frac{1}{1+(y/b)^{a}}\right)^{q}$$

    b. Use these results to obtain an expression for the hazard function for the BR12. Recall that the hazard function is defined to be pdf/(1-cdf)

    c. The hazard function for the Burr 12 with a >1 is upside down “U” shaped. Investigate the shape of the hazard function for the two cases (1) $0< a < 1$ and (2) $a = 1$. Assume that all parameter values are positive. Hint: What happens to thevalue of the hazard function as y increases for arbitrarry but fixed values of a?

13. What are the possible shapes of the hazard function for the gamma, Weibull and exponential and distributions? (Hint: This problem is not requesting a derivation. Refer to figure 4 for the generalized gamma function).

14. If $Y \sim LN(\mu, \sigma^2)$, verify that $Z = Y^2 \sim LN(2\mu, 4\sigma^2)$. Don't forget the Jacobian in your analysis of the transformation.

15. If $X \sim GA(x; \beta = 1, p)$, verify that $Y = \beta X^{1/a} \sim GG(y ; a, \beta, p)$. Don't forget the Jacobian in your analysis of the transformation.

16. Use the cumulant generating function for the normal, $N(\mu, \sigma^2)$, to show that the normalized or scalled kurtosis for a normally distributed random variable is given by 3.

17. Given that the n x 1 vector Y is distributed as $N(\mu,\Sigma)$ where $\Sigma$ is positive definite, demonstrate that the mode of the multivariate density $N(\mu,\Sigma)$ occurs at $Y = \mu$, i.e.,
demonstrate $\frac{df(Y)}{dY} = 0$ at $Y = \mu$ and note that $\frac{d^2 f(Y)}{dY^2}$ is negative definite.

of the expression of the multivariate normal density function.

18. Given $g(X) = 2x_1^2 + (5/2)x_2^2 + 4x_1x_2 + x_1 + 3x_2$:

a. Demonstrate that

$$g(X) = (x_1, x_2)\begin{pmatrix} 2 & 2 \\ 2 & 5/2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (x_1, x_2)\begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

b. Using the technique of differentiating g(X) with respect to the vector $X = (x_1, x_2)'$, determine the vector X which is associated with an optimum of g(X).

c. Evaluate $\dfrac{d^2 g(X)}{dX^2}$ and determine whether your answer to (b) corresponds to a maximum or minimum.

19. Let $\mathbf{i} = (1, 1, ..., 1)'$

a. Evaluate $\mathbf{i'i}$

b. Evaluate $\mathbf{ii'}$

c. Demonstrate that the matrix

$$\left(\frac{1}{n}\mathbf{ii'}\right)$$

is symmetric and idempotent.

20. Let $y_t$ be independently distributed as $N(O,\sigma^2)$, i.e.,

$Y = (y_1, y_2, \ldots, y_n)' \sim N[O,\sigma^2 I_n]$.

a. Determine the distribution of

$$\overline{Y} = \sum_{t=1}^{n} y_t/n.$$

Hint: $$\overline{Y} = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \left(\frac{1}{n}\right)\mathbf{i'}y$$

$= BY$ where $B = \left(\frac{1}{n}\right)\mathbf{i'}.$

Consider the first "useful" theorem in the section on multivariate statistics.

Now consider $s^2 = (\sum(y_i - \overline{Y})^2/(n - 1))$

$$= (y_1, \ldots, y_n)\left[\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & & \vdots \\ 0 & \cdots & & 1 \end{pmatrix} - \frac{1}{n}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ & & & \\ \vdots & \vdots & & \vdots \\ 1 & 1 & & 1 \end{pmatrix}\right]\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}/n-1$$

$$= Y'(I - \frac{1}{n}\mathbf{ii}')Y/(n-1)$$

$$= Y'AY/(n-1)$$

where $A = I - \frac{1}{n}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix} = I - \frac{1}{n}\mathbf{ii}'$

b. Demonstrate that A is idempotent.

c. Verify that trace (A) = n - 1.

d. Demonstrate that the distribution of $s^2(n - 1)/\sigma^2$ is $\chi^2(n - 1)$.

e. Verify that $\bar{Y}$ and $s^2$ are independently distributed.
Hint: Is BA = 0?

f. What is the distribution of $\left(\dfrac{\dfrac{\bar{Y}-0}{\sigma/n}}{\left(\dfrac{s^2(n-1)}{\sigma^2}/(n-1)\right)^{1/2}}\right) = \dfrac{\bar{Y}-0}{s/n}$

21. Method of Moments

The mean and variance of a Gamma are given by $\beta p$ $\beta^2 p$ , respectively. Let the corresponding sample moments be denoted by $\bar{Y}$ and $s^2$.

a. Derive the method of moments estimators of the parameters $\beta$ ε $p$ .

b. Derive the method of moments estimator of the parameter $\beta$ in an exponential distribution.

c. Using two sample moments discuss how you could use generalized method of moments estimation to estimate the parameter $\beta$ in an exponential distribution.

22. Optional problem. Verify

a. $EGB(\varepsilon;\delta,\sigma,c,p,q) = GB(e^{\varepsilon};\ a=1/\sigma,\ b = e^{\delta},\ c,\ p,\ q)e^{\varepsilon}$

Why is the expression for GB( ) multiplied by $e^{\varepsilon}$?

b. $EGB2(\varepsilon;\ \delta,\sigma,p,q) = (EGB(\varepsilon;\delta,\sigma,\ c=1,p,q)$

$$= \frac{e^{p(\varepsilon-\delta)/\sigma}}{\sigma B(p,q)(1+e^{(\varepsilon-\delta)/\sigma})^{p+q}}$$

for $-\infty < \varepsilon < \infty$

c. $$M_{EGB2}(t) = \frac{e^{\delta t}\Gamma(p + t\sigma)\Gamma(q - t\sigma)}{\Gamma(p)\Gamma(q)}$$

Hint: Consider the moments for the GB2 reported in Table 1.

d. Obtain the cumulant generating function for the EGB2 and show that the mean of the EGB2 is zero if $\delta = \sigma[\psi(q) - \psi(p)]$ where $\psi(x) = d\ \ln\Gamma(x)/dx$.