Lecture Notes - Econometrics: Some Statistics

Paul Söderlind[1]

8 July 2013

[1] University of St. Gallen. *Address:* s/bf-HSG, Rosenbergstrasse 52, CH-9000 St. Gallen, Switzerland. *E-mail:* Paul.Soderlind@unisg.ch. Document name: EcmXSta.TeX.

Contents

Chapter 21

Some Statistics

This section summarizes some useful facts about statistics. Heuristic proofs are given in a few cases.

Some references: Mittelhammer (1996), DeGroot (1986), Greene (2000), Davidson (2000), Johnson, Kotz, and Balakrishnan (1994).

21.1 Distributions and Moment Generating Functions

Most of the stochastic variables we encounter in econometrics are continuous. For a continuous random variable X, the range is uncountably infinite and the probability that $X \leq x$ is $\Pr(X \leq x) = \int_{-\infty}^{x} f(q)dq$ where $f(q)$ is the continuous probability density function of X. Note that X is a random variable, x is a number (1.23 or so), and q is just a dummy argument in the integral.

Fact 21.1 *(cdf and pdf) The cumulative distribution function of the random variable X is $F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(q)dq$. Clearly, $f(x) = dF(x)/dx$. Note that x is just a number, not random variable.*

Fact 21.2 *(Moment generating function of X) The moment generating function of the random variable X is $mgf(t) = \mathrm{E}\, e^{tX}$. The rth moment is the rth derivative of $mgf(t)$ evaluated at $t = 0$: $\mathrm{E}\, X^r = d\,mgf(0)/dt^r$. If a moment generating function exists (that is, $\mathrm{E}\, e^{tX} < \infty$ for some small interval $t \in (-h, h)$), then it is unique.*

Fact 21.3 *(Moment generating function of a function of X) If X has the moment generating function $mgf_X(t) = \mathrm{E}\, e^{tX}$, then $g(X)$ has the moment generating function $\mathrm{E}\, e^{tg(X)}$.*

The affine function $a + bX$ (a and b are constants) has the moment generating function $mgf_{g(X)}(t) = \mathrm{E}\, e^{t(a+bX)} = e^{ta}\, \mathrm{E}\, e^{tbX} = e^{ta} mgf_X(bt)$. By setting $b = 1$ and $a = -\mathrm{E}\, X$ we obtain a mgf for central moments (variance, skewness, kurtosis, etc), $mgf_{(X-\mathrm{E}\, X)}(t) = e^{-t\, \mathrm{E}\, X} mgf_X(t)$.

Example 21.4 *When $X \sim N(\mu, \sigma^2)$, then $mgf_X(t) = \exp\left(\mu t + \sigma^2 t^2/2\right)$. Let $Z = (X-\mu)/\sigma$ so $a = -\mu/\sigma$ and $b = 1/\sigma$. This gives $mgf_Z(t) = \exp(-\mu t/\sigma) mgf_X(t/\sigma) = \exp\left(t^2/2\right)$. (Of course, this result can also be obtained by directly setting $\mu = 0$ and $\sigma = 1$ in mgf_X.)*

Fact 21.5 *(Characteristic function and the pdf) The characteristic function of a random variable x is*

$$\begin{aligned} g(\phi) &= \mathrm{E} \exp(i\phi x) \\ &= \int_x \exp(i\phi x) f(x) dx, \end{aligned}$$

where $f(x)$ is the pdf. This is a Fourier transform of the pdf (if x is a continuous random variable). The pdf can therefore be recovered by the inverse Fourier transform as

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\phi x) g(\phi) d\phi.$$

In practice, we typically use a fast (discrete) Fourier transform to perform this calculation, since there are very quick computer algorithms for doing that.

Fact 21.6 *The charcteristic function of a $N(\mu, \sigma^2)$ distribution is $\exp(i\phi\mu - \phi^2\sigma^2/2)$ and of a lognormal(μ, σ^2) distribuion (where $\ln x \sim N(\mu, \sigma^2)$) $\sum_{j=0}^{\infty} \frac{(i\phi)^j}{j!} \exp(j\mu + j^2\sigma^2/2)$.*

Fact 21.7 *(Change of variable, univariate case, monotonic function) Suppose X has the probability density function $f_X(c)$ and cumulative distribution function $F_X(c)$. Let $Y = g(X)$ be a continuously differentiable function with $dg/dX > 0$ (so $g(X)$ is increasing for all c such that $f_X(c) > 0$. Then the cdf of Y is*

$$F_Y(c) = \Pr[Y \leq c] = \Pr[g(X) \leq c] = \Pr[X \leq g^{-1}(c)] = F_X[g^{-1}(c)],$$

where g^{-1} is the inverse function of g such that $g^{-1}(Y) = X$. We also have that the pdf of Y is

$$f_Y(c) = f_X[g^{-1}(c)] \left| \frac{dg^{-1}(c)}{dc} \right|.$$

If, instead, $dg/dX < 0$ (so $g(X)$ is decreasing), then we instead have the cdf of Y

$$F_Y(c) = \Pr[Y \leq c] = \Pr[g(X) \leq c] = \Pr[X \geq g^{-1}(c)] = 1 - F_X[g^{-1}(c)],$$

but the same expression for the pdf.

Proof. Differentiate $F_Y(c)$, that is, $F_X[g^{-1}(c)]$ with respect to c. ■

Example 21.8 *Let $X \sim U(0, 1)$ and $Y = g(X) = F^{-1}(X)$ where $F(c)$ is a strictly increasing cdf. We then get*

$$f_Y(c) = \frac{dF(c)}{dc}.$$

The variable Y then has the pdf $dF(c)/dc$ and the cdf $F(c)$. This shows how to generate random numbers from the $F()$ distribution: draw $X \sim U(0, 1)$ and calculate $Y = F^{-1}(X)$.

Example 21.9 *Let $Y = \exp(X)$, so the inverse function is $X = \ln Y$ with derivative $1/Y$. Then, $f_Y(c) = f_X(\ln c)/c$. Conversely, let $Y = \ln X$, so the inverse function is $X = \exp(Y)$ with derivative $\exp(Y)$. Then, $f_Y(c) = f_X[\exp(c)]\exp(c)$.*

Example 21.10 *Let $X \sim U(0, 2)$, so the pdf and cdf of X are then $1/2$ and $c/2$ respectively. Now, let $Y = g(X) = -X$ gives the pdf and cdf as $1/2$ and $1 + y/2$ respectively. The latter is clearly the same as $1 - F_X[g^{-1}(c)] = 1 - (-c/2)$.*

Fact 21.11 *(Distribution of truncated a random variable) Let the probability distribution and density functions of X be $F(x)$ and $f(x)$, respectively. The corresponding functions, conditional on $a < X \leq b$ are $[F(x) - F(a)]/[F(b) - F(a)]$ and $f(x)/[F(b) - F(a)]$. Clearly, outside $a < X \leq b$ the pdf is zero, while the cdf is zero below a and unity above b.*

21.2 Joint and Conditional Distributions and Moments

21.2.1 Joint and Conditional Distributions

Fact 21.12 *(Joint and marginal cdf) Let X and Y be (possibly vectors of) random variables and let x and y be two numbers. The joint cumulative distribution function of X and Y is* $H(x, y) = \Pr(X \le x, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} h(q_x, q_y) dq_y dq_x$*, where* $h(x, y) = \partial^2 F(x, y)/\partial x \partial y$ *is the joint probability density function.*

Fact 21.13 *(Joint and marginal pdf) The marginal cdf of X is obtained by integrating out Y:* $F(x) = \Pr(X \le x, Y \text{ anything}) = \int_{-\infty}^{x}\left[\int_{-\infty}^{\infty} h(q_x, q_y) dq_y\right] dq_x$*. This shows that the marginal pdf of x is* $f(x) = dF(x)/dx = \int_{-\infty}^{\infty} h(q_x, q_y) dq_y$*.*

Fact 21.14 *(Conditional distribution) The pdf of Y conditional on $X = x$ (a number) is* $g(y|x) = h(x, y)/f(x)$*. This is clearly proportional to the joint pdf (at the given value x).*

Fact 21.15 *(Change of variable, multivariate case, monotonic function) The result in Fact 21.7 still holds if X and Y are both $n \times 1$ vectors, but the derivative are now* $\partial g^{-1}(c)/\partial dc'$ *which is an $n \times n$ matrix. If g_i^{-1} is the ith function in the vector g^{-1} then*

$$\frac{\partial g^{-1}(c)}{\partial dc'} = \begin{bmatrix} \frac{\partial g_1^{-1}(c)}{\partial c_1} & \cdots & \frac{\partial g_1^{-1}(c)}{\partial c_n} \\ \vdots & & \vdots \\ \frac{\partial g_n^{-1}(c)}{\partial c_1} & \cdots & \frac{\partial g_n^{-1}(c)}{\partial c_m} \end{bmatrix}.$$

21.2.2 Moments of Joint Distributions

Fact 21.16 *(Caucy-Schwartz)* $(\mathrm{E}\,XY)^2 \le \mathrm{E}(X^2)\,\mathrm{E}(Y^2)$.

Proof. $0 \le \mathrm{E}[(aX+Y)^2] = a^2\,\mathrm{E}(X^2) + 2a\,\mathrm{E}(XY) + \mathrm{E}(Y^2)$. Set $a = -\mathrm{E}(XY)/\mathrm{E}(X^2)$ to get

$$0 \le -\frac{[\mathrm{E}(XY)]^2}{\mathrm{E}(X^2)} + \mathrm{E}(Y^2), \text{ that is, } \frac{[\mathrm{E}(XY)]^2}{\mathrm{E}(X^2)} \le \mathrm{E}(Y^2).$$

■

Fact 21.17 *($-1 \le Corr(X, y) \le 1$). Let Y and X in Fact 21.16 be zero mean variables (or variables minus their means). We then get* $[\mathrm{Cov}(X, Y)]^2 \le \mathrm{Var}(X)\,\mathrm{Var}(Y)$*, that is,* $-1 \le \mathrm{Cov}(X, Y)/[Std(X)Std(Y)] \le 1$.

21.2.3 Conditional Moments

Fact 21.18 *(Conditional moments)* $\mathrm{E}(Y|x) = \int y g(y|x) dy$ *and* $\mathrm{Var}(Y|x) = \int [y - \mathrm{E}(Y|x)] g(y|x) dy$.

Fact 21.19 *(Conditional moments as random variables) Before we observe X, the conditional moments are random variables—since X is. We denote these random variables by* $\mathrm{E}(Y|X)$, $\mathrm{Var}(Y|X)$, *etc.*

Fact 21.20 *(Law of iterated expectations)* $\mathrm{E}\,Y = \mathrm{E}[\mathrm{E}(Y|X)]$. *Note that* $\mathrm{E}(Y|X)$ *is a random variable since it is a function of the random variable X. It is not a function of Y, however. The outer expectation is therefore an expectation with respect to X only.*

Proof. $\mathrm{E}[\mathrm{E}(Y|X)] = \int \left[\int y g(y|x) dy\right] f(x) dx = \int\int y g(y|x) f(x) dy dx = \int\int y h(y,x) dy dx = \mathrm{E}\,Y$. ■

Fact 21.21 *(Conditional vs. unconditional variance)* $\mathrm{Var}(Y) = \mathrm{Var}[\mathrm{E}(Y|X)] + \mathrm{E}[\mathrm{Var}(Y|X)]$.

Fact 21.22 *(Properties of Conditional Expectations) (a)* $Y = \mathrm{E}(Y|X) + U$ *where U and* $\mathrm{E}(Y|X)$ *are uncorrelated:* $\mathrm{Cov}(X,Y) = \mathrm{Cov}[X, \mathrm{E}(Y|X) + U] = \mathrm{Cov}[X, \mathrm{E}(Y|X)]$. *It follows that (b)* $\mathrm{Cov}[Y, \mathrm{E}(Y|X)] = \mathrm{Var}[\mathrm{E}(Y|X)]$; *and (c)* $\mathrm{Var}(Y) = \mathrm{Var}[\mathrm{E}(Y|X)] + \mathrm{Var}(U)$. *Property (c) is the same as Fact 21.21, where* $\mathrm{Var}(U) = \mathrm{E}[\mathrm{Var}(Y|X)]$.

Proof. $\mathrm{Cov}(X,Y) = \int\int x(y - \mathrm{E}\,y) h(x,y) dy dx = \int x \left[\int (y - \mathrm{E}\,y) g(y|x) dy\right] f(x) dx$, but the term in brackets is $\mathrm{E}(Y|X) - \mathrm{E}\,Y$. ■

Fact 21.23 *(Conditional expectation and unconditional orthogonality)* $\mathrm{E}(Y|Z) = 0 \Rightarrow \mathrm{E}\,YZ = 0$.

Proof. Note from Fact 21.22 that $\mathrm{E}(Y|X) = 0$ implies $\mathrm{Cov}(X,Y) = 0$ so $\mathrm{E}\,XY = \mathrm{E}\,X\,\mathrm{E}\,Y$ (recall that $\mathrm{Cov}(X,Y) = \mathrm{E}\,XY - \mathrm{E}\,X\,\mathrm{E}\,Y$). Note also that $\mathrm{E}(Y|X) = 0$ implies that $\mathrm{E}\,Y = 0$ (by iterated expectations). We therefore get

$$\mathrm{E}(Y|X) = 0 \Rightarrow \begin{bmatrix} \mathrm{Cov}(X,Y) = 0 \\ \mathrm{E}\,Y = 0 \end{bmatrix} \Rightarrow \mathrm{E}\,YX = 0.$$

■

21.2.4 Regression Function and Linear Projection

Fact 21.24 *(Regression function) Suppose we use information in some variables X to predict Y. The choice of the forecasting function $\hat{Y} = k(X) = \mathrm{E}(Y|X)$ minimizes $\mathrm{E}[Y - k(X)]^2$. The conditional expectation $\mathrm{E}(Y|X)$ is also called the regression function of Y on X. See Facts 21.22 and 21.23 for some properties of conditional expectations.*

Fact 21.25 *(Linear projection) Suppose we want to forecast the scalar Y using the $k \times 1$ vector X and that we restrict the forecasting rule to be linear $\hat{Y} = X'\beta$. This rule is a linear projection, denoted $P(Y|X)$, if β satisfies the orthogonality conditions $\mathrm{E}[X(Y - X'\beta)] = \mathbf{0}_{k\times 1}$, that is, if $\beta = (\mathrm{E}\, XX')^{-1}\,\mathrm{E}\, XY$. A linear projection minimizes $\mathrm{E}[Y - k(X)]^2$ within the class of linear $k(X)$ functions.*

Fact 21.26 *(Properties of linear projections) (a) The orthogonality conditions in Fact 21.25 mean that*

$$Y = X'\beta + \varepsilon,$$

where $\mathrm{E}(X\varepsilon) = \mathbf{0}_{k\times 1}$. This implies that $\mathrm{E}[P(Y|X)\varepsilon] = 0$, so the forecast and forecast error are orthogonal. (b) The orthogonality conditions also imply that $\mathrm{E}[XY] = \mathrm{E}[XP(Y|X)]$. (c) When X contains a constant, so $\mathrm{E}\,\varepsilon = 0$, then (a) and (b) carry over to covariances: $\mathrm{Cov}[P(Y|X), \varepsilon] = 0$ and $\mathrm{Cov}[X, Y] = \mathrm{Cov}[XP, (Y|X)]$.

Example 21.27 *($P(1|X)$) When $Y_t = 1$, then $\beta = (\mathrm{E}\, XX')^{-1}\,\mathrm{E}\, X$. For instance, suppose $X = [x_{1t}, x_{t2}]'$. Then*

$$\beta = \begin{bmatrix} \mathrm{E}\, x_{1t}^2 & \mathrm{E}\, x_{1t}x_{2t} \\ \mathrm{E}\, x_{2t}x_{1t} & \mathrm{E}\, x_{2t}^2 \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{E}\, x_{1t} \\ \mathrm{E}\, x_{2t} \end{bmatrix}.$$

If $x_{1t} = 1$ in all periods, then this simplifies to $\beta = [1, 0]'$.

Remark 21.28 *Some authors prefer to take the transpose of the forecasting rule, that is, to use $\hat{Y} = \beta'X$. Clearly, since XX' is symmetric, we get $\beta' = \mathrm{E}(YX')(\mathrm{E}\, XX')^{-1}$.*

Fact 21.29 *(Linear projection with a constant in X) If X contains a constant, then $P(aY + b|X) = aP(Y|X) + b$.*

Fact 21.30 *(Linear projection versus regression function) Both the linear regression and the regression function (see Fact 21.24) minimize $\mathrm{E}[Y - k(X)]^2$, but the linear projection* *imposes the restriction that $k(X)$ is linear, whereas the regression function does not impose any restrictions. In the special case when Y and X have a joint normal distribution, then the linear projection is the regression function.*

Fact 21.31 *(Linear projection and OLS) The linear projection is about population moments, but OLS is its sample analogue.*

21.3 Convergence in Probability, Mean Square, and Distribution

Fact 21.32 *(Convergence in probability) The sequence of random variables $\{X_T\}$ converges in probability to the random variable X if (and only if) for all $\varepsilon > 0$*

$$\lim_{T\to\infty} \Pr(|X_T - X| < \varepsilon) = 1.$$

We denote this $X_T \xrightarrow{p} X$ or $\operatorname{plim} X_T = X$ (X is the probability limit of X_T). Note: (a) X can be a constant instead of a random variable; (b) if X_T and X are matrices, then $X_T \xrightarrow{p} X$ if the previous condition holds for every element in the matrices.

Example 21.33 *Suppose $X_T = 0$ with probability $(T-1)/T$ and $X_T = T$ with probability $1/T$. Note that $\lim_{T\to\infty} \Pr(|X_T - 0| = 0) = \lim_{T\to\infty}(T-1)/T = 1$, so $\lim_{T\to\infty} \Pr(|X_T - 0| = \varepsilon) = 1$ for any $\varepsilon > 0$. Note also that $\mathrm{E}\, X_T = 0 \times (T-1)/T + T \times 1/T = 1$, so X_T is biased.*

Fact 21.34 *(Convergence in mean square) The sequence of random variables $\{X_T\}$ converges in mean square to the random variable X if (and only if)*

$$\lim_{T\to\infty} \mathrm{E}(X_T - X)^2 = 0.$$

We denote this $X_T \xrightarrow{m} X$. Note: (a) X can be a constant instead of a random variable; (b) if X_T and X are matrices, then $X_T \xrightarrow{m} X$ if the previous condition holds for every element in the matrices.

Fact 21.35 *(Convergence in mean square to a constant) If X in Fact 21.34 is a constant, then then $X_T \xrightarrow{m} X$ if (and only if)*

$$\lim_{T\to\infty} (\mathrm{E}\, X_T - X)^2 = 0 \text{ and } \lim_{T\to\infty} \operatorname{Var}(X_T{}^2) = 0.$$

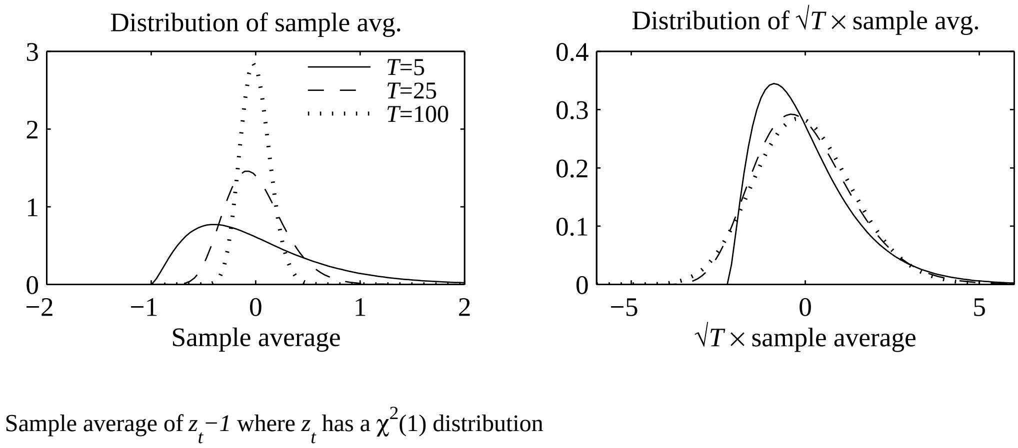

Figure 21.1: Sampling distributions

This means that both the variance and the squared bias go to zero as $T \to \infty$.

Proof. $\mathrm{E}(X_T - X)^2 = \mathrm{E}\, X_T^2 - 2X\, \mathrm{E}\, X_T + X^2$. Add and subtract $(\mathrm{E}\, X_T)^2$ and recall that $\mathrm{Var}(X_T) = \mathrm{E}\, X_T^2 - (\mathrm{E}\, X_T)^2$. This gives $\mathrm{E}(X_T - X)^2 = \mathrm{Var}(X_T) - 2X\, \mathrm{E}\, X_T + X^2 + (\mathrm{E}\, X_T)^2 = \mathrm{Var}(X_T) + (\mathrm{E}\, X_T - X)^2$. ■

Fact 21.36 *(Convergence in distribution) Consider the sequence of random variables $\{X_T\}$ with the associated sequence of cumulative distribution functions $\{F_T\}$. If* $\lim_{T\to\infty} F_T = F$ *(at all points), then F is the limiting cdf of X_T. If there is a random variable X with cdf F, then X_T converges in distribution to X: $X_T \xrightarrow{d} X$. Instead of comparing cdfs, the comparison can equally well be made in terms of the probability density functions or the moment generating functions.*

Fact 21.37 *(Relation between the different types of convergence) We have $X_T \xrightarrow{m} X \Rightarrow X_T \xrightarrow{p} X \Rightarrow X_T \xrightarrow{d} X$. The reverse implications are not generally true.*

Example 21.38 *Consider the random variable in Example 21.33. The expected value is* $\mathrm{E}\, X_T = 0(T-1)/T + T/T = 1$. *This means that the squared bias does not go to zero, so X_T does not converge in mean square to zero.*

Fact 21.39 *(Slutsky's theorem) If $\{X_T\}$ is a sequence of random matrices such that* $\operatorname{plim} X_T = X$ *and $g(X_T)$ a continuous function, then* $\operatorname{plim} g(X_T) = g(X)$.

Fact 21.40 *(Continuous mapping theorem) Let the sequences of random matrices* $\{X_T\}$ *and* $\{Y_T\}$*, and the non-random matrix* $\{a_T\}$ *be such that* $X_T \overset{d}{\to} X$, $Y_T \overset{p}{\to} Y$*, and* $a_T \to a$ *(a traditional limit). Let* $g(X_T, Y_T, a_T)$ *be a continuous function. Then* $g(X_T, Y_T, a_T) \overset{d}{\to} g(X, Y, a)$.

21.4 Laws of Large Numbers and Central Limit Theorems

Fact 21.41 *(Khinchine's theorem) Let* X_t *be independently and identically distributed (iid) with* $\mathrm{E}\, X_t = \mu < \infty$*. Then* $\Sigma_{t=1}^{T} X_t / T \overset{p}{\to} \mu$.

Fact 21.42 *(Chebyshev's theorem) If* $\mathrm{E}\, X_t = 0$ *and* $\lim_{T\to\infty} \mathrm{Var}(\Sigma_{t=1}^{T} X_t / T) = 0$*, then* $\Sigma_{t=1}^{T} X_t / T \overset{p}{\to} 0$.

Fact 21.43 *(The Lindeberg-Lévy theorem) Let* X_t *be independently and identically distributed (iid) with* $\mathrm{E}\, X_t = 0$ *and* $\mathrm{Var}(X_t) < \infty$*. Then* $\frac{1}{\sqrt{T}} \Sigma_{t=1}^{T} X_t / \sigma \overset{d}{\to} N(0, 1)$.

21.5 Stationarity

Fact 21.44 *(Covariance stationarity)* X_t *is covariance stationary if*

$$\begin{aligned} \mathrm{E}\, X_t &= \mu \text{ is independent of } t, \\ \mathrm{Cov}\,(X_{t-s}, X_t) &= \gamma_s \text{ depends only on } s, \text{ and} \\ &\text{both } \mu \text{ and } \gamma_s \text{ are finite.} \end{aligned}$$

Fact 21.45 *(Strict stationarity)* X_t *is strictly stationary if, for all* s*, the joint distribution of* $X_t, X_{t+1}, ..., X_{t+s}$ *does not depend on* t.

Fact 21.46 *(Strict stationarity versus covariance stationarity) In general, strict stationarity does not imply covariance stationarity or vice versa. However, strict stationary with finite first two moments implies covariance stationarity.*

21.6 Martingales

Fact 21.47 *(Martingale) Let* Ω_t *be a set of information in* t*, for instance* $Y_t, Y_{t-1}, ...$ *If* $\mathrm{E}\,|Y_t| < \infty$ *and* $\mathrm{E}(Y_{t+1}|\Omega_t) = Y_t$*, then* Y_t *is a martingale.*

Fact 21.48 *(Martingale difference) If Y_t is a martingale, then $X_t = Y_t - Y_{t-1}$ is a martingale difference: X_t has $\mathrm{E}\,|X_t| < \infty$ and $\mathrm{E}(X_{t+1}|\Omega_t) = 0$.*

Fact 21.49 *(Innovations as a martingale difference sequence) The forecast error $X_{t+1} = Y_{t+1} - \mathrm{E}(Y_{t+1}|\Omega_t)$ is a martingale difference.*

Fact 21.50 *(Properties of martingales) (a) If Y_t is a martingale, then $\mathrm{E}(Y_{t+s}|\Omega_t) = Y_t$ for $s \geq 1$. (b) If X_t is a martingale difference, then $\mathrm{E}(X_{t+s}|\Omega_t) = 0$ for $s \geq 1$.*

Proof. (a) Note that $\mathrm{E}(Y_{t+2}|\Omega_{t+1}) = Y_{t+1}$ and take expectations conditional on Ω_t: $\mathrm{E}[\mathrm{E}(Y_{t+2}|\Omega_{t+1})|\Omega_t] = \mathrm{E}(Y_{t+1}|\Omega_t) = Y_t$. By iterated expectations, the first term equals $\mathrm{E}(Y_{t+2}|\Omega_t)$. Repeat this for $t+3$, $t+4$, etc. (b) Essentially the same proof. ■

Fact 21.51 *(Properties of martingale differences) If X_t is a martingale difference and g_{t-1} is a function of Ω_{t-1}, then $X_t g_{t-1}$ is also a martingale difference.*

Proof. $\mathrm{E}(X_{t+1}g_t|\Omega_t) = \mathrm{E}(X_{t+1}|\Omega_t)g_t$ since g_t is a function of Ω_t. ■

Fact 21.52 *(Martingales, serial independence, and no autocorrelation) (a) X_t is serially uncorrelated if $\mathrm{Cov}(X_t, X_{t+s}) = 0$ for all $s \neq 0$. This means that a linear projection of X_{t+s} on $X_t, X_{t-1},...$ is a constant, so it cannot help predict X_{t+s}. (b) X_t is a martingale difference with respect to its history if $\mathrm{E}(X_{t+s}|X_t, X_{t-1}, ...) = 0$ for all $s \geq 1$. This means that no function of $X_t, X_{t-1}, ...$ can help predict X_{t+s}. (c) X_t is serially independent if $\mathrm{pdf}(X_{t+s}|X_t, X_{t-1}, ...) = \mathrm{pdf}(X_{t+s})$. This means than no function of $X_t, X_{t-1}, ...$ can help predict any function of X_{t+s}.*

Fact 21.53 *(WLN for martingale difference) If X_t is a martingale difference, then $\mathrm{plim}\, \Sigma_{t=1}^T X_t/T = 0$ if either (a) X_t is strictly stationary and $\mathrm{E}\,|x_t| < 0$ or (b) $\mathrm{E}\,|x_t|^{1+\delta} < \infty$ for $\delta > 0$ and all t. (See Davidson (2000) 6.2)*

Fact 21.54 *(CLT for martingale difference) Let X_t be a martingale difference. If $\mathrm{plim}\, \Sigma_{t=1}^T (X_t^2 - \mathrm{E}\,X_t^2)/T = 0$ and either*

(a) X_t is strictly stationary or
(b) $\max_{t\in[1,T]} \frac{(\mathrm{E}\,|X_t|^{2+\delta})^{1/(2+\delta)}}{\Sigma_{t=1}^T \mathrm{E}\,X_t^2/T} < \infty$ for $\delta > 0$ and all $T > 1$,

then $(\Sigma_{t=1}^T X_t/\sqrt{T})/(\Sigma_{t=1}^T \mathrm{E}\,X_t^2/T)^{1/2} \xrightarrow{d} N(0,1)$. (See Davidson (2000) 6.2)

21.7 Special Distributions

21.7.1 The Normal Distribution

Fact 21.55 *(Univariate normal distribution) If* $X \sim N(\mu, \sigma^2)$*, then the probability density function of* X*,* $f(x)$ *is*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}.$$

The moment generating function is $mgf_X(t) = \exp\left(\mu t + \sigma^2 t^2/2\right)$ *and the moment generating function around the mean is* $mgf_{(X-\mu)}(t) = \exp\left(\sigma^2 t^2/2\right)$.

Example 21.56 *The first few moments around the mean are* $\mathrm{E}(X-\mu) = 0$, $\mathrm{E}(X-\mu)^2 = \sigma^2$, $\mathrm{E}(X-\mu)^3 = 0$ *(all odd moments are zero),* $\mathrm{E}(X-\mu)^4 = 3\sigma^4$, $\mathrm{E}(X-\mu)^6 = 15\sigma^6$, *and* $\mathrm{E}(X-\mu)^8 = 105\sigma^8$.

Fact 21.57 *(Standard normal distribution) If* $X \sim N(0,1)$*, then the moment generating function is* $mgf_X(t) = \exp\left(t^2/2\right)$*. Since the mean is zero,* $m(t)$ *gives central moments. The first few are* $\mathrm{E}X = 0$, $\mathrm{E}X^2 = 1$, $\mathrm{E}X^3 = 0$ *(all odd moments are zero), and* $EX^4 = 3$*. The distribution function,* $\Pr(X \leq a) = \Phi(a) = 1/2 + 1/2\,\mathrm{erf}(a/\sqrt{2})$*, where* erf() *is the error function,* $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z \exp(-t^2)dt$*). The complementary error function is* $\mathrm{erfc}(z) = 1 - \mathrm{erf}(z)$*. Since the distribution is symmetric around zero, we have* $\Phi(-a) = \Pr(X \leq -a) = \Pr(X \geq a) = 1 - \Phi(a)$*. Clearly,* $1 - \Phi(-a) = \Phi(a) = 1/2\,\mathrm{erfc}(-a/\sqrt{2})$.

Fact 21.58 *(Multivariate normal distribution) If* X *is an* $n \times 1$ *vector of random variables with a multivariate normal distribution, with a mean vector* μ *and variance-covariance matrix* Σ*,* $N(\mu, \Sigma)$*, then the density function is*

$$f(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right].$$

Fact 21.59 *(Conditional normal distribution) Suppose* $Z_{m\times 1}$ *and* $X_{n\times 1}$ *are jointly normally distributed*

$$\begin{bmatrix} Z \\ X \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_Z \\ \mu_X \end{bmatrix}, \begin{bmatrix} \Sigma_{ZZ} & \Sigma_{ZX} \\ \Sigma_{XZ} & \Sigma_{XX} \end{bmatrix}\right).$$

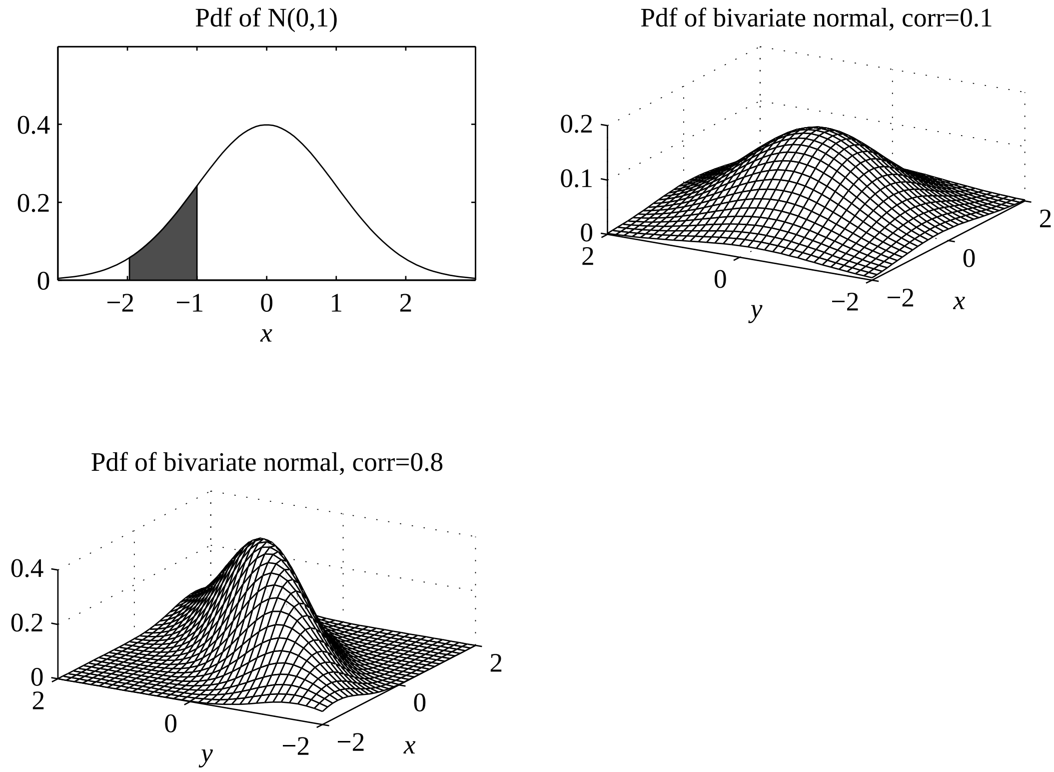

Figure 21.2: Normal distributions

The distribution of the random variable Z conditional on that $X = x$ (a number) is also normal with mean

$$\mathrm{E}\left(Z|x\right) = \mu_Z + \Sigma_{ZX}\Sigma_{XX}^{-1}\left(x - \mu_X\right),$$

and variance (variance of Z conditional on that $X = x$, that is, the variance of the prediction error $Z - \mathrm{E}\left(Z|x\right)$)

$$\mathrm{Var}\left(Z|x\right) = \Sigma_{ZZ} - \Sigma_{ZX}\Sigma_{XX}^{-1}\Sigma_{XZ}.$$

Note that the conditional variance is constant in the multivariate normal distribution ($\mathrm{Var}\left(Z|X\right)$ *is not a random variable in this case). Note also that* $\mathrm{Var}\left(Z|x\right)$ *is less than* $\mathrm{Var}(Z) = \Sigma_{ZZ}$ *(in a matrix sense) if X contains any relevant information (so Σ_{ZX} is not zero, that is,* $\mathrm{E}\left(Z|x\right)$ *is not the same for all x).*

Example 21.60 *(Conditional normal distribution) Suppose Z and X are scalars in Fact*

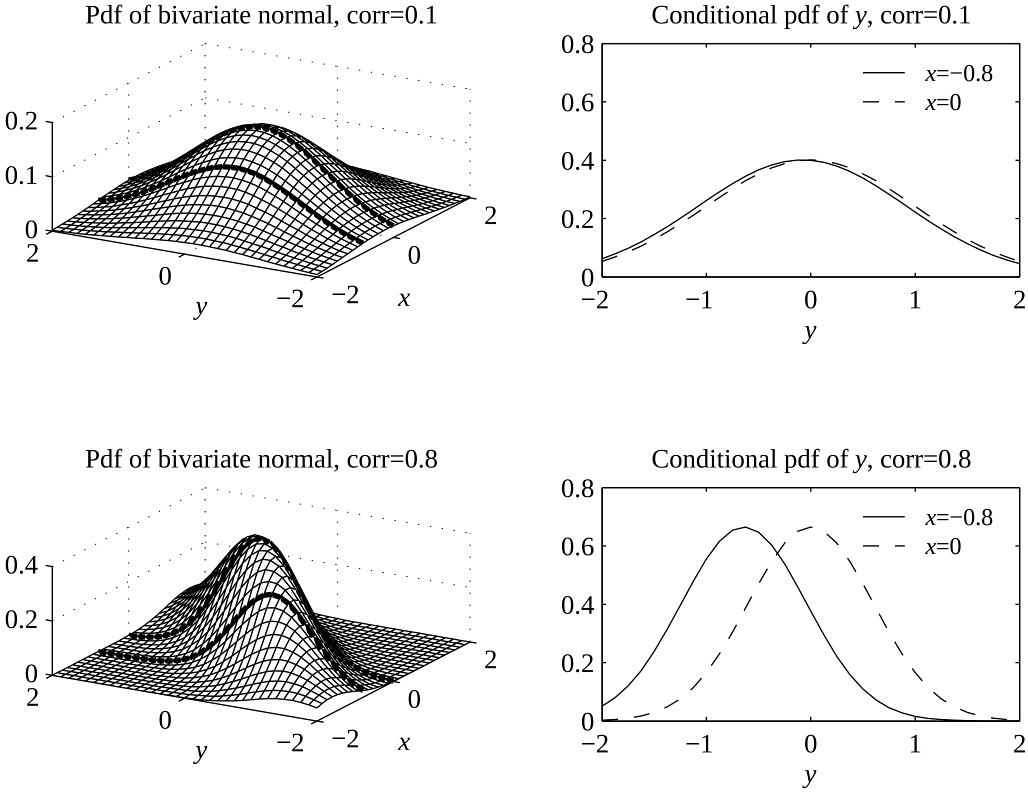

Figure 21.3: Density functions of normal distributions

21.59 and that the joint distribution is

$$\begin{bmatrix} Z \\ X \end{bmatrix} \sim N\left(\begin{bmatrix} 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 2 & 6 \end{bmatrix} \right).$$

The expectation of Z conditional on $X = x$ is then

$$\mathrm{E}\,(Z|x) = 3 + \frac{2}{6}\,(x-5) = 3 + \frac{1}{3}\,(x-5)\,.$$

Similarly, the conditional variance is

$$\mathrm{Var}\,(Z|x) = 1 - \frac{2 \times 2}{6} = \frac{1}{3}.$$

Fact 21.61 *(Stein's lemma) If Y has normal distribution and $h()$ is a differentiable function such that* $\mathrm{E}\,|h'(Y)| < \infty$, *then* $\mathrm{Cov}[Y, h(Y)] = \mathrm{Var}(Y)\,\mathrm{E}\,h'(Y)$.

Proof. $\mathrm{E}[(Y-\mu)h(Y)] = \int_{-\infty}^{\infty}(Y-\mu)h(Y)\phi(Y;\mu,\sigma^2)dY$, where $\phi(Y;\mu,\sigma^2)$ is the pdf of $N(\mu,\sigma^2)$. Note that $d\phi(Y;\mu,\sigma^2)/dY = -\phi(Y;\mu,\sigma^2)(Y-\mu)/\sigma^2$, so the integral can be rewritten as $-\sigma^2\int_{-\infty}^{\infty}h(Y)d\phi(Y;\mu,\sigma^2)$. Integration by parts ("$\int udv = uv - \int vdu$") gives $-\sigma^2\left[h(Y)\phi(Y;\mu,\sigma^2)\big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\phi(Y;\mu,\sigma^2)h'(Y)dY\right] = \sigma^2\,\mathrm{E}\,h'(Y)$. ■

Fact 21.62 *(Stein's lemma 2) It follows from Fact 21.61 that if X and Y have a bivariate normal distribution and $h()$ is a differentiable function such that $\mathrm{E}\,|h'(Y)| < \infty$, then* $\mathrm{Cov}[X,h(Y)] = \mathrm{Cov}(X,Y)\,\mathrm{E}\,h'(Y)$.

Example 21.63 *(a) With $h(Y) = \exp(Y)$ we get $\mathrm{Cov}[X,\exp(Y)] = \mathrm{Cov}(X,Y)\,\mathrm{E}\exp(Y)$; (b) with $h(Y) = Y^2$ we get $\mathrm{Cov}[X,Y^2] = \mathrm{Cov}(X,Y)2\,\mathrm{E}\,Y$ so with $\mathrm{E}\,Y = 0$ we get a zero covariance.*

Fact 21.64 *(Stein's lemma 3) Fact 21.62 still holds if the joint distribution of X and Y is a mixture of n bivariate normal distributions, provided the mean and variance of Y is the same in each of the n components. (See Söderlind (2009) for a proof.)*

Fact 21.65 *(Truncated normal distribution) Let $X \sim N(\mu,\sigma^2)$, and consider truncating the distribution so that we want moments conditional on $a < X \leq b$. Define $a_0 = (a-\mu)/\sigma$ and $b_0 = (b-\mu)/\sigma$. Then,*

$$\mathrm{E}(X|a < X \leq b) = \mu - \sigma\frac{\phi(b_0)-\phi(a_0)}{\Phi(b_0)-\Phi(a_0)} \text{ and}$$

$$\mathrm{Var}(X|a < X \leq b) = \sigma^2\left\{1 - \frac{b_0\phi(b_0)-a_0\phi(a_0)}{\Phi(b_0)-\Phi(a_0)} - \left[\frac{\phi(b_0)-\phi(a_0)}{\Phi(b_0)-\Phi(a_0)}\right]^2\right\}.$$

Fact 21.66 *(Lower truncation) In Fact 21.65, let $b \to \infty$, so we only have the truncation $a < X$. Then, we have*

$$\mathrm{E}(X|a < X) = \mu + \sigma\frac{\phi(a_0)}{1-\Phi(a_0)} \text{ and}$$

$$\mathrm{Var}(X|a < X) = \sigma^2\left\{1 + \frac{a_0\phi(a_0)}{1-\Phi(a_0)} - \left[\frac{\phi(a_0)}{1-\Phi(a_0)}\right]^2\right\}.$$

(The latter follows from $\lim_{b\to\infty} b_0\phi(b_0) = 0$.)

Example 21.67 *Suppose $X \sim N(0,\sigma^2)$ and we want to calculate $\mathrm{E}\,|x|$. This is the same as $\mathrm{E}(X|X > 0) = 2\sigma\phi(0)$.*

Fact 21.68 *(Upper truncation) In Fact 21.65, let $a \to -\infty$, so we only have the truncation $X \leq b$. Then, we have*

$$\mathrm{E}(X|X \leq b) = \mu - \sigma\frac{\phi(b_0)}{\Phi(b_0)} \text{ and}$$
$$\mathrm{Var}(X|X \leq b) = \sigma^2\left\{1 - \frac{b_0\phi(b_0)}{\Phi(b_0)} - \left[\frac{\phi(b_0)}{\Phi(b_0)}\right]^2\right\}.$$

(The latter follows from $\lim_{a\to-\infty} a_0\phi(a_0) = 0$.)

Fact 21.69 *(Delta method) Consider an estimator $\hat{\beta}_{k\times1}$ which satisfies*

$$\sqrt{T}\left(\hat{\beta} - \beta_0\right) \overset{d}{\to} N\left(0, \Omega\right),$$

and suppose we want the asymptotic distribution of a transformation of β

$$\gamma_{q\times1} = g\left(\beta\right),$$

where $g\left(.\right)$ is has continuous first derivatives. The result is

$$\sqrt{T}\left[g\left(\hat{\beta}\right) - g\left(\beta_0\right)\right] \overset{d}{\to} N\left(0, \Psi_{q\times q}\right), \text{ where}$$
$$\Psi = \frac{\partial g\left(\beta_0\right)}{\partial\beta'}\Omega\frac{\partial g\left(\beta_0\right)'}{\partial\beta}, \text{ where } \frac{\partial g\left(\beta_0\right)}{\partial\beta'} \text{ is } q \times k.$$

Proof. By the mean value theorem we have

$$g\left(\hat{\beta}\right) = g\left(\beta_0\right) + \frac{\partial g\left(\beta^*\right)}{\partial\beta'}\left(\hat{\beta} - \beta_0\right),$$

where

$$\frac{\partial g\left(\beta\right)}{\partial\beta'} = \begin{bmatrix} \frac{\partial g_1(\beta)}{\partial\beta_1} & \cdots & \frac{\partial g_1(\beta)}{\partial\beta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_q(\beta)}{\partial\beta_1} & \cdots & \frac{\partial g_q(\beta)}{\partial\beta_k} \end{bmatrix}_{q\times k},$$

and we evaluate it at β^* which is (weakly) between $\hat{\beta}$ and β_0. Premultiply by $\sqrt{T}$ and rearrange as

$$\sqrt{T}\left[g\left(\hat{\beta}\right) - g\left(\beta_0\right)\right] = \frac{\partial g\left(\beta^*\right)}{\partial\beta'}\sqrt{T}\left(\hat{\beta} - \beta_0\right).$$

If $\hat{\beta}$ is consistent ($\operatorname{plim} \hat{\beta} = \beta_0$) and $\partial g(\beta^*) / \partial \beta'$ is continuous, then by Slutsky's theorem $\operatorname{plim} \partial g(\beta^*) / \partial \beta' = \partial g(\beta_0) / \partial \beta'$, which is a constant. The result then follows from the continuous mapping theorem. ■

21.7.2 The Lognormal Distribution

Fact 21.70 *(Univariate lognormal distribution) If $x \sim N(\mu, \sigma^2)$ and $y = \exp(x)$ then the probability density function of y, $f(y)$ is*

$$f(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\ln y - \mu}{\sigma}\right)^2}, \; y > 0.$$

The rth moment of y is $\mathrm{E}\, y^r = \exp(r\mu + r^2\sigma^2/2)$. See 21.4 for an illustration.

Example 21.71 *The first two moments are $\mathrm{E}\, y = \exp\left(\mu + \sigma^2/2\right)$ and $\mathrm{E}\, y^2 = \exp(2\mu + 2\sigma^2)$. We therefore get $\mathrm{Var}(y) = \exp\left(2\mu + \sigma^2\right)\left[\exp\left(\sigma^2\right) - 1\right]$ and $\mathrm{Std}(y) / \mathrm{E}\, y = \sqrt{\exp(\sigma^2) - 1}$.*

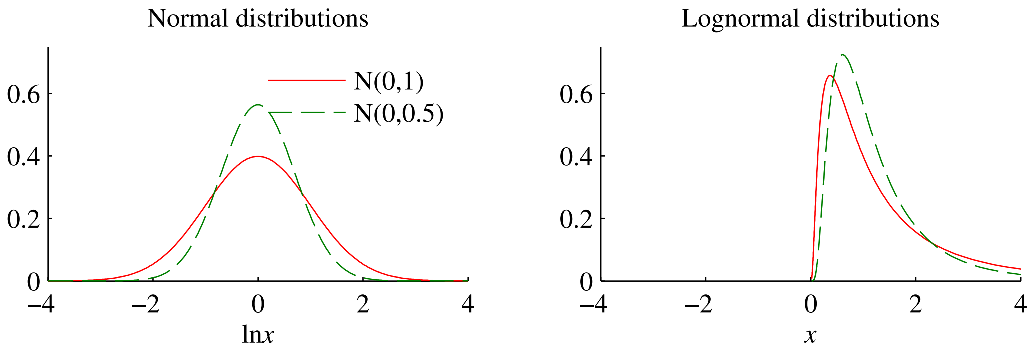

Figure 21.4: Lognormal distribution

Fact 21.72 *(Moments of a truncated lognormal distribution) If $x \sim N(\mu, \sigma^2)$ and $y = \exp(x)$ then $\mathrm{E}(y^r | y > a) = \mathrm{E}(y^r)\Phi(r\sigma - a_0)/\Phi(-a_0)$, where $a_0 = (\ln a - \mu)/\sigma$. Note that the denominator is $\Pr(y > a) = \Phi(-a_0)$. In contrast, $\mathrm{E}(y^r | y \leq b) = \mathrm{E}(y^r)\Phi(-r\sigma + b_0)/\Phi(b_0)$, where $b_0 = (\ln b - \mu)/\sigma$. The denominator is $\Pr(y \leq b) = \Phi(b_0)$. Clearly, $\mathrm{E}(y^r) = \exp(r\mu + r^2\sigma^2/2)$*

Fact 21.73 *(Moments of a truncated lognormal distribution, two-sided truncation) If* $x \sim N(\mu, \sigma^2)$ *and* $y = \exp(x)$ *then*

$$\mathrm{E}(y^r | a > y < b) = \mathrm{E}(y^r) \frac{\Phi(r\sigma - a_0) - \Phi(r\sigma - b_0)}{\Phi(b_0) - \Phi(a_0)},$$

where $a_0 = (\ln a - \mu)/\sigma$ *and* $b_0 = (\ln b - \mu)/\sigma$. *Note that the denominator is* $\Pr(a > y < b) = \Phi(b_0) - \Phi(a_0)$. *Clearly,* $\mathrm{E}(y^r) = \exp(r\mu + r^2\sigma^2/2)$.

Example 21.74 *The first two moments of the truncated (from below) lognormal distribution are* $\mathrm{E}(y|y > a) = \exp\left(\mu + \sigma^2/2\right)\Phi(\sigma - a_0)/\Phi(-a_0)$ *and* $\mathrm{E}(y^2|y > a) = \exp\left(2\mu + 2\sigma^2\right)\Phi(2\sigma - a_0)/\Phi(-a_0)$.

Example 21.75 *The first two moments of the truncated (from above) lognormal distribution are* $\mathrm{E}(y|y \leq b) = \exp\left(\mu + \sigma^2/2\right)\Phi(-\sigma + b_0)/\Phi(b_0)$ *and* $\mathrm{E}(y^2|y \leq b) = \exp\left(2\mu + 2\sigma^2\right)\Phi(-2\sigma + b_0)/\Phi(b_0)$.

Fact 21.76 *(Multivariate lognormal distribution) Let the* $n \times 1$ *vector* x *have a mulivariate normal distribution*

$$x \sim N(\mu, \Sigma), \text{ where } \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{bmatrix}.$$

Then $y = \exp(x)$ *has a lognormal distribution, with the means and covariances*

$$\begin{aligned} \mathrm{E}\, y_i &= \exp\left(\mu_i + \sigma_{ii}/2\right) \\ \mathrm{Cov}(y_i, y_j) &= \exp\left[\mu_i + \mu_j + (\sigma_{ii} + \sigma_{jj})/2\right]\left[\exp(\sigma_{ij}) - 1\right] \\ \mathrm{Corr}(y_i, y_j) &= \left[\exp(\sigma_{ij}) - 1\right] / \sqrt{\left[\exp(\sigma_{ii}) - 1\right]\left[\exp(\sigma_{jj}) - 1\right]}. \end{aligned}$$

Cleary, $\mathrm{Var}(y_i) = \exp\left[2\mu_i + \sigma_{ii}\right]\left[\exp(\sigma_{ii}) - 1\right]$. $\mathrm{Cov}(y_1, y_2)$ *and* $\mathrm{Corr}(y_1, y_2)$ *have the same sign as* $\mathrm{Corr}(x_i, x_j)$ *and are increasing in it. However,* $\mathrm{Corr}(y_i, y_j)$ *is closer to zero.*

21.7.3 The Chi-Square Distribution

Fact 21.77 *(The* χ^2_n *distribution) If* $Y \sim \chi^2_n$, *then the pdf of* Y *is* $f(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}$, *where* $\Gamma()$ *is the gamma function. The moment generating function is* $mgf_Y(t) = (1 - 2t)^{-n/2}$ *for* $t < 1/2$. *The first moments of* Y *are* $\mathrm{E}\, Y = n$ *and* $\mathrm{Var}(Y) = 2n$.

Fact 21.78 *(Quadratic forms of normally distribution random variables) If the $n \times 1$ vector $X \sim N(0, \Sigma)$, then $Y = X'\Sigma^{-1}X \sim \chi^2_n$. Therefore, if the n scalar random variables X_i, $i = 1, ..., n$, are uncorrelated and have the distributions $N(0, \sigma_i^2)$, $i = 1, ..., n$, then $Y = \Sigma_{i=1}^n X_i^2/\sigma_i^2 \sim \chi^2_n$.*

Fact 21.79 *(Distribution of $X'AX$) If the $n \times 1$ vector $X \sim N(\mathbf{0}, I)$, and A is a symmetric idempotent matrix ($A = A'$ and $A = AA = A'A$) of rank r, then $Y = X'AX \sim \chi^2_r$.*

Fact 21.80 *(Distribution of $X'\Sigma^+X$) If the $n \times 1$ vector $X \sim N(0, \Sigma)$, where Σ has rank $r \leq n$ then $Y = X'\Sigma^+X \sim \chi^2_r$ where Σ^+ is the pseudo inverse of Σ.*

Proof. Σ is symmetric, so it can be decomposed as $\Sigma = C\Lambda C'$ where C are the orthogonal eigenvector ($C'C = I$) and Λ is a diagonal matrix with the eigenvalues along the main diagonal. We therefore have $\Sigma = C\Lambda C' = C_1\Lambda_{11}C_1'$ where C_1 is an $n \times r$ matrix associated with the r non-zero eigenvalues (found in the $r \times r$ matrix Λ_{11}). The generalized inverse can be shown to be

$$\Sigma^+ = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} \Lambda_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} C_1 & C_2 \end{bmatrix}' = C_1\Lambda_{11}^{-1}C_1',$$

We can write $\Sigma^+ = C_1\Lambda_{11}^{-1/2}\Lambda_{11}^{-1/2}C_1'$. Consider the $r \times 1$ vector $Z = \Lambda_{11}^{-1/2}C_1'X$, and note that it has the covariance matrix

$$\mathrm{E}\, ZZ' = \Lambda_{11}^{-1/2}C_1' \,\mathrm{E}\, XX'C_1\Lambda_{11}^{-1/2} = \Lambda_{11}^{-1/2}C_1'C_1\Lambda_{11}C_1'C_1\Lambda_{11}^{-1/2} = I_r,$$

since $C_1'C_1 = I_r$. This shows that $Z \sim N(\mathbf{0}_{r\times 1}, I_r)$, so $Z'Z = X'\Sigma^+X \sim \chi^2_r$. ■

Fact 21.81 *(Convergence to a normal distribution) Let $Y \sim \chi^2_n$ and $Z = (Y - n)/n^{1/2}$. Then $Z \overset{d}{\to} N(0, 2)$.*

Example 21.82 *If $Y = \Sigma_{i=1}^n X_i^2/\sigma_i^2$, then this transformation means $Z = (\Sigma_{i=1}^n X_i^2/\sigma_i^2 - 1)/n^{1/2}$.*

Proof. We can directly note from the moments of a χ^2_n variable that $\mathrm{E}\, Z = (\mathrm{E}\, Y - n)/n^{1/2} = 0$, and $\mathrm{Var}(Z) = \mathrm{Var}(Y)/n = 2$. From the general properties of moment generating functions, we note that the moment generating function of Z is

$$mgf_Z(t) = e^{-t\sqrt{n}} \left(1 - 2\frac{t}{n^{1/2}}\right)^{-n/2} \text{ with } \lim_{n\to\infty} mgf_Z(t) = \exp(t^2).$$

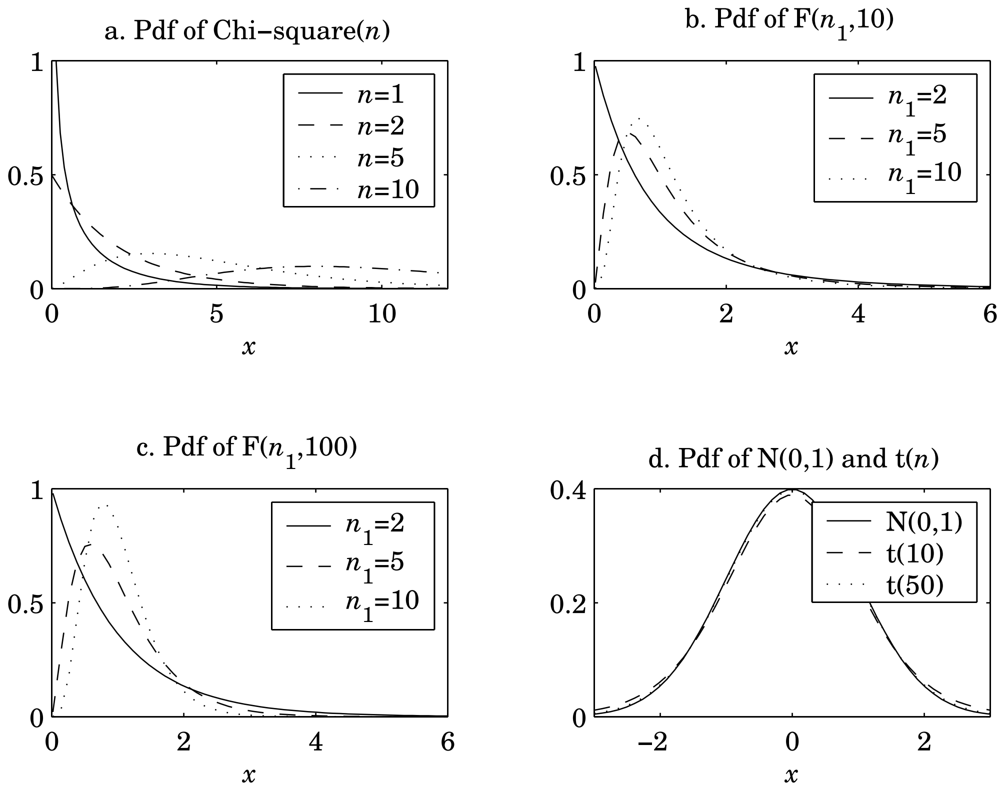

Figure 21.5: χ^2, F, and t distributions

This is the moment generating function of a $N(0,2)$ distribution, which shows that $Z \xrightarrow{d} N(0,2)$. This result should not come as a surprise as we can think of Y as the sum of n variables; dividing by $n^{1/2}$ is then like creating a scaled sample average for which a central limit theorem applies. ■

21.7.4 The t and F Distributions

Fact 21.83 *(The $F(n_1, n_2)$ distribution) If $Y_1 \sim \chi^2_{n_1}$ and $Y_2 \sim \chi^2_{n_2}$ and Y_1 and Y_2 are independent, then $Z = (Y_1/n_1)/(Y_2/n_2)$ has an $F(n_1, n_2)$ distribution. This distribution has no moment generating function, but* $\mathrm{E}\, Z = n_2/(n_2 - 2)$ *for $n > 2$.*

Fact 21.84 *(Convergence of an $F(n_1, n_2)$ distribution) In Fact (21.83), the distribution of $n_1 Z = Y_1/(Y_2/n_2)$ converges to a $\chi^2_{n_1}$ distribution as $n_2 \to \infty$. (The idea is essentially that $n_2 \to \infty$ the denominator converges to the mean, which is* $\mathrm{E}\, Y_2/n_2 = 1$. *Only the numerator is then left, which is a $\chi^2_{n_1}$ variable.)*

Fact 21.85 *(The t_n distribution) If $X \sim N(0,1)$ and $Y \sim \chi_n^2$ and X and Y are independent, then $Z = X/(Y/n)^{1/2}$ has a t_n distribution. The moment generating function does not exist, but* $\mathrm{E}\,Z = 0$ *for $n > 1$ and Var$(Z) = n/(n-2)$ for $n > 2$.*

Fact 21.86 *(Convergence of a t_n distribution) The t distribution converges to a $N(0,1)$ distribution as $n \to \infty$.*

Fact 21.87 *(t_n versus $F(1,n)$ distribution) If $Z \sim t_n$, then $Z^2 \sim F(1,n)$.*

21.7.5 The Bernouilli and Binomial Distributions

Fact 21.88 *(Bernoulli distribution) The random variable X can only take two values: 1 or 0, with probability p and $1-p$ respectively. The moment generating function is $mgf(t) = pe^t + 1 - p$. This gives* $\mathrm{E}(X) = p$ *and* $\mathrm{Var}(X) = p(1-p)$.

Example 21.89 *(Shifted Bernoulli distribution) Suppose the Bernoulli variable takes the values a or b (instead of 1 and 0) with probability p and $1-p$ respectively. Then* $\mathrm{E}(X) = pa + (1-p)b$ *and* $\mathrm{Var}(X) = p(1-p)(a-b)^2$.

Fact 21.90 *(Binomial distribution). Suppose $X_1, X_2, ..., X_n$ all have Bernoulli distributions with the parameter p. Then, the sum $Y = X_1 + X_2 + ... + X_n$ has a Binomial distribution with parameters p and n. The pdf is pdf$(Y) = n!/[y!(n-y)!]p^y(1-p)^{n-y}$ for $y = 0, 1, ..., n$. The moment generating function is $mgf(t) = [pe^t + 1 - p]^n$. This gives* $\mathrm{E}(Y) = np$ *and* $\mathrm{Var}(Y) = np(1-p)$.

Example 21.91 *(Shifted Binomial distribution) Suppose the Bernuolli variables $X_1, X_2, ..., X_n$ take the values a or b (instead of 1 and 0) with probability p and $1-p$ respectively. Then, the sum $Y = X_1 + X_2 + ... + X_n$ has* $\mathrm{E}(Y) = n[pa + (1-p)b]$ *and* $\mathrm{Var}(Y) = n[p(1-p)(a-b)^2]$.

21.7.6 The Skew-Normal Distribution

Fact 21.92 *(Skew-normal distribution) Let ϕ and Φ be the standard normal pdf and cdf respectively. The pdf of a skew-normal distribution with shape parameter α is then*

$$f(z) = 2\phi(z)\Phi(\alpha z).$$

If Z has the above pdf and

$$Y = \mu + \omega Z \text{ with } \omega > 0,$$

then Y is said to have a SN(μ, ω^2, α) distribution (see Azzalini (2005)). Clearly, the pdf of Y is

$$f(y) = 2\phi\left[(y-\mu)/\omega\right]\Phi\left[\alpha\,(y-\mu)/\omega\right]/\omega.$$

The moment generating function is $mgf_y(t) = 2\exp\left(\mu t + \omega^2 t^2/2\right)\Phi(\delta\omega t)$ where $\delta = \alpha/\sqrt{1+\alpha^2}$. When $\alpha > 0$ then the distribution is positively skewed (and vice versa)—and when $\alpha = 0$ the distribution becomes a normal distribution. When $\alpha \to \infty$, then the density function is zero for $Y \le \mu$, and $2\phi\left[(y-\mu)/\omega\right]/\omega$ otherwise—this is a half-normal distribution.

Example 21.93 *The first three moments are as follows. First, notice that $\mathrm{E}\,Z = \sqrt{2/\pi}\delta$, $\mathrm{Var}(Z) = 1 - 2\delta^2/\pi$ and $\mathrm{E}(Z - \mathrm{E}\,Z)^3 = (4/\pi - 1)\sqrt{2/\pi}\delta^3$. Then we have*

$$\begin{aligned}
\mathrm{E}\,Y &= \mu + \omega\,\mathrm{E}\,Z \\
\mathrm{Var}(Y) &= \omega^2\,\mathrm{Var}(Z) \\
\mathrm{E}\,(Y - \mathrm{E}\,Y)^3 &= \omega^3\,\mathrm{E}(Z - \mathrm{E}\,Z)^3.
\end{aligned}$$

Notice that with $\alpha = 0$ (so $\delta = 0$), then these moments of Y become μ, ω^2 and 0 respecively.

21.7.7 Generalized Pareto Distribution

Fact 21.94 *(Cdf and pdf of the generalized Pareto distribution) The generalized Pareto distribution is described by a scale parameter ($\beta > 0$) and a shape parameter (ξ). The cdf ($\Pr(Z \le z)$, where Z is the random variable and z is a value) is*

$$G(z) = \begin{cases} 1 - (1 + \xi z/\beta)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp(-z/\beta) & \xi = 0, \end{cases}$$

for $0 \le z$ and $z \le -\beta/\xi$ in case $\xi < 0$. The pdf is therefore

$$g(z) = \begin{cases} \frac{1}{\beta}(1 + \xi z/\beta)^{-1/\xi - 1} & \text{if } \xi \neq 0 \\ \frac{1}{\beta}\exp(-z/\beta) & \xi = 0. \end{cases}$$

The mean is defined (finite) if $\xi < 1$ *and is then* $\mathrm{E}(z) = \beta/(1-\xi)$*, the median is* $(2^{\xi}-1)\beta/\xi$ *and the variance is defined if* $\xi < 1/2$ *and is then* $\beta^2/[(1-\xi)^2(1-2\xi)]$.

21.8 Inference

Fact 21.95 *(Comparing variance-covariance matrices) Let* $\mathrm{Var}(\hat{\beta})$ *and* $\mathrm{Var}(\beta^*)$ *be the variance-covariance matrices of two estimators,* $\hat{\beta}$ *and* β^**, and suppose* $\mathrm{Var}(\hat{\beta}) - \mathrm{Var}(\beta^*)$ *is a positive semi-definite matrix. This means that for any non-zero vector* R *that* $R' \mathrm{Var}(\hat{\beta}) R \geq R' \mathrm{Var}(\beta^*) R$*, so every linear combination of* $\hat{\beta}$ *has a variance that is as large as the variance of the same linear combination of* β^**. In particular, this means that the variance of every element in* $\hat{\beta}$ *(the diagonal elements of* $\mathrm{Var}(\hat{\beta})$*) is at least as large as variance of the corresponding element of* β^**.*

Bibliography

Azzalini, A., 2005, "The skew-normal distribution and related Multivariate Families," *Scandinavian Journal of Statistics*, 32, 159–188.

Davidson, J., 2000, *Econometric theory*, Blackwell Publishers, Oxford.

DeGroot, M. H., 1986, *Probability and statistics*, Addison-Wesley, Reading, Massachusetts.

Greene, W. H., 2000, *Econometric analysis*, Prentice-Hall, Upper Saddle River, New Jersey, 4th edn.

Johnson, N. L., S. Kotz, and N. Balakrishnan, 1994, *Continuous univariate distributions*, Wiley, New York, 2nd edn.

Mittelhammer, R. C., 1996, *Mathematical statistics for economics and business*, Springer-Verlag, New York.

Söderlind, P., 2009, "An extended Stein's lemma for asset pricing," *Applied Economics Letters,* forthcoming, 16, 1005–1008.