B9311-020 Introduction to Econometrics

Week 2 Lecture Notes

1 Some Discrete Distributions

- Bernoulli: X can take on two values, 0 and 1, $f(1) = p$ and $f(0) = 1 - p$. Thus, $f(x) = p^x(1-p)^{1-x}$
  - The parameter p indexes the distribution
  - Exercise: Work out MGF and all moments...
- Binomial: Suppose X_i , $i = 1, ..., n$ are Independent and Identically Distributed (*IID*) Bernoulli random variables with parameter p. Let $Y = \sum_{i=1}^{n} X_i$. Then Y has a Binomial distribution with parameters n and p. Y can take on values $0, 1, ..., n$. and

$$f(y) = \binom{n}{y} p^y (1-p)^{n-y}$$

where

$$\binom{n}{y} = \frac{n!}{y!(n-y!)}$$

is the number of ways that y successes can occur in n outcomes.

- Exercise: work out MGF of Y. Use: $M_Y(t) = \Pi_{i=1}^{n} M_{X_i}(t)$ which follows from (i) $(e^{t\sum X_i} = \Pi e^{tX_i})$ and (ii) independence.
  - Poisson: X takes on the values $0, 1, 2, ...$ with

$$f_X(x) = \frac{m^x e^{-m}}{x!}$$

This distribution is useful for modeling "successes" that occur over intervals of time. (Customers walking into a store, changes in Fed Funds Rate, etc.). Let $g(x, w)$ denote the probability that x successes occur in a period of length w. Suppose

1. $g(1, h) = \lambda h + o(h)$, where λ is a positive constant, $h > 0$, and $o(h)$ means a term that satisfies $\lim_{h \to 0}[o(h)/h)] = 0$
2. $\sum_{x=2}^{\infty} g(x, h) = o(h)$
3. The number of successes in non-overlapping periods are independent.

When these postulates describe an experiment, then you can show (See Hogg and Craig Section 3.2) that the number of successes over a period of time with length w follows a Poisson distribution with parameter $m = \lambda w$.

- Exercise: You should be able to show the MGF is $e^{m(e^t - 1)}$, and that both the mean and variance are equal to m.

2 Some Continuous Distributions

- Uniform: $f(x) = (b-a)^{-1}$ for $a \leq x \leq b$ and 0 elsewhere
  - MGF is
  $$M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}$$
- Univariate Normal
  - Standard Normal (denoted $N(0,1)$):
  $$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$
  - General Normal (denoted $N(\mu, \sigma^2)$): Let $Y = \mu + \sigma Z$ where Z is standard normal and $\sigma > 0$. Then from the change-of-variables formula
  $$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$
  - To compute the MGF
  $$\begin{aligned} M_Y(t) &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2\sigma^2}(y-\mu)^2 + ty]dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2\sigma^2}\{(y-\mu)^2 - 2\sigma^2 ty\}]dy \\ &= \exp[\frac{2\sigma^2\mu t + \sigma^4 t^2}{2\sigma^2}] \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2\sigma^2}(y - \mu - \sigma^2 t)^2]dy \\ &= e^{\mu t + \frac{\sigma^2 t^2}{2}} \end{aligned}$$

    since the integral term =1 (it is the integral of the density of a random variable distributed $N(\mu + \sigma^2 t, \sigma^2)$).
  - Thus
    1. $E(Y) = \mu$
    2. $E(Y^2) = \sigma^2 + \mu^2$, so that $Var(Y) = \sigma^2$
    3. $E[(Y-\mu)^k] = 0$, for $k = 1, 3, 5, ...$
    4. $E[(Y-\mu)^4] = 3\sigma^4$
- Chi-Squared Distribution: Let Z_i, $i = 1, ..., n$ be distributed $NIID(0,1)$ (where $NIID$ denotes *Normal, Independent and Identically Distributed*) and let $Y = \sum_{i=1}^{n} Z_i^2$. Then Y_i is distributed as a χ_n^2 random variable. The parameter n is called the *degrees of freedom* of the distribution

- F Distribution: Let $Y \sim \chi^2_n$, $X \sim \chi^2_m$ and suppose that Y and X are. Then
$$Q = \frac{Y/n}{X/m}$$
is distributed $F_{n,m}$. The parameters n and m are called the numerator and denominator degrees of freedom.

- Students t distribution: Let $Z \sim N(0,1)$ and $Y \sim \chi^2_n$ and suppose Z and Y are independent. Then,
$$X = \frac{Z}{(Y/n)^{\frac{1}{2}}}$$
is distributed t_n. The parameter n is called the degrees of freedom of the distribution.

2.1 Multivariate Normal Distribution.

Definition : A p–dimensional random vector X is p–dimensionally normally distributed if the one–dimensional random variables $a'X$ are normally distributed for all $a \in R^p$. (See Rao page 518)

It follows from this definition that if X is p-dimensionally normally distributed, then X_i is normally distributed. The mean vector and covariance matrix of X therefore both exist. Let them be denoted by μ and Σ.

For any $a \in R^p$, we get
$$E[a'X] = a'\mu \qquad \text{and} \qquad V[a'X] = a'\Sigma a.$$

Therefore
$$M_X(a) = Ee^{a'X} = M_{a'X}(1) = e^{a'\mu + \frac{1}{2}a'\Sigma a},$$
where $M(\cdot)$ denotes the moment generating function. Note that this implies that the distribution of X is completely characterized by μ and Σ. We write $X \sim N(\mu, \Sigma)$ (or sometimes $X \sim N_p(\mu, \Sigma)$).

We will now state a number of results about multivariate normal distributions. None of them will be proved. (They are proved in C. R. Rao: *Linear Statistical Inference and Its Applications* pp. 185–189 and pp. 519–527.

Theorem A. Let X be an p–dimensional random vector. If there exist a vector μ and a matrix Σ such that $a'X \sim N(a'\mu, a'\Sigma a)$ for all $a \in R^p$, then $X \sim N_p(\mu, \Sigma)$.

Theorem B. Let $X \sim N_p(\mu, \Sigma)$, let B be a $k \times p$ matrix and let η denote a $k \times 1$ vector, then
$$Y \stackrel{def}{=} \eta + BX \sim N_k(\eta + B\mu, B\Sigma B').$$

Theorem C. If $X_1 \sim N_p(\mu_1, \Sigma_1)$ and $X_2 \sim N_q(\mu_2, \Sigma_2)$, and X_1 and X_2 are
then $X = (X_1', X_2')' \sim N_{p+q}(\mu, \Sigma)$, where

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \text{and} \qquad \Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}.$$

Theorem D. If $X_1 \sim N_p(\mu_1, \Sigma_1)$ and $X_2 \sim N_p(\mu_2, \Sigma_2)$, and X_1 and X_2 are
then

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2).$$

Let $X \sim N_p(\mu, \Sigma)$. Also let $X = (X_1', X_2')'$, $\mu = (\mu_1', \mu_2')'$, and

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

be the partitions of X, μ and Σ such that X_1 and μ_1 are k–dimensional and Σ_{11} is a $k \times k$ matrix, then:

Theorem E. The marginal distribution of X_1 is $N_k(\mu_1, \Sigma_{11})$.

Theorem F. If $\Sigma_{12} = 0$ then X_1 and X_2 are independent.

Theorem F can be generalized:

Theorem G. If $X \sim N_p(\mu, \Sigma)$, B is a $p \times k$ matrix, and C is a $p \times m$ matrix, then $B'X$ and $C'X$ are independent if and only if $B'\Sigma C = 0$.

Theorem H. The conditional distribution of X_1 given $X_2 = x_2$ is given by

$$X_1 | X_2 = x_2 \sim N_k \left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right).$$

REMARK: If Σ_{22} is singular then the Theorem H remains true, but Σ_{22}^{-1} should be interpreted as any matrix satisfying $\Sigma_{22}\Sigma_{22}^{-1}\Sigma_{22} = \Sigma_{22}$.

We define the rank of the distribution $N_p(\mu, \Sigma)$ to be the rank of the covariance matrix, Σ.

Theorem I. X is $N_p(\mu, \Sigma)$ of rank r if and only if

$$X = \mu + BU, \qquad BB' = \Sigma$$

where B is a $r \times p$ matrix of rank r and $U \sim N_r(0, I)$, i.e. the components $U_1, U_2, \ldots, U_r$ are i.i.d. $N(0,1)$.

Theorem J. Suppose $X \sim N_p(\mu, \Sigma)$ and Σ has rank p. Then X has density given by

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\}, \quad x \in R^p.$$

Theorem K. If $X \sim N_p(\mu, \Sigma)$ where Σ has rank p, then

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

We will call a quantity of the form $Y'AY$ a quadratic form. Without loss of generality, assume that A is symmetric. (This follows since $Y'AY = Y'A'Y$, so that $Y'AY = Y'BY$, with $B = \frac{1}{2}(A + A')$.)

Theorem L. Let $X \sim N_p(\mu, \Sigma)$. A necessary and sufficient condition that

$$Q \stackrel{def}{=} (X-\mu)'A(X-\mu) \sim \chi^2_k$$

is that $\Sigma(A\Sigma A - A)\Sigma = 0$ in which case $k = rank(A\Sigma)$.

In Theorem L, note that if Σ is of full rank, then a necessary and sufficient condition is that $A\Sigma A = A$. If $\Sigma = I$, then the condition is that $A^2 = A$ (so that A is idempotent).

Theorem M. Let $X \sim N_p(0, I)$, and assume that $Q_1 \stackrel{def}{=} X'A_1X \sim \chi^2_{a_1}$ and $Q_2 \stackrel{def}{=} X'A_2X \sim \chi^2_{a_2}$. A necessary and sufficient condition that Q_1 and Q_2 are independent is that $A_1A_2 = 0$.

Theorem N. Let $X \sim N_p(\mu, \Sigma)$. A necessary and sufficient condition that $P'X$ and $(X-\mu)'A(X-\mu)$ are independent is that

$$\Sigma A \Sigma P = 0.$$

Theorem O. Let $X \sim N_p(\mu, \Sigma)$. A necessary and sufficient condition that

$$(X-\mu)'A_1(X-\mu) \quad \text{and} \quad (X-\mu)'A_2(X-\mu)$$

are independent is that

$$\Sigma A_1 \Sigma A_2 \Sigma = 0.$$

Theorem P. Let $X \sim N_n(0, I)$. Let $Q_1, \ldots, Q_k$ be quadratic forms (in X) with matrices $A_1, \ldots, A_k$ of rank $r_1, \ldots, r_k$. Assume that we can write

$$X'X = Q_1 + Q_2 + \ldots, Q_k.$$

The following statements are then equivalent:

a. The random variables $Q_1, Q_2, \ldots, Q_k$ are mutually stochastically independent and Q_j is $\chi^2_{r_j}$

b. $\sum_{j=1}^{k} r_j = n$.

(Theorem P is sometimes called the Fisher–Cochran Theorem.)

Theorem Q. Let $X \sim N_p(0, I)$. Let $Q, Q_1, \ldots, Q_k$ be quadratic forms (in X) such that $Q = Q_1 + Q_2 + \ldots + Q_k$. Let Q be χ^2_r, let Q_i be $\chi^2_{r_i}$, $i = 1, 2, \ldots, k-1$, and let Q_k be non–negative. Then the random variables $Q_1, Q_2, \ldots, Q_k$ are mutually stochastically independent and, hence, Q_k is $\chi^2_{r_k}$ where $r_k = r - r_1 - r_2 - \ldots - r_{k-1}$.

Examples.

1. Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. So $X = (X_1, \ldots, X_n)' \sim N_n(m, \Sigma)$, where $m = (\mu, \mu, \ldots, \mu)'$ and $\Sigma = \sigma^2 I$. Define matrices P and A by

$$P = \begin{pmatrix} n^{-1} \\ n^{-1} \\ \vdots \\ n^{-1} \end{pmatrix} \quad \text{and} \quad A = \frac{1}{\sigma^2} \begin{pmatrix} 1-n^{-1} & -n^{-1} & \cdots & -n^{-1} \\ -n^{-1} & 1-n^{-1} & \cdots & -n^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ -n^{-1} & -n^{-1} & \cdots & 1-n^{-1} \end{pmatrix}$$

It is then easy to verify that $P'X = \bar{X}$ and $(X - m)'A(X - m) = X'AX = (n-1)S^2/\sigma^2$. Now observe that $\Sigma A \Sigma P = \sigma^4 AP = 0$, so by Theorem N, $\overline{X}$ and S^2 are independent. By Theorem B, $\overline{X} \sim N(\mu, \sigma^2/n)$. Finally it is clear that Σ is of full rank and that $A\Sigma A = A$, so by Theorem L, $(n-1)S^2/\sigma^2$ has a χ^2_k–distribution, where $k = trace(A\Sigma) = n - 1$. Thus,

$$\frac{\frac{(\overline{X}-\mu)}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{(n-1)S^2/\sigma^2}{n-1}}} = \frac{(\overline{X}-\mu)}{\sqrt{S^2/n}} \sim t_{n-1}.$$

2. Let $\varepsilon_1, \ldots, \varepsilon_n$ be a random sample from a normal distribution with mean 0 and variance σ^2. So $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)' \sim N_n(0, \Sigma)$, where $\Sigma = \sigma^2 I$. Let X denote a $n \times k$ matrix with rank k, and let $P_X = X(X'X)^{-1}X$ and $M_X = I_k - P_X$. Let $Q_1 = (1/\sigma^2)\varepsilon' P_X \varepsilon$ and $Q_2 = (1/\sigma^2)\varepsilon' M_X \varepsilon$. Then $Q_1 \sim \chi^2_k$, $Q_2 \sim \chi^2_{n-k}$ and Q_1 and Q_2 are independent. The result follows since P_X and M_X are idempotent, $\Sigma = \sigma^2 I$, and $P_X M_X = 0$.

1 Modes of Convergence

- In your first calculus class you discussed sequences $\{X_n\}$ and limits. Recall $\lim_{n\to\infty} X_n = X$ if for any $\varepsilon > 0\ \exists\ N(\varepsilon)$ with $|X_n - X| < \varepsilon$ for all $n > N(\varepsilon)$. We need to discuss convergence of random sequences $\{X_n(\omega)\}$ to random variables $X(\omega)$. There are a variety of notions of convergence:
  - For a given ω we can ask whether $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ using the standard definition of a limit. If the set of ω for which this limit obtains has probability 1 then we say $X_n(\omega)$ converges to $X(\omega)$ almost surely (or with probability 1).

    $$X_n(\omega) \stackrel{a.s.}{\to} X(\omega) \text{ if } P\{\omega | \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 1$$

  - For any $\varepsilon > 0$ we can calculate $p_n(\varepsilon) = P(|X_n - X| > \varepsilon)$. If for any value of $\varepsilon > 0$, this sequence converges to 0, then we say that X_n converges in probability to X.

    $$X_n \stackrel{p}{\to} X \text{ if for any } \varepsilon > 0,\ \lim_{n\to\infty} p_n(\varepsilon) = 0.$$

    This is sometimes written as $\text{plim} X_n = X$.
  - $|Y - Z|^p$ for $0 < p < \infty$ is called the L_p distance between Y and Z. X_n converges to X in the L_p norm if $\lim_{n\to\infty} \mathrm{E}(|X_n - X|^p) = 0$

    $$X_n \stackrel{L_p}{\to} X \text{ if } \lim_{n\to\infty} \mathrm{E}(|X_n - X|^p) = 0$$

    when $p = 2$, the convergence criterion is $\lim_{n\to\infty} \mathrm{E}(|X_n - X|^2) = 0$. In this case the convergence is sometimes called *mean square* convergence or *convergence in quadratic mean*, denoted as $X_n \stackrel{ms}{\to} X$ or $X_n \stackrel{qm}{\to} X$.
  - Suppose $F_{X_n}(x)$ is the CDF for X_n and $F_X(x)$ is the CDF for X. Then, in the limit X_n will have the same CDF as X if the function F_{X_n} converges to F_X. This notion of convergence is called *convergence in Distribution* or *convergence in Law.*

    $$X_n \stackrel{d}{\to} X \text{ if } \lim_{n\to\infty} F_{X_n}(x) = F_X(x) \text{ for all values of } x \text{ where } F_X(.) \text{ is continuous.}$$

  - If X_n is a vector, then $X_n \stackrel{as}{\to} X$ if each element of X_n converges *a.s.* the corresponding element of X. Convergence in probability and L_p convergence is defined analogously. $X_n \stackrel{d}{\to} X$ if the joint CDF of X_n converges the joint CDF of X.

1.1 Relationships between the modes of convergence and some useful results

- If $X_n \xrightarrow{L_p} X$ then $X_n \xrightarrow{p} X$.

Proof:
Without loss of generality set $X = 0$. Let $\varepsilon > 0$

$$\mathbf{E}(|X_n|^p) = \int_{-\infty}^{\infty} |X_n|^p dF_{X_n} \geq \int_{-\infty}^{-\varepsilon} |X_n|^p dF_{X_n} + \int_{\varepsilon}^{\infty} |X_n|^p dF_{X_n}$$

$$\geq |\varepsilon|^p [\int_{-\infty}^{-\varepsilon} dF_{X_n} + \int_{\varepsilon}^{\infty} dF_{X_n}]$$

$$= |\varepsilon|^p P(|X_n| \geq \varepsilon)$$

and thus

$$P(|X_n| \geq \varepsilon) \leq \frac{\mathbf{E}(|X_n|^p)}{|\varepsilon|^p}$$

so that $\lim_{n\to\infty} \mathbf{E}(|X_n|^p) = 0$ implies $\lim_{n\to\infty} P(|X_n| > \varepsilon) = 0$ for any $\varepsilon > 0$.

- - The result

    $$P(|X_n| \geq \varepsilon) \leq \frac{\mathbf{E}(|X_n|^p)}{|\varepsilon|^p}$$

    is known as *Markov's inequality.*

  - The result

    $$P(|X_n| \geq \varepsilon) \leq \frac{\mathbf{E}(|X_n|^2)}{|\varepsilon|^2}$$

    is known as *Chebyshev's inequality.*

- If $X_n \xrightarrow{a.s.} X$ then $X_n \xrightarrow{p} X$
  To prove this we need to show that for any $\varepsilon > 0$ and $\delta > 0$ $\exists$ $N(\varepsilon, \delta)$ such that $P(\omega \mid |X_n(\omega) - X(\omega)| > \varepsilon) < \delta$ for $n > N$. For each ω with $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ we can find a $N(\varepsilon, \omega)$ such that $|X_n(\omega) - X(\omega)| < \varepsilon$ for all $n > N(\varepsilon, \omega)$. Let $N(\varepsilon, \delta)$ be the smallest of these values such that $P(\omega \mid |X_n(\omega) - X(\omega)| < \varepsilon) > 1 - \delta$, for all $n > N(\varepsilon, \delta)$. (The existence of this value of N is guaranteed by the condition that $P\{\omega | \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 1$). Then $P(\omega \mid |X_n(\omega) - X(\omega)| > \varepsilon) < \delta$ for all $n > N$ as required.

- If $X_n \xrightarrow{p} X$ does not imply that $X_n \xrightarrow{a.s} X$.(See Amemiya, page 88 for a counterexample)

- If $X_n \xrightarrow{p} X$ then $X_n \xrightarrow{d} X$. (The proof is in Rao, page 122, result ix)

1.2 Slutsky's Theorem and the Continuous Mapping Theorem

- Slutsky's theorem (Rao page 122)
  - $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} 0$ implies $X_n Y_n \xrightarrow{p} 0$
  - Let c be a constant and suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$
    * $X_n + Y_n \xrightarrow{d} X + c$
    * $X_n Y_n \xrightarrow{d} Xc$
    * $X_n / Y_n \xrightarrow{d} X/c$, if $c \neq 0$
  - $X_n - Y_n \xrightarrow{p} 0$ and $Y_n \xrightarrow{d} Y$ then $X_n \xrightarrow{d} Y$
- Continuous Mapping Theorem (Rao, page 124)
  - Let $g(.)$ be a continuous function, then
    * $X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$
    * $X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$
    * $X_n - Y_n \xrightarrow{p} 0$ and $Y_n \xrightarrow{d} Y$ then $g(X_n) - g(Y_n) \xrightarrow{p} 0$

1.3 O_p and o_p Notation

Let $\{a_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ denote two sequences of real numbers. Recall (from your Calculus class) that

- $$a_n = o(g_n) \text{ if } \lim_{n\to\infty} \frac{a_n}{g_n} = 0$$

  and
  $$a_n = O(g_n) \text{ if } \exists \text{ a number } M \text{ such that } \frac{a_n}{g_n} < M \text{ for all } n$$

  We use similar notation for random variables. Suppose now that $\{a_n\}_{n=1}^{\infty}$ is a sequence of random variables, then
  $$a_n = o_p(g_n) \text{ if } \frac{a_n}{g_n} \xrightarrow{p} 0$$

  and

  $$a_n = O_p(g_n) \text{ if for any } \varepsilon > 0, \exists \text{ a number } M \text{ such that } P(|\frac{a_n}{g_n}| < M) > 1-\varepsilon \text{ for all } n.$$

- Let $\{f_n\}$ and $\{g_n\}$ be sequences of real numbers and let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables, then
  - If $X_n = o_p(f_n)$ and $Y_n = o_p(g_n)$, then

  - $X_nY_n = o_p(f_ng_n)$
  - $|X_n|^s = o_p(f_n^s)$ for $s > 0$
  - $X_n + Y_n = o_p(\max\{f_n, g_n\})$
- If $X_n = O_p(f_n)$ and $Y_n = O_p(g_n)$, then
  - $X_nY_n = O_p(f_ng_n)$
  - $|X_n|^s = O_p(f_n^s)$ for $s > 0$
  - $X_n + Y_n = O_p(\max\{f_n, g_n\})$
- If $X_n = o_p(f_n)$ and $Y_n = O_p(g_n)$, then
  - $X_nY_n = o_p(f_ng_n)$

2 Laws of Large Numbers

2.1 A Weak Law of Large Numbers

- Let $X_1, X_2, ...$ be a sequence of random variables with $\mathrm{E}(X_i) = \mu_i$ and $Var(X_i) = \sigma_i^2$ and $Cov(X_i, X_j) = 0$ for $i \neq j$. Let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$, $\overline{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \sigma_i^2$ and $\overline{\mu} = n^{-1}\sum_{i=1}^n \mu_i$ with $\lim_{n\to\infty} n^{-1}\overline{\sigma}_n^2 = 0$. Then $\overline{X}_n - \overline{\mu} \xrightarrow{p} 0$.
  Proof:
$$P(|\overline{X}_n - \overline{\mu}| > \varepsilon) \leq \frac{\mathrm{E}[(\overline{X}_n - \overline{\mu})^2]}{\varepsilon^2} = \frac{n^{-1}\overline{\sigma}_n^2}{\varepsilon^2} \to 0$$
  where the first inequality follows from Chebyshev's inequality

2.2 A Strong Law of Large Number

- If $X_1, X_2, ...$ are with $\mathrm{E}(X) = \mu < \infty$, then $\overline{X}_n \xrightarrow{a.s.} \mu$. (Proof: Rao, pages 114-115)

3 Central Limit Theorems

3.1 Characteristic Function (Rao, page 99-108)

Consider a random variable X with CDF $F(x)$. The characteristic function of X, denoted $C(t)$ is given by
$$C(t) = \mathrm{E}(e^{itX}) = \int e^{itx}dF(x)$$
where $i = \sqrt{-1}$. Thus, This change is useful because
$$e^{iz} = \cos(z) + i\sin(z)$$
so that $|e^{iz}| = 1$ for all z. This means that $C(t)$ will always exists, while $M(t)$ exists only for certain distributions.

Some useful results:

1. Let $\alpha_r = \mathrm{E}(X^r)$, which is assumed to exist. Then

$$\frac{d^r C(t)}{dt^r} = i^r \int x^r e^{itx} dF(x) \text{ exists}$$

2. Suppose α_r exists, then expanding $C(t)$ in a Taylor Series expansion about $C(0)$ yields:

$$C(t) = C(0) + \sum_{j=1}^{r} \alpha_j [\frac{(it)^j}{j!}] + O(t^{r+1})$$

   and $C(0) = 1$

3. Let $\phi(t) = \ln(C(t))$, then if α_r exists

$$\phi(t) = \sum_{j=0}^{r} \kappa_j [\frac{(it)^j}{j!}] + O(t^{r+1})$$

   where κ_j is called the $k'th$ cummulant. A direct calculation shows

   (a) $\kappa_1 = \alpha_1 = \mu$
   (b) $\kappa_2 = \alpha_2 - \alpha_1^2 = \sigma^2$

4. If $X \sim N(\mu, \sigma^2)$, then $C(t) = M(it) = \exp[it\mu - \frac{t^2\sigma^2}{2}]$, and thus $\kappa_0 = 0$, $\kappa_1 = \mu$, $\kappa_2 = \sigma^2$, $\kappa_j = 0$ for $j > 2$.

5. Let $Z = X + Y$, where X and Y are independent, then $C_Z(t) = C_X(t)C_Y(t)$

6. Let Z=δX, where δ is a constant. Then $C_Z(t) = C_X(\delta t)$

7. There is a 1-to-1 relation between $F(x)$ and $C(t)$

8. Let $C_n(t)$ denote the CF of X_n and $C(t)$ denote the CF of X. If $X_n \xrightarrow{d} X$, then $C_n(t) \rightarrow C(t)$ for all t. Moreover, if $C_n(t) \rightarrow C(t)$ for all t and if $C(t)$ is continuous at $t = 0$, then $X_n \xrightarrow{d} X$.

3.2 A Central Limit Theorem

(Lindberg-Levy CLT): Let $X_1, X_2, \ldots$ denote a sequence of iid random variables with $E(X_i) = \mu$ and $var(X_i) = \sigma^2 \neq 0$. Let $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Then

$$\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu) \xrightarrow{d} N(0,1)$$

Proof:
Let $Z_n = \frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu) = \sum_{i=1}^{n}(\frac{X_i - \mu}{\sqrt{n}\sigma})$. Since $\mathrm{E}(\frac{X_i-\mu}{\sigma}) = 0$ and $var(\frac{X_i-\mu}{\sigma}) = 1$,the log-CF of $\frac{X_i-\mu}{\sigma}$ is

$$\phi(t) = -\frac{1}{2}t^2 + O(t^3)$$

so that Z_n has log-CF

$$\phi_{Z_n}(t) = \sum_{i=1}^{n}\{-(\frac{1}{2})(\frac{t}{\sqrt{n}})^2 + O[(\frac{t}{\sqrt{n}})^3]$$
$$= -\frac{1}{2}t^2 + n \times O(\frac{t^3}{n^{1/2}}) \to -\frac{1}{2}t^2$$

which is the log-CF of a $N(0,1)$ random variable. Since $-\frac{1}{2}t^2$ is continuous at $t = 0$, $Z_n \xrightarrow{d} Z \sim N(0,1)$.

- Example: Suppose that X_i are *iid* Bernoulli random variables with parameter p. The CLT says that

$$\sqrt{n}\frac{(\overline{X} - p)}{(p(1-p))^{1/2}} \xrightarrow{d} N(0,1)$$

which implies that for large n

$$\sqrt{n}\frac{(\overline{X} - p)}{(p(1-p))^{1/2}} \overset{a}{\sim} N(0,1)$$

where "$\overset{a}{\sim}$" means "approximately distributed as". Thus

$$\overline{X} \overset{a}{\sim} N(p, \frac{p(1-p)}{n})$$

Suppose $p = .25$ and $n = 100$ then $P(\overline{X} \leq .20)$ is $P(Y \leq 20)$ where $Y = \sum_{i=1}^{100} X_i$ is distributed binomial with $n = 100$ and $p = .25$. A direct calculation shows: $P(Y \leq 20) = .14$. The normal approximation gives

$$P(\overline{X} \leq .20) \overset{a}{=} P(\frac{\overline{X} - .25}{(\frac{.25\times.75}{100})^{1/2}} \leq \frac{.20 - .25}{(\frac{.25\times.75}{100})^{1/2}})$$
$$= P(Z \leq -1.155) = .12;$$

3.3 A Multivariate Central Limit Theorem (Rao, p. 128)

Let X_n denote a sequence of $k \times 1$ random vectors. Let X denote a $k \times 1$ random vector with $X \sim N_k(0, \Sigma)$, and λ denote a $k \times 1$ vector of constants. Then a necessary and sufficient condition for $X_n \xrightarrow{d} X$ is that $\lambda' X_n \xrightarrow{d} \lambda' X$ for all λ.

3.4 The Delta Method

Let Y_n denote a sequence of random variables, and let $X_n = \sqrt{n}(Y_n - a)$, where a is a constant. Let $g(.)$ be a continuously differentiable function. Suppose $X_n \xrightarrow{d} X \sim N(0, \sigma^2)$. Then $\sqrt{n}(g(Y_n) - g(a)) \xrightarrow{d} V \sim N(0, [g'(a)]^2\sigma^2)$.

Proof: By the mean value theorem

$$g(Y_n) = g(a) + (Y_n - a)g'(\widetilde{Y}_n)$$