Exam 2, Thursday, March 24, 15 sheets of notes and a calculator,

You should know most of the material on Exam 1 review.

27) If $n^\delta(T_{1,n} - \theta) \xrightarrow{D} N(0, \sigma_1^2(F))$ and $n^\delta(T_{2,n} - \theta) \xrightarrow{D} N(0, \sigma_2^2(F))$, then the **asymptotic relative efficiency** of $T_{1,n}$ with respect to $T_{2,n}$ is

$$ARE(T_{1,n}, T_{2,n}) = \frac{\sigma_2^2(F)}{\sigma_1^2(F)}.$$

The "better" estimator has the smaller $\sigma_i^2(F)$. Here $0 < \delta \leq 1$.

28) An estimator T_n of $\tau(\theta)$ is **asymptotically efficient** if

$$\sqrt{n}(T_n - \tau(\theta)) \xrightarrow{D} N\left(0, \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right).$$

29) For a 1P–REF, $\frac{1}{n}\sum_{i=1}^n t(Y_i)$ is an asymptotically efficient estimator of $g(\eta) = \mu_t = E(t(Y))$.

30) If $\hat{\theta}_n$ is the MLE or UMVUE of θ, then $T_n = \tau(\hat{\theta}_n)$ is an asymptotically efficient estimator of $\tau(\theta)$ when the Standard Limit Theorem holds. Hence if $\tau'(\theta) \neq 0$, then

$$\sqrt{n}[\tau(\hat{\theta}_n) - \tau(\theta)] \xrightarrow{D} N\left(0, \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right).$$

31) Theorem: Let X_θ be a random variable with a distribution depending on θ, and $0 < \delta \leq 1$. Suppose

$$n^\delta(T_n - \tau(\theta)) \xrightarrow{D} X_\theta$$

for all $\theta \in \Theta$, then T_n is a consistent estimator of $\tau(\theta)$. If the convergence holds for a fixed θ, then $T_n \xrightarrow{P} \theta$.

Note: Often $X_\theta \sim N(0, v(\theta))$ and $\delta = 0.5$.

32) **Slutsky's Theorem**: Suppose $Y_n \xrightarrow{D} Y$ and $W_n \xrightarrow{P} c$ for some constant c. Then

a) $Y_n + W_n \xrightarrow{D} Y + c$,

b) $Y_n W_n \xrightarrow{D} cY$, and

c) $Y_n/W_n \xrightarrow{D} Y/c$ if $c \neq 0$.

Note that $Y_n \xrightarrow{B} Y$ implies $Y_n \xrightarrow{D} Y$ where $B = wp1, r,$ or P. Also $W_n \xrightarrow{P} c$ iff $W_n \xrightarrow{D} c$. If a sequence of constants $c_n \to c$ as $n \to \infty$ (everywhere convergence), then $c_n \xrightarrow{wp1} c$ and $c_n \xrightarrow{P} c$. i) So 32) a), b) and c) hold if $Y_n \xrightarrow{B} Y$. ii) If $Y = d$ where d is a constant, then $Y_n \xrightarrow{B} d$ implies that 32) a), b) and c) hold with Y replaced by d, and $\xrightarrow{D}$ can be replaced by $\xrightarrow{P}$. iii) In i) and ii), $W_n \xrightarrow{P} c$ can be replaced by $W_n \xrightarrow{A} c$ where $A = D, wp1, r,$ or P.

33) Let $g : \mathbb{R} \to \mathbb{R}$ be continuous at c.

a) If $X_n \xrightarrow{D} c$, then $g(X_n) \xrightarrow{D} c$.

b) If $X_n \xrightarrow{P} c$, then $g(X_n) \xrightarrow{P} c$.

c) If $X_n \xrightarrow{wp1} c$, then $g(X_n) \xrightarrow{wp1} c$.

Note: If $X_n \xrightarrow{r} c$, then $X_n \xrightarrow{P} c$ and $g(X_n) \xrightarrow{P} c$.

34) **Continuity Theorem:** Let Y_n be sequence of random variables with characteristic functions $c_{Y_n}(t)$. Let Y be a random variable with cf $c_Y(t)$.

a)

$$Y_n \xrightarrow{D} Y \quad \text{iff} \quad \mathrm{c_{Y_n}(t) \to c_Y(t)\ \forall t \in \mathbb{R}}.$$

b) Also assume that Y_n has mgf m_{Y_n} and Y has mgf m_Y. Assume that all of the mgfs m_{Y_n} and m_Y are defined on $|t| \leq d$ for some $d > 0$. Then if $m_{Y_n}(t) \to m_Y(t)$ as $n \to \infty$ for all $|t| < c$ where $0 < c < d$, then $Y_n \xrightarrow{D} Y$.

35) Theorem: If $\lim_{n\to\infty} c_{X_n}(t) = g(t)$ for all t where g is continuous at $t = 0$, then $g(t) = c_X(t)$ is a characteristic function for some RV X, and $X_n \xrightarrow{D} X$.

Note: Hence continuity at $t = 0$ implies continuity everywhere since $g(t) = \varphi_X(t)$ is continuous. If $g(t)$ is not continuous at 0, then X_n does not converge in distribution.

36) If $c_{Y_n}(t) \to h(t)$ where $h(t)$ is not continuous, then Y_n does not converge in distribution to any RV Y, by the Continuity Theorem and 35).

37) Let $X_1, ..., X_n$ be independent RVs with characteristic functions $c_{X_j}(t)$. Then the characteristic function of $\sum_{j=1}^n X_j$ is $c_{\sum_{j=1}^n X_j}(t) = \displaystyle\prod_{j=1}^n c_{X_j}(t)$. If the RVs also have mgfs $m_{X_j}(t)$, then the mgf of $\sum_{j=1}^n X_j$ is $m_{\sum_{j=1}^n X_j}(t) = \displaystyle\prod_{j=1}^n m_{X_j}(t)$.

38) **Helly-Bray-Pormanteau Theorem**: $X_n \xrightarrow{D} X$ iff $E[g(X_n)] \to E[g(X)]$ for every bounded, real, continuous function g.

Note: 38) is used to prove 39 b).

39) a) **Generalized Continuous Mapping Theorem:** If $X_n \xrightarrow{D} X$ and the function g is such that $P[X \in C(g)] = 1$ where $C(g)$ is the set of points where g is continuous, then $g(X_n) \xrightarrow{D} g(X)$.

Note: $P[X \in C(g)] = 1$ can be replaced by $P[X \in D(g)] = 0$ where $D(g)$ is the set of points where g is not continuous.

b) **Continuous Mapping Theorem:** If $X_n \xrightarrow{D} X$ and the function g is continuous, then $g(X_n) \xrightarrow{D} g(X)$.

Note: the function g can not depend on n since g_n is a sequence of functions rather than a single function.

40) Let X_n have pdf $f_{X_n}(x)$, and let X have pdf $f_X(x)$. If $f_{X_n}(x) \to f_X(x)$ for all x (or for x outside of a set of Lebesgue measure 0), then $X_n \xrightarrow{D} X$.

41) Suppose X_n and X are integer valued RVs with pmfs $f_{X_n}(x)$ and $f_X(x)$. Then $X_n \xrightarrow{D} X$ iff $P(X_n = k) \to P(X = k)$ for every integer k iff $f_{X_n}(x) \to f_X(x)$ for every real x.

Extensions to Random Vectors

42) Let $\boldsymbol{X} = (X_1, ..., X_k)^T \in \mathbb{R}^k$ be a $k \times 1$ **column vector**.

$$E(\boldsymbol{X}) = (E(X_1), ..., E(X_k))^T$$

and the $k \times k$ *covariance matrix*

$$\mathrm{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma} = E[(\boldsymbol{X} - E[\boldsymbol{X}])(\boldsymbol{X} - E[\boldsymbol{X}])^T] = (\sigma_{ij}).$$

That is, the ij entry of $\mathrm{Cov}(\boldsymbol{X})$ is $\mathrm{Cov}(X_i, X_j) = \sigma_{ij}$.

43) If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are $k \times 1$ random vectors, $\boldsymbol{a}$ a conformable constant vector, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are conformable constant matrices, then
$E(\boldsymbol{a} + \boldsymbol{X}) = \boldsymbol{a} + E(\boldsymbol{X})$ and $\mathrm{E}(\boldsymbol{X} + \boldsymbol{Y}) = \mathrm{E}(\boldsymbol{X}) + \mathrm{E}(\boldsymbol{Y})$ and
$E(\boldsymbol{AX}) = \boldsymbol{A}E(\boldsymbol{X})$ and $\mathrm{E}(\boldsymbol{AXB}) = \boldsymbol{A}\mathrm{E}(\boldsymbol{X})\boldsymbol{B}$. Thus
$\mathrm{Cov}(\boldsymbol{a} + \boldsymbol{AX}) = \mathrm{Cov}(\boldsymbol{AX}) = \boldsymbol{A}\mathrm{Cov}(\boldsymbol{X})\boldsymbol{A}^T$.

44) The **characteristic function** of $\boldsymbol{X}$ is $c_{\boldsymbol{X}}(\boldsymbol{t}) = E[e^{i\boldsymbol{t}^T\boldsymbol{X}}]$.

The moment generating function of $\boldsymbol{X}$ is $m_{\boldsymbol{X}}(\boldsymbol{t}) = E[e^{\boldsymbol{t}^T\boldsymbol{X}}]$ provided the expectation exists for all $\boldsymbol{t}$ in a neighborhood of $\boldsymbol{0}$.

The cumulative distribution function (cdf) $F_{\boldsymbol{X}}(\boldsymbol{x}) = P(X_1 \le x_1, ..., X_k \le x_k)$.

45) Let the Euclidean norm $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^T\boldsymbol{x}} = \sqrt{x_1^2 + \cdots + x_k^2}$.

46) Let $\boldsymbol{X}_n \in \mathbb{R}^k$ be a sequence of random vectors with joint cdfs $F_{\boldsymbol{X}_n}(\boldsymbol{x})$ and let $\boldsymbol{X} \in \mathbb{R}^k$ be a random vector with joint cdf $F_{\boldsymbol{X}}(\boldsymbol{x})$.

a) $\boldsymbol{X}_n$ **converges in distribution** to $\boldsymbol{X}$, written $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$, if $F_{\boldsymbol{X}_n}(\boldsymbol{x}) \to F_{\boldsymbol{X}}(\boldsymbol{x})$ as $n \to \infty$ for all points $\boldsymbol{x}$ at which $F_{\boldsymbol{X}}(\boldsymbol{x})$ is continuous. The distribution of $\boldsymbol{X}$ is the **limiting distribution** or **asymptotic distribution** of $\boldsymbol{X}_n$, and the limiting distribution does not depend on n.

b) $\boldsymbol{X}_n$ **converges in probability** to $\boldsymbol{X}$, written $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X}$, if for every $\epsilon > 0$, $P(\|\boldsymbol{X}_n - \boldsymbol{X}\| > \epsilon) \to 0$ as $n \to \infty$.

c) Let $r > 0$ be a real number. Then $\boldsymbol{X}_n$ **converges in** r**th mean** to $\boldsymbol{X}$, written $\boldsymbol{X}_n \xrightarrow{r} \boldsymbol{X}$, if $E(\|\boldsymbol{X}_n - \boldsymbol{X}\|^r) \to 0$ as $n \to \infty$.

d) $\boldsymbol{X}_n$ **converges with probability one** to $\boldsymbol{X}$, written $\boldsymbol{X}_n \xrightarrow{wp1} \boldsymbol{X}$, if
$P(\lim_{n\to\infty} \boldsymbol{X}_n = \boldsymbol{X}) = 1$.

e) Replace $\boldsymbol{X}$ by $\boldsymbol{c}$ for $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{c}$, $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{c}$, $\boldsymbol{X}_n \xrightarrow{r} \boldsymbol{c}$, or $\boldsymbol{X}_n \xrightarrow{wp1} \boldsymbol{c}$.

f) $\xrightarrow{D} = \xrightarrow{L}$ and $\boldsymbol{X}_n \xrightarrow{wp1} \boldsymbol{X} = \boldsymbol{X}_n \xrightarrow{as} \boldsymbol{X} = \boldsymbol{X}_n \xrightarrow{ae} \boldsymbol{X}$.

47)) **Generalized Chebyshev's Inequality** = *Generalized Markov's Inequality*: Let $u : \mathbb{R}^k \to [0, \infty)$ be a nonnegative function. If $E[u(\boldsymbol{X})]$ exists, then for any $\epsilon > 0$,

$$P[u(\boldsymbol{X}) \ge \epsilon] \le \frac{E[u(\boldsymbol{X})]}{\epsilon}.$$

48) Let $u(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{c}\|^r$ for some $r > 0$. Often $\boldsymbol{c} = \boldsymbol{0}$ or $\boldsymbol{a} = E(\boldsymbol{X}) = \boldsymbol{\mu}$. If $E[u(\boldsymbol{X})]$ exists, then for any $\epsilon > 0$,

$$P(\|\boldsymbol{X} - \boldsymbol{c}\| \ge \epsilon] = P(\|\boldsymbol{X} - \boldsymbol{c}\|^r \ge \epsilon^r] \le \frac{E[\|\boldsymbol{X} - \boldsymbol{c}\|^r]}{\epsilon^r}.$$

49) A $k \times 1$ random vector $\boldsymbol{X}$ has a $k-$dimensional *multivariate normal (MVN) distribution* $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ iff $\boldsymbol{t}^T\boldsymbol{X}$ has a univariate normal distribution for any $k \times 1$ constant vector $\boldsymbol{t}$. Then $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $\mathrm{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$. Note that $\boldsymbol{t}^T\boldsymbol{X} \sim N(\boldsymbol{t}^T\boldsymbol{\mu}, \boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t})$, a univariate normal distribution. A univariate normal distribution is a special case of a MVN distribution with $k = 1$.

50) Let $\boldsymbol{A}$ be a $q \times k$ constant matrix, b a constant, $\boldsymbol{a}$ a $k \times 1$ constant vector, and $\boldsymbol{d}$ a $q \times 1$ constant vector.

A) Suppose $\boldsymbol{X} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then
i) $\boldsymbol{AX} \sim N_q(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^T)$.
ii) $\boldsymbol{a} + b\boldsymbol{X} \sim N_k(\boldsymbol{a} + b\boldsymbol{\mu}, b^2\boldsymbol{\Sigma})$.
iii) $\boldsymbol{AX} + \boldsymbol{d} \sim N_q(\boldsymbol{A\mu} + \boldsymbol{d}, \boldsymbol{A\Sigma A}^T)$.
(Find the mean and covariance matrix of the left hand side and plug in those values for the right hand side. **Be careful with the dimension** k or q.)

B) Suppose $\boldsymbol{X}_n \xrightarrow{D} N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then
i) $\boldsymbol{AX}_n \xrightarrow{D} N_q(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^T)$.
ii) $\boldsymbol{a} + b\boldsymbol{X}_n \xrightarrow{D} N_k(\boldsymbol{a} + b\boldsymbol{\mu}, b^2\boldsymbol{\Sigma})$.
iii) $\boldsymbol{AX}_n + \boldsymbol{d} \xrightarrow{D} N_q(\boldsymbol{A\mu} + \boldsymbol{d}, \boldsymbol{A\Sigma A}^T)$.
(The behavior of convergence in distribution to a MVN distribution is much like the behavior of the MVN distributions in A).)

51) The **Multivariate Central Limit Theorem (MCLT)**: If $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ are iid $k \times 1$ random vectors with $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $\text{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$, then

$$\sqrt{n}(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$$

where the sample mean

$$\overline{\boldsymbol{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i.$$

Note: the usual CLT is a special case with $k = 1$.

52) If $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ are iid, $E(\|\boldsymbol{X}\|) < \infty$, and $E(\boldsymbol{X}) = \boldsymbol{\mu}$, then
a) WLLN: $\overline{\boldsymbol{X}}_n \xrightarrow{P} \boldsymbol{\mu}$, and
b) SLLN: $\overline{\boldsymbol{X}}_n \xrightarrow{wp1} \boldsymbol{\mu}$.

53) **Continuity Theorem**: Let $\boldsymbol{X}_n$ be a sequence of $k \times 1$ random vectors with characteristic functions $c_{\boldsymbol{X}_n}(\boldsymbol{t})$, and let $\boldsymbol{X}$ be a $k \times 1$ random vector with cf $c_{\boldsymbol{X}}(\boldsymbol{t})$. Then

$$\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X} \text{ iff } \mathrm{c}_{\boldsymbol{X}_\mathrm{n}}(\boldsymbol{t}) \to \mathrm{c}_{\boldsymbol{X}}(\boldsymbol{t})$$

for all $\boldsymbol{t} \in \mathbb{R}^k$.

54) **Theorem: Cramér Wold Device**: Let $\boldsymbol{X}_n$ be a sequence of $k \times 1$ random vectors, and let $\boldsymbol{X}$ be a $k \times 1$ random vector. Then

$$\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X} \text{ iff } \boldsymbol{t}^\mathrm{T}\boldsymbol{X}_\mathrm{n} \xrightarrow{\mathrm{D}} \boldsymbol{t}^\mathrm{T}\boldsymbol{X}$$

for all $\boldsymbol{t} \in \mathbb{R}^k$.

55) Use 53) and 54) to prove the MCLT. Use 53) to prove 54).

56) **Theorem.** a) If $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X}$, then $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$.
b)

$$\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{c} \text{ iff } \boldsymbol{X}_\mathrm{n} \xrightarrow{\mathrm{D}} \boldsymbol{c}.$$

57) **Continuous Mapping Theorem.** Let $\boldsymbol{X}, \boldsymbol{X}_n \in \mathbb{R}^k$. If $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$ and if the function $\boldsymbol{g} : \mathbb{R}^k \to \mathbb{R}^j$ is continuous, then $\boldsymbol{g}(\boldsymbol{X}_n) \xrightarrow{D} \boldsymbol{g}(\boldsymbol{X})$.

This theorem also holds if $C(\boldsymbol{g})$ is the set of points $\boldsymbol{x}$ for which $\boldsymbol{g}$ is continuous and $P(\boldsymbol{X} \in C(\boldsymbol{g})) = 1$. (Equivalently, $D(\boldsymbol{g})$ is the set of discontinuity points for $\boldsymbol{g}$ and $P(\boldsymbol{X} \in D(\boldsymbol{g})) = 0$.)

58) **Theorem**: Let $\boldsymbol{X}_n = (X_{1n}, ..., X_{kn})^T$ be a sequence of $k \times 1$ random vectors, let $\boldsymbol{Y}_n$ be a sequence of $k \times 1$ random vectors, and let $\boldsymbol{X} = (X_1, ..., X_k)^T$ be a $k \times 1$ random vector. Let $\boldsymbol{W}_n$ be a sequence of $k \times k$ nonsingular random matrices, and let $\boldsymbol{C}$ be a $k \times k$ constant nonsingular matrix.

a) $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X}$ iff $X_{in} \xrightarrow{P} X_i$ for $i = 1, ..., k$.

b) **Slutsky's Theorem**: If $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$ and $\boldsymbol{Y}_n \xrightarrow{P} \boldsymbol{a}$ for some constant $k \times 1$ vector $\boldsymbol{a}$, then i) $\boldsymbol{X}_n + \boldsymbol{Y}_n \xrightarrow{D} \boldsymbol{X} + \boldsymbol{a}$ and

ii) $\boldsymbol{Y}_n^T \boldsymbol{X}_n \xrightarrow{D} \boldsymbol{a}^T \boldsymbol{X}$.

c) If $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$ and $\boldsymbol{W}_n \xrightarrow{P} \boldsymbol{C}$, then $\boldsymbol{W}_n \boldsymbol{X}_n \xrightarrow{D} \boldsymbol{C}\boldsymbol{X}$, $\boldsymbol{X}_n^T \boldsymbol{W}_n \xrightarrow{D} \boldsymbol{X}^T \boldsymbol{C}$, $\boldsymbol{W}_n^{-1} \boldsymbol{X}_n \xrightarrow{D} \boldsymbol{C}^{-1} \boldsymbol{X}$, and $\boldsymbol{X}_n^T \boldsymbol{W}_n^{-1} \xrightarrow{D} \boldsymbol{X}^T \boldsymbol{C}^{-1}$.

59) If $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$, then $X_{in} \xrightarrow{D} X_i$ for $i = 1, ..., k$.

60) In general, $\boldsymbol{X}_{in} \xrightarrow{D} \boldsymbol{X}_i$ for $i = 1, ..., m$ **does not imply that**

$$\begin{bmatrix} \boldsymbol{X}_{1n} \\ \vdots \\ \boldsymbol{X}_{mn} \end{bmatrix} \xrightarrow{D} \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_m \end{bmatrix}.$$

That is, marginal convergence in distribution does not imply joint convergence in distribution.

61) Suppose that $\boldsymbol{X}_n \perp\!\!\!\perp \boldsymbol{Y}_n$ for $n = 1, 2, ...$ Suppose $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$, and $\boldsymbol{Y}_n \xrightarrow{D} \boldsymbol{Y}$. Then

$$\begin{bmatrix} \boldsymbol{X}_n \\ \boldsymbol{Y}_n \end{bmatrix} \xrightarrow{D} \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix}$$

where $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y}$.

If the sequence $\{\boldsymbol{X}_n\} \perp\!\!\!\perp \{\boldsymbol{Y}_n\}$ so that $\boldsymbol{X}_i \perp\!\!\!\perp \boldsymbol{Y}_j$ for every i and j, then we should have $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y}$ even if $\boldsymbol{X} = \boldsymbol{c} = \boldsymbol{Y}$. Roughly, independence is an exception to 60) since independent random vectors have a joint distribution that does not affect the marginal distributions.

62) Theorem: Let

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \in \mathbb{R}^k$$

with $\boldsymbol{X}_1 \in \mathbb{R}^{k_1}$ and $\boldsymbol{X}_2 \in \mathbb{R}^{k_2}$ where $k_1 + k_2 = k$. Let $c_{\boldsymbol{X}}, c_{\boldsymbol{X}_1}$, and $c_{\boldsymbol{X}_2}$ be the characteristic functions of $\boldsymbol{X}, \boldsymbol{X}_1$ and $\boldsymbol{X}_2$. Then $\boldsymbol{X}_1 \perp\!\!\!\perp \boldsymbol{X}_2$ iff

$$c_{\boldsymbol{X}}(\boldsymbol{t}) = c_{\boldsymbol{X}_1}(\boldsymbol{t}_1) c_{\boldsymbol{X}_2}(\boldsymbol{t}_2) \ \ \forall \boldsymbol{t} = \begin{bmatrix} \boldsymbol{t}_1 \\ \boldsymbol{t}_2 \end{bmatrix} \in \mathbb{R}^k.$$

63) If $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$ and $\boldsymbol{Y}_n \xrightarrow{D} \boldsymbol{c}$, a constant vector, then

$$\begin{bmatrix} \boldsymbol{X}_n \\ \boldsymbol{Y}_n \end{bmatrix} \xrightarrow{D} \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{c} \end{bmatrix}.$$

64) Theorem:
i) $\boldsymbol{X}_n \stackrel{wp1}{\rightarrow} \boldsymbol{X} \Rightarrow \boldsymbol{X}_n \stackrel{P}{\rightarrow} \boldsymbol{X}$.
ii) $\boldsymbol{X}_n \stackrel{r}{\rightarrow} \boldsymbol{X} \Rightarrow \boldsymbol{X}_n \stackrel{P}{\rightarrow} \boldsymbol{X}$.
iii) $\boldsymbol{X}_n \stackrel{P}{\rightarrow} \boldsymbol{X} \Rightarrow \boldsymbol{X}_n \stackrel{D}{\rightarrow} \boldsymbol{X}$.
iv) $\boldsymbol{X}_n \stackrel{P}{\rightarrow} \boldsymbol{c}$ iff $\boldsymbol{X}_n \stackrel{D}{\rightarrow} \boldsymbol{c}$.

65) **Multivariate Delta Method**: If $\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\theta}) \stackrel{D}{\rightarrow} N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$, then

$$\sqrt{n}(\boldsymbol{g}(\boldsymbol{T}_n) - \boldsymbol{g}(\boldsymbol{\theta})) \stackrel{D}{\rightarrow} N_d(\boldsymbol{0}, \boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})} \boldsymbol{\Sigma} \boldsymbol{D}^T_{\boldsymbol{g}(\boldsymbol{\theta})})$$

if $\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})} \boldsymbol{\Sigma} \boldsymbol{D}^T_{\boldsymbol{g}(\boldsymbol{\theta})}$ is nonsingular, where the $d \times k$ Jacobian matrix of partial derivatives

$$\boldsymbol{D}_{\boldsymbol{g}(\boldsymbol{\theta})} = \begin{bmatrix} \frac{\partial}{\partial \theta_1} g_1(\boldsymbol{\theta}) & \dots & \frac{\partial}{\partial \theta_k} g_1(\boldsymbol{\theta}) \\ \vdots & & \vdots \\ \frac{\partial}{\partial \theta_1} g_d(\boldsymbol{\theta}) & \dots & \frac{\partial}{\partial \theta_k} g_d(\boldsymbol{\theta}) \end{bmatrix}.$$

Here the mapping $\boldsymbol{g} : \mathbb{R}^k \rightarrow \mathbb{R}^d$ needs to be differentiable in a neighborhood of $\boldsymbol{\theta} \in \mathbb{R}^k$.

66) The smallest integer function $\lceil x \rceil$ rounds up, eg $\lceil 7.7 \rceil = 8$. The *sample α quantile* $\hat{\xi}_{n,\alpha} = Y_{(\lceil n\alpha \rceil)}$. The *population quantile* $\xi_\alpha = Q(\alpha) = \inf\{y : F(y) \geq \alpha\}$. If F is continuous at ξ_α, then $F(\xi_\alpha) = \alpha$. The 0.5 quantile $\hat{\xi}_{n,0.5}$ is asymptotically equivalent to the sample median. The α quantile is the 100αth percentile. So a sample quantile is a sample percentile.

67) **Quantile Theorem.** Let $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_k < 1$. Suppose that F has a density f that is positive and continuous in neighborhoods of $\xi_{\alpha_1}, ..., \xi_{\alpha_k}$. Then

$$\sqrt{n}[(\hat{\xi}_{n,\alpha_1}, ..., \hat{\xi}_{n,\alpha_k})^T - (\xi_{\alpha_1}, ..., \xi_{\alpha_k})^T] \stackrel{D}{\rightarrow} N_k(\boldsymbol{0}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma} = (\sigma_{ij})$ and

$$\sigma_{ij} = \frac{\alpha_i(1 - \alpha_j)}{f(\xi_{\alpha_i}) f(\xi_{\alpha_j})}$$

for $i \leq j$ and $\sigma_{ij} = \sigma_{ji}$ for $i > j$.

68) For each positive integer n, let $W_{n1}, ..., W_{nr_n}$ be independent. The probability space may change with n, giving a double array of RVs. Let $E[W_{nk}] = 0$, $V(W_{nk}) = E[W_{nk}^2] = \sigma^2_{nk}$, and $s_n^2 = \sum_{k=1}^{r_n} \sigma^2_{nk} = V[\sum_{k=1}^{r_n} W_{nk}]$. Then

$$Z_n = \frac{\sum_{k=1}^{r_n} W_{nk}}{s_n}$$

is the z-score of $\sum_{k=1}^{r_n} W_{nk}$.

69) **Lyapounov's CLT**: Under 68), assume the $|W_{nk}|^{2+\delta}$ are integrable for some $\delta > 0$. Assume Lyapounov's condition:

$$\lim_{n\rightarrow\infty} \sum_{k=1}^{r_n} \frac{E[|W_{nk}|^{2+\delta}]}{s_n^{2+\delta}} = 0.$$

Then

$$Z_n = \frac{\sum_{k=1}^{r_n} W_{nk}}{s_n} \xrightarrow{D} N(0,1).$$

70) Special cases: i) $r_n = n$ and $W_{nk} = W_k$ has $W_1, ..., W_n, ...$ independent.
ii) $W_{nk} = X_{nk} - E(X_{nk}) = X_{nk} - \mu_{nk}$ has

$$\frac{\sum_{k=1}^{r_n}(X_{nk} - \mu_{nk})}{s_n} \xrightarrow{D} N(0,1).$$

iii) Suppose $X_1, X_2, ...$ are independent with $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$. Let

$$Z_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}}$$

be the z-score of $\sum_{i=1}^n X_i$. Assume $E[|X_i - \mu_i|^3] < \infty$ for all $n \in \mathbb{N}$ and

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n E[|X_i - \mu_i|^3]}{\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}} = 0. \qquad (*)$$

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

71) The (Lindeberg-Lévy) CLT has the X_i iid with $V(X_i) = \sigma^2 < \infty$. The Lyapounov CLT in 70 iii) has the X_i independent (not necessarily identically distributed), but needs stronger moment conditions to satisfy $(*)$.

72) **Lindeberg CLT**: Let the W_{nk} satisfy 68) and Lindeberg's condition

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{E(W_{nk}^2 \; I[|W_{nk}| \ge \epsilon s_n])}{s_n^2} = 0$$

for any $\epsilon > 0$. Then

$$Z_n = \frac{\sum_{k=1}^{r_n} W_{nk}}{s_n} \xrightarrow{D} N(0,1).$$

Notes: The Lindeberg CLT is sometimes called the Lindeberg-Feller CLT. Lindeberg's condition is nearly necessary for $Z_n = \frac{\sum_{k=1}^{r_n} W_{nk}}{s_n} \xrightarrow{D} N(0,1)$. Lindeberg's condition is

$$\lim_{n\to\infty} \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{\{|W_{nk}| \ge \epsilon s_n\}} W_{nk}^2 dP = 0$$

for any $\epsilon > 0$.

73) Special case of the Lindeberg CLT: Let $r_n = n$ and let the $W_{nk} = W_k$ be independent. If

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{E(W_k^2 \; I[|W_k| \ge \epsilon s_n])}{s_n^2} = 0$$

for any $\epsilon > 0$. Then

$$Z_n = \frac{\sum_{k=1}^{n} W_k}{s_n} \xrightarrow{D} N(0,1).$$

74) a) **uniformly bounded sequence**: Let $r_n = n$ and $W_{nk} = W_k$. If there is a constant $c > 0$ such that $P(|W_k| < c) = 1 \ \forall k$, and if $s_n \to \infty$ as $n \to \infty$, then Lindeberg's CLT 73) holds.

b) Let $r_n = n$ and let the $W_{nk} = W_k$ be **iid** with $V(W_k) = \sigma^2 \in (0, \infty)$. Then Lindeberg's CLT 73) holds. (Taking $W_i = X_i - \mu$ proves the usual CLT with the Lindeberg CLT.)

c) If Lyapunov's condition holds, then Lindeberg's condition holds. Hence the Lindeberg CLT proves the Lyapounov CLT.

75) The Hájek Šidak CLT: Let $X_1, ..., X_n$ be iid with $E(X_i) = \mu$ and $V(X_i) = \sigma^2$. Let $\boldsymbol{c}_n = (c_{n1}, ..., c_{nn})^T$ be a vector of constants such that

$$\max_{1 \le i \le n} \frac{c_{ni}^2}{\sum_{j=1}^n c_{nj}^2} \to 0 \ \text{ as } \ \text{n} \to \infty.$$

Then

$$Z_n = \frac{\sum_{i=1}^n c_{ni}(X_i - \mu)}{\sigma \sqrt{\sum_{j=1}^n c_{nj}^2}} \xrightarrow{D} N(0, 1).$$

Note: $c_{ni} = 1/n$ gives the usual CLT.

76) a) A sequence of random variables W_n is *tight* or *bounded in probability,* written $W_n = O_P(1)$, if for every $\epsilon > 0$ there exist positive constants D_ϵ and N_ϵ such that

$$P(|W_n| \le D_\epsilon) \ge 1 - \epsilon$$

for all $n \ge N_\epsilon$. Also $W_n = O_P(X_n)$ if $|W_n/X_n| = O_P(1)$.

b) The sequence $W_n = o_P(n^{-\delta})$ if $n^\delta W_n = o_P(1)$ which means that

$$n^\delta W_n \xrightarrow{P} 0.$$

c) W_n has the *same order as* X_n *in probability*, written $W_n \asymp_P X_n$, if for every $\epsilon > 0$ there exist positive constants N_ϵ and $0 < d_\epsilon < D_\epsilon$ such that

$$P(d_\epsilon \le \left| \frac{W_n}{X_n} \right| \le D_\epsilon) \ge 1 - \epsilon$$

for all $n \ge N_\epsilon$.

d) Similar notation is used for a $k \times r$ matrix $\boldsymbol{A}_n = [a_{i,j}(n)]$ if each element $a_{i,j}(n)$ has the desired property. For example, $\boldsymbol{A}_n = O_P(n^{-1/2})$ if each $a_{i,j}(n) = O_P(n^{-1/2})$.

77) Let $\hat{\boldsymbol{\beta}}_n$ be an estimator of a $p \times 1$ vector $\boldsymbol{\beta}$, and let $W_n = \|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\|$.

a) If $W_n \asymp_P n^{-\delta}$ for some $\delta > 0$, then both W_n and $\hat{\boldsymbol{\beta}}_n$ have (tightness) **rate** n^δ.

b) If there exists a constant θ such that

$$n^\delta (W_n - \theta) \xrightarrow{D} X$$

for some nondegenerate random variable X, then both W_n and $\hat{\boldsymbol{\beta}}_n$ have *convergence rate* n^δ.

78) Suppose there exists a constant θ such that

$$n^{\delta}(W_n - \theta) \xrightarrow{D} X.$$

a) Then $W_n = O_P(n^{-\delta})$.
b) If X is not degenerate, then $W_n \asymp_P n^{-\delta}$.

79) a) If $W_n \asymp_P X_n$ then $X_n \asymp_P W_n$.
b) If $W_n \asymp_P X_n$ then $W_n = O_P(X_n)$.
c) If $W_n \asymp_P X_n$ then $X_n = O_P(W_n)$.
d) $W_n \asymp_P X_n$ iff $W_n = O_P(X_n)$ and $X_n = O_P(W_n)$.

80) Pratt (1959). Let $X_{1,n}, ..., X_{K,n}$ each be $O_P(1)$ where K is fixed. Suppose $W_n = X_{i_n,n}$ for some $i_n \in \{1, ..., K\}$. Then $W_n = O_P(1)$.

81) Suppose $\|T_{j,n} - \boldsymbol{\beta}\| = O_P(n^{-\delta})$ for $j = 1, ..., K$ where $0 < \delta \leq 1$. Let $T_n^* = T_{i_n,n}$ for some $i_n \in \{1, ..., K\}$ where, for example, $T_{i_n,n}$ is the $T_{j,n}$ that minimized some criterion function. Then $\|T_n^* - \boldsymbol{\beta}\| = O_P(n^{-\delta})$.

82) Let W_n, X_n, Y_n and Z_n be sequences of random variables such that $Y_n > 0$ and $Z_n > 0$. (Often Y_n and Z_n are deterministic, e.g. $Y_n = n^{-1/2}$.)

a) If $W_n = O_P(1)$ and $X_n = O_P(1)$, then $W_n + X_n = O_P(1)$ and $W_n X_n = O_P(1)$, thus $O_P(1) + O_P(1) = O_P(1)$ and $O_P(1)O_P(1) = O_P(1)$.

b) If $W_n = O_P(1)$ and $X_n = o_P(1)$, then $W_n + X_n = O_P(1)$ and $W_n X_n = o_P(1)$, thus $O_P(1) + o_P(1) = O_P(1)$ and $O_P(1)o_P(1) = o_P(1)$.

c) If $W_n = O_P(Y_n)$ and $X_n = O_P(Z_n)$, then $W_n + X_n = O_P(\max(Y_n, Z_n))$ and $W_n X_n = O_P(Y_n Z_n)$, thus $O_P(Y_n) + O_P(Z_n) = O_P(\max(Y_n, Z_n))$ and $O_P(Y_n)O_P(Z_n) = O_P(Y_n Z_n)$.

83) Consider predicting a future test value Y_f given training data $Y_1, ..., Y_n$. A large sample $100(1-\delta)\%$ *prediction interval* (PI) for Y_f has the form $[\hat{L}_n, \hat{U}_n]$ where $P(\hat{L}_n \leq Y_f \leq \hat{U}_n)$ is eventually bounded below by $1-\delta$ as the sample size $n \to \infty$. A large sample $100(1-\delta)\%$ PI is *asymptotically optimal* if it has the shortest asymptotic length: the length of $[\hat{L}_n, \hat{U}_n]$ converges to $U_s - L_s$ as $n \to \infty$ where $[L_s, U_s]$ is the *population shorth*: the shortest interval covering at least $100(1-\delta)\%$ of the mass.

84) Let $Y_1, ..., Y_n, Y_f$ be iid. Let $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$ be the order statistics of the training data. Let $k_1 = \lceil n\delta/2 \rceil$ and $k_2 = \lceil n(1-\delta/2) \rceil$ where $0 < \delta < 1$. The large sample $100(1-\delta)\%$ *percentile prediction interval* for Y_f is

$$[Y_{(k_1)}, Y_{(k_2)}]. \tag{1}$$

85) Let the shortest closed interval containing at least c of the $Y_1, ..., Y_n$ be $\mathrm{shorth(c)} = [\mathrm{Y_{(s)}, Y_{(s+c-1)}}]$.

86) Let $Y_1, ..., Y_n, Y_f$ be iid. Let $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$ be the order statistics of the training data. The large sample $100(1-\delta)\%$ shorth(c) prediction interval for Y_f is

$$[Y_{(s)}, Y_{(s+c-1)}] \quad \text{where} \quad \mathrm{c = min(n, \lceil n[1 - \delta + 1.12\sqrt{\delta/n}\ ]\ \rceil)}.$$

87) Let $Y_1, ..., Y_n, Y_f$ be iid. Let $W_{(1)} \leq W_{(2)} \leq \cdots \leq W_{(n)}$ be the order statistics of the squared training data $W_1, ..., W_n$ where $W_i = Y_i^2$ for $i = 1, ..., n$. Let $k_n = \lceil n(1-\delta) \rceil$. Let $L_n = -U_n$ and $U_n = \sqrt{W_{(k_n)}}$. Then $[L_n, U_n]$ is a large sample $100(1-\delta)\%$ PI for Y_f.

88) Let $Y_1, ..., Y_n, Y_f$ be iid. Suppose that $E(Y) = \mu$ and the standard deviation $SD(Y) = \sigma$. Let $\hat{\mu}$ and $\hat{\sigma}$ be consistent estimators of μ and σ. Let $1 - 1/k^2 \geq 1 - \delta$. Let $\mu \pm k\sigma$ be continuity points of $F_Y(y)$. Then

$$[L_n, U_n] = [\hat{\mu} - k\hat{\sigma}, \hat{\mu} + k\hat{\sigma}]$$

is a large sample $100(1-\delta)\%$ *Chebyshev PI* for Y_f.

Note often $k = 1.96$ is used which is good for a 95% PI for iid normal data, but is usually too short to be a 95% PI for iid data.

89) In a simulation for a PI, prediction region, CI, or confidence region with nominal $100(1-\delta)\%$ coverage, let the actual coverage $1 - \delta_n = P(a_n \in R)$ be $P(Y_f \in PI)$, $P(\boldsymbol{Y}_f \in$ prediction region), $P(\theta \in CI)$, or $P(\boldsymbol{\theta} \in$ confidence region). Then $P(a_n \in R) \sim bin(k, 1-\delta_n) \approx bin(k, 1-\delta)$ where k is the number of runs in the simulation. a) for $k = 5000$, simulated coverage in [0.94,0.95] suggests the actual coverage $1 - \delta_n$ is close to the nominal coverage $1 - \delta = 0.95$. b) for $k = 100$, simulated coverage in [0.89,1] suggests the actual coverage $1 - \delta_n$ is close to the nominal coverage $1 - \delta = 0.95$.