Math 581 Exam 3 is Tuesday, Nov. 30, 2:00-3:15 NO NOTES. CHECK FORMULAS: YOU ARE RESPONSIBLE FOR ANY ERRORS ON THIS HANDOUT!

Know the large sample theory 89)-92) from Exam 2 review.

93) The **characteristic function** (cf) $\varphi_X(t) = E(e^{itX}) = E[\cos(tX)] + iE[\sin(tX)]$ where the complex number $i = \sqrt{-1}$. If X is a random variable, then $\varphi_X(t)$ always exists, and completely determines the distribution of X. If the moment generating function $M_X(t)$ exists, then the mgf also completely determines the distribution of X. If X is discrete with pmf $p_X(x)$, then $\varphi_X(t) = \sum_x e^{itx} p_X(x)$.

94) **Continuity Theorem:** Let X_n be sequence of random variables with characteristic functions $\phi_{X_n}(t)$. Let X be a random variable with cf $\phi_X(t)$.

a)

$$X_n \xrightarrow{D} X \quad \text{iff} \quad \phi_{\mathrm{X_n}}(\mathrm{t}) \to \phi_{\mathrm{X}}(\mathrm{t}) \ \forall \mathrm{t} \in \mathbb{R}.$$

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

95) Some properties of a characteristic function.

i) $\varphi(0) = 1$

ii) $|\varphi(t)| \le 1$ where the modulus $|a + ib| = \sqrt{a^2 + b^2}$.

iii) $\varphi(t)$ is a continuous function.

iv) $\varphi(t)$ is uniformly continuous.

96) If $\phi_{X_n}(t) \to h(t)$ where $h(t)$ is not continuous, then X_n does not converge in distribution to any RV X, by the Continuity Theorem and 95).

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

98) **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: 98) is used to prove 99 b).

99) Assume that the function g does not depend on n.

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 $g(X_n) \xrightarrow{D} g(X)$.

100) A sequence of RVs X_n **converges in distribution to a constant** c, written $X_n \xrightarrow{D} c$, if $X_n \xrightarrow{D} X$ where $P(X = c) = 1$. (Hence X is the point mass at c.)

101) A sequence of random variables X_n *converges in probability to* X, written $X_n \xrightarrow{P} X$, if for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1 \text{ or, equivalently, } \lim_{n\to\infty} \mathrm{P}(|X_n - X| \geq \epsilon) = 0.$$

102) A sequence of random variables X_n *converges in probability to a constant* c, written $X_n \xrightarrow{P} c$, if for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - c| < \epsilon) = 1 \text{ or, equivalently, } \lim_{n\to\infty} \mathrm{P}(|X_n - c| \geq \epsilon) = 0.$$

103) In 101), X is a RV, and $X_n \xrightarrow{P} X$ iff $(X_n - X) \xrightarrow{P} 0$.

104) For a real number $r > 0$, X_n **converges in rth mean** to a random variable X, written $X_n \xrightarrow{r} X$, if

$$E(|X_n - X|^r) \to 0$$

as $n \to \infty$. In particular, if $r = 2$, X_n **converges in quadratic mean** to X, written $X_n \xrightarrow{2} X$ or $X_n \xrightarrow{qm} X$, if $E[(X_n - X)^2] \to 0$ as $n \to \infty$. If $r \geq 1$, $X_n \xrightarrow{r} X$ is often written as $X_n \xrightarrow{L^r} X$ or $X_n \xrightarrow{L_r} X$.

105) For a real number $r > 0$, X_n **converges in rth mean** to a constant c, written $X_n \xrightarrow{r} c$, if

$$E(|X_n - c|^r) \to 0$$

as $n \to \infty$.

106) A sequence of random variables X_n *converges almost everywhere* to X or *almost surely* to X, or *with probability 1* to X if

$$P(\lim_{n\to\infty} X_n = X) = 1.$$

This type of convergence will be denoted by $X_n \xrightarrow{ae} X$, or $X_n \xrightarrow{as} X$, or $X_n \xrightarrow{wp1} X$.

Note: Convergence ae is also known as strong convergence, while convergence in probability is also known as weak convergence.

107) For a constant c, $X_n \xrightarrow{ae} c$, $X_n \xrightarrow{as} c$, $X_n \xrightarrow{wp1} c$ if $P(\lim_{n\to\infty} X_n = c) = 1$.

108) a) For $X_n \xrightarrow{P} X$, $X_n \xrightarrow{r} X$, or $X_n \xrightarrow{wp1} X$, the X_n and X need to be defined on the same probability space.

b) For $X_n \xrightarrow{D} X$, the probability spaces can differ.

c) For $X_n \xrightarrow{P} c$, $X_n \xrightarrow{wp1} c$, $X_n \xrightarrow{D} c$, and $X_n \xrightarrow{r} c$, the probability spaces can differ.

109) Strong Law of Large Numbers (**SLLN**): If $X_1, ..., X_n$ are iid with $E(X_i) = \mu$ finite, then $\overline{X}_n \xrightarrow{wp1} \mu$.

110) Weak Law of Large Numbers (**WLLN**): If $X_1, ..., X_n$ are iid with $E(X_i) = \mu$ finite, then $\overline{X}_n \xrightarrow{P} \mu$.

Remark: Know the proof for when the X_i also have $V(X_i) = \sigma^2$, but do not forget that the SLLN and WLLN hold when $E(X_i) = \mu$ finite, even if $V(X_i)$ does not exist.

111) If $X_n \xrightarrow{D} X$ and $X_n \xrightarrow{D} Y$, then $X \stackrel{D}{=} Y$. If $F_n \xrightarrow{W} F$ and $F_n \xrightarrow{W} G$, then $F = G$.

112) Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function that does not depend on n.
a) If $X_n \xrightarrow{D} X$, then $g(X_n) \xrightarrow{D} g(X)$.
b) If $X_n \xrightarrow{P} X$, then $g(X_n) \xrightarrow{P} g(X)$.
c) If $X_n \xrightarrow{wp1} X$, then $g(X_n) \xrightarrow{wp1} g(X)$.

Remark: **Know the proofs for a) and c)** where the continuous mapping theorem a) is proved with the Helly-Bray-Portmanteau Theorem and the continuity theorem.

113) Theorem: Suppose X_n and X are RVs with the same probability space.
a) If $X_n \xrightarrow{wp1} X$, then $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{D} X$.
b) If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{D} X$.
c) If $X_n \xrightarrow{r} X$, then $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{D} X$.
d) $X_n \xrightarrow{P} c$ **iff** $X_n \xrightarrow{D} c$ where c is a constant.

114) a) If $E[(X_n - X)^2] \to 0$ as $n \to \infty$, then $X_n \xrightarrow{P} X$.

b) If $E(X_n) \to E(X)$ and $V(X_n - X) \to 0$ as $n \to \infty$, then $X_n \xrightarrow{P} X$.

c) Let c be a constant. If $E(X_n) \to c$ and $V(X_n) \to 0$ as $n \to \infty$, then $X_n \xrightarrow{P} c$.

Know the proofs for 114).

115) a) If X has mgf $M(t)$, then $E(X^k)$ exists for all positive integers k.

b) Let j and k be positive integers. If $E(X^k)$ is finite, then $E(X^j)$ is finite for $1 \le j \le k$.

116) if $n^\delta(X_n - c) \xrightarrow{D} X$ where $0 < \delta \le 1$, then $X_n \xrightarrow{P} c$. So if $\sqrt{n}(X_n - c) \xrightarrow{D} N(0, \sigma^2)$, then $X_n \xrightarrow{P} c$.

117) **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 $c_n \xrightarrow{ae} c$ and $c_n \xrightarrow{P} c$.

118) If $X_n \xrightarrow{r} X$, then $X_n \xrightarrow{k} X$ where $0 < k < r$.

119) 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 $X_n \xrightarrow{D} X$.

120) 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 $g(X_n) \xrightarrow{wp1} c$.

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

122) A complex RV Z has the form $Z = X + iY$ where X and Y are ordinary RVs. Then $E(Z) = E(X) + iE(Y)$, and Z is integrable if $E[|Z|] = E[\sqrt{X^2 + Y^2}] < \infty$. Linearity, LDCT, and key inequalities remain valid, including $|E[Z]| \le E[|Z|]$.

123) Theorem: If $\lim_{n\to\infty} \varphi_{X_n}(t) = g(t)$ for all t where g is continuous at $t = 0$, then $g(t) = \varphi_X(t)$ 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.

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

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

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

125) **Lyapounov's CLT**: Under 124), assume the $|W_{nk}|^{2+\delta}$ are integrable for some $\delta > 0$. Assume Lyapounov's condition:

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

Then

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

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

127) The (Lindeberg-Lévy) CLT has the X_i iid with $V(X_i) = \sigma^2 < \infty$. The Lyapounov CLT in 126 iii) has the X_i independent (not necessarily identically distributed), but needs stronger moment conditions to satisfy $(*)$.

128) **Lindeberg CLT**: Let the W_{nk} satisfy 124) and Lindeberg's condition

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

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 $\epsilon > 0$.

129) 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

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

130) 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 129) 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 129) 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.

Section 29–Large Sample Theory for Random Vectors:

131) **Change in Notation**: 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*

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

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

132) 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

$$E(\boldsymbol{a} + \boldsymbol{X}) = \boldsymbol{a} + E(\boldsymbol{X}) \quad \text{and} \quad E(\boldsymbol{X} + \boldsymbol{Y}) = E(\boldsymbol{X}) + E(\boldsymbol{Y})$$

and

$$E(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}E(\boldsymbol{X}) \quad \text{and} \quad E(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) = \boldsymbol{A}E(\boldsymbol{X})\boldsymbol{B}.$$

Thus

$$\text{Cov}(\boldsymbol{a} + \boldsymbol{A}\boldsymbol{X}) = \text{Cov}(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}\text{Cov}(\boldsymbol{X})\boldsymbol{A}^T. \tag{1}$$

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

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) $F_{\boldsymbol{X}}(\boldsymbol{x}) = P(X_1 \le x_1, ..., X_k \le x_k)$.

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

134) 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 rth 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 almost everywhere** to $\boldsymbol{X}$, written $\boldsymbol{X}_n \xrightarrow{ae} \boldsymbol{X}$, or $\boldsymbol{X}_n$ **converges almost surely** to $\boldsymbol{X}$, written $\boldsymbol{X}_n \xrightarrow{as} \boldsymbol{X}$, or $\boldsymbol{X}_n$ **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 $\boldsymbol{X}_n \xrightarrow{wp1} \boldsymbol{c}$.

135) **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 $\epsilon > 0$,

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

136) 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 $\epsilon > 0$,

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

137) 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 $\text{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$. 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 $k = 1$.

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

139) 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 $k = 1$.

140) Theorem: If $0 < \delta \leq 1$, $\boldsymbol{X}$ is a random vector, and

$$n^{\delta}(\boldsymbol{X}_n - \boldsymbol{c}) \xrightarrow{D} \boldsymbol{X},$$

then $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{c}$.

141) 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: $\overline{\boldsymbol{X}}_n \xrightarrow{ae} \boldsymbol{\mu}$.

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

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

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

143) **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

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

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

144) Use 142) and 143) to prove the MCLT. Use 142) to prove 143).

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

b)

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

146) **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 $P(\boldsymbol{X} \in D(\boldsymbol{g})) = 0$.)

147) **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 $\boldsymbol{Y}_n \xrightarrow{P} \boldsymbol{c}$ for some constant $k \times 1$ vector $\boldsymbol{c}$, then i) $\boldsymbol{X}_n + \boldsymbol{Y}_n \xrightarrow{D} \boldsymbol{X} + \boldsymbol{c}$ and

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

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 $\boldsymbol{X}_n^T \boldsymbol{W}_n^{-1} \xrightarrow{D} \boldsymbol{X}^T \boldsymbol{C}^{-1}$.

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

149) 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.

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

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

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 149) since independent random vectors have a joint distribution that does not affect the marginal distributions.

151) 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 $\varphi_{\boldsymbol{X}}, \varphi_{\boldsymbol{X}_1}$, and $\varphi_{\boldsymbol{X}_2}$ be the characteristic functions of $\boldsymbol{X}, \boldsymbol{X}_1$ and $\boldsymbol{X}_2$. Then $\boldsymbol{X}_1 \perp\!\!\!\perp \boldsymbol{X}_2$ iff

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

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

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

153) Theorem:

i) $\boldsymbol{X}_n \xrightarrow{wp1} \boldsymbol{X} \Rightarrow \boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X}$.

ii) $\boldsymbol{X}_n \xrightarrow{r} \boldsymbol{X} \Rightarrow \boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X}$.

iii) $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X} \Rightarrow \boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$.

iv) $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{c}$ iff $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{c}$.

Conditional Expectation

154) Let μ and ν be measures on $(\Omega, \mathcal{F})$. Then ν is **absolutely continuous wrt** μ if for each $A \in \mathcal{F}$, $\mu(A) = 0 \Rightarrow \nu(A) = 0$, written $\nu << \mu$.

155) **Radon-Nikodym Theorem**: If μ and ν are σ-finite measures such that $\nu << \mu$, then there exists a measurable, nonnegative f, a density, such that $\nu(A) = \int_A f d\mu$ for all $A \in \mathcal{F}$. For two such densities f and g, $\mu[f \neq g] = 0$. Hence $f = g \quad \mu$ ae.

156) The density $f = \dfrac{d\nu}{d\mu}$ is called the Radon-Nikodym derivative of ν wrt μ. Note that $\nu(A) = \displaystyle\int_A \frac{d\nu}{d\mu} d\mu = \int_A d\nu$ for all $A \in \mathcal{F}$.

157) The Radon-Nikodym Theorem is used to prove the existence of the conditional probability $P(A|\mathbb{G})$ and of the conditional expectation $E(X|\mathbb{G})$. See points 158) and 160).

158) Fix $A \in \mathcal{F}$ and let the σ-field $\mathbb{G} \subseteq \mathcal{F}$. A **conditional probability of** A **given** $\mathbb{G}$ is an $f = P[A|\mathbb{G}]$ that is i) measurable $\mathbb{G}$ and integrable, and ii) $\int_G P[A|\mathbb{G}]dP = E[P(A|\mathbb{G})I_G] = P(A \cap G)$ for any $G \in \mathbb{G}$.

159) i) Note that $f = P[A|\mathbb{G}]$ is a random variable wrt $\mathbb{G}$.
ii) $0 \leq P[A|\mathbb{G}] \leq 1$ wp1.
iii) There are many such RVs $P[A|\mathbb{G}]$ satisfying 158), but any two of them are equal wp1. A specific such RV is called a **version** of $P[A|\mathbb{G}]$.

160) Let $E(X)$ exist on $(\Omega, \mathcal{F}, P)$, and let the σ-field $\mathbb{G} \subseteq \mathcal{F}$. A **conditional expectation of** X **given** $\mathbb{G}$ is a $f = E[X|\mathbb{G}]$ that is i) measurable $\mathbb{G}$ and integrable, and ii) $\int_G E[X|\mathbb{G}]dP = E[E(X|\mathbb{G})I_G] = E[XI_G] = \int_G XdP$ for any $G \in \mathbb{G}$.

161) i) Note that $f = E[X|\mathbb{G}]$ is a random variable wrt $\mathbb{G}$.
ii) There are many such RVs $E[X|\mathbb{G}]$ satisfying 160), but any two of them are equal wp1. A specific such RV is called a **version** of $E[X|\mathbb{G}]$.

161) i) Fix $A \in \mathcal{F}$. If $X = I_A$, then $E[I_A|\mathbb{G}]$ is a version of $P[A|\mathbb{G}]$.
ii) Since $\mathbb{G} \subseteq \mathcal{F}$, often X is not measurable $\mathbb{G}$. Then X is not a version of $E[X|\mathbb{G}]$. If X is measurable $\mathbb{G}$, then X is a version of $E[X|\mathbb{G}]$.

162) Theorem: If X is measurable $\mathbb{G}$ and Y and XY are integrable, then $E[XY|\mathbb{G}] = XE[Y|\mathbb{G}]$ wp1. That is, $XE[Y|\mathbb{G}]$ is a version of $E[XY|\mathbb{G}]$.

163) Theorem: Let X, Y, and X_n be integrable. Let a and b be constants.
i) If $X = a$ wp1, then $E[X|\mathbb{G}] = a$ wp1.
ii) $E[(aX + bY)|\mathbb{G}] = aE[X|\mathbb{G}] + bE[Y|\mathbb{G}]$ wp1.
iii) If $X \leq Y$ wp1, then $E[X|\mathbb{G}] \leq E[Y|\mathbb{G}]$ wp1.
iv) $|E[X|\mathbb{G}]| \leq E[|X| \mid \mathbb{G}]$ wp1.
v) If $\lim_n X_n = X$ wp1, $|X_n| \leq Y$, and Y is integrable, then $\lim_n E[X_n|\mathbb{G}] = E[X|\mathbb{G}]$ wp1.

164) If X is integrable and $\sigma-$fields $\mathbb{G}_1 \subseteq \mathbb{G}_2 \subseteq \mathcal{F}$, then $E(E[X|\mathbb{G}_2]|\mathbb{G}_1) = E[X|\mathbb{G}_1]$ wp1.