Exam 1, Thursday, Feb. 10, 10 sheets of notes and a calculator,

0) The *gamma function* $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ for $x > 0$.

i) $\Gamma(k) = (k-1)!$ for integer $k \geq 1$. ii) $\Gamma(x+1) = x\ \Gamma(x)$ for $x > 0$.

iii) $\Gamma(x) = (x-1)\ \Gamma(x-1)$ for $x > 1$. iv) $\Gamma(0.5) = \sqrt{\pi}$.

Let $P(Y \leq y_\delta) = \delta$ if the pdf of $Y$ is positive at $y_\delta$.

a) $Y \sim$ beta$(\delta, \nu)$

$$f(y) = \frac{\Gamma(\delta+\nu)}{\Gamma(\delta)\Gamma(\nu)} y^{\delta-1}(1-y)^{\nu-1}$$

where $\delta > 0$, $\nu > 0$ and $0 \leq y \leq 1$.

$$E(Y) = \frac{\delta}{\delta+\nu}, \quad V(Y) = \frac{\delta\nu}{(\delta+\nu)^2(\delta+\nu+1)}.$$

b) Bernoulli$(\rho)$ = binomial$(k = 1, \rho)$, $\quad f(y) = \rho^y(1-\rho)^{1-y}$ for $y = 0, 1$.
$E(Y) = \rho$, $V(Y) = \rho(1-\rho)$.

$$m(t) = [(1-\rho) + \rho e^t], \quad c(t) = [(1-\rho) + \rho e^{it}].$$

c) binomial$(k, \rho)$, $Y \sim BIN(k, \rho)$,

$$f(y) = \binom{k}{y} \rho^y (1-\rho)^{k-y}$$

for $y = 0, 1, \ldots, k$ where $0 < \rho < 1$. $\quad E(Y) = k\rho, \quad V(Y) = k\rho(1-\rho)$. 1P-REF is $k$ is known, and $I_1(\rho) = \dfrac{k}{\rho(1-\rho)}$. $m(t) = [(1-\rho) + \rho e^t]^k$, $c(t) = [(1-\rho) + \rho e^{it}]^k$. If $Y_1, ..., Y_n$ are independent binomial BIN$(k_i, \rho)$ random variables, then

$$\sum_{i=1}^n Y_i \sim \text{BIN}\left(\sum_{\rm i=1}^{\rm n} \rm k_i, \rho\right).$$

Thus if $Y_1, ..., Y_n$ are iid BIN$(k, \rho)$ random variables, then $\sum_{i=1}^n Y_i \sim$ BIN(nk, $\rho$).

d) $Y \sim$ Cauchy$(\mu, \sigma)$,

$$f(y) = \frac{1}{\pi\sigma[1 + (\frac{y-\mu}{\sigma})^2]}$$

where $y$ and $\mu$ are real numbers and $\sigma > 0$. $\quad E(Y)$ and $V(Y)$ do not exist.
$c(t) = \exp(it\mu - |t|\sigma)$.

$$F(y) = \frac{1}{\pi}[\arctan(\frac{y-\mu}{\sigma}) + \pi/2].$$

e) chi-square$(p)$ = gamma$(\nu = p/2, \lambda = 2)$, $Y \sim \chi^2_p$,

$$f(y) = \frac{y^{\frac{p}{2}-1}e^{-\frac{y}{2}}}{2^{\frac{p}{2}}\Gamma(\frac{p}{2})}$$

where $y > 0$ and $p$ is a positive integer. $\quad E(Y) = p, \quad V(Y) = 2p$.

$$m(t) = \left(\frac{1}{1-2t}\right)^{p/2} = (1-2t)^{-p/2} \text{ for } \rm t < 1/2, \quad c(t) = \left(\frac{1}{1-i2t}\right)^{p/2}.$$

If $Y_1, ..., Y_n$ are independent chi–square $\chi^2_{p_i}$, then

$$\sum_{i=1}^n Y_i \sim \chi^2(\sum_{i=1}^n p_i).$$

Thus if $Y_1, ..., Y_n$ are iid $\chi^2_p$, then

$$\sum_{i=1}^n Y_i \sim \chi^2_{np}.$$

f) exponential($\lambda$)= gamma($\nu = 1, \lambda$), $Y \sim$ EXP($\lambda$)

$$f(y) = \frac{1}{\lambda} \exp\left(-\frac{y}{\lambda}\right)\ I(y \geq 0)$$

where $\lambda > 0$. $E(Y) = \lambda$, $V(Y) = \lambda^2$, and $y_\delta = -\lambda \ln(1-\delta)$. 1P-REF and $I_1(\lambda) = 1/\lambda^2$.

$$m(t) = 1/(1 - \lambda t) \ \text{ for } \ \mathrm{t} < 1/\lambda, \ \ \mathrm{c(t)} = 1/(1 - \mathrm{i}\lambda\mathrm{t}).$$

$$F(y) = 1 - \exp(-y/\lambda),\ y \geq 0.$$

If $Y_1, ..., Y_n$ are iid exponential EXP($\lambda$), then

$$\sum_{i=1}^n Y_i \sim G(n, \lambda).$$

g) gamma($\nu, \lambda$), $Y \sim G(\nu, \lambda)$,

$$f(y) = \frac{y^{\nu-1} e^{-y/\lambda}}{\lambda^\nu \Gamma(\nu)}$$

where $\nu, \lambda$, and $y$ are positive. $E(Y) = \nu\lambda$, $V(Y) = \nu\lambda^2$. 2P-REF and if $\nu$ is known, then $I_1(\lambda) = \nu/\lambda^2$.

$$m(t) = \left(\frac{1}{1 - \lambda t}\right)^\nu \ \text{ for } \ \mathrm{t} < 1/\lambda, \ \ \mathrm{c(t)} = \left(\frac{1}{1 - \mathrm{i}\lambda\mathrm{t}}\right)^\nu.$$

If $Y_1, ..., Y_n$ are independent Gamma $G(\nu_i, \lambda)$ then

$$\sum_{i=1}^n Y_i \sim G(\sum_{i=1}^n \nu_i, \lambda).$$

Thus if $Y_1, ..., Y_n$ are iid $G(\nu, \lambda)$, then $\sum_{i=1}^n Y_i \sim G(n\nu, \lambda)$.

h) $Y \sim N(\mu, \sigma^2)$

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right)$$

where $\sigma > 0$ and $\mu$ and $y$ are real. $E(Y) = \mu$, $V(Y) = \sigma^2$, and $y_\delta = \mu + \sigma z_\delta$. 2P-REF. If $\sigma^2$ is known, then $I_1(\mu) = 1/\sigma^2$. If $\mu$ is known, then $I_1(\sigma^2) = \dfrac{1}{2\sigma^4}$.

$$I_1(\mu,\sigma) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}, \quad I_1(\mu,\sigma^2) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}.$$

$$m(t) = \exp(t\mu + t^2\sigma^2/2), \quad c(t) = \exp(it\mu - t^2\sigma^2/2).$$

$$F(y) = \Phi\left(\frac{y-\mu}{\sigma}\right).$$

If $Y_1, ..., Y_n$ are independent normal $N(\mu_i, \sigma_i^2)$, then

$$\sum_{i=1}^n (a_i + b_i Y_i) \sim N(\sum_{i=1}^n (a_i + b_i\mu_i), \sum_{i=1}^n b_i^2\sigma_i^2).$$

Here $a_i$ and $b_i$ are fixed constants. Thus if $Y_1, ..., Y_n$ are iid $N(\mu, \sigma^2)$, then $\overline{Y} \sim N(\mu, \sigma^2/n)$.

i) Poisson($\theta$), $Y \sim$ POIS($\theta$)

$$f(y) = \frac{e^{-\theta}\theta^y}{y!}$$

for $y = 0, 1, \ldots,$ where $\theta > 0$. $E(Y) = \theta = V(Y)$. 1P-REF and $I_1(\theta) = 1/\theta$.

$$m(t) = \exp(\theta(e^t - 1)), \quad c(t) = \exp(\theta(e^{it} - 1)).$$

If $Y_1, ..., Y_n$ are independent POIS($\theta_i$), then

$$\sum_{i=1}^n Y_i \sim \mathrm{POIS}(\sum_{\mathrm{i=1}}^{\mathrm{n}} \theta_{\mathrm{i}}).$$

Thus if $Y_1, ..., Y_n$ are iid $POIS(\theta)$, then

$$\sum_{i=1}^n Y_i \sim \mathrm{POIS(n}\theta).$$

j) uniform($\theta_1, \theta_2$), $Y \sim U(\theta_1, \theta_2)$.

$$f(y) = \frac{1}{\theta_2 - \theta_1} I(\theta_1 \le y \le \theta_2).$$

$F(y) = (y - \theta_1)/(\theta_2 - \theta_1)$ for $\theta_1 \le y \le \theta_2$. $E(Y) = (\theta_1 + \theta_2)/2$. $V(Y) = (\theta_2 - \theta_1)^2/12$, and $y_\delta = (\theta_2 - \theta_1)\delta + \theta_1$. By definition, $m(0) = c(0) = 1$. For $t \neq 0$,

$$m(t) = \frac{e^{t\theta_2} - e^{t\theta_1}}{(\theta_2 - \theta_1)t}, \quad \text{and} \quad \mathrm{c(t)} = \frac{\mathrm{e}^{\mathrm{it}\theta_2} - \mathrm{e}^{\mathrm{it}\theta_1}}{(\theta_2 - \theta_1)\mathrm{it}}.$$

k) point mass at $c$: The distribution of $Y$ is a point mass at $c$ (or $Y$ is degenerate at $c$) if $P(Y = c) = 1$ with pmf $f(c) = 1$. Hence $Y \sim N(c, 0)$, $E(Y) = c$, $V(Y) = 0$. $m(t) = e^{tc}$. $c(t) = e^{itc}$.

**More Distributions:**

l) If $Y$ has a geometric distribution, $Y \sim \text{geom}(\rho)$ then the pmf of $Y$ is

$$f(y) = P(Y = y) = \rho(1-\rho)^y$$

for $y = 0, 1, 2, ...$ and $0 < \rho < 1$. $E(Y) = (1-\rho)/\rho$. $V(Y) = (1-\rho)/\rho^2$. $Y \sim NB(1, \rho)$. Hence the mgf of $Y$ is

$$m(t) = \frac{\rho}{1-(1-\rho)e^t}$$

for $t < -\log(1-\rho)$.

m) If $Y$ has a negative binomial distribution, $Y \sim \text{NB}(\text{r}, \rho)$, then the pmf of $Y$ is

$$f(y) = P(Y = y) = \binom{r+y-1}{y} \rho^r (1-\rho)^y$$

for $y = 0, 1, \ldots$ where $0 < \rho < 1$. $E(Y) = r(1-\rho)/\rho$, and

$$V(Y) = \frac{r(1-\rho)}{\rho^2}.$$

The moment generating function

$$m(t) = \left[\frac{\rho}{1-(1-\rho)e^t}\right]^r$$

for $t < -\log(1-\rho)$.

n) If $Y$ has an F distribution, $Y \sim F(\nu_1, \nu_2)$, then the pdf of $Y$ is

$$f(y) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \frac{y^{(\nu_1-2)/2}}{\left(1+(\frac{\nu_1}{\nu_2})y\right)^{(\nu_1+\nu_2)/2}}$$

where $y > 0$ and $\nu_1$ and $\nu_2$ are positive integers.

$$E(Y) = \frac{\nu_2}{\nu_2 - 2}, \quad \nu_2 > 2$$

and

$$V(Y) = 2\left(\frac{\nu_2}{\nu_2-2}\right)^2 \frac{(\nu_1+\nu_2-2)}{\nu_1(\nu_2-4)}, \quad \nu_2 > 4.$$

o) If $Y$ has a Student's $t$ distribution, $Y \sim t_p$, then the pdf of $Y$ is

$$f(y) = \frac{\Gamma(\frac{p+1}{2})}{(p\pi)^{1/2}\Gamma(p/2)}(1+\frac{y^2}{p})^{-(\frac{p+1}{2})}$$

where $p$ is a positive integer and $y$ is real. This family is symmetric about 0. The $t_1$ distribution is the Cauchy$(0, 1)$ distribution. If $Z$ is $N(0, 1)$ and is independent of $W \sim \chi^2_p$, then

$$\frac{Z}{(\frac{W}{p})^{1/2}}$$

is $t_p$. $E(Y) = 0$ for $p \geq 2$. $V(Y) = p/(p-2)$ for $p \geq 3$.

**Two Multivariate Distributions:**

p) point mass at $\boldsymbol{c}$: The distribution of the $p \times 1$ random vector $\boldsymbol{Y}$ is a point mass at $\boldsymbol{c}$ (or $\boldsymbol{Y}$ is degenerate at $\boldsymbol{c}$) if $P(\boldsymbol{Y} = \boldsymbol{c}) = 1$ with pmf $f(\boldsymbol{c}) = 1$. Hence $\boldsymbol{Y} \sim N_p(\boldsymbol{c}, \boldsymbol{0})$, $E(\boldsymbol{Y}) = \boldsymbol{c}$, $\text{Cov}(\boldsymbol{Y}) = \boldsymbol{0}$, $m(\boldsymbol{t}) = e^{\boldsymbol{t}^T \boldsymbol{c}}$, $c(\boldsymbol{t}) = e^{i\boldsymbol{t}^T \boldsymbol{c}}$.

q) MVN distribution: If $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $E(\boldsymbol{Y}) = \boldsymbol{\mu}$ and $\text{Cov}(\boldsymbol{Y}) = \boldsymbol{\Sigma}$.

$$m(\boldsymbol{t}) = \exp\left(\boldsymbol{t}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right), \quad c(\boldsymbol{t}) = \exp\left(i\boldsymbol{t}^T\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right).$$

If $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and if $\boldsymbol{A}$ is a $q \times p$ matrix, then $\boldsymbol{AY} \sim N_q(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^T)$. If $\boldsymbol{a}$ is a $p \times 1$ vector of constants, then $\boldsymbol{Y} + \boldsymbol{a} \sim N_p(\boldsymbol{\mu} + \boldsymbol{a}, \boldsymbol{\Sigma})$.

$$\text{Let } \boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \text{and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

**All subsets of a MVN are MVN:** $(Y_{k_1}, ..., Y_{k_q})^T \sim N_q(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$ where $\tilde{\boldsymbol{\mu}}_i = E(Y_{k_i})$ and $\tilde{\boldsymbol{\Sigma}}_{ij} = \text{Cov}(Y_{k_i}, Y_{k_j})$. In particular, $\boldsymbol{Y}_1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\boldsymbol{Y}_2 \sim N_{p-q}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. If $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are independent iff $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$.

1) **CLT**: Let $Y_1, ..., Y_n$ be iid with $E(Y) = \mu$ and $V(Y) = \sigma^2$. Then $\sqrt{n}(\overline{Y}_n - \mu) \xrightarrow{D} N(0, \sigma^2)$.

2) a) $Z_n = \sqrt{n}\left(\frac{\overline{Y}_n - \mu}{\sigma}\right) = \left(\frac{\overline{Y}_n - \mu}{\sigma/\sqrt{n}}\right) = \left(\frac{\sum_{i=1}^n Y_i - n\mu}{\sqrt{n}\sigma}\right)$ is the z–score of $\overline{X}_n$ (and the z-score of $\sum_{i=1}^n Y_i$), and $Z_n \xrightarrow{D} N(0,1)$. b) Two applications of the CLT are to give the limiting distribution of $\sqrt{n}(\overline{Y}_n - \mu)$ and the limiting distribution of $\sqrt{n}(Y_n/n - \mu_Y)$ for a random variable $Y_n$ such that $Y_n = \sum_{i=1}^n X_i$ where the $X_i$ are iid with $E(X) = \mu_X$ and $V(X) = \sigma_X^2$. See Section 1.4. c) The CLT is the Lindeberg-Lévy CLT.

3) **Delta Method**: If $g'(\theta) \neq 0$, and

$$\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2), \quad \text{then} \quad \sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2[g'(\theta)]^2).$$

4) **Second Order Delta Method.** Suppose that $g'(\theta) = 0$, $g''(\theta) \neq 0$ and

$$\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \tau^2(\theta)). \text{ Then } n[g(T_n) - g(\theta)] \xrightarrow{D} \frac{1}{2}\tau^2(\theta)g''(\theta)\chi_1^2.$$

5) $X_n \xrightarrow{D} X$ if

$$\lim_{n\to\infty} F_n(t) = F(t)$$

at each continuity point $t$ of F. **Convergence in distribution** is also known as weak convergence and convergence in law. $X$ is the limiting distribution or asymptotic distribution of $X_n$. **The limiting distribution does not depend on** the sample size $n$. $X_n \xrightarrow{D} \tau(\theta)$ if $X_n \xrightarrow{D} X$ where $P(X = \tau(\theta)) = 1$: hence $X$ is *degenerate at* $\tau(\theta)$ or the distribution of $X$ is a *point mass at* $\tau(\theta)$.

6) If $X_n \xrightarrow{D} X$ and $X_n \xrightarrow{D} Y$, then i) $X \stackrel{D}{=} Y$ and ii) $F_X(x) = F_Y(x)$ for all real $x$.

7) **Convergence in probability**: a) $X_n \xrightarrow{P} \tau(\theta)$ if for every $\epsilon > 0$,

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

b) $X_n \xrightarrow{P} X$ if for every $\epsilon > 0$,

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

8) $T_n$ is a **consistent estimator** of $\tau(\theta)$ if $T_n \xrightarrow{P} \tau(\theta)$ for every $\theta \in \Theta$.

9) Theorem: $T_n$ is a **consistent estimator** of $\tau(\theta)$ if any of the following 2 conditions holds:

i) $\lim_{n\to\infty} V_\theta(T_n) = 0$ and $\lim_{n\to\infty} E_\theta(T_n) = \tau(\theta)$ for all $\theta \in \Theta$.

ii) $MSE_{\tau(\theta)}(T_n) = E[(T_n - \tau(\theta))^2] \to 0$ for all $\theta \in \Theta$.

Here

$$MSE_{\tau(\theta)}(T_n) = V_\theta(T_n) + [\text{Bias}_{\tau(\theta)}(T_n)]^2$$

where $\text{Bias}_{\tau(\theta)}(T_n) = E_\theta(T_n) - \tau(\theta)$.

10) $Y_n$ **converges in $r$th mean** to a random variable $Y$, $Y_n \xrightarrow{r} Y$, if

$$E(|Y_n - Y|^r) \to 0$$

as $n \to \infty$. In particular, if $r = 2$, $Y_n$ **converges in quadratic mean** to $Y$, written

$$Y_n \xrightarrow{2} Y \quad \text{or} \quad \mathrm{Y_n} \xrightarrow{\mathrm{qm}} \mathrm{Y},$$

if $E[(Y_n - Y)^2] \to 0$ as $n \to \infty$. $Y_n \xrightarrow{r} \tau(\theta)$ if $E(|Y_n - \tau(\theta)|^r) \to 0$ as $n \to \infty$. If $r \geq 1$, $Y_n \xrightarrow{r} Y$ is often written as $Y_n \xrightarrow{L^r} Y$ or $Y_n \xrightarrow{L_r} Y$.

11) a) A sequence of random variables $X_n$ *converges with probability 1* to $X$, written $X_n \xrightarrow{wp1} X$ if

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

b) $X_n \xrightarrow{wp1} \tau(\theta)$ if $P(\lim_{n\to\infty} X_n = \tau(\theta)) = 1$.

Note: Convergence wp1 is also known as a) strong convergence, b) $X_n$ *converges almost everywhere* to $X$, written $X_n \xrightarrow{ae} X$, or c) $X_n$ *converges almost surely* to $X$, written $X_n \xrightarrow{as} X$.

12) Let $Y_1, ..., Y_n, ...$ be a sequence of iid random variables with $E(Y_i) = \mu$.

a) **WLLN:** Then $\overline{Y}_n \xrightarrow{P} \mu$.

b) **SLLN:** Then $\overline{Y}_n \xrightarrow{wp1} \mu$.

Note: Hence $\overline{Y}_n$ is a consistent estimator of $\mu$ if the convergence is for all $\mu$ in the parameter space.

13) $F(t) = P(X \leq t)$. a) $0 \leq F(t) \leq 1$ for all real $t$, so $\lim\limits_{n\to\infty} F_n(t) = H(t)$ has $0 \leq H(t) \leq 1$ if the limit exists. b) $F(-\infty) = \lim\limits_{n\to-\infty} F_n(t) = 0$, c) $F(\infty) = \lim\limits_{n\to\infty} F_n(t) = 1$, d) $F(t)$ is right continuous: $\lim\limits_{h \downarrow 0} F(t+h) = F(t)$ for all real $t$. e) There are at most countably many discontinuity points of $F(t)$.

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

$$P[u(Y) \geq c] \leq \frac{E[u(Y)]}{c}.$$

If $\mu = E(Y)$ exists, then taking $u(y) = |y - \mu|^r$ and $\tilde{c} = c^r$ gives
**Markov's Inequality:** for $r > 0$ and any $c > 0$,

$$P(|Y - \mu| \geq c] = P(|Y - \mu|^r \geq c^r] \leq \frac{E[|Y - \mu|^r]}{c^r}$$

if the expected value exists. If $r = 2$ and $\sigma^2 = V(Y)$ exists, then we obtain
**Chebyshev's Inequality:**

$$P(|Y - \mu| \geq c] \leq \frac{V(Y)}{c^2}.$$

15) Let $k > 0$. If $E(X^k)$ is finite, then $E(X^j)$ is finite for $0 < j \leq k$. Moments use $j, k$ positive integers. If the mgf $m_X(t)$ exists, then $E(X^k)$ is finite for every positive integer $k$.

16) **Jensen's Inequality**: $g[E(X)] \le E[g(X)]$ if the expected values exist and $g$ is convex on an interval containing the range of $X$.

Note: A sufficient condition for a function $g$ to be convex on an open interval $(a, b)$ is $g''(x) > 0$ on $(a, b)$. If $g$ is convex on $(a, b)$ and continuous on $[a, b)$, then $g$ is convex on $[a, b)$.

17) **Theorem**: If $X_n \xrightarrow{r} X$, then $X_n \xrightarrow{k} X$ where $0 < k < r$.

18) **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) $X_n \xrightarrow{P} \tau(\theta)$ **iff** $X_n \xrightarrow{D} \tau(\theta)$.

e) If $X_n \xrightarrow{D} X$ and $X_n \xrightarrow{D} Y$, then $X \stackrel{D}{=} Y$ and $F_X(x) = F_Y(x)$ for all real $x$.

Note: If $X_n \xrightarrow{a} X$ and $X_n \xrightarrow{a} Y$, then $X \stackrel{D}{=} Y$ where $a$ is $wp1$, $r$, or $P$.

19) **Standard Limit Theorem:** Let $\hat{\theta}_n$ be the MLE or UMVUE of $\theta$. If $\tau'(\theta) \neq 0$, then under strong regularity conditions,

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

Note: Suppose $Y_1, ..., Y_n$ are iid with pdf $f(y|\theta)$. Suppose $\hat{\theta}_n$ is the MLE of $\theta$. a) 19) holds for a 1P-REF with pdf $f(y|\theta)$. b) 19) holds if the first and second derivatives can be interchanged in the following integrals:

$$0 = \frac{d}{d\theta}\int_{-\infty}^{\infty} f(y|\theta)dy = \int_{-\infty}^{\infty} \frac{d}{d\theta} f(y|\theta)dy \text{ and } 0 = \frac{d^2}{d\theta^2}\int_{-\infty}^{\infty} f(y|\theta)dy = \int_{-\infty}^{\infty} \frac{d^2}{d\theta^2} f(y|\theta)dy.$$

20) See 0) for a) $I_1(\rho)$ if $Y \sim bin(k, \rho)$ with $k$ known, b) $I_1(\lambda)$ if $Y \sim EXP(\lambda)$, c) $I_1(\lambda)$ if $Y \sim G(\nu, \lambda)$ with $\nu$ known, d) $I_1(\mu)$ if $Y \sim N(\mu, \sigma^2)$ with $\sigma$ known, e) $I_1(\sigma^2)$ if $Y \sim N(\mu, \sigma^2)$ with $\mu$ known, and f) $I_1(\theta)$ if $Y \sim POIS(\theta)$. These are 1P-REFs.

21) If $Y_1, ..., Y_n$ are iid from a 1P-REF, then

a) $I_n(\theta) = nI_1(\theta)$.

b)

$$I_n(\tau(\theta)) = \frac{nI_1(\theta)}{[\tau'(\theta)]^2} \text{ so } I_1(\tau(\theta)) = \frac{I_1(\theta)}{[\tau'(\theta)]^2}.$$

c)

$$I_1(\theta) = -E_\theta\left[\frac{d^2}{d\theta^2}\log(f(Y|\theta))\right].$$

22) Let $Y_1, ..., Y_n$ be the data. Let $Y_{(1)} \le Y_{(2)} \le \cdots \le Y_{(n)}$ be the **order statistics**. The notation $Y_{i:n} = Y_{(i)}$ is often used.

23) **Limit Theorem for the Sample Median:** Let $Y_1, ..., Y_n$ be iid with a pdf that satisfies $f(MED(Y)) > 0$ where $MED(Y)$ is the population median. Let $MED(n)$ be the sample median. Then

$$\sqrt{n}(MED(n) - MED(Y)) \xrightarrow{D} N\left(0, \frac{1}{4f^2(MED(Y))}\right).$$

24) If $\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma_A^2)$, then $T_n \approx N\left(0, \dfrac{\sigma_A^2}{n}\right)$, written $T_n \sim AN\left(0, \dfrac{\sigma_A^2}{n}\right)$, an approximate distribution. The limiting distribution $N(0, \sigma_A^2)$ does not depend on the sample size $n$, but the approximate distribution does depend on $n$.

25) Under the conditions of the CLT, $\overline{Y}_n \sim AN\left(0, \dfrac{S^2}{n}\right)$, where the sample variance $S^2 = S_n^2 = \dfrac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2$. The method of moments estimator of the variance is $S_m^2 = S_{m,n}^2 = \dfrac{n-1}{n} S_n^2 = \dfrac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2$.

26) If $\sqrt{n}(T_n - \mu_T) \xrightarrow{D} N(0, \sigma_A^2)$, then the *asymptotic variance* of $T_n$ is $\sigma_A^2/n$. Here $\sigma_A^2 = \sigma_A^2(F)$ depends on the distribution of the data, e.g. $Y_1, ..., Y_n$ are iid with cdf $F$. If $S_A^2$ is a consistent estimator of $\sigma_A^2$, then the (asymptotic) *standard error* of $T_n$ is $SE(T_n) = S_A/\sqrt{n}$. Hence $T_n \sim AN\left(0, \dfrac{S_A^2}{n}\right)$.

Note: Often $nV(T_n) \to \sigma_A^2$, but not always.