EXAM3, FINAL REVIEW (and a review for some of the QUAL problems): **No notes will be allowed,** but you may bring a calculator. Memorize the pmf or pdf f, $E(Y)$ and V(Y) for the following RVs: 1) beta(δ, ν), 2) Bernoulli(ρ) = bin$(k = 1, \rho)$, 3) binomial(k, ρ), 4) Cauchy(μ, σ), 5) chi-square(p) = gamma$(\nu = p/2, \lambda = 2)$, 6) exponential(λ) = gamma$(\nu = 1, \lambda)$, 7) gamma(ν, λ), 8) $N(\mu, \sigma^2)$, 9) Poisson(θ), and 10) uniform(θ_1, θ_1). **Sufficient, minimal sufficient, and complete sufficient statistics will be on exam 2.**

Memorize the mgf of the binomial, χ^2_p, exponential, gamma, normal and Poisson distributions. You should memorize the cdf of the exponential and of the normal distribution $\Phi(\frac{y-\mu}{\sigma})$. Know how to get the uniform cdf.

Let CB stand for Casella and Berger (2002) and BD 1st ed. or BD for Bickel and Doksum (1977, 2007). Old for Olive (2008). Other references are for Olive (2014).

From chapters 1 and 2 (CB ch. 1, 2, 3, 4, 5) you should know the sample space, conditional probability, random variables, cdfs, pmfs, pdfs, how to find the distribution of a function of a RV (Th. 2.13 p. 47; Old p. 51; CB th. 2.1.5, p. 51; BD p. 486), expected values, mgfs, the kernel method, the binomial theorem, the Gamma function, location, scale and location-scale families, random vectors, joint and marginal distributions, conditional distributions, independence, the law of iterated expectations (Th. 2.10, p. 43; Old p. 45; CB th. 4.4.3, p. 164; BD p. 481), Steiner's formula (p. 43; Old p. 45; CB th. 4.4.7, p. 167; BD p. 34), that the pdf of $Y = t(X)$ is $f_Y(y) = f_X(t^{-1}(y)) \left|\frac{dt^{-1}(y)}{dy}\right|$ for $y \in \mathcal{Y}$ (Th 2.13 p. 47), covariance, correlation, multivariate distributions, random sample (CB th. 5.2.4, p. 212), (Th 2.15, p. 52; Old p. 56-7; CB lemma 5.2.5, p. 213), (CB th. 5.2.6, p. 213-4), and (Th. 3.5, p. 92; Old p. 99; CB th. 5.2.11, p. 217; BD Th. 1.6.1, p. 51).

You should know the central limit theorem (Th. 8.1, p. 215; Old p. 203; CB p. 236; BD p. 470). Know the t and F distributions (CB 5.3.2). Know how to find the pdf of the min and the max (Th 4.2 p. 105, Old p. 110). Indicator functions (p. 110; Old p. 115; CB p. 113) are extremely important. Exponential families (Ch. 3; CB ∮ 3.4; BD ∮ 1.6) are extremely important.

Suppose $W = \sum_{i=1}^n Y_i$ or $W = \overline{Y}_n$ where $Y_1, ..., Y_i$ are independent. Be able to find the distribution of W if i) $Y_i \sim N(\mu_i, \sigma_i^2)$, ii) $Y_i \sim Ber(p)$, iii) $Y_i \sim bin(n_i, p)$, iv) $Y_i \sim neg.bin.(1, p)$, v) $Y_i \sim neg.bin.(n_i, p)$, vi) $Y_i \sim pois(\lambda)$, vii) $Y_i \sim exp(\lambda)$, viii) $Y_i \sim gamma(\alpha_i, \lambda)$, ix) $Y_i \sim \chi^2_{n_i}$.

Given the pdf of Y or that Y is a brand name distribution, know how to show whether Y belongs to an exponential family or not.

If Y belongs to an exponential family, know how to find the natural parameterization and the natural parameter space Ω.

If Y belongs to an exponential family, know how to show whether Y belongs to a k-parameter regular exponential family (kP-REF). In particular, if $k = 2$ you should be able to show whether η_1 and η_2 satisfy a linearity constraint (plotted points fall on a line) and whether t_1 and t_2 satisfy a linearity constraint, to plot Ω and to determine whether Ω contains a 2-parameter rectangle.

Section 4.1 (CB 5.3) is important, especially (Th 4.1, p. 103; Old p. 108; CB p. 218; BD p. 495). If $Y_1, ..., Y_n$ are iid $N(\mu, \sigma^2)$, know that $\overline{Y}$ and S^2 are **independent,** $\overline{Y} \sim N(\mu, \sigma^2/n)$ and $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$. Use these facts to find the distribution of S^2. Know that $(\overline{Y} - \mu)/(S/\sqrt{n}) \sim t_{n-1}$.

Dr. Olive usually uses problems from EIGHT basic questions for Exam 3, the final and the qual. Other topics can be added and the qual is written by at least 2 professors and often has additional topics.

1) Minimal sufficient and complete statistics
2) MLE
3) Method of Moments
4) Minimize MSE
5) UMVUE and FCRLB
6) UMP TEST with Neyman Pearson Lemma and exponential family theory
7) LRT
8) Large Sample Theory (CLT, Limiting Distribution of the MLE, Delta method, consistency and asymptotic efficiency)

---

1) **Minimal Sufficient and Complete Statistics**

(p. 108; Old p. 113; CB p. 272): A statistic $\boldsymbol{T}(Y_1, ..., Y_n)$ is a **sufficient statistic** for $\boldsymbol{\theta}$ if the conditional distribution of $(Y_1, ..., Y_n)$ given $\boldsymbol{T}$ does not depend on $\boldsymbol{\theta}$.

(p. 111; Old p. 116; CB p. 280; BD p. 46): A sufficient statistic $\boldsymbol{T}(\boldsymbol{Y})$ is a **minimal sufficient statistic** if for any other sufficient statistic $\boldsymbol{S}(\boldsymbol{Y})$, $\boldsymbol{T}(\boldsymbol{Y})$ is a function of $\boldsymbol{S}(\boldsymbol{Y})$.

(p. 112; Old p. 116; CB p. 285): Suppose that a **statistic** $\boldsymbol{T}(\boldsymbol{Y})$ has a pmf or pdf $f(\boldsymbol{t}|\boldsymbol{\theta})$. Then $\boldsymbol{T}(\boldsymbol{Y})$ is a **complete sufficient statistic** if $E_{\boldsymbol{\theta}}[g(\boldsymbol{T})] = 0$ for all $\boldsymbol{\theta}$ implies that
$P_{\boldsymbol{\theta}}[g(\boldsymbol{T}(\boldsymbol{Y})) = 0] = 1$ for all $\boldsymbol{\theta}$.

(p. 114; Old p. 119; CB p. 280, 282): A one to one function of a sufficient, minimal sufficient, or complete sufficient statistic is sufficient, minimal sufficient, or complete sufficient respectively.

Factorization Theorem, (p. 108; Old p. 113; CB p. 276; BD p. 43): Let $f(\boldsymbol{y}|\boldsymbol{\theta})$ denote the pdf or pmf of a sample $\boldsymbol{Y}$. A statistic $\boldsymbol{T}(\boldsymbol{Y})$ is a sufficient statistic for $\boldsymbol{\theta}$ iff for all sample points $\boldsymbol{y}$ and for all parameter points $\boldsymbol{\theta} \in \Theta$,

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = g(\boldsymbol{T}(\boldsymbol{y})|\boldsymbol{\theta})h(\boldsymbol{y})$$

where both g and h are nonnegative functions.

Note: if no such factorization exists for $\boldsymbol{T}$, then $\boldsymbol{T}$ is not sufficient.

Lehmann-Scheffé (LSM) Theorem (p. 112; Old p. 116; CB p. 281): Let $f(\boldsymbol{y}|\boldsymbol{\theta})$ be the pmf or pdf of a sample $\boldsymbol{Y}$. Let $c_{\boldsymbol{x},\boldsymbol{y}}$ be a constant. Suppose there exists a function $\boldsymbol{T}(\boldsymbol{y})$ such that for any two sample points $\boldsymbol{x}$ and $\boldsymbol{y}$, the ratio $R_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{\theta}) = f(\boldsymbol{x}|\boldsymbol{\theta})/f(\boldsymbol{y}|\boldsymbol{\theta}) = c_{\boldsymbol{x},\boldsymbol{y}}$ for all $\boldsymbol{\theta}$ in Θ iff $\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{T}(\boldsymbol{y})$. Then $\boldsymbol{T}(\boldsymbol{Y})$ is a minimal sufficient statistic for $\boldsymbol{\theta}$.

Minimal and complete sufficient statistics for k-parameter exponential families (Th. 4.5, p. 112; Old p. 117; CB p. 279, 288): Let $Y_1, ..., Y_n$ be iid from an exponential family $f(y|\boldsymbol{\theta}) = h(y)c(\boldsymbol{\theta}) \exp[\sum_{j=1}^k w_j(\boldsymbol{\theta}) t_j(y)]$ with the natural parameterization $f(y|\boldsymbol{\eta}) = h(y)b(\boldsymbol{\eta}) \exp[\sum_{j=1}^k \eta_j t_j(y)]$. Then $\boldsymbol{T}(\boldsymbol{Y}) = (\sum_{i=1}^n t_1(Y_i), ..., \sum_{i=1}^n t_k(Y_i))$ is

a) a minimal sufficient statistic for $\boldsymbol{\eta}$ if the η_j do not satisfy a linearity constraint and for $\boldsymbol{\theta}$ if the $w_j(\boldsymbol{\theta})$ do not satisfy a linearity constraint.

b) a complete sufficient statistic for $\boldsymbol{\theta}$ and for $\boldsymbol{\eta}$ if $\boldsymbol{\eta}$ is a one to one function of $\boldsymbol{\theta}$ and if Ω contains a k–dimensional rectangle.

Completeness of REFs (Cor. 4.6, p. 114; Old p. 118; CB th. 6.2.25, p. 288; BD p. 123 1st ed.): Suppose that $Y_1, ..., Y_n$ are iid from a kP–REF

$$f(y|\boldsymbol{\theta}) = h(y)c(\boldsymbol{\theta}) \exp\left[w_1(\boldsymbol{\theta})t_1(y) + \cdots + w_k(\boldsymbol{\theta})t_k(y)\right]$$

with $\boldsymbol{\theta} \in \Theta$, and $f(y|\boldsymbol{\eta}) = h(y)b(\boldsymbol{\eta}) \exp[\sum_{j=1}^k \eta_j t_j(y)]$ with natural parameter $\boldsymbol{\eta} \in \Omega$. Then

$$\boldsymbol{T}(\boldsymbol{Y}) = (\sum_{i=1}^n t_1(Y_i), ..., \sum_{i=1}^n t_k(Y_i)) \ \text{ is}$$

a) a minimal sufficient statistic for $\boldsymbol{\eta}$ and for $\boldsymbol{\theta}$,

b) a complete sufficient statistic for $\boldsymbol{\theta}$ and for $\boldsymbol{\eta}$ if $\boldsymbol{\eta}$ is a one to one function of $\boldsymbol{\theta}$ and if Ω contains a k–dimensional rectangle.

For a 2-parameter exponential family ($k = 2$), η_1 and η_2 satisfy a linearity constraint if the plotted points fall on a line in a plot of η_1 versus η_2. If the plotted points fall on a nonlinear curve, then $\boldsymbol{T}$ is minimal sufficient but Ω does not contain a 2-dimensional rectangle.

Tips for finding sufficient, minimal sufficient and complete sufficient statistics. a) Typically $Y_1, ..., Y_n$ are iid so the joint distribution $f(y_1, ..., y_n) = \prod_{i=1}^n f(y_i)$ where $f(y_i)$ is the marginal distribution. Use the **factorization theorem** to find the candidate sufficient statistic $\boldsymbol{T}$.

b) Use factorization to find candidates $\boldsymbol{T}$ that might be minimal sufficient statistics. Try to find $\boldsymbol{T}$ with as small a dimension k as possible. If the support of the random variable depends on θ, often $Y_{(1)}$ or $Y_{(n)}$ will be a component of the minimal sufficient statistic. To prove that $\boldsymbol{T}$ is minimal sufficient, use the **LSM theorem**. **Alternatively prove or recognize that Y comes from a regular exponential family.** $\boldsymbol{T}$ will be minimal sufficient for $\boldsymbol{\theta}$ if Y comes from an exponential family as long as the $w_i(\boldsymbol{\theta})$ do not satisfy a linearity constraint.

c) **To prove that the statistic is complete, prove or recognize that Y comes from a regular exponential family**. Check whether $\dim(\Theta) = k$, if $\dim(\Theta) < k$, then the family is usually not a kP–REF and Th. 4.5 and Cor. 4.6 do not apply. The uniform distribution where one endpoint is known also has a complete sufficient statistic.

d) Let k be free of the sample size n. Then a $k-$dimensional complete sufficient statistic is also a minimal sufficient statistic (**Bahadur's theorem**).

e) To show that a statistic $\boldsymbol{T}$ is not a sufficient statistic, either show that factorization fails or find a minimal sufficient statistic $\boldsymbol{S}$ and show that $\boldsymbol{S}$ is not a function of $\boldsymbol{T}$.

f) To show that $\boldsymbol{T}$ is not minimal sufficient, first try to show that $\boldsymbol{T}$ is not a sufficient statistic. If $\boldsymbol{T}$ is sufficient, find a minimal sufficient statistic $\boldsymbol{S}$ and show that $\boldsymbol{T}$ is not a function of $\boldsymbol{S}$. (Of course $\boldsymbol{S}$ will be a function of $\boldsymbol{T}$.) The **Lehmann-Scheffé (LSM) theorem cannot be used to show that a statistic is not minimal sufficient.**

g) To show that a sufficient statistics $\boldsymbol{T}$ is not complete, find a function $g(\boldsymbol{T})$ such that $E_{\boldsymbol{\theta}}(g(\boldsymbol{T})) = 0$ for all $\boldsymbol{\theta}$ but $g(\boldsymbol{T})$ is not equal to the zero with probability one. Finding such a g is often hard, unless there are clues. For example, if $\boldsymbol{T} = (\overline{X}, \overline{Y}, ...)$ and $\mu_1 = \mu_2$, try $g(\boldsymbol{T}) = \overline{X} - \overline{Y}$. As a **rule of thumb**, a k–dimensional minimal sufficient statistic will generally not be complete if $k > dim(\Theta)$. In particular, if $\boldsymbol{T}$ is k–dimensional and $\boldsymbol{\theta}$ is j–dimensional with $j < k$ (especially $j = 1 < 2 = k$) then $\boldsymbol{T}$ will **generally not be complete.** If you can show that a k–dimensional sufficient statistic $\boldsymbol{T}$ is not minimal sufficient (often hard), then $\boldsymbol{T}$ is not complete by Bahadur's Theorem. Basu's Theorem can sometimes be used to show that a minimal sufficient statistic is not complete.

A **common** question takes $Y_1, ..., Y_n$ iid $U(h_l(\theta), h_u(\theta))$ where the min $= Y_{(1)}$ and the max $= Y_{(n)}$ form the 2-dimensional minimal sufficient statistic. Since θ is one dimensional, the minimal sufficient statistic is not complete. State this fact, but if you have time find $E_\theta[Y_{(1)}]$ and $E_\theta[Y_{(n)}]$. Then show that $E_\theta[aY_{(1)} + bY_{(n)} + c] \equiv 0$ so that $\boldsymbol{T} = (Y_{(1)}, Y_{(n)})$ is not complete.

---

2) **MLEs**

Def. (p. 129; Old 133; CB p. 290; BD p. 47): Let $f(\boldsymbol{y}|\boldsymbol{\theta})$ be the pmf or pdf of a sample $\boldsymbol{Y}$. If $\boldsymbol{Y} = \boldsymbol{y}$ is observed, then **the likelihood function** $L(\boldsymbol{\theta}) = f(\boldsymbol{y}|\boldsymbol{\theta})$.

Note: it is crucial to observe that the likelihood function is a function of $\boldsymbol{\theta}$ (and that $y_1, ..., y_n$ act as fixed constants).

Note: If $Y_1, ..., Y_n$ is an independent sample from a population with pdf or pmf $f(y|\boldsymbol{\theta})$ then the likelihood function

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}|y_1, ..., y_n) = \prod_{i=1}^n f(y_i|\boldsymbol{\theta}).$$

Def. (p. 129, Old 133; CB p. 316, BD p. 114): Let $\boldsymbol{Y} = (Y_1, ..., Y_n)$. For each sample point $\boldsymbol{y} = (y_1, ..., y_n)$, let $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ be a parameter value at which $L(\boldsymbol{\theta}|\boldsymbol{y})$ attains its maximum as a function of $\boldsymbol{\theta}$ with $\boldsymbol{y}$ held fixed. Then a maximum likelihood estimator (**MLE**) of the parameter $\boldsymbol{\theta}$ based on the sample $\boldsymbol{Y}$ is $\hat{\boldsymbol{\theta}}(\boldsymbol{Y})$.

Note: If the MLE $\hat{\boldsymbol{\theta}}$ exists, then $\hat{\boldsymbol{\theta}} \in \Theta$.

There are **four commonly used techniques** for finding the MLE.

- Potential candidates can be found by differentiating log $L(\boldsymbol{\theta}|\boldsymbol{y})$, the log likelihood.
- Potential candidates can be found by differentiating the likelihood.
- The MLE can sometimes be found by direct maximization of the likelihood. (Sketching the likelihood function and (CB th. 5.2.4, p. 212) can be useful for this method.)
- **Invariance Principle** (p. 130; Old 134; CB p. 320; BD p. 114): If

- $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$, then $h(\hat{\boldsymbol{\theta}})$ is the MLE of $h(\boldsymbol{\theta})$.

You should know how to find the MLE for the normal distribution (including when μ or σ^2 is known, memorize the MLEs $\overline{Y}$, $S_M^2 = \sum_{i=1}^n (Y_i - \overline{Y})^2/n$, $\sum_{i=1}^n (Y_i - \mu)^2/n$) and for the uniform distribution. Also $\overline{Y}$ is the MLE for several brand name distributions.

Know how to find the max and min of a function h that is continuous on an interval [a,b] and differentiable on (a,b). Solve $h'(x) \equiv 0$ and find the places where $h'(x)$ does not exist. These values are the **critical points**. Evaluate h at a, b, and the critical points. One of these values will be the min and one the max.

Assume h is continuous. Then a critical point θ_o is a local max of $h(\theta)$ if h is increasing for $\theta < \theta_o$ in a neighborhood of θ_o and if h is decreasing for $\theta > \theta_o$ in a neighborhood of θ_o. The first derivative test is often used.

If h is strictly concave ($\frac{d^2}{d\theta^2} h(\theta) < 0$ for all θ), then any local max of h is a global max.

Suppose $h'(\theta_o) = 0$. The 2nd derivative test states that if $\frac{d^2}{d\theta^2} h(\theta_o) < 0$, then θ_o is a local max.

(p. 131; Old 135-6; CB p. 317). If $h(\theta)$ is a continuous function on an interval with endpoints $a < b$ (not necessarily finite), and differentiable on (a,b) and if the **critical point is unique**, then the critical point is a **global maximum** if it is a local maximum (because otherwise there would be a local minimum and the critical point would not be unique). To show that $\hat{\theta}$ is the MLE (the global maximizer of $\log L(\theta)$), show that $\log L(\theta)$ is differentiable on (a,b) where Θ may contain the endpoints a and b. Then show that $\hat{\theta}$ is the unique solution to the equation $\frac{d}{d\theta} \log L(\theta) = 0$ and that the 2nd derivative evaluated at $\hat{\theta}$ is negative: $\frac{d^2}{d\theta^2} \log L(\theta)|_{\hat{\theta}} < 0$.

Suppose $X_1, ..., X_n$ are iid with pdf or pmf $f(x|\lambda)$ and $Y_1, ..., Y_n$ are iid with pdf or pmf $g(y|\mu)$. Suppose that the X's are independent of the Y's. Then

$$\sup_{(\lambda,\mu)\in\Theta} L(\lambda, \mu | \boldsymbol{x}, \boldsymbol{y}) \le \sup_{\lambda} L_{\boldsymbol{x}}(\lambda) \sup_{\mu} L_{\boldsymbol{y}}(\mu)$$

where $L_{\boldsymbol{x}}(\lambda) = \prod_{i=1}^n f(x_i|\lambda)$. Hence if $\hat{\lambda}$ is the marginal MLE of λ and $\hat{\mu}$ is the marginal MLE of μ, then $(\hat{\lambda}, \hat{\mu})$ is the MLE of (λ, μ) provided that $(\hat{\lambda}, \hat{\mu})$ is in the parameter space Θ.

Note: Finding the potential candidates for the MLE will get a lot of partial credit. Sometimes showing that the MLE is actually the global max is unreasonable. Make an attempt to show that the MLE is a global max, but do not waste much time if you get stuck. On the other hand, you should always evaluate $L(\theta)$ or $\log L(\theta)$ at the endpoints a and b of $\Theta = [a, b]$.

(CB p. 322) shows how to use the Hessian to determine that $(\hat{\theta}_1, \hat{\theta}_2)$ is a local max. This is a very involved calculation and should be avoided if possible.

MLE for a REF (Barndorff–Nielsen 1982): Suppose that the natural parameterization of the k-parameter regular exponential family is used so that Ω is a k-dimensional convex set (usually an open interval or cross product of open intervals). Then the log likelihood function log $L(\boldsymbol{\eta})$ is a strictly concave function of $\boldsymbol{\eta}$. Hence if $\hat{\boldsymbol{\eta}}$ is a critical point of log L($\boldsymbol{\eta}$) **and if** $\hat{\boldsymbol{\eta}} \in \Omega$ then $\hat{\boldsymbol{\eta}}$ is the unique MLE of $\boldsymbol{\eta}$. (The Hessian matrix of 2nd derivatives does not need to be checked!) If $\boldsymbol{\eta}$ is a one to one function of $\boldsymbol{\theta}$, then $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$ by invariance.

Note: the MLE is usually a function of the minimal sufficient statistic.

On the qual, the $N(\mu, \mu)$ and $N(\mu, \mu^2)$ distributions are common. (See problems 5.30 and 5.35.)

---

3) **Method of Moments**

See ∮ 5.2. Let $\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^n Y_i^j$, let $\mu_j = E(Y^j)$ and assume that $\mu_j = \mu_j(\theta_1, ..., \theta_k)$. Solve the system

$$\begin{aligned} \hat{\mu}_1 &\stackrel{\text{set}}{=} \mu_1(\theta_1, ..., \theta_k) \\ &\vdots \\ \hat{\mu}_k &\stackrel{\text{set}}{=} \mu_k(\theta_1, ..., \theta_k) \end{aligned}$$

for the method of moments estimator $\tilde{\boldsymbol{\theta}}$.

If g is a continuous function of the first k moments and $h(\boldsymbol{\theta}) = g(\mu_1(\boldsymbol{\theta}), ..., \mu_k(\boldsymbol{\theta}))$, then the method of moments estimator of $h(\boldsymbol{\theta})$ is
$g(\hat{\mu}_1, ..., \hat{\mu}_k)$.

If the method of moments estimator is a sum or sample mean $T = \sum_{i=1}^n W_i$ or $T = \sum_{i=1}^n W_i/n$, you may need to find the limiting distribution of $\sqrt{n}(T - E(T))$ using the central limit theorem. See 8).

---

4) **Minimizing MSE**

Def. (p. 157, Old p. 160) The **bias** of an estimator $T \equiv T(Y_1, ..., Y_n)$ of $\tau(\theta)$ is $B_{\tau(\boldsymbol{\theta})}(T) \equiv \text{Bias(T)} \equiv \text{Bias}_{\tau(\boldsymbol{\theta})}\text{(T)} = \text{E}_{\boldsymbol{\theta}}\text{(T)} - \tau(\boldsymbol{\theta})$.

Def. (p. 157, Old p. 160) The **MSE** of an estimator T for $\tau(\theta)$ is

$$MSE = E_\theta[(T - \tau(\theta))^2] = Var_{\boldsymbol{\theta}}(T) + [\text{Bias}_{\tau(\boldsymbol{\theta})}\text{(T)}]^2.$$

Def. (p. 157, Old p. 160) T is an **unbiased estimator** of $\tau(\theta)$ if $E_\theta(T) = \tau(\theta)$ for all $\theta \in \Theta$.

For this type of problem, consider a class of estimators $T_k(\boldsymbol{Y})$ of $\tau(\theta)$ where $k \in \Lambda$. Find the MSE as a function of k and then find the value $k_o \in \Lambda$ that is the global minimizer of MSE(k). This type of problem is a lot like the MLE problem except you need to find the global min rather than the global max.

If $Y_1, ..., Y_n$ are iid $N(\mu, \sigma^2)$ then $k_o = n + 1$ will minimize the MSE for estimators of σ^2 of the form

$$S^2(k) = \frac{1}{k}\sum_{i=1}^n (Y_i - \overline{Y})^2$$

where $k > 0$.

This type of problem can be done if $T_k = kS_1(\boldsymbol{Y}) + (1-k)S_2(\boldsymbol{Y})$ where both S_1 and S_2 are unbiased estimators of $\tau(\theta)$ and $0 \leq k \leq 1$.

---

5) **UMVUEs and FCRLB** for Unbiased Estimators of a real valued function $\tau(\boldsymbol{\theta})$

Def. (p. 160; Old p. 163; CB p. 334): Let $U \equiv U(Y_1, ..., Y_n)$ be an estimator of a real valued function $\tau(\theta)$. Then U is the **UMVUE** of $\tau(\theta)$ if U is an unbiased estimator of $\tau(\theta)$ and if $\text{VAR}_\theta(U) \leq \text{VAR}_\theta(\text{W})$ for all $\theta \in \Theta$ where W is any other unbiased estimator of $\tau(\theta)$.

Lehmann-Scheffé LSU Th. (p. 160; Old p. 163; CB p. 347, 369; BD p. 122, 1st ed.): If $T(\boldsymbol{Y})$ is a complete sufficient statistic for θ, then $U = g(T(\boldsymbol{Y}))$ *is the UMVUE* of $\tau(\theta) = E_\theta(U) = E_\theta[g(T(\boldsymbol{Y}))]$. In particular, if $W(\boldsymbol{Y})$ is *any unbiased estimator* of $\tau(\theta)$, then $U \equiv E[W(\boldsymbol{Y})|T(\boldsymbol{Y})]$ is the UMVUE of $\tau(\theta)$.

Note: This process is also called Rao-Blackwellization because of the following theorem.

Rao-Blackwell th: Let $W \equiv W(\boldsymbol{Y})$ be an unbiased estimator of $\tau(\theta)$ and let $T \equiv T(\boldsymbol{Y})$ be a sufficient statistic for θ. Then $\phi(T) = E[W|T]$ is an unbiased estimator of $\tau(\theta)$ and $\text{VAR}_\theta[\phi(T)] \leq \text{VAR}_\theta(W)$ for all $\theta \in \Theta$.

The following th. is sometimes useful when no complete sufficient statistic is available.

(CB Th 7.3.20, p. 344): If W is an unbiased estimator of $\tau(\boldsymbol{\theta})$ then W is the UMVUE of $\tau(\boldsymbol{\theta})$ iff W is uncorrelated with all unbiased estimators U of zero. (The underlying distribution for the expectations is the distribution of W.)

Def. (p. 162; Old 164; CB p. 338; BD p. 180): Let $\boldsymbol{Y} = (Y_1, ..., Y_n)$ have a pdf or pmf $f(\boldsymbol{y}|\theta)$. Then the **information number** or **Fisher Information** is

$$I_n(\theta) = E_\theta\left(\left[\frac{\partial}{\partial\theta}\log(f(\boldsymbol{Y}|\theta))\right]^2\right).$$

Let $\eta = \tau(\theta)$ where $\tau'(\theta) \neq 0$. Then $I_n(\eta) \equiv I_n(\tau(\theta)) = \frac{I_n(\theta)}{[\tau'(\theta)]^2}$.

Let $Y_1, ..., Y_n$ be independent with joint pdf or pmf $f(\boldsymbol{y}|\theta) = \prod_{i=1}^n f(y_i|\theta)$. Then the **information number** of **Fisher Information** is

$$I_n(\theta) = E_\theta[(\frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(Y_i|\theta))^2].$$

Fact (p. 162; Old 165; CB p. 338): If Y comes from an exponential family, then

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

Fact (p. 162; Old 165; CB p. 338): If the derivative and integral operators can be interchanged, and if $Y_1, ..., Y_n$ are iid (ie the data are iid from a REF), then $I_n(\theta) = nI_1(\theta)$.

(p. 90; Old 97; CB Lemma p. 337): If Y comes from an exponential family, then

$$\frac{d}{d\theta}\int \cdots \int g(\boldsymbol{y}) f(\boldsymbol{y}|\theta) d\boldsymbol{y} = \int \cdots \int g(\boldsymbol{y}) \frac{\partial}{\partial\theta} f(\boldsymbol{y}|\theta) d\boldsymbol{y}$$

for any function $g(\boldsymbol{y})$ with $V_\theta[w(\boldsymbol{Y})] < \infty$. Replace integrals by sums for a pmf.

Fréchet Cramér Rao Lower Bound or Information Inequality: (p. 164; Old p. 167) Let $Y_1, ..., Y_n$ be iid with joint pdf or pmf $f(\boldsymbol{x}|\theta)$ that satisfies the above lemma. Let $W(Y_1, ..., Y_n)$ be any estimator of $\tau(\theta) \equiv E_\theta[W(\boldsymbol{Y})]$. Then

$$V_\theta(W(\boldsymbol{Y})) \geq FCRLB_n(\tau(\theta)) = \frac{[\frac{d}{d\theta} E_\theta W(\boldsymbol{Y})]^2}{E_\theta[(\frac{\partial}{\partial\theta} \log f(\boldsymbol{Y}|\theta))^2]} = \frac{[\tau'(\theta)]^2}{I_n(\theta)} = \frac{[\tau'(\theta)]^2}{nI_1(\theta)}.$$

The quantity $\dfrac{[\tau'(\theta)]^2}{nI_1(\theta)} = FCRLB_n(\tau(\theta))$ is the **Fréchet Cramér Rao lower bound** (FCRLB) for the variance of unbiased estimators of $\tau(\theta)$. F and Fréchet are often omitted.

Fact: if the family is not an exponential family, the FCRLB may **not be a lower bound** on the variance of unbiased estimators of $\tau(\theta)$.

Fact: Even if the sample is from a one parameter exponential family with complete sufficient statistic T, the FCRLB will typically hold with equality for linear functions of T, but not for nonlinear functions of T. (Recall that $U = g(T)$ is the UMVUE of its expectation $E_\theta(g(T))$ by the LSU theorem.)

Finding the UMVUE given a complete sufficient statistic T: The first method for finding the UMVUE of $\tau(\theta)$ is to guess g and show that $E_\theta[U(\boldsymbol{Y})] = E_\theta[g(T(\boldsymbol{Y}))] = \tau(\theta)$ for all θ. The second method is to find **any unbiased estimator** $W(\boldsymbol{Y})$ of $\tau(\theta)$. Then $U(\boldsymbol{Y}) = E[W(\boldsymbol{Y})|T(\boldsymbol{Y})]$ is the UMVUE of $\tau(\theta)$. For **full credit**, $E[W(\boldsymbol{Y})|T(\boldsymbol{Y})]$ needs to be computed.

Note: If you are asked to find the UMVUE of $\tau(\theta)$, see if an unbiased estimator $W(\boldsymbol{Y})$ is given in the problem. Also check whether you are asked to compute $E[W(\boldsymbol{Y})|T(\boldsymbol{Y}) = t]$ anywhere.

Note: This problem is typically very hard. Write down the two methods for finding the UMVUE for partial credit. If you can not guess g, find an unbiased estimator W, or compute $E[W|T]$, come back to the problem later.

The following facts can be useful for computing the conditional expectation. Suppose $Y_1, ..., Y_n$ are iid with finite expectation.
a) Then $E[Y_1|\sum_{i=1}^n Y_i = x] = x/n$.
b) If the Y_i are iid Poisson(θ), then $(Y_1|\sum_{i=1}^n Y_i = x) \sim bin(x, 1/n)$.
c) If the Y_i are iid Bernoulli Ber(ρ), then $(Y_1|\sum_{i=1}^n Y_i = x) \sim Ber(x/n)$.
d) If the Y_i are iid N(μ, σ^2), then $(Y_1|\sum_{i=1}^n Y_i = x) \sim N[x/n, \sigma^2(1 - 1/n)]$.

---

6) **UMP TESTS via the Neyman Pearson Lemma and Exponential Family Theory**

Def. (p. 183; Old p. 182; CB p. 383): A **type I error** is rejecting H_o when H_o is true, while a **type II error** is failing to reject H_o when H_o is false.

Def. (p. 183; Old p. 182; CB def. 8.3.1, p. 383): The **power function** of a hypothesis test with rejection region R is $\beta(\theta) = P_\theta(\boldsymbol{Y} \in R)$ for $\theta \in \Theta$. More generally, $\beta(\theta) = P_\theta(H_o \text{ is rejected})$.

Def. (p. 184; Old 183; CB p. 385): Let $0 \leq \alpha \leq 1$. Then a test with power function $\beta(\theta)$ is a **level α test** if

$$\sup_{\theta \in \Theta_o} \beta(\theta) \leq \alpha.$$

Def. (p. 185; Old 183; CB p. 388; BD p. 227): Consider all level α tests of $H_o : \theta \in \Theta_o$ vs $H_1 : \theta_1 \in \Theta_1$. A **uniformly most powerful** (UMP) level α test is a test with power function $\beta_{UMP}(\theta)$ such that $\beta_{UMP}(\theta) \geq \beta(\theta)$ for every $\theta \in \Theta_1$ where β is a power function for any level α test of H_o vs H_1.

One Sided UMP Tests for one parameter REFs, (p. 186; Old 185; see CB th. 8.3.17 p. 391 and th. 5.2.11 p. 217; see BD p. 228-229): Let $Y_1, ..., Y_n$ be iid with pdf or pmf

$$f(y|\theta) = h(y)c(\theta)\exp[w(\theta)t(y)]$$

from a one parameter exponential family where θ is real and $w(\theta)$ is increasing. Here $T(\boldsymbol{y}) = \sum_{i=1}^n t(y_i)$. I) Let $\theta_1 > \theta_o$. Consider the test that rejects H_o if $T(\boldsymbol{y}) > k$ and rejects H_o with probability γ if $T(\boldsymbol{y}) = k$ where $\alpha = P_{\theta_o}(T(\boldsymbol{Y}) > k) + \gamma P_{\theta_o}(T(\boldsymbol{Y}) = k)$. This test is the UMP test for
a) $H_o : \theta = \theta_o$ vs $H_1 : \theta = \theta_1$,
b) $H_o : \theta = \theta_o$ vs $H_1 : \theta > \theta_o$, and
c) $H_o : \theta \leq \theta_o$ vs $H_1 : \theta > \theta_o$.
II) Let $\theta_1 < \theta_o$. Consider the test that rejects H_o if $T(\boldsymbol{y}) < k$ and rejects H_o with probability γ if $T(\boldsymbol{y}) = k$ where $\alpha = P_{\theta_o}(T(\boldsymbol{Y}) < k) + \gamma P_{\theta_o}(T(\boldsymbol{Y}) = k)$. This test is the UMP test for
d) $H_o : \theta = \theta_o$ vs $H_1 : \theta = \theta_1$
e) $H_o : \theta = \theta_o$ vs $H_1 : \theta < \theta_o$, and
f) $H_o : \theta \geq \theta_o$ vs $H_1 : \theta < \theta_o$. As a mnemonic, note that the **inequality used in the "rejection region" is the same as the inequality in the alternative hypothesis.**

Note: **usually $\gamma = 0$ if f is a pdf.**

The Neyman Pearson Lemma, (p. 185; Old 184; CB p. 388; BD p. 224): Consider testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ where the pdf or pmf corresponding to θ_i is $f(\boldsymbol{y}|\theta_i)$ for $i = 0, 1$. Suppose the test rejects H_0 if $f(\boldsymbol{y}|\theta_1) > kf(\boldsymbol{y}|\theta_0)$, and rejects H_0 with probability γ if $f(\boldsymbol{y}|\theta_1) = kf(\boldsymbol{y}|\theta_0)$ for some $k \geq 0$. If

$$\alpha = \beta(\theta_0) = P_{\theta_0}[f(\boldsymbol{Y}|\theta_1) > kf(\boldsymbol{Y}|\theta_0)] + \gamma P_{\theta_0}[f(\boldsymbol{Y}|\theta_1) = kf(\boldsymbol{Y}|\theta_0)],$$

then this test is a UMP level α test.

Fact: **typically $\gamma = 0$ if f is a pdf**, but usually $\gamma > 0$ if f is a pmf.

Fact: To find an UMP test with the NP lemma, often the ratio $\dfrac{f(\boldsymbol{y}|\theta_1)}{f(\boldsymbol{y}|\theta_0)}$ is computed. The test will certainly reject H_o is the ratio is large, but usually the distribution of the ratio is not easy to use. Hence try to get an equivalent test by simplifying and transforming the ratio. Ideally, the ratio can be transformed into a statistic $T \equiv T(\boldsymbol{Y})$ whose distribution is tabled. If the test rejects H_o if $T > k$ and with probability γ if $T = k$, (or if $T < k$ and with probability γ if $T = k$) the test is in **useful form** if for a given α, k is also given. If you are asked to use a table, put the test in useful form. Often it is too hard to give the test in useful form. Then simply specify when the test rejects H_o and α in terms of k (eg $\alpha = P_{H_o}(T > k) + \gamma P_{H_o}(T = k)$).

Def. (p. 187; CB p. 388, 391): A **simple hypothesis** consists of exactly one distribution for the sample. A **composite hypothesis** consists of more than one distribution for the sample.

One Sided UMP Tests via NP lemma: (p. 186; Old p. 185) Suppose that the hypotheses are of the form $H_o : \theta \leq \theta_o$ vs $H_1 : \theta > \theta_o$ or $H_o : \theta \geq \theta_o$ vs $H_1 : \theta < \theta_o$, or that the inequality in H_o is replaced by equality. Also assume that $\sup_{\theta \in \Theta_0} \beta(\theta) = \beta(\theta_o)$. Pick $\theta_1 \in \Theta_1$ and use the Neyman Pearson lemma to find the UMP test for $H_o^* : \theta = \theta_o$ vs $H_A^* : \theta = \theta_1$. Then the UMP test rejects H_o^* if $f(\boldsymbol{y}|\theta_1) > kf(\boldsymbol{y}|\theta_o)$, and rejects H_o^* with probability γ if $f(\boldsymbol{y}|\theta_1) = kf(\boldsymbol{y}|\theta_o)$ for some $k \geq 0$ where $\alpha = \beta(\theta_o)$. This test is also the UMP level α test for $H_o : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ if *k does not depend on the value of* $\theta_1 \in \Theta_1$. Note that k does depend on α and θ_o. If $R = f(\boldsymbol{Y}|\theta_1)/f(\boldsymbol{Y}|\theta_o)$, then $\alpha = P_{\theta_o}(R > k) + \gamma P_{\theta_o}(R = k)$.

The **power** $\beta(\theta) = P_\theta(\text{reject } H_o)$ is the probability of rejecting H_o when θ is the true value of the parameter. Often the power function can not be calculated, but you should be prepared to calculate the power for a sample of size one for a test of the form $H_o : f(y) = f_0(y)$ versus $H_1 : f(y) = f_1(y)$ or if the test is of the form $\sum t(Y_i) > k$ or $\sum t(Y_i) < k$ when $\sum t(Y_i)$ has an easily handled distribution under H_1, eg binomial, normal, Poisson or χ^2_p. To compute the power, you need to find k and γ for the given value of α.

---

7) **Likelihood Ratio Tests**

Def. (p. 192; Old 190; CB p. 375, 386): Let $Y_1, ..., Y_n$ be the data with pdf or pmf $f(\boldsymbol{y}|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a vector of unknown parameters with parameter space Θ. Let $\hat{\boldsymbol{\theta}}$ be the MLE of $\boldsymbol{\theta}$ and let $\hat{\boldsymbol{\theta}}_o$ be the MLE of $\boldsymbol{\theta}$ if the parameter space is Θ_0 (where $\Theta_0 \subset \Theta$). A likelihood test (LRT) statistic for testing $H_o : \boldsymbol{\theta} \in \Theta_0$ versus $H_1 : \boldsymbol{\theta} \in \Theta_0^c$ is

$$\lambda(\boldsymbol{y}) = \frac{L(\hat{\boldsymbol{\theta}}_o|\boldsymbol{y})}{L(\hat{\boldsymbol{\theta}}|\boldsymbol{y})} = \frac{\sup_{\Theta_0} L(\boldsymbol{\theta}|\boldsymbol{y})}{\sup_{\Theta} L(\boldsymbol{\theta}|\boldsymbol{y})}.$$

The **likelihood ratio test** (LRT) has a rejection region of the form

$$R = \{\boldsymbol{y}|\lambda(\boldsymbol{y}) \leq c\}$$

where $0 \leq c \leq 1$, and $\alpha = \sup_{\boldsymbol{\theta} \in \Theta_0} P_{\boldsymbol{\theta}}(\lambda(\boldsymbol{Y}) \leq c)$. Suppose $\boldsymbol{\theta}_o \in \Theta_0$ and $\sup_{\boldsymbol{\theta} \in \Theta_0} P_{\boldsymbol{\theta}}(\lambda(\boldsymbol{Y}) \leq c) = P_{\boldsymbol{\theta}_o}(\lambda(\boldsymbol{Y}) \leq c)$. Then $\alpha = P_{\boldsymbol{\theta}_o}(\lambda(\boldsymbol{Y}) \leq c)$.

Fact: often $\Theta_o = (a, \theta_o]$ and $\Theta_1 = (\theta_o, b)$ or $\Theta_o = [\theta_o, b)$ and $\Theta_1 = (a, \theta_o)$.

Asymptotic Distribution of LRT, (p. 193; Old 190; CB th 10.3.3 p. 490): Let $Y_1, ..., Y_n$ be iid. Then under strong regularity conditions, $-2 \log \lambda(\boldsymbol{y}) \approx \chi^2_j$ for large n

where $j = r - q$, r is the number of free parameters specified by $\boldsymbol{\theta} \in \Theta$, and q is the number of free parameters specified by $\boldsymbol{\theta} \in \Theta_o$. Hence the approximate LRT rejects H_o if $-2\log\lambda(\boldsymbol{y}) > c$ where $P(\chi^2_j > c) = \alpha$. Thus $c = \chi^2_{j,1-\alpha}$.

Note: to find the LRT, find the two MLEs and write $L(\theta|\boldsymbol{y})$ in terms of a sufficient statistic. Simplify the statistic $\lambda(\boldsymbol{y})$ and state that the LRT test rejects H_o if $\lambda(\boldsymbol{y}) \leq c$ where $\alpha = P_{\boldsymbol{\theta}_o}(\lambda(\boldsymbol{y}) \leq c)$.

Note: The above rejection region is not in useful form. Sometimes you do not need to put the rejection region into a useful form, but often you do. Either you will use the above asymptotic distribution, or you can simplify λ or transform λ so that the test rejects H_o if some statistic $T > k$ (or $T < k$). Getting the test into useful form can be very difficult. Monotone transformations such as log or power transformations can be useful. State the asymptotic result if you can not find a statistic T with a simple distribution.

Warning: BD uses $\psi(\boldsymbol{y}) = 1/\lambda(\boldsymbol{y})$ as the test statistic. So $-2\log\lambda(\boldsymbol{y}) = 2\log\psi(\boldsymbol{y})$ and $\lambda(\boldsymbol{y}) \leq c$ is equivalent to $\psi(\boldsymbol{y}) \geq c$.

A common LRT problem is $X_1, ..., X_n$ are iid with pdf $f(x|\theta)$ while $Y_1, ..., Y_m$ are iid with pdf $f(y|\mu)$. $H_0 : \mu = \theta$ and $H_1 : \mu \neq \theta$. Then under H_0, $X_1, ..., X_n, Y_1, ..., Y_m$ are an iid sample of size $n + m$ with pdf $f(y|\theta)$. Hence if under H_0 $f(y|\theta)$ is the $N(\mu, 1)$ pdf, then $\hat{\mu}_0 (= \hat{\theta}_0) = \dfrac{\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j}{n+m}$, the sample mean of the combined sample, while $\hat{\theta} = \overline{X}_n$ and $\hat{\mu} = \overline{Y}_m$.

---

8) **Large Sample Theory** (CLT, Limiting Distribution of the MLE or of an estimator that is a sum, Delta method, Consistency and Asymptotic Efficiency)

Central Limit Theorem (p. 215; Old 203; CB p. 236; BD p. 470): Let $Y_1, ..., Y_n$ be iid with $E(Y) = \mu$ and $V(Y) = \sigma^2$. Let the sample mean $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$. Then

$$\sqrt{n}(\overline{Y}_n - \mu) \xrightarrow{D} N(0, \sigma^2). \text{ Hence } \sqrt{n}\left(\frac{\overline{Y}_n - \mu}{\sigma}\right) = \sqrt{n}\left(\frac{\sum_{i=1}^n Y_i - n\mu}{n\sigma}\right) \xrightarrow{D} N(0,1).$$

MLE Rule of thumb (p. 226; Old 214; CB p. 472; BD p. 331) If $\hat{\theta}_n$ is the MLE or UMVUE of θ, then under strong regularity conditions $T_n = \tau(\hat{\theta}_n)$ is an asymptotically efficient estimator of $\tau(\theta)$, and if $\tau'(\theta) \neq 0$, then

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N\left(0, \frac{1}{I_1(\theta)}\right) \text{ and } \sqrt{\mathrm{n}}[\tau(\hat{\theta}_{\mathrm{n}}) - \tau(\theta)] \xrightarrow{\mathrm{D}} \mathrm{N}\left(0, \frac{[\tau'(\theta)]^2}{\mathrm{I}_1(\theta)}\right).$$

The Delta Method (p. 217; Old 205; CB p. 243; BD p. 311): Suppose that $\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2)$. Then

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

if $g'(\theta) \neq 0$ exists.

Know: Often the θ and σ^2 in the delta method are found either by using the central limit theorem with $\theta = \mu$ or by using the above MLE rule of thumb with $\sigma^2 = 1/I_1(\theta)$.

(p. 225; Old 213; CB p. 471): Let $Y_1, ..., Y_n$ be iid RVs. An estimator $T_n(\boldsymbol{x})$ is **asymptotically efficient** for $\tau(\theta)$ if

$$\sqrt{n}(T_n - \tau(\theta)) \xrightarrow{D} N\left(0, \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right) \sim N(0, FCRLB_1[\tau(\theta)])$$

where $I_1(\theta)$ is the Fisher information for θ based on a sample of size 1. Since $FCRLB_n(\tau(\theta)) = \frac{[\tau'(\theta)]^2}{nI_1(\theta)}$, an asymptotically efficient estimator T_n satisfies

$$(T_n - \tau(\theta)) \approx N(0, FCRLB_n[\tau(\theta)]).$$

Rule of thumb: in one parameter REFs, the MLE and UMVUE of $\tau(\theta)$ tend to be asymptotically efficient if $\tau'(\theta) \neq 0$ exists. For MLEs the result follows when the MLE rule of thumb holds by the delta method.

Notation: (p. 228; Old 216): If $T_n \xrightarrow{P} \tau(\theta)$ for all $\theta \in \Theta$, then T_n is a **consistent estimator** of $\tau(\theta)$.

(p. 230; Old 219) Fact: if $\sqrt{n}(T_n - \tau(\theta)) \xrightarrow{D} N(0, v(\theta))$ for all $\theta \in \Theta$, then T_n is a consistent estimator of $\tau(\theta)$.

Fact: (p. 229; Old 218; CB p. 469): If $V_\theta(T_n) \to 0$ and $E_\theta(T_n) \to \tau(\theta)$ as $n \to 0$ for all $\theta \in \Theta$ (ie if $MSE_{\tau(\theta)}(T_n) \to 0$), then T_n is a consistent estimator of $\tau(\theta)$.

Slutsky's Theorem (p. 230; Old 220; CB p. 239; BD p. 467): If $Y_n \xrightarrow{D} Y$ and $W_n \xrightarrow{P} w$ for some constant w, then $Y_n W_n \xrightarrow{D} wY$, $Y_n + W_n \xrightarrow{D} Y + w$ and $Y_n/W_n \xrightarrow{D} Y/w$ for $w \neq 0$.

(p. 223; Old 211-2; CB p. 476; BD p. 357): Let $T_{1,n}$ and $T_{2,n}$ be two estimators of a parameter τ such that

$$n^\delta(T_{1,n} - \tau) \xrightarrow{D} N(0, \sigma_1^2(F))$$

and

$$n^\delta(T_{2,n} - \tau) \xrightarrow{D} N(0, \sigma_2^2(F)),$$

then the **asymptotic relative efficiency** of $T_{1,n}$ with respect to $T_{2,n}$ is

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

Some distribution facts not on p. 1 of the review.

Suppose $Y_1, ..., Y_n$ are iid $N(\mu, \sigma^2)$. Then $Z = \frac{Y-\mu}{\sigma} \sim N(0,1)$.

$Z = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ while $a + cY_i \sim N(a + c\mu, c^2\sigma^2)$.

Suppose $Z, Z_1, ..., Z_n$ are iid N(0,1). Then $Z^2 \sim \chi_1^2$.

Also $a + cZ_i \sim N(a, c^2)$ while $\sum_{i=1}^n Z_i^2 \sim \chi_n^2$.

If X_i are independent $\chi_{k_i}^2 \equiv \chi^2(k_i)$ for $i = 1, ..., n$, then $\sum_{i=1}^n X_i \sim \chi^2(\sum_{i=1}^n k_i)$.

Let $W \sim EXP(\lambda)$ and let $c > 0$. Then $cW \sim EXP(c\lambda)$.

Let $W \sim gamma(\nu, \lambda)$ and let $c > 0$. Then $cW \sim gamma(\nu, c\lambda)$.

If $W \sim EXP(\lambda) \sim gamma(1, \lambda)$, then $2W/\lambda \sim EXP(2) \sim gamma(1, 2) \sim \chi^2(2)$.

Let $k \geq 0.5$ and let $2k$ be an integer. If $W \sim gamma(k, \lambda)$, then $2W/\lambda \sim gamma(k, 2) \sim \chi^2(2k)$.

Let $W_1, ..., W_n$ be independent $gamma(\nu_i, \lambda)$. Then $\sum_{i=1}^n W_i \sim gamma(\sum_{i=1}^n \nu_i, \lambda)$.

---

9) Note that sometimes you need the following result: the pdf of $Y = t(X)$ is $f_Y(y) = f_X(t^{-1}(y)) \left|\frac{dt^{-1}(y)}{dy}\right|$ for $y \in \mathcal{Y}$. See Jan. 2009 1a, Aug. 2012 5a, Jan. 2012 2b, and Jan. 2013 5a.

The following types of qual problems will not appear on the 3rd midterm or final.

A) Sometimes problems that require memorization of the solution appear on the qual.

a) Jan. 2004 1a): Prove that $\overline{Y}$ and S^2 are independent if $Y_1, ..., Y_n$ are iid $\mathrm{N}(\mu, \sigma^2)$ using Basu's theorem. See p. 119 (Old p. 124) and CB example 6.2.27 on p. 289).

b) Aug. 2001 1b): $Y_1, ..., Y_n$ are iid $U(0, \theta)$. Show that $\max(Y_i)$ is a complete statistic for θ. See CB example 6.2.23 on p. 286.

c) Memorization of the solution of UMVUE problems from Lehmann's Theory of Point Estimation such as (CB p. 86).

B) Other problems not from 1) - 8) occur. **Work old quals to see the types of problems.**

a) Aug. 2000 1) Steiner's formula = conditional variance identity (p. 43; Old p. 45; problem 2.68 on p. 94; Old p. 85; CB p. 167). Also see Jan. 2012 qual.

b) Jan. 2004 and Jan. 2010 5): A) a) above. Independence of $\overline{X}$ and S^2 if data is iid $N(\mu, \sigma^2)$.

c) Aug. 2003 problem 9.1b on p. 334; Old p. 269. Also see problem 9.12 from Aug. 2009 7), Sept. 2010 7), Jan 2018 6), Jan. 2020 6d). **Confidence intervals**.

d) Theorem 8.30: Suppose that g does not depend on n, $g'(\theta) = 0$, $g''(\theta) \neq 0$ and

$$\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \tau^2(\theta)).$$

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

Sept. 2005 problem 8.27c, Jan 2014 2c). Also see ex. 8.14.

e) Wald statistic for testing: Jan. 2010 problem 6bc.

f) Jan. 2007 3a: state Basu's theorem.

g) Find moment generating function of $Y = X_1 X_2$, Jan. 2012, 1.

h) Jan. 2016 1a) pdf of $X + Y$, X,Y iid U(0,1)

To pass the qual you need to satisfy the graders. Often a score of 80% or higher will pass, but the needed score may be higher or lower for a given qual. Students who have answered 3 out of 6 questions correctly (or at least with a grade of an A) and 2 questions with "right idea" but some moderate calculation errors (with a grade of high C or low B) and one question with major errors (grade of D or high F) have passed, but have also not passed.