

Math 541A Midterm 2 Cheat Sheet

Gregory Faletto

Remark. Unlike Exam 1, Exam 2 will have a reference sheet at the beginning of the exam, where the following definitions will be stated: sufficient statistic, minimal sufficient, ancillary, complete, and the definition of conditional expectation as a random variable.

Exercises to do (in relevant parts of lecture notes but not assigned in HW): 6.22 in lecture notes; problems 1 - 3 in HW6.

1 Data Reduction

1.1 Sufficient Statistics

How to show sufficiency: either use the definition (as we did in some examples; show that conditional density does not depend on θ), or use the Factorization Theorem:

Theorem 1 (Factorization Theorem, Theorem 5.4 in 541A notes). Suppose $X = (X_1, \dots, X_n)$ is a random sample of size n from a distribution f where $f \in \{f_\theta : \theta \in \Theta\}$ is a family of probability density functions or probability mass functions. Let $t : \mathbb{R}^n \rightarrow \mathbb{R}^k$, so $Y := t(X_1, \dots, X_n)$ is a statistic. Then Y is sufficient for θ if and only if there exists a nonnegative $\{g_\theta : \theta \in \Theta\}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}$, $g_\theta : \mathbb{R}^k \rightarrow \mathbb{R}$ such that

$$f_\theta(x) = g_\theta(t(x))h(x), \quad \forall x \in \mathbb{R}^n, \quad \forall \theta \in \Theta. \quad (1)$$

How to prove a minimal sufficient statistic exists: pretty much always exists by Proposition 5.12.

How to show minimal sufficiency: typically, use Theorem 5.8:

Theorem 2 (Theorem 5.8 in 541A notes). Let $\{f_\theta : \theta \in \Theta\}$ be a family of probability density functions or probability mass functions. Let X_1, \dots, X_n be a random sample from a member of the family. Let $t : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and define $Y := t(X_1, \dots, X_n)$. Y is minimal sufficient if and only if the following condition holds for every $x, y \in \mathbb{R}^n$:

There exists $c(x, y) \in \mathbb{R}$ that does not depend on θ such that $f_\theta(x) = c(x, y)f_\theta(y) \quad \forall \theta \in \Theta$

if and only if

$$t(x) = t(y).$$

Remark. If a minimal sufficient exists, it is unique up to an invertible transformation. By this uniqueness, the converse of Theorem 5.8 also holds.

1.2 Ancillary Statistics

How to show a statistic is complete: Use the definition (difficult except for discrete random variables where you can express expectation without θ , as in binomial).

Why do we care if a statistic is complete?

(1) **Complete sufficiency implies minimal sufficiency:**

Theorem 3 (Bahadur's Theorem; Theorem 5.25 in Math 541A notes). If Y is a complete sufficient statistic for a family $\{f_\theta : \theta \in \Theta\}$ of probability densities or probability mass functions, then Y is a minimal sufficient statistic for θ .

Remark. Because it is true that if a minimal sufficient exists, it is unique up to an invertible transformation, from Bahadur's Theorem we also know that a complete sufficient statistic is unique up to an invertible mapping. However, the converse of Bahadur's Theorem is false.

(2) **Complete sufficient statistics are independent from ancillary statistics:**

Theorem 4 (Basu's Theorem, Theorem 5.27 in Math 541A notes). Let $Y : \Omega \rightarrow \mathbb{R}^k$ and $Z : \Omega \rightarrow \mathbb{R}^m$ be statistics. If Y is a complete sufficient statistic for $\{f_\theta : \theta \in \Theta\}$ and Z is ancillary for θ , then for all $\theta \in \Theta$, Y and Z are independent with respect to f_θ .

(3) **With a complete sufficient statistic (and any unbiased estimator), we can find a UMRU/UMVU estimator** (see Lehmann-Scheffe below).

2 Point Estimation

2.1 Evaluating Estimators

Definition 2.1 (UMVU, sometimes called MVUE (minimum variance unbiased estimator); Definition 6.3 in Math 541A Notes). Let X_1, \dots, X_n be a random sample of size n from a family of distributions $\{f_\theta : \theta \in \Theta\}$. Let $g : \Theta \rightarrow \mathbb{R}$. Let $t : \mathbb{R}^n \rightarrow \mathbb{R}$ and let $Y := t(X_1, \dots, X_n)$ be an unbiased estimator for $g(\theta)$. We say that Y is **uniformly minimum variance unbiased (UMVU)** if, for any other unbiased estimator Z for $g(\theta)$, we have $\text{Var}_\theta(Y) \leq \text{Var}_\theta(Z)$ for all $\theta \in \Theta$.

Remark. The “uniform” property has to do with the fact that this inequality must hold for every $\theta \in \Theta$ (as opposed to for a particular θ , or averaged over all $\theta \in \Theta$, or something like that).

Remark (Remark 6.14 in Math 541A notes). Let $Z : \Omega \rightarrow \mathbb{R}^k$ be a complete sufficient statistic for $\{f_\theta : \theta \in \Theta\}$ and let $h : \mathbb{R}^k \rightarrow \mathbb{R}^m$. Let $g(\theta) := \mathbb{E}_\theta h(Z)$ for all $\theta \in \Theta$. Then $h(Z)$ is unbiased for $g(\theta)$, since $\mathbb{E}_\theta h(Z) = g(\theta) = \mathbb{E}_\theta(g(\theta))$. Applying Theorem 6 (Lehmann-Scheffe, Theorem 6.13 in Math 541A notes), we have

$$W := \mathbb{E}_\theta(h(Z) \mid Z) = \mathbb{E}_\theta(\mathbb{E}_\theta[h(Z) \mid h(Z)] \mid Z) = \mathbb{E}_\theta[h(Z) \mid h(Z)] = h(Z).$$

Therefore by Theorem 6 (Lehmann-Scheffe, Theorem 6.13 in Math 541A notes), $h(Z)$ is UMVU for $g(\theta)$. That is, **any function of a complete sufficient statistic is UMVU for its expected value**. So one way to find a UMVU is to come up with a function of a complete sufficient statistic that is unbiased for a given function $g(\theta)$.

More generally, given a family of distributions $\{f_{\tilde{\theta}} : \tilde{\theta} \in \Theta\}$, we could be given a **loss function** $\ell(\theta, y) : \Theta \times \mathbb{R}^k \rightarrow \mathbb{R}$ and be asked to minimize the **risk function** $r(\theta, Y) := \mathbb{E}_{\tilde{\theta}}(\ell(\theta, Y))$ over all possible estimators Y . In the case of mean squared error loss, we have $\ell(\theta, y) := (y - g(\theta))^2$ for all $y, \theta \in \mathbb{R}$.

Definition 2.2 (Definition 6.4 in 541A notes). Let X_1, \dots, X_n be a random sample of size n from a family of distributions $\{f_{\theta} : \theta \in \Theta\}$. Let $g : \Theta \rightarrow \mathbb{R}$. Let $t : \mathbb{R}^n \rightarrow \mathbb{R}$ and let $Y := t(X_1, \dots, X_n)$ be an unbiased estimator for $g(\theta)$. We say Y is **uniformly minimum risk unbiased** (UMRU) if for any other unbiased estimator Z for $g(\theta)$,

$$r(\theta, Y) \leq r(\theta, Z), \quad \forall \theta \in \Theta$$

The Rao-Blackwell Theorem says that we can lower the risk of an estimator Y by conditioning on a sufficient statistic Z .

Theorem 5 (Rao-Blackwell; Theorem 6.4 in Math 541A notes). Let Z be a sufficient statistic for $\{f_{\theta} : \theta \in \Theta\}$ and let Y be an estimator for $g(\theta)$. Define $W := \mathbb{E}_{\theta}(Y | Z)$. Let $\theta \in \Theta$. Then

$$\text{Var}_{\theta}(W) \leq \text{Var}_{\theta}(Y).$$

Further, let $r(\theta, y) < \infty$ and such that $\ell(\theta, y)$ is convex in y . Then

$$r(\theta, W) \leq r(\theta, Y).$$

In practice, if Z is not complete then we might not see an improvement.

Theorem 6 (Lehmann-Scheffe, Theorem 6.13 in Math 541A notes). Let Z be a complete sufficient statistic for a family of distributions $\{f_{\theta} : \theta \in \Theta\}$. Let Y be an unbiased estimator for $g(\theta)$. Define $W := \mathbb{E}_{\theta}(Y | Z)$. (Since Z is sufficient, W does not depend on θ .) Then W is UMRU for $g(\theta)$. Further, if $\ell(\theta, y)$ is strictly convex in y for all $\theta \in \Theta$, then W is unique. In particular, W is the unique UMVU for $g(\theta)$.

What if we don't have a complete sufficient statistic but we still want the UMVU?

Theorem 7 (Alternate Characterization of UMVU; Theorem 6.18 in Math 541A notes). Let $f \in \{f_{\theta} : \theta \in \Theta\}$ be a family of distributions and let $g : \Theta \rightarrow \mathbb{R}$. Let W be an unbiased estimator for $g(\theta)$ (note that the existence of an unbiased estimator is a nontrivial assumption). Let $L_2(\Omega)$ be the set of statistics with finite second moment. Then $W \in L_2(\Omega)$ is UMVU for $g(\theta)$ if and only if for any $\theta \in \Theta$,

$$\mathbb{E}_{\theta}(WU) = 0, \quad \forall U \in L_2(\Omega) \text{ that are unbiased estimators of } 0$$

Thinking of this as an inner product, we have to be orthogonal to all such U .

Remark. If we have a complete sufficient statistic, better to use the earlier methods in general (unless it is really complicated to work with). If we don't have a complete sufficient statistic, use this theorem.

Summary of methods for finding a UMVU estimator for $g(\theta)$:

(1) If we have a complete sufficient statistic Z :

- (a) **(Condition method/Rao-Blackwell):** Follow Theorem 6 (Lehmann-Scheffe, Theorem 6.13 in Math 541A notes): find an unbiased Y and let $W := \mathbb{E}_\theta(Y \mid Z)$; this is UMVU. (problem: can be hard to find an unbiased Y .)
- (b) Solve for $h : \mathbb{R}^k \rightarrow \mathbb{R}$ satisfying

$$\mathbb{E}_\theta h(Z) = g(\theta) \quad (2)$$

by the above remark, $h(Z)$ is UMVU for $g(\theta)$. By “solve”, consider that we have g and Z and somehow solve for the h satisfying (2). For example if Z is binomial the left side of (2) will be the sum of a bunch of numbers. Find the h values that satisfy (2), if possible.

(2) If we don't have a complete sufficient statistic:

- (a) For a one-parameter family of distributions, follow the equality case of Theorem 8 (Cramer-Rao/Information Inequality, Theorem 6.23 in Math 541A Notes). That is, an unbiased estimator Y that is a constant multiple of $\frac{d}{d\theta} \log f_\theta(X)$ is UMVU. (Note that this avoids any discussion of complete sufficient statistics, but it's generally not easy to do.)
- (b) For a multiple-parameter family of distributions, apply Theorem 6.18 in Math 541A notes, but it will probably be difficult to apply.

2.2 Efficiency of an Estimator

Definition 2.3 (Fisher Information, Definition 6.19 in Math 541A notes). Let $f \in \{f_\theta : \theta \in \Theta\}$ be a family of multivariate probability densities or probability mass functions. Assume $\Theta \subseteq \mathbb{R}$ (this is a one-parameter situation). Let X be a random variable with distribution f_θ . Define the **Fisher information** of the family to be

$$I(\theta) = I_X(\theta) := \mathbb{E}_\theta \left(\frac{d}{d\theta} \log f_\theta(X) \right)^2, \quad \forall \theta \in \Theta$$

if this quantity exists and is finite.

Remark. Note that if X is continuous,

$$\mathbb{E}_\theta \left(\frac{d}{d\theta} \log f_\theta(X) \right) = \int_{\mathbb{R}^n} \frac{1}{f_\theta(x)} \frac{d}{d\theta} f_\theta(x) \cdot f_\theta(x) dx = \int_{\mathbb{R}^n} \frac{d}{d\theta} f_\theta(x) dx = \frac{d}{d\theta} \int_{\mathbb{R}^n} f_\theta(x) dx = \frac{d}{d\theta} 1 = 0.$$

So we could have equivalently defined the Fisher information as

$$I_X(\theta) = \text{Var}_\theta \left(\frac{d}{d\theta} \log f_\theta(X) \right)$$

Theorem 8 (Cramer-Rao/Information Inequality, Theorem 6.23 in Math 541A Notes). Let $X : \Omega \rightarrow \mathbb{R}^n$ be a random variable with distribution from a family of multivariable probability densities or probability mass functions $\{f_\theta : \theta \in \Theta\}$ with $\Theta \subseteq \mathbb{R}$. Let $t : \mathbb{R}^n \rightarrow \mathbb{R}$ and let $Y := t(X)$ be a statistic. For any $\theta \in \Theta$ let $g(\theta) := \mathbb{E}_\theta Y$. Then

$$\text{Var}_\theta(Y) \geq \frac{|g'(\theta)|^2}{I_X(\theta)}, \quad \forall \theta \in \Theta.$$

In particular, if Y is unbiased for θ , then $g(\theta) = \theta$, so,

$$\text{Var}_\theta(Y) \geq \frac{1}{I_X(\theta)}, \quad \forall \theta \in \Theta.$$

Equality occurs for some $\theta \in \Theta$ only when $\frac{d}{d\theta} \log f_\theta(x)$ and $Y - \mathbb{E}_\theta Y$ are multiples of each other.

Definition 2.4 (Efficiency, Definition 6.25 in Math 541A notes). Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable with distribution from a family of multivariable probability densities or probability mass functions $\{f_\theta : \theta \in \Theta\}$ with $\Theta \in \mathbb{R}$. Let $t : \mathbb{R}^n \rightarrow \mathbb{R}$ and let $Y := t(X)$ be a statistic. Define the **efficiency** of Y to be

$$\frac{1}{I_X(\theta) \text{Var}_\theta(Y)}, \quad \forall \theta \in \Theta$$

if this quantity exists and is finite. If Z is another statistic, we define the **relative efficiency** of Y to Z to be

$$\frac{I_X(\theta) \text{Var}_\theta(Z)}{I_X(\theta) \text{Var}_\theta(Y)} = \frac{\text{Var}_\theta(Z)}{\text{Var}_\theta(Y)}, \quad \forall \theta \in \Theta.$$

2.3 Bayes Estimation

In Bayes estimation, the parameter $\theta \in \Theta$ is regarded as a random variable Ψ . The distribution of Ψ reflects our prior knowledge about the probable values of Ψ . Then, given that $\Psi = \theta$, the conditional distribution of $X \mid \Psi = \theta$ is assumed to be $\{f_\theta : \theta \in \Theta\}$, where $f_\theta : \mathbb{R}^n \rightarrow [0, \infty)$. Suppose $t : \mathbb{R}^n \rightarrow \mathbb{R}^k$ and we have a statistic $Y := t(X)$ and a loss function $\ell : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}$. Let $g : \Theta \rightarrow \mathbb{R}^k$.

Definition 2.5 (Bayes estimator, Definition 6.26 in Math 541A notes). A **Bayes estimator** Y for $g(\theta)$ with respect to Ψ is defined such that

$$\mathbb{E}\ell(g(\Psi), Y) \leq \mathbb{E}\ell(g(\Psi), Z)$$

for all estimators Z . Here the expectation is with respect to both Ψ and Y . Note that we have not made any assumptions about bias for Y or Z .

Remark. $t(X)$ can depend on Ψ .

In order to find a Bayes estimator, it is sufficient to minimize the conditional risk.

Proposition 9 (Proposition 6.27 in Math 541A notes). Suppose there exists $t : \mathbb{R}^k \rightarrow \mathbb{R}$ such that for almost every $x \in \mathbb{R}^n$, $Y := t(X)$ minimizes

$$\mathbb{E}(\ell(g(\Psi), Z) \mid X = x)$$

over all estimators Z . Then $t(X)$ is a Bayes estimator for $g(\theta)$ with respect to Ψ .