300 Quals Guide

June 25, 2025

1 Misc facts

Definition 1 (Convexity)
For all $0 \leqslant t \leqslant 1$ and $x_1, x_2 \in X$:

$$f(tx_1 + (1-t)x_2) \leqslant tf(x_1) + (1-t)f(x_2).$$

Alternative, check the second derivative $\geqslant 0$

Theorem 2 (Jensen's Inequality)
If ϕ convex, then $\mathbf{E}\phi(X) \geqslant \phi(EX)$. Eg $EX^2 \geqslant (EX)^2$. Inequality is strict if ϕ is strictly convex and X is not degenerate (constant). Also conditional version holds.

2 Exponential Families

Definition 3 (Exponential Family)

$$p_\theta(x) = \exp(\sum_{i=1}^{s} T_i(X)\eta_i(\theta) - A(\theta))h(x)$$

$$p_\eta(x) = \exp(\sum_{i=1}^{s} T_i(X)\eta_i - \tilde{A}(\eta))h(x)$$

Fact 4 ($E[T(X), Cov(T(X))$)

$$\mathbf{E}[T(X)] = \nabla A(\eta)$$

$$\mathbf{Cov}(T_i(X), T_j(X)) = \partial_{\eta_i}\partial_{\eta_j}A(\eta)$$

2.1 All the Expo family examples

Include curved..

3 Sufficiency

"Throwing away everything else besides this statistic entails no loss of info in estimating θ"

Definition 5 (Sufficiency)
If for all t, the distribution of $X|T = t$ does not depend on θ.

If you collect data $X_1, \dots, X_n$, then throw away data except for $T = t$, then can construct $\tilde{X}_1, \dots \tilde{X}_n$ with same distribution as $\underline{X}$ just by knowing T.

Make a table of good examples including Unif$(0, \theta)$

Theorem 6 (Factorization Theorem)
T sufficient for θ (or for the model) $\iff$ we can write $p_\theta(\underline{x}) = g_\theta(t(\underline{x}))h(\underline{x})$. Ie, can factor the density into a part that depends on statistic and θ, and another part that depends on data but not θ.

Example 7 (Expo(θ))
Can write $p_\theta(\underline{x}) = \theta^n \exp(-\theta \sum x_i)$ so $T(X) = \sum_{i=1}^n X_i$ is sufficient.

Example 8 (Exponential families)
Look at the form – T is sufficient.

Definition 9 (Minimal sufficiency)
For all T' sufficient, T is a function of T'.

Theorem 10 (Rao-Blackwell)
If T sufficient, L is any convex loss, and δ estimates $g(\theta)$, define:

$$\eta(T) = \mathbf{E}[\delta(X)|T],$$

then using Tower property and Jensen's:

$$R(\theta, \delta(X) = \mathbf{E}[g(\theta), \delta(X)] \geqslant \mathbf{E}[L(\theta, \eta(T)) = R(\theta, \eta(T))$$

Definition 11 (Ancillary Statistic)
Distribution of $A(X)$ does not depend on θ.

Definition 12 (First Order Ancillary)
If $\mathbf{E}_\theta A(X)$ does not depend on θ. Weaker than Ancillary.

4 Completeness

Definition 13 (Completeness)
$T(X)$ complete if no non-constant function of T is even 1st order ancillary, or equivalently:

$$\mathbf{E}f(T) = 0 \implies f(T) = 0 \text{ a.s.}$$

Example 14 ($X_i \sim \text{Unif}(0, \theta)$)
See coaching notes– break h into positive and negative parts to show $h = 0$ as.

Theorem 15 (Full rank exponential family)
Ie if $(\eta_1, \ldots, \eta_k) : \eta \in \Omega$ contains a k–dimensional rectangle, then T is not only sufficient, but **also complete**.

Theorem 16 (Basu's Theorem)
If T is CSS and A Ancillary, then $T \perp\!\!\!\perp A$.

A common strategy with Basu's theorem is to consider a submodel– show independence in the submodel, then based on arbitrariness of submodel, show true for full model.

Example 17 (Show sample mean and sample variance in $N(\mu, \sigma^2)$ are independent)
(Assume both unknown). then in the submodel where $\sigma = \sigma_0$ known, then $\overline{X}$ is CSS. Can show that $\sum(X_i - \overline{X})^2$ is Ancillary. So $\overline{X} \perp\!\!\!\perp \sum(X_i - \overline{X})^2$. This is true for all μ, and for $\sigma = \sigma_0$ fixed. But σ_0 arbtirary, so true for all μ, σ.

Add a table here including 2 param Unif

1. $\text{Unif}(0, \theta) : T = X_{(n)}$
2. $\text{Unif}(\theta_1, \theta_2)$: $T = (X_{(1)}, X_{(n)})$

5 UMVU

Note that we can't take existence of an unbiased estimator for granted. If $x \sim Bin(n, \theta)$. Take like $g(\theta) = \frac{1-\theta}{\theta}$ or a polynomial of degree $> n$– no unbiased estimators. Strategy to show no unbiased estimator is just write out the expectation of an estimator.

Definition 18 (UMVU)
δ^* is unbiased and for all other δ, $R(\delta^*, g(\theta)) \leqslant R(\delta, g(\theta))$ for all θ.

Note that if we have an unbiased estimator, we can Rao-Blackwellize with a sufficient statistic and still have unbiased estimator.

Theorem 19 (Lehman-Scheffe)
If unbiased estimator is function of CSS, it is UMVU. (Since there exists at most one unbiased estimator that's function of T by completeness).

Alternate statement: if there exists any unbiased estimator and a CSS T, then there is a unique unbiased estimator that's a function of T, which is UMVU (UMRU for any convex loss). Unique UMVU if strict convex loss, since this makes Jensen strict.

Recipe 20 (UMVU with CSS)
Steps:

1. Find a CSS
2. Find an unbiased estimator function of CSS. (Alternatively, find a dumb unbiased estimator and RB)
3. $\implies$ UMVU

Note that UMVU can be inadmissible - see James-Stein or the Poisson UMVU for $g(\theta)$. = $\exp(-3\lambda)$ example in Lecture 5.

Theorem 21 (Orthogonality Condition)
Let $\hat{\theta}$ be an unbiased estimator with $E_P\hat{\theta}^2 < \infty$ if $U(X)$ is an unbiased estimator of 0 for all P, then:

$$\hat{\theta} \text{ UMVU} \iff E_P[\hat{\theta}(X)U(X)] = 0 \text{ for all unbiased estimators of 0 and for all P}$$

An application is showing that the addition of UMVUs is UMVU.
Also reasonable to try to show not UMVU. Come back to Ex 2.3 in Notes

5.1 UMVU Examples

1. Basic Expo Fam examples – eg $\overline{X}$ in Bernouli or Normal with known variance
2. Empirical CDF in $\mathcal{N}(\theta, 1)$.
   - CSS is $\overline{X}$, $\delta = \mathbf{1}[X_1 < u]$ unbiased.
   - Idea to RB then add and subtract $\overline{X}$ since $X_1 - \overline{X}$ is ancillary, then apply Basu
   - UMVU is $\Phi(\frac{u-\overline{X}}{\sqrt{(n-1)/n}})$

Non parametric examples

1. $X_i \sim F \in \mathcal{F} = \{$ all distributions with density wrt Lebesgue and finite variance $\}$. $g(\theta) = E_F X_i$.
   - Note that $\overline{X}$ is unbiased in the big family and is UMVU in the normal **subfamily**
   - Order statistics CSS (always sufficient - complete by subfamily arg and bijection with sums of powers): $(X_{(1)}, \dots, X_{(n)}) \iff (\sum X_i, \dots, \sum X_i^n)$ bijection.
   - So $\overline{X}$ is UMVU
2. X_i iid symmetric about θ, $EX_i = \theta$. Finite variance.
   - Two subclasses: normal family, Unif(θ_1, θ_2) family - have different UMVUs and both are unbiased in the original class

Recipe 22 (Subfamily UMVU Argument)
If UMVU in subfamily and unbiased in big family, must be UMVU in big family **if** an UMVU exists in the big family. Because the UMVU **uniquely** minimizes variance in the subfamily.

Proof. Take the big family. Take a function such that $E_\theta f(X) = 0$ for all θ. Then true for subfamily. So $P_\theta(f(X) = 0)$ for all θ in the subfamily. Same null sets as big family, means that also $P_\theta(f(X) = 0)$ for the big family. ☐

Fact 23 (Completeness in subfamily)
If $\mathcal{F}_0 \subset \mathcal{F}$ and they have the same null sets, then completeness in $\mathcal{F}_0$ implies completeness in $\mathcal{F}$.

Recipe 24 (Non-existence of Non-parametric UMVU)
Find two different subclasses with different unique UMVU that are also unbiased in the big class – no UMVU in big class.

5.2 Non-convex loss functions

If loss is bounded, there is no UMRU estimator. This is unbiased:

$$\delta_\pi = \begin{cases} g(\theta_0) \text{ wp } 1-\pi \\ \frac{1}{\pi}[\delta_0(X) - g(\theta_0)] + g(\theta_0) \end{cases}$$

and its risk is πM, so it could be arbitrarily small.

6 MRE

Big picture– there's an easy recipe for MRE for square error or absolute error if we pick any old δ_0 equivariant function of CSS.

Definition 25 (Location models)
Density satisfies

$$f_{\theta+h}(x+h) = f_\theta(x)$$

Think: normal at its mean has same density as a shifted normal at shifted mean.
A **location invariant loss** is $\ell(a+h, \theta+h) = \ell(a, \theta)$ for all a, θ, h.

Want location equivariant estimators, ie $\hat{\theta}(x+h) = \hat{(\theta)} + h$ – if our observations all shift by some amount, our estimate of the location should also shift by that amount.

Fact 26 (Bias, variance, risk of equivariant estimators)
Do not depend on θ. Ie,

$$\mathbf{E}_\theta \hat{\theta} = \theta + b \text{ for all } \theta$$

for risk, since it's constant, we can hope to find the best among all equivariant estimators.

Definition 27 (MRE)
Satisfies equivariance condition:

$$\hat{\theta}(x+c, u) = \hat{\theta}(x, u) + c$$

and minimum risk condition amongst all equivariant estimators.

Definition 28 (Characterization of location **invariant** estimators)
Location invariant means $U(X_1+c, \ldots, X_n+c) = U(X)$. Characterize by: U location invariant $\iff U = V(y_1, \ldots, y_{n-1})$ where $y_i = X_i - X_n$.

Fact 29 (Characterization of location **equivariant** estimators)
Let δ_0 be **any** location equivariant estimator, eg $\delta_0 = \overline{X}$.

$$\delta \text{ is location equivariant } \iff \delta(X_1, \dots, X_n) = \delta_0(X_1, \dots, X_n) + U(X_1, \dots, X_n)$$

where U is location **invariant**. Then from invariant characterization:

$$\iff \delta(X_1, \dots, X_n) = \delta_0(X_1, \dots, X_n) + V(Y_1, \dots, Y_{n-1})$$

Recipe 30 (Finding the MRE)
Let X_i observations iid from location model and let $Y = (X_1 - X_n, \dots, X_{n-1} - X_n)$. Let $\hat{\theta}_0$ be any location equivariant estimator of θ_0 with finite risk. If the following is well-defined:

$$v(y) = \arg\min_v E_0[\ell(\hat{\theta}_0(X) - v, 0) \mid Y = y]$$

Then there exists an MRE $\hat{\theta}^*(X) = \hat{\theta}_0(X) - v(Y)$.

Want to minimize

$$\mathbf{E} - \theta[\rho(\delta_0(X) - V(Y) - \theta)] = \mathbf{E}_0 \rho(\delta_0(X) - V(Y))$$

Apply Tower property and minimize the inner expectation. For square error yields:

$$V = \mathbf{E}_0[\delta_0(X)|Y)]$$

So $\delta^* = \delta_0 - E_0[\delta_0|Y]$ Notice that the Y_i here are ancillary so if we pick δ_0 function of CSS, easy.
If we can make δ_0 a function of CSS, then we can apply Basu

Recipe 31 (MRE for Square Error Loss)
Choose δ_0 equivariant function of CSS. Then $\delta^* = \delta_0 - E_0\delta_0$ is MRE.

Example 32 ($X_i \sim N(\theta, 1)$)
Let $\delta_0 = \overline{X}$. Then $E_0\delta_0 = 0$. So $\overline{X}$ is MRE.

Note for absolute error, just do $\text{median}_{\theta=0}[\delta(X)|Y]$ instead of expectation. Ie find m such that $P(X \leqslant m) \geqslant 1/2$ and $P(X \geqslant m) \geqslant 1/2$.

Theorem 33 (Existence of MRE)
If loss is convex and not monotone, then MRE exists by the previous theorem. If strictly convex, unique.

Fact 34 (MRE is Unbiased (under squared error loss))
Since the bias does not depend on θ, just subtract off whatever bias.

Fact 35 (UMVU is MRE if UMVU is location equivariant)
In a location model, the UMVU is location equivariant. Since MRE is the best amongst equivariant estimators, and any competing equivariant estimators are unbiased.. Is this just with square error loss?

Definition 36 (Pitman Estimator)

Theorem 37 (Mini-convolution theorem)

7 James-Stein Estimator

Setup: $\mathcal{N}(\mu, \sigma^2)$ with σ^2 known.

Theorem 38 (SURE - Stein Unbiased Risk Estimator)
Letting $\hat{\mu}(x) = x + g(x)$ for $g : \mathbb{R}^p \longrightarrow \mathbb{R}^p$ almost differentiable, and assume that $\mathbf{E}[\sum_{i=1}^p |\partial_i g_i(X)|] < \infty$. Then:

$$\mathbf{E}_\mu[\|\hat{\mu}(X) - \mu\|^2] = p\sigma^2 + \mathbf{E}\left[\|g(X)\|^2 + 2\sigma^2 \sum_{i=1}^p \partial_i g_i(X)\right].$$

Proved using integration by Parts – see Lec 17 300c.

Fact 39 (UMVU is not admissible in $\mathcal{N}(\mu, 1)$ model)
Because James Stein renders X inadmissible.

J-S estimator is given by:

$$\hat{\mu}^{JS}(X) = \left(1 - \frac{\sigma^2(p-2)}{\|X\|_2^2}\right) X,$$

biased towards the origin. Prove that it has better risk by SURE estimator.

8 Bayes Estimators

Fact 40 (Unique Bayes is Admissible)
Idea is that if $R(\hat{\theta}', \theta) \leqslant R(\hat{\theta}, \theta)$ for all θ, it would then be Bayes.
provided the prior isn't super weird (eg continuous dist with an atom)

Fact 41 (Constant risk Bayes is minimax)
If not, some other estimator would render Bayes inadmissible, which would make that estimator the Bayes estimator.

Fact 42 (Bayes is not UMVU if $r_\Lambda < \infty$)
Under square error loss, Bayes estimators are biased

9 Minimax

A Bayes or admissible estimator is constant risk, then it is minimax.