Background Review and Convergence

Albert Dorador

UW-Madison

February 3rd, 2023

Exercise 1: Background review 1 (Properties of Expectation and Variance)

Let X and Y be independent random variables with $\mathbb{E}(X) = \mathbb{V}(X) = 1$ and $\mathbb{E}(Y) = \mathbb{V}(Y) = 2$. Optional: Can you think of a general family of distributions that could generate such random variables?

Calculate:

- $\mathbb{E}(3X + Y)$
- $\mathbb{E}(3X * Y)$
- $\mathbb{V}(3X + Y)$
- $\mathbb{V}(3X * Y)$

Exercise 2: Background review 2 (Jacobian Theorem, Lecture 10 Stat 609, 2020)

Let X_1, X_2 i.i.d. $Exp(\lambda)$ with rate parametrization. Let $Y_1 = \frac{X_1}{X_1+X_2}$ and $Y_2 = X_1 + X_2$

a) Write down the possible values of Y_1 and Y_2

b) Find the joint density for (Y_1, Y_2)

c) Find the marginal densities for Y_1 and Y_2

d) Are Y_1 and Y_2 independent?

Exercise 3: Convergence of random variables (Final Exam Stat 609, 2020)

Let $X_1, \ldots, X_n$ i.i.d. $Exp(1/2)$ with rate parametrization

Let $Y_1, \ldots, Y_n$ i.i.d. $Pois(\mu)$

Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $Z_n = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$

Find the limiting distribution of $W_n = [\sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{2})]^{-1} Z_n^2$

Convergence in probability revisited, Sufficiency

Albert Dorador

UW-Madison

February 10th, 2023

An application of convergence in probability (Consistency)

A reasonable point estimator is expected to perform better, at least on average, if more information about the unknown population is available.

T_n is a consistent estimator of $g(\theta)$ if, for every $\epsilon > 0$ and any $\theta \in \Theta$, $\lim_{n\to\infty} P(|T_n - g(\theta)| \geq \epsilon) = 0$ i.e. if it converges in probability

How do we show an estimator is consistent? 4 usual methods:

1. Evaluate $P(|T_n - g(\theta)| \geq \epsilon)$ directly and fids its limit as $n \to \infty$
2. Invoke the LLN (sample moments converge in probability to population moments, under mild assumptions)
3. Use Continuous mapping theorem: if you know S_n is consistent for θ, then, for any continuous function g, $T_n := g(S_n)$ is consistent for $g(\theta)$
4. Thm 10.1.3: T_n is a consistent estimator of $g(\theta)$ if, for any $\theta \in \Theta$, $\lim_{n\to\infty} Bias(T_n; \theta) = 0$ and $\lim_{n\to\infty} Var(T_n) = 0$

Exercise 1: Consistency

Let $X_1, \ldots, X_n$ i.i.d. from a distribution with pdf

$$f(x) = \frac{1}{2}(1 + \theta x) \cdot I(-1 < x < 1)$$

Find a consistent estimator of θ, and prove it is indeed consistent

Exercise 2: Sufficiency

Let $X_1, \ldots, X_n$ i.i.d. from a distribution with pdf

$$f(x) = \frac{e^{-(x-\mu)/\sigma}}{\sigma} \cdot I(x > \mu)$$

Find a two-dimensional sufficient statistic for (μ, σ), and prove it is indeed sufficient

Q&A

Minimal sufficiency and completeness

Albert Dorador

UW-Madison

February 17th, 2023

Exercise 1: Minimal sufficiency

Let $X_1, \ldots, X_n$ i.i.d. from $Unif(\theta, 2\theta)$, with $\theta > 0$.

Find a minimal sufficient statistic for θ. Is it complete too?

Exercise 2: Completeness

Let $X_1, \ldots, X_n$ i.i.d. from a distribution with pdf

$$f(x) = \theta x^{\theta-1} \cdot I(0 < x < 1)$$

with $\theta > 0$

Find a complete sufficient statistic for θ

Minimal sufficient M(X) vs complete C(X)

- Minimal sufficient: Max data reduction while keeping all info about θ
- Complete: abstract property of the family of distributions that C belongs to, so it's not directly tied to a likelihood function. Also, strictly speaking, completeness doesn't care about preserving all information about θ, and in fact $C(X) = 3$ is trivially complete.
- But, if C is sufficient, $M(X) = C(X)$ if they (both) exist. So M minimal sufficient but not complete implies no complete sufficient statistic exists
- Lastly, why is it called 'complete'? Recall that the vector space of real functions whose domain is $\mathbb{R}$ has inner product $\langle f, g\rangle = \int_{\mathbb{R}} fg\, dx$

Does that ring a bell? Indeed $\mathbb{E}(g(C)) = \int_{\mathbb{R}} g(C)\, f_\theta(x)\, dx = \langle g(C), f_\theta \rangle$

Then "$\mathbb{E}(g(C)) = 0\ \forall\theta \Rightarrow g = 0$ a.s. $\forall\theta$" means that the functions f_θ span the *complete* space of functions of C, so the only orthogonal vector is 0

MoM and MLE

Albert Dorador

UW-Madison

February 24th, 2023

Another look at MoM and MLE

- MoM idea: if $dim(\theta) = k$, set (at least) the first k sample moments equal to the first k population moments, and solve for $\theta_1, \ldots, \theta_k$ in terms of the sample moments.
- BUT, these equations may not have a solution! In this case, consider exploring higher moments.
- MLE idea: find the (log) likelihood of θ $L(\theta)$ (i.e. the joint pdf when $x_1, \ldots, x_n$ are viewed as fixed) and find the θ^* that maximizes it
- Vast majority of times, this is an unconstrained optimization problem: set 1st θ-derivative equal to 0. BUT sometimes that doesn't work (e.g. no stationary points or there are boundary constraints).
- Always check that the θ^* you found is indeed a maximizer! Usual way: check $d^2L(\theta)/d\theta^2 < 0$ when evaluated at $\theta = \theta^*$

Exercise 1

Let $X_1, \ldots, X_n$ i.i.d. from the pdf $f_\theta(x) = \frac{\theta}{x^2} I(x \geq \theta > 0)$

Find the MLE and the MoM estimator of θ

Exercise 2

Let $X_1, \ldots, X_n$ i.i.d. from the pdf $f_\theta(x) = \frac{1}{\theta} I(0 \leq x \leq \theta)$ with $\theta > 0$

Find the MLE and the MoM estimator of θ. Which one is better in terms of Mean Squared Error?

Review of Lectures 1-11

Albert Dorador

UW-Madison

March 3rd, 2023

Lecture 1

- Convergence in probability
- Weak Law of Large Numbers (Thm 5.5.2)
- Continuous mapping theorem (Thm 5.5.4)

Lecture 2

- Convergence in distribution
- Convergence in distribution vs in probability
- Central Limit Theorem (Thm 5.5.14)

Lecture 3

- Slutsky's theorem (Thm 5.5.17)

- 1st and 2nd order Delta method (Thm 5.5.24, 5.5.26)

- Sufficient statistic

Lecture 4

- Factorization Theorem (Thm 6.2.6)

- One-to-One functions of sufficient statistics

Lecture 5

- Exponential family

- Full-rank vs curved exponential family

- Minimal sufficient statistic + Thm 6.2.13

Lecture 6

- Ancillary statistic
- Complete statistic
- Ancillary vs Complete

Lecture 7

- Minimal sufficient vs complete sufficient

- Basu's Theorem (Thm 6.2.24)

- Thm 6.2.25

Lecture 8

- Method of Moments point estimator

- Maximum Likelihood point estimator

Lectures 9 and 10

- Advanced MLE (2D params, constraints, 0 or ∞-many stationary pts)

- Mean Squared Error (MSE) of an estimator

- Bayes estimator, prior and posterior distributions
  - Short-cut solution for CB 7.24

Lecture 11

- UMVUE

- How to find an unbiased estimator?

A couple tips

- CB page 69: Thm 2.4.1 (Leibnitz's Rule) useful when taking derivative on both sides of $E(g(T)) = 0$ to show $g = 0$ a.s. so T is complete
- CB page 229: Thm 5.4.4 (Formula for pdf of $X_{(j)}$)
- CB page 621: Table of common distributions

Midterm debrief, UMVUE, CRLB

Albert Dorador

UW-Madison

March 24th, 2023

Brief Midterm debrief (1/2)

- Overall, very good job!
- Questions 1 and 4b were the most challenging – please check the solutions carefully and get back to us should you have any questions

Quick fixes:

- Q1 and Q4: when working with absolute values, keep in mind the two cases: when the inside is non-negative and when it is negative. Common mistake: only focusing on the non-negative case.

Brief Midterm debrief (2/2)

- Q2: Always see if you can use the ‘proportional to’ shortcut when deriving the posterior distribution. If the likelihood and prior look similar in terms of the parameter of interest (like in the exam) you can drop the constants and greatly simplify your work
- Q3: Exponential family gives you sufficiency, but not minimality or completeness. For these, you need additional conditions. For example, if in addition we have full rank, then we have completeness too. In fact, T in a full rank exponential family is not only complete sufficient but minimal sufficient too.

CB 7.47 (UMVUE) + Bonus question

Suppose we make n independent measurements of the radius of a circle, each with a random error $\sim N(0, \sigma^2)$, with known σ^2. Find the best unbiased estimator (UMVUE) of the area of the circle. What if σ^2 is unknown?

Bonus question: if T is UMVUE of θ, is any one-to-one function g of T UMVUE of a) θ, or maybe b) $g(\theta)$?

CB 7.48a (Cramer-Rao Lower Bound)

Let $X_1, ..., X_n$ iid Ber(p). Show that the variance of the MLE of p attains the CRLB. Can we then conclude the MLE of p is in fact UMVUE?

RB x LS, LRT

Albert Dorador

UW-Madison

March 31st, 2023

Rao-Blackwell x Lehmann-Scheffe

RB

If W(X) is unbiased and S(X) is sufficient for θ, Var(E[W(X) | S(X)]) $\leq$ Var(W(X)), and E(W(X) | S(X)) is still unbiased (trivial proof).

LS

The unbiased estimator of θ that has the smallest variance must be a function of a complete sufficient statistic for θ.

RB x LS: If C(X) is complete sufficient for θ, E(W(X) | C(X)) is still unbiased and has smaller (or equal) variance than any other unbiased estimator, including E[W(X) | S(X)], which already was an improvement over W(X)

CB 7.52b (previous HW exercise)

$X_1, ..., X_n$ iid Poisson(λ). Prove that $\mathbb{E}(S^2|\bar{X}) = \bar{X}$.

Hypothesis testing and LRT

- Main big ideas of hypothesis testing
  - Goal: to find evidence for/against a hypothesis and reach a conclusion
  - Accept vs reject vs not accept vs not reject?

  - Acceptance and Rejection regions

- Example using LRT

$X_1, ..., X_n$ iid $N(\theta, a\theta)$, both a, θ unknown. Find the LRT of $H_0 : a = 1$ vs $H_1 : a \neq 1$, and its associated rejection region.

Power analysis in Hypothesis testing

Albert Dorador

UW-Madison

April 7th, 2023

Power function vs Power vs size vs level

Power function

$\beta(\theta) := P_{\theta \in \Theta}(X \in RR)$

Power

The probability that H_0 is correctly rejected, i.e. $P_{\theta \in \Theta_1}(X \in RR)$

Size

$sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$

Level

$sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$

- Size α implies level α, but the converse is not necessarily true.
- Intuition for including sup: worst-case scenario (conservative).

Example 1

Let $X \sim Beta(\theta, 1)$. We want to test $H_0 : \theta \leq 1$ vs $H_1 : \theta > 1$. Consider the test that rejects H_0 iff $X > 1/2$. Find the power function of this test and sketch it. What's the size of this test?

Example 2 (adapted from this site)

Performance on a math test is reported to be normally distributed with a mean of 40 and a standard deviation of 9.

You'd like to know whether the average performance in your school differs from the national average of 40, but you'd only care if the difference is big enough. How many students do you need to include in your sample to have power of 80% to detect a difference of 3 points using a two-tailed Z-test with alpha $= 0.05$? Hint: You'll need to know $\Phi^{-1}(0.975) = 1.96$ and $\Phi^{-1}(0.2) = -0.84$. What if you wanted power of 99% instead?

Uniformly Most Powerful tests

Albert Dorador

UW-Madison

April 14th, 2023

UMP and NP Lemma

UMP test

- The test with largest power in a given class of tests.
- The class we usually consider is the class of level (or size) α tests. Why?

Neyman-Pearson Lemma

- Tells you how to find the size-α UMP test when the hypothesis to be tested is of the type 'simple vs simple' i.e. $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$.

Any test that rejects H_0 when

$$\frac{f(x|\theta_1)}{f(x|\theta_0)} > k \quad \text{for some } k \geq 0 \quad \text{s.t.} \quad P_{\theta_0}(X \in RR) = \alpha$$

is a size-α UMP test. Note: if $< k$ then we don't reject. What if $= k$?

Example 1

Let $X \sim Beta(\theta, 1)$. Find the size-α UMP test of $H_0 : \theta = 1$ vs $H_1 : \theta = 2$.

MLR and Karlin-Rubin Theorem

MLR

Let $T(X)$ be a statistic. A family of pdf's or pmf's parametrized by θ has a MLR in $T(X)$ if for all $\theta_2 > \theta_1$ we have $\frac{f_{\theta_2}(x)}{f_{\theta_1}(x)}$ is a *non-decreasing* function of $T(x)$, on the set of x's s.t. at least one of $f_{\theta_2}(x)$ or $f_{\theta_1}(x)$ is positive.

Karlin-Rubin

Suppose that the parametric family of pdf's or pmf's has MLR in the statistic T. The test rejecting H_0 iff $T > c$ is a UMP test of size $\alpha = P_{\theta_0}(T > c)$ for testing $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$.
Extension: flip all inequalities.

Example 2

Let $X \sim Beta(\theta, 1)$. Find a UMP test of $H_0 : \theta \leq 1$ vs $H_1 : \theta > 1$ in case such a test exists.

MLR, p-value, Confidence Interval

Albert Dorador

UW-Madison

April 21st, 2023

Closing thoughts on MLR / Karlin-Rubin

2 equivalent approaches

- First find T sufficient and its distribution, then take ratio of likelihoods
- First take ratio of likelihoods and then find T and its distribution

MLR implies $\beta(\theta)$ is monotonic in θ

'Lemma B' in Prof. Jun Shao's notes: Suppose that the pdf or pmf of X is in a family parametrized by $\theta \in \mathbb{R}$ with MLR in $T(X)$. If $\psi(t)$ is a nondecreasing (respectively, nonincreasing) function of t, then $g(\theta) = E[\psi(T)]$ is a nondecreasing (respectively, nonincreasing) function of θ.

For the case $H_0 : \theta \leq \theta_0$, consider $\psi(t) = I(t > c)$.

For the case $H_0 : \theta \geq \theta_0$, consider $\psi(t) = I(t < c)$.

This means that we don't need to prove $\sup_{\theta \in \Theta_0} \beta(\theta) = \beta(\theta_0)$ when using Karlin-Rubin.

p-value

Thm 8.3.27

Let $W(X)$ be a statistic such that large values give evidence that H_1 is true. For each observed sample value x, define
$p(x) = \sup_{\theta \in \Theta_0} P_\theta(W(X) \geq W(x))$. Then $p(X)$ is a valid p-value.
Remarks:

- This is the most popular characterization of p-value, but there are others
- If $p(X)$ is a valid p-value, then the test that rejects H0 iff $p(X) \leq \alpha$ is a level α test

Example 1

Let $X \sim Ber(\theta)$. We wish to test $H_0 : \theta \leq 0.5$ vs $H_1 : \theta > 0.5$. We observe 7 successes out of 10 trials. Construct a reasonable test statistic $W(X)$ and calculate its associated p-value.

Confidence intervals

Main idea

Construct an interval of the form $C(X) := [L(X), U(X)]$ such that $P_\theta(\theta \in C(X))$ is large – say, 95%. Note: $C(X)$ is random (depends on your sample), while θ is fixed (but unknown).

Coverage probability

The probability that your random confidence interval contains the true parameter, i.e.

$$P_\theta(\theta \in C(X))$$

Confidence coefficient

The probability that your random confidence interval contains the true parameter in the 'worst-case scenario', i.e.

$$\inf_\theta P_\theta(\theta \in C(X))$$

Example 2

Let $X \sim Beta(\theta, 1)$. Let $Y = -[\log(X)]^{-1}$. Calculate the confidence coefficient of the interval $[y/2, y]$.

Inverting a test, pivotal quantities, shortest CI's

Albert Dorador

UW-Madison

April 28, 2023

Inverting a test

Relationship between Acceptance Region and Confidence Set

- One-to-one correspondence

$(x_1, ..., x_n) \in A(\theta_0) \iff \theta_0 \in C(x_1, ..., x_n)$
Inverting the AR of a level-α test gives you a $1 - \alpha$ confidence set

- Intuition?

The hypothesis test fixes the parameter value and asks what sample values are consistent with that fixed value θ_0 (i.e. the AR).
The confidence set fixes the sample values and asks what parameter values would make this sample values most plausible (i.e. the C(x)).

- Mechanics in practice

Start with $RR(\theta_0)$, then find $AR(\theta_0) = \{x : \theta_0 \in C(x)\} = \{x : C(x) \ni \theta_0\}$, then finally get $C(x) = \{\theta : x \in AR(\theta)\} = \{\theta : AR(\theta) \ni x\}$

Pivotal quantities

Definition

A random variable $Q(X_1, ..., X_n, \theta)$ is a pivotal quantity if its distribution is independent of any parameter.

Note that a pivotal quantity is NOT a statistic, since (in general) its definition will involve the parameter(s).

So, a pivotal quantity is in general not an ancillary statistic even if both random variables have a distribution that is independent of the parameters – unless Q happens not to involve θ.

Example 1

Let $X \sim \mathit{Beta}(\theta, 1)$. Find a pivotal quantity and use it to set up a confidence interval $C(X) = [L(X), U(X)]$ with confidence coefficient of 0.24.

Shortest CI's

Main idea

Shorter CI's are usually preferred because they tend to be more informative: saying that a real-valued parameter is between $-\infty$ and ∞ with very high confidence is pretty much useless.

Thm 9.3.2

For a unimodal pdf $f(x)$ (very common case), if the CI $[a, b]$ satisfies:

1. $\int_a^b f(x)\,dx = 1 - \alpha$
2. $f(a) = f(b) > 0$
3. $a \leq x^* \leq b$ where x^* is a mode of the pdf

then, $[a, b]$ is the shortest CI satisfying (1).
Note 1: Careful when dealing with scale families! Replace (2) by $a^2 f(a) = b^2 f(b) > 0$ (see CB page 444).
Note 2: unimodal includes flat peaks, even the uniform would be (trivially) 'unimodal'.

Example 2

Find L(X) and U(X) in Example 1 so that $C(X)$ is as short as possible.

Consistency and Efficiency of estimators

Albert Dorador

UW-Madison

May 5, 2023

Revisiting Consistency

Definition and strategies to prove consistency

See Discussion 2

Thm 10.1.5

Let T_n be a consistent estimator of θ. Then if $\lim_{n\to\infty} a_n = 1$ and $\lim_{n\to\infty} b_n = 0$,

$$U_n := a_n T_n + b_n$$

is still consistent for θ.

Consistency of MLE's

Thm 10.1.6

Under several regularity conditions, MLE's are consistent estimators.
Note: if the pdf of the data-generating process belongs to the Exponential Family (e.g. Normal, Exponential, Gamma,...) then all those conditions are automatically satisfied, and hence $\hat{\theta}_{MLE}$ is consistent for θ.
What's more, $g(\hat{\theta}_{MLE})$ is consistent for $g(\theta_{MLE})$, as long as g is continuous.

Limiting vs asymptotic variance

Def: Limiting variance

Let T_n be an estimator.
$\lim_{n\to\infty} k_n Var(T_n) = \tau^2 \in (0,\infty)$
We call τ^2 the limiting variance of T_n

Def: Asymptotic variance

Let T_n be an estimator.
Suppose $\lim_{n\to\infty} k_n(T_n - g(\theta)) = N(0,\sigma^2)$
We call σ^2 the asymptotic variance of T_n

Difference? Very important: limiting variance (morally) takes the limit of the variance as $n \to \infty$, whereas the asymptotic variance is the variance of the limiting distribution of T_n, i.e. you don't take the limit of the variance! The two 'flavors' above may coincide, but not always! Second flavor is usually more useful, e.g. it's used in the definition of asymptotic efficiency of an estimator.

Asymptotic efficiency of MLE's

Def: Asymptotic efficiency

An estimator T_n of $g(\theta)$ is asymptotically efficient if
$\sqrt{n}(T_n - g(\theta)) \to N(0, CRLB)$

Thm 10.1.12

Under a long list of conditions (which includes those in Thm 10.1.6), MLE's are asymptotically efficient.
HOWEVER, all those conditions are immediately satisfied if the pdf of the data-generating process belongs to the Exponential Family (e.g. Normal, Exponential, Gamma,. . .)

In summary, if Exponential Family, MLE estimator is BOTH consistent and asymptotically efficient.

Example: Extension of example in class

Let $X_1, ..., X_n$ iid with $E(X) = \mu < \infty$ and $V(X) = \sigma^2 < \infty$.

a) Find the limiting variance and the asymptotic variance of the MoM estimator of μ. Hint: CLT

b) Consider an arbitrary differentiable function g. What can we say about the limiting variance and the asymptotic variance of this function of the MoM estimator? Hint: Delta method

Q & A