Know For Midterm 2018

Mingyang Li

April 27, 2018

Abstract

This is meant for STAT512 by Professor Ewens at the University of Pennsylvania.

Part I
Concepts

1 Basic Aims of Statistics

- To **estimate** the range of a **parameter** optimally.
- To **test hypotheses** about the numerical value of the parameter optimally.

2 Statistics

Statistics is an inferential science bansed on observations involving randomness.

3 Quantities

- A "**random variable**", Y, follows a distribution which depends on some **parameter** θ.
  - We want to estimate the parameter θ, but -- more often -- we estimate an one-to-one function of it, $\tau(\theta)$. Whichever the case, the variable we want to estimate is called the **estimand**.
  - A function involving a R.V. Y, $f(Y, ...)$, is also a RV.
- Any function $f(Y)$ of the RV Y alone can be seen as an **estimator** for the estimand $\tau(\theta)$ associated with its distribution.
  - If the mean of this function, $\mathrm{E}[f(Y)]$, happens to be the estimand itself, then this function -- as an estimator -- is **unbiased**.
    - The **MVU ("minimal variance unbiased") estimator** of $\tau(\theta)$: The unbiased estimator of $\tau(\theta)$ whose variance is $\leq$ any other unbiased estimator of $\tau(\theta)$.
  - The value an estimator takes on (or "yields") is called an **estimate**.
- **Sufficient Statistics**, $w(Y_1, ..., Y_n)$, of a parameter, θ, is a function of the n iid RVs whose JDF will become independent of this parameter if w is given.
  - The **Minimal Non-Trivial Sufficient Statistics (MNTSS)** has two constraints over the ordinary definition of SS:
    - Minimality: Any other SS can be reduced (read: "transformed via a function") into this SS.
    - Non-triviality: The dimension of this SS should be $< n$. i.e, we have actually cut off some data / compressed the data.
- Others

- "Average" is not "mean":
  - "Mean" (μ) is a parameter.
  - "Average" can be either
    - a RV: $\bar{Y}$, or
    - a number: $\bar{y}$.
- Variance:$\mathrm{Var}(Y) = \mathrm{E}\left(Y^2\right) - \mathrm{E}^2(Y)$.

Part II
Formulas

4 Gamma Function

- Definition: $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$.
- Values:
  - $\Gamma(1) = \int_0^\infty e^{-t} dt = 1$
  - $\Gamma(2) = \int_0^\infty t \cdot e^{-t} dt = 1$
  - $\Gamma\left(\frac{1}{2}\right) = \int_0^\infty \frac{1}{\sqrt{t}} e^{-t} dt = \sqrt{\pi}$
- Recurrence Relation: $\Gamma(x) = (x-1) \cdot \Gamma(x-1)$
  - If x is integer: $\Gamma(x) = (x-1)!$
  - If $x > 0$ but is not int: Use the Recurrence Relation to strip the "x" to the lowest number $\in (1,2)$, then plug in the value as given in the table.
- Integrals involving Gamma Function:
  - $\int_0^\infty t^{x-1} e^{-ct} dt = c^{-x} \cdot \Gamma(x)$
  - $\int_0^\infty g(t) \cdot e^{-h(t)} dt$: often helpful to set $h(t) =: t'$.

5 The density functions of order statistics (OS) of n iid continuous RVs $Y_i \sim f(y)$

- The i-th OS alone: $f_{Y_{(i)}}\left(y_{(i)}\right) = \frac{n!}{(i-1)!(n-i)!} \left[F_Y\left(y_{(i)}\right)\right]^{i-1} \cdot f_Y\left(y_{(i)}\right) \cdot \left[1 - F_Y\left(y_{(i)}\right)\right]^{n-i}$
- The JDF of the i-th OS and the j-th OS: $f_{Y_{(i)},Y_{(j)}}\left(y_{(i)}, y_{(j)}\right) = \frac{n!}{(i-1)!(j-i)!(n-j)!} \left[F_Y\left(y_{(i)}\right)\right]^{i-1} \cdot f_Y\left(y_{(i)}\right) \cdot \left[F_Y\left(y_{(j)}\right) - F_Y\left(y_{(i)}\right)\right]^{j-i-1} \cdot f_Y\left(y_{(j)}\right) \cdot \left[1 - F_Y\left(y_{(j)}\right)\right]^{n-j}$

6 The Cramer-Rao Lower Bound of the Variance of an Estimator

- This Bound is **achievable**[1] iff the JDF $f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n; \theta)$ can be written in the "exponential family" form:

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n; \theta) = h(y_1, \ldots, y_n) \cdot e^{C(\theta) + D(\theta) \cdot \hat{\tau}_{\mathrm{MLU}}(y_1, \ldots, y_n)}$$

[2]

[1] "There exists an estimad of θ, $\tau(\theta)$, that has an unbiased estimator, $\hat{\tau}_{\mathrm{MLU}}(y_1, \ldots, y_n)$, whose variance is this value."
[2] As you convert it into this form, in the same time, the MVU estimator $\hat{\tau}_{\mathrm{MLU}}(y_1, \ldots, y_n)$ is identified.

- The Bound is given by:[3] $\mathrm{Var}\left[\hat{\tau}\left(y_1,...,y_n\right)\right] \geq$

$$\mathrm{Var}\left[\hat{\tau}_{\mathrm{MLE}}\left(y_1,...,y_n\right)\right] = \frac{-\left(\frac{\partial}{\partial\theta}\tau\left(\theta\right)\right)^2}{\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f_{Y_1,...,Y_n}\left(y_1,...,y_n;\theta\right)\right]} \begin{array}{l} \leftarrow \text{is } -1 \text{ if } \tau\left(\theta\right)=\theta \\ \leftarrow \text{is } n\cdot\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f_Y\left(y;\theta\right)\right] \text{ if } iid \end{array}$$

- Such estimad $\tau\left(\theta\right)$ is given by

$$\tau\left(\theta\right) = -\frac{\frac{\partial}{\partial\theta}C\left(\theta\right)}{\frac{\partial}{\partial\theta}D\left(\theta\right)}, \text{ or } = -\frac{A\left(\theta\right)}{B\left(\theta\right)}.$$

- After this estimad is found, its variance can be calculated by:
  - CR Bound
  - Traditional statistics
  - $\mathrm{Var}\left[\hat{\tau}\left(y_1,...,y_n\right)\right] = \frac{-1}{B(\theta)}\cdot\frac{d}{d\theta}\frac{A(\theta)}{B(\theta)}$

7 Sufficient Statistics (SS), $w(Y_1,...,Y_n)$, for a parameter θ

For n RVs, $Y_1,...,Y_n$, whose JDF is $f_{Y_1,...,Y_n}\left(y_1,...,y_n;\theta\right)$, a function $w := w(Y_1,...,Y_n)$ is a SS for the paramter θ iff the conditional distribution of those RVs – given w – is independent of θ: [4]

$$\begin{aligned} f_{Y_1,...,Y_n}\left(y_1,...,y_n|w;\theta\right), \text{ by definition} &\equiv \frac{f_{Y_1,...,Y_n}\left(y_1,...,y_n,w;\theta\right)}{f_W\left(w;\theta\right)} \\ \text{this is equivalently: } &= \frac{f_{Y_1,...,Y_n}\left(y_1,...,y_n;\theta\right)}{f_W\left(w;\theta\right)} \\ \text{core of this "iff"} \rightarrow &= h\left(Y_1,...,Y_n\right) \text{ (i.e., indep. of } \theta) \\ &\Leftrightarrow w(Y_1,...,Y_n) \text{ is a SS for } \theta. \end{aligned}$$

(Reason for the equivalence on the second line: Since w is a function of Y_i's, when Y_i's are all speficied, w is also determined.)

This expression is equivalent to:

$$f_{Y_1,...,Y_n}\left(y_1,...,y_n;\theta\right) = f_W\left(w;\theta\right)\cdot h\left(y_1,...,y_n\right) \Leftrightarrow w(Y_1,...,Y_n) \text{ is a SS for } \theta.$$

If the support of Y_i's is independent of the parameter θ, then this is also equivalent to:

$$f_{Y_1,...,Y_n}\left(y_1,...,y_n;\theta\right) = g\left(w;\theta\right)\cdot h\left(y_1,...,y_n\right) \Leftrightarrow w(Y_1,...,Y_n) \text{ is a SS for } \theta$$

where g is any function of w (and thus of θ).

7.1 Minimal, Non-Trivial Sufficient Statistics (MNTSS) – How To Find

7.1.1 When the support of Y_i's is independent of θ

Method 1: Factorization If:

- the JDF $f_{Y_1,...,Y_n}\left(y_1,...,y_n;\theta\right)$ can be factorized into $f_W\left(w;\theta\right)\cdot h\left(y_1,...,y_n\right)$, **and**
- $dim\left(w\right) < n$,

then w is a MNTSS of θ.

Method 2: Smith-Jones (preferred) Assuming 2 sets of readings are obtained from the same set of n RVs, $y_{11},...,y_{1n}$ and $y_{21},...,y_{2n}$, we look at the ratio of their probability:$R = \frac{f_{Y_1,...,Y_n}(y_{11},...,y_{1n};\theta)}{f_{Y_1,...,Y_n}(y_{21},...,y_{2n};\theta)}$. If this can be simplified to $\frac{g(y_{11},...,y_{1n})}{g(y_{11},...,y_{1n})}$[5], then this $g\left(Y_1,...,Y_n\right)$ is a MNTSS of θ.

[3] The MVU estimator $\hat{\tau}_{\mathrm{MLU}}\left(y_1,...,y_n\right)$ may not exist / be known by the time you evaluate this Bound.
[4] w is like a sponge on a wet plate $f_{Y_1,...,Y_n}$: it **sucks up** all the information contained in the water θ.
[5] i.e., the NUMERATOR and the DENOMINATOR are of the same form independent of θ

Method 3: Exponential Family If the JDF can be written in the "exponential family" form, then the then-called MVU estimator, $\hat{\tau}\left(Y_1, ..., Y_n\right)$is a MNTSS of θ.

7.1.2 When the support of Y_i's does depend on θ

- $(a, b\left(\theta\right))$: The only possible MNTSS is $Y_{\max}$("$Y_{(n)}$").
- $(a\left(\theta\right), b)$: The only possible MNTSS is $Y_{\min}$("$Y_{(1)}$").

Whichever the case, to confirm the MNTSS, $f_Y\left(y;\theta\right)$ should be able to be factorized into $g(y) \cdot h(\theta)$.

7.2 Rao-Blackwell Theorem

Supposing $w(Y_1, ..., Y_n)$ is a SS for the parameter θ:

1. The MVU estimator of the estimable function, $\tau\left(\theta\right)$, is some unique function of w.
2. This unique MVU estimator of $\tau\left(\theta\right)$ is $\mathrm{E}\left(\hat{\tau}|w\right)$, where $\hat{\tau}\left(Y_1, ..., Y_n\right)$ is ANY unbiased estimator of θ.

They lead to 2 approaches[6] to finding rhe MVU estimator of $\tau\left(\theta\right)$:

1. Consider only function of w as possibilities.
2. Find any unbaised estimator of $\tau\left(\theta\right)$, find its conditional expectation given w, which exactly must be the MVU estimator we want to find.

8 Maximum-Likelihood Estimation (One-Parameter Case)

- The JDF, $f_{Y_1,...,Y_n}\left(y_1, ..., y_n;\theta\right)$, without changing its expression, can be thought as a "likelihood"[7] $L\left(\theta; y_1, ..., y_n\right)$.
- The "Maximum Likelihood Estimat**or**" of θ, is denoted by $\hat{\theta}_{\mathrm{MLE}}\left(y_1, ..., y_n\right)$.
- The "Maximum Likelihood Estimat**e**" of θ, a value of $\hat{\theta}_{\mathrm{MLE}}\left(y_1, ..., y_n\right)$, is the value at which $L\left(\theta; y_1, ..., y_n\right)$is maximized (usually we look at $\ln L$ for simplicity).

8.1 Properties

- Invariance: Wraping the parameter $\theta$with a monotonic function modified its MLE-tor alike.
- Relation with SS: The MLE-tor, $\hat{\theta}_{\mathrm{MLE}}\left(y_1, ..., y_n\right)$ is the same as SS $w\left(y_1, ..., y_n\right)$.
  - thoug denot differ- ently
- Asymptotic results[8]:
  - MLE is asymptotically unbiased: As $n \to \infty$, $\mathrm{E}\left[\hat{\theta}_{\mathrm{MLE}}\left(y_1, ..., y_n\right)\right] \to \theta$.
  - MLE asymptotically attains a normal distribution: As $n \to \infty$, $\hat{\theta}_{\mathrm{MLE}}\left(y_1, ..., y_n\right) \sim N$.
  - MLE asymptotically achieves the CR Bound: As $n \to \infty$, $\mathrm{Var}\left(\hat{\theta}_{\mathrm{MLE}}\left(y_1, ..., y_n\right)\right) \to$the CR Bound.

9 Common Distributions

Name	Expression	Mean	Variance
Normal(μ, σ^2)	$\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(y-\mu)^2}{2\sigma^2}}$	μ	σ^2
Gamma(α, β)	$\frac{1}{\Gamma(\alpha)\beta^\alpha}y^{\alpha-1}e^{-\frac{y}{\beta}}$	$\alpha\beta$	$\alpha\beta^2$
Cauchy(θ, σ)	$\frac{1}{\pi\sigma}\cdot\frac{1}{1+\left(\frac{y-\theta}{\sigma}\right)^2}, \sigma > 0$	D.N.E.	D.N.E.
"Chi-2"$\chi^2(\nu)$	$\frac{1}{y^{\frac{\nu}{2}}\cdot\Gamma\left(\frac{\nu}{2}\right)}\cdot y^{\frac{\nu}{2}-1}\cdot e^{-\frac{y}{2}}$, $y > 0$	ν	2ν
Binomial(n, p)	$\mathrm{Prob}(Y = y) = \binom{n}{y}\theta^y\left(1-\theta\right)^{n-y}$, $y = 0, ..., n$	np	$np(1-p)$
Poisson(λ)	$\mathrm{Prob}(Y = y) = e^{-\lambda}\frac{\lambda^y}{y!}$, $y = 0, 1, ...$	λ	λ

[6] Neither guranteed to work.

[7] If we encountered such observation, $y_1, ..., y_n$, how likely is the parameter θ to take on a particular value of θ?

[8] Due to the Invariance Property, all $\hat{\theta}_{\mathrm{MLE}}\left(y_1, ..., y_n\right)$ here can also be a function of that.

9.1 Conversion Between Distributions

- (Any) Normal Distribution $\rightarrow$ Standard Normal Distribution: If $Y \sim N\left(\mu, \sigma^2\right)$, then $\frac{Y-\mu}{\sigma} \sim N(0,1)$.
- Standard Normal Distribution $\rightarrow$Chi-Square Distribution: If $Y \sim N(0,1)$, then $Y^2 \sim \chi^2(\nu = 1)$.

9.2 Properties of Chi-Square Distribution

- The sum of some χ^2-distributed RVs is another χ^2-distributed RV with a degree-of-freedom of the sum of those of the summand RVs: $Y_i \sim \chi^2(\nu_i)$ for $i = 1, ..., n \Rightarrow \sum_{i=1}^{n} Y_i \sim \chi^2\left(\sum_{i=1}^{n} \nu_i\right)$.

9.3 Properties of Poisson Distribution

- The sum of some Poisson-distributed RVs is another Poisson-distributed RV with a λ of the sum of those of the summand RVs: $Y_i \sim \text{Poisson}(\lambda_i)$ for $i = 1, ..., n \Rightarrow \sum_{i=1}^{n} Y_i \sim \text{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right)$.
- If the sum of some Poisson-distributed RVs is fixed, then any partial sum of these RVs is a binomially-distributed RV whose
  - index n is equal to the fixed total sum;
  - parameter p is equal to the ratio $\frac{\sum_{\text{partial sum}} \lambda_j}{\sum_{\text{total sum}} \lambda_i}$.
- (Continuing from above) When the summand RVs are iid, the partial sum of any j of them $\sim$ Binomial$\left(\text{total sum}, \frac{j}{n}\right)$.