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 based on observations involving randomness.

3 Quantities

• A "random variable", \( Y \), follows a distribution which depends on some parameter \( \theta \).
  
  – We want to estimate the parameter \( \theta \), 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, \( E[f(Y)] \), 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, \( \theta \), 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: Ŷ, or
  - a number: ŷ.

Variance: \( \text{Var}(Y) = \text{E}(Y^2) - \text{E}^2(Y) \).

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: \( f_{Y_i}(y(i)) = \frac{n!}{(i-1)!(n-i)!} \left[ F_Y(y(i)) \right]^{i-1} \cdot f_Y(y(i)) \cdot \left[ 1 - F_Y(y(i)) \right]^{n-i} \)

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

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:

\( f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n;\theta) = h(y_1,\ldots,y_n) \cdot e^{C(\theta)+D(\theta) \cdot \tau_{\text{MLU}}(y_1,\ldots,y_n)} \)

\(^1\)“There exists an estimad of \( \theta \), \( \tau(\theta) \), that has an unbiased estimator, \( \hat{\tau}_{\text{MLU}}(y_1,\ldots,y_n) \), whose variance is this value.”

\(^2\)As you convert it into this form, in the same time, the MVU estimator \( \hat{\tau}_{\text{MLU}}(y_1,\ldots,y_n) \) is identified.
• The Bound is given by:\(^3\) \(\text{Var} [\hat{\tau}(y_1, \ldots, y_n)] \geq \)
\[
\text{Var} [\hat{\tau}_{\text{MLE}} (y_1, \ldots, y_n)] = -\frac{\partial^2 \tau (\theta)}{\partial \theta^2} \left( \frac{\partial \ln f_{Y_1, \ldots, Y_n} (y_1, \ldots, y_n; \theta)}{\partial \theta} \right) \]
\[
\text{is} \ - \ 1 \ \text{if} \ \tau (\theta) = \theta
\]
• Such estimad \(\tau (\theta)\) is given by
\[
\tau (\theta) = -\frac{\partial}{\partial \theta} C (\theta) \quad \text{or} \quad -\frac{A (\theta)}{B (\theta)}
\]
• After this estimad is found, its variance can be calculated by:
  - CR Bound
  - Traditional statistics
  - \(\text{Var} [\hat{\tau}(y_1, \ldots, y_n)] = \frac{-1}{\text{dim}(\theta)} \cdot \frac{d A(\theta)}{d B(\theta)}\)

7 **Sufficient Statistics (SS), \(w(Y_1, \ldots, Y_n)\), for a parameter \(\theta\)**

For \(n\) RVs, \(Y_1, \ldots, Y_n\), whose JDF is \(f_{Y_1, \ldots, Y_n} (y_1, \ldots, y_n; \theta)\), a function \(w := w(Y_1, \ldots, Y_n)\) is a SS for the parameter \(\theta\) iff the conditional distribution of those RVs – given \(w\) – is independent of \(\theta\): \(^4\)

\[
f_{Y_1, \ldots, Y_n} (y_1, \ldots, y_n | w; \theta), \text{ by definition } \equiv \frac{f_{Y_1, \ldots, Y_n} (y_1, \ldots, y_n, w; \theta)}{f_W (w; \theta)}
\]
\[
\text{this is equivalently: } \frac{f_{Y_1, \ldots, Y_n} (y_1, \ldots, y_n; \theta)}{f_W (w; \theta)}
\]
\[
\text{core of this "iff" } \Rightarrow \quad h (Y_1, \ldots, Y_n) \quad \text{(i.e., indep. of } \theta) \Rightarrow w(Y_1, \ldots, Y_n) \text{ is a SS for } \theta.
\]

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

This expression is equivalent to:
\[
f_{Y_1, \ldots, Y_n} (y_1, \ldots, y_n; \theta) = f_W (w; \theta) \cdot h (y_1, \ldots, y_n) \quad \text{if } w(Y_1, \ldots, Y_n) \text{ is a SS for } \theta.
\]

If the support of \(Y_i\)'s is independent of the parameter \(\theta\), then this is also equivalent to:
\[
f_{Y_1, \ldots, Y_n} (y_1, \ldots, y_n; \theta) = g (w; \theta) \cdot h (y_1, \ldots, y_n) \quad \text{if } w(Y_1, \ldots, Y_n) \text{ is a SS for } \theta
\]

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

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

7.1.1 **When the support of \(Y_i\)'s is independent of \(\theta\)**

**Method 1: Factorization** If:
• the JDF \(f_{Y_1, \ldots, Y_n} (y_1, \ldots, y_n; \theta)\) can be factorized into \(f_W (w; \theta) \cdot h (y_1, \ldots, y_n)\), and
• \(\text{dim} (w) < n,\)
then \(w\) is a MNTSS of \(\theta\).

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

---

\(^3\)The MVU estimator \(\hat{\tau}_{\text{MLE}} (y_1, \ldots, y_n)\) may not exist / be known by the time you evaluate this Bound.
\(^4\)\(w\) is like a sponge on a wet plate \(f_{Y_1, \ldots, Y_n}\): it *sucks up* all the information contained in the water \(\theta\).
\(^5\)i.e., the NUMERATOR and the DENOMINATOR are of the same form independent of \(\theta\).
Method 3: Exponential Family  If the JDF can be written in the “exponential family” form, then the then-called MVU estimator, \( \hat{\tau}(Y_1, ..., Y_n) \) is a MNTSS of \( \theta \).

7.1.2 When the support of \( Y_i \)‘s does depend on \( \theta \)
- \((a, b(\theta))\): The only possible MNTSS is \( Y_{\alpha \beta} = \left(Y_{(n)} \right) \).
- \((a(\theta), b)\): The only possible MNTSS is \( Y_{\alpha \beta} = \left(Y_{(1)} \right) \).

Whichever the case, to confirm the MNTSS, \( f_Y(y; \theta) \) 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 \( \theta \):
1. The MVU estimator of the estimable function, \( \tau(\theta) \), is some unique function of \( w \).
2. This unique MVU estimator of \( \tau(\theta) \) is \( E(\hat{\tau}|w) \), where \( \hat{\tau}(Y_1, ..., Y_n) \) is ANY unbiased estimator of \( \theta \).

They lead to 2 approaches\(^6\) to finding the MVU estimator of \( \tau(\theta) \):
1. Consider only function of \( w \) as possibilities.
2. Find any unbiased estimator of \( \tau(\theta) \), 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, \( f_{Y_1, ..., Y_n}(y_1, ..., y_n; \theta) \), without changing its expression, can be thought as a “likelihood”\(^7\) \( L(\theta; y_1, ..., y_n) \).
- The “Maximum Likelihood Estimator” of \( \theta \), is denoted by \( \hat{\theta}_{MLE}(y_1, ..., y_n) \).
- The “Maximum Likelihood Estimate” of \( \theta \), a value of \( \hat{\theta}_{MLE}(y_1, ..., y_n) \), is the value at which \( L(\theta; y_1, ..., y_n) \) is maximized (usually we look at \( \ln L \) for simplicity).

8.1 Properties
- Invariance: Wrapping the parameter \( \theta \) with a monotonic function modified its MLE-tor alike.
- Relation with SS: The MLE-tor, \( \hat{\theta}_{MLE}(y_1, ..., y_n) \) is the same as SS \( w(y_1, ..., y_n) \).
- Asymptotic results\(^8\):
  - MLE is asymptotically unbiased: As \( n \to \infty \), \( E(\hat{\theta}_{MLE}(y_1, ..., y_n)) \to \theta \).
  - MLE asymptotically attains a normal distribution: As \( n \to \infty \), \( \hat{\theta}_{MLE}(y_1, ..., y_n) \sim N \).
  - MLE asymptotically achieves the CR Bound: As \( n \to \infty \), \( \text{Var}(\hat{\theta}_{MLE}(y_1, ..., y_n)) \to \text{the CR Bound} \).

9 Common Distributions

<table>
<thead>
<tr>
<th>Name</th>
<th>Expression</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal((\mu, \sigma^2))</td>
<td>( \frac{1}{\sqrt{2\pi \sigma}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} )</td>
<td>( \mu )</td>
<td>( \sigma^2 )</td>
</tr>
<tr>
<td>Gamma((\alpha, \beta))</td>
<td>( \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} y^{\alpha-1} e^{-\frac{y}{\beta}} )</td>
<td>( \alpha )</td>
<td>( \alpha \beta^2 )</td>
</tr>
<tr>
<td>Cauchy((\theta, \sigma))</td>
<td>( \frac{1}{\pi \sigma} \frac{1}{1+(\frac{y-\theta}{\sigma})^2} )</td>
<td>D.N.E.</td>
<td>D.N.E.</td>
</tr>
<tr>
<td>“Chi-2”(\chi^2(\nu))</td>
<td>( \frac{1}{\pi^{\nu/2} \Gamma(\frac{\nu}{2})} y^{\nu/2-1} e^{-\frac{y}{2}} ), ( \sigma &gt; 0 )</td>
<td>( \nu )</td>
<td>( 2\nu )</td>
</tr>
<tr>
<td>Binomial((n, p))</td>
<td>( \text{Prob}(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, y = 0, ..., n )</td>
<td>( np )</td>
<td>( np(1-p) )</td>
</tr>
<tr>
<td>Poisson((\lambda))</td>
<td>( \text{Prob}(Y = y) = e^{-\lambda} \frac{\lambda^y}{y!}, y = 0, 1, ... )</td>
<td>( \lambda )</td>
<td>( \lambda )</td>
</tr>
</tbody>
</table>

\(^6\)Neither guaranteed to work.

\(^7\)If we encountered such observation, \( y_1, ..., y_n \), how likely is the parameter \( \theta \) to take on a particular value of \( \theta \)?

\(^8\)Due to the Invariance Property, all \( \hat{\theta}_{MLE}(y_1, ..., y_n) \) here can also be a function of that.
9.1 Conversion Between Distributions

- (Any) Normal Distribution \(\rightarrow\) Standard Normal Distribution: If \( Y \sim N(\mu, \sigma^2) \), 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 \( \chi^2 \)-distributed RVs is another \( \chi^2 \)-distributed RV with a degree-of-freedom of the sum of those of the summand RVs: \( Y_i \sim \chi^2(\nu_i) \) for \( i = 1, \ldots, n \Rightarrow \sum_{i=1}^{n} Y_i \sim \chi^2(\sum_{i=1}^{n} \nu_i) \).

9.3 Properties of Poisson Distribution

- The sum of some Poisson-distributed RVs is another Poisson-distributed RV with a \( \lambda \) of the sum of those of the summand RVs: \( Y_i \sim \text{Poisson}(\lambda_i) \) for \( i = 1, \ldots, n \Rightarrow \sum_{i=1}^{n} Y_i \sim \text{Poisson}(\sum_{i=1}^{n} \lambda_i) \).
- 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 \( \sim \text{Binomial}(\text{total sum}, \frac{j}{n}) \).