STAT 611-600

Theory of Statistics - Inference
Lecture 4: Completeness & Likelihood

Tiandong Wang

Department of Statistics
Texas A&M University

Completeness

- Hard to directly test for minimality
- **Completeness**: Easier to test for and is sufficient to show minimality.

Idea: Completeness means that there is NO extra information in $T(X)$ that's unrelated to θ.

$\mapsto$ We cannot find a (non-constant) function $g(T)$ such that $\mathbb{E}_\theta[g(T)]$ does not depend on θ.

$\mapsto$ We cannot find a function $g(T)$ (that is not $= 0$ everywhere) such that $\mathbb{E}_\theta[g(T)] = 0$ for $\forall\theta$.

Definition. A statistic $T(\mathbf{X})$ is complete for θ if for any function g, $\mathbb{E}_\theta[g(T(X))] = 0$ for all θ implies

$$\mathbb{P}_\theta[g(T(X)) = 0] = 1, \qquad \forall\theta.$$

Completeness

- Hard to directly test for minimality
- **Completeness**: Easier to test for and is sufficient to show minimality.

Idea: Completeness means that there is NO extra information in $T(X)$ that's unrelated to θ.

$\mapsto$ We cannot find a (non-constant) function $g(T)$ such that $\mathbb{E}_\theta[g(T)]$ does not depend on θ.

$\mapsto$ We cannot find a function $g(T)$ (that is not $= 0$ everywhere) such that $\mathbb{E}_\theta[g(T)] = 0$ for $\forall\theta$.

Definition. A statistic $T(\mathbf{X})$ is complete for θ if for any function g, $\mathbb{E}_\theta[g(T(X))] = 0$ for all θ implies

$$\mathbb{P}_\theta[g(T(X)) = 0] = 1, \qquad \forall\theta.$$

Completeness: Exponential Family

Theorem

For an exponential family of full rank (i.e., minimal and full dimensional), $T(X)$ is complete and sufficient.

Example: For $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$,

$$T(\mathbf{X}) = \left(\sum_i X_i, \sum_i X_i^2 \right)$$

is complete and sufficient for (μ, σ^2).

Note: Any bijective transformation of a complete sufficient statistics is also complete and sufficient!

Hence,

$$T(\mathbf{X}) = \left(\overline{X}, S^2 \right)$$

is also complete and sufficient.

Completeness: Exponential Family

Theorem

For an exponential family of full rank (i.e., minimal and full dimensional), $T(X)$ is complete and sufficient.

Example: For $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$,

$$T(\mathbf{X}) = \left(\sum_i X_i, \sum_i X_i^2 \right)$$

is complete and sufficient for (μ, σ^2).
Note: Any bijective transformation of a complete sufficient statistics is also complete and sufficient!

Hence,

$$T(\mathbf{X}) = \left(\overline{X}, S^2 \right)$$

is also complete and sufficient.

Theorem

A sufficient statistic $T(\boldsymbol{X})$ is minimal if it is complete.

Proof:

- Suppose the sufficient statistic $T(\mathbf{X})$ is complete and the sufficient statistic $S(\mathbf{X})$ is minimal.
- By minimality of S, there exists a function H such that $S = H(T)$. Then
$$\mathbb{E}(T|S) = f(S) = f(H(T)) =: \tilde{f}(T).$$
- Consider
$$g(T) := T - \mathbb{E}(T|S).$$
- Clearly, $\mathbb{E}(g(T)) = 0$. Therefore, with probability 1,
$$T = \mathbb{E}(T|S) = f(S).$$
- Then T is a function of S. Since S is minimal, then so is T.

Completeness: Example

$X_1, \ldots, X_n$ are iid Unif$[0, \theta]$, $\theta > 0$:

- Sufficient stat: $T(\mathbf{X}) = \max_i X_i$.
- Is $T(\mathbf{X})$ complete?

Hints:

- Find the pdf of $T(\mathbf{X})$.
- Suppose g satisfies $\mathbb{E}_\theta[g(T)] = 0$ for all θ, then for all θ,

$$\frac{d}{d\theta}\mathbb{E}_\theta[g(T)] = 0.$$

Remarks

Complete suff is in some sense “more stringent" than min suff:

- Minimal suff statistics exist under mild conditions.
- Complete statistics do not always exist.

When both a min suff stat and a complete suff stat exist, they are equivalent.

Theorem

If T is complete suff statistic (and if a minimal suff stat exists), then T is also a minimal suff statistic.

Intuitively:

- Complete suff is our <u>ideal</u> notion of **optimal** data reduction.
- Min suff is our <u>achievable</u> notion of optimal data reduction.

Remarks

Complete suff is in some sense “more stringent" than min suff:

- Minimal suff statistics exist under mild conditions.
- Complete statistics do not always exist.

When both a min suff stat and a complete suff stat exist, they are equivalent.

Theorem

If T is complete suff statistic (and if a minimal suff stat exists), then T is also a minimal suff statistic.

Intuitively:

- Complete suff is our ideal notion of **optimal** data reduction.
- Min suff is our achievable notion of optimal data reduction.

Likelihood

Suppose $X \sim f(\cdot, \theta)$ pmf or pdf.
Now given $X = x$ is observed, the likelihood function is given by

$$L(\theta; \mathbf{x}) = f(\mathbf{x}; \theta)$$

- Same expression as pmf or pdf, but now x is fixed and θ is variable.
- Eg., binomial experiment. Decide to stop after 10 trials. 3 successes obtained.
- Eg., negative binomial experiment. Decide to stop after 3 successes. 10 trials were needed.

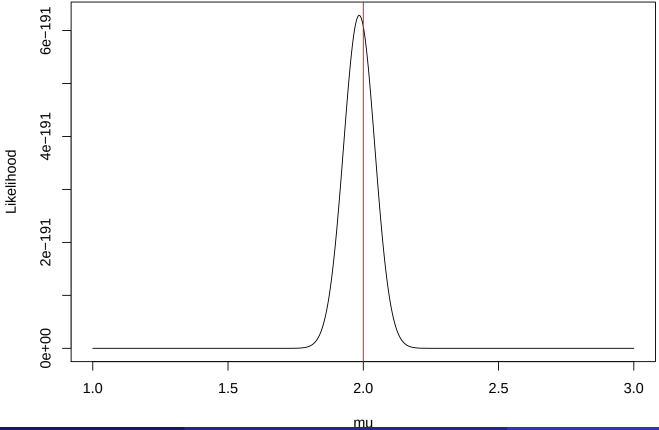
Examples of likelihood functions

Likelihood can be viewed as the degree of plausibility.

- An estimate of θ may be obtained by choosing the most plausible value i.e., where the likelihood function is maximized. This leads to one of the most important methods of estimation the maximum likelihood estimator (Chapter 7).
- For instance, in either example above, the likelihood function is maximized at 0.3.
- It is also tempting to interpret likelihood as probability about θ give the data.

Likelihood: example

Example: draw n samples from $N(\mu, 1)$, plot the likelihood function of μ.
Run Chapter6-Example.R



Likelihood Principle

Theorem

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are two sample points and there exists a constant $C(\boldsymbol{x}, \boldsymbol{y})$ such that:

$$L(\theta; \boldsymbol{x}) = C(\boldsymbol{x}, \boldsymbol{y}) L(\theta; \boldsymbol{y}) \qquad \forall \theta,$$

then the conclusions drawn from $\boldsymbol{x}$ and $\boldsymbol{y}$ should be identical.

Likelihood: Bayesian

- In the Bayesian paradigm the parameter θ is thought of as a random variable with some probability distribution $\pi(\theta)$ (the prior distribution)
- The likelihood function is considered as the conditional probability density function for x, given θ;
- Statistical analysis is just probability theory based on the joint distribution $\pi(\theta)f(\mathbf{x}|\theta)$ for θ and $\mathbf{x}$.
- Posterior distribution $f(\theta|\mathbf{x}) = c \cdot \pi(\theta)f(\mathbf{x}|\theta)$ (the normalizing constant $c = 1/\int_{\Theta} \pi(\theta)f(\mathbf{x}|\theta)d\theta$ is seldom important).