ST2132 More on Estimation

Sufficiency, Efficiency, and Duality with Testing Hypothesis

Semester 1 2022/2023

Poisson likelihood

$x_1, \ldots, x_n$ from Poisson(λ).

- $$L(\lambda) = \lambda^y e^{-n\lambda}, \qquad y = \sum_{i=1}^{n} x_i$$

  ML estimate of λ depends on data only through the sum. Data sets with the same sum give the same ML estimate.

- y is a realisation of $Y \sim$ Poisson($n\lambda$).

  $$\Pr(Y = y) = \frac{(n\lambda)^y e^{-n\lambda}}{y!}$$

  gives the same likelihood as above.

Bernoulli likelihood

$x_1, \ldots, x_n$ from Bernoulli(p).

- $$L(p) = p^y(1-p)^{n-y}, \qquad y = \sum_{i=1}^{n} x_i$$

  ML estimate of p depends on data only through the sum. If we only know y, but not the x's, we are fine.

- y is a realisation of $Y \sim$ Binomial(n, p). Seen: likelihood based on y is the same as above; Fisher information in y is the same as that in the x's.

Bernoulli vs binomial

- For ML estimation,
  a Binomial(n, p) RV is equivalent to n IID Bernoulli(p) RV's;
  a Multinomial$(n, \mathbf{p})$ is equivalent to n IID Multinomial$(1,\mathbf{p})$ RV's.

- Can simplify

  $\hat{\theta}_n$: ML estimator of θ based on either IID RV's $X_1, \ldots, X_n$ or a multinomial $(X_1, \ldots, X_r)$ with n trials.
  $\mathcal{I}(\theta)$: Fisher information in any X_i, or in the multinomial RV with 1 trial.

  to

  $\hat{\theta}_n$: ML estimator of θ based on IID RV's $X_1, \ldots, X_n$.
  $\mathcal{I}(\theta)$: Fisher information in any X_i.

Sufficient statistics

$\mathbf{x} = (x_1, \ldots, x_n)$: realisations of IID RV's with density $f(x|\theta)$.

A function $h(\mathbf{x})$ is a **sufficient statistic** for θ if the refined likelihood $L(\theta)$ depends on $\mathbf{x}$ only through $h(\mathbf{x})$.

- $\mathbf{x}$ from Poisson(λ). $y = \sum_{i=1}^n x_i$. Refined likelihood: $L(\lambda) = \lambda^y e^{-n\lambda}$, so y is sufficient for λ.
  Joint probability
  $$\frac{\lambda^y e^{-n\lambda}}{x_1! \cdots x_n!}$$
  depends on $\mathbf{x}$ through y and the product of factorials, but the latter is not in the refined likelihood.
- $\mathbf{x}$ from Bernoulli(p). $y = \sum_{i=1}^n x_i$. Refined likelihood: $L(p) = p^y(1-p)^{n-y}$, so y is sufficient for p.

Gamma and Normal

- $\mathbf{x}$ from Gamma(α, λ). Identify the sufficient statistics for α and λ.

- $\mathbf{x}$ from N(μ, σ^2). Identify the sufficient statistics for μ and σ.

More on Sufficiency

- Can a sufficient statistic involve θ? Will a sufficient statistic change its value if the entries of $\mathbf{x}$ are reordered?
- The usual definition of sufficient statistic goes through a conditional distribution. See Rice for detail.
- The refined likelihood approach is not standard, though its essence is in a factorisation theorem. See Rice for detail.

Throw raw data away?

- If the statistical model is correct, then ML estimation is fine with the sufficient statistics, even if you lose the raw data.

- Raw data is essential for checking if a model is reasonable. If you only have total number of heads in successive blocks of 100 coin tosses, you would not know if there is some systematic pattern within each block.

- We can never be certain that a model is correct (remember Box). Keep your raw data.

Definition of unbiasedness

$\hat{\theta}_n$: estimator of $\theta \in \Theta \subset \mathbb{R}$, based on n IID RV's.

- $\hat{\theta}_n$ is unbiased for θ if

$$\mathrm{E}(\hat{\theta}_n) = \theta, \qquad \forall \theta \in \Theta$$

- All our unbiased estimators satisfy the definition.

  For example, if the RV's are Bernoulli(p), $\mathrm{E}(\bar{X}) = p$ for any $p \in (0, 1)$.

Biased estimators

If $\mathrm{E}(\hat{\theta}_n) \neq \theta$ for only a single $\theta \in \Theta$, then $\hat{\theta}_n$ is biased.
This rarely happens with the usual estimators.

- For any IID RV's,
$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2, \qquad \forall \sigma > 0.$$
The bias is always negative, but can be removed: use $S^2 = \frac{n}{n-1}\hat{\sigma}^2$.
- If RV's are Bernoulli(p), let $\tilde{p} = 0.5$. $\mathrm{E}(\tilde{p}) = 0.5$: $\tilde{p}$ is perfect if $p = 0.5$, but is off if p is any other value. The bias looks impossible to fix.

Cramér-Rao inequality

$\hat{\theta}_n$: estimator of $\theta \in \Theta \subset \mathbb{R}$ based on IID RV's $X_1, \ldots, X_n$.
$\mathcal{I}(\theta)$: Fisher information in any X_i.

Theorem: If $\hat{\theta}_n$ is unbiased, then

$$\mathsf{var}(\hat{\theta}_n) \geq \frac{\mathcal{I}(\theta)^{-1}}{n}, \qquad \forall \theta \in \Theta$$

- $\frac{\mathcal{I}(\theta)^{-1}}{n}$ is called the Cramér-Rao lower bound (CRLB).
- If $\hat{\theta}_n$ is unbiased, then its variance cannot be smaller than the CRLB.
  What does the theorem tell you about a constant estimator, such as $\tilde{p}$?

Efficient estimation

Let $\hat{\theta}_n$ be unbiased for θ. It is efficient if its variance is the CRLB:

$$\mathrm{var}(\hat{\theta}_n) = \frac{\mathcal{I}(\theta)^{-1}}{n}$$

Which estimator is efficient?

- $X_1, \ldots, X_n$ IID Bernoulli(p). $\hat{p} = \bar{X}$.
- $X_1, \ldots, X_n$ IID Poisson(λ). $\hat{\lambda} = \bar{X}$.
- $X_1, \ldots, X_n$ IID $\mathrm{N}(\mu, \nu)$. $\hat{\nu} = S^2$.
- $X_1, \ldots, X_n$ IID $\mathrm{N}(\mu, \sigma^2)$. $\hat{\sigma}^2 = \frac{n-1}{n} S^2$.

ML estimators

- Since ML estimators are generally biased, efficiency is not so relevant. But they are asymptotically unbiased, allowing us to discuss asymptotic efficiency.
- For large n, the ML estimator $\hat{\theta}_n$ has

$$\mathrm{var}(\hat{\theta}_n) \approx \frac{\mathcal{I}(\theta)^{-1}}{n}$$

  Hence $\hat{\theta}_n$ is asymptotically efficient.
- ML estimator is about the best we can get. For large n, it is approximately unbiased, efficient, and normally distributed.

Duality between CI and testing hypothesis

- Introduction slide 2: Proportion p of 10,000 tickets are white. SRS of 100: 78 are white.
  Survey Sampling II slide 10: p is estimated as 0.78, estimated SE is 0.04. Approximate 90%-CI for p:

$$(0.71, 0.85)$$

- Imagine using the same data to test

$$H_0 : p = p_0, \qquad \text{vs} \qquad H_1 : p \neq p_0$$

- If p_0 in CI, P-value is > 0.1. Otherwise, P-value is < 0.1.
- Roughly speaking, a $(1-\alpha)$-CI for θ allows us to infer the P-value for $H_0 : \theta = \theta_0$, $H_1 : \theta \neq \theta_0$. If θ_0 is in CI, $P > \alpha$. Otherwise, $P < \alpha$.