Stat 301

topics:

statistical model $(P_\theta : \theta \in \Theta)$ $X_1, \cdots, X_n \overset{iid}{\sim} P_\theta$

① sufficiency & exponential family.

- factorization
- minimal sufficiency
- ancillary statistic
- completeness
- Rao-Blackwell thm.

② decision theory.

- loss $\ell(\hat{\theta}, \theta)$
- risk $\mathbb{E}\,\ell(\hat{\theta}, \theta)$.
- Bayes & minimax optimality.
- admissibility

James-Stein estimator.
(application in adaptive nonparametric estimation)

Neyman-Pearson lemma.

minimax lower bound via Le Cam two-point method.

③ estimation under constraint.

- unbiasedness (UMVUE, Lehmann-Scheffe).
- invariance. (location family, Pitman estimator).

④. likelihood & asymptotics.

- consistency of MLE.
- Fisher info & score.
- LAN & DQM.
- Cramer Rao lower bound.
- Hodges estimator.
- convolution theorem & local asymptotic minimaxity.
- Bernstein von-Mises thm.

books: ① E. Lehmann & G. Casella Theory of Point Estimation.

② E. Lehmann & J. Romano Testing Statistical Hypothesis.

③ I. Johnstone Gaussian sequence model.

④ A. van der Vaart Asymptotic Statistics

TAs: Chih-Hsuan Wu & Qing Yan.

homework: due Thursday.

midterm + final.

statistical model / experiment: $(P_\theta : \theta \in \Theta)$

data / observations: $X_1, \cdots, X_n \overset{iid}{\sim} P_\theta$

statistic: $T = T(X_1, \cdots, X_n)$ function of data.

Def 1: T is sufficient iff the conditional distribution of $X|T$ does not depend on θ.

Alice

$X_1, \cdots, X_n$

Bob.

$T = T(X_1, \cdots, X_n)$ (sufficient).

Bob's strategy: sample $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_n$ from the conditional distribution of $X|T$

$$(\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n) \stackrel{d}{=} (X_1, \ldots, X_n)$$

e.g. $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, 1)$. $T(X) = \bar{X}$ is sufficient.

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \Big| \bar{X} \sim N\left(\begin{pmatrix} \bar{X} \\ \bar{X} \\ \vdots \\ \bar{X} \end{pmatrix}, I_n - \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^T \right) \qquad I_n = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \qquad \mathbb{1}_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$I_n - \frac{1}{n} \mathbb{1}_n \mathbb{1}_n^T = \begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & & \\ -\frac{1}{n} & 1-\frac{1}{n} & & \\ & & \ddots & \\ & & & 1-\frac{1}{n} \end{pmatrix}.$$

Bob can sample $\begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \vdots \\ \tilde{X}_n \end{pmatrix} \sim$ this distribution.

$$\begin{cases} \mathbb{E}\tilde{X}_1 = \mathbb{E}\left(\mathbb{E}(\tilde{X}_1 \mid \bar{X})\right) = \mathbb{E}\bar{X} = \theta \\ \mathbb{E}\tilde{X}_1^2 = \mathbb{E}\left(\mathbb{E}(\tilde{X}_1^2 \mid \bar{X})\right) = \mathbb{E}\left(1 - \frac{1}{n} + \bar{X}^2\right) = 1 - \frac{1}{n} + \frac{1}{n} + \theta^2 = 1 + \theta^2. \\ \mathrm{Var}(\tilde{X}_1) = \mathbb{E}\tilde{X}_1^2 - (\mathbb{E}\tilde{X}_1)^2 = 1 + \theta^2 - \theta^2 = 1. \\ \mathbb{E}\tilde{X}_1\tilde{X}_2 = \mathbb{E}\left(\mathbb{E}(\tilde{X}_1\tilde{X}_2 \mid \bar{X})\right) = \mathbb{E}\left(-\frac{1}{n} + \bar{X}^2\right) = \mathbb{E}\left(-\frac{1}{n} + \frac{1}{n} + \theta^2\right) \\ \qquad = \theta^2. \\ \mathrm{Cov}(\tilde{X}_1, \tilde{X}_2) = \mathbb{E}\tilde{X}_1\tilde{X}_2 - (\mathbb{E}\tilde{X}_1)(\mathbb{E}\tilde{X}_2) = \theta^2 - \theta^2 = 0. \end{cases}$$

$$\begin{pmatrix} \tilde{X}_1 \\ \vdots \\ \tilde{X}_n \end{pmatrix} \sim N\left(\begin{pmatrix} \theta \\ \vdots \\ \theta \end{pmatrix}, I_n \right) \iff \tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n \overset{iid}{\sim} N(\theta, 1).$$

e.g. $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$. $T(X) = \sum_{i=1}^n X_i$ is sufficient.

$\sim \text{Binomial}(n, \theta)$

$$P(X=x \mid T=t) = \frac{P(X=x, T=t)}{P(T=t)}$$

$$P(X=x, T=t) = \begin{cases} P(X=x) & \sum_{i=1}^n x_i = t \\ 0 & \sum_{i=1}^n x_i \neq t \end{cases}$$

$$= \mathbb{1}\left\{\sum_{i=1}^n x_i = t\right\} P(X=x) = \mathbb{1}\left\{\sum_{i=1}^n x_i = t\right\} \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}$$

$$= \mathbb{1}\left\{\sum_{i=1}^n x_i = t\right\} \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} = \mathbb{1}\left\{\sum_{i=1}^n x_i = t\right\} \theta^t (1-\theta)^{n-t}.$$

$$P(T=t) = \binom{n}{t} \theta^t (1-\theta)^{n-t}$$

$$\mathbb{P}(X=x \mid T=t) = \mathbb{1}\left\{\sum_{i=1}^{n} x_i = t\right\} \frac{1}{\binom{n}{t}}.$$

e.g. $X_1, \cdots, X_n \overset{iid}{\sim} P_\theta$ $\quad T = \underbrace{(X_{(1)}, X_{(2)}, \cdots, X_{(n)})}_{\text{order statistic.}}$ is sufficient.

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

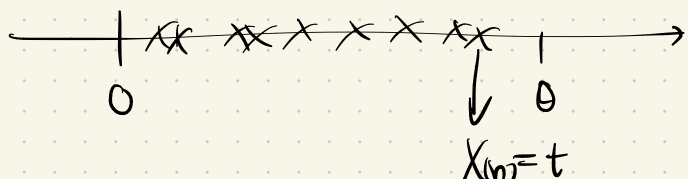
$\underbrace{X_1, \cdots, X_n \mid X_{(1)}, X_{(2)}, \cdots, X_{(n)}}_{\text{uniform } \frac{1}{n!}}$ # permutation is $n!$

e.g. $X_1, \dots, X_n \overset{iid}{\sim} \text{Unif}(0, \theta)$. $T = \max_{1 \le i \le n} X_i = X_{(n)}$ is sufficient.

$$X_{(1)}, \dots, X_{(n-1)} \mid X_{(n)} = t$$

are order statistics from $n-1$ iid samples from $\text{Unif}(0, t)$.

(homework).



Shall we always use sufficient statistic?

- information-theoret perspective — yes
- computational perspective. — sometimes no

Sampling $\tilde{X} \sim X|T$ can be NP hard.

(Montanari 2015, Bresler, Gamarnik & Shah 2014)