STAT 30100
Mathematical Statistics-1
Notes

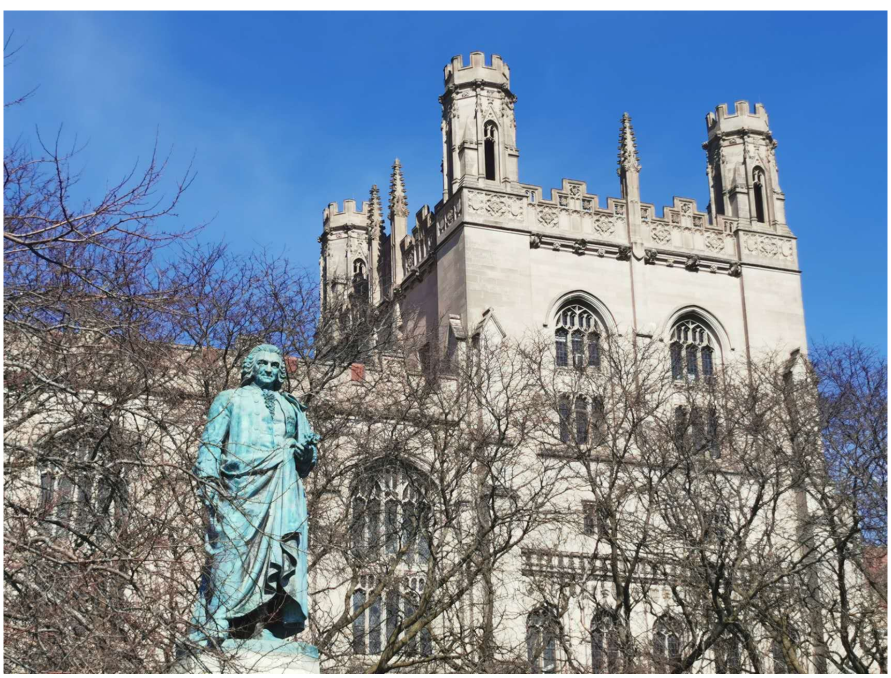
Zhiyuan Chen

November 13, 2024

Abstract

This is my class notes for UChicago STAT 30100 Mathematical Statistics-1. Given the challenging nature of the required graduate courses in statistics at UChicago, I believe it is essential to have comprehensive notes. Unfortunately, 301's notes are not provided on Canvas. And you can find most of the materials on Jinhong Du's page except 301 due to the change of the lecturer. That is why I decided to organize my notes, which may contain factual or typographic errors. If you are interested in helping refine this note, please contact me.

By the way, I would strongly recommend this course because Prof. Gao Chao taught very well: clear and full of passion. 301 is one of the best courses I have taken here, respecting Prof. Gao.



Chapter 1

Sufficiency

1.1 Sufficiency

$P_\theta : \theta \in \Theta$ is a statistical experiment, where Θ is the parameter space. Given $x_1, \ldots, x_n \overset{\text{i.i.d.}}{\sim} p_\theta$ for some $\theta \in \Theta$, our question is: can we summarize $x_1, \ldots, x_n$ by some statistic $T = T(x_1, \ldots, x_n)$ without loss of information?

Definition 1.1 (Sufficiency)**.** *$T = T(x_1, \ldots, x_n)$ is a sufficient statistic iff the conditional distribution of $x \mid T$ does not depend on $\theta \in \Theta$ i.e. $L(x \mid \tau)$ is the same for all $\theta \in \theta$*

Example 1.1. If $x_1, \ldots, x_n \sim N(\theta, 1), \quad \theta \in \mathbb{R}$, then $T = \bar{x}$ is sufficient.

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \Big| \bar{x} \sim N\left(\begin{pmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix} \cdot \begin{bmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ & 1 - \frac{1}{n} & \cdots & \vdots \\ & & \ddots & \vdots \\ & & & 1 - \frac{1}{n} \end{bmatrix} \right) = P_{\bar{x}}$$

$$\text{Draw: } \begin{pmatrix} \widetilde{x}_1 \\ \vdots \\ \widetilde{x}_n \end{pmatrix} \Big| \bar{x} \sim P_{\bar{x}}, \quad \text{Verify: } \begin{pmatrix} \widetilde{x}_1 \\ \vdots \\ \widetilde{x}_n \end{pmatrix} \overset{d}{=} N(\theta \mathbf{1}, I)$$

$$\begin{aligned}
\mathbb{E}[\widetilde{x_1}] &= \mathbb{E}\left(\mathbb{E}\left(\widetilde{x_1} \mid \bar{x}\right)\right) = \mathbb{E}[\bar{x}] = \theta \\
\mathrm{Var}(\widetilde{x_1}) &= \mathrm{Var}\left(\mathbb{E}\left(\widetilde{x_1} \mid \bar{x}\right)\right) + \mathbb{E}\left[\mathrm{Var}(\widetilde{x_1} \mid \bar{x})\right] \\
&= \mathrm{Var}(\bar{x}) + \mathbb{E}\left(1 - \frac{1}{n}\right) \\
&= \frac{1}{n} + 1 - \frac{1}{n} \\
&= 1 \\
\mathrm{Cov}(\widetilde{x_1}, \widetilde{x_2}) &= \mathbb{E}[\widetilde{x_1}\widetilde{x_2}] - \mathbb{E}[\widetilde{x_1}]\mathbb{E}[\widetilde{x_2}] \\
&= \mathbb{E}\left(\mathbb{E}\left(\widetilde{x_1}\widetilde{x_2} \mid \bar{x}\right)\right) - \theta^2 \\
&= \mathbb{E}\left[\mathrm{Cov}(\widetilde{x_1}, \widetilde{x_2} \mid \bar{x}) + \mathbb{E}(\widetilde{x_1} \mid x)\mathbb{E}(\widetilde{x_2} \mid x)\right] - \theta^2 \\
&= \mathbb{E}\left[-\frac{1}{n} + \widetilde{x}^2\right] - \theta^2 \\
&= -\frac{1}{n} + \frac{1}{n} + \theta^2 - \theta^2 \\
&= 0
\end{aligned}$$

$$\text{Thus, } \begin{pmatrix} \widetilde{x_1} \\ \vdots \\ \widetilde{x_n} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta \\ \vdots \\ \vdots \end{pmatrix}, I_n\right)$$

Example 1.2. If $x_1, \ldots, x_n \sim \mathrm{Bern}(\theta)$, $\theta \in [0,1]$, then $T = \bar{x} \cdot n$ is sufficient.

$$\begin{aligned}
\mathbb{P}(X = x \mid T = t) &= \frac{\mathbb{P}(X = x, T = t)}{\mathbb{P}(T = t)} \\
&= \frac{\mathbf{1}\left\{\sum_{i=1}^n x_i = t\right\} \theta^t (1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\
&= \mathbf{1}\left\{\sum_{i=1}^n x_i = t\right\} \cdot \binom{n}{t}^{-1} \\
&\perp \theta \\
\mathbb{P}(X = x, T = t) &= \begin{cases} \mathbb{P}(X = x), & \text{if } t = \sum_{i=1}^n x_i \\ 0, & \text{if } t \neq \sum_{i=1}^n x_i \end{cases} \\
&= \mathbf{1}\left\{\sum_{i=1}^n x_i = t\right\} \mathbb{P}(X = x) \\
&= \mathbf{1}\left\{\sum_{i=1}^n x_i = t\right\} \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i} \\
&= \mathbf{1}\left\{\sum_{i=1}^n x_i = t\right\} \theta^t (1-\theta)^{n-t} \\
T &\sim \mathrm{Binomial}(n, \theta) \\
\mathbb{P}(T = t) &= \binom{n}{t}\theta^t(1-\theta)^{n-t}
\end{aligned}$$

Theorem 1.1 (Factorization)**.** *Suppose $P_\theta, \theta \in \Theta$ is discrete or continuous, then $T = T(x)$ is sufficient iif $p(x \mid \theta) = g_\theta(T(x)) \cdot h(x)$.*

Proof. For the discrete situation, suppose $p(x \mid \theta) = g_\theta(T(x))h(x)$.

$$\begin{aligned}
\mathbb{P}(X = x \mid T = t) &= \frac{\mathbb{P}(X = x, T = t)}{\mathbb{P}(T = t)} \\
&= \frac{\mathbf{1}\{T(x) = t\} P(X = x)}{\mathbb{P}(T = t)} \\
&= \frac{\mathbf{1}\{T(x) = t\} g_\theta(T(x))h(x)}{\mathbb{P}(T = t)} \\
&= \frac{\mathbf{1}\{T(x) = t\} g_\theta(t)h(x)}{\mathbb{P}(T = t)} \\
\mathbb{P}(T = t) &= \sum_{T(x')=t} \mathbb{P}(X = x') \\
&= \sum_{T(x')=t} g_\theta(T(x'))h(x') \\
&= \sum_{T(x')=t} g_\theta(t)h(x') \\
&= g_\theta(t) \sum_{T(x')=t} h(x') \\
\mathbb{P}(X = x \mid T = t) &= \frac{g_\theta(t)\mathbf{1}\{T(x) = t\} h(x)}{g_\theta(t) \sum_{T(x')=t} h(x')} \\
&= \frac{\mathbf{1}\{T(x) = t\} h(x)}{\sum_{T(x')=t} h(x')}
\end{aligned}$$

□

Example 1.3. Suppose T is sufficient

$$\begin{aligned}
\mathbb{P}(X = x) &= \mathbb{P}(X = x, T(X) = T(x)) \\
&= \mathbb{P}(X = x \mid T(x) = T(x)) \cdot \mathbb{P}(T(x) = T(x))g_\theta(T(x))
\end{aligned}$$

where $h(x) = \mathbb{P}(X = x \mid T(x) = T(x))$, $g_\theta(T(x)) = \mathbb{P}(T(x) = T(x))g_\theta(T(x))$.

Example 1.4. Suppose $X_1 \dots X_n \sim U(0, \theta)$

$$\begin{aligned}
p(x \mid \theta) &= \prod_{i=1}^{n} \frac{1}{\theta} 1\{0 < x_i < \theta\} \\
&= \frac{1}{\theta^n} \mathbf{1}\left\{0 < x_{(1)}\right\} \mathbf{1}\, x_{(n)} < \theta\} \\
\Rightarrow \quad T &= X_{(n)}
\end{aligned}$$

1.2 Exponential Family

Distributions of exponential family has PDF like

$$p(x \mid \theta) = \exp\left(\sum_{j=1}^{d} \eta_j(\theta)T_j(x) - B(\theta)\right) \cdot h(x)$$

where d is dimensions of θ, $\eta_j(\theta)$ is natural parameter, $T_j(x)$ is sufficient statistic, $h(x)$ is the base measure of x and $B(\theta) = \log\left(\int \exp\left(\sum_{j=1}^{d} \eta_j(\theta)T_j(x)\right) h(x)\, dx\right)$.

Example 1.5. For $x \sim \exp(\theta)$

$$\begin{aligned}
p(x \mid \theta) &= \theta e^{-\theta x}\mathbf{1}\{x > 0\} \\
&= \exp\{-\theta x + \log\theta\} \perp \{x > 0\}
\end{aligned}$$

Example 1.6. for $x \sim N\left(\mu, \sigma^2\right)$

$$
\begin{aligned}
p(x \mid \theta) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \\
&= \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2} - \frac{1}{2}\log\left(2\pi\sigma^2\right)\right\} \\
&= \exp\left\{-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log\left(2\pi\sigma^2\right)\right\}
\end{aligned}
$$

where $d = 2,\ \theta = \left(\mu, \sigma^2\right)$. So

$$
\begin{cases} \eta_1(\theta) = -\frac{1}{2\sigma^2} \\ \eta_2(\theta) = \frac{\mu}{\sigma^2} \end{cases} \iff \begin{cases} T_1(x) = x^2 \\ T_2(x) = x \end{cases}
$$

$$
\begin{aligned}
h(x) &= 1 \\
B(\theta) &= -\frac{\mu}{2\sigma^2} - \frac{1}{2}\log\left(2\pi\sigma^2\right)
\end{aligned}
$$

Example 1.7. for Multinomial $x \sim M\left(p_0, \ldots, p_n; n\right)$

$$
\begin{aligned}
\mathbb{P}\left(X_0 = x_0, \ldots, X_s = x_s\right) &= \frac{n!}{x_0! \cdots x_s!} p_0^{x_0} \cdots p_s^{x_s} \\
&= \exp\left\{x_0 \log p_0 + \cdots + x_s \log p_s\right\} \cdot h(x) \\
&= \exp\left\{n \log p_0 + x_1 \log\frac{p_1}{p_0} + \cdots + x_s \log\frac{p_s}{p_0}\right\} \cdot h(x) \\
&\begin{cases} \eta_i = \log\frac{p_i}{p_0}, \quad i = 1, \ldots, s \\ T_i = x_i \\ A(\eta) = -n\log p_0 = -n\log\left(1 - \sum_{i=1}^s p_i\right) \end{cases}
\end{aligned}
$$

1.3 Joint Exponential Family

For $x_1, \ldots, x_n \overset{iid}{\sim} P_\theta$, we have

$$
\begin{aligned}
p(x \mid \theta) &= \exp\left\{\sum_{j=1}^d \eta_j(\theta) T_j(x) - B(\theta)\right\} h(x) \\
p\left(x_1, \ldots, x_n \mid \theta\right) &= \prod_{i=1}^n p\left(x_i \mid \theta\right) \\
&= \exp\left\{\sum_{j=1}^d \eta_j(\theta)\left[\sum_{i=1}^n T_j\left(x_i\right)\right] - nB(\theta)\right\} \cdot \prod_{i=1}^n h\left(x_i\right)
\end{aligned}
$$

Definition 1.2 (Canonical exponential family)**.** *P_η is exponential family of canonical form if*

$$
\begin{aligned}
P(x \mid \eta) &= \exp\left\{\sum_{j=1\eta}^d T_j(x) - A(\eta)\right\} h(x) \\
A(\eta) &= \log \int \exp\left\{\sum_{i=1}^d \eta_j T_j(x)\right\} h(x) d\mu(x)
\end{aligned}
$$

Definition 1.3 (Minimal exponential family)**.** *$P_\eta : \eta \in \mathcal{H}$ is a minimal exponential family of canonical form if the dimension cannot be reduced. i.e. $\eta_1, \ldots, \eta_d$ linear independent and $T_1(x), \ldots, T_d(x)$ linear independent.*

Remark 1.1 (Full rank and curved). For the minimal exponential family, if $\mathcal{H}$ contains an open rectangle of d dimensions, it is of full rank minimal. Otherwise, it is curved minimal.

Example 1.8 (Non-minimal). Let $\eta_2 = 3\eta_1$, then

$$\begin{aligned} P(x \mid \eta) &= \exp\left\{\eta_1 T_1(x) + \eta_2 T_2(x) - A(\eta)\right\} \cdot h(x) \\ &= \exp\left\{\eta_1 \left[T_1(x) + 3T_2(x)\right] - A(\eta)\right\} \cdot h(x) \end{aligned}$$

Example 1.9. For $\mathcal{N}(\mu, \sigma^2)$, From example 1.6 we know$\eta_1 = \frac{1}{2\sigma^2},\ \eta_2 = \frac{\mu}{\sigma^2};\ T_1(x) = -x^2,\ T_2(x) = x$. Then
1. $\mathcal{M} = \sigma^2 \Rightarrow \eta_2 = 1:$ non-minimal
2. $\mu = \sigma \Rightarrow \eta_2^2 = \frac{1}{\sigma^2} = \frac{\eta_1}{2}:$ non-linear relation $\Rightarrow$ minimal curved
3. $\mathcal{H} = \{(\eta_1, \eta_2) : \eta_1 > 0, \eta_2 \in \mathbb{R}\}:$ minimal full rank

1.4 Minimal Sufficient Statistics

Definition 1.4. *S is a minimal sufficient statistic if it is sufficient and for any sufficient $T, S = g(T)$.*

Corollary 1.1.1. *If there exists sufficient T s.t. sufficient $S \neq g(T)$, then S is not minimal sufficient.*

Lemma 1.1. *Suppose $\Theta_0 \subseteq \Theta$, if T is sufficient for $\theta \in \Theta$ and is minimal sufficient for $\theta \subseteq \Theta_0$, then T is minimal sufficient for $\theta \in \Theta$.*

Theorem 1.2. *$P_{\theta:} :\in \{\theta_0, \ldots, \theta_d\}$ have the same support, then $T = \left(\frac{P(x|\theta_1)}{P(x|\theta_0)}, \frac{P(x|\theta_2)}{P(x|\theta_0)}, \ldots, \frac{P(x|\theta_d)}{P(x|\theta_0)}\right)$ is minimal sufficient for $\theta \in \{\theta_0, \ldots, \theta_d\}$*

Proof.

$$\begin{aligned} &p\left(x \mid \theta_0\right) = 1 \cdot p\left(x \mid \theta_0\right), \\ &p\left(x \mid \theta_j\right) = T_j(x) p\left(x \mid \theta_0\right),\ j = 1, \ldots, d. \end{aligned}$$

By factorization, T is sufficient $\Leftrightarrow p(x \mid \theta) = g_\theta(T(x)) \cdot h(x)$, where

$$\begin{aligned} g_\theta(T(x)) &= \begin{cases} 1, & \theta = \theta_0 \\ T_j(x), & \theta = \theta_i, \quad i = 1, \ldots, d \end{cases} \\ h(x) &= p\left(x \mid \theta_0\right) \\ \Rightarrow &\quad T \text{ is sufficient} \\ \Rightarrow &\quad \frac{p\left(x \mid \theta_j\right)}{p\left(x \mid \theta_0\right)} = \frac{g_{\theta_j}(T(x))}{g_{\theta_0}(T(x))} \text{ is a function of } T \\ \Rightarrow &\quad \text{for } \forall \text{ sufficient } T', \frac{p\left(x \mid \theta_j\right)}{p\left(x \mid \theta_0\right)} \text{ is a function of } T' \\ \Rightarrow &\quad T \text{ is a function of } T' \end{aligned}$$

□

Example 1.10. For $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(\theta), \theta \in [0, 1] = \Theta$, let $\Theta_0 = \{0.3, 0.7\}$, then

$$T = \prod_{i=1}^{n} \frac{P\left(X_i \mid 0.7\right)}{P\left(X_i \mid 0.3\right)} = \frac{0.7^{\sum_1^n X_i} 0.3^{n - \sum_1^n X_i}}{0.7^{n - \sum_1^n X_i} 0.3^{\sum_1^n X_i}} = \left(\frac{7}{3}\right)^{\sum_1^n X_i} \left(\frac{3}{7}\right)^{n - \sum_1^n X_i}.$$

By Theorem 1.2, T is minimal sufficient for $\theta \in \Theta_0$. By Lemma 1.1 T is minimal sufficient for $\theta \in \Theta$.

1.5 Minimal sufficient exponential family

A minimal exponential family $(P_\eta : \eta \in \mathcal{H})$ can be written in form of

$$\begin{aligned} p(x \mid \eta) &= \exp\left\{\sum_{j=1}^{d} \eta_j T_j(x) - A(\eta)\right\} h(x) \\ &= \exp\left\{\langle \eta, T(x)\rangle - A(\eta)\right\} h(x), T(x) = \begin{bmatrix} T_1(x) \\ \vdots \\ T_d(x) \end{bmatrix} \end{aligned}$$

Theorem 1.3. *If $(P_\eta, \eta \in H)$ is a minimal exponential family, then $T(x) \in R^d$ is a sufficient minimum statistic.*