Statistical Inference (Tut-1)

RECAP:

(1) Basic prob.

- defn's: random variable
  - PMF/PDF
  - conditional prob.
  - independence
- basic distributions (NON-EXHAUSTIVE)
  - discrete: Bernoulli
    - Binomial
    - Poisson
    - exponential (family)
  - cts: Gaussian
    - Gamma

(2) Lectures – 1,2

- setup (parametric, random sample)
- statistic defn (coarsening sample space, equivalence reln)
- "sufficient" statistic
  - defn ⇒ Q1, 2, 3, 6
  - Factorization thm ⇒ Q4, 5, 7
  - "Indicator fun" method ⇒ Q8

"minimal" sufficient statistic.

(1) Probability basics

- Sample space : S = set of all possible outcomes of an experiment
- Event : $E \in 2^S$ ie. $E \subseteq S$.
- random variable : $X : S \to \mathbb{R}$

$$P[X=x] = P[\{s \in S : X(s) = x\}]$$

ex: Tossing a coin twice

$S = \{HH, HT, TH, TT\}$

X = # heads

ie. $X(HH) = 2$, $X(HT) = X(TH) = 1$, $X(TT) = 0$

$$P[X=1] = P[\{s \in S : X(s) = 1\}] = P[\{HT, TH\}]$$

conditional prob: $P[A|B] = \dfrac{P[A \cap B]}{P[B]}$, $P[B] \neq 0$

$$\Rightarrow P[A,B] = P[A|B] \cdot P[B] = P[B|A] \cdot P[A].$$

$$\Rightarrow P[A|B] = \frac{P[B|A] \cdot P[A]}{P[B]} \quad \text{(Bayes thm)}$$

CDF: Cumulative distribution fun

$F_X(x) = P[X \le x]$ st. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$ and non-dec., right cts.

if $F_X(\cdot)$ is cts → cts. rv.

step fun → discrete rv.

PMF of a discrete rv.

$$f_X(x) = P[X=x]$$

eg 1: Bernoulli (p)

$$f_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & x = 0 \\ 0 & \text{o/w.} \end{cases}$$

eg 2: Binomial (n,p)

$$f_X(x) = \begin{cases} {}^nC_x\, p^x (1-p)^{n-x} & x \in \{0,1,\dots,n\} \\ 0 & \text{o/w.} \end{cases}$$

for cts. distributions : $P[X=x] = 0 \ \forall x \in \mathbb{R}$.

Idea: $P[X=x] > 0$ introduces discontinuity in $F_X(\cdot)$

So, we consider $f_X(x) = \dfrac{d}{dx}[F_X(x)]$ st.

$$F_X(x) = P[X \le x] = \int_{-\infty}^{x} f_X(x)\, dx.$$

eg 1: $X \sim N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \forall x \in \mathbb{R}$$

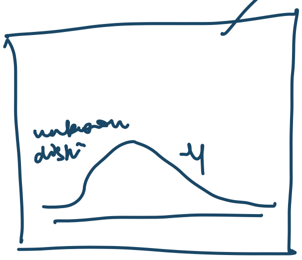
eg 2. $X \sim \text{Gamma}(\alpha, \beta)$

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \ ; \ x \geq 0, \ \alpha, \beta > 0$$

↳ "gamma fun" - generalization of factorial fun"

- $\Gamma(n) = (n-1)! \ \forall n \in \mathbb{Z}^+$
- $\Gamma(z) = z\Gamma(z-1) \ \forall z \in \mathbb{R}$

SETUP:



"sample" iid n times

ie. $X_1 \sim \mu$

$\vdots$

$X_n \sim \mu$

$\underset{\sim}{X} := (X_1, \ldots X_n) \sim \mu^n$

If we know the distribution but not the parameters ⇒ "parametric" inference.

eg 1: if $X \sim N(\mu, \sigma^2)$. We draw 'n' samples ie. $\underset{\sim}{X} = (X_1, X_2 \ldots X_n)$

1. sample space of X?
2. Parameter space of X?
3. PDF of $\underset{\sim}{X}$?
4. sample space of $\underset{\sim}{X}$?
5. Parameter space of $\underset{\sim}{X}$?

Ans:

"Statistic"

Let $X_1, X_2 \ldots X_n$ be a random sample of size 'n' and $T(x_1, x_2 \ldots x_n)$ be a real/vector valued function whose domain includes the sample space of $(X_1, X_2 \ldots X_n)$. Then the random variable/vector $Y = T(X_1, X_2 \ldots X_n)$ is called a "statistic". [shouldn't depend on unknown param].

- defines a form of "data reduction"

↓

consider $\underset{\sim}{x_1}, \underset{\sim}{x_2}$ st. $T(\underset{\sim}{x_1}) = T(\underset{\sim}{x_2})$

Do we really need to distinguish between $\underset{\sim}{x_1}$ and $\underset{\sim}{x_2}$?

- "coarsens" the sample space.

eg: coin toss: $X \sim \text{Bernoulli}(p)$ st $X \in \{0,1\}$

$T(\underset{\sim}{x}) := \sum_{i=1}^{n} x_i$ (# heads (say))

(verify this is a statistic).

sample space: $\{0,1\}^n \to$ size: 2^n

Image of x under T: ie.

$\mathcal{T} = \{t : t = T(\underset{\sim}{x}) \text{ for some } \underset{\sim}{x} \in \mathcal{X}\}$

$= \{0, 1 \ldots n\} \to$ size: $n+1$.

∴ T induces a partition of sample space.

SUFFICIENCY PRINCIPLE:

If $T(\underset{\sim}{X})$ is "sufficient" for θ, then any inference about θ should depend on $\underset{\sim}{X}$ through $T(\underset{\sim}{X})$. ie. if $\underset{\sim}{X} = \underset{\sim}{x_1}$, $\underset{\sim}{X} = \underset{\sim}{x_2}$ are observed st. $T(\underset{\sim}{x_1}) = T(\underset{\sim}{x_2})$ then inference about θ should be the same.

Defn: sufficient statistic for θ if $P(\underset{\sim}{x} \mid T(\underset{\sim}{x}) = t)$: doesn't depend on θ.

Q1: $X_1, X_2 \ldots X_n \sim \text{Poisson}(\lambda)$; $\lambda > 0$.

Is $T(\underset{\sim}{X}) = \sum X_i$ sufficient?

Ans:

$$P[\underset{\sim}{X} = \underset{\sim}{x} \mid T(\underset{\sim}{X}) = t]$$

$$= P[\underset{\sim}{X} = (x_1, x_2 \ldots x_n) \mid \textstyle\sum X_i = t]$$

$$= \begin{cases} \dfrac{P[X_1 = x_1, X_2 = x_2 \ldots X_n = x_n]}{P[\sum X_i = t]} & \text{if } \sum x_i = t \\ 0 & \text{o/w} \end{cases}$$

$X_i \sim \text{Poisson}(\lambda)$ i.i.

$$P[X=x] = \frac{\lambda^x e^{-\lambda}}{x!}$$

- $P[\underset{\sim}{X} = \underset{\sim}{x}]$ i.e.

$$P[X_1 = x_1, X_2 = x_2 \cdots X_n = x_n]$$
$$= \prod_{i=1}^{n} P[X_i = x_i] \qquad [\because \text{independent}]$$
$$= \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$
$$= \frac{\lambda^{\sum x_i} e^{-n\lambda}}{\prod x_i!}$$

- $P[\sum X_i = t]$

Fact: If $X_i \sim \text{Poisson}(\lambda)$, $\sum X_i \sim \text{Poisson}(n\lambda)$

Pf: $\underline{n=2}$ (Extend by induction).

$$P[X_1 + X_2 = x] = \sum_{k=0}^{x} P[X_1 = k, X_2 = x-k]$$

$(x \in \mathbb{Z}^+)$

$$= \sum_{k=0}^{x} P[X_1 = k] \cdot P[X_2 = x-k]$$
$$= \sum_{k=0}^{x} \frac{e^{-\lambda}\lambda^k}{k!} \cdot \frac{e^{-\lambda}\lambda^{x-k}}{(x-k)!}$$
$$= e^{-2\lambda}\lambda^x \sum_{k=0}^{x} \frac{1}{k!(x-k)!}$$
$$= \frac{e^{-2\lambda}\lambda^x}{x!} \underbrace{\sum_{k=0}^{x} {}^{x}C_k}_{=(1+1)^x}$$
$$= \frac{e^{-2\lambda}(2\lambda)^x}{x!}$$

$\therefore X_1 + X_2 \sim \text{Poisson}(2\lambda)$.

$\therefore$ Getting back to our def^n of sufficiency:

$$P[\underset{\sim}{X} = \underset{\sim}{x} \mid T(\underset{\sim}{X}) = t] = \frac{P[X_1 = x_1, \cdots X_n = x_n]}{P[T(X_1, X_2 \cdots X_n) = t]}$$

(recall $\sum x_i = t$)

$$= \frac{\left(\frac{\cancel{\lambda^{\sum x_i}}\,\cancel{e^{-n\lambda}}}{\prod x_i!}\right)}{\frac{\cancel{e^{-n\lambda}}\,\cancel{(n\lambda)^t}}{t!}} = \frac{t!}{n^t \prod x_i!} \quad (\text{No } \underline{\underline{\lambda}})$$

$\therefore$ Sufficient for λ.

$\underline{Q2}$: $X_1, X_2, X_3 \overset{iid}{\sim} \text{Bernoulli}(p)$

Show $U = X_1 X_2 + X_3$ not sufficient.

Pf: Single counterexample suffices.

Consider $P[\underset{\sim}{X} = \underset{\sim}{x} \mid U(\underset{\sim}{X}) = 0]$.

$$= \begin{cases} \dfrac{P[X_1 = x_1, X_2 = x_2, X_3 = x_3]}{P[U(\underset{\sim}{X}) = 0]} & \text{if } x_1 x_2 + x_3 = 0 \\ 0 & \text{o/w} \end{cases}$$

$U(\underset{\sim}{X}) = 0$

$\Rightarrow$ either $X_1 = 0, X_2 = 0, X_3 = 0$

$X_1 = 0, X_2 = 1, X_3 = 0$

$X_1 = 1, X_2 = 0, X_3 = 0$

$$P[U(\underset{\sim}{X}) = 0] = (1-p)^3 + 2p(1-p)^2$$

$$P[\underset{\sim}{X} = (0,0,0) \mid U(\underset{\sim}{X}) = 0]$$
$$= \frac{(1-p)^3}{(1-p)^3 + 2p(1-p)^2} \qquad \text{dependent on } p\,!$$

$\therefore$ Not sufficient.

"Exponential family"

PDF of the form:

$$f_X(x \mid \underset{\sim}{\theta}) = h(x)\, c(\underset{\sim}{\theta})\, e^{\underbrace{w(\underset{\sim}{\theta})^T T(x)}}$$

can also be written as: $\sum_{j=1}^{k} w_j(\theta) \cdot T_j(x)$

- support should be independent of θ.

$\underline{Q3}$: Verify it belongs to exponential family $\Rightarrow$ just check the form. (also L7 of notes)

sufficient statistic $\Rightarrow$ look at Q1.

FACTORIZATION TH^m

A statistic $T(\underset{\sim}{X})$ is a sufficient statistic for θ $\underline{\text{iff}}$ there exist functions $g(t \mid \theta)$ and $h(\underset{\sim}{x})$ such that $\forall\ \underset{\sim}{x} \in S$ and all $\theta \in H$

$$f(\underset{\sim}{x} \mid \theta) = g(T(\underset{\sim}{x}) \mid \theta)\, h(\underset{\sim}{x})$$

Allows us to get sufficient stats by inspecting the PDF.

Q4: $X \sim N(0, \sigma^2)$, $\sigma^2 = \theta$

To show: $T(x) = |x|$ is a sufficient stat.

Pf: Inspect PDF:

$$f_X(x) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-0)^2}{2\theta}}$$

$$= \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{x^2}{2\theta}}$$

want to get

$$f(x|\theta) = g(T(\underset{\sim}{x})|\theta) \cdot h(\underset{\sim}{x})$$

Note: $x^2 = |x|^2$

so, $f_X(x|\theta) = \left(\underbrace{\frac{1}{\sqrt{2\pi\theta}} e^{-\frac{|x|^2}{2\theta}}}_{\text{only depends on } \underset{\sim}{X} \text{ through } T(\underset{\sim}{X})}\right) \cdot 1$

$\therefore$ By Factorization th^{m}, $T(X) = |x|$ is sufficient.

Q5 $f_X(x|\theta) = \frac{\theta}{(1+x)^{1+\theta}}$, $x > 0$, $\theta > 0$

To show: $T = \prod_{i=1}^{n}(1+x_i)$ is sufficient for θ

Ans: Inspect PDF:

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}|\theta) = \prod_{i=1}^{n} \frac{\theta}{(1+x_i)^{1+\theta}}$$

$$= \frac{\theta^n}{\Big(\underbrace{\prod_{i=1}^{n}(1+x_i)}_{T(\underset{\sim}{X})}\Big)^{1+\theta}}$$

$$= \underbrace{\frac{\theta^n}{(T(\underset{\sim}{X}))^{1+\theta}}}_{g(T(\underset{\sim}{X})|\theta)} \cdot \underbrace{1}_{h(\underset{\sim}{x})}.$$

Again, by factorization th^{m}, $T = \prod_{i=1}^{n}(1+x_i)$ is sufficient.

$x \longmapsto \log x$ is a 1-1 mapping ie. preserves partitions ($\Rightarrow$ preserves sufficiency).

$\therefore$ $U = \sum \log(1+x_i)$ is also sufficient.

Q6: $X_1, X_2 \cdots X_n \overset{iid}{\sim}$

$$f_X(x|\mu,\sigma) = \frac{1}{\sigma} e^{-\frac{(x-\mu)}{\sigma}}$$

$\mu < x < \infty$

$0 < \sigma < \infty$

First, figure out PDF of $\underset{\sim}{X}$.

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}|\mu,\sigma) = \prod_{i=1}^{n} f_X(x_i|\mu,\sigma)$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma} e^{-\frac{(x_i-\mu)}{\sigma}}$$

Note: $f_X(x_i) \neq 0$ if $x_i > \mu$.

if $\exists i \in [n]$ st. $x_i \leq \mu$

$\Rightarrow f_{\underset{\sim}{X}}(\underset{\sim}{x}|\mu,\sigma) = 0$ (trivially indep. of μ, σ).

$$\therefore f_{\underset{\sim}{X}}(\underset{\sim}{x}|\mu,\sigma) = \begin{cases} \frac{1}{\sigma^n} e^{-\frac{(\Sigma x_i - n\mu)}{\sigma}} & \text{if } x_i > \mu \ \forall i \in [n] \\ 0 & \text{o/w} \end{cases}$$

Cleaner way to write:

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}|\mu,\sigma) = \frac{1}{\sigma^n} e^{-\frac{(\Sigma x_i - n\mu)}{\sigma}} \cdot \prod_{i=1}^{n} \mathbb{1}\{x_i > \mu\}$$

Notice: $\min_{i \in [n]} x_i > \mu$

$\iff x_i > \mu \quad \forall i \in [n]$

Let $\min_{i \in [n]} x_i = x_{(1)}$ ["order stats"]

$\therefore$

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}|\mu,\sigma) = \frac{1}{\sigma^n} e^{-\frac{(\Sigma x_i - n\mu)}{\sigma}} \cdot \mathbb{1}\{x_{(1)} > \mu\}.$$

Notice, we only need $(\Sigma x_i, x_{(1)})$ to find this.

$\therefore$ $T(\underset{\sim}{X}) = (\Sigma x_i, x_{(1)})$ is sufficient by factorization th^{m}. (why?)

Q7: $X_1, X_2 \dots X_n \sim \text{gamma}(\alpha, \beta)$

$$f_X(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

Again, start by writing the PDF:

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}|\alpha,\beta) = \frac{1}{(\Gamma(\alpha)\beta^\alpha)^n} (\Pi x_i)^{\alpha-1} e^{-\Sigma x_i/\beta}$$

$T(\underset{\sim}{X}) \overset{?}{=}$

$g(T(\underset{\sim}{x})|\alpha,\beta) \overset{?}{=}$

$h(\underset{\sim}{x}) \overset{?}{=}$

Minimal sufficient stats

Defn: A sufficient stat. $T(\underset{\sim}{X})$ is called a MSS if for any other sufficient stat. $T'(\underset{\sim}{X})$

$\underbrace{T(\underset{\sim}{X})}$ is a function of $T'(\underset{\sim}{x})$

$\Downarrow$

just means $T'(\underset{\sim}{x}) = T'(\underset{\sim}{y})$

$\Rightarrow T(\underset{\sim}{x}) = T(\underset{\sim}{y})$

Recall the partitioning picture. MSS is the "coarsest" possible partition of the sample space.

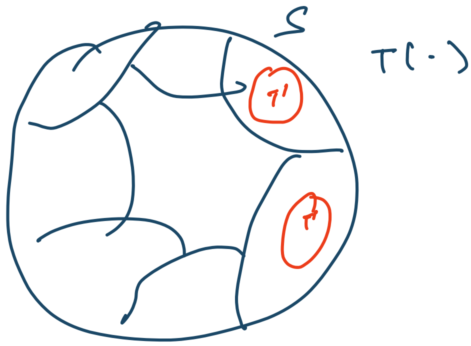
Thm (6.2.13) CB

· If $\exists\, T(\underset{\sim}{X})$ st $\forall\, \underset{\sim}{x}, \underset{\sim}{y} \in S$:

$$\frac{f(\underset{\sim}{x}|\theta)}{f(\underset{\sim}{y}|\theta)} = \text{constant in } \theta \text{ iff } T(\underset{\sim}{x}) = T(\underset{\sim}{y})$$

$\Rightarrow T(\underset{\sim}{X})$ is MSS.

Ex 1: $X_1, X_2 \dots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.
Show $(\bar{X}, S^2)$ is MSS.



$T_t := \{x \in S : T(x) = t\}$

$T'_{t'} = \{x \in S : T'(x) = t'\}$

$T'(x) = T'(y) = t'$

$\Rightarrow T(x) = T(y) = t$

$T_t \supseteq T'_{t'}$

Q8: $X_1, X_2, \dots X_n \overset{iid}{\sim} f_X(x|\theta) = \begin{cases} e^{-(x-\theta)} & \text{if } x > \theta \\ 0 & \text{o/w} \end{cases}$

Again, inspect PDF of $\underset{\sim}{X}$.

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}|\theta) = \begin{cases} e^{-(\sum x_i - n\theta)} & \text{if } x_i > \theta \ \forall i \in [n] \\ 0 & \text{o/w} \end{cases}$$

Cleaner way to write:

$$= e^{-(\sum x_i - n\theta)} \cdot \mathbb{1}\{x_{(1)} > \theta\}.$$

$$= \underbrace{e^{n\theta} \cdot \mathbb{1}\{x_{(1)} > \theta\}}_{g(T(\underset{\sim}{x})|\theta)} \cdot \underbrace{e^{-\sum x_i}}_{h(\underset{\sim}{x})}$$

where $T(\underset{\sim}{x}) := x_{(1)}$ ie

$= \min_{i \in [n]} x_i$

To show MSS:

let $\underset{\sim}{x}, \underset{\sim}{y} \in S$:

$$\frac{f_{\underset{\sim}{X}}(\underset{\sim}{x}|\theta)}{f_{\underset{\sim}{X}}(\underset{\sim}{y}|\theta)} = \frac{\cancel{e^{n\theta}} \cdot \mathbb{1}\{x_{(1)} > \theta\} \cdot e^{-\sum x_i}}{\cancel{e^{n\theta}} \cdot \mathbb{1}\{y_{(1)} > \theta\} \cdot e^{-\sum y_i}}$$

want constant in θ

iff $x_{(1)} = y_{(1)}$

$\therefore T(\underset{\sim}{x}) = x_{(1)}$ is MSS.