

Story #28

$x_1, x_2, \dots$ from $\{1, \dots, n\}$ $\Pr(X_i = 1/n)$

? $V_n = T_1 + T_2 + \dots + T_r + \dots + T_n$ where T_r is the time to see the r -th card starting from the $r-1$ card

(a) $V_r = T_1 + T_2 + \dots + T_r$ $T_i \sim \text{Geom}(p_i)$, $p_i = \frac{n-i+1}{n} = 1 - \frac{(i-1)}{n}$

$$E[V_r] = E[T_1 + \dots + T_r]$$

$$= \frac{1}{\frac{n-1+1}{n}} + \frac{1}{\frac{n-2+1}{n}} + \dots + \frac{1}{\frac{n-r+1}{n}} = 1 + \frac{n}{n-1} + \dots + \frac{n}{n-r+1}$$

(b) $f_{T_1, \dots, T_r}(t_1, \dots, t_r) = 1 \times (1 - \frac{1}{n}) (\frac{1}{n})^{t_2-1} (1 - \frac{2}{n}) (\frac{2}{n})^{t_3-1} \times \dots \times (1 - \frac{r-1}{n}) (\frac{r-1}{n})^{t_r-1}$

• $f_{\text{Geom}}(k) = (1-p)^{k-1} p$

• $\Pr(X > m+q | X > q) = \Pr(X > m)$ memorylessness

with log-likelihood $\ell_r(n) = \sum_{i=2}^r \{ (t_i - 1) \log(\frac{i-1}{n}) + \log(\frac{i-1}{n}) \}$

and sufficient statistic $V_n = T_2 + \dots + T_r$ for n

New concepts

- sufficient statistic
- (log-)likelihood

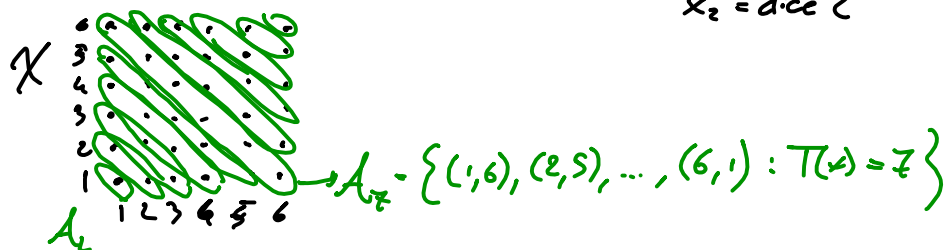
• a statistic $T(x)$ is a function $T: \mathcal{X} \rightarrow \mathbb{R}^p$ useful to summarize some quantity of interest (key feature of the ^{sample} population)

→ data reduction or data summary

→ it defines a partition of $\mathcal{X}$, $\mathcal{Z} = \{t: t = T(x) \text{ for some } x \in \mathcal{X}\}$

$T(x)$ partitions the sample space into sets $A_t = \{x: T(x) = t\}$

Example: sum of two dices $T(x) = x_1 + x_2$ $x_1 = \text{dice 1}$
 $x_2 = \text{dice 2}$



- A statistic is sufficient for a parameter θ when it summarises all information about θ that is contained in the sample

SUFFICIENCY PRINCIPLE: if $T(x)$ is a sufficient statistic for θ , then any inference about θ depends on the sample X only through $T(x)$

→ if $T(x) = T(y)$, then the inference about θ is the same whether one observes x or y

Definition 6.2.1: A statistic $T(X)$ is a sufficient statistic for θ if the conditional distribution of X given $T(x)$ does not depend on θ

Theorem 6.2.2 if $p(x|\theta)$ is the joint pmf (or pdf) of X and $q(t|\theta)$ is the pmf (or pdf) of $T(X)$, then $T(x)$ is a sufficient statistic for θ if

$\forall x \in X$, $p(x|\theta)/q(T(x)|\theta)$ does not depend on θ

Proof

$$\begin{aligned} P_{\theta}(X=x | T(X)=t) &= 0 \text{ if } t \neq T(x) \\ P_{\theta}(X=x | T(X)=T(x)) &= \frac{P_{\theta}(X=x \cap T(X)=T(x))}{P_{\theta}(T(X)=T(x))} \\ &= \frac{P_{\theta}(X=x)}{P_{\theta}(T(X)=T(x))} = \frac{p(x|\theta)}{q(T(x)|\theta)} \end{aligned}$$

$\{x=x\}$ is a subset of $T(x)=T(x)$

$(1,2,3) \rightarrow T(x)=6$
 $(2,2,2) \rightarrow T(x)=6$
 $(3,3,0) \rightarrow T(x)=6$

Example 6.2.4 $X_1, \dots, X_n$ iid $X_i \sim N(\mu, \sigma_0^2)$ σ_0^2 is known
 $T(x) = \bar{X}$ is sufficient for μ

$$\begin{aligned} f(x|\mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2\sigma_0^2}(x_i - \mu)^2\right\} \\ &= (2\pi\sigma_0^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2\right\} \\ &= (2\pi\sigma_0^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 + 0\right)\right\} \end{aligned}$$

$$q(\bar{x}|\mu) = (2\pi\sigma_0^2)^{-n/2} \exp\left\{-\frac{n}{2\sigma_0^2}(\bar{x} - \mu)^2\right\}$$

$$\frac{f(x|\mu)}{q(\bar{x}|\mu)} = n^{-1/2} (2\pi\sigma_0^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 - n(\bar{x} - \mu)^2\right)\right\}$$

this ratio does not depend on $\mu \Rightarrow \bar{X}$ is a sufficient statistic for μ

→ issues: in the story #28, we do not know the distn of V_r
 more generally, this theorem only allows us to check if a statistic is sufficient,
 no clever way to find it

Theorem 6.2.6 (Factorization theorem) Let $f(x|\theta)$ denote the joint pdf/pmf of X . A statistic $T(X)$ is sufficient for θ IFF

$$\exists g(t|\theta) \text{ and } h(x), \text{ function s.t. } \forall x \in \mathcal{X} \text{ and } \theta \in \Theta, \\ f(x|\theta) = g(T(x)|\theta) h(x)$$

Example (continued)

$$f(x|\mu) = \underbrace{(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2\right\}}_{h(x)} \underbrace{\exp\left\{-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2\right\}}_{g(\bar{x}|\mu)}$$

$$f(x|\mu) = h(x)g(T(x)|\mu) \Rightarrow T(x) = \bar{x} \text{ is a sufficient statistic for } \mu$$

Consider the case with σ^2 unknown

$$f(x|\mu) = \underbrace{(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{n-1}{2\sigma^2} S^2\right\}}_{g(\bar{x}, S^2|\mu, \sigma^2)} \underbrace{\exp\left\{-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2\right\}}_{h(x)}$$

Then $(\bar{x}, S^2)$ is a sufficient statistic for (μ, σ^2)

- all the information about the Gaussian is contained in $\bar{x}, S^2$
- the sufficient statistics are model specific

back to story #28

$$\begin{aligned} f_{t_1, \dots, t_n}(t_1, \dots, t_n) &= 1 \cdot \left(\frac{1}{n}\right)^{t_1-1} \left(1 - \frac{1}{n}\right) \cdot \left(\frac{2}{n}\right)^{t_2-1} \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{r-1}{n}\right) \left(\frac{r-1}{n}\right)^{t_r-1} \\ &= \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{r-1}{n}\right) \cdot n^{t_1-1} \cdot n^{t_2-1} \dots n^{t_r-1} \cdot 1 \cdot 2 \cdot \dots \cdot (r-1)^{t_r-1} \\ &= \prod_{i=1}^{r-1} \left(1 - \frac{i}{n}\right) \times n^{r - \sum_{i=1}^r t_i} \times \prod_{i=1}^{r-1} i^{t_i-1} \\ &= \underbrace{\prod_{i=1}^{r-1} \left(1 - \frac{i}{n}\right) \cdot n^{r - \sum_{i=1}^r t_i}}_{g(\sum_{i=1}^r t_i | n)} \cdot \underbrace{\prod_{i=1}^{r-1} i^{t_i-1}}_{h(t)} \Rightarrow \sum_{i=1}^r t_i \text{ is a sufficient statistic for } n \end{aligned}$$

Theorem 6.2.10 Let $X_1, \dots, X_n$ be iid from a pdf/pmf that belongs to the exponential family on the form $f(x|\theta) = h(x) c(\theta) \exp\left\{\sum_{i=1}^k \psi_i(\theta) t_i(x)\right\}$

Then $T(x) = \left(\sum_{j=1}^n t_1(x_j), \dots, \sum_{j=1}^n t_k(x_j)\right)$ is a sufficient statistic for $\theta = (\theta_1, \dots, \theta_k)$

For any problem, there are in general many sufficient statistics.

E.g., for the $N(\mu, \sigma^2)$

• $T(x) = X_1, \dots, X_n$

• $T(x) = X_{(1)}, \dots, X_{(n)}$

• $T(x) = (\bar{X}, S^2)$

• $T(x) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$

← the random sample is always a sufficient statistic Ex 4.20

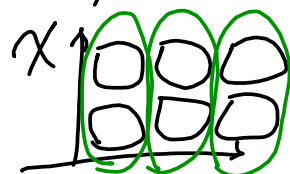
} an one-to-one function of a sufficient statistic is also a sufficient statistic

Is there a "best" sufficient statistic?

Idea: achieve the most data reduction while still retaining all information

Definition 6.2.11 A sufficient statistic $T(x)$ is called a minimal sufficient statistic if, for any other sufficient statistic $T'(x)$, $T(x)$ is a function of $T'(x)$

I.e., if $T'(x) = T'(y)$, then $T(x) = T(y)$



$B_{t'} = \{x \in X : T'(x) = t'\}$

$A_t = \{x \in X : T(x) = t\}$

→ the partition associated to the minimal sufficient statistic (A_t) is the coarsest possible partition of X

→ $B_{t'}$ is a subset of some A_t

→ any one-to-one function of a minimal sufficient statistic is also a minimal sufficient statistic

Theorem 6.2.13 Let $f(x|\theta)$ be the pdf/pmf of a r.v. X . Suppose $\exists$ a function $T(x)$, st.

$\forall x, y \in X$, the ratio $f(x|\theta)/f(y|\theta)$ does not depend on θ IFF $T(x) = T(y)$

Then $T(x)$ is a minimal sufficient statistic for θ proof ex 4.22

Example (cont.) $X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$

μ, σ^2 unknown, x, y are points of the sample

$$\frac{f(x|\mu, \sigma^2)}{f(y|\mu, \sigma^2)} = \frac{(\frac{1}{2\pi\sigma^2})^{n/2} \exp\{-\frac{1}{2\sigma^2}(n(\bar{x}-\mu)^2 + (n-1)S_x^2)\}}{(\frac{1}{2\pi\sigma^2})^{n/2} \exp\{-\frac{1}{2\sigma^2}(n(\bar{y}-\mu)^2 + (n-1)S_y^2)\}} \\ = \exp\left\{-\frac{1}{2\sigma^2}(n(\bar{x}-\bar{y})^2 + 2n(\bar{x}-\bar{y})(\bar{x}-\bar{y}) + (n-1)(S_x^2 - S_y^2))\right\}$$

the ratio is independent on (μ, σ^2) IFF $\bar{x} = \bar{y}$ and $S_x^2 = S_y^2$. Therefore $(\bar{x}, S^2)$ is a minimal sufficient statistic for (μ, σ^2)