STA732
Statistical Inference

Lecture 03: Sufficient Statistics

Yuansi Chen

Spring 2022

Duke University

https://www2.stat.duke.edu/courses/Spring22/sta732.01/

Recap from Lecture 02

Introduced exponential families: many good properties

- Natural parameter space is convex
- Easy joint density of i.i.d. random variables
- Easy sufficient statistics
- Easy moments

Goal of Lecture 03

1. Define sufficiency
2. Factorization theorem
3. Minimal sufficiency

Chap. 3 in Keener or Chap. 1.6 in Lehmann and Casella

Sufficiency

Motivation for sufficiency

Coin flipping experiment: suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim}$ Bernoulli(θ). Then the joint density is

$$\begin{aligned} p_\theta(X_1 = x_1, \ldots, X_n = x_n) &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \end{aligned}$$

Let $T(X) = \sum_{i=1}^n X_i$. $T(X)$ follows Binomial distribution Binom(n, θ), with $p_\theta(T(X) = t) = \theta^t(1-\theta)^{n-t}\binom{n}{t}$

Motivation for sufficiency

Coin flipping experiment: suppose $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim}$ Bernoulli(θ). Then the joint density is

$$p_\theta(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$
$$= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}$$

Let $T(X) = \sum_{i=1}^n X_i$. $T(X)$ follows Binomial distribution Binom(n, θ), with $p_\theta(T(X) = t) = \theta^t(1-\theta)^{n-t}\binom{n}{t}$

It seems that to estimate θ it is sufficient to know $T(X)$, but $T(X)$ did throw away data. How to justify?

Sufficient statistics

Suppose X has distribution from a family $\mathscr{P} = \{P_\theta, \theta \in \Omega\}$. $T(X)$ is a sufficient statistics if for every t and θ, the conditional distribution of X under P_θ given $T = t$ does not depend on θ.

Back to the coin flipping example

$$\begin{aligned} p_\theta(X = x \mid T = t) &= \frac{p_\theta(X = x, T = t)}{p_\theta(T = t)} \\ &= \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\mathbf{1}_{\sum x_i = t}}{\theta^t(1-\theta)^{n-t}\binom{n}{t}} \\ &= \frac{1}{\binom{n}{t}}\mathbf{1}_{\sum x_i = t} \end{aligned}$$

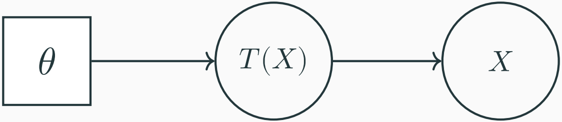
Conditioned on $T(X) = t$, X only takes values on sequences such that $\sum X_i = t$ and it is uniformly distributed. The conditional distribution does no depend on θ!

Interpretation of sufficiency

Interpretation 1: sufficiency from generative model perspective

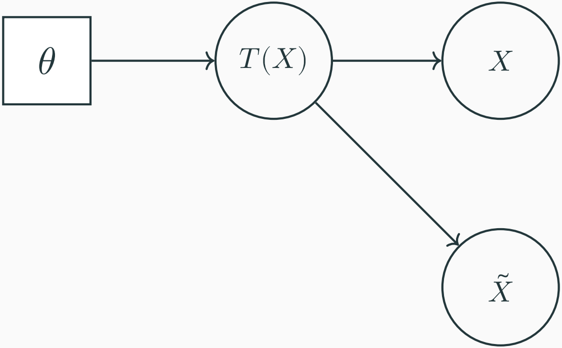
- We care about X because
  - X is generated according to P_θ
  - X can be used to infer properties of θ
- Sufficiency is saying that $T(X)$ is informative enough
- We can think of data being generated in two stages
  1. Generate T: distribution depends on θ
  2. Generate $X \mid T$: distribution does not depend on θ

A graphical model for the data generation



We lose nothing (in terms of θ estimation) by considering $T(X)$ alone.

Fake data generated from T is good enough



For the purpose of inferring properties of θ, The fake data $\tilde{X}$ is as good as X

Interpretation 2: estimator using $T(X)$ alone cannot be worse

Theorem 3.3 in Keener

Suppose $T(X)$ is sufficient. Then for any estimator $\delta(X)$ of $g(\theta)$ there exists a randomized estimator based on T that has the same risk function as $\delta(X)$.

Interpretation 2: estimator using $T(X)$ alone cannot be worse

Theorem 3.3 in Keener

Suppose $T(X)$ is sufficient. Then for any estimator $\delta(X)$ of $g(\theta)$ there exists a randomized estimator based on T that has the same risk function as $\delta(X)$.

Proof sketch: generate $\tilde{X}$ from T (random step), then $\delta(\tilde{X})$ should be as good as δX.

Factorization theorem

Motivation for factorization theorem

- The sufficiency definition is hard to work with in practice
- There is a convenient way to verify sufficiency by factorizing the density

Theorem 3.6 in Keener

Let $\mathscr{P} = \{P_\theta, \theta \in \Omega\}$ be a family of distributions dominated by μ ($P_\theta \ll \mu, \forall\theta$). T is sufficient for $\mathscr{P}$ iff there exists functions g_θ, h such that

$$p_\theta(x) = g_\theta(T(x))h(x), \text{ a.e.}$$

proof (see the rigorous proof in Kenner 6.4):

Ex1: exponential family

$T(X)$ is sufficient statistics by factorization theorem

$$p_\theta(x) = e^{\eta(\theta)^\top T(x) - B(\theta)} h(x)$$

Ex2: joint distributuon of i.i.d. uniform

Suppose $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} U[\theta, \theta + 1]$. The joint density is

$$p_\theta(x) = \prod_{i=1}^{n} \mathbf{1}_{\{\theta \le x_i \le \theta+1\}}$$
$$= \mathbf{1}_{\left\{\theta \le x_{(1)}, \dots x_{(n)} \le \theta+1\right\}}.$$

The order statistics $(X_{(i)})_{i=1}^n$ (where $X_{(k)}$ is the k-th smallest) is sufficient

Ex3: joint distribution invariant to permutations of X_i

Suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P_\theta^{(1)}$. The joint distributuon P_θ is invariant to permuations of $X = (X_1, \ldots, X_n)$. What are some sufficient statistics?

Ex3: joint distribution invariant to permutations of X_i

Suppose $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} P_\theta^{(1)}$. The joint distributuon P_θ is invariant to permuations of $X = (X_1, \dots, X_n)$. What are some sufficient statistics?

- Order statistics
- Empirical distribution

$$\hat{P}_n(\cdot) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(\cdot)$$

where $\delta_{X_i}(A) = \mathbf{1}_{X_i \in A}$.

In the above case, $p_\theta(X = x \mid T = t)$ is a combinatorial problem that does not depend on θ

Minimal sufficiency

Motivation for minimal sufficiency

Consider $X_1, \dots, X_n \overset{\text{i.i.d}}{\sim} \mathcal{N}(\theta, 1)$. Among the four sufficient statistics

- $\sum_{i=1}^n X_i$
- $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- $O(X) = \left(X_{(1)}, \dots, X_{(n)}\right)$
- $X = (X_1, \dots, X_n)$

which can be recovered from which?

Proposition

If $T(X)$ is sufficient and there exists f such that $T(X) = f(S(X))$, then $S(X)$ is sufficient

proof: factorization theorem

Minimal sufficiency

We say T is minimal sufficient if

- T is sufficient
- For any other sufficient statistics S, there exists f such that

$$T = f(S)$$

(a.e. $\mathscr{P}$)

Example: which is minimal sufficient?

Suppose $X_1, \dots, X_{2n}$ are i.i.d. from $\mathcal{N}(\theta, 1)$, which of the following statistics is sufficient? is minimal?

- X_1
- $\tilde{T} = \begin{pmatrix} \sum_{i=1}^{n} X_i \\ \sum_{i=n+1}^{2n} X_i \end{pmatrix}$
- $\sum_{i=1}^{2n} X_i$

Another way to check minimal sufficiency

Theorem 3.11 in Keener

Suppose $\mathscr{P} = \{P_\theta : \theta \in \Omega\}$ is a dominated family with densities $p_\theta(x) = g_\theta(T(x))h(x)$. If $p_\theta(x) \propto_\theta p_\theta(y)$ implies $T(x) = T(y)$, then T is minimal sufficient.

Interpretation: T is sufficient, if there is one-to-one relation between the statistics and the likelihood shape proof:

Ex1: minimal sufficient statistics in exponential family

$$p_\theta(x) = e^{\eta(\theta)^\top T(x) - B(\theta)} h(x)$$

Is $T(X)$ minimal sufficient?

Ex2: two parameter Gaussian

Suppose $X \sim \mathcal{N}(\mu(\theta), \mathbb{I}_2), \theta \in \mathbb{R}$. $\mu(\theta) = a + \theta b,$ for $a, b \in \mathbb{R}^2$.
Which is sufficient, which is minimal sufficient?

- X
- $b^\top X$

Ex3: Laplace location family

Suppose $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} p_\theta^{(1)}(x) = \frac{1}{2}e^{-|x-\theta|}$. Then the joint density is

$$p_\theta(x) = \frac{1}{2^n} \exp\left\{-\sum_{i=1}^{n} |x_i - \theta|\right\}$$

Is the order statistics sufficient?

Summary

- **Sufficient statistic T:** when T is known, no information about θ is left
- **Factorization theorem:** is the most convenient way to check sufficiency
- It is possible to order the sufficient statstics and define the minimal sufficient statistics.

What is next?

- Completeness
- Ancillarity
- Basu's theorem (relationship between sufficient and ancillary statistics)

Thank you for attending

See you on Monday over Zoom!