PSTAT 115 - Section Two

Winter 2023

Sufficient Statistics

- Step Up: We have a sequence of independent and identically distributed random variables from some distribution $p_\theta(y)$. We collect a *random sample* $Y = (Y_1, \ldots, Y_n)$ of size n.

Note. We represent $y = (y_1, \ldots, y_n)$ to represent a single sample of observations (lower case letters)!

- Goal: to draw inference (from the sample) on the parameter θ (comes from distribution $p(y)$).

Definition (Likelihood). The likelihood $L(\theta) = p(y_1, \ldots, y_n|\theta)$ represents the joint density of the data $\boldsymbol{y}$ for a given θ. **General Equ:** $\prod_{i=1}^{n} P_\theta(y_i)$ → Equ 1

Definition (Sufficient Statistic).

i) A statistic T is a $T := T(Y_1, \ldots, Y_n)$ function of the sample values.

Special Case: a statistic T is called a **sufficient statistic** if conditional dist. of $(Y_1, \ldots, Y_n)|T$ doesn't depend on T!

Theorem (factorization Thm). *A statistic T is sufficient iff on writes,*

$$p_\theta(\boldsymbol{y}) = g_\theta(T(\boldsymbol{y})) \times h(y).$$

Question 1. *Let $X_1, X_2, \ldots, X_n$ be i.i.d $N(\mu, \sigma^2)$ random variables. Find the sufficient (minimal statistic) T.*

Notes

1) Complete $L(\mu,\sigma)$. Here $\theta = (\mu, \sigma)$?

$$L(\theta) \overset{\text{Eq. 1}}{=} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2}$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left\{\frac{-1}{2\sigma^2}\left(\sum_{i=1}^{n}x_i^2 - 2\mu\sum_{i=1}^{n}x_i + n\mu^2\right)\right\}$$

$h(x_1, \ldots, x_n) = 1$

$T(x_1, \ldots, x_n) = \left(\sum_{i=1}^{n} x_i^2, \sum_{i=1}^{n} x_i\right)$ → minimal? Yes. can't simplify further. (lowest dim).