Note on Complete and Sufficient Statistics
March 26, 2024

- Recall that a distribution with a random variable Y with a parameter θ is said to belong to an *exponential family* if $f_Y(y)$ or $P(Y = y)$ can be written as

$$h(y)c(\theta)\exp(w(\theta)t(y))$$

- **Theorem:** If we have $Y_1, \ldots, Y_n$ iid random variables and Y belongs to an exponential family (with a single parameter θ), then, *under some technical conditions*, we say that

$$U = \sum_{i=1}^{n} t(Y_i)$$

  is a **complete and sufficient statistic** for θ.

- Or, if a random variable Y has a distribution with *multiple* parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, then recall that Y belongs to an exponential family if $f_Y(y)$ or $P(Y = y)$ can be written as

$$h(y)c(\boldsymbol{\theta})\exp\left(\sum_{j=1}^{k} w_j(\boldsymbol{\theta})t_j(y)\right)$$

  In this case, if $Y_1, \ldots, Y_n$ iid random variables and Y belongs to an exponential family, we say that (again, under some technical conditions)

$$(U_1, \ldots, U_k) = \left(\sum_{i=1}^{n} t_1(Y_i), \ldots, \sum_{i=1}^{n} t_k(Y_i)\right)$$

  are complete and sufficient statistics for $\boldsymbol{\theta}$.

- **UMVUE:** For a single parameter θ case, we say that $\hat{\theta}$ is the *uniformly minimum variance unbiased estimator* (UMVUE) of θ, if $\hat{\theta}$ is an unbiased estimator of θ that is also a function of complete and sufficient statistic U.

- This can be extended to the multiparameter case, where we say that $\hat{\boldsymbol{\theta}}$ is the UMVUE of $\boldsymbol{\theta}$, if $\hat{\boldsymbol{\theta}}$ are unbiased estimators of $\boldsymbol{\theta}$ that is also a function of complete and sufficient statistics $(U_1, \ldots, U_k)$.

- Some remarks and examples:
  - There is a technical definition of a complete statistic, but we will not cover this here.
  - Note that once the existence of a complete and sufficient statistic is established, we do NOT need to re-confirm that it is sufficient statistic by other means (e.g., no need to check the factorization).

– For example, recall that $Y \sim$ Exponential(θ) belongs to an exponential family, since

$$f_Y(y) = \frac{1}{\theta}e^{-y/\theta} = \frac{1}{\theta}\exp\left(-\frac{1}{\theta}y\right) = h(y)c(\theta)\exp(w(\theta)t(y)), \quad y > 0$$

where

$$h(y) = 1, \quad c(\theta) = \frac{1}{\theta}, \quad w(\theta) = -\frac{1}{\theta}, \quad t(y) = y.$$

Then if $Y_1, \ldots, Y_n \sim$ iid Exponential(θ), we have that

$$U = \sum_{i=1}^{n} t(Y_i) = \sum_{i=1}^{n} Y_i$$

is a complete and sufficient statistic for θ

Now, consider

$$\hat{\theta} = \overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{U}{n}$$

Since $\hat{\theta}$ is unbiased estimator of θ, for $E(\hat{\theta}) = E(\overline{Y}) = E(Y) = \theta$, and observing that $\hat{\theta}$ is a function of a complete and sufficient statistic U, we conclude that $\hat{\theta}$ is indeed a UMVUE of θ.

– As another example, recall that for $Y \sim N(\mu, \sigma^2)$, letting $\boldsymbol{\theta} = (\mu, \sigma^2)$, we have

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\mu^2}{2\sigma^2}\right)\exp\left(-\frac{y^2}{2\sigma^2} + \frac{\mu y}{\sigma^2}\right) \\ &= h(y)c(\boldsymbol{\theta})\exp(w_1(\boldsymbol{\theta})t_1(y) + w_2(\boldsymbol{\theta})t_2(y)), \quad -\infty < y < \infty \end{aligned}$$

where

$$h(y) = 1,\ c(\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\mu^2}{2\sigma^2}\right),\ w_1(\boldsymbol{\theta}) = \frac{-1}{2\sigma^2},\ t_1(y) = y^2,\ w_2(\boldsymbol{\theta}) = \frac{\mu}{\sigma^2},\ t_2(y) = y$$

So if $Y_1, \ldots, Y_n \sim$ iid $N(\mu, \sigma^2)$, it can be seen that

$$(U_1, U_2) = \left(\sum_{i=1}^{n} t_1(Y_i),\ \sum_{i=1}^{n} t_2(Y_i)\right) = \left(\sum_{i=1}^{n} Y_i^2,\ \sum_{i=1}^{n} Y_i\right)$$

are complete and sufficient statistics for (σ^2, μ).

Now, since

$$\hat{\boldsymbol{\theta}} = (\overline{Y}, S^2)$$

are unbiased estimators of $\boldsymbol{\theta} = (\mu, \sigma^2)$, and since $\overline{Y}$ and S^2 are functions of complete and sufficient statistics $\sum_{i=1}^{n} Y_i$ and $\sum_{i=1}^{n} Y_i$ respectively, we conclude that $\hat{\boldsymbol{\theta}} = (\overline{Y}, S^2)$ are UMVUEs of $\boldsymbol{\theta} = (\mu, \sigma^2)$.

- **Basu's Theorem:** Here we state an important result in statistics that makes use of complete and sufficient statistics. First, we start with a definition of a statistic needed for the theorem.

  - We say that W is an *ancillary statistic* for θ if the distribution of W does NOT depend on θ.

  Theorem: (Basu) If U is a complete and sufficient statistic for θ and W is an ancillary statistic for θ, then U and W are independent.

  - For example, for $Y_1, \ldots, Y_n \sim$ iid $N(\mu, \sigma^2)$, consider $\theta = \mu$. It was already established that the sample mean $\overline{Y}$ is a complete and sufficient statistic for μ. Now, the distribution of sample variance S^2 does not depend on μ, so S^2 is an ancillary statistic for μ. Therefore, $\overline{Y}$ and S^2 are independent by Basu's Theorem (giving us yet another proof of the independence between $\overline{Y}$ and S^2).