AMA563

Chapter 4

Sufficient Statistics and Exponential Family Class

Contents

4.1 Sufficient Statistics

In our work on parameter estimation, you have undoubtedly noticed the same estimators coming up again and again in different contexts. In particular, statistics such as the sum of sample values, the sum of square of sample values, and functions of them such as the sample mean and sample variance appear as maximum likelihood estimators, and also as unbiased, minimum variance estimators. You may have been wondering whether there is some underlying cause for this, and whether in general maximum likelihood statistics capture all that is needed in the raw data for estimation purposes.

Here is a brief roadmap through the important results of this section. A *sufficient statistic* is an estimator such that if its value is known, then the details of the sample can shed no more light on the true value of the parameter. By factoring the likelihood function strategically, all at once the form of the sufficient statistic can be discovered, and its sufficiency can be proved. For distributions belonging to a certain class called the *exponential class* (to which many of the distributions we have seen do belong), the sufficient statistics is easy to read off of the sample values. Probably most important for our purposes is that if minimum variance unbiased estimators exist, then they must be functions of sufficient statistics. Furthermore, under mild regularity conditions, maximum likelihood estimators are functions of sufficient statistics, and hence if an unbiased transformation of an MLE can be found, the resulting estimator will be a best estimator.

Definition 4.1.1. *Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with the pdf (or pmf) $f(x|\theta)$. A statistic $T(X_1, X_2, ..., X_n)$ is a sufficient statistic for θ if the conditional distribution of $X_1, X_2, ..., X_n$, given the value of $T(X_1, X_2, ..., X_n)$, dose not depend on θ.*

The fact that for a sufficient statistic, $f(x_1, x_2, ..., x_n|z; \theta)$ does not depend on θ says that the conditional distribution of any estimator $Y = h(X_1, X_2, ..., X_n)$ given $Z = z$ has nothing to do with θ; hence Y can provide no more statistical information about θ. Roughly speaking, a sufficient statistics carries in it all of the useful information for estimating a parameter.

Let's start with the following two trivial examples.

Example 4.1.2. *A trivial example is that the sample itself is a sufficient statistic. This is because*

$$f_{X_1,...,X_n|X_1,...,X_n}(x_1, \ldots, x_n|x_1, \ldots, x_n, \theta) = 1,$$

which says that the conditional distribution of $X_1, \ldots, X_n$, given $X_1 = x_1, \ldots, X_n = x_n$, is a degenerated distribution: $X_1 = x_1, \ldots, X_n = x_n$ with probability 1, and no probability for other values. Of course this distribution does not depend on θ.

Example 4.1.3. *Let X_1, X_2 be a random sample from $N(\mu, 1)$. Then X_1 is not a* *sufficient statistic for μ. This is because*

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2),$$

$$f_{X_2|X_1} = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)} = f_{X_2}(x_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_2 - \mu)^2}{2}\right),$$

so $f_{X_2|X_1}$ still depends on μ. Therefore the conditional distribution of (X_1, X_2), given X_1, still depends on θ.

How do we verify sufficiency?

To show that $Z = g(X_1, \ldots, X_n)$, obtained from a random sample $X_1, \ldots, X_n$, is a sufficient statistic for θ, we need to show that the ratio

$$\boxed{\frac{f(x_1|\theta) \times \cdots \times f(x_n|\theta)}{f_Z(g(x_1, \ldots, x_n)|\theta)}} \tag{4.1}$$

does not depend on θ for all $(x_1, \ldots, x_n)$.

Example 4.1.4. *Let $X_1, \ldots, X_n$ be a random sample from* $\mathsf{Bernoulli}(p)$*. Show that $Z = \sum_{i=1}^n X_i$ is a sufficient statistic for p.*

Solution. Let $z = x_1 + \cdots + x_n$ be the observed value of Z. Let f be the pmf of the Bernoulli distribution. Then

$$f(x|p) = p^x (1-p)^{1-x}, \quad x = 0, 1, \quad \text{zero elsewhere.}$$

So

$$\prod_{i=1}^n f(x_i|p) = p^z (1-p)^{n-z}.$$

We have that $Z \sim \mathsf{Binomial}(n, p)$, so $f_Z(z|p) = \binom{n}{z} p^z (1-p)^{n-z}$. We use the Formula (4.1) to give

$$\frac{f(x_1|p) \times \cdots \times f(x_n|p)}{f_Z(g(x_1, \ldots, x_n)|p)} = \frac{p^z (1-p)^{n-z}}{f_Z(z|p)} = \frac{p^z (1-p)^{n-z}}{\binom{n}{z} p^z (1-p)^{n-z}} = \frac{1}{\binom{n}{z}},$$

which does not depend on p. So Z is a sufficient statistic for p. □

Example 4.1.5. *Let $X_1, \ldots, X_n$ be a random sample from* $\mathsf{Gamma}(2, \theta)$ *with $\theta > 0$. Show that $Y = \sum_{i=1}^n X_i$ is a sufficient statistic for θ.*

Solution. Recall that $Y \sim \mathsf{Gamma}(2n, \theta)$. Recall that the pdf of Gamma distribution is

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad x > 0.$$

Therefore the ratio in Formula (4.1) is

$$\begin{aligned}\frac{\prod_{i=1}^n f(x_i|2,\theta)}{f(\sum_{i=1}^n x_i|2n,\theta)} &= \frac{\prod_{i=1}^n (\Gamma(2)\theta^2)^{-1} x_i e^{-x_i/\theta}}{(\Gamma(2n)\theta^{2n})^{-1} \left[\sum_{i=1}^n x_i\right]^{2n-1} \exp\left(-\sum_{i=1}^n x_i/\theta\right)} \\ &= \frac{\Gamma(2n)\theta^{2n} \prod_{i=1}^n x_i \exp\left[-\sum_{i=1}^n x_i/\theta\right]}{\theta^{2n} \left(\sum_{i=1}^n x_i\right)^{2n-1} \exp\left[-\sum_{i=1}^n x_i/\theta\right]} = \frac{\Gamma(2n)\left(\prod_{i=1}^n x_i\right)}{\left(\sum_{i=1}^n x_i\right)^{2n-1}},\end{aligned}$$

which is independent of θ. Therefore T is the sufficient statistic for θ.

Example 4.1.6 (for the tutorial)**.** *Show that the sum $Z = \sum_{i=1}^n X_i$ of random sample values from the $N(\mu, \sigma^2)$ distribution is sufficient for μ, where for simplicity we assume that σ^2 is known.*

Theorem 4.1.7 (Factorization Theorem)**.** *Let $X_1, X_2, ..., X_n$ be a random sample from a distribution characterized by density (or pmf) $f(x|\theta)$. A statistic $T(X_1, X_2, ..., X_n)$ is a sufficient statistic for θ if and only if there exist functions $g(t, \theta)$ and $h(x_1, \ldots, x_n)$, such that for all sample points $x_1, \ldots, x_n$ and all parameter points θ, the likelihood function L of the sample factors into the product*

$$L(\theta; x_1, x_2, ..., x_n) = g(T(x_1, \ldots, x_n), \theta) \times h(x_1, \ldots, x_n).$$

For the sake of simplicity and the focus of application in the subject, we skip the proof of the Theorem 4.1.7. □

Example 4.1.8 (Example 4.1.5 revisited)**.** *Let $X_1, \ldots, X_n$ be a random sample from* $\mathsf{Gamma}(2, \theta)$ *with $\theta > 0$. Show that $Y = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ.*

Solution. Recall that $Y \sim \mathsf{Gamma}(2n, \theta)$. Recall that the pdf of Gamma distribution is $f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$, $x > 0$. We give the decomposition.

$$\begin{aligned}
L(\theta) &= \left(\frac{1}{\Gamma(2)\theta^2}\right)^n \left(\prod_{i=1}^{n} x_i\right) \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} x_i\right) \\
&= \frac{\Gamma(2n)\theta^{2n} \prod_{i=1}^{n} x_i}{\theta^{2n} \left(\sum_{i=1}^{n} x_i\right)^{2n-1}} \times \frac{1}{\Gamma(2n)\theta^{2n}} \left(\sum_{i=1}^{n} x_i\right)^{2n-1} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} x_i\right) \\
&= \left[\frac{\Gamma(2n) \prod_{i=1}^{n} x_i}{\left(\sum_{i=1}^{n} x_i\right)^{2n-1}}\right] \times \left[\frac{1}{\Gamma(2n)\theta^{2n}} \left(\sum_{i=1}^{n} x_i\right)^{2n-1} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} x_i\right)\right].
\end{aligned}$$

We see that in the last product of the above equation, the factor in the first rectangle bracket is free of the parameter θ, and the factor in the second rectangle bracket is exactly the density function of Z. Therefore the Theorem 4.1.7 implies that Z is the sufficient statistic of θ. □

Example 4.1.9 (for the tutorial). *Find a sufficient statistic for the parameter p in the geometric distribution*

$$f(x|p) = p(1-p)^{x-1}, \quad x = 1, 2, 3, \ldots .$$

Example 4.1.10 (for the tutorial). *Show that the product of the sample observations is a sufficient statistic for $\theta > 0$ if the random sample is taken from a gamma distribution with parameters $\alpha = \theta$ and $\beta = 6$.*

Example 4.1.11 (for the tutorial). *What is the sufficient statistic for θ if the sample arises from a beta distribution in which $\alpha = \beta = \theta > 0$?*

Example 4.1.12 (for the tutorial). *Let $X_1, \ldots, X_n$ be iid with pdf*

$$f(x|\theta) = \begin{cases} \theta e^{-\theta x}, & 0 < x < \infty, \theta > 0, \\ 0, & \textit{elsewhere}. \end{cases}$$

(a) Find a sufficient statistic for θ.

(b) Find the MLE of θ.

(b) Determine the unbiased estimator based on the sufficient statistic in part (a).

Example 4.1.13 (for the tutorial). *Let* $f(x, y) = (2/\theta^2)e^{-(x+y)/\theta}, 0 < x < y < \infty$, *zero elsewhere, be the joint pdf of the random variables* X *and* Y.

(a) Show that the mean and the variance of Y *are, respectively,* $3\theta/2$ *and* $5\theta^2/4$.

(b) Show that $E(Y|X) = X + \theta$ *and* $\mathrm{Var}(X + \theta) = \theta^2/4 < \mathrm{Var}(Y)$.

Example 4.1.14 (for the tutorial). *Let* X_1, X_2, X_3 *be a random sample of size 3 from an exponential distribution with mean* $\theta > 0$. *By the Factorization Theorem, we know that* $Z = X_1+X_2+X_3$ *is a sufficient statistic for* θ. *Since* $E(Z) = 3\theta$, *then* $\phi(Z) = Z/3$ *is an unbiased estimator of* θ. *Let* $W = X_3$. *Compute* $E[\phi(Z)|W]$.

4.2 Exponential Families

Definition 4.2.1. *A family of pdf's or pmf's $f(x|\theta)$ is called an exponential family if its support set $S := \{x : f(x|\theta) > 0\}$ is independent of θ , and if it can be written in the form*[1]

$$f\,(x|\theta) = e^{K(x)p(\theta)+S(x)+q(\theta)}, \quad x \in S,$$

for some functions $K, p, S,$ and q.

Example 4.2.2 (Binomial distributions form an exponential family)**.** *Recall that the pmf of* $\mathsf{Binomial}(n, p)$ *distributions for a fixed n,*

$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, \ldots, n, \quad 0 < p < 1.$$

We have

$$f(x|p) = \exp\left(\boxed{x} \times \boxed{\log\left(\frac{p}{1-p}\right)} + \boxed{n\log(1-p)} + \boxed{\log\binom{n}{x}}\right),$$

so $f(x|p)$ forms an exponential family.

Example 4.2.3. *The family $f(x|\theta) = e^{-x^2/(2\theta)}/\sqrt{2\pi\theta}$, $\theta > 0$, is an exponential family. Indeed,*

$$f(x|\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-x^2/2\theta} = \exp\left(-\frac{1}{2\theta}x^2 - \log\sqrt{2\pi\theta}\right).$$

Example 4.2.4. *The family $f(x|\theta) = \frac{1}{\theta}$, $0 \le x \le \theta$, $\theta > 0$, is NOT an exponential family, because the support set*

$$S = \{x : f(x|\theta) > 0\} = [0, \theta],$$

depends on the parameter θ.

[1] A more general definition could be found in G. Casella and R.L. Berger, *Statistical Inference.* We do not expand the topic here.

Example 4.2.5 (for the tutorial)**.** *Prove that the Poisson distributions* $p(x|\theta) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, \ldots$*, form an exponential family.*

Theorem 4.2.6. *If the distributions* $f(x|\theta)$ *form an exponential family, then the statistic* $Z = \sum_{i=1}^{n} K(X_i)$ *is sufficient for* θ*.*

Proof. The likelihood is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \exp\left(K(x_i)\, p(\theta) + S(x_i) + q(\theta)\right) \\ &= \exp\left(\sum_{i=1}^{n} K(x_i)\, p(\theta) + \sum_{i=1}^{n} S(x_i) + nq(\theta)\right) \\ &= \exp\left(\sum_{i=1}^{n} S(x_i)\right) \times \exp\left(p(\theta) \sum_{i=1}^{n} K(x_i) + nq(\theta)\right) \end{aligned}$$

Let the left factor of the last expression be b, let the right factor be k, and let the sum in the exponent of the right factor be the function g. Then by $L(\theta; x_1, x_2, ..., x_n) = g(\sum_{i=1}^{n} K(x_i), \theta) \times h(x_i, \ldots, x_n)$, $Z = \sum_{i=1}^{n} K(X_i)$ is sufficient. $\square$

Example 4.2.7 (Example 4.1.5 revisited)**.** *Let $X_1, \ldots, X_n$ be a random sample from* $\mathsf{Gamma}(2, \theta)$ *with $\theta > 0$. Use Theorem 4.2.6 to show that $Y = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ.*

Since so many distributions can be put into $f(x|\theta) = e^{K(x)p(\theta)+S(x)+q(\theta)}$, Theorem 2 explains why the sum of simple functions of the sample values X_i arises so frequently as a sufficient statistic. Since, as we establish later, unbiased functions of sufficient statistics have optimal properties, such sums as $\sum K(X_i)$ often come up as minimum variance unbiased estimators as well.

Example 4.2.8. *Consider the estimation of the normal variance σ^2 assuming a known mean μ .The normal density can be written as*

$$\begin{aligned} f\left(x|\sigma^2\right) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \\ &= \exp\left[-\frac{1}{2}\ln\left(2\pi\sigma^2\right) - \frac{(x-\mu)^2}{2\sigma^2}\right] \end{aligned}$$

Let $K(x) = (x-\mu)^2$, $p(\sigma^2) = -\frac{1}{2\sigma^2}$, $S(x) = 0$ and $q(\sigma^2) = -\frac{1}{2}\ln(2\pi\sigma^2)$. Then $\sum_{i=1}^{n} K(X_i) = \sum_{i=1}^{n} (X_i - \mu)^2$ is sufficient for σ^2, by Theorem 4.2.6.

Exercise 4.1

1. Let X_1, X_2 be a random sample of size 2 from an exponential distribution with mean parameter θ (i.e. $\mathsf{Exponential}(\frac{1}{\theta})$).

   (a). Find the maximum likelihood estimator of θ.

   (b). Show that $Y_1 = X_1 + X_2$ is a sufficient statistic for θ.

   (c). Show that $Y_2 = X_2$ is an unbiased estimator of θ.

   (d). Find the joint pdf of Y_1 and Y_2.

   (e). Determine $\phi(Y_1) = E(Y_2|Y_1)$.

2. (7.2.4) Let $X_1, \ldots, X_n$ be a random sample from a geometric distribution that has pmf $f(x|\theta) = (1-\theta)^x\theta$, $x = 0, 1, 2 \ldots$, $0 < \theta < 1$, zero elsewhere. Show that $\sum_{i=1}^n X_i$ is a sufficient statistic for θ.

3. (7.2.6) Let $X_1, \ldots, X_n$ be a random sample from a $\mathsf{Beta}(\theta, 5)$. Show that the product $X_1 \times \cdots \times X_n$ is a sufficient statistic for θ.

4.3 Rao-Blackwell Theorem

Definition 4.3.1. *Let $X_1, \ldots, X_n$ be a random sample. The estimator $Z = g(X_1, \ldots, X_n)$ is called a minimum variance unbiased estimator (MVUE), of the parameter θ, if Z is unbised (that is, $E(Z) = \theta$), and if the variance of Z is no greater than the variance of every other unbiased estimator of θ.*

Theorem 4.3.2. *Let $Z = g(X_1, X_2, ..., X_n)$ be a sufficient statistic for θ based on the random sample $X_1, \ldots, X_n$ from a distribution $f(x|\theta)$. If Y is any unbiased estimator of θ, then $\phi(Z) = E(Y|Z)$, a function of the sufficient statistic, is also unbiased and has variance smaller than or equal to the variance of Y. Hence if a minimum variance unbiased estimator of θ exists, then there must also be a function of the sufficient statistic Z that is a minimum variance unbiased estimator.*

Proof. Since Z is sufficient for θ, by Definition 4.1.1 the distribution of Y given Z does not depend on θ. Also, since Y is unbiased for θ, we have $E[\phi(Z)] = E[E(Y|Z)] = E(Y) = \theta$, so $\phi(Z)$ is also unbiased for θ. Recall[2] that $\mathrm{Var}(Y) \geq \mathrm{Var}(E[Y]|Z) = \mathrm{Var}(\phi(Z))$. Therefore whenever Y is MVUE, so is $\phi(Z)$. □

[2] The details are shown in the Section 4.5, *Supplementary Notes: Conditional Probability.*

Theorem 4.3.2 says that we can restrict our search for MVUE to functions of a sufficient statistic. This theorem however does not mean that for obtaining the best estimator $\phi(Y_1)$, we must start from an unbiased estimator Y_2.

Since the function ϕ may not be invertible, although Z is sufficient, the unbiased estimator $\phi(Z)$ may not be sufficient any longer. For example, we know that the sample $Z = (X_1, \ldots, X_n)$ $(n \geq 2)$ itself is (as always) a sufficient estimator for parameter θ of the family $N(\theta, 1)$. On the other hand, $Y = X_1$ is unbiased as $E(Y) = E(X_1) = \theta$. We have $\phi(Z) = E(Y|Z) = X_1$, which is not at all sufficient.

The following theorem shows the relation between sufficient statistics and the maximum likelihood estimators.

Theorem 4.3.3. *If $Y = g\left(X_1, X_2, ..., X_n\right)$ is a sufficient statistic for θ, and $\widehat{\theta}$ is the unique MLE for θ, then $\widehat{\theta}$ is a function of Y.*

Proof. By Theorem 4.1.7,

$$L(\theta) = w(g(x_1, \ldots, x_n), \theta) \times h(x_1, \ldots, x_n).$$

Here $h(x_1, \ldots, x_n)$ does not depend on θ. So L and w, as functions of θ, achieve their unique maximum simultaneously, and thus their maximizer must be a function of $g(x_1, \ldots, x_n)$. Therefore $\widehat{\theta}$ is a function of $g(X_1, \ldots, X_n)$. $\square$

Example 4.3.4. *For a random sample $X_1, \ldots, X_n$ from $\mathsf{Poisson}(\lambda)$, find the MLE, $\hat{\lambda}$, of the unknown parameter λ. Find the variance of $\hat{\lambda}$. Show that $\hat{\lambda}$ is efficient and also sufficient for λ.*

Solution. The likelihood, and log-likelihood functions are

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}, \\ l(\lambda) &= -n\lambda + \sum_{i=1}^{n} x_i \log(\lambda) - \sum_{i=1}^{n} \log(x_i!). \end{aligned}$$

We have

$$l'(\lambda) = -n + \frac{1}{\lambda}\sum_{i=1}^n x_i, \qquad l''(\lambda) = -\frac{1}{\lambda^2}\sum_{i=1}^n x_i \leq 0.$$

Let $l'(\lambda) = 0$ to give $\lambda = \frac{1}{n}\sum_{i=1}^n x_i$. Therefore the MLE is $\widehat{\lambda}_{\mathsf{MLE}} = \overline{X}$. We have

$$\mathrm{Var}(\widehat{\lambda}_{\mathsf{MLE}}) = \frac{1}{n}\mathrm{Var}(X_1) = \frac{\lambda}{n}.$$

Since

$$\log f(x|\lambda) = -\lambda + x\log\lambda - \log(x!),$$
$$\frac{\partial^2 \log f(x|\lambda)}{\partial\lambda^2} = -\frac{x}{\lambda^2},$$

we have

$$I(\lambda) = -E\left[\frac{\partial^2 \log f(X|\lambda)}{\partial\lambda^2}\right] = \frac{E[X]}{\lambda^2} = \frac{1}{\lambda}.$$

Therefore $\mathrm{Var}(\widehat{\lambda}_{\mathsf{MLE}}) = \frac{1}{nI(\lambda)}$, so $\mathrm{Var}(\widehat{\lambda}_{\mathsf{MLE}})$ is efficient. On the other hand,

$$f(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!} = \exp(K(x)p(\lambda) + S(x) + q(\lambda)),$$

where $K(x) = x$, $p(\lambda) = \log\lambda$, $S(x) = \log(x!)$, and $q(\lambda) = -\lambda$. So $\mathrm{Var}(\widehat{\lambda}_{\mathsf{MLE}}) = \overline{X} = \frac{1}{n}\sum_{i=1}^n K(X_i)$ is sufficient. □

4.4 Completeness and Uniqueness

Definition 4.4.1. *Let $f(z|\theta)$ be a family of pdfs or pmfs for a statistic $Z = g(X_1, \cdots, X_n)$. The family of probability distribution is called complete if*

$$E_\theta(h(Z)) = 0 \textit{ for all } \theta \;\; \textit{implies } P_\theta(h(Z) = 0) = 1 \textit{ for all } \theta.$$

We say, Z is a complete statistic for θ.

Theorem 4.4.2. *Let $f(x|\theta), \gamma < \theta < \delta$, be a pdf or pmf of a random variable X whose distribution is of the exponential class. If $X_1, X_2, \ldots, X_n$ is a random sample from the distribution of X, then $Z = \sum_{i=1}^{n} K(X_i)$ is a complete sufficient statistic for θ.*

Theorem 4.4.3. *Let Z be a complete sufficient statistic for a parameter θ, and let $\phi(Z)$ be an unbiased estimator of θ. Then $\phi(Z)$ is the unique MVUE of θ.*

Example 4.4.4. *Let $X_1, X_2, \ldots, X_n$, $n > 1$, be a random sample from a distribution with pdf $f(x|\theta) = \theta e^{-\theta x}$, $0 < x < \infty$, zero elsewhere, and $\theta > 0$. Define $Y = \sum_{i=1}^{n} X_i$. Prove that $(n-1)/Y$ is the unique MVUE of θ.*

Example 4.4.5. *Let $X_1, X_2, \ldots, X_n$, $n > 2$, be a random sample from a binomial distribution $b(1, \theta)$.*

(a) *Show that $Y_1 = X_1 + X_2 + \cdots + X_n$ is a complete sufficient statistic for θ.*

(b) *Find the function $\phi(Y_1)$ which is the MVUE of θ.*

(c) *Let $Y_2 = (X_1 + X_2)/2$ and compute $E(Y_2)$.*

(d) *Determine $E(Y_2|Y_1 = y_1)$.*

Example 4.4.6. *Write the pdf*

$$f(x|\theta) = \frac{1}{6\theta^4}x^3e^{-x/\theta}, \quad 0 < x < \infty,\ 0 < \theta < \infty,$$

zero elsewhere, in the exponential form. If $X_1, X_2, ..., X_n$ is a random sample from this distribution, find a complete sufficient statistic Y_1 for θ and the unique function $\phi(Y_1)$ of this statistic that is the MVUE of θ. Is $\phi(Y_1)$ itself a complete sufficient statistic?

Example 4.4.7. *Write the pdf*

$$f(x|\theta) = \frac{1}{6}\theta^4x^3e^{-\theta x}, \quad 0 < x < \infty,\ 0 < \theta < \infty,$$

zero elsewhere, in the exponential form. If $X_1, X_2, ..., X_n$ is a random sample from this distribution, find a complete sufficient statistic Y_1 for θ and the unique function $\phi(Y_1)$ of this statistic that is the MVUE of θ. Is $\phi(Y_1)$ itself a complete sufficient statistic?

4.5 Supplementary Notes: Conditional Probability

Recall the definition of conditional expectation and conditional probability. Let X and Y be two random variables, W be a set, and A, B be some events. One has

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{(suppose } P(B) > 0\text{)}.$$

We let χ_A denote the indicator random variable for the event A. So χ_A takes value 1 when A happens and 0 otherwise. The random variable χ_B is similarly defined. Obviously we have $P(\chi_A = 1) = P(A)$, and $P(\chi_A = 0) = 1 - P(A)$.

If X_1 and X_2 are discrete random variable with joint pmf $p_{X_1,X_2}(x_1, x_2)$ and marginal pmf's $p_{X_1}(x_1) = \sum_{x_1} p_{X_1,X_2}(x_1, x_2)$ and $p_{X_2}(x_2) = \sum_{x_2} p_{X_1,X_2}(x_1, x_2)$, we have

$$p_{X_1|X_2}(x_1|x_2) := P(X_1 = x_1|X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)} = \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_2}(x_2)}.$$

Similarly, if X_1 and X_2 are continuous rv's, with joint density function $f_{X_1,X_2}(x_1, x_2)$ and marginal density $f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)dx_2$ and $f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)dx_1$, the conditional density function of X_1 given X_2 is defined as

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}.$$

The conditional probability of $a < X_2 < b$, given $X_1 = x_1$, is

$$P(a < X_2 < b|X_1 = x_1) = \int_a^b f_{X_2|X_1}(x_2|x_1)dx_2.$$

The conditional expectation of X_2, given $X_1 = x_1$ is defined by

$$E[X_2|X_1 = x_1] = \int_{-\infty}^{\infty} x_2 f_{X_2|X_1}(x_2|x_1)dx_2. \tag{4.2}$$

The conditional variance of X_2, given $X_1 = x_1$, is therefore defined by

$$\mathrm{Var}(X_2|X_1 = x_2) = E[X_2^2|X_1 = x_1] - (E[X_2|X_1 = x_1])^2.$$

We observed from Equation (4.2) that $E[X_2|X_1 = x_1]$ is independent of x_2 and is only a function of x_1. We denote the function as $\Omega(x_1) = E[X_2|X_1 = x_1]$. Then define

$$E[X_2|X_1] := \Omega(X_1).$$

We see that $E[X_2|X_1]$ is simply a function of the random variable X_1, and is therefore a random variable. Also

$$\begin{aligned}
E[E[X_2|X_1]] = E[\Omega(X_1)] = \int_{-\infty}^{\infty} \Omega(x_1) f_{X_1}(x_1) dx_1 &= \int_{-\infty}^{\infty} E[X_2|X_1 = x_1] f_{X_1}(x_1) dx_1 \\
&= \int_{-\infty}^{\infty} f_{X_1}(x_1) dx_1 \int_{-\infty}^{\infty} x_2 f_{X_2|X_1}(x_2|x_1) dx_2 \\
&= \int_{-\infty}^{\infty} f_{X_1}(x_1) dx_1 \int_{-\infty}^{\infty} x_2 \frac{f_{X_1,x_2}(x_1, x_2)}{f_{X_1}(x_1)} dx_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 f_{X_1,X_2}(x_1, x_2) dx_1 dx_2 = E[X_2].
\end{aligned}$$

We have shown that

$$\boxed{E[E[X_2|X_1]] = E[X_2].}$$

Here are some important properties about the conditional expectation.

1. For any continuous function[3] f,

   $$E[f(X_1)X_2|X_1] = f(X_1)E[X_2|X_1].$$

   Therefore $E[f(X)|X] = f(X)$. For example, $E[X^2|X] = X^2$, and $E[X|X^3] = X$. However, if X takes both positive and negative values, then X is not a function of X^2, and $E[X|X^2]$ may not be X. A direct consequence is that

   $$E[E[X_2|X_1]|X_1] = E[X_2|X_1],$$

[3] Let X be a random variable and f be any function. The conclusion that $f(X)$ is still a random variable is not trivial. To draw the conclusion, a sufficient condition is that f is continuous. We do not however expand the discussion in this subject.

which is because $E[X_2|X_1]$ is a function of X_1.

2. If X_1 and X_2 are independent, then $E[X_1|X_2] = E[X_1]$. Since a constant C is independent of any random variable X, we have $E[X|C] = E[X]$. On the other hand, $E[C|X] = CE[1|X] = C = E[C]$.

3. The operation of taking conditional expectation would in general reduce the variance: $\boxed{\text{Var}(E[X_2|X_1]) \leq \text{Var}(X_2)}$. In fact, we let $V = X_2 - E[X_2|X_1]$ and $W = E[X_2|X_1] - E[X_2] = E[X_2|X_1] - E[E[X_2|X_1]]$ to give: $E[V|X_1] = E[X_2|X_1] - E[X_2|X_1] = 0$, $E[W^2] = \text{Var}(E[X_2|X_1])$, and that W is a function of X_1. So, $E[VW] = E[E[VW|X_1]] = E[WE[V|X_1]] = E[W \times 0] = 0$. Therefore,

$$\begin{aligned}\text{Var}(X_2) &= E[(X_2 - E[X_2])^2] \\ &= E\{[X_2 - E[X_2|X_1] + E[X_2|X_1] - E[X_2]]^2\} \\ &= E[V^2] + E[W^2] + 2E[VW] = E[V^2] + \text{Var}(E[X_2|X_1]) + 0 \geq \text{Var}(E[X_2|X_1]).\end{aligned}$$

Exercise 4.2

1. (2.3.1) Let X_1 and X_2 have the joint pdf $f(x_1, x_2) = x_1 + x_2$, $0 < x_1 < 1$, $0 < x_2 < 1$, zero elsewhere. Find the conditional mean and variance of X_2, given $X_1 = x_1$, $0 < x_1 < 1$.

2. Let $X \sim N(0, 1)$. Find $E[X|X^2]$, and $E[X| - X]$.

3. (2.3.8) Let X and Y have the joint pdf $f(x, y = 2\exp\{-(x+y)\})$, $0 < x < y < \infty$, zero elsewhere. Find the conditional mean $E[Y|x]$ of Y, given $X = x$.

4. (2.3.10) Let X_1 and X_2 have the joint pmf $p(x_1, x_2)$ described as follows.

(x_1, x_2)	(0,0)	(0,1)	(1,0)	(1,1)	(2,0)	(2,1)
$p(x_1, x_2)$	$\frac{1}{18}$	$\frac{3}{18}$	$\frac{4}{18}$	$\frac{3}{18}$	$\frac{6}{18}$	$\frac{1}{18}$

and $p(x_1, x_2)$ is equal to zero elsewhere. Find the two marginal probability mass functions and the two conditional means.