Lecture 12

Let us give several more examples of finding sufficient statistics.

Example 1. Poisson Distribution $\Pi(\lambda)$ has p.f.

$$f(x|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda} \text{ for } x = 0, 1, 2, \ldots$$

and the joint p.f. is

$$f(x_1, \cdots, x_n|\lambda) = \frac{\lambda^{\sum x_i}}{\prod_{i=1}^n x_i!}e^{-n\lambda} = \frac{1}{\prod_{i=1}^n X_i!}e^{-n\lambda}\lambda^{\sum X_i}.$$

Therefore we can take

$$u(x_1, \ldots, x_n) = \frac{1}{\prod_{i=1}^n X_i!}, \quad T(x_1, \ldots, x_n) = \sum_{i=1}^n x_i \text{ and } v(T, \lambda) = e^{-n\lambda}\lambda^T.$$

Therefore, by Neyman-Fisher factorization criterion $T = \sum_{i=1}^n X_i$ is a sufficient statistics.

Example 2. Consider a family of normal distributions $N(\alpha, \sigma^2)$ and assume that σ^2 is a given known parameter and α is the only unknown parameter of the family. The p.d.f. is given by

$$f(x|\alpha) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\alpha)^2}{2\sigma^2}}$$

and the joint p.d.f. is

$$\begin{aligned}
f(x_1, \ldots, x_n|\alpha) &= \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\Big\{-\sum_{i=1}^n \frac{(x_i-\alpha)^2}{2\sigma^2}\Big\} \\
&= \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\Big\{-\frac{\sum x_i^2}{2\sigma^2} + \frac{\sum x_i\alpha}{\sigma^2} - \frac{n\alpha^2}{2\sigma^2}\Big\} \\
&= \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\Big\{-\frac{\sum x_i^2}{2\sigma^2}\Big\}\exp\Big\{\sum x_i\frac{\alpha}{\sigma^2} - \frac{n\alpha^2}{2\sigma^2}\Big\}.
\end{aligned}$$

If we take $T = \sum_{i=1}^n X_i$,

$$u(x_1, \ldots, x_n) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\Bigl\{-\frac{\sum x_i^2}{2\sigma^2}\Bigr\} \quad \text{and } v(T, \alpha) = \exp\Bigl\{T\frac{\alpha}{\sigma^2} - \frac{n\alpha^2}{2\sigma^2}\Bigr\},$$

then Neyman-Fisher criterion proves that T is a sufficient statistics.

12.1 Jointly sufficient statistics.

Consider

$$\left.\begin{array}{l} T_1 = T_1(X_1, \cdots, X_n) \\ T_2 = T_2(X_1, \cdots, X_n) \\ \cdots \\ T_k = T_k(X_1, \cdots, X_n) \end{array}\right\} \text{ - functions of the sample } (X_1, \cdots, X_n).$$

Very similarly to the case when we have only one function T, a vector $(T_1, \cdots, T_k)$ is called *jointly sufficient statistics* if the distribution of the sample given T's

$$\mathbb{P}_\theta(X_1, \ldots, X_n | T_1, \ldots, T_k)$$

does not depend on θ. The Neyman-Fisher factorization criterion says in this case that $(T_1, \ldots, T_k)$ is jointly sufficient if and only if

$$f(x_1, \ldots, x_n|\theta) = u(x_1, \ldots, x_n) v(T_1, \ldots, T_k, \theta).$$

The proof goes without changes.

Example 1. Let us consider a family of normal distributions $N(\alpha, \sigma^2)$, only now both α and σ^2 are unknown. Since the joint p.d.f.

$$f(x_1, \ldots, x_n|\alpha, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\Bigl\{-\frac{\sum x_i^2}{2\sigma^2} + \frac{\sum x_i\alpha}{\sigma^2} - \frac{n\alpha^2}{2\sigma^2}\Bigr\}$$

is a function of

$$T_1 = \sum_{i=1}^n X_i \text{ and } T_2 = \sum_{i=1}^n X_i^2,$$

by Neyman-Fisher criterion (T_1, T_2) is jointly sufficient.

Example 2. Let us consider a uniform distribution $U[a, b]$ on the interval $[a, b]$ where both end points are unknown. The p.d.f. is

$$f(x|a, b) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & \text{otherwise.} \end{cases}$$

The joint p.d.f. is

$$\begin{aligned} f(x_1,\cdots,x_n|a,b) &= \frac{1}{(b-a)^n} I(x_1\in[a,b])\times\ldots\times I(x_n\in[a,b]) \\ &= \frac{1}{(b-a)^n} I(x_{\min}\in[a,b])\times I(x_{\max}\in[a,b]). \end{aligned}$$

The indicator functions make the product equal to 0 if at least one of the points falls out of the range $[a,b]$ or, equivalently, if either the minimum $x_{\min}=\min(x_1,\ldots,x_n)$ or maximum $x_{\max}=\max(x_1,\ldots,x_n)$ falls out of $[a,b]$. Clearly, if we take

$$T_1=\max(X_1,\ldots,X_n) \text{ and } T_2=\min(X_1,\ldots,X_n)$$

then (T_1,T_2) is jointly sufficient by Neyman-Fisher factorization criterion.

Sufficient statistics:

- Gives a way of compressing information about underlying parameter θ.
- Gives a way of improving estimator using sufficient statistic (which takes us to our next topic).

12.2 Improving estimators using sufficient statistics. Rao-Blackwell theorem.

(Textbook, Section 6.9)

Consider $\delta=\delta(X_1,\cdots,X_n)$ - some estimator of unknown parameter θ_0, which corresponds to a true distribution $\mathbb{P}_{\theta_0}$ of the data. Suppose that we have a sufficient statistics $T=T(X_1,\cdots,X_n)$. (T can also be a vector of jointly sufficient statistics.)

One possible natural measure of the quality of the estimator δ is the quantity

$$\mathbb{E}_{\theta_0}(\delta(X_1,\ldots,X_n)-\theta_0)^2$$

which is an average squared deviation of the estimator from the parameter θ_0.

Consider a new estimator of θ_0 given by

$$\delta'(X_1,\ldots,X_n)=\mathbb{E}_{\theta_0}(\delta(X_1,\ldots,X_n)|T(X_1,\ldots,X_n)).$$

Question: why doesn't δ' depend on θ_0 even though formally the right hand side depends on θ_0?

Recall that this conditional expectation is the expectation of $\delta(x_1,\ldots,x_n)$ with respect to conditional distribution

$$\mathbb{P}_{\theta_0}(X_1,\ldots,X_n|T).$$

Since T is sufficient, by definition, this conditional distribution does not depend on the unknown parameter θ_0 and as a result δ' doesn't depend on θ_0. This point is important, since the estimate can not depend on the unknown parameter, we should be able to compute it using only the data.

Another important point is that the conditional distribution and, therefore, the conditional expectation depend only on T, which means that our new estimate δ' actually depends on the data only through T, i.e. $\delta' = \delta'(T)$.

Theorem. (Rao-Blackwell) *We have,*

$$\mathbb{E}_{\theta_0}(\delta' - \theta_0)^2 \le \mathbb{E}_{\theta_0}(\delta - \theta_0)^2$$

Proof. Given random variable X and Y, recall from probability theory that

$$\mathbb{E}X = \mathbb{E}\{\mathbb{E}(X|Y)\}.$$

Clearly, it we can prove that

$$\mathbb{E}_{\theta_0}((\delta' - \theta_0)^2|T) \le \mathbb{E}_{\theta_0}((\delta - \theta_0)^2|T)$$

then averaging both side will prove the Theorem.

First, consider the left hand side. Since $\delta' = \mathbb{E}_{\theta_0}(\delta|T)$,

$$\mathbb{E}_{\theta_0}((\delta' - \theta_0)^2|T) = \mathbb{E}_{\theta_0}((\mathbb{E}_{\theta_0}(\delta|T) - \theta_0)^2|T) = \ldots$$

Since $(\mathbb{E}_{\theta_0}(\delta|T) - \theta_0)^2$ is already a function of T we can remove the conditional expectation given T and continue

$$\ldots = (\mathbb{E}_{\theta_0}(\delta|T) - \theta_0)^2 = (\mathbb{E}_{\theta_0}(\delta|T))^2 - 2\theta_0\mathbb{E}_{\theta_0}(\delta|T) + \theta_0^2.$$

Next, we consider the right hand side. Squaring out we get,

$$\mathbb{E}_{\theta_0}((\delta - \theta_0)^2|T) = (\mathbb{E}_{\theta_0}(\delta^2|T)) - 2\theta_0\mathbb{E}_{\theta_0}(\delta|T) + \theta_0^2.$$

Therefore, to prove that LHS $\le$ RHS, we need to show that

$$(\mathbb{E}_{\theta_0}(\delta|T))^2 \le \mathbb{E}_{\theta_0}(\delta^2|T).$$

But this is the same as

$$0 \le \mathbb{E}_{\theta_0}(\delta^2|T) - (\mathbb{E}_{\theta_0}(\delta|T))^2 = \mathrm{Var}_{\theta_0}(\delta|T)$$

which is obvious since the variance $\mathrm{Var}_{\theta_0}(\delta|T)$ is always positive.

$\square$