Theorem 3.10 - Neyman -Fisher's factorization criterion

Proof:

We will prove the Neyman-Fisher's factorizatioon criterion for the discrete case only. We need to show that:

(1) $L(X, \theta) = g(T(X), \theta)h(X)$ implies that T is sufficient for θ (this is the $\Longrightarrow$ part of the theorem).

(2) T is sufficient for θ then $L(X, \theta) = g(T(X), \theta)h(X)$ holds (this is the $\Longleftarrow$ part of the theorem).

We start with (1):

Assume that $L(X, \theta) = g(T(X), \theta)h(X)$ holds. We are required to check that T is sufficient for θ. Consider

$$\begin{aligned} P(X = x | T = t) &= \frac{P((x = X) \cap (T = t))}{P(T = t)} \\ &= \begin{cases} 0 & \text{if} \qquad T(x) \neq t \\ \dfrac{P(X = x)}{P(T = t)} & \text{if} \qquad T(x) = t \end{cases} \end{aligned}$$

Now in the second case

$$\frac{P(X = x)}{P(T = t)} = \frac{g(t, \theta)h(x)}{\sum\limits_{\tilde{x}: T(\tilde{x}) = t} g(t, \theta)h(\tilde{x})} = \frac{h(x)}{\sum\limits_{\tilde{x}: T(\tilde{x}) = t} h(\tilde{x})},$$

which does **not** depend on θ (i.e. independent of θ). Hence T is sufficient for θ.

Consider (2):

If T is sufficient for θ then

$$P_\theta(X = x) = P_\theta((X = x) \cap \underbrace{(T = T(X))}_{*}) = P(X = x | T = t)(T = t)$$

where $h(x) = P(X = x | T = t)$ and it does not depend on θ by assumption; whereas $P_\theta(T = t) = g(t, \theta)$ involves the data via the value of the statistic only i.e. we have shown the factorization.

$*$ This is a sure event since T is sufficient by assumption

Example 3.20

Let $X_1, X_2, \ldots, X_n$ be *i.i.d.* uniform on the interval $(\theta, \theta+1)$ with $-\infty < \theta < \infty$. Note that the joint density of X is

$$f(X,\theta) = \begin{cases} 1 & \theta < x_i < \theta + 1 \\ 0 & \text{otherwise} \end{cases}, \qquad i = 1, 2, \ldots, n$$

Alternatively we can write the above as

$$f(X,\theta) = \begin{cases} 1 & \max\limits_i x_i - 1 < \theta < \min\limits_i x_i \\ 0 & \text{otherwise} \end{cases}, \qquad i = 1, 2, \ldots, n$$

For the likelihood ratio of two samples points x and y the ratio $\frac{L(x,\theta)}{L(y,\theta)}$ will be positive for the same values of θ if and only if $\min\limits_i x_i = \min\limits_i y_i$ and $\max\limits_i x_i = \max\limits_i y_i$. Hence when these hold, the ratio will be constant and equal to one.

Let $X_{(1)} = \min\limits_i X_i$ and $X_{(n)} = \max\limits_i X_i$. Then $T(X) = (X_{(1)}, X_{(n)})^\top$ is a minimal sufficient statistic. As previously noted, the dimension of a minimal sufficient statistic does not necessarily match the dimension of the parameter.

NOTE: a minimal sufficient statistic is not unique.

Let $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ be the order statistics from the sample. Let $R = X_{(n)} - X_{(1)}$ be the range statistic which does not depend on θ. Since $T = (X_{(1)}, X_{(n)})^\top$ is a minimal sufficient statistic, we say

$$T^* = (X_{(n)} - X_{(1)}, X_{(n)} + X_{(1)})^\top$$

is also a minimal sufficient statistic.

Let $Z_i = X_i - \theta$. The distribution of Z_i is

$$P(Z_i < z) = P(X_i - \theta < z) = P(X_i < z + \theta) = \int_\theta^{z+\theta} q \mathrm{d}x_i = z + \theta - \theta = z$$

Hence $z_i \sim U[0,1]$ *i.i.d.* and $X_{(n)} - \theta = Z_{(n)}$. Observe that Z_i do not depend on θ. Then

$$\begin{aligned} P(X_{(n)} - X_{(1)} < r) &= P((X_{(n)} - \theta) - (X_{(1)} - \theta) < r) \\ &= P(Z_{(n)} - Z_{(1)} < r) \end{aligned}$$

Hence the distribution of $Z_{(n)}$ and $Z_{(1)}$ are the distribution of the largest and the smallest order statistic from uniform $(0,1)$ distribution which does not involve θ.

Therefore, the first component of T^*, that is, $X_{(n)} - X_{(1)}$ of the minimal sufficient statistic turns out to be ancillary statistic.

Summary of the Pitman estimator $\hat{\theta}_p$

The claim is that if you consider equivariant estimator $\hat{\theta}$ that must satisfy:

$$\hat{\theta}(X_1 + c, X_2 + c, \dots, X_n + c) = \hat{\theta}(X_1, X_2, \dots, X_n) + c$$

when dealing with a location parameter estimation then you can find the best equivariant estimator with respect to the mean squared error (that is, in the class of these estimators, there is one particular one (namely $\hat{\theta}_p$) which minimizes $\mathbb{E}\big[(\theta - \hat{\theta})^2\big]$ for all θ!) (i.e. it has uniformly smallest risk with respect to quadratic loss).

What is interesting is that in its construction you utilise the ancillary statistic

$$\tilde{T}_2 = (X_2 - X_1, X_3 - X_1, \dots, X_n - X_1)$$

starting with an arbitrary equivariant estimator $\tilde{\theta}$ you construct $\hat{\theta}_p$ as

$$\hat{\theta}_p = \tilde{\theta} - \mathbb{E}_\theta(\tilde{\theta} \mid \tilde{T}_2).$$

When the loss is quadratic, you will end up with $\hat{\theta}_p$ possessing th above optimality.

It turns out that

$$\hat{\theta}_p = \frac{\displaystyle\int_{-\infty}^{\infty} \theta \prod_{i=1}^{n} f(x_i; \theta) \mathrm{d}\theta}{\displaystyle\int_{-\infty}^{\infty} \prod_{i=1}^{n} f(x_i; \theta) \mathrm{d}\theta}$$

which you can interpret as the Bayes estimator with respect to the quadratic loss and improper prior $\pi(\theta) \equiv 1$ on $\mathbb{R}^1$. Of course, the prior is improper since it does not integrate to one and therefore not a density over $(-\infty, \infty)$ but you can exploit the analogy.

If the location family $f_\theta(x) = f(x - \theta)$ and we are considering the family/class $\mathrm{N}(\theta, 1)$, the $\hat{\theta}_p$ is the familiar $\bar{X}$ - the arithmetic mean:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Indeed (by cancelling the constants):

$$\hat{\theta}_p = \frac{\displaystyle\int_{-\infty}^{\infty} \theta \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (X_i - \theta)^2 \right) \mathrm{d}\theta}{\displaystyle\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (X_i - \theta)^2 \right) \mathrm{d}\theta}$$

We will multiply a factor (that does not depend on θ) on the numerator and denominator.

In particular, we will multiply by $\exp\Big(-\frac{n}{2}\bar{X}^2\Big)$ to complete the square as follows:

$$\hat{\theta}_p = \frac{\displaystyle\int_{-\infty}^{\infty} \theta \exp\Big(-\frac{1}{2}\Big[\sum_{i=1}^{n} X_i^2 - 2\theta \sum_{i=1}^{n} X_i + n\theta^2\Big]\Big) \mathrm{d}\theta}{\displaystyle\int_{-\infty}^{\infty} \exp\Big(-\frac{1}{2}\Big[\sum_{i=1}^{n} X_i^2 - 2\theta \sum_{i=1}^{n} X_i + n\theta^2\Big]\Big) \mathrm{d}\theta}$$

Then by multiplying by the factor $\exp\Big(-\frac{n}{2}\bar{X}^2\Big)$ we get

$$\begin{aligned}\hat{\theta}_p &= \frac{\displaystyle\int_{-\infty}^{\infty} \theta \exp\Big(-\frac{n}{2}\theta^2 + \theta \sum_{i=1}^{n} X_i]\Big) \mathrm{d}\theta}{\displaystyle\int_{-\infty}^{\infty} \exp\Big(-\frac{n}{2}\theta^2 + \theta \sum_{i=1}^{n} X_i\Big) \mathrm{d}\theta} \\ &= \frac{\displaystyle\frac{\sqrt{n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \theta \exp\Big(-\frac{n}{2}(\theta - \bar{X})^2\Big) \mathrm{d}\theta}{\displaystyle\frac{\sqrt{n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\Big(-\frac{n}{2}(\theta - \bar{X})^2\Big) \mathrm{d}\theta}\end{aligned}$$

If we not interpret θ as a random variable with $\theta \sim \mathrm{N}(\theta, \frac{1}{n})$, then the numerator is the expected value of this random variable (which is $\bar{X}$) and the denominator, we have integrated out the density of this random variable (which is one).

Hence this ratio is equal to:

$$\hat{\theta}_p = \frac{\bar{X}}{1} = \bar{X}.$$

Note: PITMAN is a famous Australian statistician.