Theory of Statistics

Week 4

Let $X_1, \ldots, X_n$ be a random sample from some distribution F, which is parameterized by some parameter $\boldsymbol{\theta} \in \Theta$(*parameter space*). Goal is to draw inference about F, which is equivalent to infer about $\boldsymbol{\theta}$, using the random sample.

Let the sample space be $\mathcal{X} \subseteq \mathbb{R}$, and a statistic $T : \mathcal{X} \to \mathbb{R}$. Any statistic $T(\boldsymbol{X})$ defines a form of data reduction, i.e., it induces a partition in a sample space.

Define for a statistic realization t, the set $A_t = \{\boldsymbol{x} \,|\, T(\boldsymbol{x}) = t\}$. Now let $\mathcal{T} = \{t = T(\boldsymbol{x}) \,|\, \boldsymbol{x} \in \mathcal{X}\}$, the range of function T, with domain $\mathcal{X}$. The collection of sets $\{A_t : t \in \mathcal{T}\}$ defines a partition on X, as firstly $\bigcup\{T_t : t \in \mathcal{T}\} = \mathcal{X}$, as every point has a mapping, and secondly if $t \neq t^{'}$, then $A_t \bigcap A_{t'} = \phi$, as a function is one-to-many. The reduction due to the statistic T is equivalent to the partition.

Example. Let X, Y be a random sample from `Uniform`$(0, 1)$. Consider two statistics $T_1 = \max\{X, Y\}$ and $T_2 = \mathbb{I}(X > Y)$.

The underlying partition sets are $A_{1,t} = \{(x, y) : \max\{x, y\} = t\}$ is the collection of all the points on the two line segments $\{x = t, 0 < y \leq t\}$ and $\{y = t, 0 < x \leq t\}$, whereas $A_{2,0} = \{(x, y) : x \leq y\}$ and $A_{2,1} = \{(x, y) : x > y\}$. Thus T_2 induces a higher level of data reduction.

Higher level of data reduction might lead to over-summarizing of the data and loss of important information about the population.

The goal is to employ the highest level of reduction as long as no *important* information is lost. In parametric inference, all information relevant to the parameter $\boldsymbol{\theta}$ is important.

4.1 Sufficient Statistic

Definition 4.1.1 (Sufficient Statistic)**.** Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ be a random sample from the distribution $\{F_\theta : \theta \in \Theta\}$. A statistic $T(\boldsymbol{X})$ is a sufficient statistic for θ, if the conditional distribution of the random sample $\boldsymbol{X}$ given the value of $T(\boldsymbol{X})$ does not depend on θ.

Example. If $X_i \overset{\text{iid}}{\sim}$ `Bernoulli`(θ), then $\overline{X}_n$ is sufficient for θ.

This is because

$$f_{\mathbf{X}|\overline{X}_n = y}(\boldsymbol{z}) = \begin{cases} \frac{1}{\binom{n}{ny}} & \text{when } \sum_i z_i = ny \\ 0 & \text{otherwise} \end{cases}$$

Let $\boldsymbol{X}$ be a random variable with f_θ. Then the conditional distribution of $\boldsymbol{X}$ given for a statistic T, $T(\boldsymbol{X}) = t$ is

$$f_{\mathbf{X}|T(\mathbf{X})=t}(\boldsymbol{x}) = \frac{f_{\theta;\mathbf{X},t}(\boldsymbol{x}, t)}{f_{\theta;T}(t)} = \begin{cases} f_{\theta;\mathbf{X}}(\boldsymbol{x})/f_{\theta;T}(t) & \text{if} \quad x \in A_t \\ 0 & \text{otherwise} \end{cases}$$

$$f_{\theta;T}(t) = \textstyle\sum_{\boldsymbol{x}:T(\boldsymbol{x})=t} f_{\theta;\mathbf{X}}(\boldsymbol{x})$$

By definition of sufficient statistic, $f_{\mathbf{X}|T(\mathbf{X})=t}(\boldsymbol{x})$ is free of θ, and hence is completely known. Thus it is possible to simulate from this distribution. Suppose the random variable $\boldsymbol{Y}|T(\boldsymbol{X}) = t$ is distributed as this conditional distribution. Then the unconditional distribution of $\boldsymbol{Y}$ is same as the unconditional distribution of $\boldsymbol{X}$ regardless of the value of θ, i.e., for any measurable subset $A \subseteq \mathcal{X}$, $P_\theta(\boldsymbol{Y} \in A) = P_\theta(\boldsymbol{Y} \in A)$, regardless of the value of θ

$P_\theta(\boldsymbol{Y} \in A) = \int_A \left\{\int_t f_{\theta;\mathbf{X}|T=t}(\boldsymbol{x}) f_T(t) dt\right\} d\boldsymbol{x} = \int_A f_{\theta;\mathbf{X}(\boldsymbol{x})d\boldsymbol{x}} = P_\theta(\mathbf{X} \in A)$

Thus $Y \overset{D}{=} X$. This implies that, without knowing the distribution of $\boldsymbol{X}$ is able to generate realizations from the distribution of $\boldsymbol{X}$, regardless of the value of θ.

Example. Let $X_1, \ldots, X_n$ be a random sample from $\mathcal{N}(\mu, 1)$ distribution. Then the statistic $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ is sufficient.

To essentially figure out, $f_{\theta;\mathbf{X}}(\boldsymbol{x})/f_{\theta;T}(t)$. Firstly,

$$\begin{aligned} f_{\theta;\mathbf{X}}(\boldsymbol{x}) &= \prod_{i=1}^n f_{\theta,X_i}(x_i) \\ &= \prod_{i=1}^n (2\pi)^{-1/2} \exp\left(-(x_i - \mu)^2/2\right) \end{aligned}$$

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

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

Now, the sample mean has a distribution $\mathcal{N}(\mu, 1/n)$. Hence,

$$\frac{f_{\theta;\boldsymbol{X}}(\boldsymbol{x})}{f_{\theta;T}(t)} = \frac{(2\pi)^{-n/2}\exp\left(-(\sum_{i=1}^{n}(x_i - \overline{x}_n)^2 + n(\overline{x}_n - \mu)^2)/2\right)}{(2\pi/n)^{-1/2}\exp\left(-n(\overline{x}_n - \mu)^2/2\right)}$$
$$= n^{-1/2}(2\pi)^{-(n-1)/2}\exp\left(-\sum_{i=1}^{n}(x_i - \overline{x}_n)^2/2\right)$$

which does not depend on μ and hence it is a sufficient statistic for μ.

Alternatively, consider another solution which uses the orthonormal transformation. To show that $\overline{X}_n$ is a sufficient statistic, consider the transformation $\boldsymbol{W} = A\boldsymbol{X}$ wherein $W_1 = \frac{1}{\sqrt{n}}X_1 + \cdots + \frac{1}{\sqrt{n}}X_n$, and $W_i = a_{i1}X_1 + \cdots + a_{in}X_n$ for $2 \le i \le n$, s.t. $A^T A = I$

$$f_{\theta;\boldsymbol{W}}(\boldsymbol{w}) = f_{\theta;\boldsymbol{X}}(A^T\boldsymbol{w}).|J|$$
$$= (2\pi)^{-n/2}\exp\left(-\left\{n\mathcal{S}_n^2 + n(\overline{x}_n - \mu)^2\right\}/2\right)\cdot 1$$

Why? $n\mathcal{S}_n^2 = \sum_{i=1}^{n}(x_i - \overline{x}_n)^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2 = \sum_{i=1}^{n} w_i^2 - w_1^2 = \sum_{i=2}^{n} w_i^2$ and the jacobian is $|A^T| = 1$.

$$= (2\pi)^{-n/2}\exp\left(-\left\{\sum_{i=2}^{n} w_j^2 + n(w_1/\sqrt{n} - \mu)^2\right\}/2\right)$$

This implies that $W_1 \sim \mathcal{N}(\sqrt{n}\mu, 1)$ and $W_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ for $2 \le i \le n$ and are independent.

It now becomes evident that $\boldsymbol{W}|W_1$ is not dependent on μ,

hence, $\boldsymbol{W}|\overline{X}_n$ is not dependent on μ,

hence, $\boldsymbol{X}|\overline{X}_n$ is not dependent on μ, as A is not related to μ.

Example. Let $X_1, \ldots, X_n$ be a random sample from $\texttt{Gamma}(2,\theta)$ distribution. Then the statistic $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ is sufficient.

Recall, $T = \sum_{i=1}^{n} X_i$ follows $\texttt{Gamma}(\sum_i \alpha_i, \beta) = \texttt{Gamma}(2n, \theta)$.

Note. *$\boldsymbol{X}$ is trivially a sufficient statistic for all distributions.*

$\boldsymbol{T} = [X_{(1)}, \ldots, X_{(n)}]$ is also a sufficient statistic for all distributions (recall the joint distribution of the order statistics).

4.2 Factorization Theorem

Theorem 4.2.1. *Let $X_1, \ldots, X_n$ be a random sample from a discrete or absolutely continuous distribution with pmf/pdf as $f_{\boldsymbol{X}}(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}$. The statistic T is sufficient for $\boldsymbol{\theta}$ **if and only if**, there exists functions $g(t;\boldsymbol{\theta})$ and $h(\boldsymbol{x})$ such that for all sample points $\boldsymbol{x}$ and for all $\boldsymbol{\theta}$*

$$f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) = g(T(\boldsymbol{x});\boldsymbol{\theta}) \cdot h(\boldsymbol{x})$$

Example. $X_i \overset{\text{iid}}{\sim} \texttt{Uniform}(0,\theta)$. To show that n-th order statistic is sufficient.

$$f_{\boldsymbol{X};\theta}(\boldsymbol{z}) = \begin{cases} \frac{1}{\theta^n} & z_{(1)} > 0, z_{(n)} < \theta \\ 0 & \text{otherwise} \end{cases}$$
$$= \frac{1}{\theta^n}\mathbb{I}(z_{(1)} > 0)\mathbb{I}(z_{(n)} < \theta)$$
$$= h(\boldsymbol{z})g_\theta(z_{(n)})$$

Factorization theorem. [$\Leftarrow$]

$$f_{\boldsymbol{X}|T=t}(\boldsymbol{z}) = \frac{f_{X,T}(\boldsymbol{z},t)}{f_T(t)}$$
$$= \frac{g_\theta(T(\boldsymbol{z}))h(\boldsymbol{z})}{\sum_{\boldsymbol{z}\in A_t} f_{\boldsymbol{X}}(\boldsymbol{z})}$$
$$= \frac{g_\theta(T(\boldsymbol{z}))h(\boldsymbol{z})}{\sum_{\boldsymbol{z}'\in A_t} g_\theta(T(\boldsymbol{z}'))h(\boldsymbol{z}')}$$
$$= \frac{h(\boldsymbol{z})}{\sum_{\boldsymbol{z}'\in A_t} h(\boldsymbol{z}')}$$

The last holds because of the definition of A_t. This implies T is sufficient.

[$\Rightarrow$]

Let for some $\boldsymbol{z} \in \mathcal{X}, T(\boldsymbol{z}) = t \in \mathcal{T}$

$$f_{\boldsymbol{X}|T=t}(\boldsymbol{z}) = \underbrace{\frac{f_{\boldsymbol{X}}(\boldsymbol{z})}{f_T(t)}}_{\text{free of }\theta} = h(\boldsymbol{z})$$

$$f_{\boldsymbol{X}}(\boldsymbol{z}) = \underbrace{f_T(t)}_{g_\theta}\, h(\boldsymbol{z})$$

□

Remark. Let $X_i \overset{\text{iid}}{\sim} F_\theta$ and T be a sufficient statistic for θ. T is a function of another statistic U. This implies U is a sufficient statistic.

Proof.

$$\begin{aligned} f_{\mathbf{X}|U=u_0} &= f_{\mathbf{X},U}(\mathbf{z}, u_0)/f_U(u_0) \\ &= \begin{cases} 0 & U(\mathbf{z}) \neq u_0 \\ \frac{f_{\mathbf{X}}(\mathbf{z})}{f_U(u_0)} & U(\mathbf{z}) = u_0 \end{cases} \\ &= \frac{g_\theta(T(\mathbf{z}))h(\mathbf{z})}{\sum_{\mathbf{z}' \in A_{u_0}} f_{\mathbf{X}}(\mathbf{z}')} \\ &= \frac{g_\theta(T(\mathbf{z}))h(\mathbf{z})}{\sum_{\mathbf{z}' \in A_{u_0}} g_\theta(T(\mathbf{z}'))h(\mathbf{z}')} \end{aligned}$$

This becomes free of θ, because notice that $U(\mathbf{z}) = u_0$ and the definition of A_{u_0}. □

Note. *If $\mathbf{x} \in \mathcal{X}, U(\mathbf{x}) = u_0$ then $T(\mathbf{x}) = h'(u)$ for some $h' : \boldsymbol{\mathcal{U}} \to \mathcal{T}$, where the domain and range are respectively the range of U and T.*

Remark. Converse might not be true. If U is a sufficient statistic, then T might not be.

Remark. Let $X_1, \ldots, X_n$ be a random sample from distribution $F_{\boldsymbol{\theta}}$ and $T(\mathbf{X})$ be a sufficient statistic of $\boldsymbol{\theta}$. If $U(\mathbf{X})$ is a bijective function of $T(\mathbf{X})$. Then $U(\mathbf{X})$ is also sufficient for $\boldsymbol{\theta}$.

4.2.1 Exponential Family

A family of pmfs of pdfs is called a d-parameter exponential family if it can be expressed as

$$f_X(x; \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right\}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, $h(x) \geq 0$ and for all x and $t_1, \ldots, t_k$ are real valued functions for x, not depending upon θ. Further $c(\boldsymbol{\theta}), w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta})$ are real valued functions of $\boldsymbol{\theta}$ not depending on x.

Example. $X \sim \texttt{Beta}(\alpha, \beta)$

$$\begin{aligned} f_X(x) &= \frac{1}{\texttt{Beta}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\ &= \underbrace{\frac{1}{\texttt{Beta}(\alpha,\beta)}}_{c(\boldsymbol{\theta})} \exp\left\{\underbrace{(\alpha-1)}_{w_1}\underbrace{\log x}_{t_1} + \underbrace{(\beta-1)}_{w_2}\underbrace{\log(1-x)}_{t_2}\right\} \end{aligned}$$

Other examples are: $\texttt{Binomial}(n,p)$ with known n, $\texttt{Poisson}(\lambda), \texttt{Normal}(\mu,\sigma^2), \texttt{Exponential}(\lambda), \texttt{Gamma}(\alpha,\beta)$.

Remark. **Sufficient statistics for exponential family of distributions.**

Let $X_1 \ldots, X_n$ be a random sample from a distribution with pmf or pdf $f_X(\cdot; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$ which belongs to an exponential family with $d \leq k$. Then

$$\boldsymbol{T}(\mathbf{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i)\right)$$

is jointly sufficient for $\boldsymbol{\theta}$.

Example. As shown above for the `Beta` distribution, we get that $(\sum_i \log X_i, \sum_i \log(1-X_i))$ is jointly sufficient for α, β.

Week 5

5.1 Minimal Sufficient Statistic

For any family of distributions $\{F_\theta; \theta \in \Theta\}$, a sufficient statistic for θ captures the maximum information about θ, however often fails to have maximum level of data reduction. One would like to find the sufficient statistic which is most precise, among the pool of sufficient statistics. This leads to the concept of *minimal sufficiency.*

Definition 5.1.1 (Minimal Sufficient Statistic)**.** A minimal statistic $T(\boldsymbol{X})$ is called minimal sufficient if for every other sufficient statistic $U(\boldsymbol{X})$, $T(\boldsymbol{X})$ is a function of $U(\boldsymbol{X})$.

5.1.1 Understanding Minimal sufficiency

Suppose $T(\boldsymbol{X})$ and $U(\boldsymbol{X})$ are two sufficient statistic for a class of distributions $F_\theta; \theta \in \Theta$. Let $\mathcal{T} = \{t = T(\mathbf{z}) : \mathbf{z} \in \mathcal{X}\}$, and similarly define $\mathcal{U}$ as the ranges of U and T. Further consider the sets $A_t = \{\mathbf{z} : T(\mathbf{z}) = t\}$ and $B_u = \{\mathbf{z} : U(\mathbf{z}) = u\}$. A_t and B_u are partitions of $\mathcal{X}$.

If T is a function of U, then $\exists h : \mathcal{U} \to \mathcal{T}$, s.t. for any $\mathbf{z} \in \mathcal{X} \quad T(\mathbf{z}) = t = h(u) = h(U(\mathbf{z}))$, where $u = U(\mathbf{z})$. $C_t = \{u : h(u) = t\} \subseteq \mathcal{U}$ as the preimage of t.

Consider $\mathbf{z} \in B_{u_0}$, then $U(\mathbf{z}) = u_0$ and $T(\mathbf{z}) = h(U(\mathbf{z})) = h(u_0) = t_0$, which implies $\mathbf{z} \in A_{t_0}$, where $h(u_0) = t_0$. Hence $B_{u_0} \subseteq A_{t_0}$ when $h(u_0) = t_0$. Now suppose there exists another $u^{'}$ such that $h(u^{'}) = t_0$, then similarly $B_{u'} \subseteq A_{t_0}$. Hence $B_u \subseteq A_{t_0}$ for all $u \in C_{t_0}$.

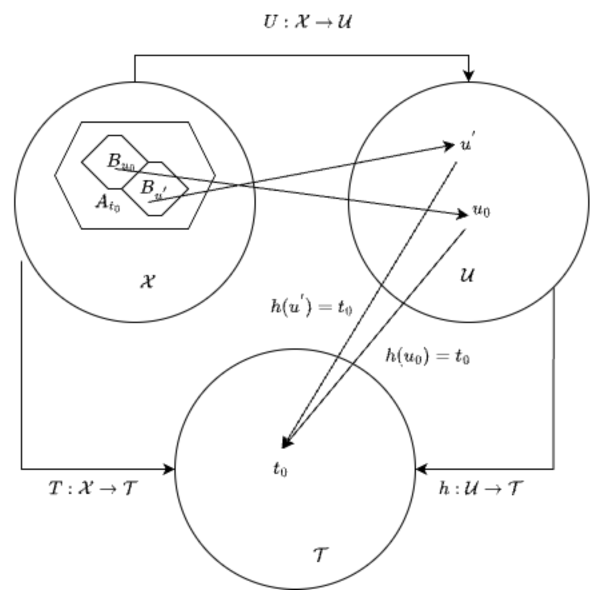
A minimal sufficient statistic T, being functions of all other sufficient statistics, provides the highest level of data reduction among the class of sufficient statistics.

How to know if a sufficient statistic is minimal sufficient?

Theorem 5.1.1. *Let $f(\boldsymbol{x}; \theta)$ be the pmf or pdf of a sample $\boldsymbol{X}$. Suppose there exists a function $T(\boldsymbol{x})$ such that for every two points x and y, the ratio $f(\boldsymbol{x}; \theta)/f(\boldsymbol{y}; \theta)$ is constant as a function of θ iff $T(\boldsymbol{x}) = T(\boldsymbol{y})$. Then $T(\boldsymbol{X})$ is a minimal sufficient statistic for θ.*

Proof. [Sufficiency of T]

Consider the conditional distribution of $\boldsymbol{X}$ given $T = t$



$$f_{\mathbf{X}|T=t}(\mathbf{z}) = \begin{cases} \frac{f_{\mathbf{X}}(\mathbf{z})}{f_T(t)} & \text{if } T(\mathbf{z}) = t \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{f_{\mathbf{X}}(\mathbf{z})}{\sum_{\mathbf{z}' \in A_t} f_{\mathbf{X}}(\mathbf{z}')} & T(\mathbf{z}) = t \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{}{\sum_{\mathbf{z}' \in A_t} \{f_{\mathbf{X}}(\mathbf{z}')/f_{\mathbf{X}}(\mathbf{z})\}} & T(\mathbf{z}) = t \\ 0 & \text{otherwise} \end{cases}$$

Now since $\mathbf{z}^{'} \in A_t$ hence $T(\mathbf{z}^{'}) = t$. Hence $f_{\mathbf{X}|T=t}(\mathbf{z})$ is free of θ hence T is sufficient.

[Minimality of T]

To show: If $\exists$ another sufficient statistic U then T is a function of U.

Claim. Let $U : \mathcal{X} \to \mathbb{R}$ and $T : \mathcal{X} \to \mathbb{R}$ be two statistics. Let $\mathcal{U} = \{U(\boldsymbol{x}) | x \in \mathcal{X}\}$ and $\mathcal{T} = \{T(\boldsymbol{x}) | x \in \mathcal{X}\}$. T is a function of U if and only if for any $x \in \mathcal{X}$ s.t. $U(\boldsymbol{x}) = u_0$ and $T(\boldsymbol{x}) = t_0$, $A_{u_0} \subseteq A_{t_0}$.

The forward direction is already established (*Why*?)

For the other direction, define $h : \mathcal{U} \to \mathcal{T}$ s.t.

$h(u_0) = t_0$. For each $u \in \mathcal{U}$ obtain A_u then for some $\boldsymbol{x} \in A_u$ obtain $T(\boldsymbol{x})$. By the claim $A_u \subseteq A_t$ hence $h(u) = t$. Note that for each $u, \exists h(u)$ and $\nexists$ any u s.t. $h(u)$ is not unique. Hence a proper function is defined. □

5.2 Ancillary Statistic

Definition 5.2.1. Let $X_1, \ldots, X_n$ be a random sample from some distribution F_θ, then a statistic $S(\boldsymbol{X})$ is ancillary for θ if the distribution of S does not depend on θ.

Example. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 1)$. Then $T = S_n^2$ is ancillary for μ.

Example. Let $X_1, \ldots, X_n$ be a random sample from $\texttt{Uniform}(\theta, \theta + 1)$, then $S(\boldsymbol{X}) = X_{(n)} - X_{(1)}$ is ancillary for θ.

Consider the transformation $\left[X_{(n)}, X_{(1)}\right] \rightarrow \left[X_{(n)} - X_{(1)}, X_{(1)}\right] = [U, V]$. $X_{(1)} = V$ and $X_{(n)} = U + V$

Since $f_{X_{(n)}, X_{(1)}}(t, s) = n(n-1)(t-s)^{n-2}$ hence $f_{U,V}(u, v) = n(n-1)u^{n-2}$.

Further $V \in (\theta, \theta + 1 - U)$. Hence $f_U(u) = n(n-1)u^{n-2}(1-u)$ which is free of θ.

5.2.1 Location family of Ancillary statistic

Let $X_1, \ldots, X_n$ be a random sample from a distribution F_θ s.t. $X_i = W_i + \theta$ and $W_1, \ldots, W_n$ are IID from the distribution F which is free of θ, then X is said to be from the location family of distribution.

For example. $X \sim \mathcal{N}(\mu, 1)$ is a member of location family.

Further if $S(\boldsymbol{X})$ be a statistic s.t. $S(\boldsymbol{X} + d\mathbf{1}) = S(\boldsymbol{X})\, \forall \boldsymbol{X} \in \mathbb{R}^n, \forall d$, then S is ancillary for θ.

Proof. $P(S(\boldsymbol{x}) \le s) = P(S(\boldsymbol{x} - \theta\mathbf{1}) \le s) = P(S(\boldsymbol{w}) \le s)$ which is free of θ, hence done. □

5.2.2 Scale family of Ancillary statistic

Let $X_1, \ldots, X_n$ be a random sample from a distribution F_θ s.t. $X_i = W_i \cdot \theta$ and $W_1, \ldots, W_n$ are IID from the distribution F which is free of θ, then X is said to be from the scale family of distribution.

For example. $X \sim \mathcal{N}(0, \sigma^2)$ is a member of scale family, and so is $X_i \sum \texttt{Uniform}(0, \theta)$.

Further if $S(\boldsymbol{X})$ be a statistic s.t. $S(c\boldsymbol{X}) = S(\boldsymbol{X})\, \forall \boldsymbol{X} \in \mathbb{R}^n, \forall c$, then S is ancillary for θ.

For example, the statistic is $S(\boldsymbol{X}) = X_{(1)}/X_{(n)}$ is one such statistic. Now if X_i follow a scale family of distribution, then S becomes ancillary.

Proof. $P(S(\boldsymbol{x}) \le s) = P(S(\frac{1}{\theta}\boldsymbol{x}) \le s) = P(S(\boldsymbol{w}) \le s)$ which is free of θ, hence done. □

5.2.3 Location-Scale family of Ancillary statistic

Let $X_1, \ldots, X_n$ be a random sample from a distribution F_θ s.t. $X_i = \theta_1 + W_i \cdot \theta_2$ and $W_1, \ldots, W_n$ are IID from the distribution F which is free of θ, then X is said to be from the scale family of distribution.

Further if $S(\boldsymbol{X})$ be a statistic s.t. $S(c\boldsymbol{X} + d\mathbf{1}) = S(\boldsymbol{X})\, \forall \boldsymbol{X} \in \mathbb{R}^n, \forall c, d$, then S is ancillary for θ.

Note. *If $S(\boldsymbol{X})$ is a sufficient statistic for θ and $T(\boldsymbol{X})$ is another statistic s.t. S and T are independently distributed, then $T(\boldsymbol{X})$ is ancillary for θ.*

This is because $f_T(t) = f_{T|S=s}(t) = \sum_{x \in A_t} f_{\boldsymbol{X}|S=s}(\boldsymbol{x})$ which is free of θ, since sufficiency. Hence T is ancillary.

Remark. Converse is generally not true, i.e., if $S(\boldsymbol{X})$ is minimal sufficient for θ and $T(\boldsymbol{X})$ is ancillary for θ then it is not true that $T(\boldsymbol{X})$ and $S(\boldsymbol{X})$ are independent.

Example. Consider $\texttt{Uniform}(\theta - 1/2, \theta + 1/2)$ which is of the location family. $T(\boldsymbol{X}) = \left[X_{(n)} - X_{(1)}\right]$ is ancillary for θ, but $S(\boldsymbol{X}) = \begin{pmatrix} X_{(1)} + X_{(n)} \\ X_{(n)} - X_{(1)} \end{pmatrix}$ is minimal sufficient, and they are clearly not independent.

Definition 5.2.2 (Completeness)**.** A family of distribution $f_T(.;\theta), \theta \in \Theta$ is called **complete** if $E_\theta\left[g(T)\right] = 0$ for all $\theta \in \Theta$ implies $P_\theta(g(T) = 0)$ for all $\theta \in \Theta$. If the family of a statistic $T(\boldsymbol{X})$ is complete, $T(\boldsymbol{X})$ is called a complete statistic.

Example. $X_1, X_2 \sim \mathcal{N}(\mu, 1)$. And let $\boldsymbol{T}(\boldsymbol{X}) = [X_1, X_2]^T$ is not complete. $E_\theta[g(T)] = 0$ does not generally implies $P(g(T) = 0) = 1$.

Let $g(\boldsymbol{T}) = X_1 - X_2$, $E_\mu[g(\boldsymbol{T})] = 0$, however $P(g(\boldsymbol{T}) = 0) = P(X_1 = X_2) = 0$, and hence it is not complete.

Example. $\texttt{Poisson}(\theta)$ is complete.
Let $T \sim \texttt{Poisson}(\theta)$.

$$
\begin{aligned}
E_\theta[g(T)] &= \sum_{t=0}^{\infty} g(T) e^{-\theta} \frac{\theta^t}{t!} = 0 \quad \forall \theta \in \Theta \\
&= g(0) + \theta g(1) + \theta^2 \frac{g(2)}{2!} + \cdots = 0 \quad \forall \theta \in \Theta \\
&\Rightarrow g(t) = 0 \forall t \in \mathbb{N} \cup \{0\}
\end{aligned}
$$

$\{g(T) = 0\} = \{t \in S : g(t) = 0\}$ are all positive integers and zero, hence $P_\theta(g(T) = 0) = 1$.

Example. $X_1, \ldots, X_n \sim \texttt{Uniform}(0, \theta)$ and $T(\boldsymbol{X}) = X_{(n)}$. This gives $f_T(t) = \frac{nt^{n-1}}{\theta^n}, \quad 0 < t < \theta$. To show that $T(\boldsymbol{X})$ is sufficient for θ.

$$
\begin{aligned}
E_\theta[g(t)] = \frac{n}{\theta^n}\int_0^\theta g(t)t^{n-1}.dt = 0, \quad \theta > 0 \\
\Rightarrow \int_0^\theta g(t)t^{n-1}.dt = 0 \\
\Rightarrow \frac{\partial}{\partial\theta}\int_0^\theta g(t)t^{n-1}.dt = 0 \\
\Rightarrow g(\theta)\underbrace{\theta^{n-1}}_{\neq 0} = 0 \\
\Rightarrow g(\theta) = \theta\forall\theta > 0
\end{aligned}
$$

Hence $P(g(T) = 0) = 1$ for all $t > 0$.

Include the examples of minimal sufficiency for $\texttt{Uniform}(0,\theta)$, and $\mathcal{N}(\mu,\sigma^2)$