1 Introduction

Statistics is concerned with the collection of data and their analysis and its interpretation. After collecting the data we are interested to look what the data tells us. Statistical inference means drawing conclusions based on data. More technically, inference may be defined as the selection of a probabilistic model to resemble the process we wish to investigate, investigation of that model’s behavior and interpretation of the results. Two major approaches in inference is classical inference and Bayesian inference where both of them have their own significance.

Classical Inference: The observations are postulated to be the values taken by random variables which are assumed to follow a joint probability distributions, $f(x|\theta)$. The distributions we assumed are indexed by an unknown value $\theta$. Then the first step of the analysis is to specify a plausible value for $\theta$. This is the problem of point estimation. In other words, one context for inference is the parametric model, in which data are supposed to come from a certain distribution family, the members of which are distinguished by differing parameter values.

As pointed out, once the data is available we take data $X = (X_1, ..., X_n)$ comes from a probability distribution $f(x|\theta)$, with $\theta$ unknown. First task is to estimate the value of $\theta$.

Definition 1 A point estimator of $\theta$ is any function $W(X) = W(X_1, ..., X_n)$ of a sample.
Definition 2  Given a sample of realized observations, the number $W(x_1, \ldots, x_n)$ is called a point estimate of $\theta$.

It is noted that an estimator is the function of the sample, while an estimate is the realized value of an estimator. Some important questions in this context are, how to construct an estimator $\theta$ using the random samples $X_1, \ldots, X_n$, how to measure or evaluate the closeness of $\theta$, how to find the best possible $\theta$. To answer the last two questions we have to define some properties of the estimators. Then construct an estimator by some method so that it posses some of the properties of the estimators. First we shall discuss the properties of the estimators.

1.1 Sufficient Statistics

The starting point of a statistical analysis, as described in the preceding section, is a random observable $X$ taking on values in a sample space, and a family of possible distributions of $X$. It often turns out that some part of the data carries no information about the unknown distribution and that $X$ can therefore be replaced by some statistic $T = T(X)$ (not necessarily real-valued) without loss of information. Hence it is good to look at statistic (sufficient statistics) which capture all the information about the parameter. The concept of sufficiency originated from Fisher (1920) and later it blossomed further, again in the hands of Fisher (1922). Now we introduce the notion of sufficiency which helps in summarizing data without any loss of information.

Definition 3  A statistic $T(X)$ is a function of the sample $X_1, \ldots, X_n$. Examples of statistics are sample mean $\bar{X}$, sample variance $S^2$, the largest order statistic $X_{(n)}$, the smallest order statistic $X_{(1)}$. Note that any statistic $T(X)$ is point estimator.

Before introducing sufficiency, we give a small description about the partition of a sample space by a statistic $T(X)$. For any possible value $t$ of $T$, consider a corresponding set

$$A_t = \{x : T(x) = t\}.$$
The collection of all set \( \{ A_t, \text{ all } t \} \) makes a partition on the sample space of \( X \). Clearly

\[
P(T(X) = t) = \sum_{x \in A_t} P(X = x).
\]

**Example 1** Toss a coin 3 times, and let \( X_1, X_2 \) and \( X_3 \) be respectively the outcome of each toss. Let \( T \) be the total number of heads obtained, clearly, \( T = \sum_{i=1}^{3} X_i \). The partition of the sample space given by \( T \) follows. Clearly \( t = 0, 1, 2, 3 \), hence

\[
A_0 = \{ x : T(x) = 0 \} = \{ TTT \}
\]

And the set \( \{ A_t \}, t = 1, 2, 3 \) can be obtain in similar way and it can be easily verified that \( \{ A_t \} \) form a partition of the sample space given by

\[
\Omega = \{ TTT, TTH, HTT, THT, THH, HTH, HHT, HHH \}.
\]

Sample \( X_1, ..., X_n \) contains information, where some are relevant for \( \theta \) and some are not. Dropping irrelevant information is desirable, but dropping relevant information is undesirable. Hence we are looking for a statistic \( T(X) \) which contain all the information in the sample about \( \theta \). Normally, the dimension of \( T(X) \) is smaller than the sample size \( n \).

If \( T(x) \) has a simpler data structure and distribution than the original sample \( X_1, ..., X_n \), so it would be nice if we can use \( T(X) \) to summarize and then replace the entire data. As mentioned, while finding a \( T(x) \) which has dimension smaller than \( n \), the following questions should be carefully addressed. Is there any loss of information due to summarization? How to compare the amount of information about \( \theta \) in the original data \( X \) and in \( T(X) \)? Is it sufficient to consider only the ‘reduced data’ \( T \)?

**Definition 4** A statistic \( T \) is called sufficient if the conditional distribution of \( X \) given \( T \) is independent of \( \theta \). That is the conditional distribution of \( X \) given \( T \) is completely known (does not involve \( \theta \)).

**Definition 5** Sufficiency Principle: Any inference procedure should depend on the data only
What sufficiency tells us? Given the value $t$ of a sufficient statistic $T$, conditionally there is no more information left in the original data regarding the unknown parameter $\theta$. Another way of interpretation, our sample $X$ trying to tell us a story about $\theta$, but once a sufficient summary $T$ becomes available, the original story then becomes superfluous.

Look at this concepts in another way, suppose that an investigator reports the value of $T$, but on being asked for the full data, admits that they have been discarded. In an effort at reconstruction, one can use a random mechanism (such as a pseudo-random number generator) to obtain a random quantity $X'$, which is distributed according to the conditional distribution of $X$ given $t$. Then the unconditional distribution of $X'$ is the same as that of $X$. Hence, from a knowledge of $T$ alone, it is possible to construct a quantity $X'$ which is completely equivalent to the original $X$. Since $X$ and $X'$ have the same distribution for all $\theta$, they provide exactly the same information about $\theta$. The following example will illustrates this concepts clear.

**Example 2** Suppose that $X_1$ and $X_2$ are two independent Bernoulli random variables with parameter $p$, $0 < p < 1$. It is given that $x_1 = 1$ and $x_2 = 1$. From the observed sample we can conclude (intuitionally) that the value of the parameter $p$ is large. For this sample the value of $t = x_1 + x_2 = 2$, it also tells us that the value of the parameter $p$ is large. Hence we are not loosing any information relevant to $p$. If we take $t = x_1 - x_2 = 0$, then $t$ tells us that the value of $p$ is either too large or very small, so that $t$ only gives a partial information about $p$. The statistic $T$ is not giving full information in the sample regarding $p$.

**Definition 6** A vector valued statistic $T = (T_1, ..., T_k)$ where $T_i = T_i(X_1, ..., X_n), i = 1, ..., k$, is called jointly sufficient $\theta$ if and only if the conditional distribution of $X$ given $T = t$ does not involve $\theta$, for all $t$.

The sufficiency can be checked using the Definition 4. We illustrated it through some simple examples.
Example 3  Suppose that $X_1,\ldots, X_n$ are iid Bernoulli(p), where $p$ is the unknown parameter, $0 < p < 1$. We will prove that the $T = \sum_{k=1}^n X_k$ is sufficient for $p$. Consider

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | T = t) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n, T = t) / P(T = t)$$

$$= \prod_{k=1}^n P(X_k = x_k) / P(T = t).$$

After some simplification it reduces to $1/\binom{n}{t}$ which is free from $p$, hence $\sum_{k=1}^n X_k$ is sufficient for $p$.

Example 4  Suppose that $X_1, X_2, X_3$ are iid Bernoulli(p), where $p$ is the unknown parameter, $0 < p < 1$. From the above example we have seen that $\sum_{k=1}^3 X_k$ is sufficient for $p$. Is, $T = X_1 X_2 + X_3$, sufficient for $p$. Since the event $\{T = 0\}$ is same as

$$\{X_1 = 0\} \cap \{X_2 = 0\} \cap \{X_3 = 0\} \cup \{X_1 = 1\} \cap \{X_2 = 0\} \cap \{X_3 = 0\} \cup \{X_1 = 0\} \cap \{X_2 = 1\} \cap \{X_3 = 0\},$$
we can write

$$P(T = 0) = P(X_1 = 0, X_2 = 0, X_3 = 0) + P(X_1 = 1, X_2 = 0, X_3 = 0) + P(X_1 = 0, X_2 = 1, X_3 = 0)$$

$$= (1 - p)^3 + 2p(1 - p)^2 = (1 - p)^2(1 + p).$$

Consider

$$P(X_1 = 0, X_2 = 0, X_3 = 0 | T = 0) = P(X_1 = 0, X_2 = 0, X_3 = 0) / P(T = 0)$$

$$= (1 - p)^3 / (1 - p)^2(1 + p) = (1 - p) / (1 + p),$$

which depends on $p$ so that $T$ is not sufficient for $p$.

Example 5  Suppose that $X_1,\ldots, X_n$ are iid $P(\lambda)$ where $\lambda$ is the unknown parameter. The similar way as above we can prove that $T = \sum_{k=1}^n X_k$ is sufficient for $\lambda$. 
Example 6 Suppose that $X_1, \ldots, X_n$ are iid $U(0, \theta)$ where $\theta > 0$. Then the $n^{th}$ order statistic $X_{(n)}$ is sufficient for $\theta$. It can be easily seen that, given the value $T = x_{(n)}$, the distribution of $X_i \quad i = 1, 2, \ldots, n - 1$ is uniformly distributed on the interval $(0, x_{(n)})$. Hence the conditional distribution of $X_i | T = x_{(n)}$ is independent of $\theta$ and $X_{(n)}$ is sufficient for $\theta$.

How to find a sufficient statistic? The determination of sufficient statistics by means of the Definition 4 is inconvenient since it requires, first, guessing a statistic $T$ that might be sufficient and, then, checking whether the conditional distributions of $X$ given $t$ is independent of $\theta$.

The following theorem will help us to find a sufficient statistic.

Theorem 1 (Factorization Criterion) A necessary and sufficient condition for a statistic $T$ to be sufficient is that $p(x|\theta)$ can be written as

$$p(x|\theta) = h(x)g(T(x); \theta),$$

where $h(x)$ is independent of $\theta$ and $g(T(x); \theta)$ depends on $x$ only through $T(X)$.

The proof was discussed in the class.

Example 7 Let $X_i \sim Exp(\theta), i = 1, 2, \ldots, n$ where $\theta > 0$. Consider

$$p(x_1, \ldots, x_n|\theta) = \theta^n exp\left(-\theta \sum_{i=1}^{n} x_i\right).$$

Taking $h(x) = 1$ and $g(T(x); \theta) = \theta^n exp(-\theta \sum_{i=1}^{n} x_i)$, using factorization criterion, we can conclude that $T = \sum_{i=1}^{n} x_i$ is sufficient for $\theta$.

Note: When the range of $X$ depends on $\theta$, one should be more careful about factorization.

Must use indicator functions explicitly while writing the pdf of $X$.

Example 8 Suppose that $X_1, \ldots, X_n$ are iid $U(-2/\theta, 2/\theta)$ where $\theta > 0$. Then

$$p(x_1, \ldots, x_n|\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I(x_i > -2/\theta) \prod_{i=1}^{n} I(x_i < 2/\theta).$$
\[
\frac{1}{\theta^n} I(x_{(1)} > -2/\theta) I(x_{(n)} < 2/\theta),
\]
where \(x_{(1)}\) and \(x_{(n)}\) are 1st and \(n\)th order statistic respectively. Using factorization theorem we can conclude that \((x_{(1)}, x_{(n)})\) is sufficient for \(\theta\).

**Exercise 1** Suppose \(X_i \sim \text{Exp}(\theta, 1), \ i = 1, 2, \ldots, n\) where \(\theta > 0\), find a sufficient statistic for \(\theta\).

**Exercise 2** Suppose \(X_i \sim \text{U}(\theta - 1/2, \theta + 1/2), \ i = 1, 2, \ldots, n\) where \(\theta > 0\), find a sufficient statistic for \(\theta\).

**Example 9** Suppose that \(X_1, \ldots, X_n\) are iid from \(N(\theta, \sigma^2)\) (both unknown). Consider

\[
p(x_1, \ldots, x_n | \theta, \sigma^2) = \frac{1}{(2\pi)^{n/2} \sigma^{n/2}} \exp \left( -\frac{1}{2} \sum_{i=1}^n x_i^2 - \theta \sum_{i=1}^n x_i - n\theta \right).
\]

Hence \((\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)\) is jointly sufficient for \((\theta, \sigma^2)\).

**Exercise 3** Suppose that \(X_1, \ldots, X_n\) are iid from \(\text{Gamma}(\alpha, \beta)\). Find sufficient statistic for \((\alpha, \beta)\).

The exponential family is extensively used in classical as well as in Bayesian inference in view of the pleasant properties it possesses in developing optimal estimators and tests. The completeness property of the family is widely exploited in unbiased estimation problem. A probability density function \(f(x | \theta)\) is said to belong to the \(k\)-parameter exponential family, if it can be written in the form

\[
f(x | \theta) = c(\theta)h(x)\exp \left[ \sum_{j=1}^k W_j(x)s_j(\theta) \right],
\]

where \(h(x)\) and \(W_j(x)\) are real valued measurable functions and \(c(\theta)\) and \(s_j(\theta)\) are real functions of \(\theta\) having continuous non-vanishing derivatives.

**Theorem 2** Let \(X_1, \ldots, X_n\) be a random sample from the exponential family. Denote the statistic \(T_j = \sum_{i=1}^n W_j(x_i)\). Then the statistic \(T = (T_1, \ldots, T_k)\) is jointly sufficient for \(\theta\).
Exercise 4 Apply the general exponential family result to all the standard distributions discussed above such as Poisson, normal, exponential, gamma.

Remark: The original data $X_1, ..., X_n$ are always sufficient for $\theta$. (They are trivial statistics, since they do not lead any data reduction). Order statistics $T = (X_{(1)}, ..., X_{(n)})$ are always sufficient for $\theta$. The dimension of order statistics is $n$, the same as the dimension of the data. Still this is a nontrivial reduction as $n!$ different values of data corresponds to one value of $T$.

Remark: If $T$ is sufficient for $\theta$, and $U = f(T)$ with $f$ being one-to-one, then $U$ is also sufficient. When one statistic is a one-one function of the other statistic and vice versa, then they carry exactly the same amount of information.

Remark: If $T$ is sufficient for $\theta$ and $T = f(U)$, a function of some other statistic $U$, then $U$ is also sufficient. Clearly knowledge of $U$ implies knowledge of $T$ and hence permits reconstruction of the original data. Furthermore, $T$ provides a greater reduction of the data than $U$ unless $f$ is 1:1, in which case $T$ and $U$ are equivalent.

Example 10 If $\sum_{i=1}^{n} x_i$ is sufficient for $\theta$ so $\bar{X}$ is sufficient for $\theta$.

Example 11 If $(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2)$ is jointly sufficient for $(\theta, \sigma^2)$ so $(\bar{X}, S^2)$ is jointly sufficient for $(\theta, \sigma^2)$.

1.2 Minimal Sufficiency and Complete

From the above discussion, it is seen that different sufficient statistic are possible. One may ask which one is the best? What is a natural way to define the 'best' sufficient summary statistic? How to get hold of such a 'best' sufficient summary statistic? Lehmann and Scheff (1950) developed a precise mathematical formulation of the concept known as minimal sufficiency and they gave a technique that helps to locate minimal sufficient statistic. In minimal sufficiency we are looking for a statistic with maximum possible reduction of the data.

Definition 7 A sufficient statistic $T$ is called minimal sufficient if $T$ is a function of any other sufficient statistic $U$. That is, if for any sufficient statistic $U$ there exists a function $f$ such that $T = f(U)$
**Remark:** A sufficient statistic $T$ is said to be minimal if of all sufficient statistics it provides the greatest possible reduction of the data. It is a tedious job to check the minimal sufficiency by using the definition directly. For exponential family we have the following result.

**Theorem 3** Let $X_1, ..., X_n$ be a random sample from the exponential family and it is full rank. Denote the statistic $T_j = \sum_{i=1}^{n} W_j(x_i)$. Then the statistic $T = (T_1, ..., T_k)$ is minimal sufficient.

Next we state a theorem due to Lehmann and Scheff (1950) which provides a direct approach to find minimal sufficient statistics for $\theta$.

**Theorem 4** Suppose, for any two sample points $x$ and $y$, $f(x; \theta)/f(y; \theta)$ is a constant function of $\theta$ (i.e. $f(x|\theta)/f(y|\theta)$ does not depend on $\theta$ if and only if $T(x) = T(y)$). Then the statistic $T$ is minimal sufficient for $\theta$.

**Proof:** For $t = T(y)$ defined the sample points $y(t) = (y_1^t, ..., y_n^t)$ as $\{y^{(t)} : t = T(y)\}$. For any sample $x_1, ..., x_n$, let $T(x) = t^*$. Consider

$$f(x|\theta) = \frac{f(x|\theta)}{f(y^{(t^*)}|\theta)} f(y^{(t^*)}|\theta).$$

Since, $f(x|\theta)/f(y^{(t^*)}|\theta)$ does not depend on $\theta$ and by factorization theorem $T$ is sufficient for $\theta$. Again for $U(x) = U(y)$ consider,

$$\frac{f(U(x)|\theta)}{f(U(y)|\theta)} = \frac{g(U(x), \theta)h(x)}{g(U(y), \theta)h(y)},$$

a constant since $U(x) = U(y)$. Hence $T(x) = T(y)$ and $T$ is a function of $U$, $T$ is minimal sufficient for $\theta$.

**Definition 8** A statistic $V(X)$ is said to be ancillary if its distribution does not depend on $\theta$, and first-order ancillary if its expectation $E_\theta[V(X)]$ is constant, independent of $\theta$.

**Example 12** If $X_{(i)} \sim N(0, \sigma^2)$, the sample mean $\bar{X}$ is ancillary for $\sigma^2$. 

9
Exercise 5  Consider independent identical distribution $U(\theta, \theta + 1)$. Show $R = X_{(n)} - X_{(1)}$ is ancillary (Hint: Show that $R \sim \text{Beta}(n - 1, 2)$).

Exercise 6  If $X_1$ and $X_2$ are independent $N(0, \sigma^2)$. Show $X_1/X_2$ is ancillary.

An ancillary statistic by itself contains no information about $\theta$, but minimal sufficient statistics may still contain much ancillary material. In Exercise 2, for instance, the differences $X_{(n)} - X_{(i)}$ ($i = 1, ..., n-1$) are ancillary despite the fact that they are functions of the minimal sufficient statistics $(X_{(1)}, X_{(n)})$.

Intuitively, one may think that ancillary contains no information about $\theta$, so it should not change the information content about $\theta$ if we add or remove of ancillary information. But this interpretation is false. See the following example.

Example 13  In Exercise 5, if $M = (X_{(1)} + X_{(n)})/2$ and $R = X_{(n)} - X_{(1)}$, then $(M, R)$ is minimal. Here $R$ is ancillary, but a part of minimal sufficient statistics.

Definition 9  A statistic $T$ is said to be complete for the family of distributions \{f(x|\theta), \theta \in \Theta\} if for any function $T(x)$, $E(T(X)) = 0$ implies $T(x) = 0$.

Since complete sufficient statistics are particularly effective in reducing the data, it is not surprising that a complete sufficient statistic is always minimal.

Theorem 5  If $T$ is complete and sufficient for the family \{f_\theta, \theta \in \Theta\}, and a minimal sufficient statistic exists, then $T$ is also minimal sufficient.

The proof will be discussed later as it uses Rao-Blackwell Theorem.

Remark: The converse of the Theorem 6 is not true as the following example will illustrate it.

Example 14  Let $X_1, ..., X_n$ be iid according to $U(\theta - 1/2, \theta + 1/2)$, $-\infty < \theta < \infty$. Then, $T = (X_{(1)}, X_{(n)})$ is minimal sufficient for $\theta$. On the other hand, $T$ is not complete since $X_{(n)} - X_{(1)}$ is ancillary. Since, $E[X_{(n)} - X_{(1)} - (n - 1)/(n + 1)] = 0$ for all $\theta$, hence $T$ is not complete.
The following theorem is due to the simple application of the factorization criteria.

**Theorem 6** If $X$ is distributed according to the exponential family and the family is of full rank, then $T = [T_1(X), \ldots, T_k(X)]$ is complete.

**Example 15** Let $X_1, \ldots, X_n$ be iid from Bernoulli(1, $p$). Then $T = \sum_{i=1}^{n} X_i$ is a complete statistic.

Since the distribution of $T$ is binomial with parameter $p$, $Eg(T) = 0 \text{ for all } p \in (0, 1)$ implies

$$\sum_{t=0}^{n} g(t)\left(\begin{array}{c} n \\ t \end{array}\right) p^t(1-p)^{n-t} = 0 \text{ for all } p \in (0, 1),$$

which can be rewritten as

$$(1-p)^n \sum_{t=0}^{n} g(t)\left(\begin{array}{c} n \\ t \end{array}\right) (\frac{p}{1-p})^t = 0 \text{ for all } p \in (0, 1).$$

This is a polynomial in $t$, each term should vanish so that $g(t) = 0$ for all $t = 1, 2, \ldots, n$.

**Example 16** Let $X_1, \ldots, X_n$ be iid from $Exp(\lambda)$ Then, $T = \sum_{i=1}^{n} X_i$ is a complete statistic. Clearly the distribution of $T$ is Gamma($n, \lambda$), $E(g(T)) = 0$ implies that

$$\frac{\lambda^n}{\Gamma(n)} \int_{0}^{\infty} g(t)t^{n-1}e^{-\lambda t} = 0.$$

Using the uniqueness properties of Laplace transform, we have, $g(t)t^{n-1} = 0 \text{ for all } t > 0$, hence $g(t) = 0$ and $T$ is complete.

**Example 17** Let $X_1, \ldots, X_n$ be iid from $U(0, \theta)$, then the family is complete.

To prove that a statistic is not complete we need to verify any of the following

1. If a non-constant function of $T$ is ancillary, then $T(X)$ is not complete.
2. If $E(T)$ does not depend on $\theta$, then $T(X)$ is not complete.

**Remark:** In the search for complete and sufficient statistics, it is enough to check the completeness of a minimal sufficient statistic.
The following fact can be verified easily

(i) If $T$ is complete and $S = f(T)$, then $S$ is also complete.

(ii) If a statistic $T$ is complete and sufficient, then any minimal sufficient statistic is complete.

(iii) Trivial (constant) statistics are complete for any family.

(iv) Non-trivial (non-constant) ancillary statistic cannot be complete.

(v) A statistic is called first order ancillary if its expectation is free of $\theta$. If a nontrivial function of statistic $T$ is first order ancillary, then $T$ cannot be complete.

**Theorem 7 (Basu’s Theorem)** If $T$ is a complete sufficient statistic for the family of distribution \( \{f(x|\theta), \theta \in \Theta\} \), then any ancillary statistic $A$ is independent of $T$.

**Proof:** To prove the theorem we will show that

\[
P(A \leq a|T = t) = P(A \leq a)
\]

Since $A$ ancillary statistic, the distribution of $A$ is independent of $\theta$, that is $P(A \leq a)$ is independent of $\theta$. Consider

\[
g_a(t) = P(A \leq a|T = t).
\]

Taking expectation on both sides, we obtain

\[
E(g_a(T)) = P(A \leq a),
\]

right hand side follows from the identity $E(E(X|Y)) = E(X)$ and the fact that $P(X < a) = E(I(X \leq a))$. Since $P(A \leq a)$ is independent of $\theta$, the above identity can be written as

\[
E(g_a(T) - P(A \leq a)) = 0.
\]

Given $T$ is complete, using the definition, we have $g_a(T) = P(A \leq a)$. That is $P(A \leq a|T = t) = P(A \leq a)$, hence the proof.

As an applications of Basu’s Theorem we can prove the following theorem
Theorem 8 If $X \sim N(\mu, \sigma^2)$, $\sigma^2$ is known, then the sample mean $\bar{X}$ and sample variance $S^2$ are independent.

The proof is simple as $\bar{X}$, is complete for the given family and $S^2$ is ancillary for $\mu$.

Example 18 Let $X_1, \ldots, X_n$ be iid from $Exp(\lambda)$. Show that $T = \sum_{i=1}^n X_i$ and $W_1, \ldots, W_n$ are independent, where $W_i = X_i/T \quad i = 1, 2, \ldots, n$

Some more examples discussed in the class.

Example 19 Let $X_1, \ldots, X_n$ be iid from the probability density function

$$f(x) = \frac{1}{\sigma} \exp\left(-\frac{(x-\lambda)}{\sigma}\right), \quad \infty < \lambda < x < \infty, \quad 0 < \sigma < \infty.$$ 

Find two dimensional sufficient statistics for $(\lambda, \sigma)$.

Example 20 Let $X_1, \ldots, X_n$ be independent random variable with probability density function

$$f(x_i|\theta) = \frac{1}{2i\theta}, \quad i(\theta - 1) < x < i(\theta + 1),$$

Find two dimensional sufficient statistics for $\theta$.

Example 21 Let $X_1, \ldots, X_n$ be iid from the probability density function

$$f(x) = \frac{\exp\left(-\frac{(x-\lambda)}{\sigma}\right)}{\left(1 + \exp\left(-\frac{(x-\lambda)}{\sigma}\right)\right)^2}, \quad -\infty < \lambda, x < \infty.$$ 

Find minimal sufficient statistics for $\lambda$.

Example 22 Let $N$, the sample size be a random variable taking the values $1, 2, \ldots$ with known probabilities $p_1, p_2, \ldots$ with $\sum p_i = 1$. Having observed $N = n$, perform $n$ Bernoulli trials with success probability $\theta$, getting $X$ success. Prove that $(X, N)$ is minimal sufficient and $N$ is ancillary for $\theta$.  

13
Example 23  Let $X_1, \ldots, X_n$ be iid from the probability density function

$$f(x|\lambda) = exp(-(x - \lambda)), \quad 0 < \lambda < x < \infty.$$ 

Find minimal sufficient statistics for $\lambda$ and show that $X_{(1)} = \min\{X_1, X_2, \ldots X_n\}$ and $S^2$ are independent.

Reference

