Chapter 6

Principles of Data Reduction

6.1 Introduction

Goal: To summarize or reduce the data $X_1, X_2, \ldots, X_n$ to get information about an unknown parameter θ.

Note:

1. $\mathbf{X}$ denotes the random variables $X_1, X_2, \ldots, X_n$ and $\mathbf{x}$ denotes the sample point $x_1, x_2, \ldots, x_n$ (i.e., a realization, observation, or observed value of $\mathbf{X}$);
2. A statistic, $T(\mathbf{X})$, defines a form of data reduction or data summary and $T(\mathbf{x})$ is an observed value of the statistic $T(\mathbf{X})$;
3. Data Reduction in terms of a statistic $T(\mathbf{X})$ can be thought of as a partition of the sample space $\mathcal{X}$, the set of possible observed values of $\mathbf{X}$.
4. $\mathcal{T} = \{t : t = T(\mathbf{x}), \text{ for } \mathbf{x} \in \mathcal{X}\}$ is the image of $\mathcal{X}$ under $T(\mathbf{x})$. Then the statistic $T(\mathbf{X})$ partitions the sample space $\mathcal{X}$ into sets A_t, $t \in \mathcal{T}$, defined by $A_t = \{\mathbf{x} : T(\mathbf{x}) = t, \ \mathbf{x} \in \mathcal{X}\}$.

Two Principles of Data Reduction

1. **Sufficiency Principle**: promotes a method of data reduction that does not discard information about parameter θ.
2. **Likelihood Principle**: describes a function of the parameter, determined by the observed sample, that contains all the information about θ that is available from the sample.

6.2 Sufficiency Principle

A ***sufficient statistic*** for a parameter θ is a statistic that, in a certain sense, captures all the information about θ contained in the sample. Any additional information in the sample, besides the value of the sufficient statistic, does not contains any more information about θ.

Sufficiency Principle: If $T(\mathbf{X})$ is a sufficient statistic for θ, then any inference about θ should depend on the sample $\mathbf{X}$ only through the value of $T(\mathbf{X})$. That is, if $\mathbf{x}$ and $\mathbf{y}$ are two sample points such that $T(\mathbf{x}) = T(\mathbf{y})$, then the inference about θ should be the same whether $\mathbf{X} = \mathbf{x}$ or $\mathbf{X} = \mathbf{y}$ is observed.

6.2.1 Sufficiency Statistics

Definition 6.2.1: A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if the conditional distribution of the sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ.

Question: Is there a simpler way to verify a sufficient statistic?

Theorem 6.2.2: If $p(\mathbf{x}|\theta)$ is the joint pdf or pmf $\mathbf{X}$ and $q(t|\theta)$ is the pdf or pmf of $T(\mathbf{X})$, then $T(\mathbf{X})$ is a sufficient statistic for θ if, for every $\mathbf{x}$ in the sample space, the ratio $p(\mathbf{x}|\theta)/q\big(T(\mathbf{x})|\theta\big)$ is a constant function of θ.

Proof.

$$\begin{aligned} P_\theta\big(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = T(\mathbf{x})\big) &= \frac{P_\theta\big(\mathbf{X} = \mathbf{x} \text{ and } T(\mathbf{X}) = T(\mathbf{x})\big)}{P_\theta\big(T(\mathbf{X}) = T(\mathbf{x})\big)} \\ &= \frac{P_\theta(\mathbf{X} = \mathbf{x})}{P_\theta\big(T(\mathbf{X}) = T(\mathbf{x})\big)} = \frac{p(\mathbf{x}|\theta)}{q\big(T(\mathbf{x})|\theta\big)} \end{aligned}$$

■

<u>Example 6.2.3</u>: (Binomial Sufficient Statistic)
Let $X_1, \dots, X_n$ be iid Bernoulli random variables with parameter θ, for $0 < \theta < 1$. Show that $T(\mathbf{X}) = X_1 + \cdots + X_n$ is a sufficient statistic for θ.

$$\frac{p(\mathbf{x}|\theta)}{q\big(T(\mathbf{x})|\theta\big)} = \frac{\prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \xlongequal{\left(t=T(\mathbf{x})=\sum\limits_{i=1}^n x_i\right)} \frac{1}{\binom{n}{t}}$$

<u>Example 6.2.4</u>: (Normal Sufficient Statistic)
Let $X_1, \dots, X_n$ be iid Normal random variables $\mathrm{n}(\mu, \sigma^2)$, where σ^2 is known. Show that $T(\mathbf{X}) = \bar{X}$ is a sufficient statistic for μ.

<u>Example</u>: (Example of a Statistic that is Not Sufficient)
Consider the model of Example 6.2.3 again with $n = 3$. Then $T(\mathbf{X}) = X_1 + X_2 + X_3$ is sufficient while $T(\mathbf{X}) = X_1 + 2X_2 + X_3$ is not sufficient because:

$$\begin{aligned}
&P(X_1 = 1, X_2 = 0, X_3 = 1 | X_1 + 2X_2 + X_3 = 2) \\
=& \frac{P(X_1 = 1, X_2 = 0, X_3 = 1)}{P(X_1 = 1, X_2 = 0, X_3 = 1) + P(X_1 = 0, X_2 = 1, X_3 = 0)} \\
=& \frac{\theta(1-\theta)\theta}{\theta(1-\theta)\theta + (1-\theta)\theta(1-\theta)} = \frac{\theta(1-\theta)\theta}{\theta(1-\theta)} = \theta.
\end{aligned}$$

Example 6.2.5: (Sufficient Order Statistic)

Let $X_1, \dots, X_n$ be iid from a pdf f and no other information about f is available. Then it follows that

$$p(\mathbf{x}) = \prod_{i=1}^{n} f(x_i) = \frac{1}{n!} \prod_{i=1}^{n} f(x_{(i)}),$$

where $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ are the order statistics.

$\Rightarrow$ By Theorem 6.2.2, the order statistics are a sufficient statistic.

$\Rightarrow$ Without additional information about f, we cannot have further reduction.

$\Rightarrow$ If f is Cauchy pdf, $f(x) = \dfrac{1}{\pi(x-\theta)^2}$, or Logistic pdf, $f(x) = \dfrac{e^{(x-\theta)}}{\left(1+e^{-(x-\theta)}\right)^2}$, the most reduction we can get are the order statistics.

Remark: Outside the exponential family of distributions, it is rare to have a sufficient statistic of smaller dimension than the size of the sample and in many cases order statistics is the best we can do.

Example: (Sufficient Statistic for Poisson Family)

Let $X_1, \dots, X_n$ be iid Poisson population with the parameter $\lambda > 0$. Then $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is a sufficient statistic for λ.

Proof. Notice that $T(\mathbf{X})$ has a Poisson distribution with the parameter $n\lambda$. ■

Question: Can we find a sufficient statistic by simple examination of the pdf or pmf?

Theorem 6.2.6: (Factorization Theorem)

Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of a sample $\mathbf{X}$. A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if and only if there exists functions $g(t|\theta)$ and $h(\mathbf{x})$ such that, for all sample points $\mathbf{x}$ and all parameter points θ,

$$f(\mathbf{x}|\theta) = h(\mathbf{x})g\big(T(\mathbf{x})|\theta\big). \tag{6.2.3}$$

Remark: To use the Factorization Theorem to find a sufficient statistic, we factor the joint pdf of the sample into two parts, with one part not depending on θ. The part that does not depend on θ constitutes the $h(x)$ function. The other part, the one that depends on θ, usually depends on the sample $\mathbf{x}$ only through some function $T(\mathbf{X})$ and this function $T(\mathbf{X})$ is the sufficient statistic of θ.

Example 6.2.7: (Continuation of Example 6.2.4)

Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$, where σ^2 is known. Show that $T(\mathbf{X}) = \bar{X}$ is a sufficient statistic for μ using the Factorization Theorem.

Example 6.2.8: (Uniform Sufficient Statistic)
Let $X_1, \ldots, X_n$ be iid from a discrete uniform on $1, \ldots, \theta$. Show that $T(\mathbf{X}) = X_{(n)} = \max_{1 \le i \le n} X_i$ is a sufficient statistic for θ.

Example 6.2.9: (Normal Sufficient Statistic, μ and σ^2 unknown)
Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$. Show that $T(\mathbf{X}) = \big(T_1(\mathbf{X}), T_2(\mathbf{X})\big) = (\bar{X}, S^2)$ is a sufficient statistic for μ and σ^2.

Remark: For a normal model $\mathrm{n}(\mu, \sigma^2)$, $\bar{X}$ and S^2 contain all information about μ and σ^2. However, if the model is not normal, this may not necessarily be true.

Example: (Sufficient Statistic for Poisson Family)
Let $X_1, \ldots, X_n$ be iid Poisson population with parameter $\lambda > 0$. Then use the Factorization Theorem to show that both $T'(\mathbf{X}) = \sum_{i=1}^n X_i$ and $T(\mathbf{X}) = (X_1, \sum_{i=2}^n X_i)$ are sufficient statistics for λ.

Question: Is there an easy way to find a sufficient statistic for an exponential family of distributions?

<u>**Theorem 6.2.10**</u>**:** Let $X_1, \ldots, X_n$ be iid from a pdf or pmf $f(x|\theta)$ that belongs to an exponential family given by

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

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, $d \leq k$. Then

$$T(\mathbf{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \sum_{j=1}^{n} t_2(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j)\right)$$

is a sufficient statistic for θ.

<u>**Example**</u>**: (Sufficient Statistic for Poisson Family)**
Let $X_1, \ldots, X_n$ be iid Poisson population with the parameter λ. Since

$$f(x|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda} = \frac{1}{x!}e^{-\lambda} \exp\big(x \log(\lambda)\big),$$

we have

$$h(x) = \frac{1}{x!},\ c(\lambda) = e^{-\lambda},\ w(p) = \log(\lambda) \text{ and } \ t(x) = x.$$

Then based on Theorem 6.2.10, we have $T(\mathbf{X}) = \sum_{i=1}^{n} t(X_i) = \sum_{i=1}^{n} X_i$ is a sufficient statistic for λ.

<u>**Note:**</u> There can be more than one sufficient statistic for a given model (e.g., $\mathbf{X}$ itself is a sufficient statistic; any one-to-one function of a sufficient statistic is also a sufficient statistic).

6.2.2 Minimal Sufficient Statistics

Definition 6.2.11: A sufficient statistic $T(\mathbf{X})$ is called minimal sufficient statistic if, for any other sufficient statistic $T'(\mathbf{X})$, $T(\mathbf{X})$ is a function of $T'(\mathbf{X})$.

Note:

1. The partition associated with a minimal sufficient statistic is the coarsest possible partition for a sufficient statistic so that it achieves the greatest possible data reduction for a sufficient statistic;
2. Minimal sufficient statistic "eliminates" all the extraneous information in the sample and leaves only that which contains information about θ;
3. How to find the minimal sufficient statistics?

Example 6.2.12: (Two Normal Sufficient Statistics)

Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$, where σ^2 is known. As seen in Example 6.2.9, $T'(\mathbf{X}) = (\bar{X}, S^2)$ is a sufficient statistic for μ (σ^2 is a known parameter in this case). However, we can reduce further $T'(\mathbf{X})$ by defining the function $r(a, b) = a$ so that if $T(\mathbf{X}) = r(\bar{X}, S^2) = \bar{X}$ is a sufficient statistic for μ (which we can find from Example 6.2.7). Note that we have not shown that $\bar{X}$ is minimal sufficient for μ in this case where σ^2 is known.

Theorem 6.2.13: (Minimal Sufficient Statistics)

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

Example 6.2.14: (Normal Minimal Sufficient Statistic)

Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$, where both μ and σ^2 are unknown. Let $\mathbf{x}$ and $\mathbf{y}$ be two sample points with corresponding sample means and variances $(\bar{\mathbf{x}}, S^2_{\mathbf{x}})$ and $(\bar{\mathbf{y}}, S^2_{\mathbf{y}})$. Show that $(\bar{X}, S^2)$ is a minimal sufficient statistic for (μ, σ^2).

Note: If the set of $\mathbf{x}$ values on which the pdf or pmf is positive depends on the parameter θ, then, for the ratio in Theorem 6.2.13 to be constant as a function of θ, the numerator and denominator must be positive for exactly the same values of θ.

Example 6.2.15: (Uniform Minimal Sufficient Statistic)

Suppose $X_1, \ldots, X_n$ are iid uniform observations on the interval $(\theta, \theta+1)$, for $-\infty < \theta < \infty$. Show that $T(X) = \big(X_{(1)}, X_{(n)}\big)$ is a minimal sufficient statistic. (In this example, the dimension of the minimal sufficient statistic does not match the dimension of the parameter.)

Note: A minimal sufficient statistic is not unique! Any one-to-one function of a minimal sufficient statistic is also minimal sufficient statistic.

Illustration:

1. Let $X_1, \ldots, X_n$ be iid uniform observations on the interval $(\theta, \theta+1)$. As shown above, $T(\mathbf{X}) = \left(X_{(1)}, X_{(n)}\right)$ is a minimal sufficient for θ. Hence, $\left(X_{(n)} - X_{(1)}, \left(X_{(1)} + X_{(n)}\right)/2\right)$ is also a minimal sufficient statistic for θ.
2. Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$, μ and σ^2 are unknown. $(\bar{X}, S^2)$ is a minimal sufficient statistic for (μ, σ^2). Hence, $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is also a minimal sufficient statistic for (μ, σ^2).

Question: Let $X_1, \ldots, X_n$ be iid observations from uniform$(\theta, \theta+1)$, for $-\infty < \theta < \infty$. $T(X) = \left(X_{(1)}, X_{(n)}\right)$ is a minimal sufficient statistic for θ, is $X_{(n)} - X_{(1)}$ also a minimal sufficient statistic?

6.2.3 Ancillary Statistics

Definition 6.2.16: A statistic $S(\mathbf{X})$ whose distribution does not depend on the parameter θ is called an ancillary statistic.

Example 6.2.17: (Uniform Ancillary Statistic)
Let $X_1, \ldots, X_n$ be iid uniform observations on the interval $(\theta, \theta + 1)$, for $-\infty < \theta < \infty$. Show that the range statistic, $R = X_{(n)} - X_{(1)}$, is an ancillary statistic.

Example 6.2.18: (Location Family Ancillary Statistic)
Suppose $X_1, \ldots, X_n$ are iid observations from a location parameter family with cdf $F(x-\theta)$, $-\infty < \theta < \infty$. Show the range statistic, $R = X_{(n)} - X_{(1)}$, is an ancillary statistic.

Example 6.2.19: (Scale Family Ancillary Statistic)
Suppose $X_1, \ldots, X_n$ are iid observations from a scale parameter family with cdf $F(x/\sigma)$, $\sigma > 0$. Then any statistic that depends on the sample only through the $n-1$ values $(X_1/X_n, \ldots, X_{n-1}/X_n)$ is an ancillary statistic.

Remark: From Chapter 4 (Example 4.3.6), it was shown that if X_1 and X_2 are iid $\mathrm{n}(0, \sigma^2)$, where $\sigma = 1$, then X_1/X_2 is Cauchy$(0, 1)$. In fact, this also holds for any $\sigma > 0$.

6.2.4 Sufficient, Ancillary and Complete Statistics

Question: Is an ancillary statistic not related at all to minimal sufficient statistic"?

Recall tat if $X_1, \ldots, X_n$ are iid uniform observations on the interval $(\theta, \theta + 1)$, then $\left(X_{(1)}, X_{(n)}\right)$ and $\left(X_{(n)} - X_{(1)}, \left(X_{(1)} + X_{(n)}\right)/2\right)$ are minimal sufficient statistics for θ. However, we also know that $R = X_{(n)} - X_{(1)}$ is an ancillary statistic. Hence, in this case the minimal sufficient and ancillary statistics are related.

Question: When is a minimal sufficient statistic independent of every ancillary statistic?

Definition 6.2.21: Let $f(t|\theta)$ be a family of pdfs or pmfs for a statistic $T(\mathbf{X})$. The family of probability distributions $f(t|\theta)$ is called *complete* if

$$\mathrm{E}_\theta\big(g(T)\big) = 0 \text{ for all } \theta \in \Theta \text{ implies } P_\theta\big(g(T) = 0\big) = 1 \text{ for all } \theta \in \Theta.$$

Equivalently, $T(\mathbf{X})$ is called a complete statistic.

Illustration: Consider the family of distributions $\mathrm{n}(\theta, 1)$, $-\infty < \theta < \infty$. If $g(X) = X$, then $\mathrm{E}_\theta g(X) = E_\theta X = 0$ when $\theta = 0$ but $P\big(g(X) = 0\big) = P(X = 0) = 0$ since X is a continuous random variable. So this family of distributions is complete for $-\infty < \theta < \infty$.

Example: Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\theta, 1)$. Show that $T(\mathbf{X}) = (X_1, X_2)$ is not complete.

Example: Let $X_1, \ldots, X_n$ be iid observations from uniform$(\theta, \theta+1)$, for $-\infty < \theta < \infty$. Show that $T(\mathbf{X}) = \left(X_{(1)}, X_{(n)}\right)$ is not complete.

Example 6.2.22: (Binomial Complete Sufficient Statistic)
Suppose that T has a binomial(n, p) distribution with $0 < p < 1$. Show that T is a complete statistic.

Example 6.2.23: (Uniform Complete Sufficient Statistic)
Let $X_1, \ldots, X_n$ be iid uniform$(0, \theta)$ observations, $0 < \theta < \infty$. Show that $T(\mathbf{X}) = X_{(n)}$ is a complete statistic

Question: Is there an easier way to find a complete statistic?

Theorem 6.2.25: (Complete Statistics in Exponential Family)
Let $X_1, \ldots, X_n$ be iid observations from an exponential family with pdf or pmf of the form

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

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. Then the statistic

$$T(\mathbf{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \sum_{j=1}^{n} t_2(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j)\right)$$

is complete if $\Big\{\big(w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta})\big) : \boldsymbol{\theta} \in \Theta\Big\}$ contains an open set in $\mathfrak{R}^k$.

Example: The distribution $\mathrm{n}(\mu, \mu^2)$ (recall from Example 3.4.8 that this distribution is a member of the curved exponential family of distributions) does not contain a two-dimensional open set because it contains only points on the parabola. Hence, this distribution would not satisfy the conditions of Theorem 6.2.25.

Theorem 6.2.24: (Basu's Theorem)
Let $T(\mathbf{X})$ is a complete and minimal sufficient statistic, then $T(\mathbf{X})$ is independent of every ancillary statistic.

Remark: Basu's Theorem allows us to deduce the independence of two statistics without ever finding the joint distribution of the two statistic.

<u>Example 6.2.26</u>: (Using Basu' Theorem - I)

Let $X_1, \ldots, X_n$ be iid exponential(θ) observations. Compute $E_\theta g(\mathbf{X})$ where $g(\mathbf{X}) = \dfrac{X_n}{X_1 + \cdots + X_n}$.

<u>Example 6.2.27</u>: (Using Basu' Theorem - II)

Let $X_1, \ldots, X_n$ be iid observations from n(μ, σ^2) population. Using Basu's Theorem, show that $\bar{X}$ and S^2 are independent.

<u>Theorem 6.2.2.28</u>: (Bahadur's Theorem)

If a minimal sufficient statistic exists, then any complete sufficient statistic is also a minimal sufficient statistic.

<u>Example</u>: (A Minimal Sufficient Statistic NOT Complete)

A minimal sufficient statistic is not necessarily a complete statistic. Let $X_1, \ldots, X_n$ be iid observations from uniform$(\theta, \theta + 1)$, for $-\infty < \theta < \infty$. From Example 6.2.15, we know that $T(\mathbf{X}) = \big(X_{(1)}, X_{(n)}\big)$ is a minimal sufficient statistic. However, $T(\mathbf{X}) = \big(X_{(1)}, X_{(n)}\big)$ is not complete.

6.3 Likelihood Principle

We study a specific, important statistic called the likelihood function that can also be used to summarize data.

Definition 6.3.1: Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of the sample $\mathbf{X} = (X_1, \ldots, X_n)$. Then give that $\mathbf{X} = \mathbf{x}$ is observed, the function of θ defined by

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$$

is called the likelihood function.

Question: What is the difference between the likelihood function and the pdf (or pmf)?
Answer: They have the same formula. The distinction between them is which variable is fixed and which is varying.

Likelihood Principle: If $\mathbf{x}$ and $\mathbf{y}$ are two sample points such that $L(\theta|\mathbf{x})$ is proportional to $L(\theta|\mathbf{y})$, i.e., there exists a constant $C(\mathbf{x}, \mathbf{y})$ such that $L(\theta|\mathbf{x}) = C(\mathbf{x}, \mathbf{y})L(\theta|\mathbf{y})$, then the conclusions drawn from $\mathbf{x}$ and $\mathbf{y}$ for θ should be identical.

Remark:

1. Likelihood principle states that even if two sample points have only proportional likelihoods, then they will contain equivalent information about θ.
2. Given two parameter values θ_1 and θ_2, the likelihood function tells us if θ_1 is a more plausible (not probable) parameter value than θ_2 in light of the data gathered.
3. Fiducial inference (Fisher, 1930) interprets likelihoods as probabilities for θ, called *inverse probabilities*, without calling on prior probability distributions required in Bayesian inference.

Example 6.3.2: (Negative Binomial Likelihood)

Let X have a negative binomial distribution with $r = 3$ and success probability p. If $x = 2$ is observed, then the likelihood function is the fifth-degree polynomial on $0 \geq p \geq 1$ defined by

$$L(p|2) = P_p(X = 2) = \binom{4}{2} p^3 (1-p)^2.$$

In general, if $X = x$ is observed, then the likelihood function is polynomial of degree $3 + x$,

$$L(p|x) = \binom{3+x-1}{x} p^3 (1-p)^x.$$

Example 6.3.3: (Normal Fiducial Distribution)

Let $X_1, \ldots, X_n$ be iid $\mathrm{n}(\mu, \sigma^2)$, σ^2 known. Then

$$\begin{aligned} L(\mu|\mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \\ &= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2}{2\sigma^2}\right). \end{aligned}$$

First, note that $C(\mathbf{x}, \mathbf{y})$ exists if and only if $\bar{x} = \bar{y}$, in which case

$$C(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma^2} + \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{2\sigma^2}\right).$$

Thus, the Likelihood Principle states that the conclusions about μ drawn from any sample points satisfying $\bar{x} = \bar{y}$ should be identical.

Second, the fiducial distribution, $M(\mathbf{x})L(\mu|\mathbf{x})$, has a normal distribution $\mathrm{n}(\bar{x}, \sigma^2/n)$, where

$$M(\mathbf{x}) = \frac{1}{\int_{-\infty}^{\infty} L(\mu|\mathbf{x}) d\mu}.$$

Thus we have

$$\begin{aligned} 0.95 &= P(-1.96 < \frac{\mu - \bar{x}}{\sigma/\sqrt{n}} < 1.96) \\ &= P(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}) \end{aligned}$$

Example: (Likelihood Function for Uniform Distribution)

Let $X_1, \ldots, X_n$ be iid uniform$(0, \theta)$, then the likelihood function is

$$L(\theta|\mathbf{x}) = \frac{1}{\theta^n} I_{[0<x_{(n)}<\theta]}(x_1, \ldots, x_n)$$