第2章 数据简化原理

《统计推断》 第6章

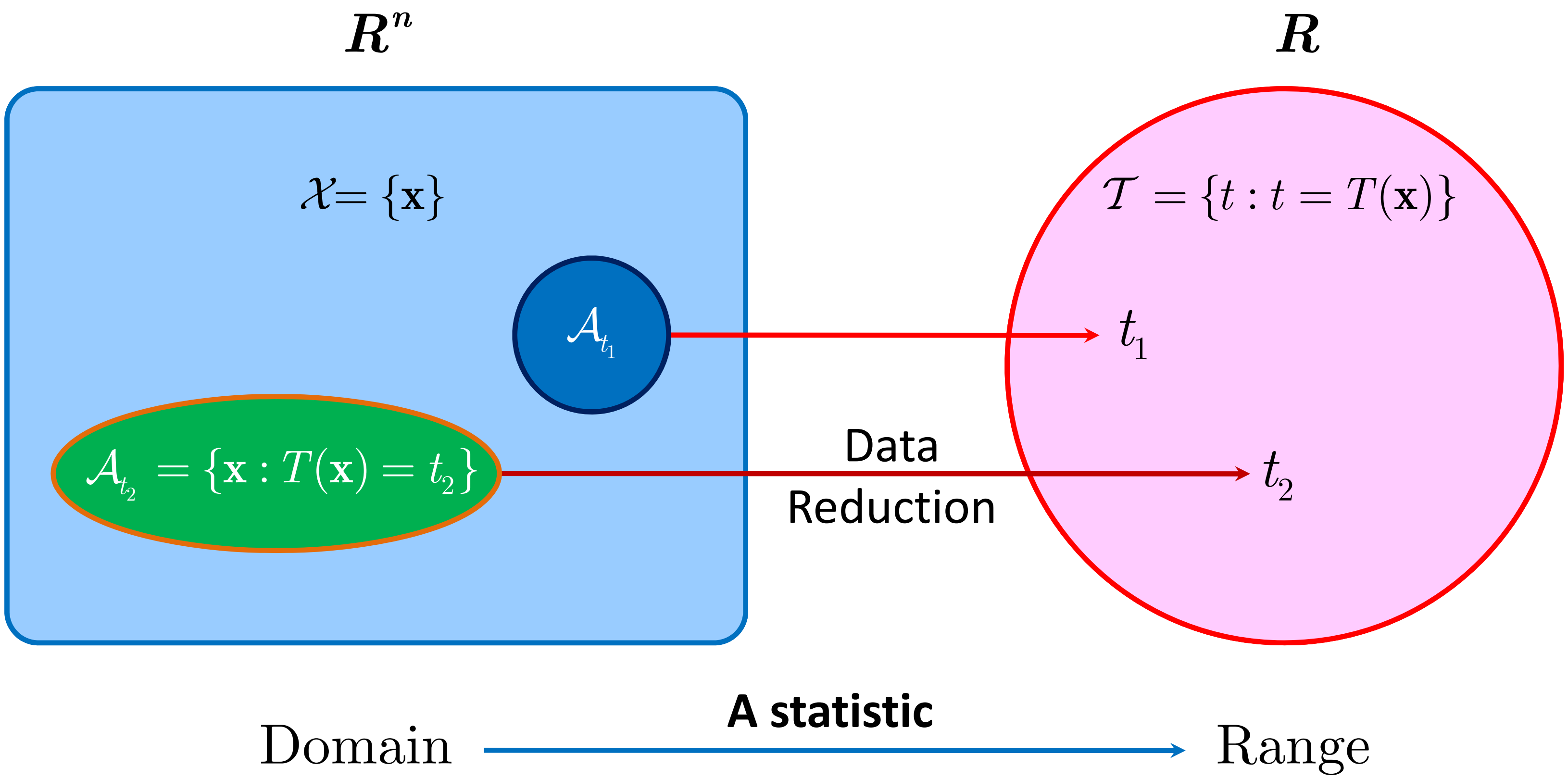
感谢清华大学自动化系江瑞教授提供PPT

Key points of statistics

- Population
  - A distribution that we are unable to see but interested in
- Sample
  - A set of iid random variables sampled from the population
- Statistic
  - Summary of the sample, reduction of the data
  - Identical observations of samples lead to equal values of statistics
  - Equal values of statistics do not mean identical observations of samples

$$\begin{array}{lcl} \mathbf{x} = \mathbf{y} & \Rightarrow & T(\mathbf{x}) = T(\mathbf{y}) \\ T(\mathbf{x}) = T(\mathbf{y}) & \not\Rightarrow & \mathbf{x} = \mathbf{y} \end{array}$$

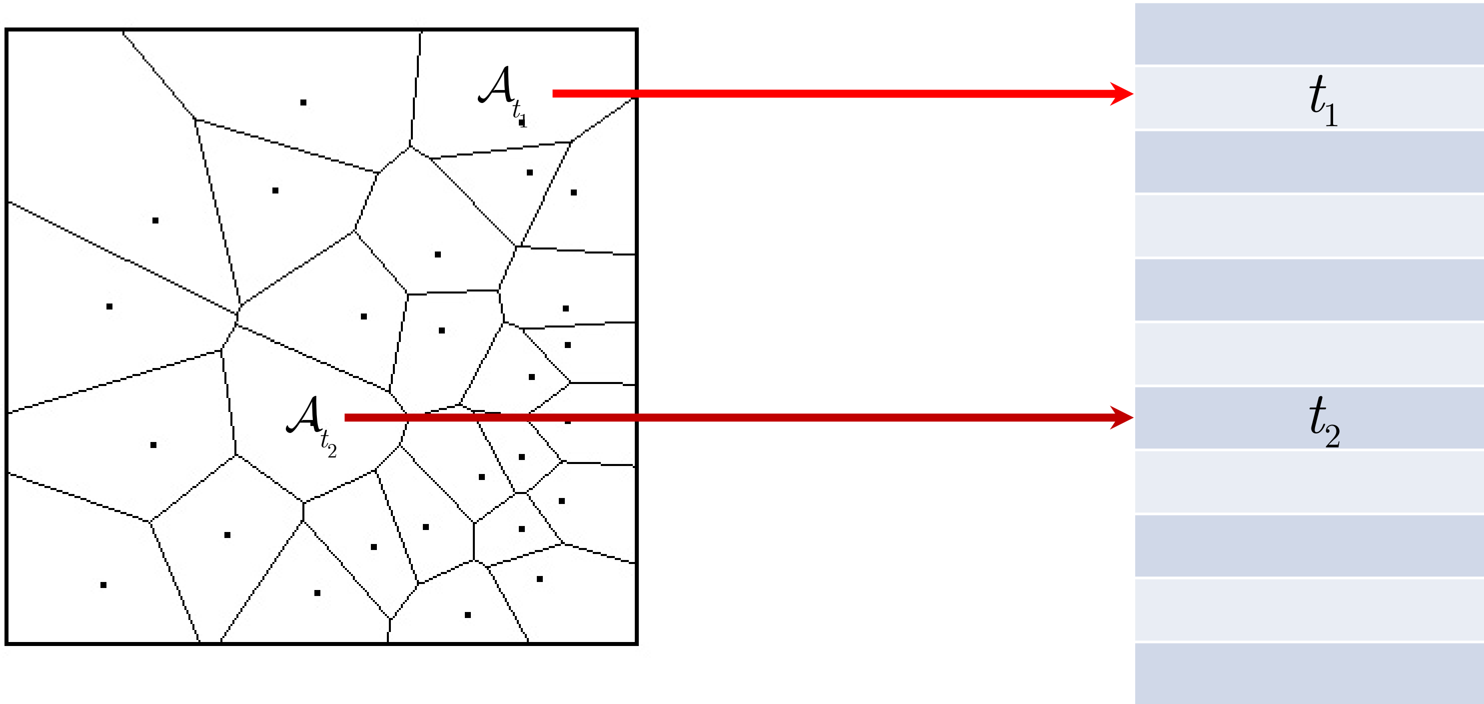
Data reduction



Report a small number of data instead of a large number of data

Sample space partition

$\mathcal{X} = \{\mathbf{x}\}$ $\qquad\qquad$ $\mathcal{T} = \{t : t = T(\mathbf{x})\}$



A statistic

$$\text{Domain} \longrightarrow \text{Range}$$

$$\mathcal{A}_{t_i} = \{\mathbf{x} : T(\mathbf{x}) = t_i\}$$

A statistic implies a partition of the sample space

The sufficiency principle

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 $T(\mathbf{X})$. 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.

A sufficient statistic captures **ALL** the information about the parameter contained in the sample. Any additional information in the sample, besides the value of the sufficient statistic, does **not** contain any more information about the parameter.

Sufficient statistics

Sufficient statistics

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.

$$P_\theta(\mathbf{X}=\mathbf{x} \mid T(\mathbf{X})=T(\mathbf{x})) = \frac{P_\theta(\mathbf{X}=\mathbf{x} \text{ and } T(\mathbf{X})=T(\mathbf{x}))}{P_\theta(T(\mathbf{X})=T(\mathbf{x}))}$$

$$= \frac{P_\theta(\mathbf{X}=\mathbf{x})}{P_\theta(T(\mathbf{X})=T(\mathbf{x}))}$$

$$P(\mathbf{X}=\mathbf{x} \mid T(\mathbf{X})=T(\mathbf{x})) = \frac{p(\mathbf{x} \mid \theta)}{q(T(\mathbf{x}) \mid \theta)}$$

说明：上面的写法只对离散分布适用。

Sufficient statistics

Sufficient condition

If $p(\mathbf{x} \mid \theta)$ is the joint pdf or pmf of $\mathbf{X}$ and $q(t \mid \theta)$ 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

$$\frac{p(\mathbf{x} \mid \theta)}{q(T(\mathbf{x}) \mid \theta)}$$

is constant as a function of θ.

需要先给定T，然后再验证其充分性

Binomial sufficient statistic

Let $X_1, \ldots, X_n$ be iid Bernoulli random variables with parameter θ, where $0 < \theta < 1$. Define the statistic $T(\mathbf{X}) = X_1 + \cdots + X_n = \sum_{i=1}^{n} X_i$. Then,

$$p(\mathbf{x} \mid \theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i} = \theta^{t} (1-\theta)^{n-t}$$

$$q(T(\mathbf{x}) \mid \theta) = \binom{n}{t} \theta^{t} (1-\theta)^{1-t}$$

$$\frac{p(\mathbf{x} \mid \theta)}{q(T(\mathbf{x}) \mid \theta)} = \frac{\theta^{t} (1-\theta)^{n-t}}{\binom{n}{t} \theta^{t} (1-\theta)^{1-t}} = \binom{n}{t}^{-1} = \binom{n}{\sum_{i=1}^{n} x_i}^{-1}$$

The total number of successes in a Bernoulli sample is a sufficient statistic for the ratio of success.

Multinomial sufficient statistic

Let $X_1,\ldots,X_n$ be iid random variables from a multinomial trial, for which parameters are $0<\theta_k<1$ and $\sum_{k=1}^{m}\theta_k=1$. Define $T(\mathbf{X})=(T_1(\mathbf{X}),\ldots,T_m(\mathbf{X}))$, where $T_k(\mathbf{X})=\sum_{i=1}^{n}I(X_i=k)$ is the number of X_is that are equal to the k-th outcome. Then,

$$p(\mathbf{x}\mid\theta) = \prod_{i=1}^{n}\prod_{k=1}^{m}\theta_k^{I(x_i=k)} = \prod_{k=1}^{m}\prod_{i=1}^{n}\theta_k^{I(x_i=k)} = \prod_{k=1}^{m}\theta_k^{\sum_{i=1}^{n}I(x_i=k)} = \prod_{k=1}^{m}\theta_k^{n_k}$$

$$q(T(\mathbf{x})\mid\theta) = \frac{n!}{n_1!\cdots n_m!}\prod_{k=1}^{m}\theta_k^{n_k} = \frac{(\sum_{k=1}^{m}n_k)!}{\prod_{k=1}^{m}n_k!}\prod_{k=1}^{m}\theta_k^{n_k}$$

$$\frac{p(\mathbf{x}\mid\theta)}{q(T(\mathbf{x})\mid\theta)} = \frac{\prod_{k=1}^{m}\theta_k^{n_k}}{\frac{(\sum_{k=1}^{m}n_k)!}{\prod_{k=1}^{m}n_k!}\prod_{k=1}^{m}\theta_k^{n_k}} = \left[\frac{(\sum_{k=1}^{m}n_k)!}{\prod_{k=1}^{m}n_k!}\right]^{-1}$$

The vector of the number of occurrence of each outcome is a sufficient statistic for the multinomial distribution.

Normal sufficient statistic

Let $X_1,\ldots,X_n$ be iid random variables with common pdf $N(\mu,\sigma^2)$, where σ^2 is known. Define the statistic $T(\mathbf{X}) = \bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. Then,

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

$$
q(T(\mathbf{x}) \mid \mu) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right]
$$

$$
\frac{p(\mathbf{x} \mid \mu)}{q(T(\mathbf{x}) \mid \mu)} = n^{-\frac{1}{2}} (2\pi\sigma^2)^{-\frac{n-1}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]
$$

The sample mean is a sufficient statistic for the population mean when population variance is known.

Sufficient order statistics

Let $X_1, \ldots, X_n$ be iid random variables from a certain pdf $f(x)$, about which we are unable to specific any more information. Define the statistic $T(\mathbf{X}) = (X_{(1)}, \ldots, X_{(n)})$.

Then,

$$q(T(\mathbf{x})) = f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = n!\, f_X(x_1) f_X(x_2) \cdots f_X(x_n) \propto p(\mathbf{x})$$

The vector of all order statistics is a sufficient statistic for the unknown population $f(x)$.

Factorization theorem

Sufficient and necessary condition

Let $f(\mathbf{x} \mid \theta)$ 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 exist functions $g(t \mid \theta)$ and $h(\mathbf{x})$ such that, for all sample points $\mathbf{x}$ and all parameter points θ,

$$f(\mathbf{x} \mid \theta) = g(T(\mathbf{x}) \mid \theta)h(\mathbf{x}).$$

因子分解定理告诉我们如何去构造充分统计量

Sufficiency

If there exist functions $g(t \mid \theta)$ and $h(\mathbf{x})$ such that

$$f(\mathbf{x} \mid \theta) = g(T(\mathbf{x}) \mid \theta)h(\mathbf{x}),$$

then $T(\mathbf{X})$ is a sufficient statistic for θ.

Let $q(t \mid \theta)$ be the pmf of $T(\mathbf{X})$, examine the ratio $f(\mathbf{x} \mid \theta) \,/\, q(T(\mathbf{x}) \mid \theta)$.

$$\begin{aligned}
\frac{f(\mathbf{x} \mid \theta)}{q(T(\mathbf{x}) \mid \theta)} &= \frac{g(T(\mathbf{x}) \mid \theta)h(\mathbf{x})}{q(T(\mathbf{x}) \mid \theta)} \\
&= \frac{g(T(\mathbf{x}) \mid \theta)h(\mathbf{x})}{\sum_{A_{T(\mathbf{x})}} g(T(\mathbf{y}) \mid \theta)h(\mathbf{y})} \;, \quad A_{T(\mathbf{x})} = \{\mathbf{y} : T(\mathbf{y}) = T(\mathbf{x})\} \\
&= \frac{g(T(\mathbf{x}) \mid \theta)h(\mathbf{x})}{g(T(\mathbf{x}) \mid \theta)\sum_{A_{T(\mathbf{x})}} h(\mathbf{y})} \\
&= \frac{h(\mathbf{x})}{\sum_{A_{T(\mathbf{x})}} h(\mathbf{y})}
\end{aligned}$$

Independent of θ, therefore, $T(\mathbf{X})$ is a sufficient statistic for θ.

Necessity

If $T(\mathbf{X})$ is a sufficient statistic for θ, then there exist functions $g(t \mid \theta)$ and $h(\mathbf{x})$ such that

$$f(\mathbf{x} \mid \theta) = g(T(\mathbf{x}) \mid \theta) h(\mathbf{x})$$

Choose

$$\begin{aligned} g(T(\mathbf{x}) \mid \theta) &= P_\theta(T(\mathbf{X}) = T(\mathbf{x})), \text{ the pmf of } T(\mathbf{X}) \\ h(\mathbf{x}) &= P(\mathbf{X} = x \mid T(\mathbf{X}) = T(\mathbf{x})) \end{aligned}$$

Since $T(\mathbf{X})$ is a sufficient statistic for θ, $h(\mathbf{x})$ does not depend on θ. We now show that the product of the above valid choice yields the pmf of $\mathbf{X}$.

$$\begin{aligned} f(\mathbf{x} \mid \theta) &= P_\theta(\mathbf{X} = \mathbf{x}) \\ &= P_\theta(\mathbf{X} = \mathbf{x} \textbf{ AND } T(\mathbf{X}) = T(\mathbf{x})) \\ &= P_\theta(T(\mathbf{X}) = T(\mathbf{x})) P_\theta(\mathbf{X} = \mathbf{x} \mid T(\mathbf{X}) = T(\mathbf{x})) \\ &= P_\theta(T(\mathbf{X}) = T(\mathbf{x})) P(\mathbf{X} = \mathbf{x} \mid T(\mathbf{X}) = T(\mathbf{x})) \\ &= g(T(\mathbf{x}) \mid \theta) h(\mathbf{x}) \end{aligned}$$

Normal sufficient statistic

Let $X_1,\ldots,X_n$ be iid random variables with common pdf $N(\mu,\sigma^2)$. Define statistics

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Then,

$$\begin{aligned}
f(\mathbf{x} \mid \mu,\sigma^2) &= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right] \\
&= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \overline{x} + \overline{x} - \mu)^2\right] \\
&= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\overline{x})^2 + \sum_{i=1}^{n}(\overline{x}-\mu)^2 + \underbrace{2\sum_{i=1}^{n}(x_i-\overline{x})(\overline{x}-\mu)}_{=0}\right]\right\} \\
&= \underbrace{(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{n}{2\sigma^2}(\overline{x}-\mu)^2 - \frac{n-1}{2\sigma^2}s^2\right]}_{g\left(\overline{x},s^2\mid\mu,\sigma^2\right)} \times \underbrace{1}_{h(\mathbf{x})}
\end{aligned}$$

The vector of the sample mean and the sample variance is a sufficient statistic in the case that the population variance is unknown.

Uniform sufficient statistic

Let $X_1,\ldots,X_n$ be iid random variables from a uniform $(0,\theta)$ distribution.

$$f(x)=\begin{cases}1/\theta & 0<x<\theta\\ 0 & \text{otherwise}\end{cases}$$

Then,

$$f(\mathbf{x}\mid\theta)=\prod_{i=1}^{n}\left[\theta^{-1}I_{(0,\theta)}(x_i)\right]=\theta^{-n}\prod_{i=1}^{n}I_{(0,\theta)}(x_i)=\theta^{-n}I_{(0,\theta)}(x_{(n)})\prod_{i=1}^{n}I_{(0,\infty)}(x_i)$$

Choose

$$g(T(\mathbf{x})\mid\theta)=\theta^{-n}I_{(0,\theta)}(x_{(n)})$$

$$h(\mathbf{x})=\prod_{i=1}^{n}I_{(0,\infty)}(x_i),\ \text{independent of }\theta$$

By the factorization theorem,

$$T(\mathbf{X})=X_{(n)}=\max_{1\le i\le n}X_i.$$

is a sufficient statistic for θ.

The largest order statistic is a sufficient statistic for the uniform population.

Exponential family

Sufficient statistic for the exponential family

Let $X_1, \ldots, X_n$ be iid random variables from a pdf or pmf that belongs to an exponential family given by

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

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d), d \leq k$. Then,

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

is a sufficient statistic for $\boldsymbol{\theta}$.

Normal sufficient statistics

Normal pdf

$$\varphi(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Written it as exponential family,

$$\varphi(x \mid \mu, \sigma^2) = \underbrace{1}_{h(x)} \cdot \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2}\right)}_{c(\mu,\sigma^2)} \exp\left[\underbrace{\frac{\mu}{\sigma^2}}_{w_1(\mu,\sigma^2)} \underbrace{x}_{t_1(x)} + \left(\underbrace{-\frac{1}{2\sigma^2}}_{w_2(\mu,\sigma^2)} \underbrace{x^2}_{t_2(x)}\right)\right].$$

Thus for a sample $X_1, \ldots, X_n$, a sufficient statistic for (μ, σ^2) is

$$\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right).$$

Sufficient statistic is not unique

Let $X_1,\ldots,X_n$ be iid random variables with common pdf $N(\mu,\sigma^2)$, where σ^2 is known. Define

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Then,

$$f(\mathbf{x} \mid \mu) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \mu)^2\right]$$

$$= \underbrace{(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2 - \frac{n-1}{2\sigma^2}s^2\right]}_{g(\bar{x},s^2|\mu)} \times \underbrace{1}_{h(\mathbf{x})}$$

$$= \underbrace{\exp\left[-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right]}_{g(\bar{x}|\mu)}\underbrace{(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]}_{h(x)}$$

Two trivial sufficient statistics

Let $X_1, \ldots, X_n$ be iid random variables from a certain pdf $f(x \mid \theta)$.

Define the statistic $T(\mathbf{X}) = (X_1, \ldots, X_n)$, then,

$$f(\mathbf{x} \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta) = \underbrace{\prod_{i=1}^{n} f(x_i \mid \theta)}_{g(T(x)|\theta)} \times \underbrace{1}_{h(x)}$$

Define the statistic $T(\mathbf{X}) = (X_{(1)}, \ldots, X_{(n)})$, then,

$$f(\mathbf{x} \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta) = \underbrace{\prod_{i=1}^{n} f(x_{(i)} \mid \theta)}_{g(T(x)|\theta)} \times \underbrace{1}_{h(x)}$$

The complete sample is always a sufficient statistic!
The vector of all order statistics is always a sufficient statistic!

Functions of a sufficient statistic

Suppose $T(\mathbf{X})$ is a sufficient statistic, by the Factorization Theorem, there exist g and h such that

$$f(\mathbf{x} \mid \theta) = g(T(\mathbf{x}) \mid \theta)h(\mathbf{x}).$$

Now, define $T^\star(\mathbf{x}) = r(T(\mathbf{x}))$ for all $\mathbf{x}$, where r is a one-to-one function with inverse r^{-1}. Then,

$$f(\mathbf{x} \mid \theta) = g(T(\mathbf{x}) \mid \theta)h(\mathbf{x}) = g(r^{-1}(T^\star(\mathbf{x})) \mid \theta)h(\mathbf{x}).$$

Define a new function $g^\star(t \mid \theta) = g(r^{-1}(t) \mid \theta)$, we see that

$$f(\mathbf{x} \mid \theta) = g^\star(T^\star(\mathbf{x}) \mid \theta)h(\mathbf{x}).$$

Again, by the Factorization Theorem, $T^\star(\mathbf{x})$ is a sufficient statistic.

Any one-to-one function of a sufficient statistic is a sufficient statistic

Minimal sufficient statistics

Minimal sufficient statistics

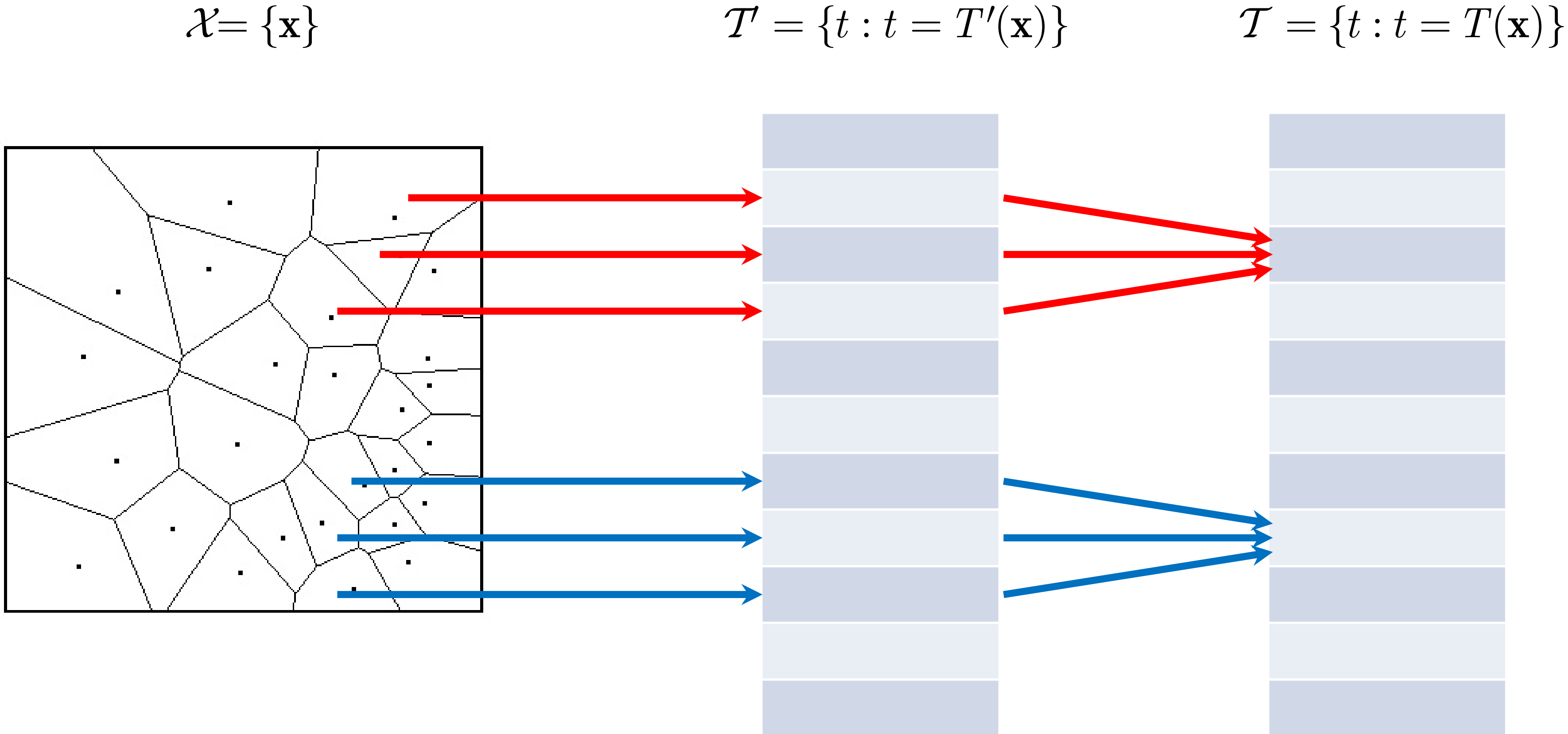
A sufficient statistic $T(\mathbf{X})$ is called a minimal sufficient statistic if, for any other sufficient statistic $T'(\mathbf{X})$, $T(\mathbf{X})$ is a function of $T'(\mathbf{X})$. Or simply,

$$\text{if } T'(\mathbf{x}) = T'(\mathbf{y}), \text{ then } T(\mathbf{x}) = T(\mathbf{y}).$$

极小的意义：最大程度的简化

直观上：每做一次简化，是进行一次函数约化。对应到样本空间上，涵盖的样本更多。

Coarsest partition of the sample space



Two normal sufficient statistic

Let $X_1,\ldots,X_n$ be iid random variables with common pdf $N(\mu,\sigma^2)$,
where σ^2 is known.

Define

$$T(\mathbf{X}) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S(\mathbf{X}) = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Then, both

$$T_1(\mathbf{X}) = T(\mathbf{X}) = \bar{X}$$

and

$$T_2(\mathbf{X}) = (T(\mathbf{X}), S(\mathbf{X})) = (\bar{X}, S^2)$$

are sufficient statistics of the population mean.

However, if we define a function $\varphi(a,b) = a,$ then,

$$T_1(\mathbf{x}) = \bar{x} = \varphi(\bar{x}, s^2) = \varphi(T_2(\mathbf{x})).$$

Minimal sufficient statistics

Sufficient condition

Let $f(\mathbf{x} \mid \theta)$ be the pmf or pdf of a 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 $f(\mathbf{x} \mid \theta) \,/\, f(\mathbf{y} \mid \theta)$ is constant as a funciton of θ if and only if $T(\mathbf{x}) = T(\mathbf{y})$. Then $T(\mathbf{X})$ is a minimal sufficient statistic for θ.

定理证明 (I)

为简化证明，不妨设 $f(x|\theta) > 0, \forall x \in \mathcal{X}.$

现证T(x)是充分统计量。令

$$A_t = \{x : T(x) = t\}$$

对任意 $x, y \in A_t, f(x|\theta)/f(y|\theta)$ 是 θ 的常函数，于是可定义 $\mathcal{X}$ 上的函数

$$h(x) = \frac{f(x|\theta)}{f(A_{T(x)}|\theta)}$$

再令 $g(t|\theta) = f(A_t|\theta)$,于是

$$f(x|\theta) = \frac{f(x|\theta)f(A_t|\theta)}{f(A_{T(x)}|\theta)} = g(T(x)|\theta)h(x)$$

由因子分解定理得T(x)是充分统计量。

定理证明(II)

- 下面证明T(x)是极小充分统计量。设$T'(x)$ 是任一充分统计量。由因子分解定理

$$f(x|\theta) = g'(T'(x)|\theta)h'(x)$$

- 若x与y满足：$T'(x) = T'(y)$

$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{g'(T'(x)|\theta)h'(x)}{g'(T'(y)|\theta)h'(y)} = \frac{h'(x)}{h'(y)}$$

- 上式与 θ 无关。由定理假设知$T(x) = T(y)$, 从而T(x)是 $T'(x)$ 的函数。因 $T'(x)$ 任意，故T(x)是极小充分统计量

Normal minimal sufficient statistic

Let $X_1,\ldots,X_n$ be iid random variables from a normal population $N(\mu,\sigma^2)$, both μ, σ^2 are unknown. Let $\mathbf{x}$ and $\mathbf{y}$ denote two sample points, and let $(\bar{x}, s_x^2)$ and $(\bar{y}, s_y^2)$ be the sample means and variances corresponding to the sample points of $\mathbf{x}$ and $\mathbf{y}$, respectively.

Then,

$$f(\mathbf{x} \mid \mu,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2 - \frac{n-1}{2\sigma^2}s_x^2\right],$$

$$f(\mathbf{y} \mid \mu,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2 - \frac{n-1}{2\sigma^2}s_y^2\right],$$

$$\frac{f(\mathbf{x} \mid \mu,\sigma^2)}{f(\mathbf{y} \mid \mu,\sigma^2)} = \exp\left\{-\frac{1}{2\sigma^2}\left[-n(\bar{x}^2-\bar{y}^2) + 2n\mu(\bar{x}-\bar{y}) - (n-1)(s_x^2 - s_y^2)\right]\right\},$$

which will be constant as a function of (μ,σ^2) if and only if $\bar{x} = \bar{y}$ and $s_x^2 = s_y^2$.

Therefore, $(\bar{X},\ S^2)$ is a minimal sufficient statistic of (μ,σ^2).

Normal minimal sufficient statistic

Since

$$
\begin{aligned}
(n-1)s^2 &= \sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}x_i^2 - n\bar{x}^2 \\
\frac{f(\mathbf{x}\mid\mu,\sigma^2)}{f(\mathbf{y}\mid\mu,\sigma^2)} &= \exp\left\{-\frac{1}{2\sigma^2}\left[-n(\bar{x}^2-\bar{y}^2)+2n\mu(\bar{x}-\bar{y})-(n-1)(s_x^2-s_y^2)\right]\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\left[-n(\bar{x}^2-\bar{y}^2)+2n\mu(\bar{x}-\bar{y})-\left(\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)+\left(\sum_{i=1}^{n}y_i^2-n\bar{y}^2\right)\right)\right]\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}\left[2n\mu(\bar{x}-\bar{y})-\left(\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}y_i^2\right)\right]\right\}
\end{aligned}
$$

which will be constant as a function of (μ,σ^2) if and only if $\bar{x}=\bar{y}$ and $\sum_{i=1}^{n}x_i^2=\sum_{i=1}^{n}y_i^2$.

Therefore, $\left(\sum_{i=1}^{n}X_i,\sum_{i=1}^{n}X_i^2\right)$ is also a minimal sufficient statistic of (μ,σ^2).

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

Uniform minimal sufficient statistic

Let $X_1, \ldots, X_n$ be iid random variables from a uniform population whose pdf is

$$f(x \mid \theta) = \begin{cases} 1 & \text{if } \theta < x < \theta + 1 \\ 0 & \text{otherwise} \end{cases}. \text{ Then, } f(\mathbf{x} \mid \theta) = \begin{cases} 1 & \text{if } \theta < x_i < \theta + 1,\ i = 1,2,\ldots n \\ 0 & \text{otherwise} \end{cases}.$$

In other words,

$$f(\mathbf{x} \mid \theta) = \begin{cases} 1 & \text{if } \max\limits_{1 \le i \le n} x_i - 1 < \theta < \min\limits_{1 \le i \le n} x_i \\ 0 & \text{otherwise} \end{cases}.$$

Consequently,

$$\frac{f(\mathbf{x} \mid \theta)}{f(\mathbf{y} \mid \theta)} = 1, \text{ if and only if } \max_{1 \le i \le n} x_i - 1 < \theta < \min_{1 \le i \le n} x_i \text{ and } \max_{1 \le i \le n} y_i - 1 < \theta < \min_{1 \le i \le n} y_i.$$

In other words, if and only if $\min\limits_{1 \le i \le n} x_i = \min\limits_{1 \le i \le n} y_i$ and $\max\limits_{1 \le i \le n} x_i = \max\limits_{1 \le i \le n} y_i$.

Therefore, $(X_{(1)}, X_{(n)})$ is a minimal sufficient statistic of θ.

Actually, $(X_{(n)} - X_{(1)}, (X_{(n)} + X_{(1)}) / 2)$ is also a minimal sufficient statistic of θ.

It has been shown (example 5.4.7) that $f_R(r \mid \theta) = n(n-1)r^{n-2}(1-r) = \text{Beta}(n-1, 2)$,

and thus the distribution of the range statistic, $f_R(r \mid \theta)$, is the same for all θ.

In other words, $f_R(r)$ is independent of the parameter θ.

Ancillary statistics

Ancillary statistics

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

Alone, an ancillary statistic contains **no** information about the parameter. When used in conjunction with other statistics, however, an ancillary statistic sometimes does contain valuable information for inferences about the parameter.

不含有待估参数的任何信息的统计量

均匀辅助统计量(I)

设 $X_1, \cdots, X_n \sim U(\theta, \theta+1)$, 那么极差R是辅助统计量. 其中极差定义为 $R = X_{(n)} - X_{(1)}$.

证明：

$$F(x|\theta) = \begin{cases} 0 & x \le \theta \\ 2-\theta & \theta < x \le \theta+1 \\ 1 & x \ge \theta+1 \end{cases}$$

根据次序统计量的密度公式$X_{(1)}$和$X_{(n)}$的联合密度

$$g(x_{(1)}, x_{(n)}) = \begin{cases} n(n-1)(x_{(n)} - x_{(1)})^{n-2} & \theta < x_{(1)} < \theta_{(n)} < \theta+1 \\ 0 & \text{other.} \end{cases}$$

均匀辅助统计量(II)

做坐标变换

$$\begin{cases} R = X_{(n)} - X_{(1)} \\ M = \frac{1}{2}(X_{(n)} + X_{(1)}) \end{cases} \qquad \begin{cases} X_{(1)} = \frac{1}{2}(2M - R) \\ X_{(n)} = \frac{1}{2}(2M + R) \end{cases}$$

得R和M的联合概率密度

$$h(r, m|\theta) = \begin{cases} n(n-1)r^{n-2} & 0 < r < 1, \theta + (r/2) < m < \theta + 1 - (r/2) \\ 0 & \text{other} \end{cases}$$

积分得R的概率密度函数为

$$h(r|\theta) = \int_{\theta+1-(r/2)}^{\theta+(r/2)} n(n-1)r^{n-2} dm$$
$$= n(n-1)r^{n-2}(1-r), 0 < r < 1$$

均匀辅助统计量(III)

- 上面的推导说明了极差统计量是辅助统计量。

- 本质的原因是所考察的概率模型是位置参数模型，所考察的参数恰好是位置参数，与X_i的均匀分布并无关系。

Location family ancillary statistic

Let $X_1,\ldots,X_n$ be iid random variables from a location parameter family with cdf $F(x-\theta),\ \ -\infty<\theta<\infty$. Let $Z_1=X_1-\theta,\ldots,Z_n=X_n-\theta$. We have that $Z_1,\ldots,Z_n$ are iid random variables from $F(x)$. Now, consider the range statistic $R=X_{(n)}-X_{(1)}$.

$$\begin{aligned}
F_R(r\mid\theta) &= P(R\le r\mid\theta)\\
&= P\left(\max_{1\le i\le n} X_i-\min_{1\le i\le n} X_i\le r\mid\theta\right)\\
&= P\left(\max_{1\le i\le n}(Z_i+\theta)-\min_{1\le i\le n}(Z_i+\theta)\le r\mid\theta\right)\\
&= P\left(\max_{1\le i\le n} Z_i-\min_{1\le i\le n} Z_i+\theta-\theta\le r\mid\theta\right)\\
&= P\left(\max_{1\le i\le n} Z_i-\min_{1\le i\le n} Z_i\le r\mid\theta\right)\\
&= P\left(\max_{1\le i\le n} Z_i-\min_{1\le i\le n} Z_i\le r\right)
\end{aligned}$$

The range statistic is an ancillary statistic for the location parameter.

Scale family ancillary statistic

Let $X_1,\ldots,X_n$ be iid random variables from a location parameter family with cdf $F(x/\sigma),\ \sigma>0$. Let $Z_1 = X_1/\sigma,\ldots,Z_n = X_n/\sigma$. We have that $Z_1,\ldots,Z_n$ are iid random variables from $F(x)$. Now, consider the statistic $T(\mathbf{X}) = (X_1/X_n,\ldots,X_{n-1}/X_n)$. Let $Y_i = X_i/X_n$. Then,

$$\begin{aligned}
F(y_1,\ldots,y_{n-1}\mid\sigma) &= P(Y_1 \le y_1,\ldots,Y_{n-1}\le y_{n-1}\mid\sigma)\\
&= P\left(X_1/X_n \le y_1,\ldots,X_{n-1}/X_n \le y_{n-1}\mid\sigma\right)\\
&= P\left((\sigma Z_1/\sigma Z_n)\le y_1,\ldots,(\sigma Z_{n-1}/\sigma Z_n)\le y_{n-1}\mid\sigma\right)\\
&= P\left(Z_1/Z_n \le y_1,\ldots,Z_{n-1}/Z_n \le y_{n-1}\mid\sigma\right)\\
&= P\left(Z_1/Z_n \le y_1,\ldots,Z_{n-1}/Z_n \le y_{n-1}\right)
\end{aligned}$$

Any statistic that depends on the sample only through the n-1 values X_1/X_n, …, $X_{n\text{-}1}/X_n$ is an ancillary statistic for the scale parameter.

Complete statistics

Let $f(t \mid \theta)$ be a family of pdfs or pmfs for a statistic $T(\mathbf{X})$. The family of probability distribution is called complete if $\mathrm{E}_\theta g(T) = 0$ for all θ implies $P_\theta(g(T) = 0) = 1$ for all θ. Equivalently, $T(\mathbf{X})$ is called a complete statistic.

Binomial complete sufficient statistic

Suppose that T has a binomial(n,p) distribution, $0<p<1$. Let $g(\cdot)$ be a function such that $\mathrm{E}_p g(T)=0$. Then,

$$\mathrm{E}_p g(T)=\sum_{t=0}^{n} g(t)\binom{n}{t}p^t(1-p)^{n-t}=(1-p)^n\sum_{t=0}^{n} g(t)\binom{n}{t}\left(\frac{p}{1-p}\right)^t$$

In order to ensure $\mathrm{E}_p g(T)=0$ for all $0<p<1$, $g(t)$ must be 0 for all t.

In other words,

$$P_p(g(T)=0)=1.$$

Therefore, T is a complete statistic.

The probability that $g(T)$=0 must be 1.

Exponential family

Let $X_1, \ldots, X_n$ be iid random variables from a pdf or pmf that belongs to an exponential family given by

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

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d), d \leq k$. Then,

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

is a complete statistic as long as the parameter space Θ contains an open set in $\Re^k$.

Open set

A subset U of the Euclidean n-space R^n is called open if, given any point $x \in U,$ there exists a real number $\varepsilon > 0$ such that, given any point y in R^n whose Euclidean distance from x is smaller than ε, y also belongs to U. Equivalently, U is open if every point in U has a neighbourhood contained in U.

The parameter space for a normal distribution $N(\mu, \sigma^2)$ is $(-\infty, \infty) \times (0, \infty)$, which is obvioiusly a open set.
The parameter space for a curved normal distribution $N(\theta, \theta^2)$ is a parabola, which does not contain a two-dimensional open set.

Basu's theorem

If $T(\mathbf{X})$ is a complelte and minimal sufficient statistic, then $T(\mathbf{X})$ is independent of every ancillary statistic.

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

Normal complete statistic

Let $X_1,\ldots,X_n$ be iid random variables with common pdf $N(\mu,\sigma^2)$, where μ is unknown but σ^2 is known. Then,

$\bar{X}$ is a sufficient statistic for μ.

$\bar{X}$ is a minimal sufficient statistic for μ.

$\bar{X}$ is a complete statistic.

S^2 is an ancillary statistic for μ.

By Basu's theorem,

The complete and minimal sufficient statistic $\bar{X}$ is independent of the ancillary statistic S^2.

The Likelihood Principle

Let $X_1, \ldots, X_n$ be iid random variables from a Bernoulli (θ) population.

Then the joint pdf of $X_1, \ldots, X_n$ is

$$f(\mathbf{x} \mid \theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i} = \theta^{n_1} (1-\theta)^{n-n_1},$$

where, $n_1 = \sum_{i=1}^{n} x_i$.

Now, we have two guesses, θ_1 and θ_2, about the true parameter θ.
Which one is more likely to be true?

Since $f(\mathbf{x} \mid \theta) = P(\mathbf{X} = \mathbf{x} \mid \theta)$, we may like to compare the two probabilities

$$f(\mathbf{x} \mid \theta_1) \text{ vs. } f(\mathbf{x} \mid \theta_2).$$

If $f(\mathbf{x} \mid \theta_1) > f(\mathbf{x} \mid \theta_2)$, θ_1 is more likely to be true.

If $f(\mathbf{x} \mid \theta_1) = f(\mathbf{x} \mid \theta_2)$, θ_1 and p_2 are equally likely to be true.

If $f(\mathbf{x} \mid \theta_1) < f(\mathbf{x} \mid \theta_2)$, θ_2 is more likely to be true.

The Likelihood Function

Let $f(\mathbf{x} \mid \theta)$ denote the joint pmf or pdf of the sample $\mathbf{X} = (X_1, \dots X_n)$. Then, given that $\mathbf{X} = \mathbf{x}$ is observed, the function of θ defined by

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

is called the **likelihood function**.

The likelihood function measures the plausibility that the sample is observed under a certain parameter. Larger likelihood means the sample that we observed is more likely to have occurred due to the given parameter.

Bernoulli Likelihood Function

Let $X_1, \ldots, X_n$ be iid random variables from a Bernoulli (θ) population. Then the joint pdf of $X_1, \ldots, X_n$ is

$$f(\mathbf{x} \mid \theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i} = \theta^{n_1} (1-\theta)^{n-n_1},$$

where $n_1 = \sum_{i=1}^{n} x_i$.

Therefore, the likelihood function for p is given by

$$L(\theta|\mathbf{x}) = \theta^{n_1} (1-\theta)^{n-n_1}.$$

In $f(\mathrm{x}|\theta)$, θ is fixed, and x is varying over all possible sample points.
In $L(\theta|\mathrm{x})$, however, x is fixed, and θ is varying over all possible parameter values.

Normal Likelihood Function

Let $X_1,\ldots,X_n$ be iid random variables from a normal population $N(\mu,\sigma^2)$, where σ^2 is already know and the only parameter is μ. Then the joint pdf of $X_1,\ldots,X_n$ is

$$f(\mathbf{x} \mid \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right],$$

Therefore, the likelihood function for μ is given by

$$L(\mu|\mathbf{x}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right].$$

In $f(\mathrm{x}|\mu)$, μ is fixed, and x is varying over all possible sample points.
In $L(\mu|\mathrm{x})$, however, x is fixed, and μ is varying over all possible parameter values.

Normal Likelihood Function

Let $X_1,\ldots,X_n$ be iid random variables from a normal population $N(\mu,\sigma^2)$.

Then the joint pdf of $X_1,\ldots,X_n$ is

$$f(\mathbf{x} \mid \mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right],$$

Therefore, the likelihood function for (μ,σ^2) is given by

$$L(\mu,\sigma^2|\mathbf{x}) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right].$$

In $f(\mathrm{x}|\mu,\ \sigma^2)$, $(\mu,\ \sigma^2)$ is fixed, and x is varying over all possible sample points.
In $L(\mu,\ \sigma^2|\mathrm{x})$, however, x is fixed, and $(\mu,\ \sigma^2)$ is varying over all possible parameter values.

Binomial Likelihood Function

Let $X_1, \dots, X_n$ be iid random variables from a Bernoulli (p) population. From previoius results, we know that $Y = \sum_{i=1}^{n} X_i$ is a sufficient statistic of p, and more importantly,

$$Y \sim \text{binomial}(n, p).$$

Because the pdf of Y is

$$f(y \mid p) = \binom{n}{y} p^y (1-p)^{n-y},$$

the likelihood function for p is given by

$$L(p \mid y) = \binom{n}{y} p^y (1-p)^{n-y}.$$

Multinomial Likelihood Function

Let $X_1,\ldots,X_n$ be iid random variables from a multinomial trial population with cell probability $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_m)$, where $0<\theta_i<1$ and $\sum_{i=1}^m \theta_i = 1.$

Then $\mathbf{n} = (n_1,\ldots,n_m)$ has a multinomial distribution,

$$f(\mathbf{n} \mid \boldsymbol{\theta}) = \frac{(\sum_{i=1}^m n_i)!}{\prod_{i=1}^m n_i!} \prod_{i=1}^m \theta_i^{n_i},$$

where $n_i = \sum_{j=1}^n I(x_j = i), i = 1,\ldots,m.$

Therefore, the likelihood function for $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_m)$ is given by

$$L(\boldsymbol{\theta} \mid \mathbf{n}) = \frac{(\sum_{i=1}^m n_i)!}{\prod_{i=1}^m n_i!} \prod_{i=1}^m \theta_i^{n_i}.$$

Normal Likelihood Function

Let $X_1,\ldots,X_n$ be iid random variables from a normal population $N(\mu,\sigma^2)$, where σ^2 is already known and the only parameter is μ. From previoius results, we know that $\bar{X}$ is a sufficient statistic of μ, and more importantly, $\bar{X} \sim N(\mu,\sigma^2 / n)$.

Then the pdf of $\bar{X}$ is

$$f(\overline{x} \mid \mu) = \frac{1}{\sqrt{2\pi}\sigma / \sqrt{n}} \exp\left[-\frac{(\overline{x}-\mu)^2}{2\sigma^2 / n}\right],$$

and the likelihood function for μ is given by

$$L(\mu \mid \overline{x}) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{n(\overline{x}-\mu)^2}{2\sigma^2}\right].$$

Calculations of Likelihoods

Suppose that 1000 Bernoulli trials have been done, n=1000, n_1=500. Then the likelihood for p=0.5 is

$$0.5^{500}(1-0.5)^{1000-500} = 0.5^{1000} \approx 9.33\times 10^{-302}$$

Suppose that 800 observations have been obtained from a standard normal population, and their squares add up to 800. Then the likelihood for (μ, σ^2)=(0, 1) is

$$(2\pi)^{-400}e^{-400} \approx 5.35\times 10^{-320}\times 1.92\times 10^{-174} \approx 1.03\times 10^{-493}$$

Log Likelihoods

Let $X_1,\ldots,X_n$ be iid random variables from a Bernoulli (θ) population. Then the likelihood function for θ is

$$L(\theta|\mathbf{x}) = \theta^{n_1}(1-\theta)^{n-n_1}.$$

Therefore, the log likelihood is

$$l(\theta|\mathbf{x}) = \log L(\theta|\mathbf{x}) = n_1 \log\theta + (n-n_1)\log(1-\theta).$$

Let $X_1,\ldots,X_n$ be iid random variables from a normal (μ,σ^2) population. Then the likelihood function for (μ,σ^2) is

$$L(\mu,\sigma^2|\mathbf{x}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}\right].$$

Therefore, the log likelihood is

$$l(\mu,\sigma^2|\mathbf{x}) = \log L(\mu,\sigma^2|\mathbf{x}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}(\sigma^2)^{-1}\sum_{i=1}^n (x_i-\mu)^2.$$

Likelihood Ratio

Let $X_1,\ldots,X_n$ be iid random variables from a multinomial trial population with cell probability $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_m)$. Then the joint pdf of $X_1,\ldots,X_n$ is

$$L(\boldsymbol{\theta}|\mathbf{x}) = \prod_{j=1}^{m} \theta_j^{n_j} \text{ and } l(\boldsymbol{\theta}|\mathbf{x}) = L(\boldsymbol{\theta}|\mathbf{x})=\sum_{j=1}^{m} n_j \log \theta_j,$$

where $n_j = \sum_{i=1}^{n} I(x_i = j), j = 1,\ldots m.$

Suppose that we have two guesses for $\boldsymbol{\theta}$, say, $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$. Then,

$$\frac{L(\boldsymbol{\theta}^{(1)}|\mathbf{x})}{L(\boldsymbol{\theta}^{(2)}|\mathbf{x})} = \frac{\prod_{j=1}^{m}\left(\theta_j^{(1)}\right)^{n_j}}{\prod_{j=1}^{m}\left(\theta_j^{(2)}\right)^{n_j}} = \prod_{j=1}^{m}\left(\frac{\theta_j^{(1)}}{\theta_j^{(2)}}\right)^{n_j}.$$

Obvioiusly,

$$\log\frac{L(\boldsymbol{\theta}^{(1)}|\mathbf{x})}{L(\boldsymbol{\theta}^{(2)}|\mathbf{x})} = l(\boldsymbol{\theta}^{(1)}|\mathbf{x}) - l(\boldsymbol{\theta}^{(2)}|\mathbf{x})=\sum_{j=1}^{m} n_j\left(\log\theta_j^{(1)} - \log\theta_j^{(2)}\right).$$

Likelihood Ratio for Comparing Parameters

Intuitively, the likelihood ratio provides a means of measuring the goodness of $\theta^{(1)}$ and $\theta^{(2)}$.

If $L(\theta^{(1)}|\mathbf{x})/L(\theta^{(2)}|\mathbf{x}) > 1$, $\theta^{(1)}$ is more likely to be the true.

If $L(\theta^{(1)}|\mathbf{x})/L(\theta^{(2)}|\mathbf{x}) = 1$, $\theta^{(1)}$ and $\theta^{(2)}$ are equally likely to be true.

If $L(\theta^{(1)}|\mathbf{x})/L(\theta^{(2)}|\mathbf{x}) < 1$, $\theta^{(2)}$ is more likely to be the true.

But how about we have another sample point $\mathbf{y}$ instead of $\mathbf{x}$, in what condition we would have the same inference results?

The Likelihood Principle

If $\mathbf{x}$ and $\mathbf{y}$ are two sample points such that $L(\theta \mid \mathbf{x})$ is proportional to $L(\theta \mid \mathbf{y})$, that is, there exists a constant $C(\mathbf{x}, \mathbf{y})$ such that

$$L(\theta \mid \mathbf{x}) = C(\mathbf{x}, \mathbf{y}) L(\theta \mid \mathbf{y}) \quad \text{for all } \theta,$$

then the conclusions drawn from $\mathbf{x}$ and $\mathbf{y}$ should be identical.

$$\frac{L(\theta^{(1)} \mid \mathbf{x})}{L(\theta^{(2)} \mid \mathbf{x})} = \frac{C(\mathbf{x}, \mathbf{y}) L(\theta^{(1)} \mid \mathbf{y})}{C(\mathbf{x}, \mathbf{y}) L(\theta^{(2)} \mid \mathbf{y})} = \frac{L(\theta^{(1)} \mid \mathbf{y})}{L(\theta^{(2)} \mid \mathbf{y})}$$

Summary Statistics

数据的概括就是根据数据简约的原理，设计出描述统计量来描述试验数据。

我们处理的是样本的观测值！

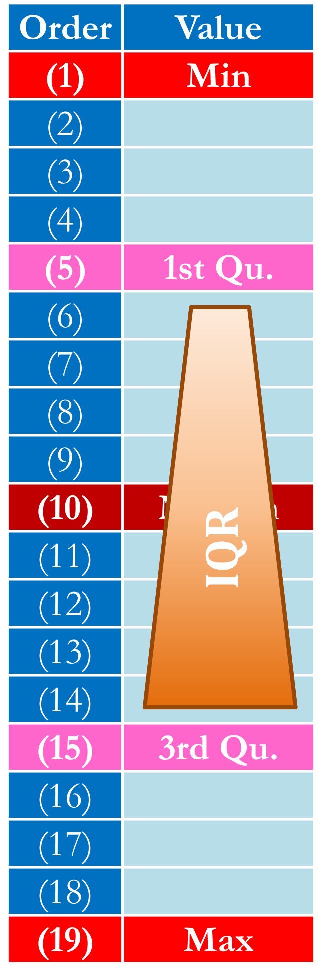
$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

Introductory statistics with R, page 57-80

Quantiles

- 四分位数 (Quartiles)
  - 1st quartile, Median, 3rd quartile
  - Interquartile range (IQR)
- 十分位数 (Centiles)
- 百分位数 (Percentiles)

```
> quantile(x)
> quantile(x, seq(0, 1, 0.1))
> quantile(x, seq(0, 1, 0.01))
> quantile(x, type=2)
```



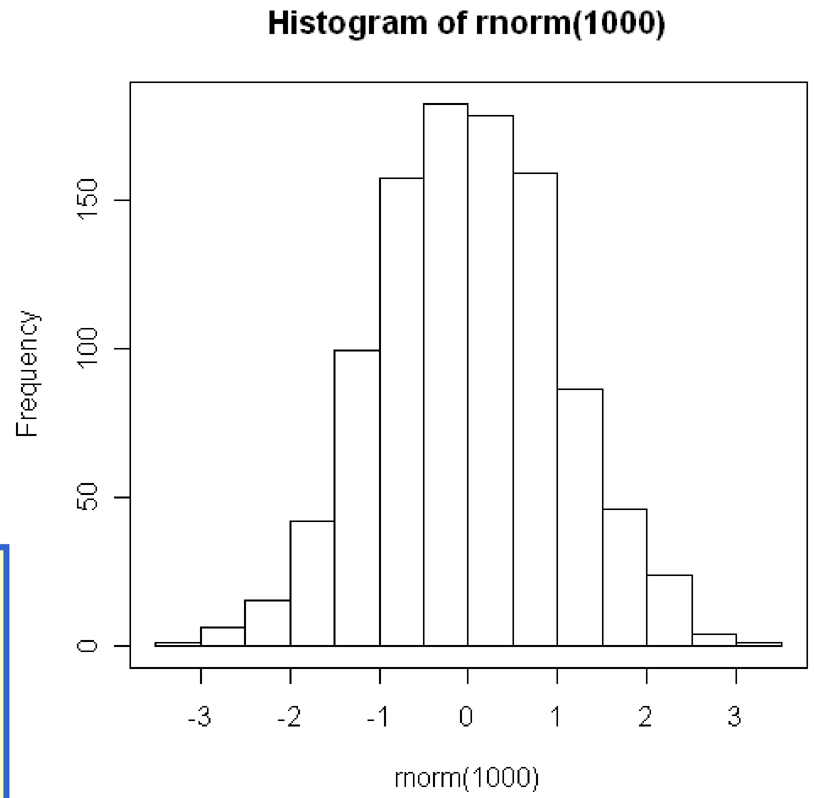
Summary Statistics

```
> fivenum(x)
> summary(x)
```

Index	Statistic
1	Min
2	1st Qu.
3	Median
4	Mean
5	3rd Qu.
6	Max
7	NAs

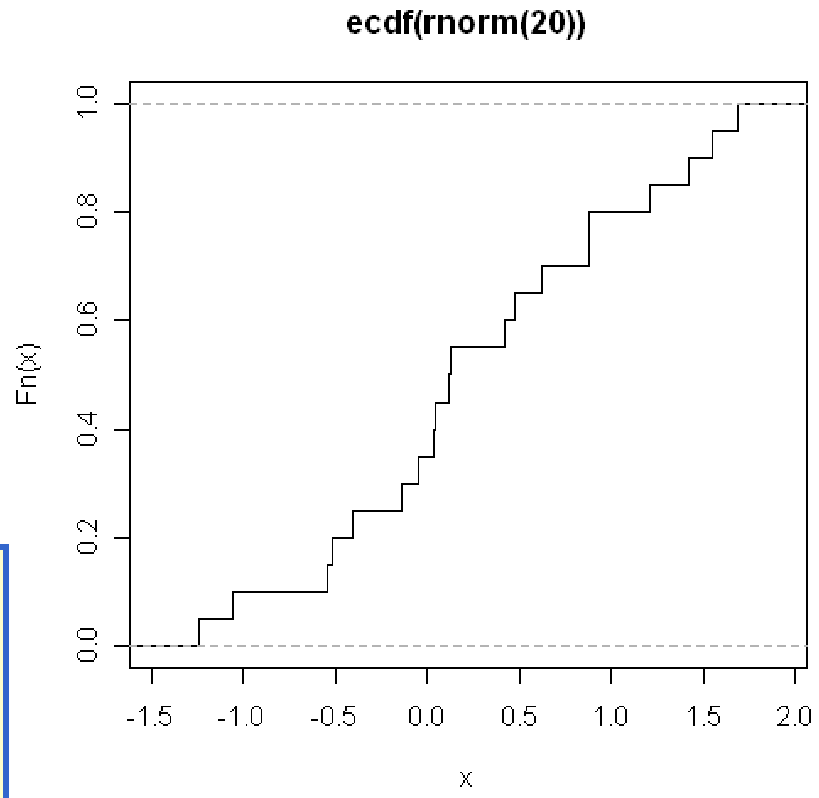
fivenum(): return Tukey's five number summary (minimum, lower-hinge, median, upper-hinge, maximum) for the input data

Histograms



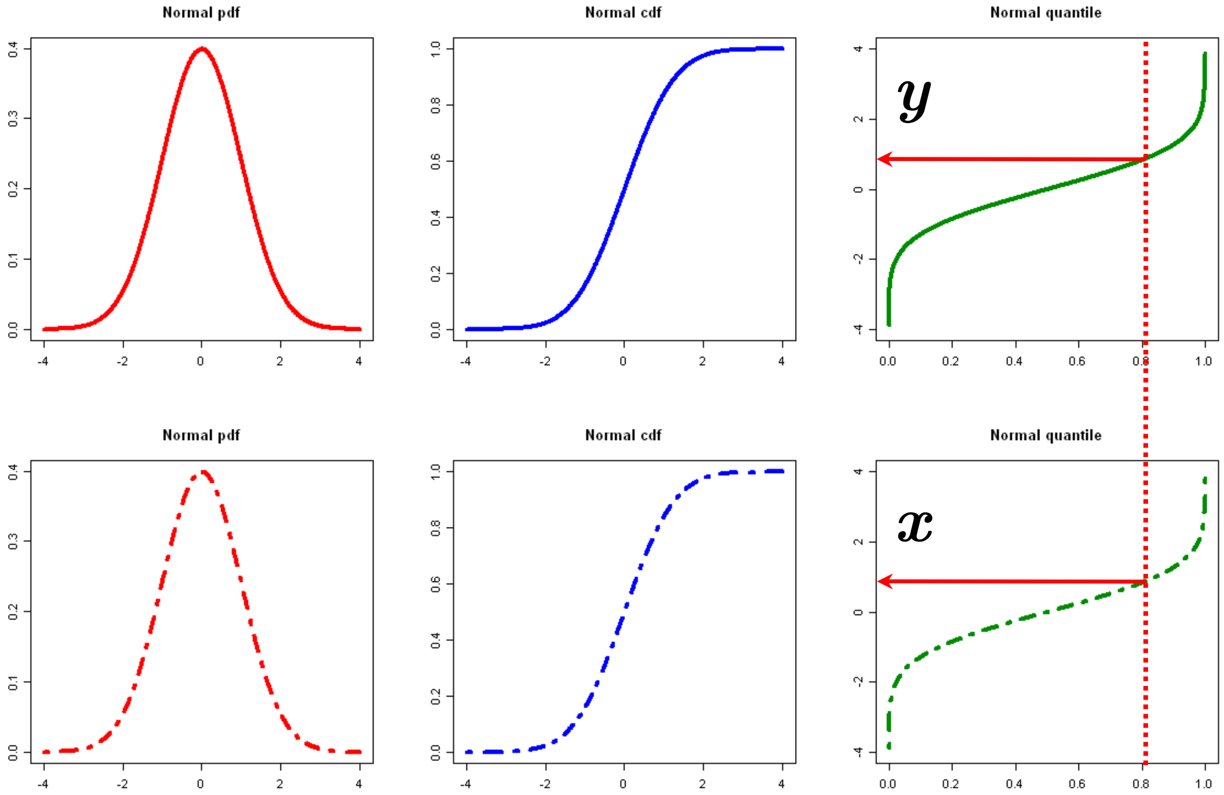

```
> hist(x)
> hist(x, freq=F)
> hist(x, freq=F, col="red")
> H <- hist(x)
```

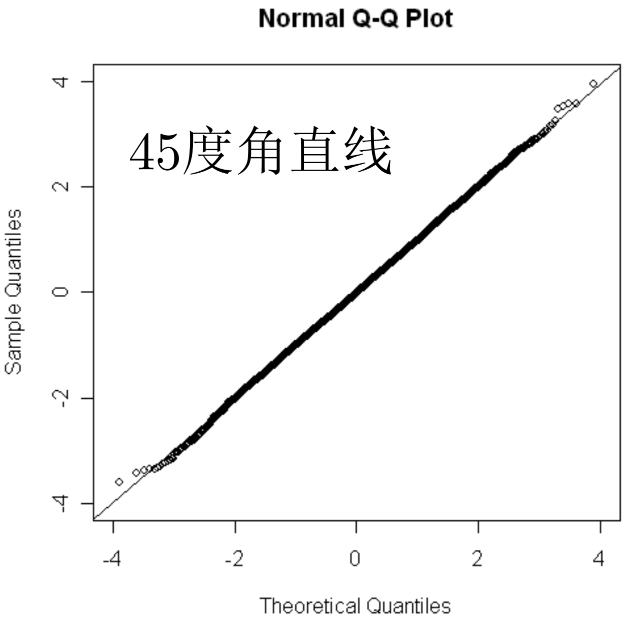
Empirical cdf (ecdf)



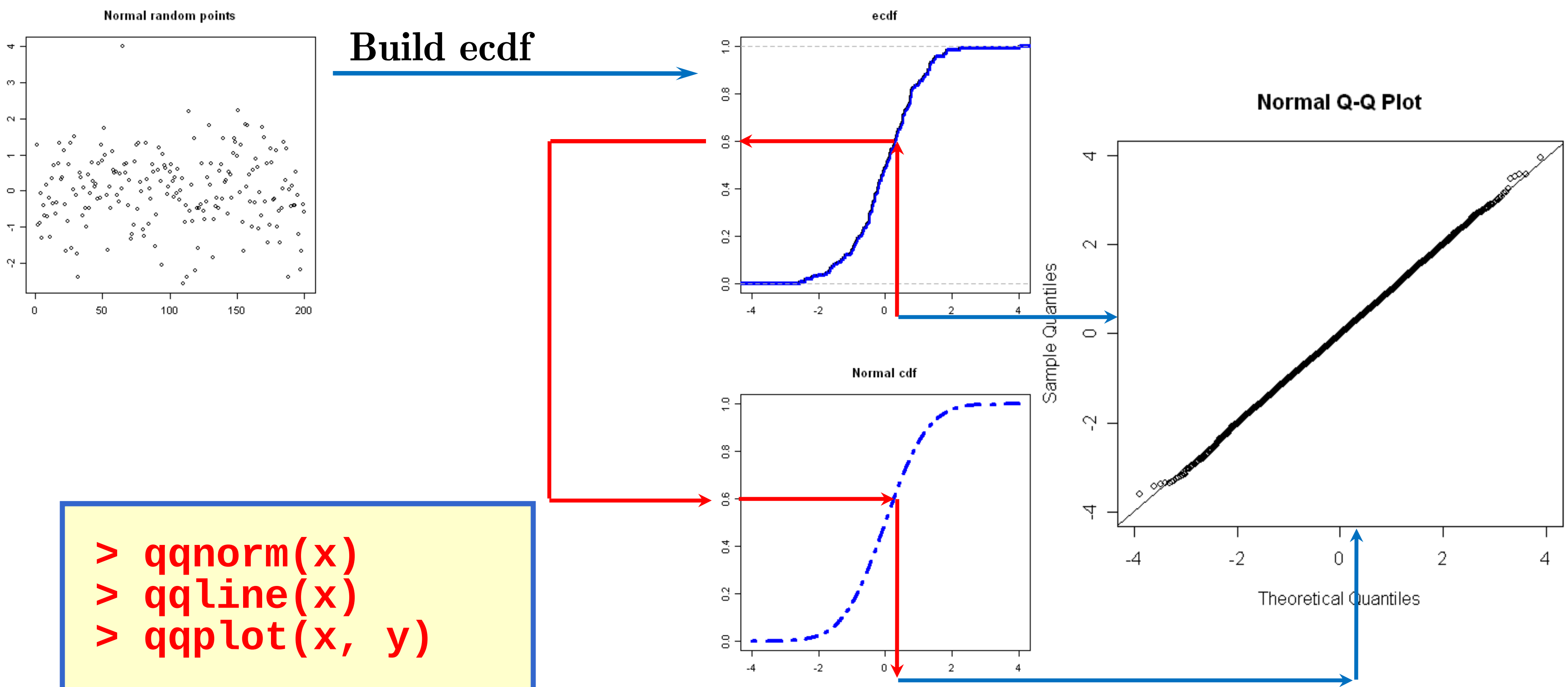

```
> F <- ecdf(x)
> plot(ecdf(x))
```

Q-Q plots


Normal pdf
Normal cdf
Normal quantile
y
Normal pdf
Normal cdf
Normal quantile
x


Normal Q-Q Plot
45度角直线
Sample Quantiles
Theoretical Quantiles

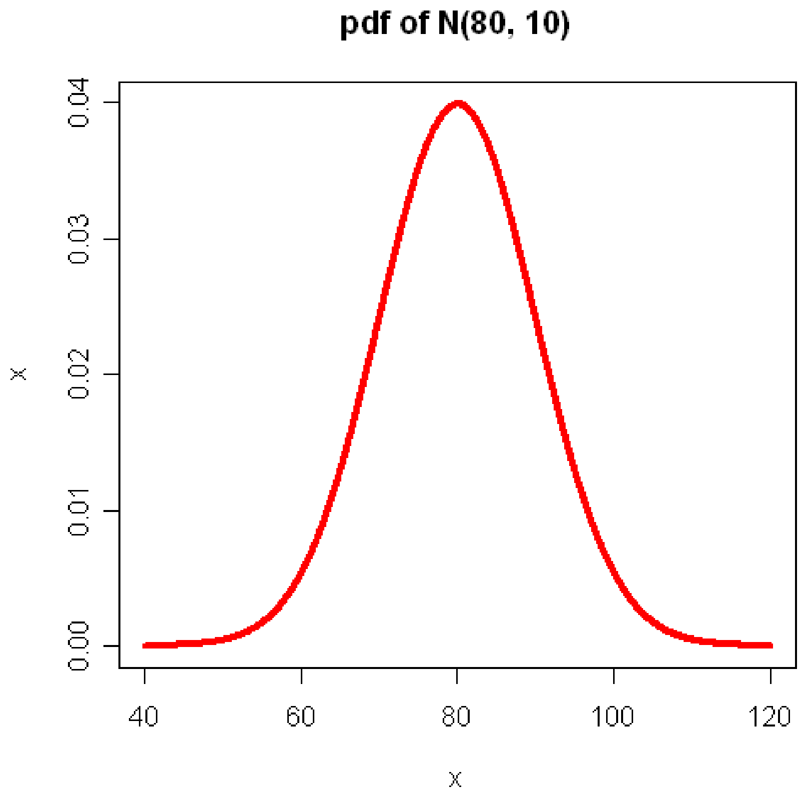
How to build a Q-Q plot

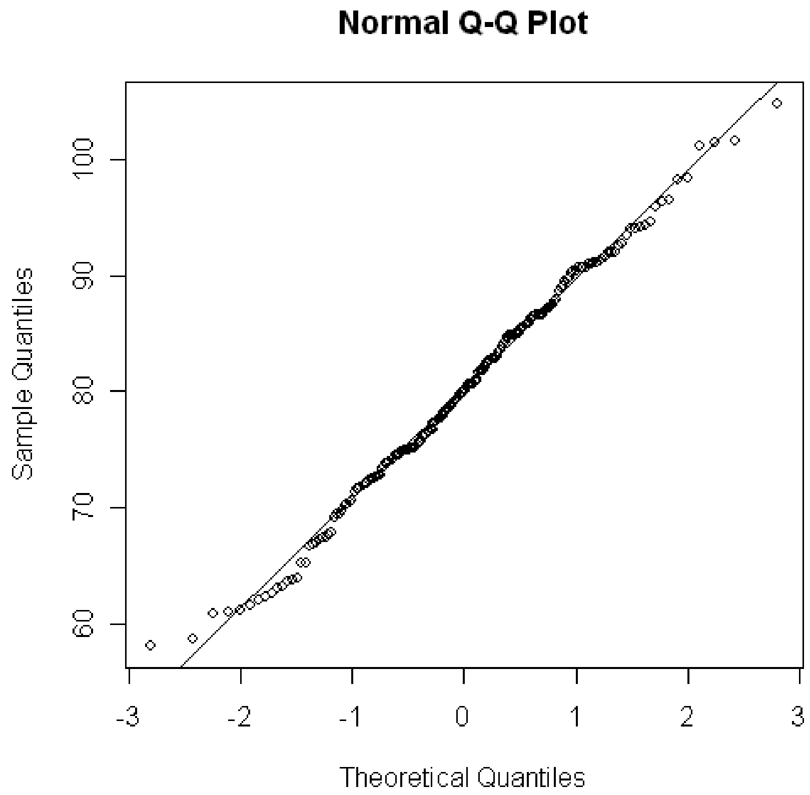



```
> qqnorm(x)
> qqline(x)
> qqplot(x, y)
```

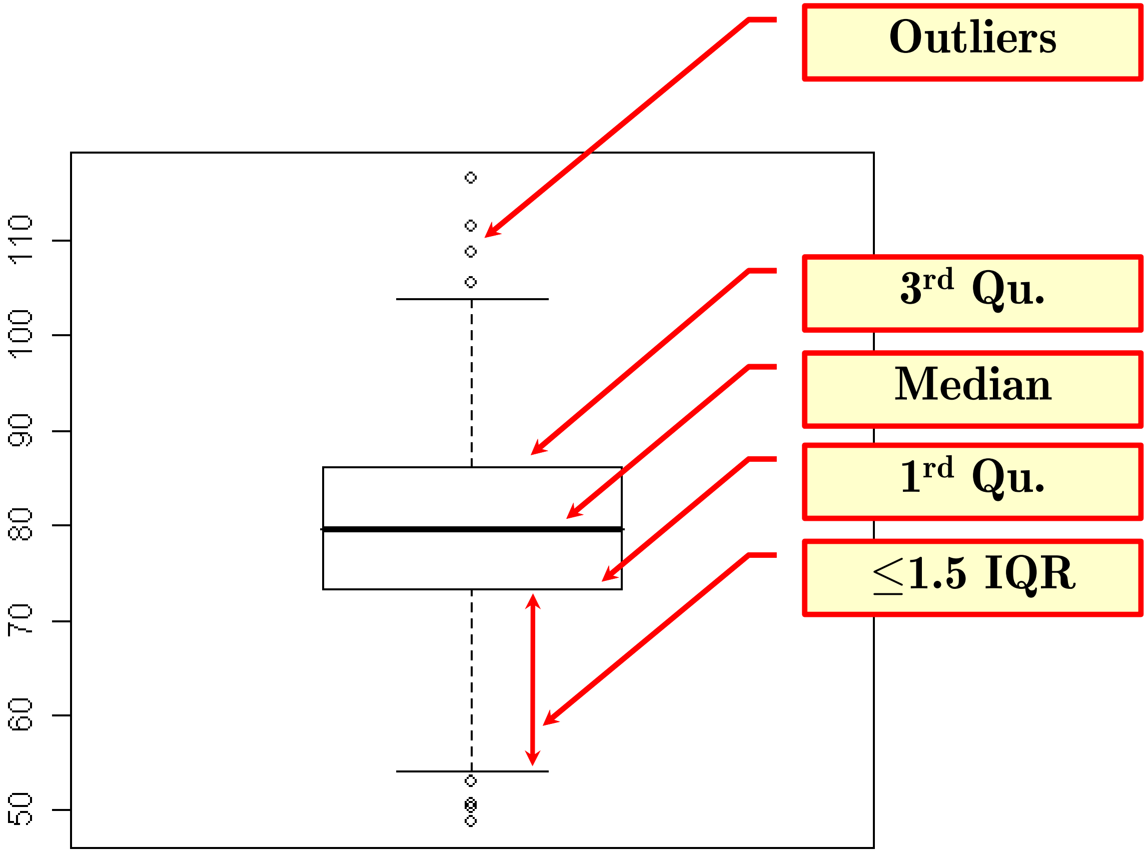
Digging into Q-Q plots — I

- Samples from normal populations





Box plots




```
> boxplot(x)
> boxplot.stats(x)
```