Mathematical Statistics and Data Analysis

Lecture 7: Statistics and their distributions

Lyu Ni

DaSE@ECNU
(lni@dase.ecnu.edu.cn)

October 25, 2019

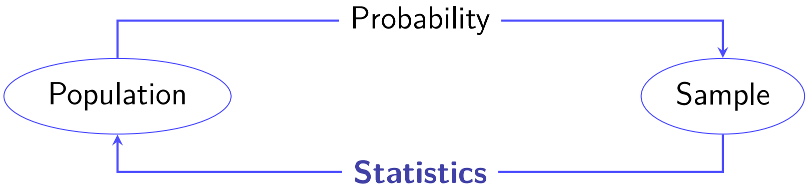
Outlines

Reading Material

Textbook:

- Rice: Chapter 3.7, 6, 7, 8.8, 10;
- Mao: Chapter 5;

Sample



Before	After

Observed

Consider x_i's as **Random Variables**	Consider x_i's as **Observed Values**

Sample

$x_1, x_2, \cdots, x_n$

Sample

Definition

The random variables $x_1, x_2, \cdots, x_n$ are called a **simple random sample** of size n from the population $F(x)$ if $x_1, x_2, \cdots, x_n$ are mutually independent random variables and the marginal c.d.f. of each X_i is the same function $F(x)$.

Remark

- $x_1, x_2, \cdots, x_n$ are independently and identically distributed. The joint c.d.f. of $(x_1, x_2, \cdots, x_n)$ is

$$F(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} F(x_i)$$

- $F(x)$ is also called **population distribution**.

The Empirical Cumulative Distribution Function

Question:

How to find the population distribution $F(x)$?

Definition

Suppose that $x_1, x_2, \cdots, x_n$ are a simple random sample. $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ is called the **ordered sample** if the sample are sorted from the smallest to the largest, that is,

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}.$$

The Empirical Cumulative Distribution Function

Question:

How to find the population distribution $F(x)$?

Definition

Suppose that $x_1, x_2, \cdots, x_n$ are a simple random sample. $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ is called the **ordered sample** if the sample are sorted from the smallest to the largest, that is,

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}.$$

The Empirical Cumulative Distribution Function

Definition

The **empirical cumulative distribution function (e.c.d.f.)** $F_n(x)$ is defined by

$$F_n(x) = \begin{cases} 0, & \text{if } x < x_{(1)}; \\ k/n, & \text{if } x_{(k)} \le x < x_{(k+1)}, k = 1, 2, \cdots, n-1; \\ 1, & \text{if } x \ge x_{(n)}; \end{cases}$$

Property

The e.c.d.f. $F_n(x)$ is a c.d.f., that is, $F_n(x)$ satisfies that

- $F_n(x)$ is non-decreasing and right-continuous;
- $F_n(-\infty) = 0$ and $F_n(\infty) = 1$;

The Empirical Cumulative Distribution Function

Example

- Aim of study: to investigate chemical methods for detecting the presence of synthetic waxes that had been added to beeswax.
- The addition of microcrystalline wax **raises the melting point of beeswax**.
- All pure beeswax had the same melting point;
- However, the melting point and other chemical properties of beeswax vary from one beehive to another.

The Empirical Cumulative Distribution Function

Example (Con'd)

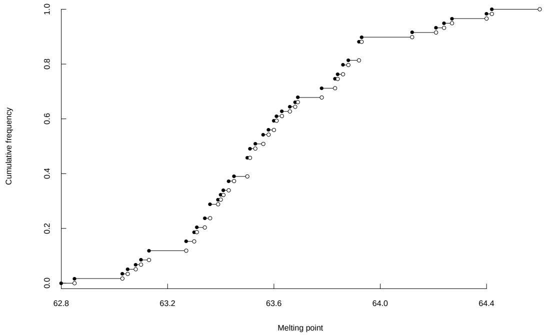
- Samples of pure beeswax are obtained from 59 sources.
- The 59 melting points (in $^\circ\mathrm{C}$) are listed as follows:

63.78	63.45	63.58	63.08	63.40	64.42	63.27	63.10
63.34	63.50	63.83	63.63	63.27	63.30	63.83	63.50
63.36	63.86	63.34	63.92	63.88	63.36	63.36	63.51
63.51	63.84	64.27	63.50	63.56	63.39	63.78	63.92
63.92	63.56	63.43	64.21	64.24	64.12	63.92	63.53
63.50	63.30	63.86	63.93	63.43	64.40	63.61	63.03
63.68	63.13	63.41	63.60	63.13	63.69	63.05	62.85
63.31	63.66	63.60					

The Empirical Cumulative Distribution Function

Example (Con'd)

- The e.c.d.f. is plotted as follows:



The Empirical Cumulative Distribution Function

$F_n(x)$ has another formula:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty,x)}(x_i)$$

where

$$I_{(-\infty,x)}(x_i) = \begin{cases} 1, & x_i \leq x; \\ 0, & x_i > x; \end{cases}$$

The random variables $I_{(-\infty,x)}(x_i)$ are independent Bernoulli random variables:

$$I_{(-\infty,x)}(x_i) = \begin{cases} 1, & \text{with probability } F(x) \\ 0, & \text{with probability } 1 - F(x); \end{cases}$$

The Empirical Cumulative Distribution Function

Thus, $nF_n(x)$ is a binomial random variable $b(n, F(x))$ and so

$$\begin{aligned} E(F_n(x)) &= F(x) \\ Var(F_n(x)) &= \frac{1}{n}F(x)(1-F(x)) \end{aligned}$$

Theorem

Suppose that $x_1, x_2, \cdots, x_n$ are a sample from a population c.d.f $F(x)$ and $F_n(x)$ is e.c.d.f. Then,

$$P\left(\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \to 0\right) = 1$$

as $n \to \infty$

Statistic

Definition

Suppose that $x_1, x_2, \cdots, x_n$ are a sample from an unknown population. A **statistic** T is defined by a function of the sample $T = T(x_1, x_2, \cdots, x_n)$ without any unknown parameters.

Remark:

- Statistics: $\sum_{i=1}^{n} x_i$, $\sum_{i=1}^{n} x_i^2$ and $F_n(x)$;
- A statistic does not depend on unknown parameters;
- The distribution of the statistic often depend on unknown parameters;

Sample Mean

Definition

Let $x_1, x_2, \cdots, x_n$ be a sample. The **sample mean** $\bar{x}$ is defined as the arithmetic mean of a sample, i.e.

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Property

- $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$;
- $\bar{x} = \underset{c}{\operatorname{argmin}} \sum_{i=1}^{n} (x_i - c)^2$, where c is a constant;

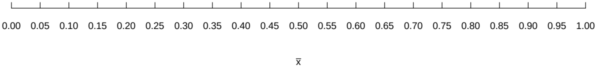
Sample Mean

Example

Suppose that $x_1, x_2, \cdots, x_{10}$ from a uniform distribution $U(0,1)$. At the i sampling, calculate the sample mean as

$$\bar{x}_i = \frac{\sum_{j=1}^{10} x_{i,j}}{10}, i = 1, 2, \cdots, 500.$$

What is the distribution of the sample mean?



Sample Mean

Theorem

Suppose that $\{x_i\}_{i=1}^n$ are a sample and $\bar{x}$ is the sample mean.

- If the population distribution is $N(\mu, \sigma^2)$, then the exact distribution of $\bar{x}$ is $N(\mu, \sigma^2/n)$;
- Suppose the population distribution is unknown. But $E(x) = \mu$ and $Var(x) = \sigma^2$. The asymptotic distribution of $\bar{x}$ is $N(\mu, \sigma^2/n)$. Denote $\bar{x} \dot{\sim} N(\mu, \sigma^2/n)$.

Proof:

- Since $\sum_{i=1}^n x_i \sim N(n\mu, n\sigma^2)$, we have

$$\bar{x} \sim N(\mu, \sigma^2/n).$$

- By CLT, $\sqrt{n}(\bar{x} - \mu)/\sigma \xrightarrow{L} N(0, 1)$. Thus, the asymptotic distribution of $\bar{x}$ is $N(\mu, \sigma^2/n)$.

Sample Variance

Definition

Suppose that $x_1, x_2, \cdots, x_n$ are a sample. The sample variance is defined by

$$s_*^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \text{ or } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Remark:

- s^2 is also called **unbiased variance**;
- The different formula for the sample variance is

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

Sample Variance

Theorem

Suppose that the population X has first- and second- order moment, that is, $E(X) = \mu$ and $Var(X) = \sigma^2 < \infty$. Let $x_1, x_2, \cdots, x_n$ be a sample from the population. $\bar{x}$ and s^2 are, respectively, the sample mean and sample variance. Then,

$$E(\bar{x}) = \mu, \quad Var(\bar{x}) = \sigma^2/n, \quad E(s^2) = \sigma^2.$$

Proof: It is obvious that

$$\begin{aligned} E(\bar{x}) &= \frac{1}{n} E\left(\sum_{i=1}^{n} x_i\right) = \frac{n\mu}{n} = \mu, \\ Var(\bar{x}) &= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} x_i\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \end{aligned}$$

Sample Variance

Theorem (Con'd)

We know

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\bar{x})^2 &= \sum_{i=1}^{n}(x_i^2-2\bar{x}x_i+\bar{x}^2)\\ &= \sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{x_i}+n\bar{x}^2=\sum_{i=1}^{n}x_i^2-n\bar{x}^2.\end{aligned}$$

Since $E(x_i^2)=Var(x_i)+\left(E(x_i)\right)^2=\sigma^2+\mu^2$ and $E(\bar{x}^2)=Var(\bar{x})+\left(E\bar{x}\right)^2=\sigma^2/n+\mu^2$, we have

$$E\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)=n(\mu^2+\sigma^2)-n(\mu^2+\sigma^2/n)=(n-1)\sigma^2.$$

Thus, $E(s^2)=\sigma^2$.

Sample Standard Deviation

Definition

Suppose that $x_1, x_2, \cdots, x_n$ are a sample. The **sample standard deviation** is defined by

$$s_* = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

or

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Sample Moment

Definition

Suppose that $x_1, x_2, \cdots, x_n$ are a sample.

- The k**th-order sample moment** is defined by

$$a_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k$$

  Particularly, $a_1 = \bar{x}$.

- The k**th-order sample central moment** is defined by

$$b_k = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

  Particularly, $b_2 = s_*^2$.

Sample Moment

Definition

Suppose that $x_1, x_2, \cdots, x_n$ are a sample.

- The **sample coefficient of skewness** is

$$\hat{\beta}_s = \frac{b_3}{b_2^{3/2}}$$

- The **sample kurtosis** is defined by

$$\hat{\beta}_k = \frac{b_4}{b_2^2} - 3$$

Order Statistics

Definition

Suppose that $x_1, \cdots, x_n$ are a sample. The **ith order statistic** is defined by $x_{(i)}$. Particularly,

- the **minimum statistic** is defined by $x_{(1)} = \min\{x_1, \cdots, x_n\}$;
- the **maximum statistic** is defined by $x_{(n)} = \max\{x_1, \cdots, x_n\}$.

Theorem

Suppose the p.d.f. is $f(x)$ and the c.d.f. is $F(x)$. Let $x_1, x_2, \cdots, x_n$ be a sample. Then the p.d.f. of the kth order statistic $x_{(k)}$ is

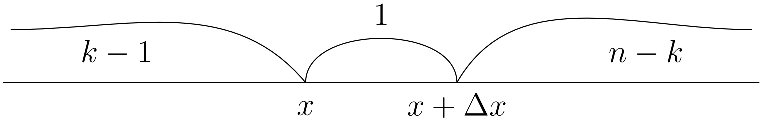
$$f_k(x) = \frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1}(1-F(x))^{n-k}f(x).$$

Proof: For any x, the event $x \le x_{(k)} \le x + \Delta x$ occurs.

Order Statistics

Theorem (Con'd)

This is equivalent to that $k-1$ observations are less than x, one observation is in the interval $[x, x+\Delta x]$, and $n-k$ observations are greater than $x+\Delta x$.



Then, for each $x_{(i)}$, we have

$$\begin{aligned}
P(x_{(i)} \le x) &= F(x) \\
P(x < x_{(i)} \le x + \Delta x) &= F(x+\Delta x) - F(x) \\
P(x_{(i)} > x + \Delta x) &= 1 - F(x+\Delta x)
\end{aligned}$$

Theorem (Con'd)

There are $\frac{n!}{(k-1)!1!(n-k)!}$ such arrangements. Let $F_k(x)$ be the c.d.f. of $x_{(k)}$. Thus, by the multinomial distribution,

$$F_k(x+\Delta x) - F_k(x) \approx \frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1} \cdot (F(x+\Delta x) - F(x))(1 - F(x+\Delta x))^{n-k}$$

Both sides are divided by Δx, and let $\Delta x \to 0$, that is,

$$\begin{aligned} f_k(x) &= \lim_{\Delta x \to 0} \frac{F_k(x+\Delta x) - F_k(x)}{\Delta x} \\ &= \frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1} f(x)(1 - F(x))^{n-k}, \end{aligned}$$

where the non-zero intervals of $f_k(x)$ and $f(x)$ are the same.

Order Statistic

Remark:

- The p.d.f. of $x_{(1)}$ is

$$f_1(x) = n(1 - F(x))^{n-1} f(x);$$

- The p.d.f. of $x_{(n)}$ is

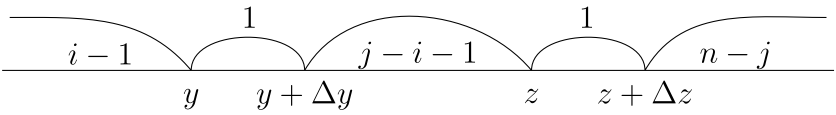
$$f_n(x) = n(F(x))^{n-1} f(x).$$

Order Statistic

Theorem

The p.d.f. of the order statistics $(x_{(i)}, x_{(j)})$ is

$$
\begin{aligned}
f_{i,j}(y,z) \quad = \quad & \frac{n!}{(i-1)!(j-i-1)!(n-j)!}(F(y))^{i-1} \\
& \cdot (F(z)-F(y))^{j-i-1}(1-F(z))^{n-j} f(y) f(z), y \leq z
\end{aligned}
$$



Example

Suppose that $x_1, x_2, \cdots, x_n$ are a sample from a uniform distribution $U(0,1)$. Then the p.d.f. of the kth order statistic is

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}, 0 < x < 1.$$

Thus, $x_{(k)} \sim Be(k, n-k+1)$ and $E(x_{(k)}) = \frac{k}{n+1}$.
The joint p.d.f. of $(Y,Z) = (x_{(1)}, x_{(n)})$ is

$$f(y,z) = n(n-1)(z-y)^{n-2}, 0 < y < z < 1.$$

Let $R = Z - Y$. Since $R > 0$ and $0 < Y < Z < 1$,

$$0 < Y = Z - R \leq 1 - R.$$

Order Statistic

Example

Suppose that $x_1, x_2, \cdots, x_n$ are a sample from a uniform distribution $U(0,1)$. Then the p.d.f. of the kth order statistic is

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}, 0 < x < 1.$$

Thus, $x_{(k)} \sim Be(k, n-k+1)$ and $E(x_{(k)}) = \frac{k}{n+1}$.
The joint p.d.f. of $(Y, Z) = (x_{(1)}, x_{(n)})$ is

$$f(y,z) = n(n-1)(z-y)^{n-2}, 0 < y < z < 1.$$

Let $R = Z - Y$. Since $R > 0$ and $0 < Y < Z < 1$,

$$0 < Y = Z - R \leq 1 - R.$$

Example (Con'd)

The joint p.d.f. of R is

$$f(y,r) = n(n-1)r^{n-2}, y > 0, r > 0, y + r < 1,$$

Then the marginal p.d.f. of R is

$$\begin{aligned} f(r) &= \int_0^{1-r} n(n-1)r^{n-2}\mathrm{d}y \\ &= n(n-1)r^{n-2}(1-r), 0 < r < 1 \end{aligned}$$

Thus, $R \sim Be(n-1, 2)$.

Sample Quantiles & Sample Median

Definition

Suppose that $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ are a ordered sample. The **pth sample quantile** is defined by

$$m_p = \begin{cases} x_{([np+1])}, & \text{if } np \text{ is not an integer;} \\ \frac{1}{2}(x_{(np)} + x_{(np+1)}), & \text{if } np \text{ is an integer;} \end{cases}$$

Particularly, the **sample median** is defined by

$$m_{0.5} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{if } n \text{ is odd;} \\ \frac{1}{2}\left(x_{\left(\frac{1}{2}\right)} + x_{\left(\frac{1}{2}+1\right)}\right), & \text{if } n \text{ is even;} \end{cases}$$

Sample Quantiles & Sample Median

Theorem

Suppose that the p.d.f. of a population is $f(x)$ and x_p is the pth sample quantile. $f(x)$ is continuous at the point $x = x_p$ and $f(x_p) > 0$. The asymptotic distribution of the pth sample quantile m_p is

$$m_p \dot{\sim} N\left(x_p, \frac{p(1-p)}{n \cdot f^2(x_p)}\right).$$

Particularly, the asymptotic distribution of the sample median is

$$m_{0.5} \dot{\sim} N\left(x_{0.5}, \frac{1}{4n \cdot f^2(x_{0.5})}\right)$$

Sample Quantiles & Sample Median

Example

The population distribution is Cauchy distribution. The p.d.f. is

$$f(x) = \frac{1}{\pi(1 + (x - \theta))^2}, -\infty < x < \infty$$

Then the c.d.f. is

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x - \theta)$$

It is obvious that θ is the median of the Cauchy distribution, that is, $x_{0.5} = \theta$. Let $x_1, x_2, \cdots, x_n$ be a sample. Then, the asymptotic distribution of the sample median is

$$m_{0.5} \dot{\sim} N(\theta, \frac{\pi^2}{4n}).$$

χ^2 Distributions

Review

The p.d.f. of Z is

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

Since $U = Z^2 \geq 0$, $F_U(u) = 0$ if $u \leq 0$. Thus, $f_U(u) = 0$ if $u \leq 0$. If $u > 0$, we have

$$\begin{aligned} F_U(u) &= P(U \leq u) = P(Z^2 \leq u) = P(-\sqrt{u} \leq Z \leq \sqrt{u}) \\ &= 2\Phi(\sqrt{y}) - 1 \end{aligned}$$

Then, the c.d.f. of U is

$$F_U(u) = \begin{cases} 2\Phi(\sqrt{y}) - 1, & y > 0, \\ 0, & y \leq 0. \end{cases}$$

Review (Con'd)

The p.d.f. of Y is

$$
\begin{aligned}
f_U(u) &= \begin{cases} \phi(\sqrt{y})y^{-1/2}, & y > 0, \\ 0, & y \leq 0, \end{cases} \\
&= \begin{cases} \dfrac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2}, & y > 0, \\ 0, & y \leq 0. \end{cases}
\end{aligned}
$$

Thus, $U \sim Ga(1/2, 1/2)$.

Definition

If Z is a standard normal r.v., the distribution of $U = Z^2$ is called **Chi-squared (χ^2)** distribution with 1 degree of freedom.

χ^2 Distributions

Review (Con'd)

The p.d.f. of Y is

$$
\begin{aligned}
f_U(u) &= \begin{cases} \phi(\sqrt{y})y^{-1/2}, & y > 0, \\ 0, & y \leq 0, \end{cases} \\
&= \begin{cases} \dfrac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2}, & y > 0, \\ 0, & y \leq 0. \end{cases}
\end{aligned}
$$

Thus, $U \sim Ga(1/2, 1/2)$.

Definition

If Z is a standard normal r.v., the distribution of $U = Z^2$ is called **Chi-squared (χ^2)** distribution with 1 degree of freedom.

χ^2 Distributions

Review

If $U_1 \sim Ga(\alpha_1, \lambda)$, $U_2 \sim Ga(\alpha_2, \lambda)$ and U_1 and U_2 are independent, then $V = U_1 + U_2 \sim Ga(\alpha_1 + \alpha_2, \lambda)$.
Since $V = U_1 + U_2 \geq 0$, the p.d.f. of V is $f_V(v) = 0$ if $v \leq 0$.
If $v > 0$, the p.d.f. of

$$\begin{aligned}
f_V(v) &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^z (z-y)^{\alpha_1-1} e^{-\lambda(z-y)} y^{\alpha_2-1} e^{-\lambda y} \mathrm{d}y \\
&= \frac{\lambda^{\alpha_1+\alpha_2} e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^z (z-y)^{\alpha_1-1} y^{\alpha_2-1} \mathrm{d}y \\
&= \frac{\lambda^{\alpha_1+\alpha_2} e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} z^{\alpha_1+\alpha_2-2} \int_0^z \left(1 - \frac{y}{z}\right)^{\alpha_1-1} \left(\frac{y}{z}\right)^{\alpha_2-1} \mathrm{d}y \\
&= \frac{\lambda^{\alpha_1+\alpha_2} e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} z^{\alpha_1+\alpha_2-1} \int_0^1 (1-t)^{\alpha_1-1} (t)^{\alpha_2-1} \mathrm{d}t
\end{aligned}$$

χ^2 Distributions

Review (Con'd)

$$
\begin{aligned}
f_V(v) &= \frac{\lambda^{\alpha_1+\alpha_2} e^{-\lambda z}}{\Gamma(\alpha_1+\alpha_2)} z^{\alpha_1+\alpha_2-1} \\
&\quad \cdot \int_0^1 \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} (1-t)^{\alpha_1-1} (t)^{\alpha_2-1} \, \mathrm{d}t \\
&= \frac{\lambda^{\alpha_1+\alpha_2} e^{-\lambda z}}{\Gamma(\alpha_1+\alpha_2)} z^{\alpha_1+\alpha_2-1}
\end{aligned}
$$

Thus, $V \sim Ga(\alpha_1+\alpha_2, \lambda)$.

- Z_i's are independently and identically distributed Gamma random variables $Ga(\alpha_i, \lambda)$. Then, $\sum_{i=1}^n Z_i \sim Ga(\sum_{i=1}^n \alpha_i, \lambda)$.

χ^2 Distributions

Definition

If $Z_1, Z_2, \cdots, Z_n$ are independently and identically distributed standard normal r.v.s, then $Z_1^2 + Z_2^2 + \cdots + Z_n^2$ is distributed as **Chi-squared (χ^2)** distribution with n degrees of freedom.

Remarks

- In fact, $Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim Ga(n/2, 1/2)$.
- The χ^2 distribution is a special case of the Gamma distribution.
- Properties:

$$E(Z_1^2 + Z_2^2 + \cdots + Z_n^2) = n$$

  and

$$Var(Z_1^2 + Z_2^2 + \cdots + Z_n^2) = 2n.$$

χ^2 Distributions

Example

Suppose that $x_1, x_2, \cdots, x_n$ is a sample from a normal population $N(\mu, \sigma^2)$, where the expectation μ is known. What is the distribution of

$$T = \sum_{i=1}^{n} (x_i - \mu)^2.$$

Solution: Let $y_i = (x_i - \mu)/\sigma, i = 1, 2, \cdots, n$. Then $y_1, y_2, \cdots, y_n$ are independently and identically distributed random variables. The distribution of y_1 is $N(0, 1)$. From the definition,

$$\frac{T}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^2 = \sum_{i=1}^{n} y_i^2 \sim \chi^2(n).$$

χ^2 Distributions

Example (Con'd)

Then, the p.d.f. of T is

$$f_T(t) = \frac{1}{(2\sigma^2)^{n/2}\Gamma(n/2)} \exp\left\{-\frac{t}{2\sigma^2}\right\} t^{\frac{n}{2}-1}$$

So,

$$T \sim Ga\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right).$$

χ^2 Distributions

Theorem

Suppose that $x_1, x_2, \cdots, x_n$ is a sample from a normal distribution $N(\mu, \sigma^2)$. The sample mean and sample variance is respectively

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Then,

- $\bar{x}$ and s^2 are independent;
- $\bar{x} \sim N(\mu, \sigma^2/n)$;
- $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$.

χ^2 Distributions

Theorem (Con'd)

Proof: The joint p.d.f. of

$$\begin{aligned} f(x_1, x_2, \cdots, x_n) &= (2\pi\sigma^2)^{-n/2} \exp\left\{ -\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} \right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{\sum_{i=1}^{n} x_i^2 - 2\bar{x}n\mu + n\mu^2}{2\sigma^2} \right\} \end{aligned}$$

Let $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)'$.

Theorem (Con'd)

Proof:

$$A = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2\cdot 1}} & -\frac{1}{\sqrt{2\cdot 1}} & 0 & \cdots & 0; \\ \frac{1}{\sqrt{3\cdot 2}} & \frac{1}{\sqrt{3\cdot 2}} & -\frac{2}{\sqrt{3\cdot 2}} & \cdots & 0; \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{\sqrt{n\cdot(n-1)}} & \frac{1}{\sqrt{n\cdot(n-1)}} & \frac{1}{\sqrt{n\cdot(n-1)}} & \cdots & -\frac{n-1}{\sqrt{n\cdot(n-1)}} \end{pmatrix}$$

As we know, the matrix A is orthogonal. Let $\boldsymbol{y} = A\boldsymbol{x}$. The Jacobian determinant is 1. Then,

$$\bar{x} = \frac{1}{\sqrt{n}} y_1 \text{ and } \sum_{i=1}^{n} y_i^2 = \boldsymbol{y}'\boldsymbol{y} == \boldsymbol{x}'A'A\boldsymbol{x} = \sum_{i=1}^{n} x_i^2$$

χ^2 Distributions

Theorem (Con'd)

The joint p.d.f. of $y_1, y_2, \cdots, y_n$ is

$$\begin{aligned} f(y_1, y_2, \cdots, y_n) &= (2\pi\sigma)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^{n} y_i - 2\sqrt{n}y_1\mu + n\mu^2}{2\sigma^2}\right\} \\ &= (2\pi\sigma)^{-n/2} \exp\left\{-\frac{\sum_{i=2}^{n} y_i + (y_1 - \sqrt{n}\mu)^2}{2\sigma^2}\right\} \end{aligned}$$

Then, $y_1, y_2, \cdots, y_n$ are independent and are distributed as a normal distribution with the variance σ^2. Thus, the mean of $y_2, y_3, \cdots, y_n$ is 0 and the mean of y_1 is $\sqrt{n}\mu$.

χ^2 distribution

Theorem (Con'd)

Since

$$\begin{aligned}(n-1)s^2 &= \sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n} x_i^2 - (\sqrt{n}\bar{x})^2 \\ &= \sum_{i=1}^{n} y_1^2 - y_1^2 = \sum_{i=2}^{n} y_i^2.\end{aligned}$$

Then, $y_2, \cdots, y_n$ are independent and identically distributed. And X_i's are distribution $N(0,1)$. Therefore,

$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=2}^{n}\left(\frac{y_i}{\sigma}\right)^2 \sim \chi^2(n-1)$$

F distribution

Definition

Let U and V be independent Chi-square random variables with m and n degrees of freedom, respectively. The distribution of

$$F = \frac{U/m}{V/n}$$

is called the **F distribution** with m and n degrees of freedom and is denoted by $F_{m,n}$ or $F(m,n)$.

Proposition

The p.d.f. of F is given by

$$f(y) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{\frac{m}{2}} y^{\frac{m}{2}-1} \left(1 + \frac{m}{n}y\right)^{-\frac{m+n}{2}}, w > 0$$

F distribution

How to derive the p.d.f. of the F distribution?

First, we derive the p.d.f. of $Z = \frac{U}{V}$. Let the $f_U(u)$ and $f_V(v)$ be respectively the p.d.f. of U and V. Then, the p.d.f. of Z is

$$\begin{aligned}
f_Z(z) &= \int_0^\infty v f_U(zv) f_V(v) \mathrm{d}v \\
&= \frac{z^{\frac{m}{2}-1}}{\Gamma(m/2)\Gamma(n/2)\cdot 2^{\frac{m+n}{2}}} \int_0^\infty v^{\frac{m+n}{2}-1} e^{-\frac{v}{2}(1+z)} \mathrm{d}v \\
&= \frac{z^{\frac{m}{2}-1}}{\Gamma(m/2)\Gamma(n/2)\cdot 2^{\frac{m+n}{2}}} \frac{\Gamma((m+n)/2)}{((1+z)/2)^{\frac{m+n}{2}}} \\
&= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} z^{\frac{m}{2}-1}(1+z)^{-\frac{m+n}{2}}, z > 0
\end{aligned}$$

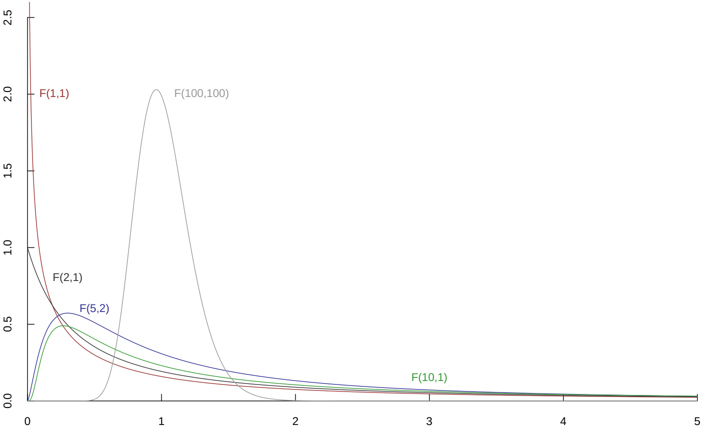
How to derive the p.d.f. of the F distribution? (Con'd)

Second, let $F = \frac{n}{m} Z$. For any $w > 0$, we have

$$
\begin{aligned}
f_F(y) &= p_Z\left(\frac{m}{n} y\right) \cdot \frac{m}{n} \\
&= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n} y\right)^{\frac{m}{2}-1} \left(1 + \left(\frac{m}{n} y\right)\right)^{-\frac{m+n}{2}} \cdot \frac{m}{n} \\
&= \frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} y^{\frac{m}{2}-1} \left(1 + \frac{m}{n} y\right)^{-\frac{m+n}{2}}
\end{aligned}
$$

F distribution

The p.d.f.s of F distribution are shown as follows:



F distribution

Proposition

Suppose that $x_1, x_2, \cdots, x_m$ is a sample from $N(\mu_1, \sigma_1^2)$ and $y_1, y_2, \cdots, y_n$ is a sample from $N(\mu_2, \sigma_2^2)$. Two samples are independent. Let

$$s_x^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2, \quad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

where $\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. Then

$$F = \frac{s_x^2/\sigma_1^2}{s_y^2/\sigma_2^2} \sim F(m-1, n-1).$$

Particularly, if $\sigma_1^2 = \sigma_2^2$, then $F = s_x^2/s_y^2 \sim F(m-1, n-1)$.

t distribution

Definition

If $Z \sim N(0,1)$ and $U \sim \chi^2_n$ and Z and U are independent, then the distribution of

$$t = \frac{Z}{\sqrt{U/n}}$$

is called the t **distribution** with n degrees of freedom.

How to derive the t distribution?

t distribution

How to derive the p.d.f. of the t distribution?

Z and $-Z$ are identically distributed for the p.d.f. of a standard normal distribution is symmetric. Then, t and $-t$ are also identically distributed. For any y,

$$P(0 < t < y) = P(0 < -t < y) = P(-y < -t < 0)$$

Thus,

$$P(0 < t < y) = \frac{1}{2} P(t^2 < y^2)$$

where

$$t^2 = \frac{Z^2}{U/n} \sim F(1, n).$$

How to derive the p.d.f. of the t distribution? (Con'd)

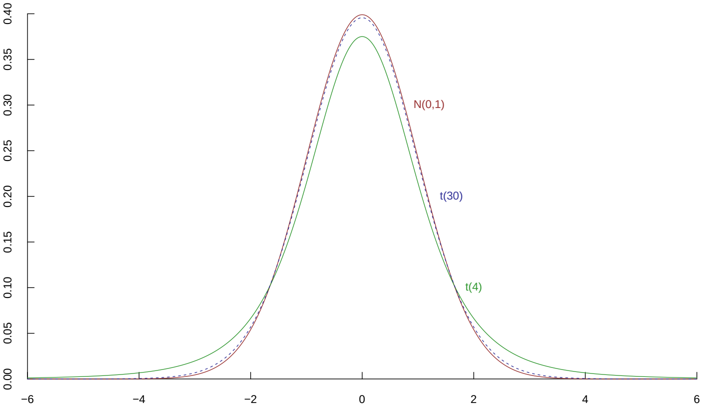
$$\begin{aligned} f_t(y) &= y f_F(y^2) = \frac{\Gamma\left(\frac{1+n}{2}\right)\left(\frac{1}{n}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}(y^2)^{\frac{1}{2}-1}\left(1+\frac{1}{n}y^2\right)^{-\frac{1+n}{2}} \cdot y \\ &= \frac{\Gamma\left(\frac{1+n}{2}\right)\left(\frac{1}{n}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{1}{n}y^2\right)^{-\frac{1+n}{2}}, -\infty < y < \infty \end{aligned}$$

Remark

- If $n = 1$, then it is a standard Cauchy distribution;
- If $n > 1$, then the expectation exists and equals 0;
- If $n > 2$, then the variance exists and equals $n/(n-2)$;
- If $n \geq 30$, then $N(0,1)$ can be used as an approximate distribution.

t distribution

The p.d.f.s of t distribution are shown as follows:



t distribution

Proposition

Suppose that $x_1, x_2, \cdots, x_n$ is a sample from a normal population $N(\mu, \sigma^2)$, and $\bar{x}$ and s^2 are respectively the sample mean and sample variance. Then

$$t = \frac{\sqrt{n}(\bar{x} - \mu)}{s} \sim t(n-1)$$

Proof: Since

$$\frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \sim N(0,1)$$

then

$$\frac{\sqrt{n}(\bar{x} - \mu)}{s} = \frac{\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}}{\sqrt{\frac{(n-1)s^2/\sigma^2}{n-1}}} \sim t(n-1)$$

t distribution

Background

Figure: W. S. Gosset

- Guinness Brewing Company;
- Mathematics and Chemistry;
- Measure how much yeast was in a given jar?
- Biometrika & Student;
- The Probable Error of the Mean (1908)

Proposition

Suppose that $x_1, x_2, \cdots, x_m$ is a sample from $N(\mu_1, \sigma_1^2)$ and $y_1, y_2, \cdots, y_n$ is a sample from $N(\mu_2, \sigma_2^2)$. Two sample are independent. In addition, suppose that $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Let

$$s_w^2 = \frac{(m-1)s_x^2 + (n-1)s_y^2}{m+n-2} = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2}{m+n-2}.$$

Then

$$\frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{s_w \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2)$$

Proof: As we know,

$$\frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{s_w \sqrt{\frac{1}{m} + \frac{1}{n}}} = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{m} + \frac{1}{n}}} \cdot \frac{1}{\sqrt{s_w^2/\sigma^2}}.$$

Proposition (Con'd)

$$\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}} = \boxed{\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}}} \cdot \frac{1}{\sqrt{s_w^2/\sigma^2}}.$$

- The 1st part on the RHS is a standard normal variable. Since $\bar{x} \sim N(\mu_1, \sigma^2/m)$, $\bar{y} \sim N(\mu_2, \sigma^2/n)$ and $\bar{x}$ and $\bar{y}$ are independent. Then,

$$\bar{x}-\bar{y} \sim N\left(\mu_1-\mu_2, \left(\frac{1}{m}+\frac{1}{n}\right)\sigma^2\right).$$

Proposition (Con'd)

$$\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}} = \frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}} \cdot \boxed{\frac{1}{\sqrt{s_w^2/\sigma^2}}}.$$

- s_w^2/σ^2 could be thought to be a χ^2 variable divided by its degree of freedom. Since $\frac{(m-1)s_x^2}{\sigma^2} \sim \chi^2(m-1)$, $\frac{(n-1)s_y^2}{\sigma^2} \sim \chi^2(n-1)$ and they are independent. Then,

$$\frac{(m+n-2)s_w^2}{\sigma^2} = \frac{(m-1)s_x^2+(n-1)s_y^2}{\sigma^2} \sim \chi^2(m+n-2)$$

Thus,

$$s_w^2/\sigma^2 = \frac{(m+n-2)s_w^2}{\sigma^2}/(m+n-2)$$

t distribution

Proposition (Con'd)

$$\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}} = \boxed{\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}}} \cdot \boxed{\frac{1}{\sqrt{s_w^2/\sigma^2}}}.$$

- Two parts on the RHS are independent. It is a fact that $\bar{x}$ and s_x^2 are independent, and $\bar{y}$ and $\boldsymbol{s}_y^2$ are independent. Two sample are independent. Then, $\bar{x}-\bar{y}$ and s_w^2 are independent.

Therefore, from the definition of t distribution,

$$\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t(m+n-2)$$

χ^2, F & t distribution

Remark

- If $F \sim F(m,n)$, then $\frac{1}{F} \sim F(n,m)$.
- If $t \sim t(n)$, then $t^2 \sim F(1,n)$.
- If $X \sim F_{m,n}$, then $\frac{(m/n)X}{1+(m/n)X} \sim Be(m/2, n/2)$.
- Suppose that $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_m$ are two independent samples from the standard normal population.

Distribution	Structure	Expectation	Variance
$\chi^2(n)$	$x_1^2 + x_2^2 + \cdots + x_n^2$	n	$2n$
$F(m,n)$	$\frac{y_1^2+y_2^2+\cdots+y_m^2}{x_1^2+x_2^2+\cdots+x_n^2}$	$\frac{n}{n-2}$ $(n>2)$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ $(n>4)$
$t(n)$	$\frac{y_1}{\sqrt{(x_1^2+x_2^2+\cdots+x_n^2)/n}}$	0 $(n>1)$	$\frac{n}{n-2}$ $(n>2)$

Sufficient Statistics

Background

Here is a practical problem: how to estimate σ?

Debate: Standard Deviation vs Mean Deviation

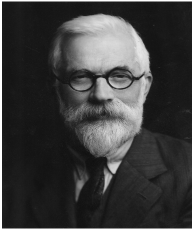


Figure: R. A. Fisher

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

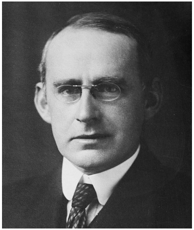


Figure: A. Eddington

$$d = \sqrt{\frac{\pi}{2}}\frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$$

Sufficient Statistics

Example

We would like to study the free throw percentage θ of a basketball player. Suppose the player attempted to shoot ten times from the foul line. He made eight free throws and he only missed the third and sixth shots. This includes two pieces of information:

- Out of 10 attempts, he made eight free throws;
- He missed the third and sixth shots;

Sufficient Statistics

Example

We would like to study the free throw percentage θ of a basketball player. Suppose the player attempted to shoot ten times from the foul line. He made eight free throws and he only missed the third and sixth shots. This includes two pieces of information:

- Out of 10 attempts, he made eight free throws; **Useful**
- He missed the third and sixth shots; **Useless**

For example,

- Result One : 1101101111;
- Result Two : 0011111111;

Sufficient Statistics

Suppose that $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ is a sample and the distribution of $\boldsymbol{x}$ is $F_\theta(\boldsymbol{x})$, which contains all the information about θ. Let $T = T(x_1, x_2, \cdots, x_n)$ be a statistic. The distribution of T is denoted as $F_\theta^T(t)$. The sample is expected to be replaced by the statistic T without loss of information. Equivalently, $F_\theta^T(t)$ contains all the information about θ as well as $F_\theta(\boldsymbol{x})$.
In other words, given the value of T,

- $F_\theta(\boldsymbol{x}|T = t)$ depends on θ;
- $F_\theta(\boldsymbol{x}|T = t)$ does not depend on θ;

The later statement means '**the statistic T contains all the information about θ**' Equivalently, we could envision keeping only T and throwing away all the $\boldsymbol{x}$ without any loss of information.

Sufficient Statistics

Definition

A statistic $T(x_1, x_2, \cdots, x_n)$ is said to be **sufficient** for θ if the conditional distribution of $x_1, x_2, \cdots, x_n$, given $T = t$, does not depend on θ for any value of t.

Example

Let $x_1, x_2, \cdots, x_n$ be a sequence of independent Bernoulli random variables with $P(x_i = 1) = \theta$. We will verify that $T = \sum_{i=1}^n x_i$ is sufficient for θ.
Given $T = t$, the conditional p.m.f. of $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ is

$$f(x_1, \cdots, x_n | T = t) = \frac{P(X_1 = x_1, \cdots, X_n = x_n, T = t)}{P(T = t)}$$

$$= \frac{P(X_1 = x_1, \cdots, X_{n-1} = x_{n-1}, X_n = t - \sum_{i=1}^{n-1} x_i)}{P(\sum_{i=1}^n X_i = t)}$$

Example (Con'd)

$$\geq\square\, f(x_1, \cdots, x_n | T = t) = \frac{\prod_{i=1}^{n-1} P(X_i = x_i) \cdot P(X_n = t - \sum_{i=1}^{n-1} x}{\binom{n}{t} \theta^t (1-\theta)^{n-t}}$$

$$= \frac{\prod_{i=1}^{n-1} \theta^{x_i} (1-\theta)^{1-x_i} \cdot \theta^{t - \sum x_i} (1-\theta)^{1-t+\sum x_i}}{\binom{n}{t} \theta^t (1-\theta)^{n-t}}$$

$$= \frac{\theta^{\sum x_i} (1-\theta)^{(n-1)-\sum x_i} \cdot \theta^{t-\sum x_i} (1-\theta)^{1-t+\sum x_i}}{\binom{n}{t} \theta^t (1-\theta)^{n-t}}$$

$$= \binom{n}{t}^{-1}$$

which does not depend on θ.

Example (Con'd)

Let $S = x_1 + x_2$ and $n > 2$. Given $S = s$, the condition p.m.f. of $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ is

$$
\begin{aligned}
& f(x_1, \cdots, x_n | S = s) = \frac{P(X_1 = x_1, \cdots, X_n = x_n, S = s)}{P(S = s)} \\
= & \frac{P(X_1 = x_1, X_2 = s - x_1, X_3 = x_3, \cdots, X_n = x_n)}{P(X_1 + X_2 = s)} \\
= & \frac{\theta^{s + \sum_{i=3}^n x_i} (1-\theta)^{n - s - \sum_{i=3}^n x_i}}{\binom{2}{s} \theta^s (1-\theta)^{2-s}} \\
= & \frac{\theta^{\sum_{i=3}^n x_i} (1-\theta)^{n-2-\sum_{i=3}^n x_i}}{\binom{2}{s}}, \text{ which depends on } \theta.
\end{aligned}
$$

Factorization Theorem

Theorem

Suppose that $f(x_1, x_2, \cdots, x_n; \theta)$ is the joint p.d.f. or p.m.f. of the sample $x_1, x_2, \cdots, x_n$. A necessary and sufficient condition for $T(x_1, x_2, \cdots, x_n)$ to be sufficient for a parameter θ is that the joint p.d.f. or p.m.f. factors in the form

$$f(x_1, x_2, \cdots x_n; \theta) = g(T(x_1, x_2, \cdots, x_n), \theta) \cdot h(x_1, x_2, \cdots, x_n)$$

Remark

- $g(t, \theta)$ depends on the statistic T and the parameter θ;
- $h(\cdot)$ does not depend on the parameter θ.

Factorization Theorem

Example: Uniform Distribution

Suppose that $x_1, x_2, \cdots, x_n$ is a sample from a uniform population $U(0, \theta)$. The p.d.f. is

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

Then, the joint p.d.f. of $(x_1, x_2, \cdots, x_n)$ is

$$f(x_1, x_2, \cdots, x_n; \theta) = \begin{cases} \left(\dfrac{1}{\theta}\right)^n, & 0 < \min\{x_i\} \le \max\{x_i\} < \theta, \\ 0, & \text{otherwise.} \end{cases}$$

Factorization Theorem

Example: Uniform Distribution (Con'd)

Since all the $x_i > 0$, the joint p.d.f. could be written as

$$f(x_1, x_2, \cdots, x_n; \theta) = (1/\theta)^n I\{x_{(n)} < \theta\}$$

Let

$$T = x_{(n)}, \quad g(t, \theta) = (1/\theta)^n I\{t < \theta\}$$

and

$$h(x_1, \cdots, x_n) = 1.$$

Then, $T = x_{(n)}$ is sufficient for θ.

Factorization Theorem

Example: Normal Distribution

Suppose that $x_1, x_2, \cdots, x_n$ is a sample from the normal distribution $N(\mu, \sigma^2)$ with two unknown parameters μ and σ^2. Let $\theta = (\mu, \sigma^2)$. The joint p.d.f. of $(x_1, x_2, \cdots, x_n)$ is

$$
\begin{aligned}
f(x_1, x_2, \cdots, x_n; \theta) &= (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\} \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{n\mu^2}{2\sigma^2} \right\} \\
&\quad \cdot \exp\left\{ -\frac{\sum_{i=1}^{n} x_i^2 - 2\mu \sum_{i=1}^{n} x_i}{2\sigma^2} \right\}
\end{aligned}
$$

Factorization Theorem

Example: Normal Distribution (Con'd)

Let $t_1 = \sum_{i=1}^{n} x_i$, $t_2 = \sum_{i=1}^{2} x_i^2$,

$$g(t_1, t_2, \theta) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\} \cdot \exp\left\{-\frac{1}{2\sigma^2}(t_2 - 2\mu t_1)\right\}$$

and $h(x_1, x_2, \cdots, x_n) = 1$. Thus, $T = (t_1, t_2) = (\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2)$ is sufficient for $\theta = (\mu, \sigma^2)$.

Factorization Theorem

Proof of Theorem

We give a proof for the discrete case. (The proof for the general case is beyond the scope of this course.)

Let $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ and $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$. The p.m.f. can be written as

$$f(\boldsymbol{x}; \theta) = P(\boldsymbol{X} = \boldsymbol{x}; \theta)$$

($\Rightarrow$) Suppose that T is a sufficient statistic.

- Given that $T = t$, the conditional probability function $P(\boldsymbol{X} = \boldsymbol{x} | T = t)$ does not depend on the parameter θ, which is denoted as $h(\boldsymbol{x})$.
- Let $A(t) = \{\boldsymbol{x} | T(\boldsymbol{x}) = t\}$. If $\boldsymbol{x} \in A(t)$,

$$\{T = t\} \supseteq \{\boldsymbol{X} = \boldsymbol{x}\}.$$

Factorization Theorem

Proof of Theorem (Con'd)

Thus,

$$\begin{aligned} P(\boldsymbol{X} = \boldsymbol{x}) &= P(\boldsymbol{X} = \boldsymbol{x}, T = t) \\ &= P(\boldsymbol{X} = \boldsymbol{x} | T = t) \cdot P(T = t) \\ &= h(\boldsymbol{x}) \cdot g(t, \theta) \end{aligned}$$

where $g(t, \theta) = P(T = t)$ and $h(\boldsymbol{x}) = P(X = x | T = t)$ does not depend on θ.

Factorization Theorem

Proof of Theorem (Con'd)

($\Leftarrow$) Suppose that

$$P(\boldsymbol{X} = \boldsymbol{x}; \theta) = g(t, \theta) \cdot h(\boldsymbol{x})$$

It is a fact that

$$\begin{aligned} P(T = t; \theta) &= \sum_{\{\boldsymbol{x}: T(\boldsymbol{x}) = t\}} P(\boldsymbol{X} = \boldsymbol{x}; \theta) \\ &= \sum_{\{\boldsymbol{x}: T(\boldsymbol{x}) = t\}} g(t, \theta) h(\boldsymbol{x}) \\ &= g(t, \theta) \sum_{\{\boldsymbol{x}: T(\boldsymbol{x}) = t\}} h(\boldsymbol{x}) \end{aligned}$$

Factorization Theorem

Proof of Theorem (Con'd)

For any t and $\boldsymbol{x} \in A(t)$, we have

$$
\begin{aligned}
P(\boldsymbol{X} = \boldsymbol{x} | T = t) &= \frac{P(\boldsymbol{X} = \boldsymbol{x}, T = t; \theta)}{P(T = t; \theta)} \\
&= \frac{P(\boldsymbol{X} = \boldsymbol{x}; \theta)}{P(T = t; \theta)} \\
&= \frac{g(t, \theta) h(\boldsymbol{x})}{g(t, \theta) \sum_{\{\boldsymbol{x}: T(\boldsymbol{x}) = t\}} h(\boldsymbol{x})} \\
&= \frac{h(\boldsymbol{x})}{\sum_{\{\boldsymbol{x}: T(\boldsymbol{x}) = t\}} h(\boldsymbol{x})}
\end{aligned}
$$

which does not depend on θ.

Factorization Theorem

Example: Normal Distribution (Revisit)

Suppose that $x_1, x_2, \cdots, x_n$ is a sample from the normal distribution $N(\mu, \sigma^2)$ with two unknown parameters μ and σ^2. Let $\theta = (\mu, \sigma^2)$. As we know,

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

is sufficient for $\theta = (\mu, \sigma^2)$. From the definition,

$$\begin{aligned} \bar{x} &= n^{-1} \sum_{i=1}^{n} x_i \\ s^2 &= (n-1)^{-1} \sum_{i=1}^{n} (x_i^2 - \bar{x})^2 \end{aligned}$$

Factorization Theorem

Example: Normal Distribution (Revisit)

The joint p.d.f. of $x_1, x_2, \cdots, x_n$ could be written as

$$f(x_1, x_2, \cdots, x_n; \theta) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{n(\bar{x}-\mu)^2 + (n-1)s^2}{2\sigma^2}\right\}$$

Thus, (μ, s^2) is also sufficient for $\theta = (\mu, \sigma^2)$.

Theorem

Suppose T is a sufficient statistic. If the statistic S is one-to-one corresponding to the statistic T, then S is also a sufficient statistic.

Exponential Family

Review

The p.d.f. or p.m.f. of a member of the exponential family is

$$f(x;\theta) = h(x)\exp\{\eta(\theta)^\tau T(x) - \zeta(\theta)\}\,.$$

Suppose that $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ is a sample from a member of the exponential family. The joint p.d.f. or p.m.f. of x is

$$\begin{aligned} f(\boldsymbol{x};\theta) &= \prod_{i=1}^n f(x_i;\theta) = \prod_{i=1}^n \left(h(x_i)\exp\{\eta(\theta)^\tau T(x_i) - \zeta(\theta)\}\right) \\ &= \prod_{i=1}^n h(x_i)\cdot\exp\left\{\eta(\theta)^\tau \sum_{i=1}^n T(x_i) - n\zeta(\theta)\right\} \end{aligned}$$

Thus, $\sum_{i=1}^n T(x_i)$ is a sufficient statistic.