Statistical Inference

Topic 2: Fundamentals of Statistics

Fall 2019

Instructor: Dr. Jiangdian Wang

清华大学统计学研究中心

Outline

Sampling Distribution

Exponential Family

Sufficient Statistics and Complete Statistics

Section 1. Sampling Distribution

Definition 1. The random variables $X_1, \cdots, X_n$ are called a random sample of size n from the population $f(x)$ if

- $X_1, \cdots, X_n$ are mutually independent random variables, and
- the marginal pdf or pmf of each X_i is the same function $f(x)$.

Alternatively, $X_1, \cdots, X_n$ are called independent and identically distributed random variables with pdf or pmf $f(x)$. Commonly abbreviated to i.i.d. random variables.

Definition 2. Let $X_1, \cdots, X_n$ be a random sample of size n from a population and let $T(x_1, \cdots, x_n)$ be a real-valued function whose domain includes the sample space of $(X_1, \cdots, X_n)$. Then the random variable or random vector $Y = \mathrm{T}(X_1, \cdots, X_n)$ is called a statistic. The probability distribution of a statistic Y is called the sampling distribution of Y.

Caution: The statistic CANNOT be a function of a parameter!

Definition 3. The sample mean is the arithmetic average of the values in a random sample. Denoted by

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Definition 4. The sample variance is the statistic defined by

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

Theorem 1. Let $x_1, \cdots, x_n$ be any numbers and $\bar{x} = (x_1 + \cdots + x_n)/n$. Then

a. $\min_a \sum_{i=1}^{n} (x_i - a)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2,$

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

Proof.

Theorem 2. Let $X_1, \cdots, X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Then

a. $\mathrm{E}(\bar{X}) = \mu$

b. $\mathrm{Var}(\bar{X}) = \sigma^2/n$

c. $\mathrm{E}(S^2) = \sigma^2$

The a. and c. are examples of unbiased statistics.

Proof.

Sum of Independent Normal Random Variables

Theorem 3. If $X_1, \cdots, X_n$ are mutually independent normal random variables with mean $\mu_1, \cdots, \mu_n$ and variances $\sigma_1^2, \cdots, \sigma_n^2$, then the linear combination

$$Y = \sum_{i=1}^{n} c_i X_i \sim N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right)$$

Proof.

Example 1: 2017年我国18岁及以上成年男性平均身高167.1cm。History also suggests that adult male height are normally distributed with a variance of 15cm. Select two male adults at random. Let X denote the first man's height, and let Y denote the second man's height. What is $P(X > Y)$?

Solution:

Sample Mean and Variance from Normal Distribution

Theorem 4. If $X_1, \cdots, X_n$ are independent random sample from a N(μ, σ^2) population, then

a. $\bar{X}$ and S^2 are independent random variables

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

c. $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$

Proof.

Important Distributions: t, χ^2, F

Definition 5 (Chi-Square distribution).

Let $X_1, X_2, \cdots, X_r$ *i.i.d.* $\sim N(0,1)$, then the distribution of the r.v.,

$$\xi = \sum_{i=1}^{r} X_i^2,$$

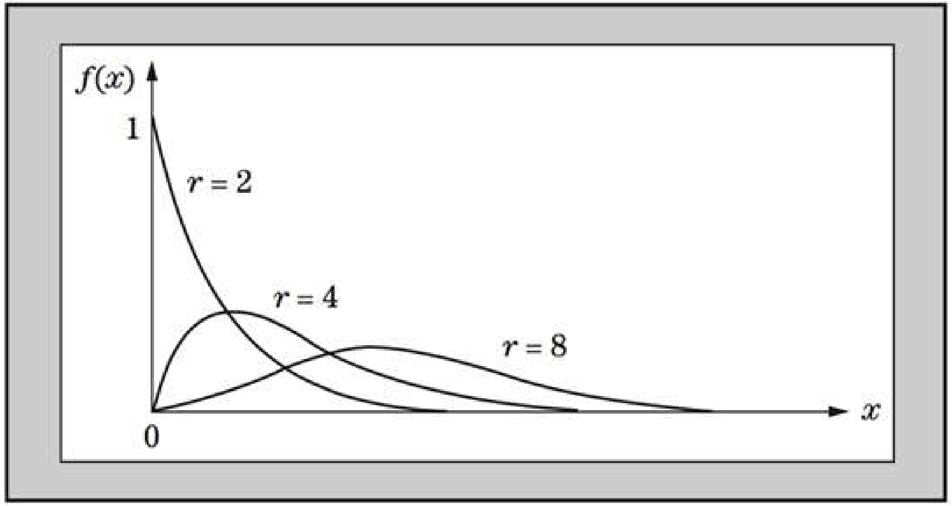
is known as the *Chi-Square* distribution with r *degrees of freedom*, and it is denoted as $\xi \sim \chi_r^2$.

Density

$$f(x) = \frac{1}{\Gamma\left(\frac{r}{2}\right) 2^{r/2}} x^{(r/2)-1} e^{-x/2}, \qquad x > 0, \qquad\qquad \Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\, dy, \qquad \alpha > 0.$$

Example: Sample variance of Normal distribution; likelihood ratio test statistics; Goodness-of-fit: Pearson χ^2 test (testing whether the sample is from a given distribution)

Chi-Square Distribution



FIGURE

Graph of the p.d.f. of the Chi-Square distribution for several values of r.

The support set (支撑集) of the p.d.f. of Chi-Square is $(0, +\infty)$
The larger n, the more symmetric the curve (asymptotic normal by CLT)

Properties of Chi-Square Distribution

Let $\xi \sim \chi_r^2$, then

(1) The c.f. of ξ is $\varphi(t) = (1-2it)^{-n/2}$;

(2) $E\xi = r$, and $Var(\xi) = 2r$;

(3) Let $\xi_1 \sim \chi^2_{r_1}$, $\xi_2 \sim \chi^2_{r_2}$ and ξ_1, ξ_2 are independent, then $\xi_1 + \xi_2 \sim \chi^2_{r_1+r_2}$.

(4) Let $\xi_i \sim \chi^2_{r_i}$, $i = 1, 2, \cdots, k$ and $\xi_1, \xi_2, \cdots, \xi_k$ are independent, then $\sum_{i=1}^k \xi_i \sim \chi^2_{r_1+\cdots+r_k}$.

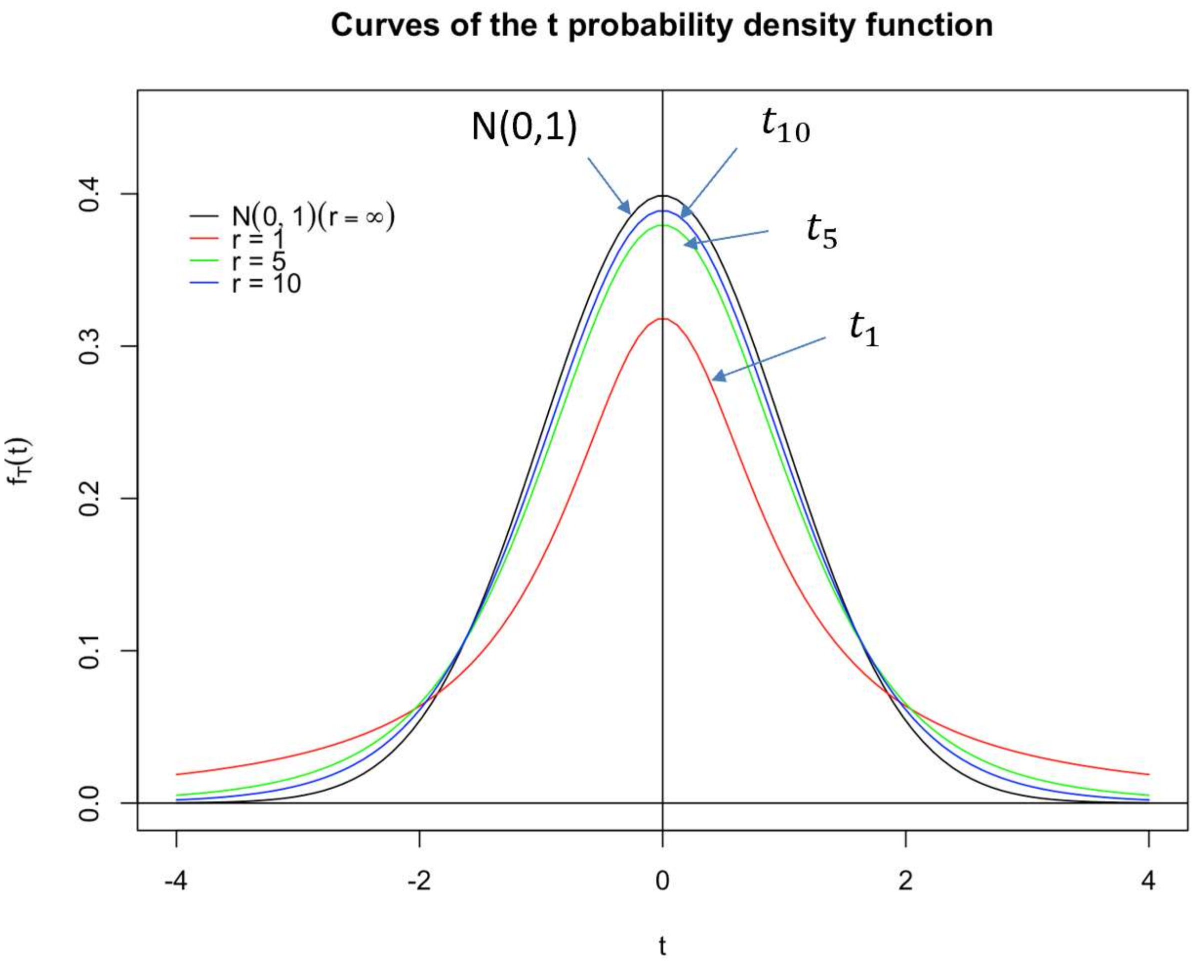
Student t Distribution

- W. S. Gosset (Student), 1908, also called Student's distribution (学生氏分布)

Definition 6 Let X and Y be two independent r.v.'s distributed as follows: $X \sim N(0,1)$ and $Y \sim \chi^2_r$, and define the r.v. T by: $T = \frac{X}{\sqrt{Y/r}}$. The r.v. T is said to have the (Student's) *t-distribution* with r *degrees of freedom*. The notation used is: $T \sim t_r$.

The p.d.f. of T, f_T, is given by the formula:

$$f_T(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\sqrt{\pi r}}\left(1+\frac{t^2}{r}\right)^{-(r+1)/2}, \quad t \in R$$



Curves of the t probability density function
N(0,1)
t_{10}
t_5
t_1
N(0, 1)(r = ∞)
r = 1
r = 5
r = 10
$f_T(t)$
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
t

Properties of t Distribution

(1) If r.v. $T \sim t_r$, then $E(T^k)$ exits only if $k < r$ $(r > 1)$ and

$$E(T^k) = \begin{cases} r^{\frac{k}{2}} \frac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{r-k}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\Gamma\left(\frac{1}{2}\right)}, & k \text{ is even,} \\ 0, & k \text{ is odd.} \end{cases}$$

In particular, when $r \geq 2$, $E(T) = 0$ and when $r \geq 3$, $Var(T) = r/(r-2)$;

(2) When $r = 1$, t_1 is the Cauchy distribution and its p.d.f. is

$$f_1(t) = \frac{1}{\pi(1+t^2)}, \quad -\infty < t < \infty;$$

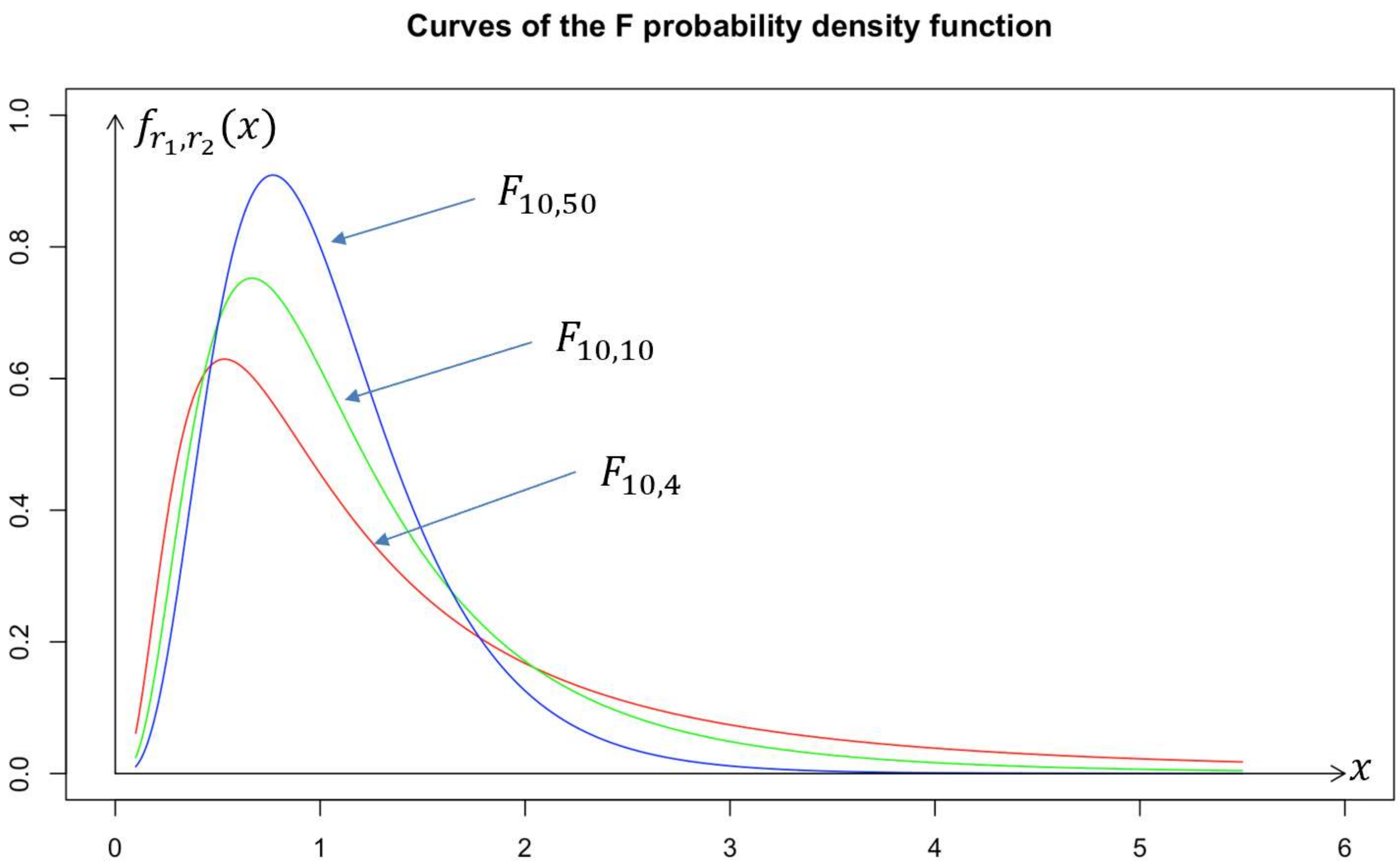
(3) As $r \to \infty$, t_r converges to $N(0, 1)$.

F Distribution

Definition 7 Let X and Y be two independent r.v.'s distributed as follows: $X \sim \chi^2_{r_1}$ and $Y \sim \chi^2_{r_2}$, and define the r.v. F by: $F = \frac{X/r_1}{Y/r_2}$. The r.v. F is said to have the *F-distribution with r_1 and r_2 degrees of freedom*. The notation often used is: $F \sim F_{r_1,r_2}$.

The p.d.f. of F, f_{r_1,r_2}, is given by the formula:

$$f_{r_1,r_2}(x) = \begin{cases} \frac{\Gamma\left(\frac{r_1+r_2}{2}\right)}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} \left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} x^{\frac{r_1}{2}-1} \left(1 + \frac{r_1}{r_2}x\right)^{-\frac{r_1+r_2}{2}}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$



Curves of the F probability density function
$f_{r_1,r_2}(x)$
$F_{10,50}$
$F_{10,10}$
$F_{10,4}$
x
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6

Properties of F Distribution

(1) If $F \sim F_{r_1,r_2}$, then $1/F \sim F_{r_2,r_1}$;

(2) If $T \sim t_r$, then $T^2 \sim F_{1,r}$;

(3) If $F \sim F_{r_1,r_2}$, then for $k > 0$ and $2k < r_2$,

$$E(F^k) = \left(\frac{r_2}{r_1}\right)^k \frac{\Gamma\left(\frac{r_1}{2}+k\right)\Gamma\left(\frac{r_2}{2}-k\right)}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)}.$$

In particular,

$$E(F) = \frac{r_2}{r_2-2},\ r_2 \geq 3, \quad \text{and} \quad Var(F) = \frac{2r_2^2(r_1+r_2-2)}{r_1(r_2-2)^2(r_2-4)},\ r_2 \geq 5;$$

(4) $F_{r_1,r_2}(1-\alpha) = 1/F_{r_2,r_1}(\alpha)$.

Section 2. Exponential Family

Definition 7 (Exponential family). Let $\mathcal{F} = \{f(x,\theta) : \theta \in \Theta\}$ is a distribution family defined on a sample space $\mathcal{X}$, where Θ is the parameter space. If the p.d.f. or p.m.f. $f(x,\theta)$ has the following form:

$$f(x,\theta) = C(\theta)\exp\left\{\sum_{i=1}^{k} Q_i(\theta)T_i(x)\right\} h(x),$$

where k is a positive integer, $C(\theta) > 0$ and $Q_i(\theta)$ are functions defined on the parameter space Θ, $h(x) > 0$ and $T_i(x)$ are functions defined on the sample space $\mathcal{X}$, then $\mathcal{F}$ is said to be *exponential family.*

Normal Distribution

Example 2. Let $\boldsymbol{X} = (X_1, \cdots, X_n)$ be a random sample from N(μ, σ^2), then the sample distribution family belongs to exponential family.

Remark. When $n = 1$, the p.d.f. of X_1 belongs to exponential family. $\{N(\mu, \sigma^2) : -\infty < \mu < +\infty, \sigma^2 > 0\}$ is exponential family and this does not depends on the sample size n.

Binomial Distribution

Example 3. Binomial distribution family $\{B(n, \theta)\}$ belongs to exponential family.

Example Distributions Not In Exponential Family

Example 4. Uniform on $[0, \theta], \theta > 0$ is not exponential family

$$f(x;\theta) = \frac{1}{\theta}, \quad x \in [0, \theta], \quad \theta > 0$$

Remark. The support set of exponential family does not depend on θ.

Example 5. The Cauchy distribution family is not exponential family

$$f(x;\theta) = \frac{1}{\pi[1 + (x - \theta)^2]}, \quad x \in R$$

Natural (Canonical) Form

$$f(x,\theta) = C(\theta)\exp\left\{\sum_{i=1}^{k} Q_i(\theta)T_i(x)\right\}h(x)$$

Let $\varphi_i = Q_i(\theta)$, transform $C(\theta)$ to $C^*(\varphi)$, $\varphi = (\varphi_1, \varphi_2, \cdots, \varphi_k)$, then change φ_i to θ_i

Definition 8. If the exponential family has the following form:

$$f(x,\theta) = C^*(\theta)\exp\left\{\sum_{i=1}^{k} \theta_i T_i(x)\right\}h(x),$$

then it is said to be the *natural (canonical) form*. The parameter space

$$\Theta^* = \left\{(\theta_1, \theta_2, \cdots, \theta_k) : \int_{\mathcal{X}} \exp\left\{\sum_{i=1}^{k} \theta_i T_i(x)\right\} h(x)dx < \infty\right\},$$

is said to be the *natural parametric space*.

Normal Distribution

Example 6. Normal distribution family

Binomial Distribution

Example 7. Binomial distribution family

Property of Exponential Family

1. All the distributions in the exponential family **does not depend on $\boldsymbol{\theta}$**.

2. The natural parametric space is a convex set.

Proof.

Property of Exponential Family

3. **Exponential Family preserved under transformations**. A smooth invertible transformation of a r.v. from the Exponential family is also within the Exponential family. If $X \to Y, Y = Y(X)$, then

$$\begin{aligned} f_Y(y;\theta) &= f_X(x(y);\theta)|\partial X/\partial Y| \\ &= C(\theta)\exp\left\{\sum_{i=1}^{k} Q_i(\theta)T_i(x(y))\right\} h(x(y))|\partial X/\partial Y| \end{aligned}$$

The Jacobian matrix $|\partial X/\partial Y|$ depends only on y, so $C(\theta), Q_i(\theta)$ do not change, while

$$T_i \to T_i(x(y))$$

$$h \to h(x(y))|\partial X/\partial Y|$$

Property of Exponential Family

4. Suppose there exists an inner point in the natural parametric space and let Θ_0 be the set of inner points. Let $g(x)$ be a real-valued function such that the following integration exits and is finite

$$G(\theta) = \int_{\mathcal{X}} g(x) \exp\left\{\sum_{i=1}^{k} \theta_i T_i(x)\right\} h(x) dx$$

Then $G(\theta)$ has any order partial derivatives in Θ_0 which are given by

$$\frac{\partial^m G(\theta)}{\partial\theta_1^{m_1}\cdots\partial\theta_k^{m_k}} = \int_{\mathcal{X}} \frac{\partial^m}{\partial\theta_1^{m_1}\cdots\partial\theta_k^{m_k}}\left[g(x)\exp\left\{\sum_{i=1}^{k}\theta_i T_i(x)\right\}h(x)\right]dx$$

Reference: 陈希孺. 数理统计引论. 北京: 科学出版社, 1981, 1998. (Theorem 1.2.1 on Page 21)

Property of Exponential Family

Property 4 can be used to **compute the expectation and covariance** of $T_i(X)$

Solution.

Since

$$\int_R f(x;\theta)dx = 1$$

we have

$$\int_R C(\theta)\exp\left\{\sum_{i=1}^k \theta_i T_i(x)\right\} h(x)dx = 1$$

Let $D(\theta) = \log(C(\theta))$, then $C(\theta) = \exp\{D(\theta)\}$

$$\int_R \exp\left\{\sum_{i=1}^k \theta_i T_i(x)\right\} h(x)dx = 1/C(\theta) = \exp\{-D(\theta)\}$$

$$\int_R \exp\left\{\sum_{i=1}^{k}\theta_i T_i(x)\right\} h(x)dx = 1/C(\theta) = \exp\{-D(\theta)\}$$

Differentiate with respect to θ_i

$$\int_R T_i(x)\exp\left\{\sum_{i=1}^{k}\theta_i T_i(x)\right\} h(x)dx = -\frac{\partial}{\partial\theta_i}D(\theta)\exp\{-D(\theta)\}$$

$$\int_R \underline{T_i(x)\exp\{D(\theta)\}\exp\left\{\sum_{i=1}^{k}\theta_i T_i(x)\right\} h(x)}dx = -\frac{\partial}{\partial\theta_i}D(\theta)$$

$$E[T_i(X)] = -\frac{\partial}{\partial\theta_i}D(\theta)$$

Exercise:

$$Var[T_i(X)] = -\frac{\partial^2}{\partial\theta_i^2}D(\theta); \quad Cov[T_i(X), T_j(X)] = -\frac{\partial^2}{\partial\theta_i\partial\theta_j}D(\theta)$$

Example 8 (Binomial Dist.) We already known that if $X \sim B(n, \theta)$, then $E(X) = n\theta$.

Solution.

$$f(x, \theta) = (1-\theta)^n \exp\left\{x \log \frac{\theta}{1-\theta}\right\} C_n^x$$

$$\varphi = \log[\theta/(1-\theta)]; \quad \theta = e^{\varphi}/(1+e^{\varphi})$$

$$f(x; \varphi) = (1+e^{\varphi})^{-n} \exp\{\varphi x\} C_n^x; \quad \Theta^* = \{\varphi : -\infty < \varphi < +\infty\}$$

Thus,

$$T_1(x) = x, \quad D(\varphi) = \log C(\varphi) = -n \log(1+e^{\varphi})$$

Therefore,

$$E(X) = E(T_1(X)) = -\frac{\partial}{\partial \varphi} D(\varphi) = n \frac{e^{\varphi}}{1+e^{\varphi}} = n\theta$$