

Stat 709: Mathematical Statistics

Lecture 6

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Lecture 6: p.d.f. and transformation

Example 1.12.

Let X be a random variable on (Ω, \mathcal{F}, P) whose c.d.f. F_X has a Lebesgue p.d.f. f_X and $F_X(c) < 1$, where c is a fixed constant.

Let $Y = \min\{X, c\}$, i.e., Y is the smaller of X and c .

Note that $Y^{-1}((-\infty, x]) = \Omega$ if $x \geq c$ and $Y^{-1}((-\infty, x]) = X^{-1}((-\infty, x])$ if $x < c$.

Hence Y is a random variable and the c.d.f. of Y is

$$F_Y(x) = \begin{cases} 1 & x \geq c \\ F_X(x) & x < c. \end{cases}$$

This c.d.f. is discontinuous at c , since $F_X(c) < 1$.

Thus, it does not have a Lebesgue p.d.f.

It is not discrete either.

Does P_Y , the probability measure corresponding to F_Y , have a p.d.f. w.r.t. some measure?

Example 1.12 (continued)

Consider the point mass probability measure on $(\mathcal{R}, \mathcal{B})$:

$$\delta_c(A) = \begin{cases} 1 & c \in A \\ 0 & c \notin A \end{cases} \quad A \in \mathcal{B}$$

Then $P_Y \ll m + \delta_c$, where m is the Lebesgue measure, and the p.d.f. of P_Y is

$$f_Y(x) = \frac{dP_Y}{d(m + \delta_c)}(x) = \begin{cases} 0 & x > c \\ 1 - F_X(c) & x = c \\ f_X(x) & x < c. \end{cases}$$

To show this, it suffices to show that

$$\int_{(-\infty, x]} f_Y(t) d(m + \delta_c) = P_Y((-\infty, x]) \quad \text{for any } x \in \mathcal{R}$$

(why?)

Example 1.12 (continued)

For $x < c$,

$$\begin{aligned}\int_{(-\infty, x]} f_Y(t) d(m + \delta_c) &= \int_{(-\infty, x]} f_X(t) dm + \int_{(-\infty, x]} f_X(t) \delta_c \\ &= \int_{(-\infty, x]} f_X(t) dm = P_X((-\infty, x]) = P_Y((-\infty, x])\end{aligned}$$

For $x \geq c$,

$$\begin{aligned}\int_{(-\infty, x]} f_Y(t) d(m + \delta_c) &= \int_{(-\infty, c]} f_Y(t) d(m + \delta_c) \\ &= \int_{(-\infty, c)} f_X(t) d(m + \delta_c) + \int_{\{c\}} [1 - F_X(c)] d(m + \delta_c) \\ &= \int_{(-\infty, c)} f_X(t) dm + \int_{\{c\}} [1 - F_X(c)] d\delta_c \\ &= F_X(c) + [1 - F_X(c)] = 1 = P_Y((-\infty, x])\end{aligned}$$

Example 1.14.

Let X be a random variable with c.d.f. F_X and Lebesgue p.d.f. f_X .

Let $Y = X^2$.

Since $Y^{-1}((-\infty, x])$ is empty if $x < 0$ and equals

$Y^{-1}([0, x]) = X^{-1}([-\sqrt{x}, \sqrt{x}])$ if $x \geq 0$, the c.d.f. of Y is

$$\begin{aligned}F_Y(x) &= P \circ Y^{-1}((-\infty, x]) \\ &= P \circ X^{-1}([-\sqrt{x}, \sqrt{x}]) \\ &= F_X(\sqrt{x}) - F_X(-\sqrt{x})\end{aligned}$$

if $x \geq 0$ and $F_Y(x) = 0$ if $x < 0$.

Clearly, the Lebesgue p.d.f. of F_Y is

$$f_Y(x) = \frac{1}{2\sqrt{x}} [f_X(\sqrt{x}) + f_X(-\sqrt{x})] I_{(0, \infty)}(x).$$

Example 1.14 (continued)

In particular, if

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

which is the Lebesgue p.d.f. of the standard normal distribution $N(0, 1)$, then

$$f_Y(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2} I_{(0, \infty)}(x),$$

which is the Lebesgue p.d.f. for the chi-square distribution χ_1^2 (Table 1.2).

This is actually an important result in statistics.

Proposition 1.8

Let X be a random k -vector with a Lebesgue p.d.f. f_X and let $Y = g(X)$, where g is a Borel function from $(\mathcal{R}^k, \mathcal{B}^k)$ to $(\mathcal{R}^k, \mathcal{B}^k)$. Let A_1, \dots, A_m be disjoint sets in \mathcal{B}^k such that $\mathcal{R}^k - (A_1 \cup \dots \cup A_m)$ has Lebesgue measure 0 and g on A_j is one-to-one with a nonvanishing Jacobian, i.e., the determinant $\text{Det}(\partial g(x)/\partial x) \neq 0$ on A_j , $j = 1, \dots, m$. Then Y has the following Lebesgue p.d.f.:

$$f_Y(x) = \sum_{j=1}^m |\text{Det}(\partial h_j(x)/\partial x)| f_X(h_j(x)),$$

where h_j is the inverse function of g on A_j , $j = 1, \dots, m$.

In Example 1.14, $A_1 = (-\infty, 0)$, $A_2 = (0, \infty)$, $g(x) = x^2$, $h_1(x) = -\sqrt{x}$, $h_2(x) = \sqrt{x}$, and $|dh_j(x)/dx| = 1/(2\sqrt{x})$.

Proposition 1.8

Let X be a random k -vector with a Lebesgue p.d.f. f_X and let $Y = g(X)$, where g is a Borel function from $(\mathcal{R}^k, \mathcal{B}^k)$ to $(\mathcal{R}^k, \mathcal{B}^k)$. Let A_1, \dots, A_m be disjoint sets in \mathcal{B}^k such that $\mathcal{R}^k - (A_1 \cup \dots \cup A_m)$ has Lebesgue measure 0 and g on A_j is one-to-one with a nonvanishing Jacobian, i.e., the determinant $\text{Det}(\partial g(x)/\partial x) \neq 0$ on A_j , $j = 1, \dots, m$. Then Y has the following Lebesgue p.d.f.:

$$f_Y(x) = \sum_{j=1}^m |\text{Det}(\partial h_j(x)/\partial x)| f_X(h_j(x)),$$

where h_j is the inverse function of g on A_j , $j = 1, \dots, m$.

In Example 1.14, $A_1 = (-\infty, 0)$, $A_2 = (0, \infty)$, $g(x) = x^2$, $h_1(x) = -\sqrt{x}$, $h_2(x) = \sqrt{x}$, and $|dh_j(x)/dx| = 1/(2\sqrt{x})$.

Example 1.15

Let $X = (X_1, X_2)$ be a random 2-vector having a joint Lebesgue p.d.f. f_X . Consider first the transformation $g(x) = (x_1, x_1 + x_2)$. Using Proposition 1.8, one can show that the joint p.d.f. of $g(X)$ is

$$f_{g(X)}(x_1, y) = f_X(x_1, y - x_1),$$

where $y = x_1 + x_2$ (note that the Jacobian equals 1). The marginal p.d.f. of $Y = X_1 + X_2$ is then

$$f_Y(y) = \int f_X(x_1, y - x_1) dx_1.$$

In particular, if X_1 and X_2 are independent, then

$$f_Y(y) = \int f_{X_1}(x_1) f_{X_2}(y - x_1) dx_1.$$

Example 1.15 (continued)

Next, consider the transformation $h(x_1, x_2) = (x_1/x_2, x_2)$, assuming that $X_2 \neq 0$ a.s.

Using Proposition 1.8, one can show that the joint p.d.f. of $h(X)$ is

$$f_{h(X)}(z, x_2) = |x_2| f_X(zx_2, x_2),$$

where $z = x_1/x_2$.

The marginal p.d.f. of $Z = X_1/X_2$ is

$$f_Z(z) = \int |x_2| f_X(zx_2, x_2) dx_2.$$

In particular, if X_1 and X_2 are independent, then

$$f_Z(z) = \int |x_2| f_{X_1}(zx_2) f_{X_2}(x_2) dx_2.$$

Example 1.16A (F-distribution)

Let X_1 and X_2 be independent random variables having the chi-square distributions $\chi_{n_1}^2$ and $\chi_{n_2}^2$ (Table 1.2), respectively.

The p.d.f. of $Z = X_1/X_2$ is

$$\begin{aligned} f_Z(z) &= \frac{z^{n_1/2-1} I_{(0,\infty)}(z)}{2^{(n_1+n_2)/2} \Gamma(n_1/2) \Gamma(n_2/2)} \int_0^\infty x_2^{(n_1+n_2)/2-1} e^{-(1+z)x_2/2} dx_2 \\ &= \frac{\Gamma[(n_1+n_2)/2]}{\Gamma(n_1/2) \Gamma(n_2/2)} \frac{z^{n_1/2-1}}{(1+z)^{(n_1+n_2)/2}} I_{(0,\infty)}(z) \end{aligned}$$

Using Proposition 1.8, one can show that the p.d.f. of

$$Y = (X_1/n_1)/(X_2/n_2) = (n_2/n_1)Z$$

is the p.d.f. of the F-distribution F_{n_1, n_2} given in Table 1.2.

Example 1.16B (t-distribution)

Let U_1 be a random variable having the standard normal distribution $N(0, 1)$ and U_2 a random variable having the chi-square distribution χ_n^2 . Using the same argument, one can show that if U_1 and U_2 are independent, then the distribution of $T = U_1 / \sqrt{U_2/n}$ is the t-distribution t_n given in Table 1.2.

Noncentral chi-square distribution

Let X_1, \dots, X_n be independent random variables and $X_i = N(\mu_i, \sigma^2)$. The distribution of $Y = (X_1^2 + \dots + X_n^2) / \sigma^2$ is called the *noncentral chi-square* distribution and denoted by $\chi_n^2(\delta)$, where $\delta = (\mu_1^2 + \dots + \mu_n^2) / \sigma^2$ is the noncentrality parameter. $\chi_k^2(\delta)$ with $\delta = 0$ is called a *central* chi-square distribution. It can be shown (exercise) that Y has the following Lebesgue p.d.f.:

$$e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} f_{2j+n}(x)$$

where $f_k(x)$ is the Lebesgue p.d.f. of the chi-square distribution χ_k^2 .

Example 1.16B (t-distribution)

Let U_1 be a random variable having the standard normal distribution $N(0, 1)$ and U_2 a random variable having the chi-square distribution χ_n^2 . Using the same argument, one can show that if U_1 and U_2 are independent, then the distribution of $T = U_1 / \sqrt{U_2/n}$ is the t-distribution t_n given in Table 1.2.

Noncentral chi-square distribution

Let X_1, \dots, X_n be independent random variables and $X_i = N(\mu_i, \sigma^2)$.

The distribution of $Y = (X_1^2 + \dots + X_n^2) / \sigma^2$ is called the *noncentral chi-square* distribution and denoted by $\chi_n^2(\delta)$, where

$\delta = (\mu_1^2 + \dots + \mu_n^2) / \sigma^2$ is the noncentrality parameter.

$\chi_k^2(\delta)$ with $\delta = 0$ is called a *central* chi-square distribution.

It can be shown (exercise) that Y has the following Lebesgue p.d.f.:

$$e^{-\delta/2} \sum_{j=0}^{\infty} \frac{(\delta/2)^j}{j!} f_{2j+n}(x)$$

where $f_k(x)$ is the Lebesgue p.d.f. of the chi-square distribution χ_k^2 .

Noncentral chi-square distribution

If Y_1, \dots, Y_k are independent random variables and Y_i has the noncentral chi-square distribution $\chi_{n_i}^2(\delta_i)$, $i = 1, \dots, k$, then $Y = Y_1 + \dots + Y_k$ has the noncentral chi-square distribution $\chi_{n_1 + \dots + n_k}^2(\delta_1 + \dots + \delta_k)$.

Noncentral t-distribution and F-distribution will be introduced in discussion session

Theorem 1.5. (Cochran's theorem)

Suppose that $X = N_n(\mu, I_n)$ and

$$X^T X = X^T A_1 X + \dots + X^T A_k X,$$

where I_n is the $n \times n$ identity matrix and A_i is an $n \times n$ symmetric matrix with rank n_i , $i = 1, \dots, k$.

A necessary and sufficient condition that $X^T A_i X$ has the noncentral chi-square distribution $\chi_{n_i}^2(\delta_i)$, $i = 1, \dots, k$, and $X^T A_i X$'s are independent is $n = n_1 + \dots + n_k$, in which case $\delta_i = \mu^T A_i \mu$ and $\delta_1 + \dots + \delta_k = \mu^T \mu$.

Noncentral chi-square distribution

If Y_1, \dots, Y_k are independent random variables and Y_i has the noncentral chi-square distribution $\chi_{n_i}^2(\delta_i)$, $i = 1, \dots, k$, then $Y = Y_1 + \dots + Y_k$ has the noncentral chi-square distribution $\chi_{n_1 + \dots + n_k}^2(\delta_1 + \dots + \delta_k)$.

Noncentral t-distribution and F-distribution will be introduced in discussion session

Theorem 1.5. (Cochran's theorem)

Suppose that $X = N_n(\mu, I_n)$ and

$$X^T X = X^T A_1 X + \dots + X^T A_k X,$$

where I_n is the $n \times n$ identity matrix and A_i is an $n \times n$ symmetric matrix with rank n_i , $i = 1, \dots, k$.

A necessary and sufficient condition that $X^T A_i X$ has the noncentral chi-square distribution $\chi_{n_i}^2(\delta_i)$, $i = 1, \dots, k$, and $X^T A_i X$'s are independent is $n = n_1 + \dots + n_k$, in which case $\delta_i = \mu^T A_i \mu$ and $\delta_1 + \dots + \delta_k = \mu^T \mu$.

Noncentral chi-square distribution

If Y_1, \dots, Y_k are independent random variables and Y_i has the noncentral chi-square distribution $\chi_{n_i}^2(\delta_i)$, $i = 1, \dots, k$, then $Y = Y_1 + \dots + Y_k$ has the noncentral chi-square distribution $\chi_{n_1 + \dots + n_k}^2(\delta_1 + \dots + \delta_k)$.

Noncentral t-distribution and F-distribution will be introduced in discussion session

Theorem 1.5. (Cochran's theorem)

Suppose that $X = N_n(\mu, I_n)$ and

$$X^T X = X^T A_1 X + \dots + X^T A_k X,$$

where I_n is the $n \times n$ identity matrix and A_i is an $n \times n$ symmetric matrix with rank n_i , $i = 1, \dots, k$.

A necessary and sufficient condition that $X^T A_i X$ has the noncentral chi-square distribution $\chi_{n_i}^2(\delta_i)$, $i = 1, \dots, k$, and $X^T A_i X$'s are independent is $n = n_1 + \dots + n_k$, in which case $\delta_i = \mu^T A_i \mu$ and $\delta_1 + \dots + \delta_k = \mu^T \mu$.