Lecture 3: Densities, moments, inequalities, and generating functions

Example 1.12.

Let X be a random variable on $(\Omega, \mathscr{F}, P)$ 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 $(\mathscr{R}, \mathscr{B})$:

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

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

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

(why?)

Example 1.12 (continued)

For $x < c$,

$$\int_{(-\infty,x]} f_Y(t) d(m+\delta_c) = \int_{(-\infty,x]} f_X(t) dm + \int_{(-\infty,x]} f_X(t)\delta_c$$

$$= \int_{(-\infty,x]} f_X(t) dm = P_X((-\infty,x]) = P_Y((-\infty,x])$$

For $x \geq c$,

$$\int_{(-\infty,x]} f_Y(t) d(m+\delta_c) = \int_{(-\infty,c]} f_Y(t) d(m+\delta_c)$$

$$= \int_{(-\infty,c)} f_X(t) d(m+\delta_c) + \int_{\{c\}} [1-F_X(c)] d(m+\delta_c)$$

$$= \int_{(-\infty,c)} f_X(t) dm + \int_{\{c\}} [1-F_X(c)] d\delta_c$$

$$= F_X(c) + [1-F_X(c)] = 1 = P_Y((-\infty,x])$$

Example 1.14.

Let X be a random variable with c.d.f. F_X and Lebesgue p.d.f. f_X, and $Y = X^2$.

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

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

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

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

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

In particular, if

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

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 (Transformation)

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 $(\mathscr{R}^k, \mathscr{B}^k)$ to $(\mathscr{R}^k, \mathscr{B}^k)$. Let $A_1, ..., A_m$ be disjoint sets in $\mathscr{B}^k$ such that $\mathscr{R}^k - (A_1 \cup \cdots \cup A_m)$ has Lebesgue measure 0 and g on A_j is one-to-one with a nonvanishing Jacobian, i.e., the determinant $\mathrm{Det}(\partial g(x)/\partial x) \neq 0$ on A_j, $j = 1, ..., m$. Then Y has the following Lebesgue p.d.f.:

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

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

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

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

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 $\chi^2_{n_1}$ and $\chi^2_{n_2}$ (Table 1.2), respectively.
The p.d.f. of $Z = X_1/X_2$ is

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

$$= \frac{\Gamma[(n_1+n_2)/2]}{\Gamma(n_1/2)\Gamma(n_2/2)} \frac{z^{n_1/2-1}}{(1+z)^{(n_1+n_2)/2}} I_{(0,\infty)}(z)$$

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^2_n.
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 $X_1, ..., X_n$ be independent random variables and $X_i = N(\mu_i, \sigma^2)$. The distribution of $Y = (X_1^2 + \cdots + X_n^2)/\sigma^2$ is called the *noncentral chi-square* distribution and denoted by $\chi_n^2(\delta)$, where $\delta = (\mu_1^2 + \cdots + \mu_n^2)/\sigma^2$ 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.

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

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

Moments

- If EX^k is finite, where k is a positive integer, EX^k is called the kth *moment* of X or P_X.
- If $E|X|^a < \infty$ for some real number a, $E|X|^a$ is called the ath *absolute moment* of X or P_X.
- If $\mu = EX$, $E(X-\mu)^k$ is called the kth *central moment* of X or P_X.
- $\mathrm{Var}(X) = E(X-EX)^2$ is called the *variance* of X or P_X.
- For random matrix $M = (M_{ij})$, $EM = (EM_{ij})$
- For random vector X, $\mathrm{Var}(X) = E(X-EX)(X-EX)^\tau$ is its covariance matrix, whose (i,j)th element, $i \neq j$, is called the covariance of X_i and X_j and denoted by $\mathrm{Cov}(X_i, X_j)$.
- $[\mathrm{Cov}(X_i, X_j)]^2 \leq \mathrm{Var}(X_i)\mathrm{Var}(X_j), \quad i \neq j$
- For random vector X, $\mathrm{Var}(X)$ is nonnegative definite
- If $\mathrm{Cov}(X_i, X_j) = 0$, then X_i and X_j are said to be uncorrelated.
- Independence implies uncorrelation, not converse
- If X is random and c is fixed, then $E(c^\tau X) = c^\tau E(X)$ and $\mathrm{Var}(c^\tau X) = c^\tau \mathrm{Var}(X) c$.

Three useful inequalities

- Cauchy-Schwartz inequality: $[E(XY)]^2 \leq EX^2 EY^2$ for random variables X and Y
- Jensen's inequality: $f(EX) \leq Ef(X)$ for a random vector X and convex function f ($f'' \geq 0$)
- Chebyshev's inequality: Let X be a random variable and φ a nonnegative and nondecreasing function on $[0,\infty)$, $\varphi(-t) = \varphi(t)$. Then, for each constant $t \geq 0$,

$$\varphi(t)P(|X| \geq t) \leq \int_{\{|X|\geq t\}} \varphi(X)dP \leq E\varphi(X)$$

Example 1.18.

If X is a nonconstant positive random variable with finite mean, then

$$(EX)^{-1} < E(X^{-1}) \quad \text{and} \quad E(\log X) < \log(EX),$$

since t^{-1} and $-\log t$ are convex functions on $(0,\infty)$.

If f and g are positive integrable functions on a measure space with a σ-finite measure ν and $\int f d\nu \geq \int g d\nu > 0$, then

$$\int f \log(f/g) d\nu \geq 0.$$

Definition 1.5 (Moment generating and characteristic functions)

Let X be a random k-vector.

(i) The *moment generating function* (m.g.f.) of X or P_X is defined as

$$\psi_X(t) = Ee^{t^\tau X}, \quad t \in \mathscr{R}^k.$$

(ii) The *characteristic function* (ch.f.) of X or P_X is defined as

$$\phi_X(t) = Ee^{\sqrt{-1}t^\tau X} = E[\cos(t^\tau X)] + \sqrt{-1}\,E[\sin(t^\tau X)], \quad t \in \mathscr{R}^k$$

Properties of m.g.f. and ch.f.

- If the m.g.f. is finite in a neighborhood of $0 \in \mathscr{R}^k$, then
  - moments of X of any order are finite,
  - $\phi_X(t)$ can be obtained by replacing t in $\psi_X(t)$ by $\sqrt{-1}t$
- If $Y = A^\tau X + c$, where A is a $k \times m$ matrix and $c \in \mathscr{R}^m$, then

$$\psi_Y(u) = e^{c^\tau u}\psi_X(Au) \quad \text{and} \quad \phi_Y(u) = e^{\sqrt{-1}c^\tau u}\phi_X(Au), \quad u \in \mathscr{R}^m$$

- For independent $X_1, ..., X_k$,

$$\psi_{\sum_i X_i}(t) = \prod_i \psi_{X_i}(t) \quad \text{and} \quad \phi_{\sum_i X_i}(t) = \prod_i \phi_{X_i}(t), \quad t \in \mathscr{R}^k$$

- For $X = (X_1, ..., X_k)$ with m.g.f. ψ_X finite in a neighborhood of 0,

$$\psi_X(t) = \sum_{(r_1,\dots,r_k)} \frac{\mu_{r_1,\dots,r_k} t_1^{r_1}\cdots t_k^{r_k}}{r_1!\cdots r_k!} \qquad \mu_{r_1,\dots,r_k} = E(X_1^{r_1}\cdots X_k^{r_k})$$

$$E(X_1^{r_1}\cdots X_k^{r_k}) = \left.\frac{\partial^{r_1+\cdots+r_k}\psi_X(t)}{\partial t_1^{r_1}\cdots\partial t_k^{r_k}}\right|_{t=0}$$

$$\left.\frac{\partial\psi_X(t)}{\partial t}\right|_{t=0} = EX, \qquad \left.\frac{\partial^2\psi_X(t)}{\partial t\partial t^\tau}\right|_{t=0} = E(XX^\tau)$$

- If $E|X_1^{r_1}\cdots X_k^{r_k}| < \infty$ for nonnegative integers $r_1,\dots,r_k$, then

$$\left.\frac{\partial^{r_1+\cdots+r_k}\phi_X(t)}{\partial t_1^{r_1}\cdots\partial t_k^{r_k}}\right|_{t=0} = (-1)^{(r_1+\cdots+r_k)/2}E(X_1^{r_1}\cdots X_k^{r_k})$$

$$\left.\frac{\partial\phi_X(t)}{\partial t}\right|_{t=0} = \sqrt{-1}EX, \quad \left.\frac{\partial^2\phi_X(t)}{\partial t\partial t^\tau}\right|_{t=0} = -E(XX^\tau)$$

- Special case of $k=1$:

$$\psi_X(t) = \sum_{i=0}^{\infty}\frac{E(X^i)t^i}{i!} \qquad \text{if } \psi(t) < \infty$$

$$E(X^i) = \psi^{(i)}(0) = \left.\frac{d\psi_X^i(t)}{dt^i}\right|_{t=0}, \qquad \phi_X^{(i)}(0) = (-1)^{i/2}E(X^i)$$

Example 1.19.

$X = N(\mu, \sigma^2)$

$$\psi_X(t) = \frac{1}{\sqrt{2\pi}\sigma} \int e^{tx} e^{-(x-\mu)^2/2\sigma^2} dx \qquad \frac{x-\mu}{\sigma} = y$$

$$= \frac{1}{\sqrt{2\pi}} \int e^{t(\sigma y+\mu)} e^{-y^2/2} dy = \frac{e^{\mu t+\sigma^2 t^2/2}}{\sqrt{2\pi}} \int e^{-(y-\sigma t)^2/2} dy = e^{\mu t+\sigma^2 t^2/2}$$

A direct calculation shows that

$EX = \psi_X'(0) = \mu$

$EX^2 = \psi_X''(0) = \sigma^2 + \mu^2$

$EX^3 = \psi_X^{(3)}(0) = 3\sigma^2\mu + \mu^3$

$EX^4 = \psi_X^{(4)}(0) = 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4$

If $\mu = 0$, then $EX^p = 0$ when p is an odd integer

$EX^p = (p-1)(p-3)\cdots 3 \cdot 1\sigma^p$ when p is an even integer

The cumulant generating function of X is

$$\kappa_X(t) = \log \psi_X(t) = \mu t + \sigma^2 t^2/2$$

$\kappa_1 = \mu$, $\kappa_2 = \sigma^2$, and $\kappa_r = 0$ for $r = 3, 4, \ldots$.

Example 1.19 (continued): A random variable X has finite $E(X^k)$ for $k = 1, 2...$, but $\psi_X(t) = \infty$, for any $t \neq 0$

P_n: the probability measure for $N(0, n^2)$ with p.d.f. f_n, $n = 1, 2, ...$
$P = \sum_{n=1}^{\infty} 2^{-n} P_n$ is a probability measure with Lebesgue p.d.f. $\sum_{n=1}^{\infty} 2^{-n} f_n$ (Exercise 35)
Let X be a random variable having distribution P.
It follows from Fubini's theorem that X has finite moments of any order; for even k,

$$E(X^k) = \int x^k dP = \int \sum_{n=1}^{\infty} x^k 2^{-n} dP_n = \sum_{n=1}^{\infty} 2^{-n} \int x^k dP_n$$

$$= \sum_{n=1}^{\infty} 2^{-n} (k-1)(k-3) \cdots 1 n^k < \infty$$

and $E(X^k) = 0$ for odd k.
By Fubini's theorem again, for any $t \neq 0$,

$$\psi_X(t) = \int e^{tx} dP = \sum_{n=1}^{\infty} 2^{-n} \int e^{tx} dP_n = \sum_{n=1}^{\infty} 2^{-n} e^{n^2 t^2/2} = \infty$$

Theorem 1.6. (Uniqueness)

Let X and Y be random k-vectors.

(i) If $\phi_X(t) = \phi_Y(t)$ for all $t \in \mathscr{R}^k$, then $P_X = P_Y$.

(ii) If $\psi_X(t) = \psi_Y(t) < \infty$ for all t in a neighborhood of 0, then $P_X = P_Y$.

Proof

See the textbook.

Example 1.20

Let X_i, $i = 1, ..., k$, be independent random variables and X_i have the gamma distribution $\Gamma(\alpha_i, \gamma)$ (Table 1.2), $i = 1, ..., k$.

From Table 1.2, X_i has the m.g.f. $\psi_{X_i}(t) = (1 - \gamma t)^{-\alpha_i}$, $t < \gamma^{-1}$, $i = 1, ..., k$.

Then, the m.g.f. of $Y = X_1 + \cdots + X_k$ is equal to

$$\psi_Y(t) = \prod_i \psi_{X_i}(t) = \prod_i (1 - \gamma t)^{-\alpha_i} = (1 - \gamma t)^{-(\alpha_1 + \cdots + \alpha_k)}, \quad t < \gamma^{-1}.$$

From Table 1.2, the gamma distribution $\Gamma(\alpha_1 + \cdots + \alpha_k, \gamma)$ has the m.g.f. $\psi_Y(t)$ and, hence, is the distribution of Y (by Theorem 1.6).

Can the moments determine a distribution?

Can two random variables with different distributions have the same moments of any order?

$$X_1 \text{ has pdf} \qquad f_1(x) = \frac{1}{\sqrt{2\pi}x} e^{-(\log x)^2/2}, \qquad x \ge 0$$

$$X_2 \text{ has pdf} \quad f_2(x) = f_1(x)[1+\sin(2\pi \log x)], \qquad x \ge 0$$

For any positive integer n,

$$E(X_1^n) = \frac{1}{\sqrt{2\pi}} \int_0^\infty x^{n-1} e^{-(\log x)^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{ny-y^2/2} dy = e^{n^2/2}$$

$$E(X_2^n) = E(X_1^n) + \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-s^2/2} \sin(2\pi s) ds = E(X_1^n)$$

This shows that X_1 and X_2 have the same moments of order $n = 1, 2, \ldots$, but they have different distributions.

$$M_X(t) = \int_0^\infty \frac{e^{tx}}{\sqrt{2\pi}x} e^{-(\log x)^2/2} dx = \infty, \quad t > 0$$

$$M_X(t) = \int_0^\infty \frac{e^{tx}}{\sqrt{2\pi}x} e^{-(\log x)^2/2} dx \le \int_0^\infty \frac{1}{\sqrt{2\pi}x} e^{-(\log x)^2/2} dx = 1, \quad t \le 0$$