Mathematical Statistics and Data Analysis

Lecture 4: Review of Probability - Part III

Lyu Ni

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

September 23, 2019

Outlines

Reading Material

Textbook:

- Rice: 2.4, Chapter 4;
- Mao: 2.6, 2.2, 2.3, 2.7;

Function of a discrete r.v.

Suppose that X is a discrete random variable and the p.m.f of X is

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_n)$	$\cdots$

Let $Y = g(X)$. Then

- Y is also a discrete r.v.
- The p.m.f of Y is

Y	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_n)$	$\cdots$
P	$f(x_1)$	$f(x_2)$	$\cdots$	$f(x_n)$	$\cdots$

- If $g(x_i) = g(x_j)$, then

$$P(Y = g(x_i)) = P(Y = g(x_j)) = f(x_i) + f(x_j)$$

Function of a continuous r.v.

Special Case: $g(x)$ is strictly monotonic

Theorem

Let X be a continuous random variable with density $f(x)$ and $Y = g(X)$ where $g(\cdot)$ is strictly monotonic and its inverse function $h(y)$ has a continuous derivate.
The p.d.f. of Y is

$$f(y) = \begin{cases} f[h(y)]|h'(y)|, & a < y < b, \\ 0, & \text{otherwise} \end{cases}$$

where

$a = \min\{g(-\infty), g(\infty)\}$ and $b = \max\{g(-\infty), g(+\infty)\}$.

Function of a continuous r.v.

Example 1

Suppose $X \sim N(\mu, \sigma^2)$. Then the p.d.f. of $Y = e^X$ is

$$f(y) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}y} \exp\left\{-\frac{(\ln y-\mu)^2}{2\sigma^2}\right\}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

Solution: As we know, $y = e^x$ is a strictly increasing function of x and the inverse function $x = \ln y$. Apply the theorem, and we have

- When $y \leq 0$, $F_Y(y) = 0$ and thus $f_Y(y) = 0$;
- When $y > 0$, the p.d.f of Y is

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\ln y - \mu)^2}{2\sigma^2}\right\} \cdot \frac{1}{y}$$

Function of a continuous r.v.

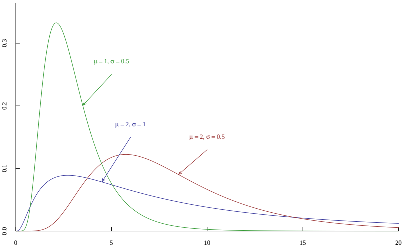
Example 1(Con'd)

Suppose $X \sim N(\mu, \sigma^2)$. Then the p.d.f. of $Y = e^X$ is

$$f(y) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}y} \exp\left\{-\frac{(\ln y-\mu)^2}{2\sigma^2}\right\}, & y > 0 \\ 0, & y \le 0 \end{cases}$$

Remark:

- This distribution is said to be log-normal distribution $LN(\mu, \sigma^2)$;
- It is a skewed distribution;



Function of a continuous r.v.

Example 2

Suppose $X \sim Ga(n, 1/\beta)$ and the p.d.f. of X is

$$f(y) = \begin{cases} \frac{1}{(n-1)!\beta^n} x^{n-1} e^{-x/\beta}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

Suppose we want to find the p.d.f. of $g(X) = \frac{1}{X}$.
Solution: Let $y = g(x)$. Then $g^{-1}(y) = 1/y$ and $\frac{\mathrm{d}}{\mathrm{d}_y} g^{-1}(y) = -\frac{1}{y^2}$. Apply the theorem, for $y > 0$, we have

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) \right| = \frac{1}{(n-1)!\beta^n} \left(\frac{1}{y} \right)^{n-1} e^{-1/(y\beta)} \frac{1}{y^2}$$

$$= \frac{1}{(n-1)!\beta^n} \left(\frac{1}{y} \right)^{n+1} e^{-1/(y\beta)}$$

Function of a continuous r.v.

Theorem

Let X have a continuous c.d.f. $F_X(x)$ and define the random variable Y as $Y = F_X(X)$. Then Y is uniformly distributed on $(0,1)$, that is,

$$P(Y \leq y) = y, 0 < y < 1.$$

Solution: Let $F_X^{-1}(y) = \inf\{x : F(x) \geq y\}$. For $Y = F_X(X)$, we have , for $0 < y < 1$,

$$\begin{aligned} P(Y \leq y) &= P(F_X(X) \leq y) \\ &= P(F_X^{-1}(F_X(X)) \leq F_X^{-1}(y)) \\ &= P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) = y \end{aligned}$$

Function of a continuous r.v.

Solution (Con'd): At the endpoints we have $P(Y \leq y) = 1$ for $y \geq 1$ and $P(Y \leq y) = 0$ for $y \leq 0$, showing that Y has a uniform distribution.

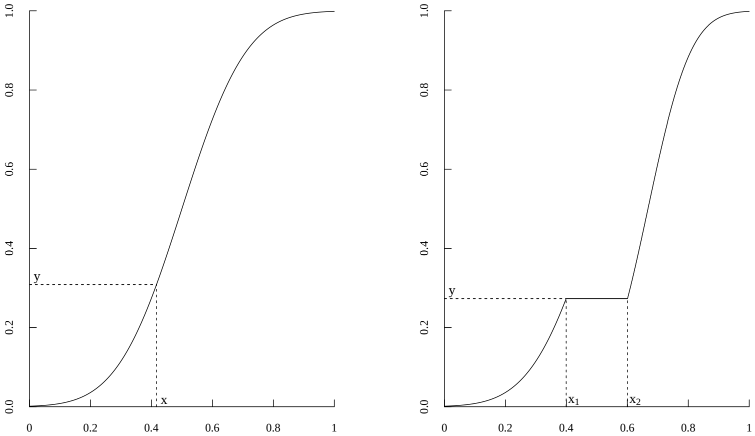
The reasoning behind the equality

$$P(F_X^{-1}(F_X(X)) \leq F_X^{-1}(y)) = P(X \leq F_X^{-1}(y))$$

is somewhat subtle and deserves additional attention.

- If F_X is strictly increasing, then it is true that $F_X^{-1}(F_X(x)) = x$.
- Suppose that F_X is flat in a certain interval, i.e. $F_X^{-1}(F_X(x)) \neq x, x \in [x_1, x_2]$. Then $F_X^{-1}(F_X(x)) = x_1$ for any x in this interval. Even in this case, though, the probability equality holds, since $P(X \leq x) = P(X \leq x_1)$ for any $x \in [x_1, x_2]$. The flat c.d.f. denotes a region of 0 probability $P(x_1 < X \leq x) = F_X(x) - F_X(x_1) = 0$.

Function of a continuous r.v.



(a) (b)

Function of a continuous r.v.

In many applications, the function g may be neither increasing nor decreasing.

Example

Suppose $X \sim N(0,1)$. What is the p.d.f. of $Y = X^2$?
Solution: Since $Y = X^2 \geq 0$, $F_Y(y) = 0$ when $y \leq 0$. Then $p_Y(y) = 0$. When $y > 0$, we have

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= 2\Phi(\sqrt{y}) - 1 \end{aligned}$$

Thus the c.d.f of Y is

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

Function of a continuous r.v.

Example (Con'd)

Solution: The p.d.f. of Y can be obtained from the c.d.f. by differentiation, that is,

$$\begin{aligned} f_Y(y) &= \begin{cases} \varphi(\sqrt{y})y^{-\frac{1}{2}}, & y > 0 \\ 0, & y \leq 0 \end{cases} \\ &= \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, & y > 0 \\ 0, & y \leq 0 \end{cases} \end{aligned}$$

Therefore, $Y \sim \chi^2(1)$.

Expectation

Definition

- Suppose that X is a discrete r.v. and the p.m.f. of X is $f(x_i) = P(X = x_i), i = 1, 2, \cdots, n$. If

$$\sum_{i=1}^{\infty} |x_i| f(x_i) < \infty,$$

then

$$E(X) = \sum_{i=1}^{\infty} x_i f(x_i)$$

is said to be the **expectation** of X.

Expectation

Definition

- Suppose that X is a continuous r.v. and the p.d.f. of X is $f(x)$. If

$$\int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x < \infty,$$

then

$$E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$$

is said to be the **expectation** of X.

Property

- Suppose X is a r.v. with a p.m.f. or p.d.f. $f(x)i$. The p.d.f. of a function of X is

$$E(g(X)) = \begin{cases} \sum_i g(x_i)f(x_i), & X \text{ is a discrete r.v.} \\ \int_{\infty}^{\infty} g(x)f(x)\mathrm{d}x & X \text{ is a continuous r.v.} \end{cases}$$

- Suppose that c is a constant. Then $E(c) = c$;
- For each a, we have

$$E(aX) = aE(X);$$

- For every two functions $g_1(x)$ and $g_2(x)$, we have

$$E(g_1(X) \pm g_2(X)) = E[g_1(X)] \pm E[g_2(X)]$$

Variance

Definition

- Suppose that there exists the expectation of the random variable X^2.
  - The **variance** of X is $E(X-EX)^2$, i.e.,

$$
\begin{aligned}
Var(X) &= E(X-E(X))^2 \\
&= \begin{cases} \sum_i (x_i - E(X))^2 f(x_i) & X \text{ is a discrete r.v.} \\ \int_{-\infty}^{\infty} (x - E(X))^2 f(x) \mathrm{d}x & X \text{ is a continuous r.v.} \end{cases}
\end{aligned}
$$

  - The **standard deviation** of X is the square root of the variance.

Variance

Property

- $Var(X) = E(X^2) - (E(X))^2$;
- If $Var(X)$ exists and $Y = a + bX$, then

$$Var(Y) = b^2 Var(X)$$

Solution: Since $E(Y) = a + bE(X)$,

$$\begin{aligned} E[(Y - E(Y))^2] &= E[(a + bX - a - bE(X))^2] \\ &= E[(bX - bE(X))^2] \\ &= b^2 E[(X - E(X))^2] \\ &= b^2 Var(X) \end{aligned}$$

Variance

Property

- (**Chebyshev's Inequality**) Let X be a random variable with mean μ and variance σ^2. Then, for any $t > 0$,

$$P(|X-\mu| > t) \leq \frac{\sigma^2}{t^2}.$$

Solution: Suppose X is a continuous random variable and the p.d.f. is $f(x)$. Then

$$\begin{aligned} P(|X-\mu| \geq t) &= \int\limits_{\{x:|x-\mu|\geq t\}} f(x)\mathrm{d}x \leq \int\limits_{\{x:|x-\mu|\geq t\}} \frac{(x-\mu)^2}{t^2} f(x)\mathrm{d}x \\ &\leq \frac{1}{t^2}\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\mathrm{d}x = \frac{Var(X)}{t^2} = \frac{\sigma^2}{t^2} \end{aligned}$$

Variance

Theorem

Suppose X is a random variable and the variance exists. Thus, $Var(X) = 0$ if and only if $P(X = c) = 1$ for c is a constant.

Solution:It is obvious that $Var(X) = 0$ if $P(X = c) = 1$. Then, we prove the necessity. Suppose that $Var(X) = 0$. It means that $E(X)$ exists. Since

$$\{|X - E(X)| > 0\} = \bigcup_{n=1}^{\infty} \left\{|X - E(X)| \geq \frac{1}{n}\right\},$$

we have

$$P(|X - E(X)| > 0) = P\left(\bigcup_{n=1}^{\infty} \left\{|X - E(X)| \geq \frac{1}{n}\right\}\right).$$

Variance

Solution (Con'd):

$$
\begin{aligned}
P(|X-E(X)|>0) &= P\left(\bigcup_{n=1}^{\infty}\left\{|X-E(X)|\geq\frac{1}{n}\right\}\right)\\
&\leq \sum_{n=1}^{\infty} P\left(|X-E(X)|\geq\frac{1}{n}\right)\\
&\leq \sum_{n=1}^{\infty}\frac{Var(X)}{(1/n)^2}=0
\end{aligned}
$$

Thus,

$$P(|X-E(X)|=0)=1$$

Let $c=E(X)$. The desired result is obtained.

Moment

Definition

Suppose X is a random variable and k is a positive integer. If the expectation exists, then

- The **kth moment** of X, μ_k, is

$$\mu_k = E(X^k)$$

- The **kth central moment** of X, ν_k, is

$$\nu_k = E(X - E(X))^k$$

Obvious, μ_1 is the expectation and v_2 is the variance.

Moment

Property

There is the relationship between moments and central moments, that is,

$$\nu_k = E(X - E(X))^k = E(X - \mu_1)^k = \sum_{i=0}^{k} \binom{k}{i} \mu_i (-\mu_1)^{k-i}$$

Then, the first, second, third and forth central moments are presented as follows:

$$\begin{aligned}
\nu_1 &= 0 \\
\nu_2 &= \mu_2 - \mu_1^2 \\
\nu_3 &= \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 \\
\nu_4 &= \mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4
\end{aligned}$$

Moment

Example

If $X \sim N(0, \sigma^2)$, then

$$
\begin{aligned}
\mu_k &= E(X^k) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^k \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \mathrm{d}x \\
&= \frac{\sigma^k}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^k \exp\left\{-\frac{u^2}{2}\right\} \mathrm{d}u
\end{aligned}
$$

- If k is odd, the integrand is an odd function. Thus, $\mu_k = 0$, $k = 1, 3, 5, \cdots$.

Example (Con'd)

- If k is even, the integrand is an even function. Let $z = u^2/2$. Then

$$\begin{aligned}\mu_k &= \sqrt{2}\pi\sigma^k 2^{(k-1)/2}\int_0^\infty z^{(k-1)/2}e^{-z}\mathrm{d}z \\ &= \sqrt{\frac{2}{\pi}}\sigma^k 2^{(k-1)/2}\Gamma\left(\frac{k+1}{2}\right) \\ &= \sigma^k(k-1)(k-3)\cdots 1, k = 2,4,6,\cdots .\end{aligned}$$

Then

$$\mu_1 = 0, \mu_2 = \sigma^2, \mu_3 = 0, \mu_4 = 3\sigma^2$$

Since $E(X) = 0$, the central moment is equal to the moment, i.e. $\nu_k = \mu_k$, $k = 1,2,3,\cdots$.

Coefficient of Variation

Definition

Suppose X is a random variable and the second-order moment exists. The **coefficient of variation, CV** is

$$C_v = \frac{\sqrt{Var(X)}}{E(X)}$$

Why we use Coefficient of Variation?

Suppose X is a random variable and $Y = bX$ where $b > 0$. Then $E(Y) = bE(X)$ and $\sqrt{Var(Y)} = \sqrt{b^2 Var(X)} = b\sqrt{Var(X)}$. However,

$$\frac{\sqrt{Var(Y)}}{E(Y)} = \frac{\sqrt{Var(X)}}{E(X)}$$

Coefficient of Variation

Definition

Suppose X is a random variable and the second-order moment exists. The **coefficient of variation, CV** is

$$C_v = \frac{\sqrt{Var(X)}}{E(X)}$$

Why we use Coefficient of Variation?

Suppose X is a random variable and $Y = bX$ where $b > 0$. Then $E(Y) = bE(X)$ and $\sqrt{Var(Y)} = \sqrt{b^2Var(X)} = b\sqrt{Var(X)}$. However,

$$\frac{\sqrt{Var(Y)}}{E(Y)} = \frac{\sqrt{Var(X)}}{E(X)}$$

Coefficient of Variation

Definition

Suppose X is a random variable and the second-order moment exists. The **coefficient of variation, CV** is

$$C_v = \frac{\sqrt{Var(X)}}{E(X)}$$

Why we use Coefficient of Variation?

Suppose X is a random variable and $Y = bX$ where $b > 0$. Then $E(Y) = bE(X)$ and $\sqrt{Var(Y)} = \sqrt{b^2 Var(X)} = b\sqrt{Var(X)}$. However,

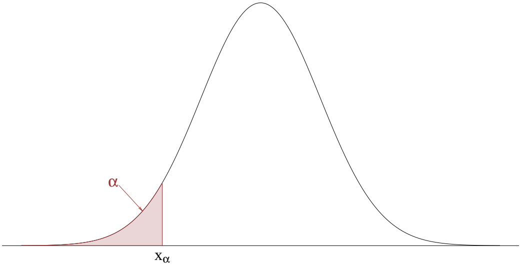
$$\frac{\sqrt{Var(Y)}}{E(Y)} = \frac{\sqrt{Var(X)}}{E(X)}$$

Quantiles

Definition

Suppose X is a continuous random variable, the c.d.f. of X is $F(x)$ and the p.d.f. of X is $f(x)$. For each $\alpha \in (0,1)$, the **αth (lower) quantile**, x_α, is

$$F(x_\alpha) = \int_{-\infty}^{x_\alpha} f(x)\mathrm{d}x = \alpha$$

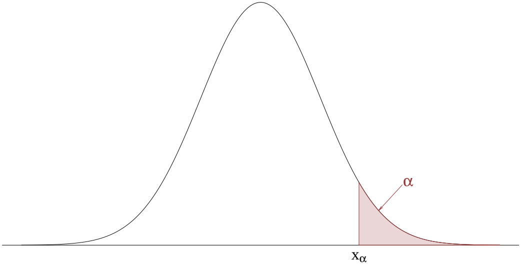


Quantiles

Definition

Suppose X is a continuous random variable, the c.d.f. of X is $F(x)$ and the p.d.f. of X is $f(x)$. For each $\alpha \in (0, 1)$, the **αth upper quantile**, x'_α, is

$$F(x_\alpha) = \int_{x'_\alpha}^{\infty} f(x)\mathrm{d}x = \alpha$$



Quantiles

Example

Suppose $Z \sim N(0,1)$, z_α is the αth quantile of Z and $\Phi(\cdot)$ is the c.d.f. of Z. Then

$$\Phi(z_\alpha) = \alpha,$$

Let $\Phi^{-1}(\cdot)$ be the inverse function of $\Phi(\cdot)$. Thus,

$$z_\alpha = \Phi^{-1}(\alpha).$$

Example (Con'd)

Suppose $X \sim N(\mu, \sigma^2)$ and x_α is the αth quantile of X. Then,

$$\Phi\left(\frac{x_\alpha - \mu}{\sigma}\right) = \alpha \Rightarrow \frac{x_\alpha - \mu}{\sigma} = z_\alpha$$

So, the relationship between x_α and z_α is

$$x_\alpha = \mu + \sigma z_\alpha$$

Median

Definition

Suppose X is a continuous random variable with a c.d.f. $F(x)$ and a p.d.f. $f(x)$. $x_{0.5}$ is the $\alpha = 0.5$ quantile, and is also said to be **median**, i.e.

$$F(0.5) = \int_{-\infty}^{x_{0.5}} f(x)\mathrm{d}x = 0.5$$

Exmaple

Suppose $X \sim Exp(\lambda)$. $x_{0.5}$ is the solution of the equation

$$1 - e^{-\lambda x_{0.5}} = 0.5.$$

Then,

$$x_{0.5} = \ln 2/\lambda$$

Expectation, Variance and Median

Remark

- The expectation, μ, is obtained by minimizing the function
$$\mu = \arg\min_a E(X-a)^2$$
- The median, $x_{0.5}$, is obtained by minimizing the function
$$x_{0.5} = \arg\min_a E|X-a|$$

Skewness

Definition

Suppose X is a random variable and the first, second and third moments exist. The **coefficient of skewness** or **skewness** is

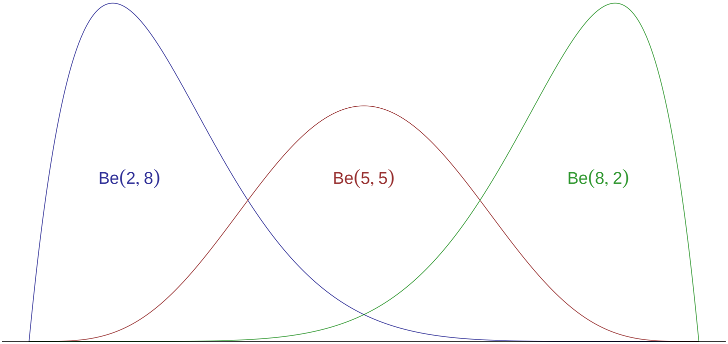
$$\beta_s = \frac{\nu_3}{\nu_4^{3/2}} = \frac{E(X - EX)^3}{[Var(X)]^{3/2}}$$

- If $\beta_s < 0$. the distribution of X is left-skewed;
- If $\beta_s > 0$, the distribution of X is right-skewed;
- If $\beta_s = 0$, the distribution of X is symmetric.

Skewness

Example

Here we consider the c.d.f.s of $Be(2, 8)$, $Be(5, 5)$ and $Be(8, 2)$ as follows:



Kurtosis

Definition

Suppose X is a random variable and the first, second, third and forth moments exist. The **coefficient of kurtosis** or **kurtosis** is

$$\beta_k = \frac{\nu_4}{\nu_2^2} - 3 = \frac{E(X - EX)^4}{[Var(X)]^2} - 3$$

Why is "3" in the formula?

- If $\beta_k > 0$, the distribution has a heavy tail;
- If $\beta_k < 0$, the distribution has a thin tail;