STA2001 Probability and Statistics (I)

Lecture 14

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

Review of Last Lecture

Key concepts and/or techniques:

- bivariate normal distribution and its properties

1. marginal distributions are normal
2. conditional distributions are normal
3. independence $\Longleftrightarrow$ Uncorrelation

- find the distribution of the function of RVs, i.e., determine the pmf or pdf of the functions of RVs

Review of Last Lecture

Definition

Let X and Y be 2 continuous RVs and have the joint pdf

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp[-\frac{1}{2}q(x,y)], x \in \mathbb{R}, y \in \mathbb{R},$$

$$q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right] \geq 0$$

where $\mu_X, \mu_Y \in \mathbb{R}$, $\sigma_X, \sigma_Y > 0$ and $|\rho| < 1$. Then X and Y are said to be bivariate normally distributed.

Key components: Scaled exponential function with a quadratic and negative function as its exponent.

Review of Last Lecture

1. Marginal distributions are normal:

$$X \sim N(\mu_X, \sigma_X^2), Y \sim N(\mu_Y, \sigma_Y^2)$$

2. Conditional distributions are normal:

$$X|Y = y \sim N\left(\mu_X + \frac{\sigma_X}{\sigma_Y}\rho(y - \mu_Y), (1 - \rho^2)\sigma_X^2\right)$$

$$Y|X = x \sim N\left(\mu_Y + \frac{\sigma_Y}{\sigma_X}\rho(x - \mu_X), (1 - \rho^2)\sigma_Y^2\right)$$

3. Independence $\Longleftrightarrow$ Uncorrelation

Review of Last Lecture

Key concepts and/or techniques:

[Function of One Random Variable]

Let X be a RV of either discrete or continuous type with its pmf or pdf denoted by $f(x)$. Consider a function of X, say $Y = u(X)$. Then Y is also a RV and has its pmf or pdf.

How to compute the pmf or pdf of Y?

1. $Y = u(X)$ is one-to-one
2. Random number generator: $F(x)$ is a strictly increasing cdf of a random distribution and $X = F^{-1}(Y)$ with $Y \sim U(0, 1)$
3. If $Y = u(X)$ is NOT one-to-one, then there are no general results and we can only rely on the definition and properties of pmf or pdf.

Review of Last Lecture

1. For discrete RV, when $Y = u(X)$ be a one-to-one mapping with inverse $X = v(Y)$. Then the pmf of Y is

$$g(y) = f[v(y)] \quad \text{for} \quad y \in \overline{S_Y}$$

2. For continuous RV, when $Y = u(X)$ is continuous, strictly decreasing or increasing and has inverse function $X = v(Y)$, whose derivative $\frac{dv(y)}{dy}$ exists, the pdf of Y, denoted by $g(y)$,

$$g(y) = f(v(y))\left|\frac{dv(y)}{dy}\right|$$

Review of Last Lecture

[Theorem: Random Number Generator]

Let $Y \sim U(0,1)$ and $F(x)$ have the properties of a cdf of a continuous RV with $F(\mathrm{a}) = 0, F(b) = 1$. Moreover, $F(x)$ is strictly increasing such that $F(x) : (a,b) \to [0,1]$, where a could be $-\infty$, b could be ∞. Then $X = F^{-1}(Y)$ is continuous RV with cdf $F(x)$.

[Algorithm: Random number generator from a random distribution with strictly increasing cdf $F(x)$]

1. generator a random number y from $U(0,1)$
2. Take $x = F^{-1}(y)$

Then x is a random number generated from the continuous RV with cdf $F(x)$.

Histogram for continuous distribution

The simplest form of a histogram is constructed as follows

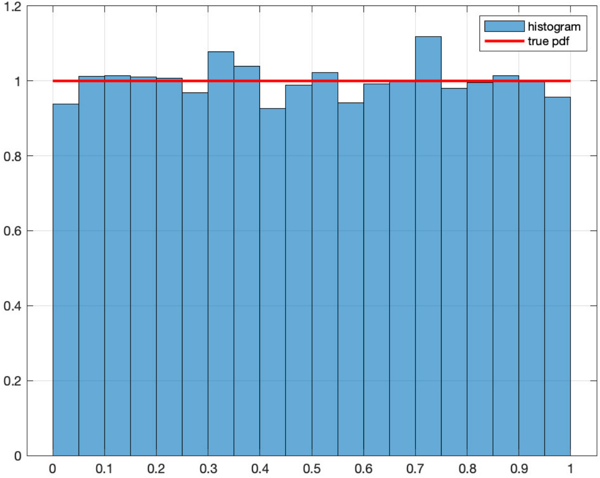
1. divide (or "bin") the sample space of the distribution into a sequence of adjacent, non-overlapping and equally spaced subintervals.
2. treat each subinterval as an event, then count how many observed numerical outcomes fall into each subinterval and calculate the relative frequency
3. draw a rectangle erected over the bin with height equal to the relative frequency divided by the width of each subinterval.

Remark:

- Note that the area of the histogram is equal to 1, thus histogram gives an approximation of the probability density function of the underlying random variable.

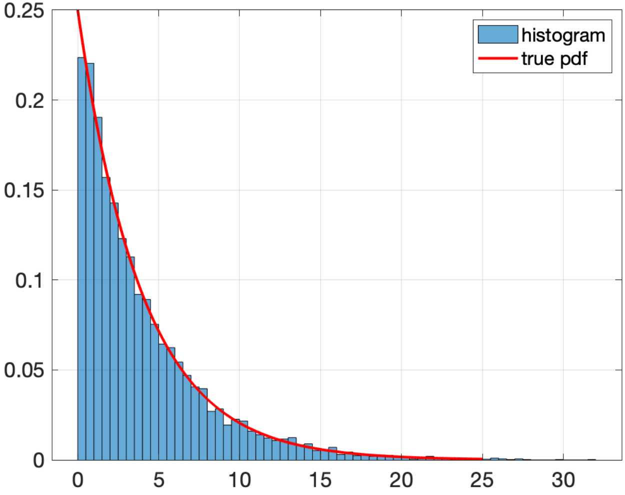
Example 3

The histogram of 10000 random numbers y generated from $U(0, 1)$



Example 3

The histogram of 10000 random numbers x for exponential distribution with $\theta = 4$



Example 3

Matlab script for the random number generator:

```
figure(1);
y=rand(10000,1); % generate 10000 random number y from U(0,1)
histogram(y,'normalization','pdf') % draw the histogram of
10000 y
hold on;
plot(0:0.01:1,ones(1,length(0:0.01:1)),'r',
'linewidth',2) % plot the true pdf of U(0,1)
grid on;
legend('histogram', 'true pdf')
figure(2);
theta=4;
x=-theta*log(1-y); % generate 10000 random number x
histogram(x,'normalization','pdf') % draw the histogram of
10000 x
hold on;
plot(0:0.01:25,exp(-(0:0.01:25)/theta)/theta,'r',
'linewidth',2) % plot the true pdf of exponential distribution with
theta=4
grid on;
```

Theorem 5.1-2, page 176

[Theorem]

Suppose that X is a continuous RV with $\overline{S_X} = (a, b)$, and moreover, its cdf $F(x)$ is strictly increasing. Then the RV Y, defined by $Y = F(X)$, has a uniform distribution, that is, $Y \sim U(0, 1)$.

Theorem 5.1-2, page 176

Proof:

Since $F(\mathrm{a}) = 0$, $F(b) = 1$ and $F(x)$ is strictly increasing,

$Y = F(X)$ with the range $\overline{S_Y} = (0, 1)$.

Consider the cdf of Y:

$$P(Y \leq y) = P(F(X) \leq y), \quad y \in (0, 1)$$

Theorem 5.1-2, page 176

Since $F(x)$ is strictly increasing, $\{F(X) \le y\} \iff \{X \le F^{-1}(y)\}$.

$$P(Y \le y) = P(F(X) \le y) = P(X \le F^{-1}(y)), \quad 0 < y < 1.$$

Since $P(X \le x) = F(x)$, we have

$$P(Y \le y) = F(F^{-1}(y)) = y, \quad 0 < y < 1 \longrightarrow \quad \text{cdf of } U(0,1)$$

Example 2, page 174, continued

Let X have the pdf

$$f(x) = 3(1-x)^2, \quad 0 < x < 1$$

Consider $Y = (1-X)^3$ and then $Y \sim U(0,1)$.

The result can be obtained from Theorem 5.1-2. Since

$$F(x) = 1 - (1-x)^3$$

is strictly increasing, then

$$F(X) = 1 - (1-X)^3 \sim U(0,1)$$

which implies $(1-X)^3 = [1 - F(X)] \sim U(0,1)$.

The case: $Y = u(X)$ not one-to-one

[Function of One Random Variable]

Let X be a RV of either discrete or continuous type with its pmf or pdf denoted by $f(x)$. Consider a function of X, say $Y = u(X)$. Then Y is also a RV and has its pmf or pdf.

How to compute the pmf or pdf of Y?

When $Y = u(X)$ is not one-to-one, there is no general result.

Example 4

Question

Assume that X is a continuous RV with pdf

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

Let $Y = X^2$. Find the pdf of Y.

Clearly, $\overline{S_Y} = [0, \infty)$

Let the cdf of Y be $G(y)$. Then

$$\begin{aligned} G(y) &= P(Y \leq y), \quad y \in [0, \infty) \\ &= P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f(x)dx \end{aligned}$$

Example 4

$$
\begin{aligned}
G(y) &= P(Y \leq y), \quad y \in [0, \infty) \\
&= P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f(x)dx \\
&= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\pi(1+x^2)} dx = 2\int_0^{\sqrt{y}} \frac{1}{\pi(1+x^2)} dx
\end{aligned}
$$

$$\Rightarrow g(y) = G'(y) = \tfrac{2}{\pi(1+y)} \times \tfrac{1}{2} \times \tfrac{1}{\sqrt{y}} = \tfrac{1}{\pi(1+y)\sqrt{y}}$$

Section 5.3 Several Random Variables (Multivariate RVs)

Motivation

Random Experiment: Any procedure that can be repeated infinitely times and has more than one possible outcomes.

Performing a random experiment one time, the outcome may contain

a scalar $\longrightarrow$ univariate RV: X,$f(x)$, pmf or pdf

a pair of two scalars $\longrightarrow$ bivariate RV:$(X, Y), f(x, y)$,pmf or pdf

a tuple of several scalars $\longrightarrow$ multivariate RV:$(X_1, X_2, \cdots, X_n)$

the corresponding joint pmf or pdf $f(x_1, x_2, \cdots, x_n)$

Joint pmf or pdf for multivariate RV

- Discrete type: $X_1, X_2, \cdots, X_n$ are all discrete

  joint pmf $f(x_1, \cdots, x_n) : \overline{S} \to (0, 1]$

  1. $f(x_1, \cdots, x_n) > 0, \quad (x_1, \cdots, x_n) \in \overline{S}$
  2. $\sum_{x_1, \cdots, x_n \in \overline{S}} f(x_1, \cdots, x_n) = 1$
  3. $P((X_1, \cdots, X_n) \in A) = \sum_{(x_1, \cdots, x_n) \in A} f(x_1, \cdots, x_n)$

- Continuous type: $X_1, X_2, \cdots, X_n$ are all continuous

  joint pdf $f(x_1, \cdots, x_n) : \overline{S} \to (0, \infty)$

  1. $f(x_1, \cdots, x_n) > 0, \quad (x_1, \cdots, x_n) \in \overline{S}$
  2. $\int_{\overline{S}} f(x_1, \cdots, x_n) dx_1 \cdots dx_n = 1.$
  3. $P((X_1, \cdots, X_n) \in A) = \int_A f(x_1, \cdots, x_n) dx_1 \cdots dx_n.$

Derivation of Joint pmf or pdf

A critical problem is how to derive the joint pmf or pdf of multivariate RVs. However, there is no general solution.

Only in some special cases, it is easy to derive the joint pmf of pdf of multivariable RVs.

Derivation of Joint pmf or pdf

A critical problem is how to derive the joint pmf or pdf of multivariate RVs. However, there is no general solution.

Only in some special cases, it is easy to derive the joint pmf of pdf of multivariable RVs.

Note that the multivariate RVs arise in many different ways.

For example, we can perform a random experiment n times and let $X_i, i = 1, \cdots, n$ denote the RV for the ith repetition of the random experiment. Then $(X_1, \cdots, X_n)$ is a multivariate RV.

If the n repetitions of the random experiment are **independent** then the joint pmf or pdf is easy to derive.

Example 1, page 188

Roll a fair die twice. Let X_1 denote the point of the first roll and

X_2 the point of the second roll.

For $X_1 = x_1$, its pmf

$$f_{X_1}(x_1) = P(X_1 = x_1) = \frac{1}{6}, \quad x_1 = 1, 2, 3, 4, 5, 6.$$

Example 1, page 188

Roll a fair die twice. Let X_1 denote the point of the first roll and

X_2 the point of the second roll.

For $X_1 = x_1$, its pmf

$$f_{X_1}(x_1) = P(X_1 = x_1) = \frac{1}{6}, \quad x_1 = 1, 2, 3, 4, 5, 6.$$

For $X_2 = x_2$, its pmf

$$f_{X_2}(x_2) = P(X_2 = x_2) = \frac{1}{6}, \quad x_2 = 1, 2, 3, 4, 5, 6.$$

Example 1, page 188

Assume that the two rolls are independent, then X_1 and X_2 are independent, and thus for $X_1 = x_1, X_2 = x_2$, the joint pmf of X_1 and X_2,

$$\begin{aligned} f(x_1, x_2) &= P(X_1 = x_1, X_2 = x_2) \\ &= P(X_1 = 1)P(X_2 = 2) \\ &= f_{X_1}(x_1) \cdot f_{X_2}(x_2) \end{aligned}$$

n Independent RVs

Definition

The n RVs $X_1, \cdots, X_n$ are said to be (mutually) independent if

$$f(x_1, \cdots, x_n) = f_{X_1}(x_1) \cdot \cdots \cdot f_{X_n}(x_n),$$

where $f(x_1, \cdots, x_n)$ is the joint pmf or pdf of $X_1, \cdots, X_n$, and $f_{X_i}(x_i)$ is the marginal pmf or pdf of X_i, $i = 1, \cdots, n$.

A necessary condition for the independence of the n RVs $X_1, \cdots, X_n$ is

$$\overline{S} = \overline{S_{X_1}} \times \cdots \times \overline{S_{X_n}}.$$

Remark: If $X_1, \cdots, X_n$ are independent, then any pair of them, any triple of them, $\cdots$, any $(n-1)$ of them are also independent.

Random Sample of Size n From a Common Distribution

Definition

Independently and identically distributed (i.i.d.) RVs $X_1, X_2, \cdots, X_n$, are also called random sample of size n from a common distribution.

In this case,

$$f(x_1, \cdots, x_n) = f_X(x_1) \cdot \cdots \cdot f_X(x_n)$$

where $f_X(x)$ is the pmf or pdf of the common random distribution.

Example 2, page 190

Question

Let X_1, X_2, X_3 be a random sample of size 3 from a distribution with pdf

$$f(x) = e^{-x}, x \in (0, \infty)$$

Q1: Derive the joint pdf of X_1, X_2 and X_3?

Q2: $P(0 < X_1 < 1, \quad 2 < X_2 < 4, \quad 3 < X_3 < 7)$?

Example 2, page 190

Q1: Derive the joint pdf of X_1, X_2 and X_3?

$$g(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3) = e^{-x_1-x_2-x_3},$$

$$x_i \in (0, \infty), \quad i = 1, 2, 3.$$

Example 2, page 190

Q2: $P(0 < X_1 < 1, \quad 2 < X_2 < 4, \quad 3 < X_3 < 7)$?

$$\begin{aligned} P(0 < X_1 < 1, 2 < X_2 < 4, 3 < X_3 < 7) \\ &= P(0 < X_1 < 1)P(2 < X_2 < 4)P(3 < X_3 < 7) \\ &= \int_0^1 e^{-x_1} dx_1 \cdot \int_2^4 e^{-x_2} dx_2 \int_3^7 e^{-x_3} dx_3 \end{aligned}$$

Calculation would be much more complicated if otherwise.

Mathematical Expectation

Let $X_1, X_2, \cdots, X_n$ be multivariate RVs and have the joint pmf or pdf given by $f(x_1, x_2, \cdots, x_n), (x_1, \cdots, x_n) \in \overline{S}$. For a function $u(X_1, X_2, \cdots, X_n)$, its mathematical expectation is

$$E[u(X_1, X_2, \cdots, X_n)] = \begin{cases} \displaystyle\sum_{(x_1,\cdots,x_n)\in\overline{S}} u(x_1, \cdots, x_n) \cdot f(x_1, \cdots, x_n) & \text{“discrete RVs”} \\ \int_{\overline{S}} u(x_1, \cdots, x_n) f(x_1, \cdots, x_n) dx_1, \cdots, dx_n & \text{“continuous RV”} \end{cases}$$

Note: Mathematical Expectation is a linear operator.

Mathematical Expectation

In the case where $X_1, \cdots X_n$, are independent,

$$f(x_1, \cdots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n), \text{ and } \overline{S} = \overline{S_{X_1}} \times \cdots \times \overline{S_{X_n}}.$$

$$E[u(X_1, X_2, \cdots, X_n)] =$$
$$\begin{cases} \sum_{x_1 \in \overline{S_{X_1}}} \cdots \sum_{x_n \in \overline{S_{X_n}}} u(x_1, \cdots, x_n) \cdot f_{X_1}(x_1) \cdots f_{X_n}(x_n) & \text{discrete} \\ \int_{\overline{S_{X_1}}} \cdots \int_{\overline{S_{X_n}}} u(x_1, \cdots, x_n) \cdot f_{X_1}(x_1) \cdots f_{X_n}(x_n) dx_1, \cdots, dx_n & \text{continuous} \end{cases}$$

Theorem 5.3-1, page 191

[Theorem 5.3-1, page 191]

Assume that $X_1, X_2, \cdots, X_n$ are independent *RVs* and

$$Y = u_1(X_1)u_2(X_2)\cdots u_n(X_n)$$

If $E[u_i(X_i)], i = 1, \cdots, n$ exist, then

$$E[Y] = E[u_1(X_1)u_2(X_2)\cdots u_n(X_n)]$$

$$= E[u_1(X_1)]E[u_2(X_2)]\cdots E[u_n(X_n)]$$

Remark: This is an extension of the result that when X and Y are independent, $E(XY) = E(X)E(Y)$.

Proof Theorem 5.3-1, page 191

$X_1, X_2, \cdots, X_n$ are independent

$$\Longrightarrow \begin{cases} 1. f(x_1, \cdots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\ 2. \overline{S} = \overline{S_{X_1}} \times \cdots \times \overline{S_{X_n}} \end{cases}$$

where $f(x_1 \cdots x_n)$ is the joint pmf, $f_{X_i}(x_i)$ is the marginal pmf or pdf of $X_i, i = 1, \cdots, n$.

Proof of Theorem 5.3-1, page 191

We only consider the discrete case. (The continuous case is left as an exercise)

$$E[u_1(X_1)\cdots u_n(X_n)]$$

$$= \sum_{(x_1,x_2,\cdots,x_n)\in\overline{S}} u_1(x_1)u_2(x_2)\cdots, u_n(x_n)f(x_1,x_2,\cdots,x_n)$$

$$= \sum_{x_1\in\overline{S_{X_1}}}\sum_{x_2\in\overline{S_{X_2}}}\cdots\sum_{x_n\in\overline{S_{X_n}}} u_1(x_1)u_2(x_2)\cdots u_n(x_n)f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n)$$

$$= \sum_{x_1\in\overline{S_{X_1}}} u_1(x_1)f_{X_1}(x_1)\sum_{x_2\in\overline{S_{X_2}}} u_2(x_2)f_{X_2}(x_2)\cdots\sum_{x_n\in\overline{S_{X_n}}} u_n(x_n)f_{X_n}(x_n)$$

$$= E[u_1(X_1)]\cdot E[u_2(X_2)]\cdot\cdots\cdot E[u_n(X_n)]$$

Theorem 5.3-2, page 192

[Theorem 5.3-2, page 192]

Assume that $X_1, X_2, \cdots, X_n$ are independent RVs with respective mean $\mu_1, \mu_2, \cdots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$, respectively. Consider $Y = \sum_{i=1}^n a_i X_i$, where $a_1, a_2, \cdots, a_n$ are real constants. Then

$$E(Y) = \sum_{i=1}^n a_i \mu_i \quad \text{and} \quad Var(Y) = \sum_{i=1}^n a_i^2 \sigma_i^2 .$$

Proof of Theorem 5.3-2, page 192

$$E(Y) = E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i) = \sum_{i=1}^{n} a_i \mu_i,$$

by that expectation is a linear operator.

$$Var(Y) = E[(Y - E(Y))^2] = E\left[\left(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i \mu_i\right)^2\right]$$

$$= E\left[\left(\sum_{i=1}^{n} a_i (X_i - \mu_i)\right)^2\right] = E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (X_i - \mu_i)(X_j - \mu_j)\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E[(X_i - \mu_i)(X_j - \mu_j)] = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

Proof of Theorem 5.3-2, page 192

When $i \neq j$, since X_i and X_j are independent

$$E[(X_i - \mu_i)(X_j - \mu_j)] = 0$$

When $i = j$,

$$E[(X_i - \mu_i)(X_i - \mu_i)] = \sigma_i^2$$

Example 2, page 193

When $X_1, X_2, \cdots, X_n$ are independent and identically distributed RV with mean μ and variance σ^2. Consider

$$\overline{X} = \sum_{i=1}^{n} \frac{1}{n} X_i.$$

- $\overline{X}$ is a function of $X_1, X_2, \cdots, X_n$.
- $\overline{X}$ has the following mean and variance:

$$E(\overline{X}) = \sum_{i=1}^{n} \frac{1}{n}\mu = \mu,$$

$$Var(\overline{X}) = \sum_{i=1}^{n} (\frac{1}{n})^2 \sigma^2 = \frac{\sigma^2}{n}$$