FALL 2013
Stat 8003: Statistical Methods I
Lecture 4

Jichun Xie

1 Multivariate Random Variables

1.1 Discrete Random Variables

For a pair of discrete rv's X and Y, the joint mass function is

$$f_{X,Y}(x,y) = \mathtt{P}(X = x, Y = y).$$

In the discrete case:

1. $\mathtt{P}(X = x, Y = y) \geq 0$ for all (x, y).
2. $\sum_x \sum_y \mathtt{P}(X = x, Y = y)\, \mathrm{d}x\, \mathrm{d}y = 1$.
3. For any subset A and B containing discrete values,

$$\mathtt{P}(X \in A, Y \in B) = \sum_{x \in A, y \in B} \mathtt{P}(X = x, Y = y).$$

The two rv's could be

- two separate realization on the same process, *e.g.* ethnicity of two different individuals, usually denoted at X_1 and X_2, or Y_1 and Y_2.

- two completely different rv, *e.g.* X is ethnicity and Y is the presence of diabetes.

In this case $\mathtt{P}(X = 1, Y = 1) = 12/1100 = 0.011$. What is $\mathtt{P}(X = 1, Y = 0)$?

D /AA	Yes ($X = 1$)	No ($X = 0$)	Total
Yes ($Y = 1$)	12	40	52
No ($Y = 0$)	88	960	1048
Total	100	1000	1100

Table 1: Prevalence of Diabetes in African-American and Non-African Americans

The joint CDF of (X, Y) can be defined as

$$\mathtt{P}(X \leq a, Y \leq b) = \sum_{x \leq a} \sum_{y \leq b} \mathtt{P}(X = x, Y = y).$$

1.2 Continuous Random Variables

In the continuous case,

1. $f(x, y) \geq 0$ for all (x, y).
2. $\int_x \int_y f(x, y)\, \mathrm{d}x\, \mathrm{d}y = 1$.
3. For any sets $A, B \subset \mathbb{R}$, $\mathtt{P}(X \in A, Y \in B) = \int_{x \in A, y \in B} f(x, y)\, \mathrm{d}x\, \mathrm{d}y$.

Example. Let X and Y have joint density 1 on the interval $(0, 1)^2$, Then what is $\mathtt{P}(X < 1/2, Y < 1/2)$ and $\mathtt{P}(X < Y)$?

$$\begin{aligned}
\mathtt{P}(X < 1/2, Y < 1/2) &= \int_0^{1/2} \int_0^{1.2} f_{X,Y}\, \mathrm{d}x\, \mathrm{d}y \\
&= \int_0^{1/2} \int_0^{1/2} 1\, \mathrm{d}x\, \mathrm{d}y \\
&= \int_0^{1/2} \frac{1}{2}\, \mathrm{d}y \\
&= 1/4
\end{aligned}$$

And

$$\begin{aligned}
\mathtt{P}(X < Y) &= \int_0^1 \int_0^y f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^1 \int_0^y 1\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^1 y\,\mathrm{d}y \\
&= \frac{y^2}{2}\bigg|_0^1 \\
&= \frac{1}{2}
\end{aligned}$$

Example. Let

1.3 Marginal Distribution

For discrete random variables, suppose X and Y have bivariate distribution $f_{X,Y}(x,y)$. Then the marginal mass function for X is

$$f_X(x) = \mathtt{P}(X = x) = \sum_y \mathtt{P}(X = x, Y = y)$$

Example. In Table 1, Y is the indicator of whether a patient in the sample has diabete.

$$\begin{aligned}
\mathtt{P}(Y = 1) &= \mathtt{P}(X = 0, Y = 1) + \mathtt{P}(X = 0, Y = 1) \\
&= 40/1100 + 12/1100 = 52/1100 = 0.047
\end{aligned}$$

For continuous random variables,

$$f_X(x) = \int_y f_{X,Y}(x,y)\,\mathrm{d}y$$

Example. Suppose that for $X, Y \geq 0$, $f_{X,Y}(x,y) = \exp\{-(x+y)\}$. Then

$$f_X(x) = \int_0^\infty f_{X,Y}(x,y)\,\mathrm{d}y = e^{-x}$$

For two independent rvs,

$$\mathtt{P}(X \leq x, Y \leq y) = \mathtt{P}(X \leq x)\mathtt{P}(Y \leq y).$$

Question. Are X and Y in Table 1 independent?

1.4 Conditional Distribution

Discrete rv (when $\mathtt{P}(Y = y) > 0$):

$$\mathtt{P}(X = x \mid Y = y) = \frac{\mathtt{P}(X = x, Y = y)}{\mathtt{P}(Y = y)}$$

Continuous rv (at the point y where $f_Y(y) > 0$):

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Example. X has a Unif$(0, 1)$ distribution. Conditioning on X, Y has a Unif$(x, 1)$ distribution. What is the marginal distribution of Y?

Now let's back to the discussion of independence. Two rvs X and Y are independent if and only if

$$\mathtt{P}(X \leq x \mid Y) = \mathtt{P}(X \leq x)$$

2 Features of Distributions

2.1 Expectation

Expectation can be viewed as a weighted mean of all possible values of a random variable.

Definition.

For discrete random variable,

$$\mathtt{E}X = \sum_x x\mathtt{P}(X = x).$$

For Continuous random variable,

$$\mathtt{E}X = \int_x x f_X(x)\,\mathrm{d}x.$$

For symmetric distribution, $\mathtt{E}X$ is at the center of the range of X; for asymmetric heavy tail distribution, $\mathtt{E}X$ goes with the tail.

Example.

1. Binomial distribution: $X \sim \mathrm{Binom}(n, p)$. Then

$$\mathtt{E}X = \sum_k k\binom{n}{k}p^k(1-p)^{n-k} = np$$

2. Poisson distribution: $X \sim \mathrm{Poisson}(\lambda)$. Then

$$\mathtt{E}X = \sum_k k \cdot \frac{\lambda^k}{k!}e^{-\lambda} = \lambda$$

3. Exponential distribution: $X \sim \mathrm{Exp}(\lambda)$, with $f(x) = \lambda\exp(-\lambda x)$. Then

$$\mathtt{E}X = \int_0^\infty x f(x)\,\mathrm{d}x = \int_0^\infty \lambda x \exp(-\lambda x)\,\mathrm{d}x = \frac{1}{\lambda}$$

4. Normal distribution: $X \sim \mathrm{N}(\mu, \sigma^2)$. Then

$$\mathtt{E}X = \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,\mathrm{d}x = \mu$$

Properties of expectation.

1. for constant a, b, $\mathtt{E}(aX + b) = a\mathtt{E}X + b$.
2. if X and Y are independent, $\mathtt{E}(XY) = \mathtt{E}X\mathtt{E}Y$.

Expectation of a function of a random variable.

- Discrete: $\mathtt{E}\{g(X)\} = \sum_x g(x)\mathtt{P}(X = x)$.
- Continuous: $\mathtt{E}\{g(X)\} = \int_x g(x) f_X(x)\, \mathrm{d}x$.

Please note that usually, $f(\mathtt{E}X) \neq E(f(X))$.

Example. Suppose we toss coin A three times. At each toss, coin A faces up with probability 1/3. Let X_1 denote the number of heads that we get in the three tosses. And, suppose we toss coin B two times. At each toss, coin B faces up with probability 1/2. Let X_2 denote the number of heads we get in those two tosses. Let Y to be the number of total heads in five tosses. What is $\mathtt{E}Y$? And what is $\mathtt{E}X_1^2 X_2$?

Conditional Expectation.

- Discrete: $\mathtt{E}(X \mid Y = y) = \sum_x x\mathtt{P}(X = x \mid Y = y)$.
- Continuous: $\mathtt{E}(X \mid Y = y) = \int_x x f_{X|Y}(x \mid y)\, \mathrm{d}x$.

Let $\mathtt{E}X = \mu$. X is a rv and μ is a parameter. However, $\mathtt{E}(X \mid Y)$ dependes on Y, which is still a rv.

Example. Suppose in a population of interest, men takes up 52% of the population. The height of men follows $\mathrm{N}(\mu = 174, \sigma^2 = 25)$, and the height of women follows $\mathrm{N}(\mu = 162, \sigma^2 = 36)$. What would be the expectation of the height in the population?

2.2 Other Location Parameters

Median. For a continuous rv ν, median ν is defined as a value that satisfied

$$\mathtt{P}(X \geq \nu) = \mathtt{P}(X \leq \nu) = 1/2.$$

Population Mode. For a random variable X, the population mode γ is defined as

$$\gamma = inf\{x : f_X(t) \leq f_X(x),\ \forall\ t\}.$$

Why sometimes we consider median or mode rather than expectation? Since some distributions are highly skewed.

Example. Suppose in a population in a developing country, 999 out of 1000 persons earn 1 dollar a day, and 1 out of 1000 persons earn 100000 dollars a day. What is the expectation of population income? What is the median? What is the mode?

2.3 Variance

Variance is a measure of spread of the distribution in the population.

Definition.

$$\mathtt{Var}(X) = \mathtt{E}\{X - \mathtt{E}(X)\}^2.$$

- Discrete: $\mathtt{Var}(X) = \sum_x \{X - \mathtt{E}(X)\}^2 \mathtt{P}(X = x)$.
- Continuous: $\mathtt{Var}(X) = \int_x \{X - \mathtt{E}(X)\}^2 f_X(x)\,\mathrm{d}x$.

Properties.

1. For some constants a, b, $\texttt{Var}(aX + b) = a^2\texttt{Var}(X)$.
2. If Y_i, $i = 1, \dots, n$, are independent of each other, $\texttt{Var}(\sum_{i=1}^n Y_i) = \sum_{i=1}^n \texttt{Var}(Y_i)$.
3. For a pair of random variables X and Y,

$$\texttt{Var}(X + Y) = \texttt{Var}(X) + \texttt{Var}(Y) + 2\texttt{Cov}(X, Y),$$

where

$$\texttt{Cov}(X, Y) = \texttt{E}\{(X - \texttt{E}X)(Y - \texttt{E}Y)\} = \texttt{E}(XY) - \texttt{E}X\texttt{E}Y.$$

Why are we interested in variance? It provides a description of the precision of a statistic. For two unbiased estimates ($\texttt{E}X = \theta$, the parameter of interest) of the same parameter, a statistic with lower variance is more interesting than the one with a higher variance. In addition, sometime variance itself is of interest since it characterize the population.

Example. At one genetic location, suppose there are two possible nucleotide types, denoted as “A” and “a”. For each person, the possible genetic combination at this location is “AA”, “Aa” or “aa”. “A” is called a wide allele and “a” is called a minor allele. Let

$$X = \begin{cases} 0 & \text{if “AA”} \\ 1 & \text{if “Aa”} \\ 2 & \text{if “aa”} \end{cases}$$

Now suppose there are two genetic locations. At location 1, the allele types are “AA”, “Aa” and “aa”; at location 2, the allele types are “BB”, “Bb” and “bb”. Now suppose in the population, at one allele, $\texttt{P}(A) = 0.6$ and $\texttt{P}(B) = 0.95$. X_1 is the count of “a” alleles at location 1 and X_2 is the count of “b” alleles at location 2. What is $\texttt{E}(X_1)$, $\texttt{E}(X_2)$, $\texttt{Var}(X_1)$ and $\texttt{Var}(X_1)$? How to interprete the results?