Introductory Econometrics

Lecture 2: Review of Probability

Instructor: Ma, Jun

Renmin University of China

September 8, 2021

Randomness

- Random experiment: an experiment the outcome of which cannot be predicted with certainty, even if the experiment is repeated under the same conditions.
- Event: a collection of outcomes of a random experiment.
- Probability: a function (denoted by Pr) from events to $[0, 1]$ interval.
  - If Ω is a collection of all possible outcomes, $\Pr(\Omega) = 1$.
  - If A is an event, $\Pr(A) \geq 0$.
  - If $A_1, A_2, \ldots$ is a sequence of *disjoint* events, $\Pr(A_1 \text{ or } A_2 \text{ or } \ldots) = \Pr(A_1) + \Pr(A_2) + \ldots$.

Randomness

- Random variable: a numerical representation of a random experiment.
- Coin-flipping example:

Outcome	X	Y	Z
Heads	0	1	-1
Tails	1	0	1

- Rolling a dice:

Outcome	X	Y
1	1	0
2	2	1
3	3	0
4	4	1
5	5	0
6	6	1

Summation operator

- Let $\{x_i : i = 1, \ldots, n\}$ be a sequence of numbers.

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n.$$

- For a constant c:

$$\begin{aligned}
\sum_{i=1}^{n} c &= nc. \\
\sum_{i=1}^{n} c x_i &= c x_1 + c x_2 + \ldots + c x_n \\
&= c\left(x_1 + x_2 + \ldots + x_n\right) \\
&= c \sum_{i=1}^{n} x_i.
\end{aligned}$$

Summation operator

- Let $\{y_i : i = 1, \ldots, n\}$ be another sequence of numbers, and a, b be two constants:

$$\sum_{i=1}^{n} (ax_i + by_i) = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} y_i .$$

- But:

$$\sum_{i=1}^{n} x_i y_i \neq \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i .$$

$$\sum_{i=1}^{n} \frac{x_i}{y_i} \neq \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i} .$$

$$\sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i \right)^2 .$$

Discrete random variables

We often distinguish between discrete and continuous random variables.

- A discrete random variable takes on only a finite or countably infinite number of values.
- The distribution of a discrete random variable is a list of all possible values and the probability that each value would occur:

Value	x_1	x_2	$\ldots$	x_n
Probability	p_1	p_2	$\ldots$	p_n

Here p_i denotes the probability of a random variable X taking on value x_i:

$$p_i = \Pr[X = x_i] \text{ Probability Mass Function (PMF).}$$

Each p_i is between 0 and 1, and $\sum_{i=1}^{n} p_i = 1$.

Discrete random variables

- Indicator function:

$$1\left(x_i \leq x\right) = \begin{cases} 1 & \text{if } x_i \leq x \\ 0 & \text{if } x_i > x \end{cases}$$

- Cumulative Distribution Function (CDF):

$$F\left(x\right) = \Pr\left[X \leq x\right] = \sum_i p_i 1\left(x_i \leq x\right).$$

- For discrete random variables, the CDF is a step function.

Continuous random variable

- A random variable is continuously distributed if the range of possible values it can take is uncountable infinite (for example, a real line).
- A continuous random variable takes on any real value with zero probability.
- For continuous random variables, the CDF is continuous and differentiable.
- The derivative of the CDF is called the Probability Density Function (PDF):

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x} \text{ and } F(x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u;$$

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1.$$

Joint distribution (discrete)

- When we have two random variables X and Y, the distribution of the random vector (X, Y) is called the joint distribution and the distributions of the random variables X and Y the marginal distributions.
- Two random variables X, Y

	y_1	y_2	$\ldots$	y_m	
x_1	p_{11}	p_{12}	$\ldots$	p_{1m}	$p_1^X = \sum_{j=1}^m p_{1j}$
x_2	p_{21}	p_{22}	$\ldots$	p_{2m}	$p_2^X = \sum_{j=1}^m p_{2j}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	p_{n1}	p_{n2}	$\ldots$	p_{nm}	$p_n^X = \sum_{j=1}^m p_{nj}$

Joint PMF: $p_{ij} = \Pr\left[X = x_i, Y = y_j\right]$.
Marginal PMF: $p_i^X = \Pr[X = x_i] = \sum_{j=1}^m p_{ij}$.

Joint distribution (discrete)

- Imagine the distribution of (X, Y) (the characteristics of a ball drawn from an urn) is given by the table:

			Y		
		metal	glass	wood	
	red	$1/30$	$1/15$	$2/15$	$7/30$
X	white	$1/15$	$1/10$	$1/6$	$1/3$
	black	$1/10$	$3/10$	$1/5$	$13/30$
		$1/5$	$3/10$	$1/2$	1

- The central 3×3 table is the joint distribution. In the right "margin" is the marginal distribution of X. In the bottom margin is the marginal distribution of Y.
- Suppose we are given the joint PMF of (X, Y), to obtain the marginal PMF of X, we just "sum out" x: $\Pr[X = x_i] = \sum_{j=1}^{m} \Pr\left[X = x_i, Y = y_j\right]$.

Joint distribution (continuous)

- Joint PDF: $f_{X,Y}(x,y)$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x\mathrm{d}y = 1$.
- Marginal PDF: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y$.
- One can recover marginal PDFs from the joint PDF, but the reverse is not true. Knowing marginal PDFs does not mean that the joint PDF is also known.

Independence

- Two (discrete) random variables are independent if for all x, y:

$$\Pr[X = x, Y = y] = \Pr[X = x] \Pr[Y = y].$$

- Two continuous random variables are independent if for all x, y:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

- If X and Y are independent then $f(X)$ and $g(Y)$ are independent for all functions f, g.

Expected value

- Let g be some function:

$$\mathrm{E}[g(X)] = \sum_i g(x_i)\, p_i \text{ (discrete).}$$

$$\mathrm{E}[g(X)] = \int g(x)\, f_X(x)\, \mathrm{d}x \text{ (continuous).}$$

  Expectation is a constant!
- Mean (measuring center of a distribution):

$$\mathrm{E}[X] = \sum_i x_i p_i \text{ or } \mathrm{E}[X] = \int x f(x)\, \mathrm{d}x.$$

- Variance (measuring spread of a distribution):
  $\mathrm{Var}[X] = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right]$

$$\mathrm{Var}[X] = \sum_i (x_i - \mathrm{E}[X])^2\, p_i \text{ or } \mathrm{Var}[X] = \int (x - \mathrm{E}[X])^2 f(x)\, \mathrm{d}x.$$

- Standard deviation: $\sqrt{\mathrm{Var}(X)}$.

Joint and marginal distributions

- Suppose we want to calculate $\mathrm{E}[g(X)]$.
- We calculate:

$$\begin{aligned}\mathrm{E}[g(X)] &= \int g(x)\, f_X(x)\,\mathrm{d}x \\ &= \int g(x)\left(\int f_{X,Y}(x,y)\,\mathrm{d}y\right)\mathrm{d}x \\ &= \int\int g(x)\, f_{X,Y}(x,y)\,\mathrm{d}y\mathrm{d}x.\end{aligned}$$

The first line is the definition of $\mathrm{E}[g(X)]$. The last line is the definition of $\mathrm{E}[g(X)]$ if we think of $g(X)$ as a function of (X,Y). They must agree.

Properties

- If c is a constant, $\mathrm{E}[c] = c$, and $\mathrm{Var}[c] = \mathrm{E}\left[(c - Ec)^2\right] = (c-c)^2 = 0.$
- Linearity:

$$\begin{aligned} \mathrm{E}[a + bX] &= \sum_i (a + bx_i)\, p_i \\ &= a \sum_i p_i + b \sum_i x_i p_i \\ &= a + b\mathrm{E}[X]. \end{aligned}$$

- Suppose $X_1, ..., X_k$ are k random variables, and $a_1, ..., a_k$ are k constants, then we have $\mathrm{E}\left[\sum_{i=1}^k a_i X_i\right] = \sum_{i=1}^k a_i \mathrm{E}[X_i]$.
- Re-centering: a random variable $X - \mathrm{E}[X]$ has mean zero: $\mathrm{E}(X - \mathrm{E}[X]) = \mathrm{E}[X] - \mathrm{E}[\mathrm{E}[X]] = \mathrm{E}[X] - \mathrm{E}[X] = 0.$

Properties

- Variance formula: $\mathrm{Var}(X) = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2$

$$\begin{aligned}
\mathrm{Var}[X] &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\
&= \mathrm{E}[(X - \mathrm{E}[X])(X - \mathrm{E}[X])] \\
&= \mathrm{E}[(X - \mathrm{E}[X])X - (X - \mathrm{E}[X]) \cdot \mathrm{E}[X]] \\
&= \mathrm{E}[(X - \mathrm{E}[X])X] - \mathrm{E}[(X - \mathrm{E}[X]) \cdot \mathrm{E}[X]] \\
&= \mathrm{E}\left[X^2 - X \cdot \mathrm{E}[X]\right] - \mathrm{E}[X] \cdot \mathrm{E}[(X - EX)] \\
&= \mathrm{E}\left[X^2\right] - \mathrm{E}[X] \cdot \mathrm{E}[X] - \mathrm{E}[X] \cdot 0 \\
&= \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2
\end{aligned}$$

- If $\mathrm{E}[X] = 0$ then $\mathrm{Var}[X] = \mathrm{E}\left[X^2\right]$.

Properties

- $\mathrm{Var}[a + bX] = b^2 \mathrm{Var}[X]$

$$
\begin{aligned}
\mathrm{Var}[a + bX] &= \mathrm{E}[(a + bX) - \mathrm{E}[a + bX]]^2 \\
&= \mathrm{E}[a + bX - a - b\mathrm{E}[X]] \\
&= \mathrm{E}[bX - b\mathrm{E}[X]]^2 \\
&= \mathrm{E}\left[b^2 (X - \mathrm{E}[X])^2\right] \\
&= b^2 \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\
&= b^2 \mathrm{Var}[X].
\end{aligned}
$$

- Re-scaling: Let $\mathrm{Var}[X] = \sigma^2$, so the standard deviation is σ:

$$
\mathrm{Var}\left[\frac{X}{\sigma}\right] = \frac{1}{\sigma^2} \mathrm{Var}[X] = 1.
$$

Covariance

- Covariance: Let X, Y be two random variables.

$$\operatorname{Cov}[X, Y] = \operatorname{E}[(X - \operatorname{E}[X])(Y - \operatorname{E}[Y])].$$

$$\begin{aligned}
\operatorname{Cov}[X, Y] &= \sum_i \sum_j (x_i - \operatorname{E}[X])\left(y_j - \operatorname{E}[Y]\right) \cdot \operatorname{P}\left[X = x_i, Y = y_j\right].\\
\operatorname{Cov}[X, Y] &= \int \int (x - \operatorname{E}[X])(y - \operatorname{E}[Y])\, f_{X,Y}(x, y)\, \mathrm{d}x\mathrm{d}y.
\end{aligned}$$

- $\operatorname{Cov}[X, Y] = \operatorname{E}[XY] - \operatorname{E}[X]\operatorname{E}[Y]$.

$$\begin{aligned}
\operatorname{Cov}[X, Y] &= \operatorname{E}[(X - \operatorname{E}[X])(Y - \operatorname{E}[Y])]\\
&= \operatorname{E}[(X - \operatorname{E}[X])Y] - \operatorname{E}(X - \operatorname{E}[X]) \cdot \operatorname{E}Y\\
&= \operatorname{E}[XY] - \operatorname{E}[X]\operatorname{E}[Y].
\end{aligned}$$

Properties of covariance

- $\text{Cov}[X, c] = 0$.
- $\text{Cov}[X, X] = \text{Var}[X]$.
- $\text{Cov}[X, Y] = \text{Cov}[Y, X]$.
- $\text{Cov}[X, Y + Z] = \text{Cov}[X, Y] + \text{Cov}[X, Z]$.
- $\text{Cov}[a_1 + b_1 X, a_2 + b_2 Y] = b_1 b_2 \text{Cov}[X, Y]$.
- If X and Y are independent then $\text{Cov}[X, Y] = 0$.
- $\text{Var}[X \pm Y] = \text{Var}[X] + \text{Var}[Y] \pm 2\text{Cov}[X, Y]$.
- X and Y are independent if and only if $\text{E}[f(X) g(Y)] = \text{E}[f(X)] \text{E}[g(Y)]$ for all functions f, g.

Correlation

- Correlation coefficient:

$$\mathrm{Corr}[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}}.$$

- Cauchy-Schwartz inequality: $|\mathrm{Cov}[X,Y]| \leq \sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}$ and therefore

$$-1 \leq \mathrm{Corr}[X,Y] \leq 1.$$

- $\mathrm{Corr}[X,Y] = \pm 1 \Leftrightarrow Y = a + bX.$

Proof of the Cauchy-Schwartz Inequality*

- Claim: $|\mathrm{E}[XY]| \leq \sqrt{\mathrm{E}[X^2]\,\mathrm{E}[Y^2]}$.
  Proof: Consider the following two random variables:
  $\frac{X}{\sqrt{\mathrm{E}[X^2]}} + \frac{Y}{\sqrt{\mathrm{E}[Y^2]}}$ and $\frac{X}{\sqrt{\mathrm{E}[X^2]}} - \frac{Y}{\sqrt{\mathrm{E}[Y^2]}}$.

$$
\begin{aligned}
& \mathrm{E}\left[\left(\frac{X}{\sqrt{\mathrm{E}[X^2]}} + \frac{Y}{\sqrt{\mathrm{E}[Y^2]}}\right)^2\right] \\
= \;& \mathrm{E}\left[\frac{X^2}{\mathrm{E}[X^2]} + \frac{Y^2}{\mathrm{E}[Y^2]} + 2\frac{XY}{\sqrt{\mathrm{E}[X^2]\,\mathrm{E}[Y^2]}}\right] \\
= \;& \frac{\mathrm{E}[X^2]}{\mathrm{E}[X^2]} + \frac{\mathrm{E}[Y^2]}{\mathrm{E}[Y^2]} + 2\frac{\mathrm{E}[XY]}{\sqrt{\mathrm{E}[X^2]\,\mathrm{E}[Y^2]}} \\
= \;& 2 + 2\frac{\mathrm{E}[XY]}{\sqrt{\mathrm{E}[X^2]\,\mathrm{E}[Y^2]}} \geq 0, \text{ or} - \sqrt{\mathrm{E}[X^2]\,\mathrm{E}[Y^2]} \leq \mathrm{E}[XY].
\end{aligned}
$$

Proof of the Cauchy-Schwartz Inequality*

Similarly,

$$
\begin{aligned}
& \mathrm{E}\left[\left(\frac{X}{\sqrt{\mathrm{E}\left[X^2\right]}} - \frac{Y}{\sqrt{\mathrm{E}\left[Y^2\right]}}\right)^2\right] \\
= \quad & \frac{\mathrm{E}\left[X^2\right]}{\mathrm{E}\left[X^2\right]} + \frac{\mathrm{E}\left[Y^2\right]}{\mathrm{E}\left[Y^2\right]} - 2\frac{\mathrm{E}\left[XY\right]}{\sqrt{\mathrm{E}\left[X^2\right]\mathrm{E}\left[Y^2\right]}} \\
= \quad & 2 - 2\frac{\mathrm{E}\left[XY\right]}{\sqrt{\mathrm{E}\left[X^2\right]\mathrm{E}\left[Y^2\right]}} \geq 0, \text{ or}
\end{aligned}
$$

$$
\mathrm{E}\left[XY\right] \leq \sqrt{\mathrm{E}\left[X^2\right]\mathrm{E}\left[Y^2\right]}.
$$

Proof of the Cauchy-Schwartz Inequality*

Together:

$$-\sqrt{\mathrm{E}\left[X^2\right]\mathrm{E}\left[Y^2\right]} \le \mathrm{E}[XY] \le \sqrt{\mathrm{E}\left[X^2\right]\mathrm{E}\left[Y^2\right]},$$

or

$$|\mathrm{E}[XY]| \le \sqrt{\mathrm{E}\left[X^2\right]\mathrm{E}\left[Y^2\right]}.$$

Let $U = X - \mathrm{E}[X]$ and $V = Y - \mathrm{E}[Y]$. Then

$$|\mathrm{E}[UV]| \le \sqrt{\mathrm{E}\left[U^2\right]\mathrm{E}\left[V^2\right]}$$

or,

$$|\mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]| \le \sqrt{\mathrm{E}\left[(X - \mathrm{E}[X])^2\right]\mathrm{E}\left[(Y - \mathrm{E}[Y])^2\right]},$$

or

$$|\mathrm{Cov}[X, Y]| \le \sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}.$$

Normal distribution

- A normal random variable is a continuous random variable that can take on any value. The PDF of a normal random variable X is

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \text{ where} \\
\mu &= \mathrm{E}[X] \text{ and } \sigma^2 = \mathrm{Var}[X].
\end{aligned}
$$

  We usually write $X \sim \mathrm{N}(\mu, \sigma^2)$.
- If $X \sim \mathrm{N}(\mu, \sigma^2)$, then $a + bX \sim \mathrm{N}(a + b\mu, b^2\sigma^2)$.

Standard normal distribution

- Standard Normal random variable has $\mu = 0$ and $\sigma^2 = 1$. Its PDF is $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$.
- Symmetric around zero (mean): if $Z \sim \mathrm{N}(0, 1)$, $\Pr[Z > z] = \Pr[Z < -z]$.
- Thin tails: $\Pr[-1.96 \leq Z \leq 1.96] = 0.95$.
- If $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$, then $(X - \mu)/\sigma \sim \mathrm{N}(0, 1)$.