Lecture 2: Review of Probability

Economics 326 — Introduction to Econometrics II

Vadim Marmer, UBC

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 from events to $[0, 1]$ interval.
  - If Ω is a collection of all possible outcomes, $P(\Omega) = 1$.
  - If A is an event, $P(A) \geq 0$.
  - If $A_1, A_2, ...$ is a sequence of *disjoint* events, $P(A_1 \text{ or } A_2 \text{ or } ...) = P(A_1) + P(A_2) + ...$

Random variable

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

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

- For a constant c:

$$\sum_{i=1}^{n} c = nc.$$

$$\sum_{i=1}^{n} cx_i = c(x_1 + x_2 + ... + x_n) = c \sum_{i=1}^{n} x_i.$$

Summation operator (continued)

- Let $\{y_i : i = 1, \dots, n\}$ 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	...	x_n
Probability	p_1	p_2	...	p_n

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

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

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

Example: Bernoulli distribution

- Consider a single trial with two outcomes: "success" (with probability p) or "failure" (with probability $1 - p$).
- Define the random variable:

$$X = \begin{cases} 1 & \text{if success} \\ 0 & \text{if failure} \end{cases}$$

- Then X follows a **Bernoulli distribution**: $X \sim Bernoulli(p)$.
- PMF:

$$P(X = x) = p^x(1-p)^{1-x}, \quad x \in \{0, 1\}.$$

Discrete random variables (continued)

- Indicator function:

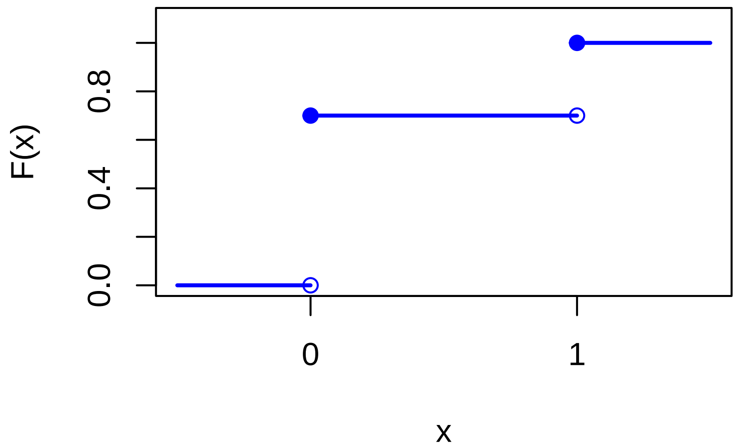
$$\mathbf{1}(x_i \leq x) = \begin{cases} 1 & \text{if } x_i \leq x \\ 0 & \text{if } x_i > x \end{cases}$$

- **Cumulative Distribution Function (CDF)**:

$$F(x) = P(X \leq x) = \sum_i p_i \mathbf{1}(x_i \leq x).$$

- $F(x)$ is non-decreasing.
- For discrete random variables, the CDF is a step function.

Example: CDF of Bernoulli(0.3)



$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1-p & \text{if } 0 \le x < 1 \\ 1 & \text{if } x \ge 1 \end{cases}$$

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)**:

$$f(x) = \frac{dF(x)}{dx} \text{ and } F(x) = \int_{-\infty}^{x} f(u)du;$$

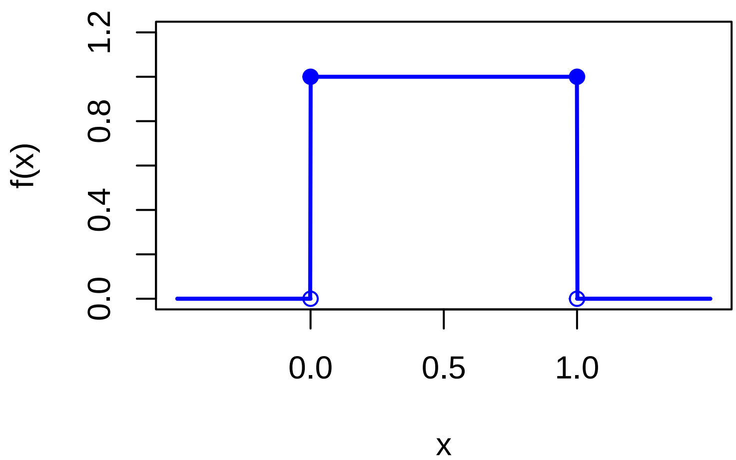
$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

- Since $F(x)$ is non-decreasing, $f(x) \ge 0$ for all x.

Example: Uniform distribution

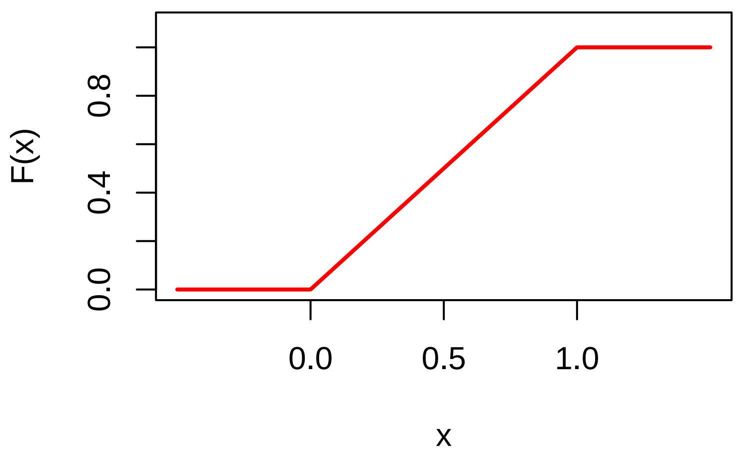
- A random variable X follows a **Uniform distribution** on $[0,1]$, written $X \sim Uniform(0,1)$, if it is equally likely to take any value in $[0,1]$.
- PDF:

$$f(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$



- CDF:

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}$$



Joint distribution (discrete)

- Two random variables X, Y

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

Joint PMF: $p_{ij} = P(X = x_i, Y = y_j)$.

Marginal PMF: $p_i^X = P(X = x_i) = \sum_{j=1}^m p_{ij}$.

- **Conditional Distribution**: If $P(X = x_1) \neq 0$,

$$p_j^{Y|X=x_1} = P(Y = y_j | X = x_1) = \frac{P(Y = y_j, X = x_1)}{P(X = x_1)} = \frac{p_{1,j}}{p_1^X}$$

Joint distribution (continuous)

- Joint PDF: $f_{X,Y}(x, y)$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1$.

- Marginal PDF: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$.
- Conditional PDF: $f_{Y|X=x}(y|x) = f_{X,Y}(x,y)/f_X(x)$.

Independence

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

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

- If independent:

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = P(Y = y).$$

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

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

- If independent, $f_{Y|X}(y|x) = f_Y(y)$ for all x.

Expected value

- Let g be some function:

$$Eg(X) = \sum_i g(x_i)p_i \text{ (discrete)}.$$

$$Eg(X) = \int g(x)f(x)dx \text{ (continuous)}.$$

  Expectation is a transformation of a distribution (PMF or PDF) and is a **constant**!

- **Mean** (center of a distribution):

$$EX = \sum_i x_i p_i \text{ or } EX = \int xf(x)dx.$$

- **Variance** (spread of a distribution): $Var(X) = E(X - EX)^2$

$$Var(X) = \sum_i (x_i - EX)^2 p_i \text{ or } Var(X) = \int (x - EX)^2 f(x)dx.$$

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

Example: Bernoulli distribution (continued)

- Recall: $X \sim Bernoulli(p)$ takes values $\{0, 1\}$ with $P(X = 1) = p$ and $P(X = 0) = 1 - p$.
- Mean:

$$E(X) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

- Variance:

$$E(X^2) = 0^2 \cdot (1 - p) + 1^2 \cdot p = p.$$

$$Var(X) = E(X^2) - (EX)^2 = p - p^2 = p(1 - p).$$

Example: Uniform distribution (continued)

- Recall: $X \sim Uniform(0,1)$ has PDF $f(x) = 1$ for $x \in [0,1]$.
- Mean:

$$E(X) = \int_0^1 x \cdot 1 \, dx = \left. \frac{x^2}{2} \right|_0^1 = \frac{1}{2}.$$

- Variance:

$$E(X^2) = \int_0^1 x^2 \cdot 1 \, dx = \left. \frac{x^3}{3} \right|_0^1 = \frac{1}{3}.$$

$$Var(X) = E(X^2) - (EX)^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Properties

- If c is a constant, $Ec = c$, and

$$Var(c) = E(c - Ec)^2 = (c - c)^2 = 0.$$

- Linearity:

$$E(a + bX) = \sum_i (a + bx_i)p_i = a \sum_i p_i + b \sum_i x_i p_i = a + bEX.$$

- Re-centering: a random variable $X - EX$ has mean zero:

$$E(X - EX) = EX - E(EX) = EX - EX = 0.$$

Properties (continued)

- Variance formula: $Var(X) = EX^2 - (EX)^2$

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

- If $EX = 0$ then $Var(X) = EX^2$.

Properties (continued)

- $Var(a + bX) = b^2 Var(X)$

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

- Re-scaling: Let $Var(X) = \sigma^2$, so the standard deviation is σ:

$$Var\left(\frac{X}{\sigma}\right) = \frac{1}{\sigma^2} Var(X) = 1.$$

Covariance

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

$$Cov(X,Y) = E[(X - EX)(Y - EY)].$$

$$Cov(X,Y) = \sum_i \sum_j (x_i - EX)(y_j - EY) \cdot P(X = x_i, Y = y_j).$$

$$Cov(X,Y) = \int \int (x - EX)(y - EY) f_{X,Y}(x,y) dx dy.$$

- $Cov(X,Y) = E(XY) - E(X)E(Y)$.

$$Cov(X,Y) = E[(X - EX)(Y - EY)] = E(XY) - EX \cdot EY.$$

Properties of covariance

- $Cov(X,c) = 0$.
- $Cov(X,X) = Var(X)$.
- $Cov(X,Y) = Cov(Y,X)$.
- $Cov(X,Y+Z) = Cov(X,Y) + Cov(X,Z)$.
- $Cov(a_1 + b_1 X, a_2 + b_2 Y) = b_1 b_2 Cov(X,Y)$.
- If X and Y are independent then $Cov(X,Y) = 0$.
- $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X,Y)$.

Correlation

- Correlation coefficient:

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}.$$

- Cauchy-Schwartz inequality: $|Cov(X,Y)| \leq \sqrt{Var(X)Var(Y)}$ and therefore

$$-1 \leq Corr(X,Y) \leq 1.$$

- $Corr(X,Y) = \pm 1 \Leftrightarrow Y = a + bX$.

Conditional expectation

- Suppose you know that $X = x$. You can update your expectation of Y by **conditional expectation**:

$$E(Y|X = x) = \sum_i y_i P(Y = y_i | X = x) \text{ (discrete)}$$

$$E(Y|X = x) = \int y f_{Y|X}(y|x) dy \text{ (continuous).}$$

- $E(Y|X = x)$ is a constant.
- $E(Y|X)$ is a function of X and is a **random variable** and a function of X (Uncertainty about X has not been realized yet):

$$E(Y|X) = \sum_i y_i P(Y = y_i | X) = g(X)$$

$$E(Y|X) = \int y f_{Y|X}(y|X) dy = g(X),$$

for some function g that depends on PMF (PDF).

Properties of conditional expectation

- Conditional expectations satisfies all properties of unconditional expectation.
- Once you condition on X, you can treat any function of X as a constant:

$$E(h_1(X) + h_2(X)Y|X) = h_1(X) + h_2(X)E(Y|X),$$

  for any functions h_1 and h_2.
- **Law of Iterated Expectation (LIE)**:

$$E[E(Y|X)] = E(Y),$$

$$E(E(Y|X, Z)|X) = E(Y|X).$$

- Conditional variance:

$$Var(Y|X) = E[(Y - E(Y|X))^2|X].$$

- **Mean independence**:

$$E(Y|X) = E(Y) = \text{constant}.$$

Relationship between different concepts of independence

$$\begin{gathered} X \text{ and } Y \text{ are independent} \\ \Downarrow \\ E(Y|X) = \text{constant (mean independence)} \\ \Downarrow \\ Cov(X, Y) = 0 \text{ (uncorrelatedness)} \end{gathered}$$

Normal distribution

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

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \text{ where}$$

$$\mu = EX \text{ and } \sigma^2 = Var(X).$$

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

Standard Normal distribution

- Standard Normal rv 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 $Z \sim N(0, 1)$, $P(Z > z) = P(Z < -z)$.
- Thin tails: $P(-1.96 \leq Z \leq 1.96) = 0.95$.
- If $X \sim N(\mu, \sigma^2)$, then $(X - \mu)/\sigma \sim N(0, 1)$.

Bivariate Normal distribution

- X and Y have a bivariate normal distribution if their joint PDF is given by:

$$f(x, y) = \frac{1}{2\pi\sqrt{(1-\rho)^2\sigma_X^2\sigma_Y^2}} \exp\left[-\frac{Q}{2(1-\rho)^2}\right],$$

  where $Q = \frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}$,

  $\mu_X = E(X), \mu_Y = E(Y), \sigma_X^2 = Var(X), \sigma_Y^2 = Var(Y)$, and $\rho = Corr(X, Y)$.

Properties of Bivariate Normal distribution

If X and Y have a bivariate normal distribution:

- $a + bX + cY \sim N(\mu^*, (\sigma^*)^2)$, where

$$\mu^* = a + b\mu_X + c\mu_Y, \quad (\sigma^*)^2 = b^2\sigma_X^2 + c^2\sigma_Y^2 + 2bc\rho\sigma_X\sigma_Y.$$

- $Cov(X,Y) = 0 \Longrightarrow X$ and Y are independent.
- $E(Y|X) = \mu_Y + \frac{Cov(X,Y)}{\sigma_X^2}(X - \mu_X)$.
- Can be generalized to more than 2 variables (multivariate normal).

Appendix: The Cauchy-Schwartz Inequality

- Claim: $|Cov(X,Y)| \leq \sqrt{Var(X)Var(Y)}$.
- **Proof**: Define

$$U = Y - \beta X,$$

  where

$$\beta = \frac{Cov(X,Y)}{Var(X)},$$

  - Note that β is a constant!
  - Also note the connection to regression and OLS in the definition of β.

- Since variances are always non-negative:

$$\begin{aligned}
0 &\leq Var(U) \\
&= Var(Y - \beta X) && \text{(def. of } U\text{)} \\
&= Var(Y) + Var(\beta X) - 2Cov(Y, \beta X) && \text{(prop. of var.)} \\
&= Var(Y) + \beta^2 Var(X) - 2\beta Cov(X,Y) && \text{(prop. of var., cov.)} \\
&= Var(Y) + \underbrace{\left(\frac{Cov(X,Y)}{Var(X)}\right)^2}_{=\beta^2} Var(X) \\
&\qquad - 2\underbrace{\left(\frac{Cov(X,Y)}{Var(X)}\right)}_{=\beta} Cov(X,Y) && \text{(def. of } \beta\text{)} \\
&= Var(Y) + \frac{Cov(X,Y)^2}{Var(X)} - 2\frac{Cov(X,Y)^2}{Var(X)} \\
&= Var(Y) - \frac{Cov(X,Y)^2}{Var(X)}.
\end{aligned}$$

- Rearranging:

$$Cov(X,Y)^2 \leq Var(X)Var(Y)$$

- or

$$|Cov(X,Y)| \leq \sqrt{Var(X)Var(Y)}.$$