Economics 326
Methods of Empirical Research in Economics

Lecture 2: Review of Probability

Vadim Marmer
University of British Columbia

January 13, 2010

Randomness

- Random experiment: an experiment the outcome of which cannot be predicted with certainty, even if the experiment is repeated under the same conditions.

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.

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 $A_1, A_2, \ldots$ is a sequence of *disjoint* events,
  $P(A_1 \text{ or } A_2 \text{ or } \ldots) = P(A_1) + P(A_2) + \ldots$.

Randomness

- Random variable: a numerical representation of a random experiment.

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

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.

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

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:

$$\begin{aligned}
\sum_{i=1}^{n} c &= nc. \\
\sum_{i=1}^{n} cx_i &= cx_1 + cx_2 + \ldots + cx_n \\
&= c\left(x_1 + x_2 + \ldots + x_n\right) \\
&= c\sum_{i=1}^{n} x_i.
\end{aligned}$$

Summation operator

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

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

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:

$$\begin{aligned} \sum_{i=1}^{n} x_i y_i &\neq \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i. \\ \sum_{i=1}^{n} \frac{x_i}{y_i} &\neq \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}. \\ \sum_{i=1}^{n} x_i^2 &\neq \left(\sum_{i=1}^{n} x_i \right)^2. \end{aligned}$$

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.

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:

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

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

Discrete random variables

- Indicator function:

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

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):

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

Discrete random variables

- Indicator function:

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

- Cumulative Distribution Function (CDF):

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

- 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).

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.

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.

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):

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

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

Joint distribution (discrete)

- Two random variables X, Y

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

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

Joint distribution (discrete)

- Two random variables X, Y

	y_1	y_2	$\dots$	y_m	
x_1	p_{11}	p_{12}	$\dots$	p_{1m}	$p_1^X = \sum_{j=1}^m p_{1j}$
x_2	p_{21}	p_{22}	$\dots$	p_{2m}	$p_2^X = \sum_{j=1}^m p_{2j}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	p_{n1}	p_{n2}	$\dots$	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 $P(X = x_1) \neq 0$,

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

Joint distribution (continuous)

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

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

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:

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

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:

$$\begin{aligned} P(Y = y \mid X = x) &= \frac{P(X = x, Y = y)}{P(X = x)} \\ &= \frac{P(X = x)\, P(Y = y)}{P(X = x)} \end{aligned}$$

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:

$$\begin{aligned} P(Y = y \mid X = x) &= \frac{P(X = x, Y = y)}{P(X = x)} \\ &= \frac{P(X = x)\, P(Y = y)}{P(X = x)} \\ &= P(Y = y). \end{aligned}$$

- Two cont. 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!

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):

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

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):

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

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

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

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):

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

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

$$Var(X) = \sum_i (x_i - EX)^2 p_i \text{ or}$$

$$Var(X) = \int (x - EX)^2 f(x)\, dx.$$

Properties

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

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

Properties

- If c is a constant, $Ec = c$, and
$$Var\left(c\right) = E\left(c - Ec\right)^2 = \left(c - c\right)^2 = 0.$$
- Linearity:
$$E\left(a + bX\right) = \sum_i \left(a + bx_i\right) p_i$$

Properties

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

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

- Linearity:

$$\begin{aligned} E\left(a + bX\right) &= \sum_i \left(a + bx_i\right) p_i \\ &= \sum_i a p_i + \sum_i b x_i p_i \end{aligned}$$

Properties

- If c is a constant, $Ec = c$, and
$$Var\left(c\right) = E\left(c - Ec\right)^2 = \left(c - c\right)^2 = 0.$$
- Linearity:
$$\begin{aligned} E\left(a + bX\right) &= \sum_i \left(a + bx_i\right) p_i \\ &= \sum_i ap_i + \sum_i bx_i p_i \\ &= a \sum_i p_i + b \sum_i x_i p_i \end{aligned}$$

Properties

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

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

- Linearity:

$$\begin{aligned} E\left(a + bX\right) &= \sum_i \left(a + bx_i\right) p_i \\ &= \sum_i ap_i + \sum_i bx_i p_i \\ &= a\sum_i p_i + b\sum_i x_i p_i \\ &= a + bEX. \end{aligned}$$

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

$$E\left(X - EX\right) = EX - E\left(EX\right) = EX - EX = 0.$$

Properties

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

Properties

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

$$Var\left(X\right) = E\left(X - EX\right)^2$$

Properties

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

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

Properties

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

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

Properties

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

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

Properties

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

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

Properties

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

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

Properties

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

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

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

Properties

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

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

Properties

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

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

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

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

Covariance

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

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

$$\begin{aligned} Cov(X, Y) &= \sum_i \sum_j (x_i - EX)(y_j - EY) \cdot P(X = x_i, Y = y_j). \\ Cov(X, Y) &= \int\int (x - EX)(y - EY) f_{X,Y}(x, y)\, dxdy. \end{aligned}$$

Covariance

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

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

$$\begin{aligned}Cov\left(X,Y\right) &= \sum_i\sum_j\left(x_i-EX\right)\left(y_j-EY\right)\cdot P\left(X=x_i,Y=y_j\right).\\ Cov\left(X,Y\right) &= \int\int\left(x-EX\right)\left(y-EY\right)f_{X,Y}\left(x,y\right)dxdy.\end{aligned}$$

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

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

Properties of covariance

- $Cov(X, c) = 0$.

Properties of covariance

- $Cov\left(X, c\right) = 0.$
- $Cov\left(X, X\right) = Var\left(X\right).$

Properties of covariance

- $Cov\left(X,c\right)=0.$
- $Cov\left(X,X\right)=Var\left(X\right).$
- $Cov\left(X,Y\right)=Cov\left(Y,X\right).$

Properties of covariance

- $Cov(X, c) = 0.$
- $Cov(X, X) = Var(X).$
- $Cov(X, Y) = Cov(Y, X).$
- $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z).$

Properties of covariance

- $Cov(X, c) = 0.$
- $Cov(X, X) = Var(X).$
- $Cov(X, Y) = Cov(Y, X).$
- $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z).$
- $Cov(a_1 + b_1 X, a_2 + b_2 Y) = b_1 b_2 Cov(X, Y).$

Properties of covariance

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

Properties of covariance

- $Cov\left(X,c\right)=0.$
- $Cov\left(X,X\right)=Var\left(X\right).$
- $Cov\left(X,Y\right)=Cov\left(Y,X\right).$
- $Cov\left(X,Y+Z\right)=Cov\left(X,Y\right)+Cov\left(X,Z\right).$
- $Cov\left(a_1+b_1X,a_2+b_2Y\right)=b_1b_2Cov\left(X,Y\right).$
- If X and Y are independent then $Cov\left(X,Y\right)=0.$
- $Var\left(X\pm Y\right)=Var\left(X\right)+Var\left(Y\right)\pm 2Cov\left(X,Y\right).$

Correlation

- Correlation coefficient:

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

Correlation

- Correlation coefficient:

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

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

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

Correlation

- Correlation coefficient:

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

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

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

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

Conditional expectation

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

$$\begin{aligned} E\left(Y|X=x\right) &= \sum_i y_i P\left(Y=y_i|X=x\right) \text{ (discrete)} \\ E\left(Y|X=x\right) &= \int y f_{Y|X}\left(y|x\right) dy \text{ (continuous).} \end{aligned}$$

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.

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:

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

for any functions h_1 and h_2.

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:

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

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

$$\begin{aligned} EE\left(Y|X\right) &= E\left(Y\right), \\ E\left(E\left(Y|X,Z\right)|X\right) &= E\left(Y|X\right). \end{aligned}$$

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:

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

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

$$\begin{aligned} EE\left(Y|X\right) &= E\left(Y\right), \\ E\left(E\left(Y|X,Z\right)|X\right) &= E\left(Y|X\right). \end{aligned}$$

- Conditional variance:

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

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:

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

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

$$\begin{aligned} EE(Y|X) &= E(Y), \\ E\left(E(Y|X,Z)\,|X\right) &= E(Y|X). \end{aligned}$$

- Conditional variance:

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

- Mean independence:

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

Relationship between different concepts of independence

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

Normal distribution

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

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

We usually write $X \sim N(\mu, \sigma^2)$.

Normal distribution

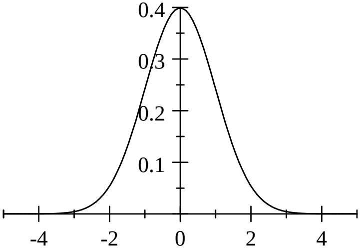
- A normal rv is a continuous rv that can take on any value. The PDF of a normal rv X is

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

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

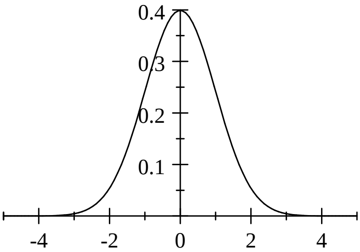
Standard Normal distribution

- Standard Normal rv has $\mu = 0$ and $\sigma^2 = 1$. Its PDF is $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$:



Standard Normal distribution

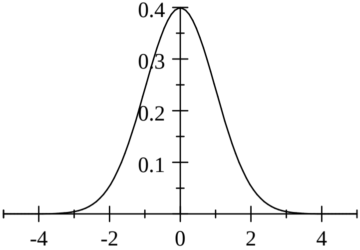
- Standard Normal rv has $\mu = 0$ and $\sigma^2 = 1$. Its PDF is $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$:



- Symmetric around zero (mean): if $Z \sim N(0, 1)$, $P(Z > z) = P(Z < -z)$.

Standard Normal distribution

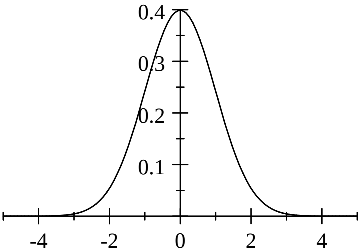
- Standard Normal rv has $\mu = 0$ and $\sigma^2 = 1$. Its PDF is $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$:



- Symmetric around zero (mean): if $Z \sim N(0, 1)$, $P(Z > z) = P(Z < -z)$.
- Thin tails: $P(-1.96 \leq Z \leq 1.96) = 0.95$.

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:

$$
\begin{aligned}
f(x, y) \; = \; & \frac{1}{2\pi\sqrt{(1-\rho)^2\sigma_X^2\sigma_Y^2}} \exp\left[-\frac{1}{2(1-\rho)^2}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \right.\right. \\
& \left.\left. + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right)\right],
\end{aligned}
$$

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

Properties of Bivariate Normal distribution

If X and Y have a bivariate normal distribution:

- $$\begin{aligned} a + bX + cY &\sim N\left(E\left(a + bX + cY\right), Var\left(a + bX + cY\right)\right) \\ &= N\left(a + b\mu_X + c\mu_Y, b^2\sigma_X^2 + c^2\sigma_Y^2 + 2bc\rho\sigma_X\sigma_Y\right) \end{aligned}$$

Properties of Bivariate Normal distribution

If X and Y have a bivariate normal distribution:

- $$\begin{aligned} a + bX + cY &\sim N\left(E\left(a + bX + cY\right), Var\left(a + bX + cY\right)\right) \\ &= N\left(a + b\mu_X + c\mu_Y, b^2\sigma_X^2 + c^2\sigma_Y^2 + 2bc\rho\sigma_X\sigma_Y\right) \end{aligned}$$
- $Cov\left(X, Y\right) = 0 \Longrightarrow X$ and Y are independent.

Properties of Bivariate Normal distribution

If X and Y have a bivariate normal distribution:

- $$\begin{aligned} a + bX + cY &\sim N\left(E\left(a + bX + cY\right), Var\left(a + bX + cY\right)\right) \\ &= N\left(a + b\mu_X + c\mu_Y, b^2\sigma_X^2 + c^2\sigma_Y^2 + 2bc\rho\sigma_X\sigma_Y\right) \end{aligned}$$
- $Cov\left(X, Y\right) = 0 \Longrightarrow X$ and Y are independent.
- $E\left(Y|X\right) = \mu_Y + \frac{Cov(X,Y)}{\sigma_X^2}\left(X - \mu_X\right)$.

Properties of Bivariate Normal distribution

If X and Y have a bivariate normal distribution:

- $$\begin{aligned} a + bX + cY &\sim N\left(E\left(a + bX + cY\right), Var\left(a + bX + cY\right)\right) \\ &= N\left(a + b\mu_X + c\mu_Y, b^2\sigma_X^2 + c^2\sigma_Y^2 + 2bc\rho\sigma_X\sigma_Y\right) \end{aligned}$$
- $Cov\left(X, Y\right) = 0 \Longrightarrow X$ and Y are independent.
- $E\left(Y|X\right) = \mu_Y + \frac{Cov(X,Y)}{\sigma_X^2}\left(X - \mu_X\right).$
- Can be generalized to more than 2 variables (multivariate normal).

Appendix: The Cauchy-Schwartz Inequality

- Claim: $|E(XY)| \leq \sqrt{EX^2 EY^2}$.

Appendix: The Cauchy-Schwartz Inequality

- Claim: $|E(XY)| \leq \sqrt{EX^2 EY^2}$.

Proof: Consider the following two random variables: $\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}$ and $\frac{X}{\sqrt{EX^2}} - \frac{Y}{\sqrt{EY^2}}$.

Appendix: The Cauchy-Schwartz Inequality

- Claim: $|E(XY)| \leq \sqrt{EX^2 EY^2}$.

Proof: Consider the following two random variables: $\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}$ and $\frac{X}{\sqrt{EX^2}} - \frac{Y}{\sqrt{EY^2}}$.

$$E\left(\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}\right)^2 =$$

Appendix: The Cauchy-Schwartz Inequality

- Claim: $|E(XY)| \leq \sqrt{EX^2 EY^2}$.

Proof: Consider the following two random variables: $\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}$ and $\frac{X}{\sqrt{EX^2}} - \frac{Y}{\sqrt{EY^2}}$.

$$
\begin{aligned}
E\left(\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}\right)^2 &= E\left(\frac{X^2}{EX^2} + \frac{Y^2}{EY^2} + 2\frac{XY}{\sqrt{EX^2 EY^2}}\right) \\
&=
\end{aligned}
$$

Appendix: The Cauchy-Schwartz Inequality

- Claim: $|E(XY)| \leq \sqrt{EX^2 EY^2}$.

Proof: Consider the following two random variables: $\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}$ and $\frac{X}{\sqrt{EX^2}} - \frac{Y}{\sqrt{EY^2}}$.

$$
\begin{aligned}
E\left(\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}\right)^2 &= E\left(\frac{X^2}{EX^2} + \frac{Y^2}{EY^2} + 2\frac{XY}{\sqrt{EX^2 EY^2}}\right) \\
&= \frac{EX^2}{EX^2} + \frac{EY^2}{EY^2} + 2\frac{E(XY)}{\sqrt{EX^2 EY^2}} \\
&=
\end{aligned}
$$

Appendix: The Cauchy-Schwartz Inequality

- Claim: $|E(XY)| \leq \sqrt{EX^2 EY^2}$.

Proof: Consider the following two random variables: $\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}$ and $\frac{X}{\sqrt{EX^2}} - \frac{Y}{\sqrt{EY^2}}$.

$$
\begin{aligned}
E\left(\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}\right)^2 &= E\left(\frac{X^2}{EX^2} + \frac{Y^2}{EY^2} + 2\frac{XY}{\sqrt{EX^2 EY^2}}\right) \\
&= \frac{EX^2}{EX^2} + \frac{EY^2}{EY^2} + 2\frac{E(XY)}{\sqrt{EX^2 EY^2}} \\
&= 2 + 2\frac{E(XY)}{\sqrt{EX^2 EY^2}}
\end{aligned}
$$

Appendix: The Cauchy-Schwartz Inequality

- Claim: $|E(XY)| \leq \sqrt{EX^2 EY^2}$.

Proof: Consider the following two random variables: $\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}$ and $\frac{X}{\sqrt{EX^2}} - \frac{Y}{\sqrt{EY^2}}$.

$$
\begin{aligned}
E\left(\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}\right)^2 &= E\left(\frac{X^2}{EX^2} + \frac{Y^2}{EY^2} + 2\frac{XY}{\sqrt{EX^2 EY^2}}\right) \\
&= \frac{EX^2}{EX^2} + \frac{EY^2}{EY^2} + 2\frac{E(XY)}{\sqrt{EX^2 EY^2}} \\
&= 2 + 2\frac{E(XY)}{\sqrt{EX^2 EY^2}} \geq 0,
\end{aligned}
$$

Appendix: The Cauchy-Schwartz Inequality

- Claim: $|E(XY)| \leq \sqrt{EX^2 EY^2}$.

Proof: Consider the following two random variables: $\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}$ and $\frac{X}{\sqrt{EX^2}} - \frac{Y}{\sqrt{EY^2}}$.

$$
\begin{aligned}
E\left(\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}\right)^2 &= E\left(\frac{X^2}{EX^2} + \frac{Y^2}{EY^2} + 2\frac{XY}{\sqrt{EX^2 EY^2}}\right) \\
&= \frac{EX^2}{EX^2} + \frac{EY^2}{EY^2} + 2\frac{E(XY)}{\sqrt{EX^2 EY^2}} \\
&= 2 + 2\frac{E(XY)}{\sqrt{EX^2 EY^2}} \geq 0, \text{ or}
\end{aligned}
$$

Appendix: The Cauchy-Schwartz Inequality

- Claim: $|E(XY)| \leq \sqrt{EX^2 EY^2}$.

Proof: Consider the following two random variables: $\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}$ and $\frac{X}{\sqrt{EX^2}} - \frac{Y}{\sqrt{EY^2}}$.

$$
\begin{aligned}
E\left(\frac{X}{\sqrt{EX^2}} + \frac{Y}{\sqrt{EY^2}}\right)^2 &= E\left(\frac{X^2}{EX^2} + \frac{Y^2}{EY^2} + 2\frac{XY}{\sqrt{EX^2 EY^2}}\right) \\
&= \frac{EX^2}{EX^2} + \frac{EY^2}{EY^2} + 2\frac{E(XY)}{\sqrt{EX^2 EY^2}} \\
&= 2 + 2\frac{E(XY)}{\sqrt{EX^2 EY^2}} \geq 0, \text{ or}
\end{aligned}
$$

$$
-\sqrt{EX^2 EY^2} \leq E(XY).
$$

Appendix: The Cauchy-Schwartz Inequality

Similarly,

$$E\left(\frac{X}{\sqrt{EX^2}} - \frac{Y}{\sqrt{EY^2}}\right)^2 \quad =$$

Appendix: The Cauchy-Schwartz Inequality

Similarly,

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

$$E(XY) \leq \sqrt{EX^2EY^2}.$$

Appendix: The Cauchy-Schwartz Inequality

Together:

$$-\sqrt{EX^2 EY^2} \leq E\left(XY\right) \leq \sqrt{EX^2 EY^2},$$

Appendix: The Cauchy-Schwartz Inequality

Together:

$$-\sqrt{EX^2EY^2} \leq E(XY) \leq \sqrt{EX^2EY^2},$$

or

$$|E(XY)| \leq \sqrt{EX^2EY^2}.$$

Let $U = X - EX$ and $V = Y - EY$.

Appendix: The Cauchy-Schwartz Inequality

Together:

$$-\sqrt{EX^2EY^2} \leq E\left(XY\right) \leq \sqrt{EX^2EY^2},$$

or

$$\left|E\left(XY\right)\right| \leq \sqrt{EX^2EY^2}.$$

Let $U = X - EX$ and $V = Y - EY$. Then,

$$\left|E\left(UV\right)\right| \leq \sqrt{EU^2EV^2},$$

Appendix: The Cauchy-Schwartz Inequality

Together:

$$-\sqrt{EX^2 EY^2} \leq E(XY) \leq \sqrt{EX^2 EY^2},$$

or

$$|E(XY)| \leq \sqrt{EX^2 EY^2}.$$

Let $U = X - EX$ and $V = Y - EY$. Then,

$$|E(UV)| \leq \sqrt{EU^2 EV^2},$$

or

$$|E\{(X - EX)(Y - EY)\}| \leq \sqrt{E(X - EX)^2 E(Y - EY)^2},$$

Appendix: The Cauchy-Schwartz Inequality

Together:

$$-\sqrt{EX^2 EY^2} \leq E(XY) \leq \sqrt{EX^2 EY^2},$$

or

$$|E(XY)| \leq \sqrt{EX^2 EY^2}.$$

Let $U = X - EX$ and $V = Y - EY$. Then,

$$|E(UV)| \leq \sqrt{EU^2 EV^2},$$

or

$$|E\{(X - EX)(Y - EY)\}| \leq \sqrt{E(X - EX)^2 E(Y - EY)^2},$$

or

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