Random Variables

Random Variables

Definition *A random variable $X(\omega)$ for $\omega \in \Omega$ is a real-valued function $X : \Omega \to R$.*

A discrete random variable takes on only a finite or countably infinite number of values. A continuous variable takes uncountably infinite number of values.

- *Discrete RVs have probability mass functions*
- *Continuous RVs have probability density functions.*

- *To find the pmf of a discrete RV we need:*
- *The support of the RV. (possible values x)*
- *The probabilities $P(X = x)$ for every x in the support.*

Example

What is the distribution of the sum of two dice?

- *The support of the RV. (possible values x):*

- *The probabilities $P(X = x)$ for every x in the support.:*

Joint Distributions

x

	1	2	3	4	
3	1/20	2/20	2/20	2/20	7/20
2	2/20	4/20	1/20	1/20	8/20
y 1	0/20	2/20	0/20	0	2/20
0	0/20	1/20	2/20	0	3/20
	3/20	9/20	5/20	3/20	

X:Grade
Y:Number of times in NHMI

Joint Distribution:
$P_{X,Y}(x, y) = P(X = x, Y = y)$ for all x, y

$$\sum_x \sum_y P_{X,Y}(x, y) = 1$$

Marginal Distribution:
$P_X(x) = \sum_y P_{X,Y}(x, y)$ for all x

Conditional :
$P_{X|Y}(x| y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}$ for all x

Joint Distributions

x

y		1	2	3	4	
	3	1/20	2/20	2/20	2/20	7/20
	2	2/20	4/20	1/20	1/20	8/20
y	1	0/20	2/20	0/20	0	3/20
	0	0/20	1/20	2/20	0	3/20
		3/20	9/20	5/20	2/20	

X:Grade
Y:Number of times in NHMI

Joint Distribution:
$P_{X,Y}(x, y) = P(X = x, Y = y)$ for all x, y

$$\sum_x \sum_y P_{X,Y}(x, y) = 1$$

Marginal Distribution:
$P_X(\, x) = \sum_y P_{X,Y}(x, y)$ for all x

Conditional :
$P_{X|Y}(x|\, y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}$ for all x

Random Variables

Given two RVs X and Y the **conditional distribution** is:

$$p(X \mid Y) = \frac{p(X,Y)}{p(Y)} = \frac{p(X,Y)}{\sum_x p(X=x,Y)}$$

Multiply both sides by $p(Y)$ to obtain the **probability chain rule**:

$$p(X,Y) = p(Y)p(X \mid Y)$$

For N RVs $X_1, X_2, \ldots, X_N$:

$$p(X_1, X_2, \ldots, X_N) = p(X_1)p(X_2 \mid X_1) \ldots p(X_N \mid X_{N-1}, \ldots, X_1)$$

$$= p(X_1) \prod_{i=2}^{N} p(X_i \mid X_{i-1}, \ldots, X_1)$$

Chain rule valid for any ordering

Fundamental Rules of Probability

Law of total probability

$$p(Y) = \sum_x p(Y, X = x)$$

Proof

$$\sum_x p(Y, X = x) = \sum_x p(Y)p(X = x \mid Y) \quad \text{(chain rule)}$$

$$= p(Y) \sum_x p(X = x \mid Y) \quad \text{(distributive property)}$$

$$= p(Y) \quad \text{(axiom of probability)}$$

Generalization for conditionals:

$$p(Y \mid Z) = \textstyle\sum_x p(Y, X = x \mid Z)$$

Independence of RVs

__Question:__ Roll two dice and let their outcomes be $X_1, X_2 \in \{1, \ldots, 6\}$ *for die 1 and die 2, respectively. Recall the definition of conditional probability,*

$$p(X_1 \mid X_2) = \frac{p(X_1, X_2)}{p(X_2)}$$

Which of the following are true?

a) $p(X_1 = 1 | X_2 = 1) > p(X_1 = 1)$

b) $p(X_1 = 1 | X_2 = 1) = p(X_1 = 1)$ Outcome of die 2 doesn't *affect* die 1

c) $p(X_1 = 1 | X_2 = 1) < p(X_1 = 1)$

Independence of RVs

<u>*Question:*</u> *Let* $X_1 \in \{1, \ldots, 6\}$ *be outcome of die 1, as before. Now let* $X_3 \in \{2, 3, \ldots, 12\}$ *be the* sum of both dice. Which of the following are true?

a) $p(X_1 = 1 | X_3 = 3) > p(X_1 = 1)$

b) $p(X_1 = 1 | X_3 = 3) = p(X_1 = 1)$

c) $p(X_1 = 1 | X_3 = 3) < p(X_1 = 1)$

Only 2 ways to get $X_3 = 3$ *, each with equal probability:*

$(X_1 = 1, X_2 = 2)$ *or* $(X_1 = 2, X_2 = 1)$

so

$$p(X_1 = 1 \mid X_3 = 3) = \frac{1}{2} > \frac{1}{6} = p(X_1 = 1)$$

Independence of RVs

Definition *Two random variables* X *and* Y *are* *independent* *if and only if,*

$$p(X = x, Y = y) = p(X = x)p(Y = y)$$

for all values x and y, and we say $X \perp Y$.

Definition RVs $X_1, X_2, \ldots, X_N$ are mutually independent if and only if,

$$p(X_1 = x_1, \ldots, X_N = x_N) = \prod_{i=1}^{N} p(X_i = x_i)$$

- Independence is *symmetric:* $X \perp Y \Leftrightarrow Y \perp X$
- Equivalent definition of independence: $p(X \mid Y) = p(X)$

Independence of RVs

- *N RVs are <u>pairwisely independent</u> if and only if:*
- $P\left(X_i, X_j\right) = P(X_i)P\left(X_j\right)$ *for every pair* i, j.
- *Pairwise independence* $\nrightarrow$ *Mutual Independence*
- $X = 1\ if\ coin\ 1\ is\ heads, 0\ otherwise$
- $Y = 1\ if\ coin\ 2\ is\ heads, 0\ otherwise$
- $Z = X\ (XOR)\ Y$

Independence of RVs

Definition *Two random variables* X *and* Y *are* *<u>conditionally independent</u>* *given* Z *if and only if,*

$$p(X = x, Y = y \mid Z = z) = p(X = x \mid Z = z)p(Y = y \mid Z = z)$$

for all values x, y, and z, and we say that $X \perp Y \mid Z$.

- N RVs conditionally independent, given Z, if and only if:

$$p(X_1, \ldots, X_N \mid Z) = \prod_{i=1}^{N} p(X_i \mid Z)$$

Shorthand notation
Implies for all *x, y, z*

- Equivalent def'n of conditional independence: $p(X \mid Y, Z) = p(X \mid Z)$
- Symmetric: $X \perp Y \mid Z \Leftrightarrow Y \perp X \mid Z$

Conditional Independence

x

y	1	2	3	4	
3	1/20	2/20	2/20	2/20	7/20
2	2/20	4/20	1/20	1/20	8/20

X:Grade
Y:Number of times in NHMI

Let's say you know that the kid has visited the museum more than once.

Are X and Y independent in this new universe?

Functions of RVs

Functions of RVs are also RVs

Let's say X has the following distribution:

$$P_X(x) = \frac{x^2}{a}, x \in \{-3, -2, -1, 1, 2, 3\}$$

1. *Find the pmf of* X
2. *Find the pmf of* $Z = X^2$

Practice

- *A fair coin is tossed 3 times*
- X: *The number of heads on the first toss*
- Y: *Total number of heads*

- *Find*
  - *The joint distribution of* X, Y
  - *The marginal distributiions of* X, Y
  - *The conditional distributions for* $Y|X = 1, X|Y = 3$
  - *Are they independent?*

Expected value of a discrete random variable

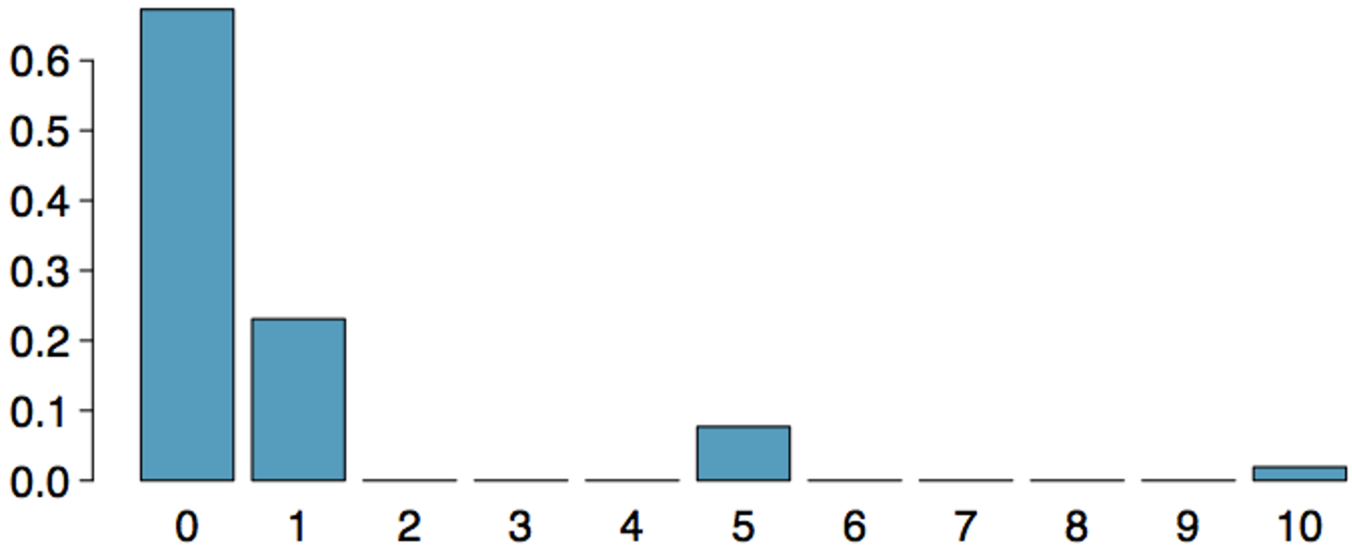
In a game of cards you win \$1 if you draw a heart, \$5 if you draw an ace (including the ace of hearts), \$10 if you draw the king of spades and nothing for any other card you draw.

Write the probability model for your winnings, and calculate your expected winning.

Event	X	$P(X)$
Heart (not ace)	1	$\frac{12}{52}$
Ace	5	$\frac{4}{52}$
King of spades	10	$\frac{1}{52}$
All else	0	$\frac{35}{52}$
Total		

Expected value of a discrete random variable (cont.)

Below is a visual representation of the probability distribution of winnings from this game:



Expected value of a discrete random variable

In a game of cards you win \$1 if you draw a heart, \$5 if you draw an ace (including the ace of hearts), \$10 if you draw the king of spades and nothing for any other card you draw.

Write the probability model for your winnings, and calculate your expected winning.

Event	X	$P(X)$	$X\,P(X)$
Heart (not ace)	1	$\frac{12}{52}$	$\frac{12}{52}$
Ace	5	$\frac{4}{52}$	$\frac{20}{52}$
King of spades	10	$\frac{1}{52}$	$\frac{10}{52}$
All else	0	$\frac{35}{52}$	0
Total			$E(X) = \frac{42}{52} \approx 0.81$

Expectation

Definition *The <u>expectation</u> of a discrete RV* X*, denoted by* $\mathbf{E}[X]$*, is:*

$$\mathbf{E}[X] = \sum_x x\, p(X = x)$$

Theorem (Linearity of Expectations) *For any finite collection of discrete RVs* $X_1, \ldots, X_n$ *with finite expectations,*

$$\mathbf{E}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathbf{E}[X_i]$$

Corollary *For any constant c*

$$\mathbf{E}[cX] = c\mathbf{E}[X]$$

Theorem: *If* $X \perp Y$ *then* $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$*.*

Conditional Expectation

Definition *The <u>conditional expectation</u> of a discrete RV* X*, given* Y *is:*

$$\mathbf{E}[X \mid Y = y] = \sum_{x} x\, p(X = x \mid Y = y)$$

Example Roll two standard six-sided dice and let X be the result of the first die and let Y be the sum of both dice, then:

$$\mathbf{E}[X_1 \mid Y = 5] = \sum_{x=1}^{4} x\, p(X_1 = x \mid Y = 5)$$

$$= \sum_{x=1}^{4} x \quad \frac{\quad}{p(Y = 5)} \quad = \sum_{x=1}^{4} x \frac{\quad}{4/36} = \frac{5}{2}$$

$E[X|Y = y]$ is a number

$E[X|Y]$ is a function

Conditional expectation follows properties of expectation (linearity, etc.)

Law of Total Expectation

Law of Total Expectation *Let X and Y be discrete RVs with finite expectations, then:*

$$\mathbf{E}[X] = \mathbf{E}_Y[\mathbf{E}_X[X \mid Y]]$$

Proof

$$\begin{aligned}
\mathbf{E}_Y[\mathbf{E}_X[X \mid Y]] &= \mathbf{E}_Y\left[\sum_x x \cdot p(x \mid Y)\right] \\
&= \sum_y \left[\sum_x x \cdot p(x \mid y)\right] \cdot p(y) && \textbf{(Definition of expectation)} \\
&= \sum_y \sum_x x \cdot p(x, y) && \textbf{(Probability chain rule)} \\
&= \sum_x x \sum_y \cdot p(x, y) && \textbf{(Linearity of expectations)} \\
&= \sum_x x \cdot p(x) = \mathbf{E}[X] && \textbf{(Law of total probability)}
\end{aligned}$$

Law of the Unconscious Statistician

- *If* $Y = g(X)$,
- $E[Y] = \sum_x g(x) P_X(x)$

Proof:

$$E[Y] = \sum_y y P_Y(y) =$$

$$\sum_y y P\left(x = g^{-1}(y)\right) =$$

$$\sum_y y \sum_{x=g^{-1}(y)} P_X(x) =$$

$$\sum_y \sum_{x=g^{-1}(y)} g(x) P_X(x) =$$

$$\sum_x g(x) P_X(x)$$

Variance

We are also often interested in the variability in the values of a random variable.

Definition *The variance of a RV   is defined as,*

$$\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$$

$$\sigma[X] = \sqrt{\mathbf{Var}[X]}$$

Lemma An equivalent form of variance is:

$$\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$$

Variance of a discrete random variable

For the previous card game example, how much would you expect the winnings to vary from game to game?

Variance of a discrete random variable

For the previous card game example, how much would you expect the winnings to vary from game to game?

X	$P(X)$	$X\,P(X)$	$(X - E(X))^2$	$P(X)\,(X - E(X))^2$
1	$\frac{12}{52}$	$1 \times \frac{12}{52} = \frac{12}{52}$	$(1 - 0.81)^2 = 0.0361$	$\frac{12}{52} \times 0.0361 = 0.0083$
5	$\frac{4}{52}$	$5 \times \frac{4}{52} = \frac{20}{52}$	$(5 - 0.81)^2 = 17.5561$	$\frac{4}{52} \times 17.5561 = 1.3505$
10	$\frac{1}{52}$	$10 \times \frac{1}{52} = \frac{10}{52}$	$(10 - 0.81)^2 = 84.4561$	$\frac{1}{52} \times 84.0889 = 1.6242$
0	$\frac{35}{52}$	$0 \times \frac{35}{52} = 0$	$(0 - 0.81)^2 = 0.6561$	$\frac{35}{52} \times 0.6561 = 0.4416$
		$E(X) = 0.81$		

Variance of a discrete random variable

For the previous card game example, how much would you expect the winnings to vary from game to game?

X	$P(X)$	$X\,P(X)$	$(X - E(X))^2$	$P(X)\,(X - E(X))^2$
1	$\frac{12}{52}$	$1 \times \frac{12}{52} = \frac{12}{52}$	$(1 - 0.81)^2 = 0.0361$	$\frac{12}{52} \times 0.0361 = 0.0083$
5	$\frac{4}{52}$	$5 \times \frac{4}{52} = \frac{20}{52}$	$(5 - 0.81)^2 = 17.5561$	$\frac{4}{52} \times 17.5561 = 1.3505$
10	$\frac{1}{52}$	$10 \times \frac{1}{52} = \frac{10}{52}$	$(10 - 0.81)^2 = 84.4561$	$\frac{1}{52} \times 84.0889 = 1.6242$
0	$\frac{35}{52}$	$0 \times \frac{35}{52} = 0$	$(0 - 0.81)^2 = 0.6561$	$\frac{35}{52} \times 0.6561 = 0.4416$
		$E(X) = 0.81$		$V(X) = 3.4246$ $SD(X) = \sqrt{3.4246} = 1.85$

Variance of Sums of Distributions

Question: *What is the variance of their sum?*

$$\begin{aligned}\mathbf{Var}[X_1 + X_2] &= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{Cov}(X_1, X_2) \\ &= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{E}[(X_1 - \mathbf{E}[X_1])(X_2 - \mathbf{E}[X_2])] \\ &= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{E}[(X_1 - \mathbf{E}[X_1])]\mathbf{E}[(X_2 - \mathbf{E}[X_2])] \\ &= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\left(\mathbf{E}[X_1] - \mathbf{E}[X_1]\right)\left(\mathbf{E}[X_2] - \mathbf{E}[X_2]\right) \\ &= \mathbf{Var}[X_1] + \mathbf{Var}[X_2]\end{aligned}$$

Theorem: *If* $X \perp Y$ *then* $\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y]$

Corollary: *If* $X \perp Y$ *then* $\mathbf{Cov}(X, Y) = 0$

Corollary: *For collection of RVs* $X_1, X_2, \ldots, X_N$ *:* $\mathbf{Var}(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} \mathbf{Var}(X_i)$

Covariance

Definition *The covariance of two RVs* X *and* Y *is defined as,*

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

Lemma *For any two RVs* X *and* Y*,*

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X,Y)$$

e.g. variance is not a linear operator.

Proof $\mathbf{Var}[X+Y] = \mathbf{E}[(X+Y-\mathbf{E}[X+Y])^2]$

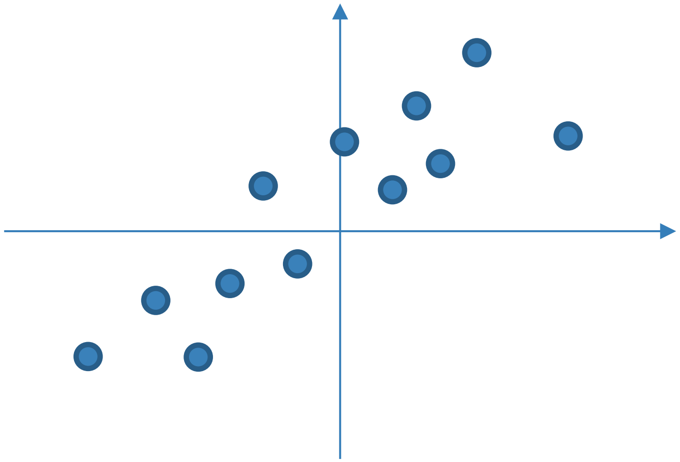
(Linearity of expectation) $= \mathbf{E}[(X+Y-\mathbf{E}[X]-\mathbf{E}[Y])^2]$

(Distributive property) $= \mathbf{E}[(X-\mathbf{E}[X])^2 + (Y-\mathbf{E}[Y])^2 + 2(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]$

(Linearity of expectation) $= \mathbf{E}[(X-\mathbf{E}[X])^2] + \mathbf{E}[(Y-\mathbf{E}[Y])^2] + 2\mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]$

(Definition of Var / Cov) $= \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X,Y)$

Covariance



Assume zero means

If $XY > 0 \Rightarrow E(XY) > 0$
If $XY < 0 \Rightarrow E(XY) < 0$

If you have equal number of points in all quarters?

If X, Y are independent:

$$E[(X - E[X])(Y - E[Y])] = E\big[X - E[X]\big]E\big[Y - E[Y]\big] = 0$$

The opposite is not necessarily true

Example:
$X \in \{-1, 0, 1\}, P(X = x) = \frac{1}{3}$ for all x
Find the covariance of X, X^2
Are X, X^2 dependent?

Correlation

- X: *Distribution of heights of kids in Greece measured in centimeters.*
- Y: *Distribution of heights of kids in Greece measured in meters.*
- Z: *Distribution of head circumference in Greece measured in centimeters*

- *Which pair has larger covariance:* X, Z *or* Y, Z?

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