Chapter 3: Probability

Yonghyun Kwon
Department of Mathematics, Korea Military Academy

Defining probability

Random processes

- A *random process* is a situation in which we know what outcomes could happen, but we don't know which particular outcome will happen.
- Examples: coin tosses, die rolls, whether the stock market goes up or down tomorrow, etc.

MP3 Players > Stories > **iTunes: Just how random is random?**

iTunes: Just how random is random?

By David Braue on 08 March 2007

- **Introduction**
- **Say You, Say What?**
- **A role for labels?**
- **The new random**

Think that song has appeared in your playlists just a few too many times? David Braue puts the randomness of Apple's song shuffling to the test -- and finds some surprising results.

Quick -- think of a number between one and 20. Now think of another one, and another, and another.

Starting to repeat yourself? No surprise: in practice, many series of random numbers are far less random than you would think.

Computers have the same problem. Although all systems are able to pick random numbers, the method they use is often tied to specific other numbers -- for example, the time -- that means you could get a very similar series of 'random' numbers in different situations.

This tendency manifests itself in many ways. For anyone who uses their iPod heavily, you've probably noticed that your supposedly random 'shuffling' iPod seems to be particularly fond of the Bee Gees, Melissa Etheridge or Pavarotti. Look at a random playlist that iTunes generates for you, and you're likely to notice several songs from one or two artists, while other artists go completely unrepresented.

http://www.cnet.com.au/

itunes-just-how-random-is-random-339274094.htm

Probability

- There are several possible interpretations of probability but they (almost) completely agree on the mathematical rules probability must follow.
  - $P(A)$ = Probability of event A
  - $0 \leq P(A) \leq 1$
- The *probability* of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.

Practice

Which of the following events would you be most surprised by?

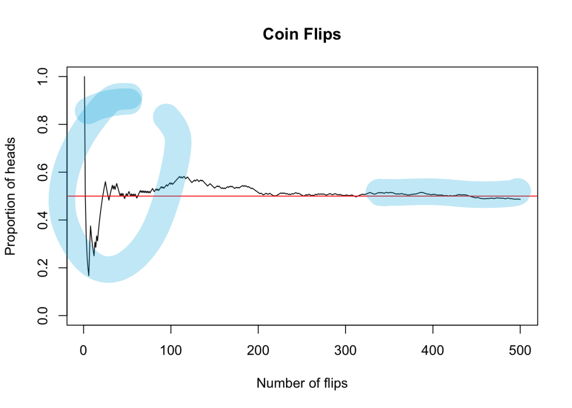
(a) exactly 3 heads in 10 coin flips

(b) exactly 3 heads in 100 coin flips

(c) exactly 3 heads in 1000 coin flips

Law of large numbers

Law of large numbers states that as more observations are collected, the proportion of occurrences with a particular outcome, $\hat{p}_n$, converges to the probability of that outcome, p.



Disjoint and non-disjoint outcomes

Disjoint (mutually exclusive) outcomes: Cannot happen at the same time.

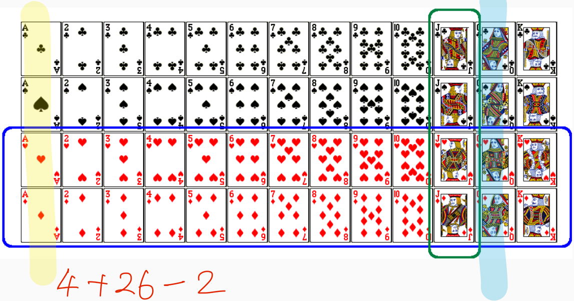
- The outcome of a single coin toss cannot be a head and a tail.
- A student both cannot fail and pass a class.
- A single card drawn from a deck cannot be an ace and a queen.

Non-disjoint outcomes: Can happen at the same time.

- A student can get an A in Stats and A in Econ in the same semester.

Union of non-disjoint events

What is the probability of drawing a jack or a red card from a well shuffled full deck? $13 \times 4 = 52$



$$\frac{4+26-2}{52}$$

Figure from *http://www.milefoot.com/math/discrete/counting/cardfreq.htm*.

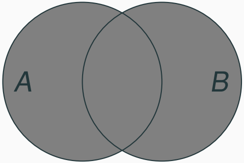
General addition rule

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

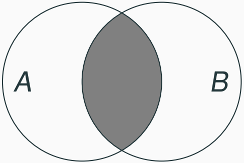
Note: *For disjoint events* $P(A \text{ and } B) = 0$*, so the above formula simplifies to* $P(A \text{ or } B) = P(A) + P(B)$.

Note: *Union of two events* A *and* B: $A \text{ or } B = A \cup B$.

Intersection of two events A *and* B: $A \text{ and } B = A \cap B = AB$.



$A \cup B$



$A \cap B$

Sample space and complements

Sample space(S) is the collection of all possible outcomes of a trial.

- A couple has one kid, what is the sample space for the gender of this kid? $S = \{M, F\}$
- A couple has two kids, what is the sample space for the gender of these kids? $S =$ $\{MM, MF, FM, FF\}$

The *Complement* of an event, A^c, is the event that A does not occur.

- A couple has one kid. If we know that the kid is not a boy, what is gender of this kid? $A = \{M\} \rightarrow A^c = \{F\}$
- A couple has two kids, if we know that they are not both girls, what are the possible gender combinations for these kids?

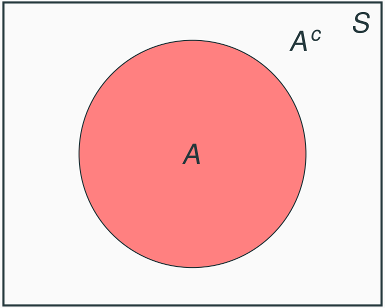
$A = \{FF\}$, $A^c = \{FM, MF, MM\}$

The complement of set A within the sample space S

Complement

A and A^c are mathematically related as:

$$P(A) + P(A^c) = 1, \quad \text{i.e.} P(A) = 1 - P(A^c).$$



Independence

Two processes are *independent* if knowing the outcome of one provides no useful information about the outcome of the other.

- The outcome of one coin toss does not affect the next. → Coin tosses are independent
- The outcome of drawing a card (*without replacement*) affects the next draw. → Card draws (*without replacement*) are dependent

Checking for independence

If P(A occurs, given that B is true) = $P(A \mid B) = P(A)$, then A and B are independent.

Product rule for independent events

$$P(A \text{ and } B) = P(A) \times P(B)$$

Or more generally, $P(A_1 \text{ and } \cdots \text{ and } A_k) = P(A_1) \times \cdots \times P(A_k)$

You toss a coin twice, what is the probability of getting two tails in a row?

$$P(T_1 T_2) = \underbrace{P(T_1)}_{1/2} \underbrace{P(T_2)}_{1/2}$$

$$P(\text{T on the first toss}) \times P(\text{T on the second toss}) = \frac{1}{4}$$

Disjoint vs. complementary

Do the sum of probabilities of two disjoint events always add up to 1?

No

Do the sum of probabilities of two complementary events always add up to 1?

Yes

Probability of "at least one"

To find the probability of at least one of something, calculate the probability of none and then subtract that result from 1.

$$P(\textit{at least one}) = 1 - P(\textit{none})$$

Practice

Roughly 20% of undergraduates at a university are vegetarian or vegan. What is the probability that, among a random sample of 3 undergraduates, at least one is vegetarian or vegan?

(a) $1 - 0.2 \times 3$

(b) $1 - 0.2^3$

(c) 0.8^3

(d) $1 - 0.8 \times 3$

(e) $1 - 0.8^3$

A : at least one is veg.

A^c : none of them are veg.

$P(A^c) = 0.8^3$

$P(A) = 1 - 0.8^3$

Conditional probability

Photo Classification Results

A classifier was tested on 1822 photos from a photo-sharing website to determine whether they depict fashion. Each photo was classified by:

- `mach_learn`: ML classifier output (`pred_fashion` or `pred_not`).
- `truth`: Human-verified label (`fashion` or `not`).

		truth		
		fashion(A)	not(A^c)	Total
mach_learn	pred_fashion(B)	197	22	219
	pred_not(B^c)	112	1491	1603
	Total	309	1513	1822

What is the probability that the ML classifier identified the photo as being about fashion?

		truth		
		fashion(A)	not(A^c)	Total
mach_learn	pred_fashion(B)	197	22	219
	pred_not(B^c)	112	1491	1603
	Total	309	1513	1822

(Probability that mach_learn is pred_fashion)

$= P(B) = \frac{219}{1822}$

Joint probability

What is the probability that the ML classifier predicted the photo as being about fashion <u>and</u> the photo is truly about fashion?

		truth		
		fashion(A)	not(A^c)	Total
mach_learn	pred_fashion(B)	197	22	219
	pred_not(B^c)	112	1491	1603
	Total	309	1513	1822

(Probability that mach_learn is pred_fashion and truth is fashion)

$= \quad P(AB) = \dfrac{197}{1822}$

Conditional probability

Conditional probability

The conditional probability of the outcome of interest A given condition B is calculated as

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \quad \begin{array}{l} \leftarrow \text{joint probability} \\ \leftarrow \text{marginal probability} \end{array}$$

	`fashion`(A)	`not`(A^c)	Total
`pred_fashion`(B)	197	22	219
`pred_not`(B^c)	112	1491	1603
Total	309	1513	1822

(Probability that `truth` is `fashion` given `mach_learn` is `pred_fashion`)

A B

$$= P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{197/1822}{219/1822} = \frac{197}{219}$$

Conditional probability (cont.)

If we knew the ML classifier predicted the photo was about fashion, what is the probability that the photo is actually about fashion?

	$\texttt{fashion}(A)$	$\texttt{not}(A^c)$	Total
$\texttt{pred\_fashion}(B)$	*197*	22	*219*
$\texttt{pred\_not}(B^c)$	112	1491	1603
Total	309	1513	1822

$$\mathrm{P}(\texttt{fashion} \mid \texttt{pred\_fashion}) = \frac{197}{219} \approx 0.900$$

$$\mathrm{P}(\texttt{fashion} \mid \texttt{pred\_not}) = \frac{112}{1603} \approx 0.070$$

Conditional probability (cont.)

What is the probability that the ML prediction was correct if the photo was about fashion?

	$\texttt{fashion}(A)$	$\texttt{not}(A^c)$	Total
$\texttt{pred\_fashion}(B)$	*197*	22	219
$\texttt{pred\_not}(B^c)$	112	1491	1603
Total	*309*	1513	1822

$$\mathrm{P}(\texttt{pred\_fashion} \mid \texttt{fashion}) = \frac{197}{309} \approx 0.638$$

$$\mathrm{P}(\texttt{pred\_fashion} \mid \texttt{not fashion}) = \frac{22}{1513} \approx 0.015$$

General multiplication rule

- Earlier we saw that if two events are independent, their joint probability is simply the product of their probabilities.
- If the events are not believed to be independent, the joint probability is calculated slightly differently.

General multiplication rule

If A and B represent two outcomes or events, then

$$P(A \text{ and } B) = P(A|B) \times P(B) = P(B|A) \times P(A)$$

- Note that this formula is simply the conditional probability formula, rearranged.

$P(A|B) = \frac{P(AB)}{P(B)} \Rightarrow P(AB) = P(A|B)P(B)$

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

		SS social science	SS^c non-social science	total
F	female	30	20	50
M	male	30	20	50
	total	60	40	100

- The probability that a randomly selected student is a social science major is $P(SS) = \frac{60}{100}$
- The probability that a randomly selected student is a social science major given that they are female is $P(SS|F) = \frac{30}{50}$

Does major of students in this class depend on their gender?

$\Rightarrow$ No, independent

Independence and conditional probabilities (cont.)

Generically, if $P(A|B) = P(A)$ then the events A and B are said to be independent.

- Conceptually: Giving B doesn't tell us anything about A.
- Mathematically: We know that if events A and B are independent, $P(A \text{ and } B) = P(A) \times P(B)$. Then,

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A) \times P(B)}{P(B)} = P(A)$$

Inverting probabilities

$P(A)$

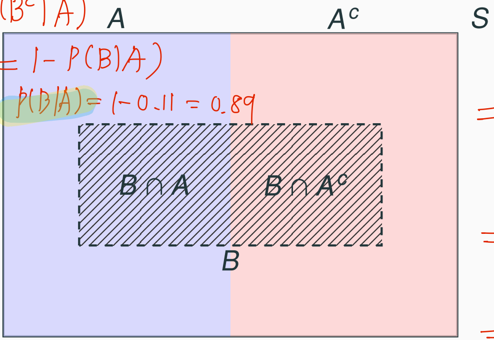
In Canada, 0.35% of women over 40 develop breast cancer each year. A mammogram, a common screening test, has a false negative rate of 11% and a false positive rate of 7%. If a randomly selected woman over 40 tests positive(event B), what is the probability that she actually has breast cancer(event A)?

$P(B|A^c)$

$= P(B^c|A)$

$= 1 - P(B|A)$

$P(B|A) = 1 - 0.11 = 0.89$



$P(AB) = P(B|A)P(A)$

$P(A|B)$

$$= \frac{P(AB)}{P(B)} = \frac{P(AB)}{P(AB) + P(A^cB)}$$

$$= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

$$= \frac{0.89 \times 0.0035}{0.89 \times 0.0035 + 0.07 \times (1 - 0.0035)} = 4.27\%$$

Bayes' Theorem

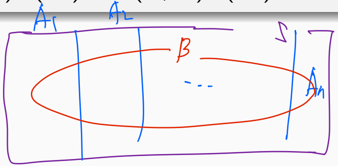
The conditional probability formula we have seen so far is a special case of the *Bayes' Theorem*, which is applicable even when events have more than just two outcomes.

Bayes' Theorem:

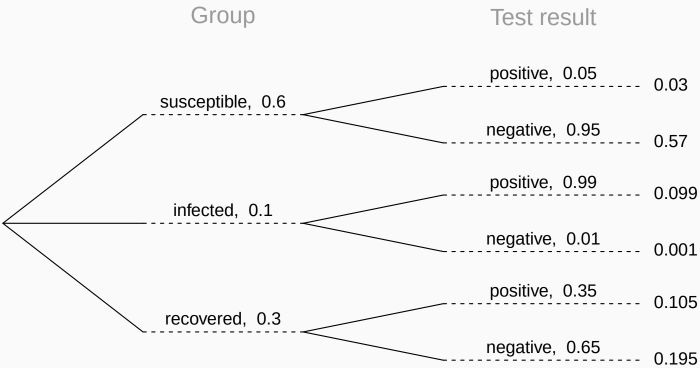
Partition of sample space

Let $A_1, \cdots, A_n$ and B be events where A_i are disjoint, $\cup_{i=1}^{n} A_i = S$, and $P(A_i) > 0$ for all $i = 1, \cdots, k$. Then

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)}.$$



Application activity: Inverting probabilities



$$P(inf|pos) = \frac{P(inf\ and\ pos)}{P(pos)} = \frac{0.099}{0.03 + 0.099 + 0.105} \approx 0.423$$

Random variables

Random variables

- A *random variable(RV)* is a numeric quantity whose value depends on the outcome of a random event
  - We use a capital letter, like X, to denote a random variable
  - The values of a random variable are denoted with a lowercase letter, in this case x
  - For example, $P(X = x)$
- There are two types of random variables:
  - Countable *Discrete random variables* often take only integer values
    - Example: Number of credit hours, Difference in number of credit hours this term vs last
  - uncountable *Continuous random variables* take real (decimal) values
    - Example: Cost of books this term, Difference in cost of books this term vs last

Let a random variable X denote the number of heads in two coin tosses. Find the probability distribution of X.

- Probability distribution function:

$$f(x) = \begin{cases} 1/4 & x=0 \\ 1/2 & x=1 \\ 1/4 & x=2 \\ 0 & \text{o.w.} \end{cases}$$

- Probability distribution table:

x	0	1	2	o.w.
$f(x)$	1/4	1/2	1/4	0

HH TH HT TT

X 2 1 1 0

$P(X=2) = \frac{1}{4}$

$P(X=1) = \frac{2}{4}$

$P(X=0) = \frac{1}{4}$

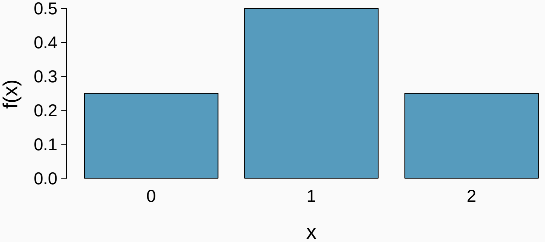
$f(x) = P(X=x)$

Probability mass function(pmf), $f(x)$ (e.g. $x_1=0,\ x_2=1,\ x_3=2$)

$$f(x) = \begin{cases} P(X = x_i), & x = x_i \ (i = 1, \cdots, k) \\ 0, & o.w. \end{cases}$$

Properties of probability mass function

- $0 \leq f(x) \leq 1, \quad \sum_{i=1}^{k} f(x_i) = 1.$
- $P(a < X < b) = \sum_{a<x<b} f(x).$



Continuous random variable

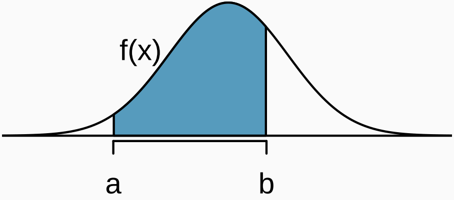
Probability density function(pdf), $f(x)$

$$P(a \le X \le b) = \int_a^b f(x)dx$$

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

Properties of probability density function

- $f(x) \ge 0$, $\int_{-\infty}^{\infty} f(x)dx = 1$.
- $P(a < X < b) = P(a \le X \le b) = \int_a^b f(x)dx$.



Practice

Consider a continuous RV X with a probability density function $f(x) = \begin{cases} cx(1-x), & 0 \le x \le 1 \\ 0, & o.w. \end{cases}$, where c is a constant.

- Find a constant c so that $f(x)$ is a probability density function.

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \underbrace{\int_{-\infty}^{0} f(x)\,dx}_{0=} + \int_{0}^{1} f(x)\,dx + \underbrace{\int_{1}^{\infty} f(x)\,dx}_{=0} = \int_{0}^{1} cx(1-x)\,dx$$

$$= c\left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_0^1 = \frac{c}{6} \Rightarrow c = 6$$

- Find $P(X \le 1/2)$.

$$= P(-\infty \le X \le 1/2)$$

$$= \int_{-\infty}^{1/2} f(x)\,dx = \int_{0}^{1/2} 6x(1-x)\,dx = \frac{1}{2}$$

If you repeat tossing two coins, how many heads would you expect to appear on average?

Expectation

We are often interested in *average* outcome of a random variable.

Expectation(mean) of a RV X

$$\mu = E(X) = \begin{cases} \sum_{i=1}^{k} x_i f(x_i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

Expectation of a function of RV $g(X)$

$$E\big(g(X)\big) = \begin{cases} \sum_{i=1}^{k} g(x_i) f(x_i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x) f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

Practice

Let a random variable X denote the number of heads in two coin tosses. Find the expectation of X.

The probability distribution of X is given as

x	0	1	2
$f(x)$	1/4	2/4	1/4

- $E(X) = \sum_{\forall x} x f(x) = 0 \cdot f(0) + 1 \cdot f(1) + 2 \cdot f(2) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{2}{4} + 2 \cdot \frac{1}{4} = 1$

Find the expectation of $Y = (X-1)^2$.

- $E(Y) = E\left[(X-1)^2\right]$

$$= \sum_{\forall x} (x-1)^2 f(x) = (0-1)^2 f(0) + (1-1)^2 f(1) + (2-1)^2 f(2)$$

$$= 1 \times \frac{1}{4} + 1 \times \frac{1}{4} = \frac{1}{2}$$

Variance and standard deviation

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

Variance and standard deviation of a RV X

$$\sigma^2 = Var(X) = E[(X-\mu)^2] = \begin{cases} \sum_{i=1}^{k}(x_i-\mu)^2 f(x_i) & X \text{ discrete} \\ \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx & X \text{ continuous} \end{cases}$$

$$\sigma = SD(X) = \sqrt{Var(X)}$$

Computing variance

$$Var(X) = E(X^2) - E(X)^2$$

Note:

$$Var(X) = E[(X-\mu)^2] = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - E(X)^2$$

Practice

Tony is waiting for a bus. Suppose that waiting time X (hours) has the following a probability density function $f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & o.w. \end{cases}$

- Find $E(X)$: $= \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x\,dx = \left[\frac{1}{2}x^2\right]_0^1 = \frac{1}{2}$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 x^2\,dx = \left[\frac{1}{3}x^3\right]_0^1 = \frac{1}{3}$$

- Find $Var(X)$:

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

Properties of expectation and variance

Let a and b be constants(not random).

Property of expectation

$E(aX + b) = aE(X) + b$

Properties of variance

- $Var(aX + b) = a^2 Var(X)$
- $Var(X) \geq 0$

$\underbrace{Var(X)}_{\geq 0} = E(\underbrace{(X-\mu)^2}_{\geq 0})$

Note: If X is a continuous random variable(discrete X can be shown similarly),

$$E(aX + b) = \int_{-\infty}^{\infty} (ax + b) f(x) dx = a \int_{-\infty}^{\infty} x f(x) dx + b \int_{-\infty}^{\infty} f(x) dx = aE(X) + b$$

$$\begin{aligned} Var(aX + b) &= E\big(aX + b - E(aX + b)\big)^2 = E\big(aX + b - aE(X) - b\big)^2 \\ &= a^2 E\big(X - E(X)\big)^2 = a^2 Var(X) \end{aligned}$$

Practice

Tony is waiting for a bus. Suppose that waiting time X (hours) has the following a probability density function $f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & o.w. \end{cases}$

Let $Y = 2X + 1$.

$E(X) = \frac{1}{2}$ $Var(X) = \frac{1}{12}$

- Find $E(Y)$: $= E(2X+1) = 2E(X)+1 = 2 \times \frac{1}{2} + 1 = 2$
- Find $Var(Y)$: $= Var(2X+1) = 4Var(X) = 4 \times \frac{1}{12} = \frac{1}{3}$

Joint probability distribution

Recall: Conditional probability

The conditional probability of the outcome of interest A given condition B is calculated as

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \quad \begin{array}{l} \leftarrow \text{joint probability} \\ \leftarrow \text{marginal probability} \end{array}$$

Joint probability: Probability for two events(variables).

Marginal probability: Probability for a single event(variable).

$1 = P(-\infty \le X \le \infty, -\infty \le Y \le \infty)$

$= \sum_{\forall x} \sum_{\forall y} f(x,y)$

Joint pmf of two *discrete* RVs X and Y

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

- $0 \le f(x,y) \le 1$, $\sum_{\forall x} \sum_{\forall y} f(x,y) = 1$.
- $P(a < X \le b, c < Y \le d) = \sum_{a<x\le b} \sum_{c<y\le d} f(x,y)$

Joint pdf of two *continuous* RVs X and Y

$$\int_a^b \int_c^d f(x,y)dydx = P(a < X < b, c < Y < d)$$

- $0 \le f(x,y)$, $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy = 1$.

$\iint_{(x,y)\in A} f(x,y)\, dxdy = P((X,Y) \in A)$

Marginal probability distribution

$f_X(x) = P(X = x, -\infty \le Y \le \infty)$

$= \sum_{\forall y} f(x,y) \quad \left(= \sum_{-\infty \le y \le \infty} f(x,y)\right).$

Marginal pmf of a *discrete* RV

$$f_X(x) = \sum_{\forall y} f(x,y), \quad f_Y(y) = \sum_{\forall x} f(x,y)$$

Marginal pdf of a *continuous* RV

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

Practice

In a certain suburban area, each household reported the number of cars X and the number of television sets Y that they owned. Joint pmf of X and Y is given below.

$f_X(x) = \sum_{\forall y} f(x,y)$

X \ Y	1	2	3	4
1	0.1	0.0	0.1	0.0
2	0.3	0.0	0.1	0.2
3	0.0	0.2	0.0	0.0

- Find the marginal pmf of X:

x	1	2	3
$f_X(x)$	0.2	0.6	0.2

- Find the marginal pmf of Y:

y	1	2	3	4
$f_Y(y)$	0.4	0.2	0.2	0.2

Independence of two variables

Recall: Independence of two events

Two events A and B are independent if

$$P(A \text{ and } B) = P(A) \times P(B).$$

Independence of two random variables

X and Y are (mutually) independent if

$$f(x, y) = f_X(x) f_Y(y).$$

Joint pdf of X and Y is given by

$$f(x,y) = \begin{cases} \dfrac{3x-y}{9} & 1 < x < 3,\ 1 < y < 2 \\ 0, & o.w. \end{cases}$$

Determine whether X and Y are independent.

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_1^2 \frac{3x-y}{9}\,dy = \int_1^2 \frac{3x}{9}\,dy - \int_1^2 \frac{y}{9}\,dy$$

$$= \frac{3x}{9}\int_1^2 1\,dy - \frac{1}{9}\int_1^2 y\,dy = \frac{3x}{9} - \frac{1}{9}\times\frac{3}{2} = \frac{1}{3}x - \frac{1}{6} \quad (1<x<3)$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \frac{4}{3} - \frac{2}{9}y \quad (1<y<2)$$

$f(x,y) \neq f_X(x) f_Y(y) \Rightarrow X$ and Y are not independent

Covariance

Expectation of a function of RVs $g(X, Y)$

$$E\left(g(X,Y)\right) = \begin{cases} \sum_{\forall x} \sum_{\forall y} g(x,y) f(x,y) & \text{if } X, Y \text{ discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dxdy & \text{if } X, Y \text{ continuous} \end{cases}$$

Covariance measures the joint variability of two random variables.

Covariance

Covariance of two random variables X and Y is

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

Note: If X and Y are independent, $Cov(X, Y) = 0$, but not vice versa.

Practice

Joint and marginal pmf of *X* and *Y* are given below. $E(X) = \sum_{\forall x} x f(x)$

$X \backslash Y$	1	2	3	4	$f(x)$
1	0.1	0.0	0.1	0.0	0.2
2	0.3	0.0	0.1	0.2	0.6
3	0.0	0.2	0.0	0.0	0.2
$f(y)$	0.4	0.2	0.2	0.2	1.0

- Find $E(X)$ and $E(Y)$:

$E(X) = 1\times 0.2 + 2\times 0.6 + 3\times 0.2 = 2$

$E(Y) = 1\times 0.4 + 2\times 0.2 + 3\times 0.2 + 4\times 0.2 = 2.2$

- Find $E(XY)$:

$E(XY) = \sum_{\forall x} \sum_{\forall y} xy\, f(x,y)$

$= 1\times 1\times 0.1 + 1\times 3\times 0.1$

$+ 2\times 1\times 0.3 + 2\times 3\times 0.1 + 2\times 4\times 0.2$

$+ 3\times 2\times 0.2 = 4.4$

- Find $Cov(X, Y)$:

$= E(XY) - E(X)E(Y) = 4.4 - 2\times 2.2$

$= 0$

Note: *Are X and Y independent?*

$f(x,y) = f_X(x)\, f_Y(y)$?

$f(1,1) \neq f_X(1)\cdot f_Y(1) \Rightarrow X, Y$ are not independent

Correlation

Correlation of two random variables X and Y is

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

Properties of correlation

- $-1 \leq \rho(X, Y) \leq 1$
- $\rho(X, Y) > 0$: positive linear relationship between X and Y.
- $\rho(X, Y) < 0$: negative linear relationship between X and Y.
- $\rho(X, Y) = 0$: no linear relationship between X and Y.

Linear combinations

A *linear combination* of random variables X and Y is given by

$$aX + bY$$

$E(aX+b) = aE(X)+b$

where a and b are some fixed constants. $Var(aX+b) = a^2 Var(X)$

The average value of a linear combination of RVs is given by

Expectation of a linear combination of two RVs

$$E(aX + bY) = a \times E(X) + b \times E(Y)$$

The variance of a linear combination of two *independent* RVs is

Variance of a linear combination of two RVs

$$Var(aX + bY) = a^2 \times Var(X) + b^2 \times Var(Y)$$

Practice

On average a notebook costs \$10 and a pen costs \$15. You plan to buy 5 notebooks (all at the same random price X) and 4 pens (all at the same random price Y). What is the expected total cost $E(5X + 4Y)$?

$$E(5X+4Y) = 5E(X) + 4E(Y) = 5\times 10 + 4\times 15 = 110$$

The standard deviation of the price is \$1.5 for a notebook and \$2 for a pen. Assuming independence between X and Y, what is the standard deviation of the total cost $SD(5X + 4Y)$?

$$SD(X) = \sqrt{Var(X)}$$

$$Var(5X+4Y) = 25\,Var(X) + 16\,Var(Y)$$

$$= 25\times 1.5^2 + 16\times 2^2 = 120.25$$

$$\Rightarrow SD(5X+4Y) = \sqrt{120.25} = 10.97$$

KOREA MILITARY ACADEMY

Simplifying random variables

$2X = X + X$ can be understood as a linear combination of two random variables X and X. Thus,

$$E(2X) = E(X + X) = E(X) + E(X) = 2E(X),$$

but

$$Var(X + X) \neq Var(X) + Var(X),$$

because X and X themselves are not independent. Instead,

$$Var(X + X) = Var(2X) = \underline{4\,Var(X)}$$

Exercises

Exercises in OpenIntro Statistics 4th ed.

- **Conditional probability:**
  - Exercise 3.22, 3.42.
- **Random variables:**
  - Exercise 3.34, 3.44 (a) and (b), 3.47(or 3.45)

1. Suppose that X and Y have a joint probability density function

$$f(x, y) = \begin{cases} \frac{3}{2}y^2, & 0 \le x \le 2, 0 \le y \le 1 \\ 0, & o.w. \end{cases}$$

1. Find the marginal pdf of Y, $f_Y(y)$.
2. Find $E(Y)$ and $Var(Y)$.

2. Suppose that X and Y have a joint probability mass function

$$f(x,y) = \begin{cases} \frac{1}{12}(x+y), & x = 0,2, \text{ and } y = 0,1,2, \\ 0, & o.w. \end{cases}$$

1. Find the marginal pmf of X, $f_X(x)$. (Hint: Use a table in slide 40).
2. Find the marginal pmf of Y, $f_Y(y)$.
3. Find $E(X)$ and $E(Y)$.
4. Find $E(XY)$ and $Cov(X,Y)$.

3. Suppose that X and Y have a joint probability density function

$$f(x,y) = \begin{cases} e^{-x-y}, & x > 0,\ y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

1. Find the marginal pdfs $f_X(x)$ and $f_Y(y)$.
2. Find $E(X)$ and $E(Y)$.
3. Find $E(XY)$ and $\mathrm{Cov}(X, Y)$.
4. Find $P(X < Y)$.