Probability - Session 2

Bayes's Theorem, Random variables, Expectation and Variance

Elizabeth Williamson
with thanks to Jennifer Rogers

Foundations of Medical Statistics

Session objectives

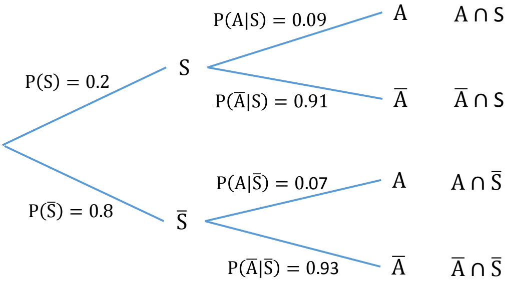
By the end of this session you should be able to:

- state and apply Bayes' Theorem
- define sensitivity, specificity and the predictive value of a screening/diagnostic test
- explain the concepts of a discrete random variable and its distribution function
- calculate the expectation and variance of a discrete random variable
- define a joint distribution and independence

Outline

Example: Smoking and asthma

Reminder:
this is the information we have about smoking and asthma



Example: Reversing conditioning

We often wish to reverse the conditioning in probabilities.

For example, suppose an asthma clinic wants to know how many resources to dedicate to anti-smoking measures.

- the clinic would want to know what proportion of asthmatic patients smoke, i.e. $P(S|A)$
- we only have information about $P(A|S)$

Bayes theorem - I

- We can express $P(A \cap S)$ in two different ways

$$\begin{aligned} P(A \cap S) &= P(A|S)P(S) \\ &\text{or} \\ P(A \cap S) &= P(S|A)P(A). \end{aligned}$$

- Equating the two, we have

$$P(S|A)P(A) = P(A|S)P(S)$$

$$\Rightarrow P(S|A) = \frac{P(A|S)P(S)}{P(A)}.$$

Bayes theorem - II

We have:

$$P(S|A) = \frac{P(A|S)P(S)}{P(A)}$$

We are now going to express the denominator in terms of conditional probabilities:

- $\{S, \bar{S}\}$ is a partition of the sample space
- From the theorem of total probability we know that

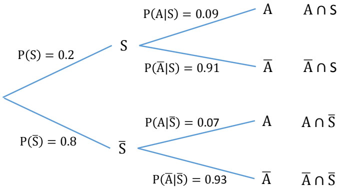
$$P(A) = P(A|S)P(S) + P(A|\bar{S})P(\bar{S})$$

Putting this into the equation above gives:

$$P(S|A) = \frac{P(A|S)P(S)}{P(A|S)P(S) + P(A|\bar{S})P(\bar{S})}$$

This is *Bayes' Theorem*.

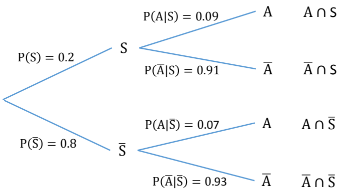
Bayes theorem: Example



For someone with asthma, applying Bayes' Theorem, the probability of being a smoker is:

$$P(S|A) \quad =$$

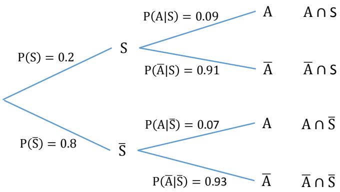
Bayes theorem: Example



For someone with asthma, applying Bayes' Theorem, the probability of being a smoker is:

$$P(S|A) = \frac{P(A|S)P(S)}{P(A|S)P(S) + P(A|\bar{S})P(\bar{S})}$$

Bayes theorem: Example



For someone with asthma, applying Bayes' Theorem, the probability of being a smoker is:

$$\begin{aligned} P(S|A) &= \frac{P(A|S)P(S)}{P(A|S)P(S) + P(A|\bar{S})P(\bar{S})} \\ &= \frac{0.09 \times 0.2}{0.09 \times 0.2 + 0.07 \times 0.8} \\ &= 0.243. \end{aligned}$$

Bayes theorem: General statement

Let A be some event and let $B_1, .., B_n$ be a partition of the sample space.

Suppose we are interested in $P(B_i|A)$, for some i.

Bayes' Theorem states that:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}.$$

Example: Cystic Fibrosis screening tests

- Cystic fibrosis (CF) is an inherited disease
- The sweat chloride test is the gold standard diagnostic test.
- In low income settings the simpler sweat conductivity test may be more appropriate.

Properties of the new sweat conductivity test:

- Among people with CF
  - $P(+ve|CF) = 0.875$
  - We say the **sensitivity** of the test is 87.5%
- Among people without CF
  - $P(+ve|\bar{CF}) = 0.004$, i.e. $P(-ve|\bar{CF}) = 0.996$
  - We say the **specificity** of the test is 99.6%

Example: Screening tests

The question we really want to answer is: "If a patient has a positive sweat conductivity test, what is the probability that the patient has CF?"

$$P(+ve|CF) = 0.875 \qquad P(CF) = 0.0325$$
$$P(+ve|\bar{CF}) = 0.004 \qquad \{CF, \bar{CF}\} \text{ is a partition}$$

[Note: For this calculation, $P(CF)$ is the probability of CF in the population the test is being used in.]

We want the **Positive Predictive Value (PPV)**, $P(CF|+ve)$

$$\begin{aligned} P(CF|+ve) &= \frac{P(+ve|CF)P(CF)}{P(+ve|CF)P(CF) + P(+ve|\bar{CF})P(\bar{CF})} \\ &= \frac{0.875 \times 0.0325}{0.875 \times 0.0325 + 0.004 \times (1 - 0.0325)} = 0.88 \end{aligned}$$

A patient whose sweat test comes back positive has an 88% chance of having CF.

Outline

Bayes theorem

Random variables

Expectation and variance

Joint distributions

Random variables

A **random variable** X is a variable which takes a numerical value which depends on the outcome of the random experiment under consideration.

Discrete random variables take values in:

- a finite set

  e.g. the number of boys in a 4-child family $(0, 1, 2, 3, 4)$
- or a countably infinite set (e.g. set of positive integers).

  e.g. the number of hospital admissions in a day $(0, 1, 2, 3, ...)$.

In contrast, **continuous random variables** take values in an uncountable set (e.g. positive real numbers).

Today we will focus on discrete random variables.

The probability distribution function

A discrete random variable can be characterised by its probability distribution function.

The **probability distribution** of the discrete random variable X is a function which tells us, for any value x which X might take, the probability X will take this value.

- $P(X = x)$ for all the x values which X can take.

The probability distribution function has the following properties:

- $0 \leq P(X = x) \leq 1$
- $\sum_x P(X = x) = 1$,
  where the sum is over all possible values of X

Example: boys in a 4-child family

Let X be the number of boys in a 4-child family.

The probability distribution function for X is:

x	$P(X = x)$
0	0.06
1	0.24
2	0.37
3	0.26
4	0.07

[Note: we will see where these numbers come from in Session 3]

Cumulative distribution function

- The cumulative distribution function (CDF) is an alternative way of characterising a random variable.
- For a random variable X, the **cumulative distribution function** (CDF) is given by:

$$F(x) = P(X \leq x)$$

For our example of the number of boys:

x	$P(X = x)$	$F(x) = P(X \leq x)$
0	0.06	0.06
1	0.24	0.30
2	0.37	0.67
3	0.26	0.93
4	0.07	1

Outline

Expectation of a random variable

- The **expectation (or mean)** of a random variable X is one measure of the centre of its distribution (another is the median).
- For discrete random variables X, it is defined as:

$$E(X) = \sum_x x\, P(X = x),$$

  the summation is over all possible values x that X can take.
- One way to think of $E(X)$ is the average value of X over a large number of repetitions of the experiment or random process that produces X.
- The Greek letter μ is often used for $E(X)$.

Example: Expected number of boys

For our example of the number of boys in a 4-child family:

x	$P(X = x)$	$F(x) = P(X \le x)$	$x \times P(X = x)$
0	0.06	0.06	0×0.06
1	0.24	0.30	1×0.24
2	0.37	0.67	2×0.37
3	0.26	0.93	3×0.26
4	0.07	1	4×0.07

$$E(X) = 0 \times 0.06 + 1 \times 0.24 + 2 \times 0.37 + 3 \times 0.26 + 4 \times 0.07 = 2.04$$

- Note - we do not actually expect to find 2.04 boys in a 4 child family!

Properties of expectation

Expectations of functions of random variables satisfy certain rules.

If a and b are constants,

- $E(X+b) = E(X) + b$
  - Adding b just shifts the distribution of X by b.
- $E(aX) = aE(X)$
  - For each value x which X takes, aX takes the value ax.
- $E(aX+b) = aE(X) + b$.
  - By combining the results above.

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

- The expectation of a function, $g(X)$, of a discrete random variable X is

$$E(g(X)) = \sum_x g(x)P(X = x)$$

Proof:

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

- The expectation of a function, $g(X)$, of a discrete random variable X is

$$E(g(X)) = \sum_x g(x)P(X = x)$$

Proof:

- Setting $g(X) = aX + b$ gives,

$$E(aX + b) = \sum_x (ax + b)P(X = x)$$

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

- The expectation of a function, $g(X)$, of a discrete random variable X is

$$E(g(X)) = \sum_x g(x)P(X = x)$$

Proof:

- Setting $g(X) = aX + b$ gives,

$$\begin{aligned} E(aX+b) &= \sum_x (ax+b)P(X=x) \\ &= \sum_x axP(X=x) + \sum_x bP(X=x) \end{aligned}$$

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

- The expectation of a function, $g(X)$, of a discrete random variable X is

$$E(g(X)) = \sum_x g(x)P(X = x)$$

Proof:

- Setting $g(X) = aX + b$ gives,

$$\begin{aligned} E(aX + b) &= \sum_x (ax + b)P(X = x) \\ &= \sum_x axP(X = x) + \sum_x bP(X = x) \\ &= a\sum_x xP(X = x) + b\sum_x P(X = x) \end{aligned}$$

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

- The expectation of a function, $g(X)$, of a discrete random variable X is

$$E(g(X)) = \sum_x g(x)P(X = x)$$

Proof:

- Setting $g(X) = aX + b$ gives,

$$\begin{aligned} E(aX + b) &= \sum_x (ax + b)P(X = x) \\ &= \sum_x axP(X = x) + \sum_x bP(X = x) \\ &= a\sum_x xP(X = x) + b\sum_x P(X = x) \\ &= aE(X) + b. \end{aligned}$$

Variance of a random variable

- The **variance** of a random variable X is one measure of the variable's dispersion, or 'variability'
- It is defined as:

$$Var(X) = E((X - \mu)^2),$$

where $\mu = E(X)$.

- Equivalently, it can be expressed as:

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

Variance of a random variable

We have two definitions:

$$Var(X) = E((X-\mu)^2), \qquad Var(X) = E(X^2) - E(X)^2$$

To see that these two definitions are equivalent,

$$Var(X) \quad = \quad E((X-\mu)^2)$$

Variance of a random variable

We have two definitions:

$$Var(X) = E((X-\mu)^2), \qquad Var(X) = E(X^2) - E(X)^2$$

To see that these two definitions are equivalent,

$$\begin{aligned} Var(X) &= E((X-\mu)^2) \\ &= E(X^2 - 2X\mu + \mu^2) \end{aligned}$$

Variance of a random variable

We have two definitions:

$$Var(X) = E((X-\mu)^2), \qquad Var(X) = E(X^2) - E(X)^2$$

To see that these two definitions are equivalent,

$$\begin{aligned} Var(X) &= E((X-\mu)^2) \\ &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \end{aligned}$$

Variance of a random variable

We have two definitions:

$$Var(X) = E((X-\mu)^2), \qquad Var(X) = E(X^2) - E(X)^2$$

To see that these two definitions are equivalent,

$$\begin{aligned} Var(X) &= E((X-\mu)^2) \\ &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \end{aligned}$$

Variance of a random variable

We have two definitions:

$$Var(X) = E((X-\mu)^2), \qquad Var(X) = E(X^2) - E(X)^2$$

To see that these two definitions are equivalent,

$$\begin{aligned} Var(X) &= E((X-\mu)^2) \\ &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

Variance of a random variable

We have two definitions:

$$Var(X) = E((X-\mu)^2), \qquad Var(X) = E(X^2) - E(X)^2$$

To see that these two definitions are equivalent,

$$\begin{aligned} Var(X) &= E((X-\mu)^2) \\ &= E(X^2 - 2X\mu + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - E(X)^2. \end{aligned}$$

Properties of variance

Variances of functions of random variables satisfy certain rules.

If a and b are constants,

- $Var(X + b) = Var(X)$
  – Adding b doesn't affect the spread of the distribution.
- $Var(aX) = a^2 Var(X)$
  – Constant multipliers have a squared effect on the variance
- $Var(aX + b) = a^2 Var(X)$.
  – By combining the results above.

Proof: $Var(aX + b) = a^2 Var(X)$

- The variance of a function, $g(X)$, is given by

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

Proof:

$$Var(aX + b) \quad = \quad E(\{ (aX + b) - E(aX + b) \}^2)$$

Proof: $Var(aX + b) = a^2 Var(X)$

- The variance of a function, $g(X)$, is given by

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

Proof:

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

Proof: $Var(aX + b) = a^2 Var(X)$

- The variance of a function, $g(X)$, is given by

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

Proof:

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

Proof: $Var(aX + b) = a^2 Var(X)$

- The variance of a function, $g(X)$, is given by

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

Proof:

$$\begin{aligned} Var(aX + b) &= E(\,\{\,(aX + b) - E(aX + b)\,\}^2\,) \\ &= E(\,\{\, aX + b - a\mu - b\,\}^2\,) \\ &= E(\,\{\, a(X - \mu)\,\}^2\,) \\ &= \sum_x a^2 (x - \mu)^2 P(X = x) \end{aligned}$$

Proof: $Var(aX + b) = a^2 Var(X)$

- The variance of a function, $g(X)$, is given by

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

Proof:

$$\begin{aligned}
Var(aX + b) &= E(\{ (aX + b) - E(aX + b) \}^2) \\
&= E(\{ aX + b - a\mu - b \}^2) \\
&= E(\{ a(X - \mu) \}^2) \\
&= \sum_x a^2 (x - \mu)^2 P(X = x) \\
&= a^2 \sum_x (x - \mu)^2 P(X = x)
\end{aligned}$$

Proof: $Var(aX+b) = a^2 Var(X)$

- The variance of a function, $g(X)$, is given by

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

Proof:

$$\begin{aligned}
Var(aX+b) &= E(\{\,(aX+b) - E(aX+b)\,\}^2\,) \\
&= E(\{\, aX + b - a\mu - b\,\}^2\,) \\
&= E(\{\, a(X-\mu)\,\}^2\,) \\
&= \sum_x a^2 (x-\mu)^2 P(X=x) \\
&= a^2 \sum_x (x-\mu)^2 P(X=x) \\
&= a^2 E((X-\mu)^2\,)
\end{aligned}$$

Proof: $Var(aX + b) = a^2 Var(X)$

- The variance of a function, $g(X)$, is given by

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

Proof:

$$\begin{aligned}
Var(aX + b) &= E(\,\{\,(aX + b) - E(aX + b)\,\}^2\,) \\
&= E(\,\{\, aX + b - a\mu - b\,\}^2\,) \\
&= E(\,\{\, a(X - \mu)\,\}^2\,) \\
&= \sum_x a^2 (x - \mu)^2 P(X = x) \\
&= a^2 \sum_x (x - \mu)^2 P(X = x) \\
&= a^2 E(\,(X - \mu)^2\,) \\
&= a^2 Var(X).
\end{aligned}$$

The Bernoulli distribution

The Bernoulli distribution corresponds to a single binary outcome (1=success, 0=failure), with $P(X = 1) = \pi$.

- The expectation is:

$$E(X) \quad =$$

- Note that $x = x^2$ for $x = 0, 1$, so $E[X^2] = E[X]$
- So the variance is:

$$Var(X) \quad =$$

The Bernoulli distribution

The Bernoulli distribution corresponds to a single binary outcome (1=success, 0=failure), with $P(X = 1) = \pi$.

- The expectation is:

$$\begin{aligned} E(X) &= \sum_x xP(X = x) \\ &= 0 \times (1 - \pi) + 1 \times \pi \\ &= \pi. \end{aligned}$$

- Note that $x = x^2$ for $x = 0, 1$, so $E[X^2] = E[X]$
- So the variance is:

$$Var(X) \;=$$

The Bernoulli distribution

The Bernoulli distribution corresponds to a single binary outcome (1=success, 0=failure), with $P(X = 1) = \pi$.

- The expectation is:

$$\begin{aligned} E(X) &= \sum_x x P(X = x) \\ &= 0 \times (1 - \pi) + 1 \times \pi \\ &= \pi. \end{aligned}$$

- Note that $x = x^2$ for $x = 0, 1$, so $E[X^2] = E[X]$
- So the variance is:

$$\begin{aligned} Var(X) &= E[X^2] - E(X)^2 \\ &= \pi - \pi^2 = \pi(1 - \pi) \end{aligned}$$

Outline

Joint distributions

- So far we have considered a single random variable X.
- Often we are interested in the relationship between two (or more) variables.

Let X and Y be two discrete random variables.

- To consider the relationship between X and Y, we need to define their *joint* distribution.

Joint distribution function

The **joint distribution function** of X and Y is given by,

- $P(X = x, Y = y)$ for values x, y which X and Y can take
- This is defined as the probability $P(X = x \cap Y = y)$.
- We often abbreviate this to $P(x, y)$.

- The joint distribution function must satisfy:

$$\begin{aligned} P(x,y) &\geq 0 \text{ for all } x, y \\ \sum_x \sum_y P(x,y) &= 1. \end{aligned}$$

Example: Obesity and exercise

We are interested in the relationship between exercise and obesity

- Let X be the typical number of days per week a person does vigorous exercise (grouped)
- Let Y be the (grouped) weight of a person

The joint distribution of obesity (Y) and exercise (X) is:

	Obesity			
Exercise (days/week)	Underweight ($y = 1$)	Normal weight ($y = 2$)	Overweight ($y = 3$)	Total
0-1 ($x = 0$)	0.05	0.1	0.15	0.3
2-4 ($x = 1$)	0	0.05	0.05	0.1
5-7 ($x = 2$)	0.05	0.35	0.2	0.6
Total	0.1	0.5	0.4	1

Marginal distributions

- The **marginal** distribution of X, $P(X = x)$, is simply the distribution of X, ignoring Y.

- The marginal distribution of X can be found from the joint distribution as:

$$P(X = x) = \sum_{y} P(x, y).$$

Conditional distributions

- The **conditional** distribution of X given $Y = y$, is the distribution of X given that $Y = y$.

- It can be expressed as a function of the joint distribution function:

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

Example: Obesity and exercise

Suppose we want to know the overall distribution of obesity.

- What is the marginal distribution of obesity (Y)?

Suppose we want to know how much exercise (X) overweight people ($Y = 3$) are doing.

- What is the conditional distribution X given $Y = 3$?

Exercise (days/week)	**Obesity** Underweight ($y = 1$)	Normal weight ($y = 2$)	Overweight ($y = 3$)	Total
0-1 ($x = 0$)	0.05	0.1	0.15	0.3
2-4 ($x = 1$)	0	0.05	0.05	0.1
5-7 ($x = 2$)	0.05	0.35	0.2	0.6
Total	0.1	0.5	0.4	1

Example: Obesity and exercise

Suppose we want to know the overall distribution of obesity.

- What is the **marginal** distribution of obesity (Y)?

Suppose we want to know how much exercise (X) overweight people ($Y = 3$) are doing.

- What is the conditional distribution X given $Y = 3$?

Exercise (days/week)	**Obesity** Underweight ($y = 1$)	Normal weight ($y = 2$)	Overweight ($y = 3$)	Total
0-1 ($x = 0$)	0.05	0.1	0.15	0.3
2-4 ($x = 1$)	0	0.05	0.05	0.1
5-7 ($x = 2$)	0.05	0.35	0.2	0.6
Total	**0.1**	**0.5**	**0.4**	1

Example: Obesity and exercise

Suppose we want to know the overall distribution of obesity.

- What is the marginal distribution of obesity (Y)?

Suppose we want to know how much exercise (X) overweight people ($Y = 3$) are doing.

- What is the **conditional** distribution X given $Y = 3$?

Exercise (days/week)	**Obesity** Underweight ($y = 1$)	Normal weight ($y = 2$)	Overweight ($y = 3$)	Total
0-1 ($x = 0$)	0.05	0.1	**0.15**	0.3
2-4 ($x = 1$)	0	0.05	**0.05**	0.1
5-7 ($x = 2$)	0.05	0.35	**0.2**	0.6
Total	0.1	0.5	0.4	1

Cumulative distribution function

- It can be expressed as:

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

Independence between two random variables

- If X and Y have no association/dependency, we say they are *independent*.
- X and Y are independent if

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

for all possible values x and y that X and Y take.

Some properties of the expectation and variance involving two random variables

Let X and Y be discrete random variables. Then,

- $E(X+Y) = E(X) + E(Y)$

If X and Y are **independent**, then also

- $E[XY] = E(X)E(Y)$, and
- $Var[X+Y] = Var(X) + Var(Y)$

Proof: $E(X + Y) = E(X) + E(Y)$

- X and Y are discrete random variables
- Let Z be their sum, $Z = X + Y$
- Then

$$E(X + Y) = E(Z) = \sum_z z\, P(Z = z)$$

Proof: $E(X+Y) = E(X) + E(Y)$

- X and Y are discrete random variables
- Let Z be their sum, $Z = X + Y$
- Then

$$E(X+Y) = E(Z) = \sum_z z\,P(Z=z)$$

$$= \sum_x \sum_y (x+y)\,P(X=x, Y=y)$$

Proof: $E(X + Y) = E(X) + E(Y)$

- X and Y are discrete random variables
- Let Z be their sum, $Z = X + Y$
- Then

$$E(X+Y) = E(Z) = \sum_z z\,P(Z=z)$$

$$= \sum_x \sum_y (x+y)\,P(X=x, Y=y)$$

$$= \sum_x \sum_y x\,P(X=x, Y=y) + \sum_x \sum_y y\,P(X=x, Y=y)$$

Proof: $E(X+Y) = E(X) + E(Y)$

- X and Y are discrete random variables
- Let Z be their sum, $Z = X + Y$
- Then

$$
\begin{aligned}
E(X+Y) = E(Z) &= \sum_z z\, P(Z = z) \\
&= \sum_x \sum_y (x+y)\, P(X = x, Y = y) \\
&= \sum_x \sum_y x\, P(X = x, Y = y) + \sum_x \sum_y y\, P(X = x, Y = y) \\
&= \sum_x x\, P(X = x) + \sum_y y\, P(Y = y)
\end{aligned}
$$

Proof: $E(X+Y) = E(X) + E(Y)$

- X and Y are discrete random variables
- Let Z be their sum, $Z = X + Y$
- Then

$$
\begin{aligned}
E(X+Y) = E(Z) &= \sum_z z\,P(Z=z) \\
&= \sum_x \sum_y (x+y)\,P(X=x, Y=y) \\
&= \sum_x \sum_y x\,P(X=x, Y=y) + \sum_x \sum_y y\,P(X=x, Y=y) \\
&= \sum_x x\,P(X=x) + \sum_y y\,P(Y=y) \\
&= E(X) + E(Y).
\end{aligned}
$$

Summary

- Bayes theorem
  - Useful for reversing the conditioning.
- Random variables
  - Quantitative variables whose value depends on outcome of random experiment.
- Expectation and variance
  - We have defined the expectation and variance of discrete random variables, and looked at some of the properties of these.
- Joint distributions
  - We have introduced the notion of joint, marginal, and conditional distribution functions, and of independence.