STAT340 Lecture 02: Independence, Conditional Probability and Bayes' Rule

Brian Powers

9/9/2024

Overview

In these notes, we return to a very basic idea: independence of random variables. We will discuss in more detail what it means for variables to be independent, and we will discuss the related notion of correlation.

We will also see how independence relates to "conditional probability", and how we can use this different "kind" of probability to answer questions that frequently arise in statistics (especially in medical trials) by appealing to Bayes' rule.

Learning objectives

After this lesson, you will be able to

- Explain what it means for variables to be dependent or independent and assess how reasonable independence assumptions are in simple statistical models.
- Explain expectations and variances of sums of variables are influenced by the dependence or independence of those random variables.
- Explain correlation, compute the correlation of two random variables, and explain the difference between correlation and dependence.
- Define the conditional probability of an event A given an event B and calculate this probability given the appropriate joint distribution.
- Use Bayes' rule to compute $\Pr[B \mid A]$ in terms of $\Pr[A \mid B]$, $\Pr[A]$ and $\Pr[B]$.

Recap: RVs events and independence. I

A set of possible *outcomes* denoted Ω.

A subset $E \subseteq \Omega$ of the outcome space is called an *event*.

E: "$X \leq 3$"

A *probability* is a function that maps events to numbers such that:

ex) $X \sim Bin(7, .2)$

$Pr[X \leq 3]$

- $\Pr[E] \in [0, 1]$ for all events E
- $\Pr[\Omega] = 1$
- For $E_1, E_2 \in \Omega$ with $\underline{E_1 \cap E_2 = \emptyset}$, $\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2]$

Note: If $\Pr[E_1 \cap E_2] = 0$, we say that E_1 and E_2 are *mutually exclusive* (or *disjoint*)

"and"

$\cap$ intersection

Recap: RVs events and independence. II

Two events E_1 and E_2 are *independent* if $\Pr[E_1 \cap E_2] = \Pr[E_1]\Pr[E_2]$.

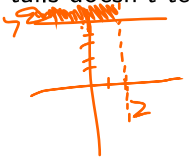
We may write $\Pr[E_1, E_2]$ to mean $\Pr[E_1 \cap E_2]$, read as "the probability of events E_1 and E_2" or "the probability that E_1 and E_2 occur".

Two *random variables* X and Y are independent if for all sets S_1, S_2, we have $\Pr[X \in S_1,\ Y \in S_2] = \Pr[X \in S_1]\Pr[Y \in S_2]$.

Two random variables are independent if learning information about one of them doesn't tell you anything about the other.

- For example, if each of us flip a coin, it is reasonable to model them as being independent.
- Learning whether my coin landed heads or tails doesn't tell us anything about your coin.

ex) $\Pr[X < 2, Y \geq 1]$



Example: dice and coins

Suppose that you roll a die and I flip a coin. Let D denote the (random) outcome of the die roll, and let C denote the (random) outcome of the coin flip. So $D \in \{1, 2, 3, 4, 5, 6\}$ and $C \in \{\overset{1}{\cancel{H}}, \overset{0}{\cancel{T}}\}$. Suppose that for all $d \in \{1, 2, 3, 4, 5, 6\}$ and all $c \in \{\underset{1}{\cancel{H}}, \underset{0}{\cancel{T}}\}$, $\Pr[D = d, C = c] = 1/12$.

Question: Verify that the random variables D and C are independent, or at least check that it's true for two particular events $E_1 \subseteq \{1, 2, 3, 4, 5, 6\}$ and $E_2 \subseteq \{\underset{1}{\cancel{H}}, \underset{0}{\cancel{T}}\}$.

$$(d,c) \in \{(1,0), (2,0), (3,0), (4,0), (5,0), (6,0), (1,1), (2,1), (3,1), (4,1), (5,1), (6,1)\}$$

$$\Pr[D \le 3, C = 1] = \frac{3}{12} = \frac{1}{4}$$

$$P[D \le 3] \cdot P[C = 1] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$D \le 3$ and $C = 1$ are independent.

Example: more dice I

Suppose that we roll a fair six-sided die. Consider the following two events:

$$E_1 = \{\text{The die lands on an even number}\}$$
$$E_2 = \{\text{The die lands showing 3}\}.$$

Are these two events independent? If I tell you that the die landed on an even number, then it's certainly impossible that it landed showing a 3, since 3 isn't even. These events should not be independent.

Let's verify that

$$\Pr[E_1 \cap E_2] \neq \Pr[E_1] \Pr[E_2].$$

There are six sides on our die, numbered 1, 2, 3, 4, 5, 6, and three of those sides are even numbers, so $\Pr[E_1] = 1/2$.

The probability that the die lands showing 3 is exactly $\Pr[E_2] = 1/6$.

Putting these together, $\Pr[E_1] \Pr[E_2] = 1/12$.

Example: more dice II

E_1 E_2

On the other hand, let's consider $E_1 \cap E_2$ (the die is even *and* it lands showing three). These two events cannot both happen!

That means that $E_1 \cap E_2 = \emptyset$. Thus $\Pr[\emptyset] = 0$.

(**Aside:** why? **Hint:** $\Pr[\Omega] = 1$ and $\emptyset \cap \Omega = \emptyset$; now use the fact that the probability of the union of disjoint events is the sum of their probabilities).

So we have

$$\Pr[E_1 \cap E_2] = 0 \neq \boxed{\frac{1}{12} = \Pr[E_1]\Pr[E_2].}$$

Our two events are indeed not independent.

$\Pr[E_1] = \frac{1}{2}$

$\Pr[E_2] = \frac{1}{6}$

Independent Random Variables I

Informally, we'll say that two random variables X and Y are independent if **any two events** concerning those random variables are independent.

That is, for *any* event E_X concerning X (i.e., $E_X = \{X \in S\}$ for $S \subseteq \Omega$) and any event E_Y concerning Y, the events E_X and E_Y are independent.

i.e., if two random variables X and Y are independent, then for any two sets $S_1, S_2 \subset \Omega$,

$$\Pr[X \in S_1, Y \in S_2] = \Pr[X \in S_1] \Pr[Y \in S_2].$$

Independent Random Variables II

If X and Y are both discrete, then for any k and ℓ,

$$\Pr[X = k, Y = \ell] = \Pr[X = k] \Pr[Y = \ell].$$

pmf

Similarly, if X and Y are continuous, then the *joint density* has the same property:

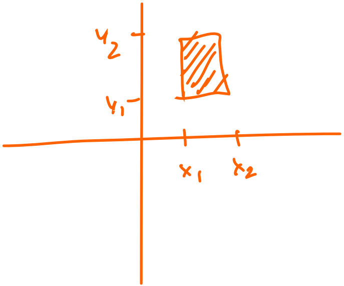
$$f_{X,Y}(s,t) = f_X(s) f_Y(t).$$

pdf

f_X pdf of X

f_Y pdf of Y

$f_{X,Y}$ joint pdf of (X,Y)



(in)dependence, expectation and variance I

Recall the definition of the expectation: If X is continuous with density f_X,

$$\boxed{\mathbb{E}X = \int_\Omega t f_X(t)dt,}$$

and if X is discrete with probability mass function $\Pr[X = k]$,

$$\mathbb{E}X = \boxed{\sum_{k\in\Omega} k \Pr[X = k]}$$

for discrete

$$\boxed{E(g(X)) = \sum_{k\in\Omega} g(k) \Pr[X=k]}$$

With the expectation defined, we can also define the variance,

$$\operatorname{Var} X = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - \mathbb{E}^2 X.$$

$$E[(X - EX)^2]$$

That second equality isn't necessarily obvious– we'll see why it's true in a moment.

Note: we often write $\mathbb{E}^2 X$ as short for $(\mathbb{E}X)^2$. similar to $\sin^2 x = (\sin x)^2$

(in)dependence, expectation and variance II

A basic property of expectation is that it is *linear*. For any constants (i.e., non-random) $a, b \in \mathbb{R}$,

$$\mathbb{E}(aX + b) = a\mathbb{E}X + b.$$

If X, Y are random variables, then

$$\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y.$$

$E(X-Y) = EX - EY$

Note that derivatives and integrals are linear, too. For example,

$$(a\ f(t) + b\ g(t))' = a\ f'(t) + b\ g'(t)$$

and

$$\int (af(t) + bg(t))dt = a \int f(t)dt + b \int g(t)dt$$

Because expected value is simply an integral (or summation), the linearity of expectation follows directly from the definition.

(in)dependence, expectation and variance III

Exercise: prove that $\mathbb{E}(aX + b) = a\mathbb{E}X + b$ for discrete r.v. X.

Exercise: prove that $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$ for discrete X and Y.

Exercise: Use the linearity of expectation to prove that $\mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - \mathbb{E}^2X$.
Hint: $\mathbb{E}(X\mathbb{E}X) = \mathbb{E}^2X$ because $\mathbb{E}X$ is NOT random– it pops right out of the expectation just like a does in the equation above.

(in)dependence, expectation and variance IV

The definition of variance and the linearity of expectation are enough to give us a property of variance:

For any constants (i.e., non-random) $a, b \in \mathbb{R}$,

$$\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X).$$

Exercise: Use the definition $\operatorname{Var}(X) = \mathbb{E}X^2 - \mathbb{E}^2 X$ to prove the above.

$$\operatorname{Var}(aX+b) = E\left(aX+b\right)^2 - E^2\left(aX+b\right)$$

$$\longrightarrow a^2 \operatorname{Var} X$$

(in)dependence, expectation and variance V

This linearity property implies that the expectation of a sum is the sum of the expectations:

$$\mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}X_1 + \mathbb{E}X_2 + \cdots + \mathbb{E}X_n.$$

- does not rely on same distribution or independence

However, the variance of the sum the is not always the sum of the variances.

Consider RVs X and Y.

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathbb{E}[X+Y-\mathbb{E}(X+Y)]^2 \\ &= \mathbb{E}[(X-\mathbb{E}X)+(Y-\mathbb{E}Y)]^2,\end{aligned}$$

where the second equality follows from applying linear of expectation to write $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$.

$\mathrm{Var}(X) = E(X - EX)^2$

(in)dependence, expectation and variance VI

Now, let's expand the square in the expectation.

$$
\begin{aligned}
\operatorname{Var}(X+Y) &= \mathbb{E}[(X-\mathbb{E}X)+(Y-\mathbb{E}Y)]^2 \\
&= \mathbb{E}[(X-\mathbb{E}X)^2+2(X-\mathbb{E}X)(Y-\mathbb{E}Y)+(Y-\mathbb{E}Y)^2] \\
&= \underbrace{\mathbb{E}(X-\mathbb{E}X)^2}_{\mathrm{Var}(X)}+2\underbrace{\mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y)}_{??}+\underbrace{\mathbb{E}(Y-\mathbb{E}Y)^2}_{\mathrm{Var}(Y)},
\end{aligned}
$$

where the last equality is just using the linearity of expectation.

Now, the first and last terms there are the variances of X and Y:

$$\operatorname{Var} X = \mathbb{E}(X-\mathbb{E}X)^2, \quad \operatorname{Var} Y = \mathbb{E}(Y-\mathbb{E}Y)^2.$$

(in)dependence, expectation and variance VII

So

$$\mathrm{Var}(X+Y) = \mathrm{Var}\,X + \underbrace{2\mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y)} + \mathrm{Var}\,Y.$$

The middle term is (two times) the *covariance* of X and Y, often written

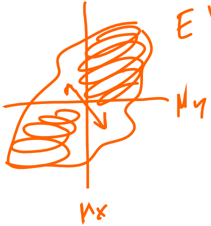
denote $EX = \mu_x$, $EY = \mu_y$

$$\mathrm{Cov}(X,Y) = \mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y).$$

$E\left((X-\mu_x)(Y-\mu_y)\right)$

If $\mathrm{Cov}(X,Y) = 0$, *then*

$$\mathrm{Var}(X+Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y.$$



But when does $\mathrm{Cov}(X,Y) = 0$?

If X and Y are independent random variables, then $\mathrm{Cov}(X,Y) = 0$ (but causality does not work the other way)

X, Y indp $\Longrightarrow \mathrm{Cov}(X,Y) = 0$

Note: We will skip the proof that independence of X and Y implies $Cov(X,Y) = 0$, but you can find this proof in many places online.

(Un)correlation and independence

Covariance might look familiar to you from a quantity that you saw in STAT240 (and a quantity that is very important in statistics!). The (Pearson) correlation between random variables X and Y is defined to be

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{(\mathrm{Var}\, X)(\mathrm{Var}\, Y)}}. = \frac{\mathrm{Cov}(X,Y)}{SD(X) \cdot SD(Y)}$$

Note that if X and Y are independent, then $\rho_{X,Y} = 0$ and we say that they are *uncorrelated*.

But the converse isn't true– it is possible to cook up examples of random variables that are uncorrelated (i.e., $\rho_{X,Y} = 0$), but which are *not* independent.

$$-1 \leq \rho_{X,Y} \leq 1$$

Example: Uncorrelated but not independent

Suppose $X \sim Unif(-1, 1)$ and $Y = X^2$. You can see from the definition that Y is most definitely dependent on X. If, for example, you know that $x = .5$, then you know that $y = .5^2 = .25$. A proof that the covariance is zero is going to be difficult, but we can simulate some data to help us be confident that the claim is true.

```
x <- runif(10000, -1, 1); y <- x^2
cov(x,y)
## [1] -0.001065749
```

The covariance of this sample of 10,000 observations of X and Y is very close to 0 (probably a little below or a little above due to randomness). The sample correlation is

```
cor(x,y)
## [1] -0.006152223
```

Again - this is a number that is very close to zero.

Example: sums of independent normals I

$\mu_{x_2} = 2$

$\sigma^2_{x_2} = 2^2$

$sd = 2$

Let's consider two *independent* normals:

$$X_1 \sim \text{Normal}(1, 1^2) \text{ and } X_2 \sim \text{Normal}(2, 2^2).$$

Since X_1 and X_2 are independent, $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$

1. the variance of their sum should be the sum of their variances, and
2. their correlation should be zero

Let's check both of those facts in simulation. We'll generate lots of copies of X_1 and X_2, and then we'll compute their

1. (sample) variances separately,
2. (sample) variance of their sum, and
3. (sample) correlation

Example: sums of independent normals II

Of course, all three of these quantities will be *estimated* from samples. The law of large numbers tells us that these sample statistics will be close to (but not preceisely equal to) the true quantities.

Example: sums of independent normals III

Okay, let's generate data.

```
M <- 1e5; # Generate 100K Monte Carlo samples
X1 <- rnorm( n=M, mean=1, sd=sqrt(1) );
X2 <- rnorm( n=M, mean=2, sd=sqrt(2) );

# Compute the (sample) variances of the copies of X1 and X2.
v1 <- var(X1);
v2 <- var(X2);

# v1 should be close to 1=Var X_1, v2 close to 2=Var X_2.
c( v1, v2 )
## [1] 1.001805 2.002920
```

Example: sums of independent normals IV

And let's check that these two independent variables have correlation (approximately) zero.

```
# cor( x, y) computes the (sample) correlation between
# the entries of vectors x and y.
# See ?cor for details.
cor( X1, X2 );
## [1] -0.0002642746
```

Again, those *sample-based* quantities will never be precisely equal to 1, 2, and 0, but they will be very close!

Example: sums of independent normals V

Finally, let's check that the variance of the *sum* $X_1 + X_2$ is the sum of variances, as it should be if the RVs are independent. So we should see

$$\operatorname{Var}(X_1 + X_2) = \operatorname{Var} X_1 + \operatorname{Var} X_2 = 1 + 2 = 3.$$

Okay, let's check.

```r
var(X1 + X2)
## [1] 3.003975
```

As we predicted!

Example: multivariate normal I

Remember that the multivariate normal is a way of generating multiple normal random variables that are correlated with one another.

Here's our code from our example modeling the voter shares in Wisconsin and Michigan.

```
mu <- c(.5,.5); # Vector of means; both W_p and M_p are mean 1/2.
# Make a two-by-two symmetric matrix.
CovMx <- matrix( c(.05^2,.04^2,.04^2,.05^2), nrow = 2);
CovMx;
##          [,1]   [,2]
## [1,] 0.0025 0.0016
## [2,] 0.0016 0.0025
```

positive

$$\begin{bmatrix} Var(W) & Cov(W,M) \\ Cov(W,M) & Var(M) \end{bmatrix}$$

The code above generates a multivariate normal with two entries. Both will have means 0.5, encoded in the vector `mu`.

Example: multivariate normal II

The variances and covariance between the two normals is encoded by `CovMx`.

It encodes a matrix (fancy word for an array of numbers), which looks like

$$\Sigma = \begin{bmatrix} 0.05^2 & 0.04^2 \\ 0.04^2 & 0.05^2 \end{bmatrix}.$$

That 0.4^2 in the off-diagonal entries is the *covariance* of the two normals.

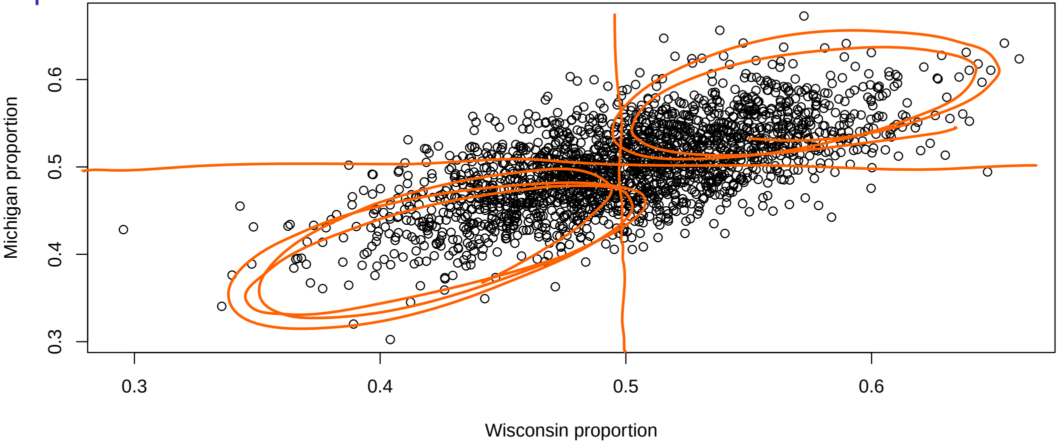
So this will generate two normal random variables, both having mean 0.5, and variance 0.05^2, but these two normals will be correlated, with covariance 0.04^2.

Example: multivariate normal III

Let's have a look:

```
library(MASS); # This library includes a multivariate normal
#mvrnorm is the multivariate version of rnorm.
WpMp = mvrnorm(n=2000, mu=mu, Sigma=CovMx);
plot(WpMp, xlab = "Wisconsin proportion",
     ylab = "Michigan proportion");
```

Example: multivariate normal IV



Example: multivariate normal V

It's clear that the Wisconsin and Michigan voter shares are correlated– we can see it in the plot!

But just to be sure:

```
# WpMp is an array with two columns and 500 rows.
# If we call cov on it directly, we get something shaped
# like our covariance matrix.
cov(WpMp )
##             [,1]        [,2]
## [1,] 0.002610808 0.001665707
## [2,] 0.001665707 0.002500532
```

Example: multivariate normal VI

The diagonal entries are the (sample) variances computed along the columns. The off-diagonal entries (note that they are both the same) tell us the (sample) covariance. Unsurprisingly, the off-diagonal is close to the true covariance 0.0016.

Also worth noting is the fact that the on-diagonal entries are (approximately) 0.025. The on-diagonal entries are computing covariances of our two columns of data *with themselves*. That is, these are computing something like $\mathrm{Cov}(X, X)$.

Example: multivariate normal VII

What is a random variable's covariance with itself? Let's plug in the definition:

$$\mathrm{Cov}(X, X) = \mathbb{E}(X - \mathbb{E}X)(X - \mathbb{E}X) = \mathbb{E}(X - \mathbb{E}X)^2$$

Hey, that's the variance! So $\mathrm{Cov}(X, X) = \mathrm{Var}\, X$.

Fact

$$\mathrm{cov}(X, Y) \le SD(X) \cdot SD(Y)$$

b/c

$$\frac{\mathrm{cov}(X, Y)}{SD(X)\, SD(Y)} = \rho_{X,Y} \le 1$$

How reasonable is independence? I

In most applications, it is pretty standard that we assume that our data are drawn *independently and identically distributed* according to some distribution. We say "i.i.d.". For example, if $X_1, X_2, \ldots, X_n$ are all continuous uniform random variables between 0 and 1, we would say

$$X_i \overset{\text{iid}}{\sim} \text{Uniform}(0, 1), \text{ for } i = 1, \ldots, n$$

This notation is common to denote iid random variables.

As another example, when we perform regression (as you did in STAT240, and which we'll revisit in more detail later this semester), we imagine that the observations (i.e., predictor-response pairs) $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are independent.

How reasonable is independence? II

Most standard testing procedures (e.g., the t-test) assume that data are drawn i.i.d.

How reasonable are these assumptions?

It depends on where the data comes from! We have to draw on what we know about the data, either from our own knowledge or from that of our clients, to assess what assumptions are and aren't reasonable.

Like most modeling assumptions, we usually acknowledge that independence may not be exactly true, but it's often a good approximation to the truth!

How reasonable is independence? III

Example: suppose we are modeling the value of a stock over time. We model the stock's price on days 1, 2, 3, etc as $X_1, X_2, X_3, \ldots$. What is wrong with modeling these prices as being independent of one another? Why might it still be a reasonable modeling assumption?

What if instead we look at the change in stock price from day to day? For example, let $Y_i = X_{i+1} - X_i$. In other words, $X_{i+1} = X_i + Y_i$. Would it be more reasonable to assume that the Y_i's are independent?

What if instead of considering a stock's returns on one day after another, we look at a change in stock price on one day, then at the change 10 days from that, and 10 days from that, and so on? Surely there is still dependence, but a longer time lag between observations *might* make us more willing to accept that our observations are *close* to independent (or at least have much smaller covariance!).

How reasonable is independence? IV

Note: Tobler's first law of geography states 'Everything is related to everything else, but near things are more related than distant things.' Does that ring true in this context?

How reasonable is independence? V

Example: suppose we randomly sample 1000 UW-Madison students to participate in a survey, and record their responses as $X_1, X_2, \ldots, X_{1000}$. What might be the problem with modeling these responses as being independent? Why might be still be a reasonable modeling assumption?

Conditional probability

We can't talk about events and independence without discussing *conditional probability*.

To motivate this, consider the following: suppose I roll a six-sided die. What is the probability that the die lands showing 2?

Now, suppose that I don't tell you the number on the die, but I *do* tell you that the die landed on an even number (i.e., one of 2, 4 or 6). Now what is the probability that the die is showing 2?

We can work out the probabilities by simply counting possible outcomes. Are the probabilities the same?

Example: disease screening

Here's a more real-world (and more consequential example): suppose we are screening for a rare disease. A patient takes the screening test, and tests positive. What is the probability that the patient has the disease, *given* that they have tested positive for it?

We will need to establish the rules of conditional probability before we can tackle a problem such as this.

Introducing conditional probability I

These kinds of questions, in which we want to ask about the probability of an event *given* that something else has happened, require that we be able to define a "new kind" of probability, called *conditional probability*.

Let A and B be two events.

- Example: A could be the event that a die lands showing 2 and B is the event that the die landed on an even number.
- Example: A could be the event that our patient has a disease and B is the event that the patient tests positive on a screening test.

Provided that $\Pr[B] > 0$, we define the *conditional probability* of A given B, written $\Pr[A \mid B]$, according to

given

$$\boxed{\Pr[A \mid B]} = \frac{\Pr[A \cap B]}{\Pr[B]}.$$

Note that if $\Pr[B] = 0$, then the ratio on the right-hand side is not defined, hence why we demanded that $\Pr[B] > 0$.

Introducing conditional probability II

Let's try computing one of these conditional probabilities: what is the probability that the die is showing 2 *conditional on the fact* that it landed on an even number?

Well,

- $\Pr[\text{ even }] = 1/2$, because there are three even numbers on the die, and all six numbers are equally likely: $3/6 = 1/2$.
- $\Pr[\text{ die lands 2 } \cap \text{even}] = \Pr[\text{ die lands 2 }]$, since 2 is an even number.

So the conditional probability is

$$\begin{aligned}
\Pr[\text{ die lands 2 } \mid \text{ even }] &= \frac{\Pr[\text{ die lands 2 } \cap \text{even}]}{\Pr[\text{ even }]} \\
&= \frac{\Pr[\text{ die lands 2 }]}{\Pr[\text{ even }]} \\
&= \frac{1/6}{1/2} = 1/3.
\end{aligned}$$

Introducing conditional probability III

This makes sense– *given* that the die lands on an even number, we are choosing from among three outcomes: $\{2, 4, 6\}$. The probability that we choose 2 from among these three possible equally-likely outcomes is $1/3$.

Disease screening

What about our disease testing example? What is the probability that our patient has the disease given that they tested positive?

Well, applying the definition of conditional probability,

$$\Pr[\text{ disease } | \text{ positive test }] = \frac{\Pr[\text{ disease} \cap \text{ positive test }]}{\Pr[\text{positive test}]}$$

Okay, but what is Pr[positive test]? I guess it's just the probability that a random person (with the disease or not) tests positive? For that matter, what is Pr[disease ∩ positive test]? These can be hard events to assign probabilities to! Luckily, there is a famous equation that often gives us a way forward.

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]} \rightarrow \begin{array}{l} \Pr[A \cap B] = \Pr[B] \cdot \Pr[A|B] \\ \Pr[A \cap B] = \Pr[A] \cdot \Pr[B|A] \end{array}$$

Bayes' rule I

The Reverend Thomas Bayes was the first to suggest an answer to this issue. Bayes' rule, as it is now called, tells us how to relate $\Pr[A \mid B]$ to $\Pr[B \mid A]$:

$$\frac{\cancel{\Pr[B]}\Pr[A \mid B]}{\cancel{\Pr[B]}} = \qquad \Pr[A \mid B] = \frac{\overbrace{\Pr[B \mid A]\Pr[A]}^{\Pr[A \cap B]}}{\Pr[B]}.$$

This is useful, because it is often easier to write one or the other of these two probabilities.

Applying this to our disease screening example,

power of the test
PPV
prevelance in the population

$$\boxed{\Pr[\text{ disease} \mid \text{ positive test }] = \frac{\Pr[\text{ positive test } \mid \text{ disease}]\Pr[\text{ disease}]}{\Pr[\text{ positive test }]}}$$

Bayes' rule II

The advantage of using Bayes' rule in this context is that the probabilities appearing on the right-hand side are all straight-forward to think about (and estimate!).

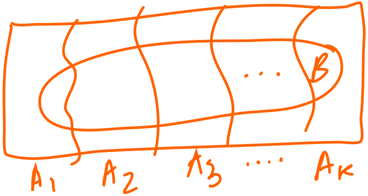
- $\Pr[\text{ disease}]$ is the just the probability that a randomly-selected person has the disease. This is known as the *prevelance* of the diseases in the population. We could estimate this probability by randomly selecting a random group of people and determining if they have the disease (hopefully not using the screening test we are already using...).
- $\Pr[\text{ positive test } \mid \text{ disease}]$ is the probability that when we give our screening test to a patient who has the disease in question, the test returns positive. This is often called the sensitivity of a test, a term you may recall hearing frequently in the early days of the COVID-19 pandemic.

Bayes' rule III

- Pr[positive test] is just the probability that a test given to a (presumably randomly selected) person returns a positive result. We just said about that this is the hard thing to estimate. In your homework, ~~you'll explore one way to get~~ at this quantity, but for now we'll have to just assume that we can estimate it somehow or other.

$$\Pr[\text{Positive test}] = \Pr\left[(\text{Positive test} \cap \text{Disease}) \text{ or } (\text{Pos test} \cap \text{No Disease})\right]$$

If $A_1, A_2, \ldots, A_k$ are m.e.



$$B = (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_3)$$

$$\Rightarrow \Pr(B) = \Pr(B \cap A_1) + \Pr(B \cap A_2) + \cdots + \Pr[B \cap A_3]$$

Example: testing for a rare disease I

Suppose that we are testing for a rare disease, say,

$$\Pr[\text{ disease}] = \frac{1}{10^6},$$

and suppose that a positive test is also rare, in keeping with the fact that our disease is rare and our test presumably has a low false positive rate:

$$\Pr[\text{ positive test}] = 1.999 * 10^{-6}$$

Note that this probability actually depends on the *sensitivity* $\Pr[\text{ positive test } \mid \text{ disease}]$ and the *specificity* $1 - \Pr[\text{ positive test } \mid \text{ healthy}]$ of our test. You'll explore this part more on your homework, but we're just going to take this number as given for now.

Example: testing for a rare disease II

Finally, let's suppose that our test is 99.99% accurate:

$$\Pr[\text{ positive test } \mid \text{ disease}] = 0.9999 = 1 - 10^{-4}$$

To recap,

$$\begin{aligned}
\Pr[\text{ disease}] &= \frac{1}{10^6} \\
\Pr[\text{ positive test}] &= 1.999 * 10^{-6} \\
\Pr[\text{ positive test } \mid \text{ disease}] &= 0.9999.
\end{aligned}$$

Now, suppose that a patient is given the screening test and receives a positive result. Bayes' rule tells us

$$\begin{aligned}
\Pr[\text{ disease} \mid \text{ positive test }] &= \frac{\Pr[\text{ positive test } \mid \text{ disease}] \Pr[\text{ disease}]}{\Pr[\text{ positive test }]} = \frac{0.9999 * 10^{-6}}{1.999 * 10^{-6}} \\
&= 0.5002001.
\end{aligned}$$

Example: testing for a rare disease III

So even in light of our positive screening test result, the probability that our patient has the disease in question is still only about 50%!

This is part of why, especially early on in the pandemic when COVID-19 was especially rare, testing for the disease in the absence of symptoms was not considered especially useful.

More generally, this is why most screenings for rare diseases are not done routinely–doctors typically screen for rare diseases only if they have a reason to think a patient is more likely to have that disease for other reasons (e.g., family history of a genetic condition or recent exposure to an infectious disease).

Calculating the denominator in Bayes' Rule I

The denominator can be decomposed into two parts using a property known as the Law of Total Probability.

$$\Pr[\text{ positive test}] = \Pr[\text{ positive test} \cap \text{disease}] + \Pr[\text{ positive test} \cap \text{no disease}]$$

In other words, all positive results are either true positives or false positives. Because these are mutually exclusive events, the total probability of a positive result is the probability of a true positive plus the probability of a false positive. We can expand each of these terms using the conditional probability rule.

$$\Pr[\text{ positive test} \cap \text{disease}] = \Pr[\text{ positive test } \mid \text{ disease}]\Pr[\text{ disease}]$$

$$\Pr[\text{ positive test} \cap \text{no disease}] = \Pr[\text{ positive test } \mid \text{ no disease}]\Pr[\text{no disease}]$$

Calculating the denominator in Bayes' Rule II

For example, suppose that a genetic condition occurs in roughly 1 out of 800 individuals. A simple saliva test is available. If a person has the gene, the test is positive with 97% probability. If a person does not have the gene, a false positive occurs with 4% probability.

To simplify notation, let G represent "the individual has the gene" and G' be the complementary event that "the individual does not have the gene." Furthermore, let *Pos* and *Neg* represent the test results.

If a random person from the population takes the test and gets a positive result, what is the probability they have the genetic condition?

Bayes' Rule to the rescue:

$$\begin{aligned} P[G|Pos] &= \frac{P[Pos|G]P[G]}{P[Pos|G]P[G] + P[Pos|G']P[G']} \\ &= \frac{(.97)(1/800)}{(.97)(1/800) + (.04)(799/800)} \end{aligned}$$

Calculating the denominator in Bayes' Rule III

In other words, a positive test result would raise the likelihood of the gene being present from $1/800 = 0.00125$ up to .0295.

Dependent free throw shots

Suppose a basketball player's likelihood of making a basket when making a free throw depends on the previous attempt. On the first throw, they have a probability of 0.67 of making the basket. On the second throw, following a basket the probability goes up to .75. If the first throw is a miss, the probability of a basket on the second throw goes down to 0.62.

Exercise: If the second throw is a basket, what is the likelihood the first throw is a basket?

Exercise: Given that the player scores at least 1 point, what is the probability that they score 2 points total?

Review:

In these notes we covered:

- The concept of independent events
- Independent random variables
- Definition of variance
- Expectation of a linear combination of r.v.s
- Variance of a linear combination of r.v.s
- Covariance and correlation
- Relationship between correlation and independence
- When the independence assumption is reasonable
- Conditional probability & the general multiplication rule
- Bayes' rule