Introduction to Probability
Conditional probability and independence

Conditional probability

The **conditional probability** of the event E, given the event F, is defined, if $P(F) > 0$, as

$$P(E \mid F) = \frac{P(EF)}{P(F)}.$$

Conditional probability

The interpretation of this: if you are “given” the information that the event F has occured, then your probability structure changes to shrink the sample space down to only the outcomes in F.

For uniform probabilities, this is clear in the calculation:

$$P(E \mid F) = \frac{P(EF)}{P(F)} = \frac{\frac{|EF|}{|\Omega|}}{\frac{|F|}{|\Omega|}} = \frac{|EF|}{|F|}.$$

(For non-uniform probabilities, the definition still holds, but this counting argument doesn't work.)

Conditional probability is a probability!

$$P(E \mid F) = \frac{P(EF)}{P(F)}$$

If we define a new function $Q_F : 2^\Omega \to \mathbb{R}$ by $Q_F(E) = P(E \mid F)$, then Q_F is also a probability measure.

We can verify this by checking the axioms of probability.

Conditional probability is a probability!

1. $\forall E \in 2^{\Omega}$, $0 \leq Q_F(E) = \frac{P(EF)}{P(F)} \leq 1$
   since $P(EF) \leq P(F)$ by monotonicity. ✓
2. **(normalization)** $Q_F(\Omega) = \frac{P(\Omega F)}{P(F)} = \frac{P(F)}{P(F)} = 1$ ✓
3. **(countable additivity)** If E_1, E_2, ... are pairwise disjoint events, then E_1F, E_2F, ... are also pairwise disjoint. Hence,

$$Q_F\left(\bigcup_{n=1}^{\infty} E_n\right) = \frac{P\left(\bigcup_{n=1}^{\infty} E_nF\right)}{P(F)} = \sum_{n=1}^{\infty} \frac{P\left(E_nF\right)}{P(F)} = \sum_{n=1}^{\infty} Q_F(E_n). \checkmark$$

Example: coin flips

Example

I flip a fair coin twice.

1. What is the probability it flips heads twice?
2. What is the probability it flips heads twice, *given that at least one flip is heads*?

Let the experiment be "flip a fair coin twice".

The sample space

$$\Omega = \{HH, HT, TH, TT\}.$$

Example: coin flips

1. What is the (classical, uniform) probability of flipping heads twice? Let $E = \{HH\}$.

$$P(E) = \frac{1}{4}.$$

2. What is the probability of flipping HH, given that at least one flip is heads?

The event

$$F = \{\text{“at least one flip is heads”}\} = \{HT, TH, HH\}$$

removes the possibility of TT. Thus, $EF = \{HH\}$, and so

$$P(E \mid F) = \frac{P(EF)}{P(F)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Partial information about the flip conditioned our probability.

Example: coin flips

Example

What is the probability of flipping heads twice, *given that at least one flip is tails*?

Intuitively, you should know the probability is zero, since you can't have two heads and a tail when you only have two flips.

The theory bears this out.

Example: coin flips

Let $E = \{HH\}$ and $G = \{\text{at least one tail}\} = \{TH, HT, TT\}$.

$EG = \{\text{two heads and at least one tail in two flips}\} = \emptyset$.

Thus, $P(EG) = 0$, and therefore $P(E \mid G) = \frac{P(EG)}{P(G)} = 0$.

Venn diagrams and conditional probability

A Venn diagram can immediately get across the notion of conditional probability: instead of considering the entire box (the "universe" sample space Ω), we effectively shrink the sample space down to just the evidence event F.

We'll see a couple examples using the magazine survey example from last time.

Example

In an old survey of 75 college students on their reading of three magazines, it was found that:

- 23 read Time (and possibly more)
- 18 read Newsweek (and possibly more)
- 14 read US News (and possibly more)
- 10 read Time and Newsweek (and possibly more)
- 9 read Time and US News (and possibly more)
- 8 read Newsweek and US News (and possibly more)
- 5 read all three.

Examples

1. Given that a random student reads Newsweek, what is the probability that student also reads Time?
2. What is the probability that a student reads US News, given that the student reads only one magazine?
3. Given that a student reads exactly two magazines, what is the probability the one not read is Time?
4. Given that a random student reads Time, what is the probability that student also reads Newsweek?

Notice that questions 1 and 4 ask about the same events, but the conditioning is in the opposite order. We'll address this issue shortly.

Examples

1. Given that a random student reads Newsweek, what is the probability that student also reads Time? Answer: $\frac{10}{18}$
2. What is the probability that a student reads US News, given that the student reads only one magazine? Answer: $\frac{2}{15}$
3. Given that a student reads exactly two magazines, what is the probability the one not read is Time? Answer: $\frac{3}{12}$
4. Given that a random student reads Time, what is the probability that student also reads Newsweek? Answer: $\frac{10}{23}$

Notice that questions 1 and 4 ask about the same events, but the conditioning is in the opposite order. We'll address this issue shortly.

$P(EF) = P(F)P(E \mid F)$

If we know a conditional probability and its evidence probability, we can use these to find the intersection probability.

$$P(E \mid F) = \frac{P(EF)}{P(F)} \implies P(EF) = P(F)P(E \mid F).$$

Also note that, if $P(E \mid F) = P(E)$, then we say E and F are **independent events**.

We'll come back to independence later on; for now, think of it as "the evidence doesn't tell us anything new".

In other words, the "evidence" doesn't change our "probability".

Example: $P(EF) = P(F)P(E \mid F)$

Flip a fair coin three times. What is the probability the first flip is H, given that the third flip is T?

Intuitively: The first and third flips have nothing to do with each other! They are independent, identical trials, and so should have the same probability structure!

Formally: Let S be the sample space of all 3-flip sequences, $A = \{$ H on flip 1 $\}$, and $B = \{$ T on flip 3 $\}$. Then

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\},$$
$$A = \{HHH, HHT, HTH, HTT\},$$
$$B = \{HHT, HTT, THT, TTT\}; \; AB = \{HHT, HTT\}.$$

$$\therefore P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{2/8}{4/8} = \frac{1}{2} = P(A).$$

$P(EF) = P(F)P(E \mid F)$ multiplication rule

In general, if $E_1, E_2, ..., E_n$ are n events, with nonempty intersections for any pairing, we have the general rule

$$P(E_1 E_2 \cdots E_n) = P(E_1) \cdot P(E_2 \,|E_1) \cdot P(E_3 \,|E_1 E_2) \cdots P(E_n \mid E_1 E_2 \cdots E_{n-1}).$$

To see this for $n = 3$:

$$\begin{aligned} P(E_1) \cdot P(E_2 \,|E_1) \cdot P(E_3 \,|E_1 E_2) &= P(E_1) \cdot \frac{P(E_1 E_2)}{P(E_1)} \cdot \frac{P(E_1 E_2 E_3)}{P(E_1 E_2)} \\ &= P(E_1 E_2 E_3). \end{aligned}$$

$P(E^C \mid F) = 1 - P(E \mid F)$

Recall that $Q_F(E) = P(E \mid F)$ is a probability measure, different from P. However, all the regular probability properties hold for Q_F. Hence, complements in S give the same probability property:

$$Q_F(E) + Q_F(E^C) = 1 \implies Q_F(E^C) = 1 - Q_F(E).$$

Example

Roll 2 dice. What is the probability that the roll does *not* sum to 2, given that the sum is at most 4?

Let $E = \{\text{sum is } 2\}$, $F = \{\text{sum is} \leq 4\} = \{\text{sum is } 2,\ 3, \text{ or } 4\}$.
Then $Q_F(E^C) = 1 - Q_F(E) = 1 - \frac{P(EF)}{P(F)} = 1 - \frac{\frac{1}{36}}{\frac{6}{36}} = \frac{5}{6}$.

Partition of a Sample Space

Recall that an event and its complement are disjoint, and their union is the whole sample space. That is,

$$E \cup E^C = \Omega$$
$$E \cap E^C = \emptyset$$

Thus, if we know conditional probabilities but not a full probability, we can decompose a full probability into intersections to calculate.

This is the simplest case of a **partition** of a sample space.

Partition of a Sample Space

A **partition** of a sample space Ω is a set of disjoint subsets whose union are all of Ω.

$\{E_1, E_2, ..., E_n\}$ partitions Ω if, for any $i \neq j$, $E_i \cap E_j = \emptyset$, and

$$\bigcup_{k=1}^{n} E_k = \Omega.$$

Law of total probability: general case

The Law of Total Probability uses a partition to decompose the probability of an event into probabilities of intersections of that event with the pieces of a partition.

Law of Total Probability: If $F \subseteq \Omega$ is an event, and $\{E_1, E_2, ..., E_n\}$ partitions Ω, then

$$P(F) = \sum_{k=1}^{n} P(FE_k) = \sum_{k=1}^{n} P(F \mid E_k) P(E_k).$$

This means you can compute the probability of F based on different "pieces of evidence".

Law of total probability: $n = 2$

In the simplest case, $n = 2$, the Law of Total Probability uses the partition $\{E, E^C\}$, and reads

$$P(F) = P(FE) + P(FE^C) = P(F \mid E)P(E) + P(F \mid E^C)P(E^C).$$

This simplifies further since $P(E^C) = 1 - P(E)$, although we cannot easily compare $P(F \mid E)$ and $P(F \mid E^C)$.

Example: Law of total probability: $n = 2$

Example

I will have my umbrella 95% of the times when it is raining.
I will have my umbrella 10% of the times when it is not raining.
It rains 25% of the time.

What is the probability I have my umbrella with me (no matter what the weather is like)?

Example: Law of total probability: $n = 2$

Let
$R =$ the event that it is raining, and
$U =$ the event that I have my umbrella.

We know:

$$P(U \mid R) = 0.95, \ P(U \mid R^C) = 0.1, \ P(R) = 0.25.$$

Then

$$\begin{aligned} P(U) &= P(U \mid R)P(R) + P(U \mid R^C)P(R^C) \\ &= 0.95(0.25) + 0.1(0.75) = 0.3125. \end{aligned}$$

Under these rules, I have my umbrella 31.25% of all time.

$P(E \mid F) \neq P(F \mid E)$

Note that $P(EF)$ shows up in both $P(E \mid F)$ and $P(F \mid E)$:

$P(E \mid F) = \frac{P(EF)}{P(F)}$ is "the probability of E, given F".

$P(F \mid E) = \frac{P(EF)}{P(E)}$ is "the probability of F, given E".

These are different probabilities!

$P(E \mid F) \neq P(F \mid E)$

In fact, there are only two special cases where $P(E \mid F) = P(F \mid E)$:

$$
\begin{aligned}
P(E \mid F) = P(F \mid E) \implies \frac{P(EF)}{P(F)} &= \frac{P(EF)}{P(E)} \\
\implies P(F) = P(E) &\text{ or } P(EF) = 0.
\end{aligned}
$$

We should understand the difference between these two conditional probabilities, since these special cases are not the general situaiton.

$P(E \mid F) \neq P(F \mid E)$

Example

What is the probability that I have a rare disease,
given that a test comes up positive? (A *false positive* is possible.)

is a very different question from

What is the probability that a test comes up positive,
given that I have a rare disease? (A *false negative* is possible.)

Bayes' formula (theorem, law)

How do we compare $P(E \mid F)$ and $P(F \mid E)$ in general?

$$\begin{aligned} P(EF) &= P(E \mid F)P(F) \\ P(EF) &= P(F \mid E)P(E) \\ \implies P(E \mid F)P(F) &= P(F \mid E)P(E) \end{aligned}$$

Bayes' Law:

$$P(E \mid F) = P(F \mid E) \cdot \frac{P(E)}{P(F)}.$$

When does $P(E \mid F) = P(F \mid E)$?

$$P(E) = P(F) \text{ or } P(EF) = 0.$$

Example: Coins

Flip a fair coin three times. As usual in this case,

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

1. What is the probability that you flip at least two H, given that you flip at least one T?
2. What is the probability that you flip at least one T, given that you flip at least two H?

Example: Coins

Let E = "at least two H" and F = "at least one T". We'll list:

$$\begin{aligned} E &= \{HHH, HHT, HTH, THH\} \\ F &= \{HHT, HTH, HTT, THH, THT, TTH, TTT\} \\ EF &= \{HHT, HTH, THH\} \end{aligned}$$

1. $P(E \mid F) = \frac{3}{7}$
2. $P(F \mid E) = P(E \mid F) \cdot \frac{P(F)}{P(E)} = \frac{3}{7} \cdot \frac{7}{4} = \frac{3}{4}$.

Example: Urns

An urn contains five green balls, seven yellow balls, and nine orange balls. Draw three without replacement.

$$|\Omega| = \binom{5+7+9}{3} = \binom{21}{3}.$$

1. What is the probability that you draw a yellow, given that you do not draw a green?
2. What is the probability that you do not draw a green, given that you draw a yellow?

Example: Urns

Let E = "draw at least one yellow" and F = "do not draw a green". Then EF = "at least one yellow, and no green".

$$
\begin{aligned}
|E| &= \binom{21}{3} - \binom{21-7}{3} \\
&= \frac{21(20)(19) - 14(13)(12)}{6}; \\
|F| &= \binom{21-5}{3} = \frac{16(15)(14)}{6}; \\
|EF| &= \binom{21-5}{3} - \binom{21-7-5}{3} \\
&= \frac{16(15)(14) - 9(8)(7)}{6}.
\end{aligned}
$$

Example: Urns

Then

$$P(E \mid F) = \frac{16(15)(14) - 9(8)(7)}{16(15)(14)} = 1 - \frac{9}{60} = \frac{17}{20} = 0.85.$$

$$\begin{aligned} P(F \mid E) &= P(E \mid F) \cdot \frac{P(F)}{P(E)} \\ &= \frac{16(15)(14) - 9(8)(7)}{21(20)(19) - 14(13)(12)} = \frac{34}{69} \approx 0.4928. \end{aligned}$$

Example: Fair vs Fake Coin?

I have two coins in my pocket: a fair coin and a double-headed coin. I take one out (fairly) at random and flip it.

I tell you it came up H.

Given this information, what is the probability I flipped the fair coin?

Example: Fair vs Fake Coin?

Let $F =$ "fair coin" and $D =$ "double-headed coin".

Since I pull one of the two fairly from my pocket, $D = F^C$ and

$$P(F) = P(D) = \frac{1}{2}.$$

Clearly, if I have the double-headed coin, it will always come up heads. The fair coin is ... fair.

$$P(H \mid F) = \frac{1}{2};\ P(H \mid D) = 1.$$

Example: Fair vs Fake Coin?

Hence, by Bayes' Law and the Law of Total Probability,

$$\begin{aligned}
P(F \mid H) &= \frac{P(H \mid F)P(F)}{P(H)} \\
&= \frac{P(H \mid F)P(F)}{P(H \mid F)P(F) + P(H \mid F^C)P(F^C)} \\
&= \frac{P(H \mid F)P(F)}{P(H \mid F)P(F) + P(H \mid D)P(D)} \\
&= \frac{P(H \mid F)}{P(H \mid F) + P(H \mid D)} \qquad \left(P(F) = P(D) = \frac{1}{2}\right) \\
&= \frac{\frac{1}{2}}{\frac{1}{2} + 1} = \frac{1}{3}.
\end{aligned}$$

Note that $P(F \mid H) < \frac{1}{2}$, but $P(F) = \frac{1}{2}$.

Example: Fair vs Fake Coin?

I flip the same coin again. H a second time... now, what is the (conditional) probability it is the fair coin?

$$
\begin{aligned}
P(HH \mid F) &= \frac{1}{4};\ P(HH \mid D) = 1 \\
P(F \mid HH) &= \frac{P(HH \mid F)P(F)}{P(HH \mid F)P(F) + P(HH \mid D)P(D)} \\
&= \frac{P(HH \mid F)}{P(HH \mid F) + P(HH \mid D)} \qquad \left(P(F) = P(D) = \frac{1}{2}\right) \\
&= \frac{\frac{1}{4}}{\frac{1}{4} + 1} = \frac{1}{5}.
\end{aligned}
$$

$P(F \mid HH) < P(F \mid H) < P(F)$. What does this mean?

Can you generalize this for n flips of H in a row?

Example: Test Screening

A rare disease has **prevalence**, i.e. “sick rate”, s in the general population; a randomly selected person from the population has the disease with probability s.

A blood test to screen for the disease has the following statistics:

- **true positive** rate **(sensitivity)** of p,
- **true negative** rate **(specificity)** of n,
- **false positive (Type I error)** rate of $1 - n$,
- **false negative (Type II error)** rate of $1 - p$.

You are given this test; it comes up positive. Given (only) this information, what is the probability you have the disease?

Example: Test Screening

Let

- $S =$ "you are sick",
- $H = S^C =$ "you are healthy",
- $T^+ =$ "you test positive", and
- $T^- = T^{+C} =$ "you test negative".

We are given

- **prevalence**, i.e. population sick rate, is $P(S) = s$,
- **sensitivity**, i.e. **true positive** rate of test, is $P(T^+ \mid S) = p$,
- **specificity**, i.e. **true negative** rate of test, is $P(T^- \mid H) = n$,

and calculate

- **false positive** rate of test, is $P(T^+ \mid H) = 1 - n$,
- **false negative** rate of test, is $P(T^- \mid S) = 1 - p$.

Example: Test Screening

We want to know, given a positive test result, what is the probability you are sick, i.e. $P(S \mid T^+)$.

Using Bayes' Law and the Law of Total Probability, this is

$$
\begin{aligned}
P(S \mid T^+) &= \frac{P(T^+ \mid S)P(S)}{P(+)} \\
&= \frac{P(T^+ \mid S)P(S)}{P(T^+ \mid S)P(S) + P(T^+ \mid H)P(H)} = \frac{ps}{ps + (1-n)(1-s)},
\end{aligned}
$$

a very different number than the true positive rate p (which is what a person typically assumes is the correct probability to use in this situation).

Example: Test Screening

One example of possible values are: $s = 1\%$ incidence, $p = 95\%$ true positive, $n = 99\%$ true negative. Then

$$P(S \mid +) = \frac{ps}{ps + (1-n)(1-s)} = \frac{0.95(0.01)}{0.95(0.01) + (0.01)(0.99)} \approx 48.97\%.$$

Other applications for this setting include drug tests, criminal profiling, insurance fraud, ...

Independent events

If we know a conditional probability and its evidence probability, we can use these to find the intersection probability.

$$P(E \mid F)P(F) = P(EF) = P(F \mid E)P(E)$$

If $P(E \mid F) = P(E)$, then we call E and F **independent events**. This is, by symmetry, equivalent to $P(F \mid E) = P(F)$.

Independent events

Think of independence as "the evidence doesn't tell us anything new", i.e. "the evidence doesn't change our probability".

In general, when not considering conditioning, E and F are called **independent events** if

$$P(EF) = P(E) \cdot P(F).$$

Pairwise, General Independence

Two events E and F are **independent** if $P(EF) = P(E)P(F)$, and *dependent* otherwise.

In general, n events E_1, ..., E_n are called **pairwise independent** if $P(E_iE_j) = P(E_i)P(E_j)$ for any pair $i \neq j$, $i, j = 1, 2, ..., n$.

Pairwise, General Independence

For E_1, ..., E_n to be considered **independent** as a full collection of events, they need to satisfy this "factoring" property for any possible combination of them: all triples need to be independent:

$$P(E_i E_j E_k) = P(E_i)P(E_j)P(E_k) \ \forall i, j, k = 1, 2, ..., n, i < j < k,$$

all sets of four, ..., all the way up to

$$P(E_1 E_2 \cdots E_n) = P(E_1)P(E_2) \cdots P(E_n).$$

IID trials

Certain examples are obvious: any repeated, identical trials are independent. We even put that fact in their name:

independent, identically distributed (iid) random trials.

Any type of compound events that can be counted solely via the **multiplication principle** are independent.

Successive coin flips, die rolls, card draws or ball-in-urn draws *with replacement*, ...

Example: independent events

Example

Show that, on a roll of two six-sided dice, the first roll coming up 3 and the second 4 are independent events.

Solution Let the sample space be $\Omega = \{(i,j) : i,j \in \{1,2,3,4,5,6\}\}$ and the events

$$A = \{(i,j) \in \Omega : i = 3\}, \ B = \{(i,j) \in \Omega : j = 4\}.$$

Example: independent events

Then $|\Omega| = 36$, $A \cap B = \{(3,4)\}$, $|A| = 6$, $|B| = 6$, $|A \cap B| = 1$, and the probabilities in question are

$$P(A)P(B) = \frac{6}{36} \cdot \frac{6}{36} = \frac{1}{36} = P(A \cap B).$$

Therefore, A and B are independent.

Examples

Show that, on a roll of two six-sided dice, the first roll coming up 3 and the *sum* 5 are *not* independent events.

Solution Using Ω and A from before, let

$$C = \{(i, j) \in \Omega : i + j = 5\}.$$

Then

$$C = \{(1,4), (2,3), (3,2), (4,1)\},\ A \cap C = \{(3,2)\},$$

and $|C| = 4$ and $|A \cap C| = 1$. Hence,

$$P(A)P(C) = \frac{6}{36} \cdot \frac{4}{36} = \frac{1}{54} \neq \frac{1}{36} = P(A \cap C).$$

Therefore, A and C are not independent.

Examples

Show that, on a roll of two six-sided dice, the first coming up 3 and the *sum* 3 are *not* independent events.

Solution Using Ω and A from before, let

$$D = \{(i,j) \in \Omega : i + j = 3\} = \{(1,2),(2,1)\}.$$

Then, clearly, $A \cap D = \emptyset$, but A and D are both nonempty, so without even calculating probabilities, we can see

$$P(A \cap D) = 0 \neq P(A)P(D).$$

Hence, A and D are not independent.

Examples

Show that, on three coin flips, show that each coming up H is independent of the others. Of course, we start with

$$\Omega = \{f_1 f_2 f_3 : f_i \in \{H, T\}, i = 1, 2, 3\}, \ |\Omega| = 8.$$

Solution Let $A_1 = \{Hf_2 f_3\}$, $A_2 = \{f_1 H f_3\}$, $A_3 = \{f_1 f_2 H\}$. Then

$$P(A_1)P(A_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(A_1 \cap A_2)$$
$$P(A_1)P(A_3) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(A_1 \cap A_3)$$
$$P(A_2)P(A_3) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(A_2 \cap A_3)$$

which shows that the three are pairwise independent. For full,

$$P(A_1)P(A_2)P(A_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} = P(A_1 \cap A_2 \cap A_3).$$

Independence implies independence of complements

Proposition

If E and F are independent events,
then so are the pairs E^C and F, E and F^C, and E^C and F^C.

Proof

$$\begin{aligned} P(EF) &= P(E)P(F) \\ \implies P(E^C F) &= P(F) - P(EF) \text{ (by monotonicity)} \\ &= P(F) - P(E)P(F) \text{ (by independence)} \\ &= P(F)(1 - P(E)) = P(F)P(E^C). \end{aligned}$$

The rest of the proof follows similarly to result in

$$P(EF^C) = P(E)P(F^C),$$
$$P(E^C F^C) = P(E^C)P(F^C). \blacksquare$$

Example: Independence via conditioning

You draw two cards face down and are told (given) that at least one card is a 2. Does this evidence affect the event that both cards are diamonds?

Solution

Let $D = \{\text{both cards are } \diamondsuit\}$, $T = \{\text{at least one card is a } 2\}$.

We want to check if D and T are independent, when T was given as evidence.

Example: Independence via conditioning

We will do this by the following reasoning: if we can show that D and T^C are independent, then by the previous proposition, D and T are independent as well.

If D and T^C are not independent, then neither are D and T.

We'll use $T^C = \{\text{neither card is a 2}\}$ as evidence in our calculation since it is easier to calculate.

Example: Independence via conditioning

We know the sample space is of size

$$|\Omega| = \binom{52}{2} = \frac{52(51)}{2} = 1326$$

since we are drawing two cards.

$$|D| = \binom{13}{2} = \frac{13(12)}{2} = 78, \ |T^C| = \binom{48}{2} = \frac{48(47)}{2} = 1128$$

$$|D \cap T^C| = \binom{12}{2} = \frac{12(11)}{2} = 66.$$

We check if $P(D \mid T^C) =$ or $\neq P(D)$:

$$P(D \mid T^C) = \frac{P(D \cap T^C)}{P(T^C)} = \frac{66}{1128} \neq \frac{78}{1326} = P(D).$$

Hence, D and T^C are not independent, and therefore D and T are not independent.

Bernoulli random variables

X is called a **Bernoulli random variable** *with parameter* p (written $\text{Bern}(p)$) if its PMF is

$$p_X(1) = p,\ p_X(0) = 1 - p.$$

This is the RV of a (biased) coin that flips 1 on H with probability p and 0 on T.

Example

"Roll a 5 on a die" has success probability $\frac{1}{6}$, and so failure probability $\frac{5}{6}$. Thus, for $X \sim \text{Bern}(\frac{1}{6})$, the probability you roll a 5 is

$$p_X(1) = \frac{1}{6}.$$

Binomial random variables

X is called a **binomial random variable** *with parameters* n, p (written $\text{Bin}(n, p)$) if its PMF is

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, 2, ..., n.$$

This is the RV that adds up n Bernoulli RVs above: if X_1, X_2, ..., X_n are IID $\text{Bern}(p)$, then $X = \sum_{i=1}^n X_i \sim Bin(n, p)$.

Example

"Roll a 5 on a die" has success probability $\frac{1}{6}$, and so failure probability $\frac{5}{6}$. Thus, for $X \sim \text{Bin}(7, \frac{1}{6})$, the probability you roll a 5 exactly three times out of seven is

$$p_X(3) = \binom{7}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^4.$$

Geometric random variables

X is called a **geometric random variable** *with parameter* p (written $\mathrm{Geom}(p)$) if its PMF is

$$p_X(k) = p(1-p)^{k-1},\ k = 1, 2, \ldots.$$

This RV represents the number of trials up to a "success" in a run of repeated IID experiments with "success" probability p. That is, $k-1$ "failures", and then "success" on trial k.

Example

"Roll a 5 on a die" has success probability $\frac{1}{6}$, and so failure probability $\frac{5}{6}$. Thus, for $X \sim \mathrm{Geom}(\frac{1}{6})$, the probability it takes exactly 4 rolls to get the first 5 is

$$p_X(4) = \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^3.$$