Chapter 2. Discrete Probability

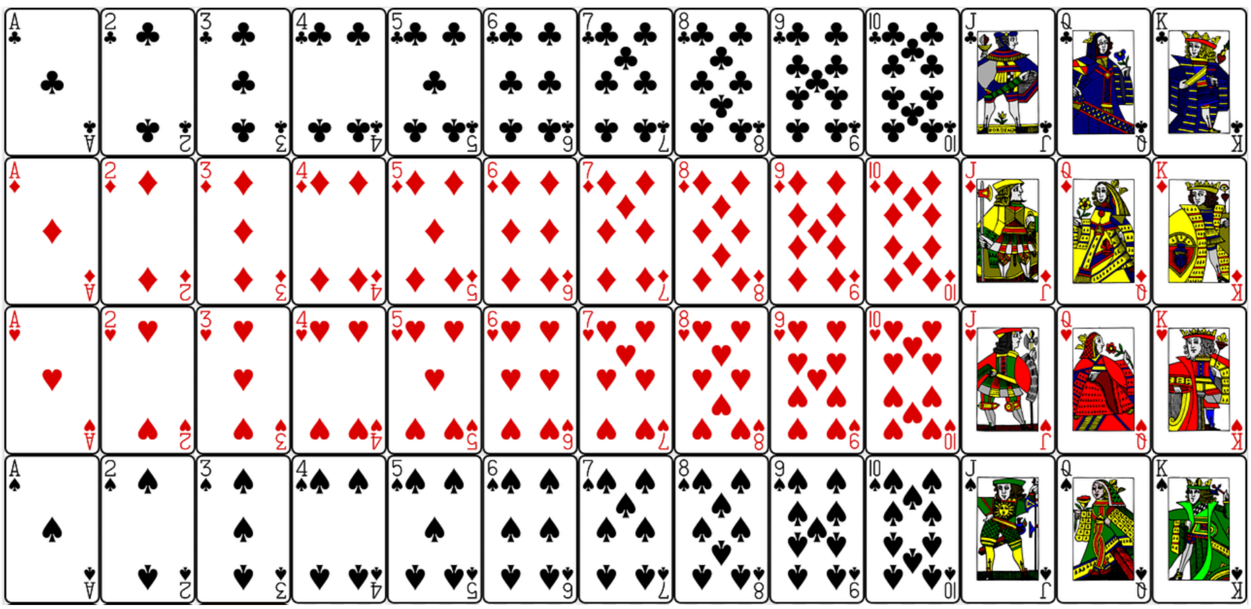
2.3: Independence

Slides (Google Drive) Alex Tsun Video (YouTube)

2.3.1 Chain Rule

We learned several tools already to compute probabilities (equally likely outcomes, Bayes Theorem, LTP). Now, we will learn how to handle the probability of several events occurring simultaneously: that is, $\mathbb{P}(A \cap B \cap C \cap D)$ for example. To compute the probability that at least one of several events happens: $\mathbb{P}(A \cup B \cup C \cup D)$, you would use inclusion-exclusion! We'll see an example which builds intuition first.

Consider a standard 52 card deck. This has four suits (clubs, spades, hearts, and diamonds). Each of the four suits has 13 cards of different rank (A, 2, 3, 4, 5, 6, 7, 8, 9, 10 J, Q, K).



Now, suppose that we shuffle this deck and draw the top three cards. Let's define:

1. A to be the event that we get the Ace of spades as our **first** card.
2. B to be the event that we get the 10 of clubs as our **second** card.
3. C to be the event that we get the 4 of diamonds as our **third** card.

What is the probability that all three of these events happen? We can write this as $\mathbb{P}(A, B, C)$ (sometimes we use commas as an alternative to using the intersection symbol, so this is equivalent to $\mathbb{P}(A \cap B \cap C)$). Note that this is equivalent to $\mathbb{P}(C, B, A)$ or $\mathbb{P}(B, C, A)$ since order of intersection does not matter.

Intuitively, you might say that this probability is $\frac{1}{52} \cdot \frac{1}{51} \cdot \frac{1}{50}$, and you would be correct.

1. The first factor comes from the fact that there are 52 cards that could be drawn, and only one ace of spades. That is, we computed $\mathbb{P}(A)$.

2. The second factor comes from the fact that there are 51 cards after we draw the first card and only one 10 of clubs. That is, we computed $\mathbb{P}(B \mid A)$.

3. The final factor comes from the fact that there are 50 cards left after we draw the first two and only one 4 of diamonds. That is, we computed $\mathbb{P}(C \mid A, B)$.

To summarize, we said that

$$\mathbb{P}(A, B, C) = \mathbb{P}(A) \cdot \mathbb{P}(B \mid A) \cdot \mathbb{P}(C \mid A, B) = \frac{1}{52} \cdot \frac{1}{51} \cdot \frac{1}{50}$$

This brings us to the chain rule:

Theorem 2.3.1: Chain Rule

Let $A_1, \dots, A_n$ be events with nonzero probabilities. Then:

$$\mathbb{P}(A_1, \dots, A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_2 \mid A_1)\,\mathbb{P}(A_3 \mid A_1 A_2) \cdots \mathbb{P}(A_n \mid A_1, \dots, A_{n-1})$$

In the case of two events, A, B (this is just the alternate form of the definition of conditional probability from 2.2):

$$\mathbb{P}(A, B) = \mathbb{P}(A)\,\mathbb{P}(B \mid A)$$

An easy way to remember this, is if we want to observe n events, we can observe one event at a time, and condition on those that we've done thus far. And most importantly, since the order of intersection **doesn't matter**, you can actually decompose this into any of $n!$ orderings. Make sure you "do" one event at a time, conditioning on the intersection of ALL past events like we did above.

Proof of Chain Rule. Remember that the definition of conditional probability says $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B \mid A)$. We'll use this repeatedly to break down our $\mathbb{P}(A_1, \dots, A_n)$. Sometimes it is easier to use commas, and sometimes it is easier to use the intersection sign $\cap$: for this proof, we'll use the intersection sign. We'll prove this for four events, and you'll see how it can be easily extended to any number of events!

$$\begin{aligned}
\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) &= \mathbb{P}((A_1 \cap A_2 \cap A_3) \cap A_4) && [\text{treat } A_1 \cap A_2 \cap A_3 \text{ as one event}] \\
&= \mathbb{P}(A_1 \cap A_2 \cap A_3)\,\mathbb{P}(A_4 \mid A_1 \cap A_2 \cap A_3) && [\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B \mid A)] \\
&= \mathbb{P}((A_1 \cap A_2) \cap A_3)\,\mathbb{P}(A_4 \mid A_1 \cap A_2 \cap A_3) && [\text{treat } A_1 \cap A_2 \text{ as one event}] \\
&= \mathbb{P}(A_1 \cap A_2)\,\mathbb{P}(A_3 \mid A_1 \cap A_2)\,\mathbb{P}(A_4 \mid A_1 \cap A_3 \cap A_3) && [\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B \mid A)] \\
&= \mathbb{P}(A_1)\,\mathbb{P}(A_2 \mid A_1)\,\mathbb{P}(A_3 \mid A_1 \cap A_2)\,\mathbb{P}(A_4 \mid A_1 \cap A_3 \cap A_3) && [\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B \mid A)]
\end{aligned}$$

Note how we keep "chaining" and applying the definition of conditional probability repeatedly!

□

Example(s)

Consider the 3-stage process. We roll a 6-sided die (numbered 1-6), call the outcome X. Then, we roll a X-sided die (numbered 1-X), call the outcome Y. Finally, we roll a Y-sided die (numbered 1-Y), call the outcome Z. What is $P(Z = 5)$?

Solution There are only three things that could have happened for the triplet (X, Y, Z) so that Z takes on the value 5: $\{(6, 6, 5), (6, 5, 5), (5, 5, 5)\}$. So

$$\begin{aligned}
\mathbb{P}(Z = 5) &= \mathbb{P}(X = 6, Y = 6, Z = 5) + \mathbb{P}(X = 6, Y = 5, Z = 5) + \mathbb{P}(X = 5, Y = 5, Z = 5) && \text{[cases]} \\
&= \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{5} + \frac{1}{6} \cdot \frac{1}{5} \cdot \frac{1}{5} && \text{[chain rule 3x]}
\end{aligned}$$

How did we use the chain rule? Let's see for example the last term:

$$\mathbb{P}(X = 5, Y = 5, Z = 5) = P(X = 5)P(Y = 5 \mid X = 5)P(Z = 5 \mid X = 5, Y = 5)$$

$P(X = 5) = \dfrac{1}{6}$ because we rolled a 6-sided die.

$P(Y = 5 \mid X = 5) = \dfrac{1}{5}$ since we rolled a $X = 5$-sided die.

Finally, $P(Z = 5 \mid X = 5, Y = 5) = P(Z = 5 \mid Y = 5) = \dfrac{1}{5}$ since we rolled a $Y = 5$-sided die. Note we didn't need to know $X = 5$ once we knew $Y = 5$!

$\square$

2.3.2 Independence

Let's say we flip a fair coin 3 times *independently* (whatever that means) - what is the probability of getting all heads? You may be inclined to say $(1/2)^3 = 1/8$ because the probability of getting heads each time is just $1/2$. However, we haven't learned such a rule to compute the joint probability $\mathbb{P}(H_1 \cap H_2 \cap H_3)$ except the chain rule.

Using only what we've learned, we could consider equally likely outcomes. There are $2^3 = 8$ possible outcomes when flipping a coin three times (by product rule), and only one of those (HHH) makes up the event we care about: $H_1 \cap H_2 \cap H_3$. Since the outcomes are equally likely,

$$\mathbb{P}(H_1 \cap H_2 \cap H_3) = \frac{|\, H_1 \cap H_2 \cap H_3 \,|}{|\, \Omega \,|} = \frac{|\, \{HHH\} \,|}{2^3} = \frac{1}{8}$$

We'd love a rule to say $\mathbb{P}(H_1 \cap H_2 \cap H_3) = \mathbb{P}(H_1) \cdot \mathbb{P}(H_2) \cdot \mathbb{P}(H_3) = 1/2 \cdot 1/2 \cdot 1/2 = 1/8$ - and it turns out this is true when the events are independent!

But first, let's consider the smaller case: does $\mathbb{P}(A, B) = \mathbb{P}(A)\mathbb{P}(B)$ in general? No! How do we know this though? Well recall that by the chain rule, we know that:

$$\mathbb{P}(A, B) = \mathbb{P}(A)\mathbb{P}(B \mid A)$$

So, unless $\mathbb{P}(B \mid A) = \mathbb{P}(B)$ the equality does not hold. However, when this equality does hold, it is a special case, which brings us to independence.

Definition 2.3.1: Independence

Events A and B are **independent** if any of the following equivalent statements hold:

1. $\mathbb{P}(A \mid B) = \mathbb{P}(A)$
2. $\mathbb{P}(B \mid A) = \mathbb{P}(B)$
3. $\mathbb{P}(A, B) = \mathbb{P}(A)\mathbb{P}(B)$

Intuitively what it means for $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ is that: given that we know B happened, the probability of observing A is the same as if we didn't know anything. So, event B has no influence on event A. The last statement

$$\mathbb{P}(A, B) = \mathbb{P}(A)\mathbb{P}(B)$$

is the most often applied to problems where we are allowed to assume independence.

What about independence of more than just two events? We call this concept "mutual independence" (but most of the time we don't even say the word "mutual"). You might think that for events A_1, A_2, A_3, A_4 to be (mutually) independent, by extension of the definition of two events, we would just need

$$\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3) \cdot \mathbb{P}(A_4)$$

But it turns out, we need this property to hold for *any* subset of the 4 events. For example, the following must be true (in addition to others):

$$\mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_3)$$

$$\mathbb{P}(A_2 \cap A_3 \cap A_4) = \mathbb{P}(A_2) \cdot \mathbb{P}(A_3) \cdot \mathbb{P}(A_4)$$

For all 2^n subsets of the 4 events ($2^4 = 16$ in our case), the probability of the intersection must simply be the product of the individual probabilities.

As you can see, it would be quite annoying to check even if three events were (mutually) independent. Luckily, most of the time we are told to assume that several events are (mutually) independent and we get all of those statements to be true for free. We are rarely asked to demonstrate/prove mutual independence.

Definition 2.3.2: Mutual Independence

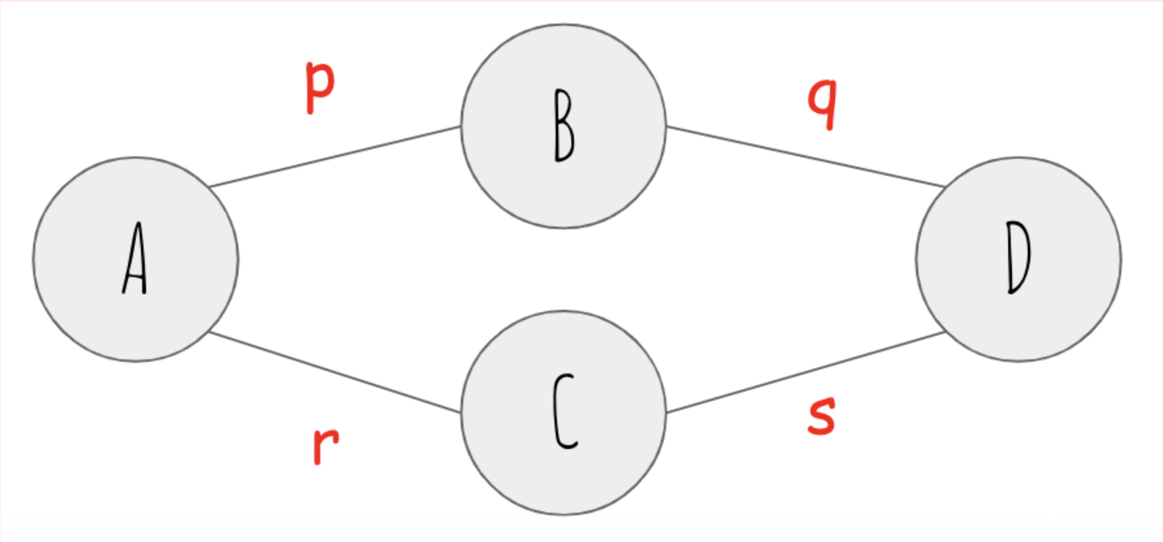
We say n events $A_1, A_2, \ldots, A_n$ are **(mutually) independent** if, for *any* subset $I \subseteq [n] = \{1, 2, \ldots, n\}$, we have

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i)$$

This is very similar to the last formula $\mathbb{P}(A, B) = \mathbb{P}(A)\mathbb{P}(B)$ in the definition of independence for two events, just extended to multiple events. It must hold for any subset of the n events, and so this equation is actually saying 2^n equations are true!

Example(s)

Suppose we have the following network, in which circles represents a node in the network (A, B, C, and D) and the links have the probabilities p, q, r and s of successfully working. That is, for example, the probability of successful communication from A to B is p. Each link is independent of the others though.



Now, let's consider the question, what is the probability that A and D can successfully communicate?

Solution There are two ways in which it can communicate: (1) in the top path via B or (2) in the bottom path via C. Let's define the event *top* to be successful communication in the top path and the event *bottom* to be successful communication in the bottom path. Let's first consider the probabilities of each of these being successful communication. For the top to be a valid path, *both* links AB and BD must work.

$$\begin{aligned}
\mathbb{P}(\text{top}) &= \mathbb{P}(AB \cap BD) \\
&= \mathbb{P}(AB)\,\mathbb{P}(BD) && \text{[by independence]} \\
&= pq
\end{aligned}$$

Similarly:

$$\begin{aligned}
\mathbb{P}(\text{bottom}) &= \mathbb{P}(AC \cap CD) \\
&= \mathbb{P}(AC)\,\mathbb{P}(CD) && \text{[by independence]} \\
&= rs
\end{aligned}$$

So, to calculate the probability of successful communication between A and D, we can take the union of *top* and *bottom* (we just need at least one of the two to work), and so we have:

$$\begin{aligned}
\mathbb{P}(\text{top} \cup \text{bottom}) &= \mathbb{P}(\text{top}) + \mathbb{P}(\text{bottom}) - \mathbb{P}(\text{top} \cap \text{bottom}) && \text{[by inclusion-exclusion]} \\
&= \mathbb{P}(\text{top}) + \mathbb{P}(\text{bottom}) - \mathbb{P}(\text{top})\,\mathbb{P}(\text{bottom}) && \text{[by independence]} \\
&= pq + rs - pqrs
\end{aligned}$$

□

2.3.3 Conditional Independence

In the example above for the chain rule, we made this step:

$$\mathbb{P}(Z=5 \mid X=5, Y=5) = \mathbb{P}(Z=5 \mid Y=5)$$

This is actually another form of independence, called conditional independence! That is, *given* that $Y = 5$, the events $X = 5$ and $Z = 5$ are independent (the above equation looks exactly like $\mathbb{P}(Z=5 \mid X=5) = \mathbb{P}(Z=5)$ except with extra conditioning on $Y = 5$ on both sides.

Definition 2.3.3: Conditional Independence

Events A and B are **conditionally independent given an event** C if any of the following equivalent statements hold:

1. $\mathbb{P}(A \mid B, C) = \mathbb{P}(A \mid C)$
2. $\mathbb{P}(B \mid A, C) = \mathbb{P}(B \mid C)$
3. $\mathbb{P}(A, B \mid C) = \mathbb{P}(A \mid C)\,\mathbb{P}(B \mid C)$

Recall the definition of A and B being (unconditionally) independent below:

1. $\mathbb{P}(A \mid B) = \mathbb{P}(A)$
2. $\mathbb{P}(B \mid A) = \mathbb{P}(B)$
3. $\mathbb{P}(A, B) = \mathbb{P}(A)\,\mathbb{P}(B)$

Notice that this is very similar to the definition of independence. There is no difference, except we have just added in conditioning on C to every probability.

Example(s)

Suppose there is a coin C_1 with $\mathbb{P}(head) = 0.3$ and a coin C_2 with $\mathbb{P}(head) = 0.9$. We pick one randomly with equal probability and will flip that coin 3 times *independently*. What is the probability we get all heads?

Solution Let us call HHH the event of getting three heads, C_1 the event of picking the first coin, and C_2 the event of getting the second coin. Then we have the following:

$$\begin{aligned}
\mathbb{P}(HHH) &= \mathbb{P}(HHH \mid C_1)\,\mathbb{P}(C_1) + \mathbb{P}(HHH \mid C_2)\,\mathbb{P}(C_2) && \text{[by the law of total probability]} \\
&= (\mathbb{P}(H \mid C_1))^3\mathbb{P}(C_1) + (\mathbb{P}(H \mid C_2))^3\mathbb{P}(C_2) && \text{[by conditional independence]} \\
&= (0.3)^3\frac{1}{2} + (0.9)^3\frac{1}{2} = 0.378
\end{aligned}$$

It is important to note that getting heads on the first and second flip are NOT independent. The probability of heads on the second, given that we got heads on the first flip, is much higher since we are more likely to have chosen coin C_2. However, *given which coin we are flipping*, the flips are conditionally independent. Hence, we can write $\mathbb{P}(HHH \mid C_1) = \mathbb{P}(H \mid C_1)^3$. $\square$

2.3.4 Exercises

1. Corrupted by their power, the judges running the popular game show America's Next Top Mathematician have been taking bribes from many of the contestants. During each of two episodes, a given contestant is either allowed to stay on the show or is kicked off. If the contestant has been bribing the judges, she will be allowed to stay with probability 1. If the contestant has not been bribing the judges, she will be allowed to stay with probability 1/3, independent of what happens in earlier episodes. Suppose that 1/4 of the contestants have been bribing the judges. The same contestants bribe the judges in both rounds.

   (a) If you pick a random contestant, what is the probability that she is allowed to stay during the first episode?

   (b) If you pick a random contestant, what is the probability that she is allowed to stay during both episodes?

   (c) If you pick a random contestant who was allowed to stay during the first episode, what is the probability that she gets kicked off during the second episode?

   (d) If you pick a random contestant who was allowed to stay during the first episode, what is the probability that she was bribing the judge?

   Solution:

   (a) Let S_i be the event a contestant stays in the i^{th} episode, and B be the event a contestant is bribing the judges. Then, by the law of total probability,

   $$\mathbb{P}(S_1) = \mathbb{P}(S_1 \mid B)\,\mathbb{P}(B) + \mathbb{P}\left(S_1 \mid B^C\right)\mathbb{P}\left(B^C\right) = 1 \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{2}$$

   (b) Again by the law of total probability,

   $$\begin{aligned}\mathbb{P}(S_1 \cap S_2) &= \mathbb{P}(S_1 \cap S_2 \mid B)\,\mathbb{P}(B) + \mathbb{P}\left(S_1 \cap S_2 \mid B^C\right)\mathbb{P}\left(B^C\right) && \text{[LTP]}\\ &= \mathbb{P}(S_1 \mid B)\,\mathbb{P}(S_S \mid B)\,\mathbb{P}(B) + \mathbb{P}\left(S_1 \mid B^C\right)\mathbb{P}\left(S_2 \mid B^C\right)\mathbb{P}\left(B^C\right) && \text{[conditional independence]}\\ &= 1 \cdot 1 \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{3}{4}\\ &= \frac{1}{3}\end{aligned}$$

   Again, it's important to note that staying on the first and second episode are NOT independent. If we know she stayed on the first episode, then it is more likely she stays on the second (since she's more likely to be bribing the judges). However, conditioned on whether or not we are bribing the judges, S_1 and S_2 are independent.

   (c)

   $$\mathbb{P}\left(S_2^C \mid S_1\right) = \frac{\mathbb{P}\left(S_1 \cap S_2^C\right)}{\mathbb{P}(S_1)}$$

   The denominator is our answer to (a), and the numerator can be computed in the same way as (b).

   (d) By Bayes Theorem,

   $$\mathbb{P}(B \mid S_1) = \frac{\mathbb{P}(S_1 \mid B)\,\mathbb{P}(B)}{\mathbb{P}(S_1)}$$

   We computed all these quantities in part (a).

2. A parallel system functions whenever at least one of its components works. Consider a parallel system of n components and suppose that each component works with probability p independently

   (a) What is the probability the system is functioning?

   (b) If the system is functioning, what is the probability that component 1 is working?

   (c) If the system is functioning and component 2 is working, what is the probability that component 1 is working?

   Solution:

   (a) Let C_i be the event component i is functioning, for $i = 1, \ldots, n$. Let F be the event the system functions. Then,

$$\begin{aligned}
\mathbb{P}(F) &= 1 - \mathbb{P}\left(F^C\right) \\
&= 1 - \mathbb{P}\left(\bigcap_{i=1}^{n} C_i^C\right) && \text{[def of parallel system]} \\
&= 1 - \prod_{i=1}^{n} \mathbb{P}\left(C_i^C\right) && \text{[independence]} \\
&= 1 - (1-p)^n && \text{[prob any fails is } 1-p\text{]}
\end{aligned}$$

   (b) By Bayes Theorem, and since $\mathbb{P}(F \mid C_1) = 1$ (system is guaranteed to function if C_1 is working),

$$P(C_1 \mid F) = \frac{\mathbb{P}(F \mid C_1)\,\mathbb{P}(C_1)}{\mathbb{P}(F)} = \frac{1 \cdot p}{1-(1-p)^n}$$

   (c)

$$\begin{aligned}
\mathbb{P}(C_1 \mid C_2, F) &= \mathbb{P}(C_1 \mid C_2) && \text{[if given } C_2\text{, already know } F \text{ is true]} \\
&= \mathbb{P}(C_1) && [C_1, C_2 \text{ independent}] \\
&= p
\end{aligned}$$