Probability I

Daniel Yu

September 24, 2024

Contents

1 Sums of Random Variables

Definition 1. Let X, Y be R.V. that map $\Omega \to \mathbb{R}$. What is the distribution of $X+Y$? If X, Y are discrete, we need to describe:

$$P[X+Y=k] = \sum_{a \in range(x), b \in range(b), a+b=k} P[x=a, Y=b].$$

Note. Example
Ω = two rolls of 4 sided dice

P = uniform

X = outcome of d_1

M = maximum of d_1, d_2

$$\begin{aligned} P[X+M=4] &= P[X=1, M=3] + P[X=2, M=2] + P[X=3, M=1] \\ &= P[\{(1,3)\}] + P[\{(2,1),(2,2)\}] + 0 \\ &= \frac{1}{16} + \frac{1}{8} \\ &= \frac{3}{16}. \end{aligned}$$

We need the joint distributions for probability of sum of two R.V! However, for expected value

$$E[X+M] = E[X] + E[M].$$

, we can compute sum based on marginals alone.

Definition 2. If X, Y independent,

$$\begin{aligned} P[X+Y=k] &= \sum_{a,b;a+b=k} P[X=a, Y=b] \\ &= \sum_{a,b;a+b=k} P[X=a] \cdot P[Y=b] \\ &= \sum_{a,b;a+b=k} P[X=a] \cdot P[Y=k-a]. \end{aligned}$$

Note. Sum of Uniform Distributions
Is not what you think it is ...

Let X, Y be uniform on $\{1,2,3,4\}$ and independent. What is the distribution of $X+Y$?

$$range(X+Y) = \{2,3,4,\ldots,8\}.$$

But $X+Y$ is not uniform.

$$\begin{aligned} &P[X+Y=2] = P[\{1,1\}] = \frac{1}{16} = P[X+Y=8] \\ &P[X+Y=3] = P[\{(1,2)\,(2,1) = \frac{1}{8} = P[X+Y=7]\} \\ &\ldots \\ &P[X+Y=5] = \frac{1}{4} \end{aligned}$$

The distribution is not uniform.

Definition 3. In the continuous case, if X, Y are independent with pdf $f_x(s), f_y(t)$. Then,

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_x(s) \cdot f_y(t-s) ds.$$

Note. Example
Let $X, Y \sim \exp(1)$ i.e. $f_X(t) = e^{-t} \forall t \geq 0$.

$$\begin{aligned} f_{X+Y}(t) &= \int_{-\infty}^{\infty} f_X(s) f_Y(t-s) ds \\ &= \int_0^t e^{-t} e^{-(t-s)} ds \\ &= e^{-t} \int_0^t 1 ds \\ &= te^{-t}, t \geq 0. \end{aligned}$$

Note. Conditional Probability with Sum Ex
Let X, Y be independent $Geo(p)$ R.V. Given $X + Y = n$ what is distribution of X?

First, what is the posssible range of X given $\{X + Y = n\}$? $= \{1, 2, 3, \ldots, n-1\}$

Second, we know, $\forall k \in \{1, 2, 3, \ldots, n-1\}$

$$\begin{aligned} P[X = k | X + Y = n] &= \frac{P[X = k, X + Y = n]}{P[X + Y = n]} \\ &= \frac{P[X = k] \cdot P[Y = n - k]}{\sum_{a=1}^{n-1} P[X = a] P[Y = n - a]} \\ &= \frac{\left((1-p)^{k-1} \cdot p\right)\left((1-p)^{n-k-1} \cdot p\right)}{\sum_{a=1}^{n-1} \left((1-p)^{a-1} \cdot p \cdot (1-p)^{n-a-1} \cdot p\right)} \\ &= \frac{(1-p)^{n-2} \cdot p^2}{\sum_{a=1}^{n-1} (1-p)^{n-2} p^2} \\ &= \frac{1}{n-1}. \end{aligned}$$

Thus notice that the probability doesn't depend on k!, only n. This is due to symmetry!

2 Bayesian Thinking

Proposition 1. Law of Total Probability
Let $\{B_i\}_{i=1}^{\infty}$ be a partition of Ω. Then for any event A,

$$P[A] = \sum_{i=1}^{\infty} P[A|B_i] \cdot P[B_i].$$

Definition 4. Bayes Rule

$$P[B|A] = \frac{P[A|B] \cdot P[B]}{P[A|B] \cdot P[B] + P[A|B^c] \cdot P[B^c]}.$$

proof:

$$\begin{aligned} P[B|A] &= \frac{P[B \cap A]}{P[A]} \\ &\text{using law of total prop} \\ &= \frac{P[B \cap A]}{P[A|B] \cdot P[B] + P[A|B^c] \cdot P[B^c]} \\ &= \frac{P[A|B] \cdot P[B]}{P[A|B] \cdot P[B] + P[A|B^c] \cdot P[B^c]}. \end{aligned}$$

□

Note. Example
Let $X = 1$, flip a coin with a probability $\frac{3}{4}$ to get H's and $X = 0$, flip a coin with prob $\frac{1}{4}$ to get H's.

Let $A = \{$the two coin tosses are heads$\}$. What is $P[X = 1|A]$

Using Bayesian probability:

$$\begin{aligned} P[X = 1|A] &= \frac{P[A|X = 1] \cdot P[X = 1]}{P[A|X = 1] \cdot P[X = 1] + P[A|X = 0] \cdot P[X = 0]} \\ &= \frac{\frac{9}{16} \cdot \frac{1}{2}}{\frac{9}{16}\frac{1}{2} + \frac{1}{16}\frac{1}{2}} \\ &= \frac{9}{10} \end{aligned}$$

What are the priors and what are the posteriors???

2.1 False Negatives and Positives

We have a disease with prevalence of .1.

The test for the disease has false positive of .05.

The test has false negative of .02

You test positive, what is the probability you are sick?

Note. Answer
Define $A = \{sick\}$, $B = \{t_{\text{positive}}\}$. We know:

$$P[B|A] \cdot P[A] = \frac{49}{50} \cdot \frac{1}{10} \text{ (true positive rate is 1- false positve rate).}$$

$$P[B|A^c] \cdot P[A^c] = \frac{1}{20}\frac{9}{10} \text{ (false positive times not sick).}$$

$$P[A|B] = \frac{P[B|A] \cdot P[A]}{P[B|A] \cdot P[A] + P[B|A^c] \cdot P[A^c]} = \frac{\frac{49}{500}}{\frac{49}{500} + \frac{9}{200}} = .685.$$

This is because the prevalance of the disease is so low so overwelhming likely the person is healthy. This is why the **prior** matters!

2.2 Monty Hall Problem

You are picking one of 3 boxes, 1 of which has a prize, 2 of which have a goat. After your intial choice, you are offered a choice by the host to switch under the following assumptions? Should you switch? The host must always open a door that was not selected by the contestant. The host must always open a door to reveal a goat and never the car. The host must always offer the chance to switch between the door chosen originally and the closed door remaining.

Note. Answer
Yes! Even though one would intially assume that the probability of $\frac{1}{2}$ either way for staying or switching, the key is that the host's answer is NOT independent of if you're choice is correct or not! It gives you information, so in fact switching would have a probability of $\frac{2}{3}$ to get the right box but staying would only have probability $\frac{1}{2}$. Thus, the saying, always confirm your priors rears its head again.

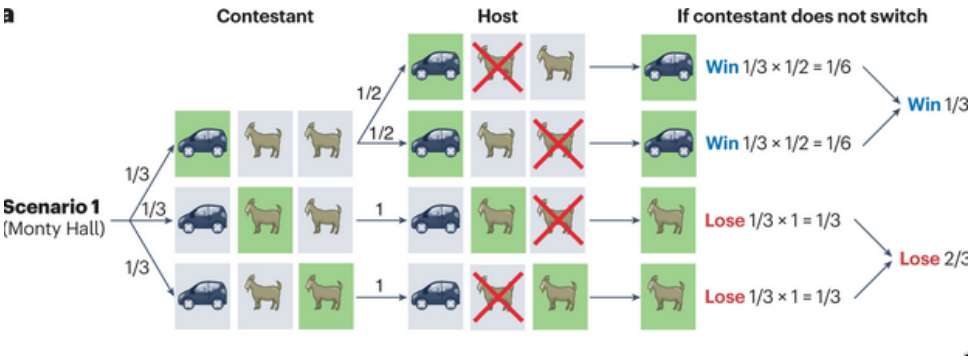

Figure 1: Monty Hall Decision Tree

When you select a goat, the host has to open the other goat box, restricting the choices (and thus influencing the probability of the events)

2.3 Coupon Collector Problem

There are n prizes distributed in cereal boxes randomly, independently, and uniformly. You want randomly open cereal boxes with replacement until you have seen every prize at least once. Let T_n = number of boxes you needed to open until you've seen all prizes. What is $E[T_n]$?

Proof. Proof. Set T_n = first time that I collect all n coupons. Instead of considering the number of the coupon box (i.e. original ordering, do we see box 3 first or box 5 with a coupon?), we consider whether we have recognized a previously collected coupon. Ex: Coupon number sequence $\rightarrow$ new/old sequence

$$(3, 3, 1, 2, 1, \ldots) \rightarrow (N, O, N, N, 0, \ldots).$$

. Given such a sequence, let R_k = increment of the time it takes to collect kth coupon = $T_k - T_{k-1}$ where T_k is the first time k number of coupons have been collected. For the example above:

$$(N \rightarrow T_1, O, N \rightarrow T_2, N \rightarrow T_3, 0, \ldots).$$

where $R_1 = T_1 - 0 = T_1$, $R_2 = T_2 - T_1$, ... What is the distribution of R_i? (The key is that the probability R_i is independent of each other i.e. the time it takes to draw a new coupon is independent for each new coupon)

$$R_1 = 1 \text{ with probability 1.}$$

$$R_2 \sim Geo(\frac{n-1}{n}) \text{ i.e } P[R_2 = k] = \frac{n-1}{n}\left(\frac{1}{n}\right)^{k-1}.$$

$$R_3 \sim Geo(\frac{n-1}{n}).$$

etc.

$$R_k \sim Geo(\frac{n-k+1}{n}).$$

and each R_i is independent of all others. Note that this does not answer the question of what is the probability of the nth toss is new or old (dependent on the previous history since the probabilities change depending on how many coupons have already been seen). What is the expected value of $E[T]$? To find this, consider the $E[Geo(p)]$ in general:

$$\begin{aligned} E[Geo(p)] &= \sum_{k=1}^{\infty} k \cdot P[Geo(p) = k] \\ &= \sum_{k=1}^{\infty} k \cdot p \cdot (1-p)^{k-1} \\ &= \frac{1}{p}. \end{aligned}$$

What about the variance of $Var(Geo(p))$? Insert algebra here for $\sum_{k=1}^{\infty} k^2 p \cdot (1-p)^{k-1}$

$$\begin{aligned} Var(Geo(p)) &= E[Geo(p)^2] - E[Geo(p)]^2 \\ &= \sum_{k=1}^{\infty} k^2 p \cdot (1-p)^{k-1} - \frac{1}{p}^2 \\ &= \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}. \end{aligned}$$

Answer
We know that $T_n = R_1 + R_2 + \ldots R_n$, so

$$\begin{aligned} E[T] &= E[R_1] + E[R_2] + \ldots + E[R_n] \\ &= E[Geo(1)] + E[Geo(\frac{n-1}{n})] + \ldots + E[Geo(\frac{1}{n})] \\ &= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \ldots + \frac{n}{2} + n \\ &= n \cdot \left(\sum_{k=1}^{n} \frac{1}{k} \right) \text{ a divergent series.} \end{aligned}$$

Further analysis reveals $\sum_{k=1}^{n} \frac{1}{k} \geq \int_1^{n+1} \frac{1}{x} dx = \ln(n+1)$. We know $\sum_{k=1}^{n} \frac{1}{k} - \ln(n-1) = \gamma \approx 0.5772$ (Euler-Masrteroni constant). For super large n, $E[T] \sim n\ln(n) + \gamma n$, i.e. the expected value of the time to collect n coupons grows supra-linearly!

What about $Var(T)$? Since R_i independent:

$$\begin{aligned} Var(T) &= \sum_{k=1}^{n} Var(R_k) \\ &= \frac{1 - \left(\frac{n-1}{n}\right)}{\frac{n-1}{n}^2} + \frac{1 - \left(\frac{n-2}{n}\right)}{\frac{n-2}{n}^2} + \ldots + \frac{1 - \frac{1}{n}}{\frac{1}{n}^2} \\ &= \sum_{k=1}^{n} \frac{1 - \left(\frac{k-n}{n}\right)}{\left(\frac{k}{n}^2\right)} \\ &= n^2 \left(\sum_{x=1}^{n} \frac{1}{k^2} - \frac{1}{n} \sum_{n-1}^{n} \frac{1}{k} \right). \end{aligned}$$

Notice that now $\sum_{x=1}^{n} \frac{1}{k^2}$ is a convergent series! So as we take $n \to \infty$, $Var(T) \approx \frac{n^2 - \pi^2}{6} - n\ln(n) - \gamma n$. This means that the central tendency (i.e. expected value) or mean of the time to collect n coupons is some value and the standard deviation is much smaller, so the distribution of the time it takes is concentrated within a narrow range! □

This is all about independence

Proposition 2. To find this, consider the $E[Geo(p)]$ in general:

$$\begin{aligned} E[Geo(p)] &= \sum_{k=1}^{\infty} k \cdot P[Geo(p) = k] \\ &= \sum_{k=1}^{\infty} k \cdot p \cdot (1-p)^{k-1} \\ &= \frac{1}{p}. \end{aligned}$$

What about the variance of $Var(Geo(p))$? Insert algebra here for $\sum_{k=1}^{\infty} k^2 p \cdot (1-p)^{k-1}$

$$\begin{aligned} Var(Geo(p)) &= E[Geo(p)^2] - E[Geo(p)]^2 \\ &= \sum_{k=1}^{\infty} k^2 p \cdot (1-p)^{k-1} - \frac{1}{p}^2 \\ &= \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}. \end{aligned}$$

2.4 Secretary Problem

We have n candidates, who are ordered uniformly at random (i.e. all $n!$ options are equally likely). Every time you interview a candidate, you must choose to hire them, or let them go forever. What strategy maximizes the probability of hiring the best candidate?

For the sake of this problem we will restrict ourselves to a set of strategies where we reject the first k candidates and then hire the next candidate that's better than the k rejected candidates. Let $S_k =$ { this strategy hsa the best candidate}. This is not all possible strategies. It is, however, an optimal one (which will not be proven).

Proof. Let's find out which S_k maximizes probability of hiring best candidate. To find this, we use the law of total probability: $\{B_i\}_{i=1}^n$ is a partition,

$$P[S_x] = \sum_{i=1}^{n} P[S_x|B_i]P[B_i].$$

We set $B_i =$ { Best candidate shows up at time i}. Then $P[B_i] = \frac{1}{n}$ (best person equally likely to show up at any time).

$$P[S_k|B_i] = \begin{cases} 0 \text{ if i} \leq k \text{ i.e. the best candidate was eliminated for coming within first k candidates} \\ ??? \ \ i > k \end{cases}.$$

Realize,

$$P[S_k|B_i] = \begin{cases} 0, \text{ if i} \leq k \text{ i.e. the best candidate was eliminated for coming within first k candidates} \\ P[\text{ Best candidate among first } \{B_1, \ldots, B_{i-1}\} \text{ shows up in 1st k candidates}], i > k \end{cases}.$$

Conditional on B_i, all possible orderings of the remaining candidates are equally likely. In particular all $(i-1)!$ orderings of the best candidate among $\{1, \ldots, i-1\}$ are equally likely.

$$P[\text{ Best candidate among first } \{B_1, \ldots, B_{i-1}\} \text{ shows up in 1st k candidates}] = \frac{k}{i-1}.$$

Then,

$$\begin{aligned} P[S_x] &= \sum_{i=1}^{n} P[S_x|B_i] \cdot P[B_i] \\ &= \frac{1}{n} \sum_{i=k+1}^{n} \frac{k}{i-1} \\ &= \frac{k}{n} \left(\sum_{i=2}^{n} \frac{1}{i-1} - \sum_{i=2}^{k} \frac{1}{i-1} \right). \end{aligned}$$

For n, k large,

$$P[S_x] \approx \frac{k}{n} \cdot \log\left(\frac{n}{k}\right).$$

If I set $k = c \cdot n$, then $P[S_x] = c \cdot \ln(\frac{1}{c})$ when $c \in [0, 1]$ what value of c should be chosen to maximize the probability?

$$\begin{aligned} f(c) &= c \ln(\frac{1}{c}) \\ f'(c) &= \ln(\frac{1}{c}) - 1 = 0 \\ \ln(\frac{1}{2}) &= 1 \\ c &= \frac{1}{e} \\ k &= \frac{n}{e}. \end{aligned}$$

and $P[S_x] = \frac{1}{e} \cdot \ln(\frac{1}{\frac{1}{e}}) = \frac{1}{e}$. □