Some Class

Daniel Yu

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 Thinkin

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:

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

$\square$

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 $A = \{\text{the two coin tosses are heads}\}$. 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.

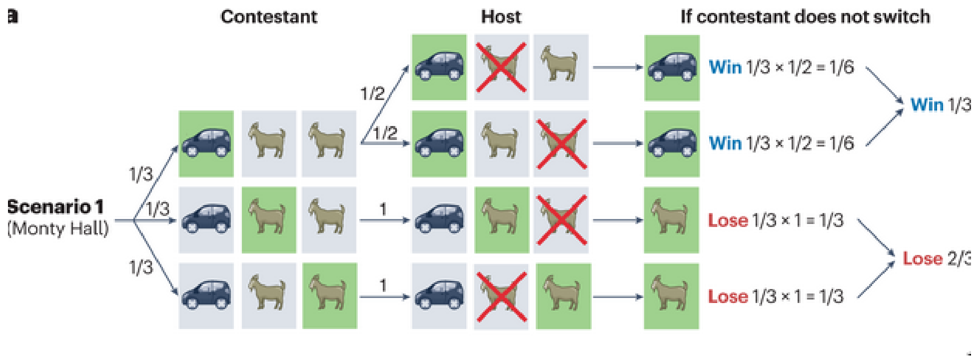

Figure 1: Monty Hall Decision Tree

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. In other words, $T_n = \{$ first time new coupons = n$\}$. We know that the time to collect first new coupon = 1.

TODO when I have time.

□