CS109 February 15, 2024

Conditional Expectation, Introductory Inference

Before you leave lab, make sure you click here so that you're marked as having attended this week's section. The CA leading your discussion section can enter the password needed once you've submitted.

1 Warmups

1.1 Why Multiple Random Variables?

What is a probabilistic model with multiple random variables? What does the term inference mean? What do you call the probability of an assignment to all variables in a probabilistic model? Why is that useful? Why can it be hard to represent?

A probabilistic model is a way of defining the relationship between many random variables. Inference is the act of computing a belief in one (or more) variables based on an observation. The probability of an assignment to all variables in a probabilistic model is called the joint. The joint can be used to solve any inference task. The number of ways of assigning values to variables is exponential in the number of random variables.

1.2 Joint Random Variables Statistics

True or False? The symbol Cov is covariance, the symbol $\wedge$ is logical-and, the symbol ρ is Pearson correlation, the symbol $\implies$ is logical implication, and $X \perp Y$ is just a fancy way to say that X and Y are independent.

A statement like "$A \sim Bin(10, 0.5) \wedge B \sim Bin(10, 0.5) \wedge A \perp B \implies A + B \sim Bin(20, 0.5)$" reads "If A and B are both distributed as Binomials with the same parameters, then $A + B$ is a Binomial as well with the same p parameter and an n parameter that's the sum of the those for A and B."

$X \perp Y \implies Cov(X,Y) = 0$	$Var(X + X) = 2Var(X)$
$Cov(X,Y) = 0 \implies X \perp Y$	$X \sim \mathcal{N}(0,1) \wedge Y \sim \mathcal{N}(0,1) \implies \rho(X,Y) = 1$
$Y = X^2 \implies \rho(X,Y) = 1$	$Y = 3X \implies \rho(X,Y) = 3$

True or False?

True	False ($\ldots = 4Var(X)$)
False (antecedent necessary, not sufficient)	False (don't know how independent X & Y are)
False ($Y = X \implies \ldots$)	False ($\ldots = 1$)

1.3 Random Number of Random Variables

Let N be a non-negative integer-valued random variable—that is, a random variable that takes on values in $\{0, 1, 2, \ldots\}$. Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of independent and identically distributed random variables (independent of N), each with mean μ, and $X = \sum_{i=1}^{N} X_i$ be the sum of the first N of them.

Before doing any work, what do you *think* $E[X]$ will turn out to be? Then show it mathematically to see if your intuition is correct.

$$E[X] = E\left[\sum_{i=1}^{N} X_i\right] = \sum_n E\left[\sum_{i=1}^{N} X_i \mid N = n\right] p_N(n) = \sum_n E\left[\sum_{i=1}^{n} X_i \mid N = n\right] p_N(n)$$

$$= \sum_n E\left[\sum_{i=1}^{n} X_i\right] p_N(n) = \sum_n n\mu p_N(n) = \mu \sum_n n p_N(n) = \mu E[N]$$

Alternatively,

$$E[X] = E[E[X|N]] = E[N\mu] = \mu E[N]$$

2 Problems

2.1 CS106A Is Popular

CS106A is Stanford's introductory programming course and largely considered the primary gateway to our undergraduate major. Assume next quarter's offering of CS106A is exactly 600 people, that each of the four undergraduate classes is compromised of 1750 students, and that next quarter's CS106A roster is just some random sample of the 7000 undergraduates. Let A, B, C, and D count the number of freshman, sophomores, juniors, and seniors in the class of 600.

- Present the joint probability mass function of A, B, C, D? Restated, present an expression for $P(A = a, B = b, C = c, D = d)$.

$$P(A = a, B = b, C = c, D = d) = \frac{\binom{1750}{a}\binom{1750}{b}\binom{1750}{c}\binom{1750}{d}}{\binom{7000}{600}}$$

- Does your PMF from part a) describe a multinomial random variable? Intuitively justify your answer.

No, a multinomial random variable requires that trials be independent. If we define a trial in this case as picking one student to be in CS106A, independence would require us to select students with replacement, which isn't the case here.

- What is the conditional probability distribution of A given that $B + C = 300$? Restated, what is $P(A = a|B + C = 300)$?

This is the same type of problem as that in part a), but we are now distributing 300 students across two groups, freshmen and seniors.

$$P(A = a|B + C = 300) = \frac{\binom{1750}{a}\binom{1750}{300-a}}{\binom{3500}{300}}$$

- Do you expect $Cov(A, B)$ to be positive, zero, or negative? Justify your answer.

We expect $Cov(A, B)$ to be negative. If there are more freshmen in the class, there will be fewer spots left for sophomores. Incidentally, the $\rho(A, B)$ value—that is, the normalized covariance—would be very, very slightly negative because of the large number of students and the large class size.

2.2 Managing Screen Time

Push notifications light up our phones at a rate that's guided by a Poisson process with an constant average rate of 5 notifications per hour at all hours, night and day. In an effort to maximize productivity, you put your phone down and ignore it as much as possible. You do, however, periodically check it to clear all notifications. You check at 7am when you wake up, noon when you grab lunch, 5pm as you wrap up classes for the day, and then again at 10pm before you go to sleep.

Let W, X, Y, and Z be Poisson random variables that count the number of push notifications that have accumulated from 10pm to 7am, 7am to noon, noon to 5pm, and 5pm to 10pm, respectively. We assume that the number of push notifications that arrive within each interval are all mutually independent of all other intervals.

- Compute the joint PMF on W, X, Y, and Z.

Since all four variables are independent, the joint PMF is the product of the four single-dimensional PMFs. That means that:

$$P(W = w, X = x, Y = y, Z = z) = \frac{e^{-45}45^w}{w!}\frac{e^{-25}25^x}{x!}\frac{e^{-25}25^y}{y!}\frac{e^{-25}25^z}{z!}$$

Ship it!

- Compute the conditional joint PMF on W, X, Y, and Z given that W + X + Y + Z = 150.

When adding Poisson random variables, it's nice to have another random variable modelling that sum. Let's define $S = W + X + Y + X$ with the understanding that the sum of independent Poisson random variables is itself a Poisson where the one parameter is the sum of the individual ones. That means $S \sim Poi(120)$ and that $P(S = s) = \frac{e^{-120}120^s}{s!} = \frac{e^{-120}120^{w+z+y+z}}{s!}$.

What's the conditional joint PMF of interest? Check this out:

$$P(W = w, X = x, Y = y, Z = z | S = s)$$

$$\begin{aligned}
&= \frac{P(S = s, W = w, X = x, Y = y, Z = z)}{P(S = s)} \\
&= \frac{P(W = w, X = x, Y = y, Z = z)}{P(S = s)} \\
&= \frac{\frac{e^{-45}45^w}{w!}\frac{e^{-25}25^x}{x!}\frac{e^{-25}25^y}{y!}\frac{e^{-25}25^z}{z!}}{\frac{e^{-120}120^s}{s!}} \\
&= \frac{\frac{45^w}{w!}\frac{25^x}{x!}\frac{25^y}{y!}\frac{25^z}{z!}}{\frac{120^{w+z+y+z}}{s!}} \\
&= \binom{s}{w, x, y, z}\left(\frac{9}{24}\right)^w\left(\frac{5}{24}\right)^x\left(\frac{5}{24}\right)^y\left(\frac{5}{24}\right)^z
\end{aligned}$$

The transition from the first line to the second is just stating that $S = s$'s presence in the joint PMF is redundant, since $s = w + x + y + z$. The bottom line here is that the conditional PMF devolves into a multinomial random variable. When s is 150, our probability is just

$$P(W = w, X = x, Y = y, Z = z | S = 150) = \binom{150}{w, x, y, z}\left(\frac{9}{24}\right)^w\left(\frac{5}{24}\right)^x\left(\frac{5}{24}\right)^y\left(\frac{5}{24}\right)^z$$

- Compute the conditional PMF of X + Y + Z—that's the number of notifications that arrive while you're awake—given that W + X + Y + Z = 150.

This can be derived from first principles much as we derived the conditional joint PMF in the previous part to this question. Or we can simply merge the three intervals of the multinomial to arrive at a binomial, i.e., $X + Y + Z | S = 150 \sim Bin(150, \frac{15}{24}) = Bin(150, \frac{5}{8})$.

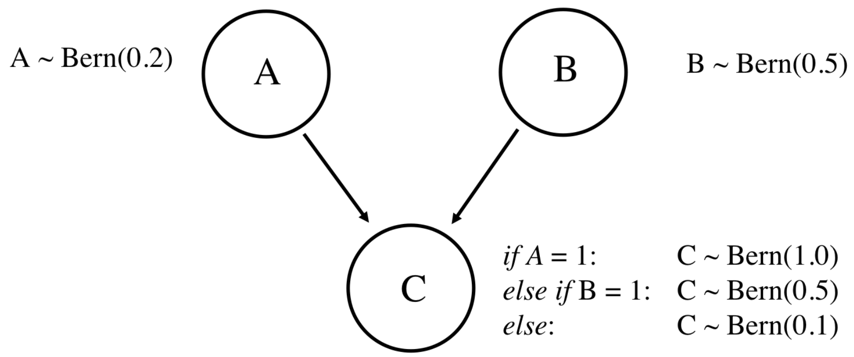
- Compute $E[X + Y + Z | W + X + Y + Z = 150]$ and $Var(X + Y + Z | W + X + Y + Z = 150)$.

$E[X + Y + Z | W + X + Y + Z = 150] = 150 \cdot \frac{5}{8} = 93.75$ and $Var(X + Y + Z | W + X + Y + Z = 150) = 150 \cdot \frac{5}{8} \cdot \frac{3}{8} = 35.15625$

2.3 Understanding Bayes Nets

	A = 0		**A = 1**	
	B = 0	B = 1	B = 0	B = 1
C = 0	0.36	0.20	0.00	0.00
C = 1	0.04	0.20	0.10	0.10

The **joint probability table (above)** for random variables A, B and C is equivalent to the **Bayes network (below)**. Both give the probability of any combination of the random variables. In the Bayes network the probability of each random variable is provided given its causal parents.



- Use the Bayes network to explain why $P(A = 0, B = 1, C = 1) = 0.20$

$P(A = 0, B = 1, C = 1) = P(A = 0)P(B = 1)P(C = 1|A = 0, B = 1) = 0.8*0.5*0.5 = 0.2.$

- What is $P(A = 1|C = 1)$?

Using the table, we see that

$$P(A = 1|C = 1) = \frac{0.1 + 0.1}{0.1 + 0.1 + 0.2 + 0.04} = \frac{0.2}{0.44} = \frac{5}{11}$$

- Is A independent of B? Explain your answer.

Yes. This follows directly from the structure of the bayesian network, because A and B have no shared ancestors. Alternatively, note that $P(A = a, B = b) = P(A = a)P(B = b)$, which satisfies the definition of independence.

- Is A independent of B **given** $C = 1$? Explain your answer.

No. From the table, we can see that $P(B = 1|A = 0, C = 1) = \frac{0.2}{0.4+0.2} \neq P(B = 1|A = 1, C = 1) = \frac{0.1}{0.1+0.1}$. So given $C = 1$, knowing the value of A informs us about the value of B, and therefore A and B are not conditionally independent given C.

Note: This phenomenon is sometimes called ”Explaining Away” if you’re curious to read more.