CS109 May 2nd, 2024

Continuous Random Variables, Joint Distributions

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 *Reviewing the Basics*

a. Given a Normal RV $X \sim N(\mu, \sigma^2)$, how can we compute $P(X \leq x)$ from the standard Normal distribution Z with CDF ϕ?

b. What is a continuity correction and when should we use it?

c. If we have a joint PMF for discrete random variables $p_{X,Y}(x, y)$, how can we compute the marginal PMF $p_X(x)$?

a. First, we apply a linear transformation to arrive at $\Phi((x - \mu)/\sigma)$. We then look up the value we've computed in the Standard Normal Table (or we rely on Python to compute the probability for us).

b. Continuity correction is used when a Normal distribution is used to approximate a Binomial. Since a Normal is continuous and Binomial is discrete, we have to use a continuity correction to discretize the Normal. The continuity correction makes it so that the normal variable is evaluated from + or - 0.5 increments from the desired k value.

c. The marginal distribution is $p_X(x) = \sum_y p_{X,Y}(x, y)$

1.2 *Independent Random Variables*

a. What distribution does the sum of two independent binomial RVs $X + Y$ have, where $X \sim Bin(n_1, p)$ and $Y \sim Bin(n_2, p)$? Include any parameters with your answer.

b. What distribution does the sum of two independent Poisson RVs $X + Y$ have, where $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$? Include any parameters with your answer.

a. Binomial: $X + Y \sim Bin(n_1 + n_2, p)$

b. Poisson: $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$

2 Marguerite Gets Some Competition

In the late 1880s, Stanford began running a horse and twelve-person buggy service from the Stanford Quad to the train station just off campus. The name of this shuttling service was chosen to be **Marguerite**, which was the name of the favorite horse of some Stanford bigwig of the time. The horse-and-carriage operation was retired around 1910 and replaced with electric streetcars, which themselves were replaced with buses around 1930. The service has grown substantially since, and the buses have been upgraded several times. The service, however, has retained its name since the very beginning.

Several Stanford horse enthusiasts have recently revived the horse-and-buggy service to compete with Marguerite, and they've given it the name **Hildegard**. Now, when you need a ride from the Quad to the train station, you have two options!

You arrive at the Quad, headed to the train station, and you're equally happy to take either of the two independent services. You arrive precisely at 8:00am, which is the time that both services start for the day. The number of minutes you need to wait for a Marguerite bus is modeled by a **discrete** Uniform random variable $M \sim Uni(0, 20)$, whereas the number of minutes you need to wait for a Hildegard horse-and-buggy is modeled by a discrete Poisson random variable $H \sim Poi(10)$. (Yes, it's technically possible that Hildegard never arrives.)

a. What is the probability that Marguerite and Hildegard both arrive at $t = 6$ minutes?

We can represent the events that the Marguerite and Hildegard arrive at $t = 6$ minutes as $M = 6$ and $H = 6$, respectively. Write

$$\begin{aligned} P(M = 6, H = 6) &= P(M = 6)P(H = 6) && (M \perp H) \\ &= \frac{1}{21} \cdot \frac{10^6 e^{-10}}{6!}. \end{aligned}$$

b. What is the conditional probability that $H < M$, given $M = m$—that is, what is $P(H < M|M = m)$? Express your answer as a sum.

We were looking for something along the lines of this:

$$\begin{aligned} P(H < M|M = m) &= P(H < m) && \text{(specifying value of } M\text{)} \\ &= \sum_{h=0}^{m-1} P(H = h) \\ &= \sum_{h=0}^{m-1} \frac{10^h e^{-10}}{h!}. \end{aligned}$$

c. What is the unconditional probability that $H < M$, i.e., what is $P(H < M)$? Express your answer as a double sum that leverages your answer to part b.

We were looking for something like this:

$$\begin{aligned}
P(H < M) &= \sum_m P(H < M, M = m) && \text{(Law of Total Probability)} \\
&= \sum_m P(H < M | M = m) P(M = m) && \text{(Chain Rule)} \\
&= \sum_{m=0}^{20} \frac{1}{21} \cdot P(H < M | M = m) \\
&= \frac{1}{21} \sum_{m=0}^{20} \sum_{h=0}^{m-1} \frac{10^h e^{-10}}{h!} && \text{(from part b)}
\end{aligned}$$

d. What is the CDF of your waiting time for the first of the two to arrive? You should leave your answer in summation form.

Let S be the random variable representing your waiting time for the first of the two to arrive. Then we have $S = \min\{M, H\}$, and accordingly that $S \in \{0, 1, \ldots, 20\}$. Letting F_S be the CDF of S and $s \in \{0, 1, \ldots, 20\}$, write

$$\begin{aligned}
F_S(s) &= P(S \le s) && \text{(definition of CDF)} \\
&= P(\min\{M, H\} \le s) && \text{(definition of } S\text{)} \\
&= P(M \le s \cup H \le s) \\
&= 1 - P\Big(\big(M \le s \cup H \le s\big)^C\Big) \\
&= 1 - P(M > s, H > s) && \text{(DeMorgan's Law)} \\
&= 1 - P(M > s) P(H > s) && (M \perp H) \\
&= 1 - \left(\sum_{m=s+1}^{20} P(M = m)\right)\left(\sum_{h=s+1}^{\infty} P(H = h)\right) \\
&= 1 - \left(\sum_{m=s+1}^{20} \frac{1}{21}\right)\left(\sum_{h=s+1}^{\infty} \frac{10^h e^{-10}}{h!}\right) \\
&= 1 - \left(\frac{20 - s}{21}\right)\left(1 - \sum_{h=0}^{s} \frac{10^h e^{-10}}{h!}\right).
\end{aligned}$$

3 Burrow Smoke Detectors and Joint Probability Distributions

Burrow Labs has taken on other startups in the home safety and security space and has recently started marketing a new smoke detector. Burrow's smoke detectors rely on CO_2 sensors that eventually fail, and that failure time dictates the average product lifetime of the smoke detector.

Burrow manufactures three quarters of its smoke detectors in central Idaho, and the rest are manufactured in suburban Maine. Any single smoke detector's product lifetime can be modeled as a Exponential random variable.

Each of the two locations sources its CO_2 sensors from different suppliers, so the smoke detectors manufactured in Maine have an average product lifetime of 7 years and the smoke detectors manufactured in Idaho have an average product lifetime of 6 years. All smoke detectors are sold online, so aside from the fact that a smoke detector is three times more likely to ship from the Idaho facility, you can't tell by looking at a single smoke detector where it was manufactured.

Let T model the amount of time that passes until the CO_2 sensor (and therefore the smoke detector) fails, and let M be a discrete random variable that takes on the value of 1 for a smoke detector manufactured in Maine, and 0 otherwise.

a. Present the cumulative distribution and probability density functions for the random variable T. Both your CDF and your PDF should be analytic functions on t.

The Law of Total Probability applies to all probabilities, including cumulative ones relevant to continuous distributions. That means that:

$$\begin{aligned} F_T(t) = P(T \leq t) &= P(T \leq t | M = 1) \cdot P(M = 1) + P(T \leq t | M = 0) \cdot P(M = 0) \\ &= (1 - e^{-t/7}) \cdot \frac{1}{4} + (1 - e^{-t/6}) \cdot \frac{3}{4} \\ &= 1 - \frac{1}{4}e^{-t/7} - \frac{3}{4}e^{-t/6} \\ f_T(t) = F_T'(t) &= \frac{1}{28}e^{-t/7} + \frac{3}{24}e^{-t/6} \end{aligned}$$

Of course, these are all defined for non-negative values of t.

b. Compute the probability that a smoke detector was manufactured in Maine, given that it lasts more than 15 years. If needed, you can keep your answer in terms of $F_T(15)$ or $f_T(15)$ from part (a). However, any conditional expression of the form $P(\cdot|\cdot)$ or $f(\cdot|\cdot)$ must be evaluated.

We once again rely on a hybrid form of Bayes's Theorem, although this time the probabilities require we integrate an accumulation of probability densities on T for t greater than 15 hours.

$$\begin{aligned}
P(M = 1|T > 15) &= \frac{P(T > 15|M = 1)P(M = 1)}{P(T > 15)} \\
&= \frac{1}{4} * \frac{1 - P(T \leq 15|M = 1)}{1 - P(T \leq 15)} \\
&= \frac{1}{4} * \frac{1 - (1 - e^{-15/7})}{1 - F_T(15)} \\
&= \frac{1}{4} * \frac{e^{-15/7}}{1 - F_T(15)} \\
&= 0.32268
\end{aligned}$$

4 Elections

We would like to see how we could predict an election between two candidates in France (A and B), given data from 10 polls. For each of the 10 polls, we report below their sample size, how many people said they would vote for candidate A, and how many people said they would vote for candidate B. Not all polls are created equal, so for each poll we also report a value ”weight” which represents how accurate we believe the poll was. The data for this problem can be found on the class website in polls.csv:

Poll	N samples	A votes	B votes	Weight
1	862	548	314	0.93
2	813	542	271	0.85
3	984	682	302	0.82
4	443	236	207	0.87
5	863	497	366	0.89
6	648	331	317	0.81
7	891	552	339	0.98
8	661	479	182	0.79
9	765	609	156	0.63
10	523	405	118	0.68
Totals:	**7453**	**4881**	**2572**	

a. First, assume that each sample in each poll is an independent experiment of whether or not a random person in France would vote for candidate A (disregard weights).

   - Calculate the probability that a random person in France votes for candidate A.
   - Assume each person votes for candidate A with the probability you’ve calculated and otherwise votes for candidate B. If the population of France is 64,888,792, what is the probability that candidate A gets more than half of the votes?

b. Nate Silver at fivethirtyeight pioneered an approach called the ”Poll of Polls” to predict elections. For each candidate A or B, we have a random variable S_A or S_B which represents their strength on election night (like ELO scores). The probability that A wins is $P(S_A > S_B)$.

- Identify the parameters for the random variables S_A and S_B. Both S_A and S_B are defined to be normal with the following parameters:

  $$S_A \sim \mathcal{N}\left(\mu = \sum_i p_{A_i} \cdot \text{weight}_i,\ \sigma^2\right) \qquad S_B \sim \mathcal{N}\left(\mu = \sum_i p_{B_i} \cdot \text{weight}_i,\ \sigma^2\right)$$

  where p_{A_i} is the ratio of A votes to N samples in poll i, p_{B_i} is the ratio of B votes to N samples in poll i, weight$_i$ is the weight of poll i, m_i is the N samples in poll i and:

  $\sigma = \frac{K}{\sqrt{\sum_i m_i}}$ s.t. $K = 350$; thus σ = 4.054.

- We will calculate $P(S_A > S_B)$ by simulating 100,000 fake elections. In each fake election, we draw a random sample for the strength of A from S_A and a random sample for the strength of B from S_B. If S_A is greater than S_B, candidate A wins. What do we expect to see if we simulate so many times? What do we actually see?

c. Which model, the one from (a) or the model from (b) seems more appropriate? Why might that be the case? On election night candidate A wins. Was your prediction from part (b) ”correct”?

a. $P(\text{random person votes for A}) = \frac{votesforA}{totalvotes} = \frac{4881}{7453} = 0.655$
Now, let X be the number of votes for candidate A. We assume that X ˜$Bin(64888792, 0.655)$.

- Since n is so large, we can approximate X using a normal Y ˜$N(np, np(1-p))$.
- $\mu = np = 42502158.76$, Variance $= np(1-p) = 14663244.77$ Std Dev = 3829.26
- Votes to win = $\frac{64888792}{2} = 32444396$
- $P(\text{A gets enough votes}) = P(X > 32444396) \approx P(Y > 32444396.5) = 1.00$

b. S_A ˜$N(5.324, 16.436)$
S_B ˜$N(2.926, 16.436)$
$P(S_A > S_B) \approx 0.66$
We can figure this out through simulation by drawing from S_A and S_B 100,000 times and seeing how often the S_A value is greater than the S_B value. Later in the quarter, when we learn the convolution of independent Gaussians, you will be able to figure this out mathematically, without sampling.

c. Algorithm (a) makes very few assumptions, and simplicity can be useful, but it does assume that each voter is independent, which we definitely know isn’t the case in real elections. Algorithm (b) allows us to model bias (using the weights we incorporated), and doesn’t think of each voter as necessarily independent.