Math 170E Week 8

Aidan Jan

March 6, 2023

Independence of Multiple Random Variables

Suppose A, B, and C are random variables, and $A \perp B$, $B \perp C$, and $C \perp A$. The following are true:

$$A^c \perp B^c$$

$$A \cap B \perp C$$

However, that does not necessarily imply that the three are independent from each other.
Now, suppose that X, Y, Z are independent. By definition, the following is true:

$$\forall k, l, m, \mathbb{P}(X = k \cap Y = l \cap Z = m) = \mathbb{P}(X = k) \cdot \mathbb{P}(Y = l) \cdot \mathbb{P}(Z = m)$$

The same can be expanded to more variables.

$$x_1, x_2, x_3, x_4 \leftrightarrow k_1, k_2, k_3, k_4$$

$$\mathbb{P}\left(\bigcap_{n=1}^{4} x_n(k_n)\right) = \prod_{n=1}^{4} (p_{x_n}(k_n))$$

Lemma:

$$\{X, Y **\} \perp \{Z **\}$$

Recall: if $X \perp Y$, then $f(X) \perp g(Y)$. Thus, if X, Y, Z are independent, then $f(X), g(Y), h(Z)$ are independent.
Example:

$$X, Y, Z \text{ are independent. Thus, } \sin(X), \cos(Y), e^Z \text{ are independent}$$

Application - the Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be independent random variables that all have the same distribution. Thus,

$$\mathbb{E}X_1 = \mu, \mathrm{Var}(X_1) = \sigma^2$$

Also, the sum, S_n is defined by the equation:

$$S_n = \sum_{k=1}^{n} X_n$$

1. To find the expected value of the sum, $\mathbb{E}S_n$, we can do

$$\mathbb{E}\sum_{k=1}^{n} X_k = \sum_{k=1}^{n}(\mathbb{E}X_k) = n \cdot \mu$$

2. We may also want to know expected value of variance, which is $\mathbb{E}(S_n - n \cdot \mu)^2$. This can simplify:

$$\begin{aligned}
&= \left(\sum_{k=1}^{n} X_k - n \cdot \mu\right)^2 \\
&= \left(\sum_{k=1}^{n} X_k - \mu\right)^2 \\
&= \big((X_1 - \mu)^2 + (x_1 - \mu)(x_2 - \mu) + (x_1 - \mu)(x_3 - \mu) \cdot \ldots\big)^2 \\
&= \sum_{k=1}^{n}\sum_{l=1}^{n} \mathbb{E}(X_k - \mu)(X_l - \mu)
\end{aligned}$$

If $(k + l)$ e.g. $\mathbb{E}(X_2 - \mu)(X_3 - \mu)$ Since $X_2 \perp X_3$, $X_2 - \mu \perp X_3 - \mu$. Thus,

$$\mathbb{E}(X_2 - \mu) = \mathbb{E}X_2 - \mathbb{E}\mu = \mu - \mu = 0$$

Therefore,

$$\mathbb{E}(S_n - \mathbb{E}S_n)^2 = n \cdot \sigma^2$$

Let's extend this example further. Consider $\mathbb{E}(S_n - \mathbb{E}S_n)^3$. This is very similar to the one above, but instead we are summing

$$\sum_{k,l,m} \mathbb{E}(X_1 - \mu)(X_2 - \mu)(X_3 - \mu)$$

This simplifies to

$$\mathbb{E}(X_1 - \mu) \cdot \mathbb{E}(X_2 - \mu) \cdot \mathbb{E}(X_3 - \mu) = 0$$

Lemma:

$$\mathbb{E}(XYZ) = \mathbb{E}(XY) \cdot \mathbb{E}(Z)$$

This is because $XY \perp Z \implies \mathbb{E}X \cdot \mathbb{E}Y \cdot \mathbb{E}Z$.
For example, $\mathbb{E}(X_1-\mu)^2(X_2-\mu) = 0$. This is because

$$\begin{aligned} &X_1 \perp X_2 \\ \implies &(X_1-\mu)^2 \perp (X_2-\mu) \\ =&\mathbb{E}(X_1-\mu)^2(X_2-\mu) \\ &= \mathbb{E}(X_1-\mu)(X_1-\mu)(X_1-\mu) \end{aligned}$$

So we can calculate

$$\mathbb{E}(S_n - n\mu)^2 = (n\sigma)^2$$
$$\left(\frac{\mathbb{E}(S_n - n\mu)}{\sqrt{n}}\right)^2 = \sigma^2$$

This can be expanded for any power of k.

$$EX^k \forall k \implies \text{Distribution of } X$$

As a side note,

$$\frac{S_n - n \cdot \mu}{\sqrt{n}}$$

is the Gaussian distribution.

Conditional Distributions

$$p_{X|Y}(k|l) = \mathbb{P}(X=k|Y=l)$$

Example:
Toss a dice. $Y = \#$
Toss Y coins. $X = \#$ heads

$$\mathbb{P}(X=k|Y=l) = \binom{l}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{l-k}$$

The formula:

$$\boxed{p_{X,Y}(k,l) = p_{X|Y}(k|l) \cdot \mathbb{P}(Y=l) = p_{X|Y}(k|l) \cdot p_Y(l)}$$

Also, the formula can be written like this:

$$\boxed{p_X(k) = \sum_l p_{X,Y}(k,l) = \sum_l p_{X|Y}(k|l) \cdot p_Y(l)}$$

Non-discrete Random Variables

So far, we have only looked at discrete random variables.

$$\mathbb{P}(X = 1) = \frac{1}{2}$$

This is a discrete random variable, since it can be assumed that X will only ever be integers.

Cantor Set are sets that are half-discrete, half-continuous, and are defined as a set of points on a line segment. (Search this up) They are used very rarely.

An example of a continuous random variable is $X \sim \text{unif}[0, 1]$, since X can be any value in a range, all with equal chance.

Coin game

For each round, a coin is flipped. Everyone either chooses heads or tails and those who guess right move to the next round. Continue until there is only one winner.

(a). If there are 70 students, what is $\mathbb{P}(\text{You win the game})$? This is an easy problem since at the start before any predictions have been made, everyone has an equal chance of winning. Thus, it is $\frac{1}{70}$.
(b). Now, what is $\mathbb{P}(\text{Winner is a girl})$? This will be $\frac{|G|}{|S|}$, where G is the number of girls, S is the total number of students (70).
(c). Chance that you guess correctly in the first three rounds? For each round you have a $\frac{1}{2}$ chance of guessing correctly. Thus, the probability is $\frac{1}{8}$.
(d). What about the chance that out of the 70 students, exactly 40 students stay after the first round? This is a binomial distribution since each trial (student) has a half chance (p-value). So the solution would be

$$\text{Bin}\left(70, \frac{1}{2}\right) = \binom{70}{40} \cdot \left(\frac{1}{2}\right)^{40} \cdot \left(\frac{1}{2}\right)^{70-40}$$

(e). Finally, what is the expected number of rounds before a winner? This can be written as $\mathbb{E}(\#\text{ rounds})$. To solve this, thinking about the game, a person only wins when there is only one person who picks correctly. Therefore, a round is successful if and only if one person picks correctly.
If two people are left in the round, then the chance that exactly one person guesses correctly is $\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$.
This is a geometric distribution, X represents the number of rounds the game goes on for, and it would follow a Geometric distribution, $X \sim \text{Geo}(\frac{1}{2})$. Therefore, if there are only two people in the round, the expected number of rounds it will be before the game ends is 2.
Now, let's consider the case with $n = 3$ people. Let B_0, B_1, B_2, B_3 be a partition, where B_k represents the case where k people guess correctly in the first round.
The Total Expected Value Theorem states that

$$\mathbb{E}X = \mathbb{E}(X|B_0) \cdot \mathbb{P}(B_0) + \mathbb{E}(X|B_1) \cdot \mathbb{P}(B_1) + \mathbb{E}(X|B_2) \cdot \mathbb{P}(B_2) + \mathbb{E}(X|B_3) \cdot \mathbb{P}(B_3)$$

It turns out from here, this problem is easy. This is because we know the distributions of all the partitions - they follow binomial distributions, of $\text{Bin}(n, k)$. We can also find $\mathbb{E}(X|B_k)$. $\mathbb{E}(X|B_1) = 1$, since if only one

person guesses correctly, then the game is over. $\mathbb{E}(X|B_2) = 3$, since it takes one round to get to the situation with only two people, plus the two rounds expected for the case with two people. $\mathbb{E}(X|B_0)$ and $\mathbb{E}(X|B_3)$ are similar, since they would both recursively lead to the same case. Thus, $\mathbb{E}(X|B_0) = \mathbb{E}(X|B_3) = \mathbb{E}X + 1$.
This method not only works for $n = 3$, but for any general n. As n gets bigger, large compared to the $\frac{1}{2}$ chance of success, the binomial distributions can be very closely approximated by Poisson distributions, since $\text{Bin}(n,p) \approx \text{Poi}(np)$.

Continuous Random Variables

Continuous random variables have non-discrete sample spaces.

Example:
Suppose you have a spinner divided into many wedges, one of which is labeled "stop." If you spin the spinner, what is the probability that it lands on "stop?" In this case, the value of the variable is the angle of the spinner. This can be calculated by

$$\mathbb{P}(\text{Stop at "stop"}) = \frac{\text{Length of Arc containing "stop"}}{\text{Circumference of spinner}} = \frac{\theta}{2\pi}$$

Example:
Suppose you have a unit square A, and you pick a random point inside the square, $P \in A$. X is the minimum distance between P and the boundaries of the square. Thus,

$$\text{range}(X) = [0, \frac{1}{2}$$

In this case, X is a continuous random variable. This question is not very difficult to write a distribution for, since there is only one point. However, what if you had two points? Consider a point Q randomly selected, and let its distance to a boundary be denoted by Y. What is $Z = \min(X, Y)$? Another thing we can ask is what is $\mathbb{P}(X \geq \frac{1}{3}$? To solve this, draw another boundary inside the square where all the sides are $\frac{1}{3}$ from the square boundary. Thus, the small boundary that represents all the points where $X \geq \frac{1}{3}$ has an area of $\frac{1}{9}$. $\mathbb{P}(X \geq \frac{1}{3} = \frac{1}{9}$.

Example:
Another example of a continuous random variable would be, consider a rectangle with infinite height, and infinite width. The rectangle is divided into pieces by red lines spaced 1 unit apart. Now, pick a random point, P on the rectangle and let X denote the minimum distance from P to a red line. What is $\mathbb{P}(X > 0.2)$? To do this, we can do the same thing as the previous problem, asking where X was greater than $\frac{1}{3}$ from the boundary of the square. In this case, for X to be greater than 0.2, draw boundaries next to the each red line, 0.2 away from it. As a result, the red space has a height of 0.4, while the area between the red areas is 0.6. Therefore, the probability that $X > 0.2$ would be $\frac{0.6}{1} = 0.6$.

Probability Mass Functions of Continuous Random Variables

Probability mass functions are only for discrete random variables; they do not make sense for continuous random variables. However, we can still calculate probabilities of ranges.
Suppose you have X that is uniformly distributed between $[0, 2]$. What is $\mathbb{P}(X = 1)$? It makes intuitive

sense that this probability has a chance of 0, but we can also prove this. First, $\mathbb{P}(X = 1) \leq \mathbb{P}(1 \leq X \leq 1.1)$. Therefore, $\mathbb{P}(X = 1) \leq \frac{0.1}{2} = 0.05$. Now, consider a smaller interval. $\mathbb{P}(X = 1) \leq \mathbb{P}(1 \leq X \leq 1.01$. From this, we can calculate that $\mathbb{P}(X = 1) \leq 0.005$. This process can continue forever, using smaller and smaller intervals. Therefore, the probability that $X = n$ where n is any number, is 0. If this is true, then the probability mass function would state that for any number, the probability is 0. There is an infinite number of numbers X can be, so the probabilities still add to 1 (indeterminate), but this is not useful.

Instead, we use **probability density functions**, which describe which ranges of numbers X is more likely to land on compared to others. We also use (cumulative) distribution functions to describe continuous random variables, denoted as F_X.

F_X is a function from $\mathbb{R} \to \mathbb{R}$.

$$\boxed{F_X(a) := \mathbb{P}(X \leq a)}$$

F_X **ALWAYS** compares X being **less than or equal** to a value. We also define $G_X(a) = \mathbb{P}(X < a)$. (This is not the same as F_X even though $\mathbb{P}(X = a) = 0$.)

Example:

Consider unif$[0, 2] \to X$.

$$F_X(a) = \begin{cases} 0 & a < 0 \\ \frac{a}{2} & 0 < a < 2 \\ 1 & a > 2 \end{cases}$$