CS 70 Discussion 12A

November 20, 2024

Continuous Random Variable

Problem: Let's say we are no longer dealing with random variables that take a countable set of values (ex. Geometric, Binomial, and Poisson random variables), and want to now represent a continuous space.

Solution: A **continuous random variable** is a random variable X where for each $x \in \mathbb{R}$, $\mathbb{P}[X = x] = 0$, but you have $\mathbb{P}[X \in \mathbb{R}] = 1$ (i.e. the probability of our random variable taking a *specific* value is infinitely small, but the probability that it can take some *range* of values is not necessarily zero).

Cumulative Density Function (CDF)

Problem: If we can no longer meaningfully define PMFs (i.e. $\mathbb{P}[X = x] = 0$ for any $x \in \mathbb{R}$) for a continuous random variable, how do we define a continuous random variable and its distribution?

Solution: We define a **cumulative density function (CDF)**, which is just a function $F_X(x) = \mathbb{P}[X \leq x]$. Any CDF F_X needs the following properties to hold:

- Non-decreasing: $(\forall a \leq b)(F_X(a) \leq F_X(b))$
- Limits: $\lim_{x \to \infty} F_X(x) = 1$ and $\lim_{x \to -\infty} F_X(x) = 0$

Probability Density Function (PDF)

Problem: Is there an analogue to PMFs for continuous random variables?
Solution: We define the **probability density function (PDF)** as a function $f_X(x) = \frac{d}{dx} F_X(x)$. Any PDF f_X needs the following to hold:

- Non-negativity: $(\forall x \in \mathbb{R})(f_X(x) \geq 0)$
- Integrates to 1: $\int_{-\infty}^{\infty} f_X(x)dx = 1$

An important fact to note is:

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$$

Important: PDFs are not probabilities, just densities!

Joint Distributions & Independence

Problem: What about joint distributions and independence with continuous random variables?
Solution: $f_{X,Y}(x, y)$ is the PDF of the joint distribution for random variables X and Y, and $F_{X,Y}(x, y) = \mathbb{P}[X \leq x, Y \leq y]$. Our definition of independence is that continuous random variables X and Y are independent iff:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y), F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

You only need to show one of the above to prove independence (because either both hold or neither hold). You can also use the total probability rule with continuous random variables (consider event A and continuous random variable X):

$$\mathbb{P}[A] = \int_{-\infty}^{\infty} \mathbb{P}[A|X = x]f_X(x)dx$$

Expectation, Variance, Covariance, etc.

For continuous random variables, a rule of thumb for calculations is that the math is generally the same as with discrete random variables except you use integrals and PDFs instead of summations and PMFs:

- $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx$
- $\mathbb{E}[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) dy dx$
- $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$
- $\text{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$
- $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

Exponential Distribution

Problem: Consider the continuous-time analogue for the geometric random variable. What is the expected amount of time until we get a success?

Solution: We define a random variable $X \sim \text{Exponential}(\lambda)$ (where $\lambda > 0$) as a random variable following these properties:

- PDF: $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$
- CDF: $F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$
- Memorylessness: $\mathbb{P}[X \geq x + c | X \geq x] = \mathbb{P}[X \geq c]$
- $\mathbb{E}[X] = \frac{1}{\lambda}, \text{Var}(X) = \frac{1}{\lambda^2}$

Note: Continuous distributions are uniquely defined by their CDF or PDF