Statistics 1 : Continuous Distributions and Moments

Anand Systla

Masters in Financial Engineering Bootcamp
UCLA Anderson

August 13, 2022

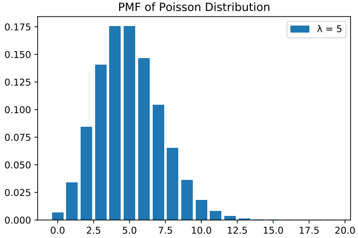
Today's Agenda

- Discuss the previous class take home questions
- Recap Poisson Distribution
- Uniform Distribution
- Exponential Distribution
- Normal Distribution
- Moment Generating Function

Poisson Distribution Recap

- Recall that a Poisson RV is a count of the number of occurrences of an event in a given unit of time/distance/area/volume, etc. Two assumptions in the Poisson distribution are - (1) rate at which events occur is constant and (2) occurrence of one event does not affect occurrence of a subsequent event. The PMF gives the probability of observing x independent and random events in a unit time. λ is average number of events

$$p(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots \qquad \text{and} \qquad E[X] = \sum_{x=0}^{\infty} x \cdot e^{-\lambda}\frac{\lambda^x}{x!} = \lambda$$



- What do some real life examples look like?

Binomial To Poisson

- Say we are interested in number of people visiting a website. Lets call the RV $X = \{\text{No. of visits per hour}\}$. Prior data tells us that we see about 20 visits per hour.
- If X was Binomial, $E[X] = np = 20$. We can think of $n = 60$ for example, as conducting 60 independent bernoulli trials each minute
- If X was poisson, $E[X] = \lambda = 20$
- Going from Bernoulli to Poisson as $n \to \infty$, $p = \lambda/n \to 0$

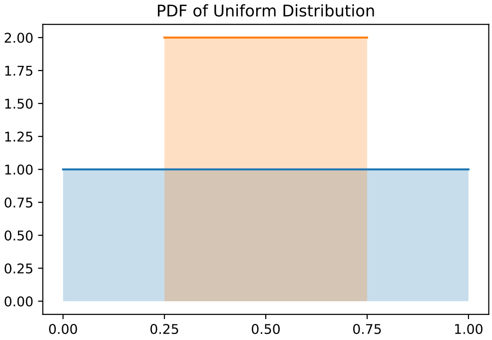
$$\lim_{n\to\infty} P(X = k) = \lim_{n\to\infty} {}^nC_k p^k (1-p)^{n-k} = \lim_{n\to\infty} {}^nC_k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}$$

- Intuition is that by going from hour to minutes, we are conducting 60 independent Bernoulli trials each minute. As we keep going to finer time intervals, we are sending $n \to \infty$ and $p \to 0$. At the limit, the Binomial gives us Poisson distribution. So the Poisson can also be thought of as conducting ∞ Bernoulli trials with a probability of success $p \sim 0$.

Uniform Distribution

- The Bernoulli of continuous distributions.

$$f(x) = \frac{1}{b-a}, \quad x \in [a, b] \qquad \text{and} \quad F(x) = \frac{x-a}{b-a}$$



- Another interesting fact is that the uniform distribution is present in every CDF. How?

Exponential Distribution

- Times between Poisson jumps are independent. The intensity of the exponential distribution is $\beta = 1/\lambda$. The intensity is the inverse of that of the Poisson distribution. It is a continuous distribution

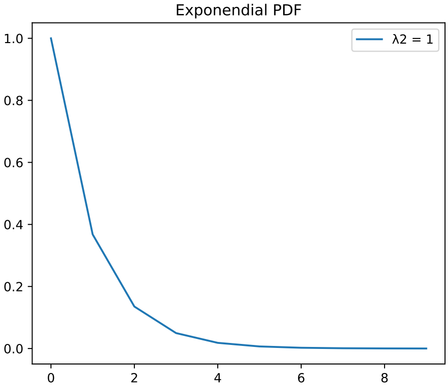
$$\text{Poisson} = \frac{\text{Number of events}}{\text{1 unit of time}} \qquad \text{Exponential} = \frac{\text{Amount of time}}{\text{1 Poisson event}}$$

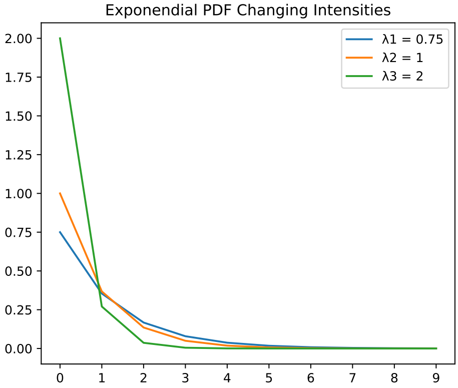
$$N \sim Pois(\lambda) \iff X \sim Exp(\beta = \frac{1}{\lambda})$$

$$X \sim Exp(\beta = \frac{1}{\lambda}), \qquad f(x) = \lambda e^{-\lambda x}, \qquad F(x) = 1 - e^{-\lambda x}, \qquad E[X] = \frac{1}{\lambda}$$

- What are some real life examples?

Exponential Distribution Example





Question

- Say a patient arrives every 20 minutes, what is

  1. What is the Poisson intensity (λ)?
  2. What is probability that a patient will arrive in 30 minutes?
  3. What is the probability that a patient will arrive exactly on the 10th minute?
  4. What is the probability that we will have 3 customers in 1 hour?

- Now consider $\beta = 1/\lambda = 20$. Graphical illustration of the question for intuition

  1. What is $P(0 \leq X \leq 1)$?
  2. What is $P(1 \leq X \leq 2)$?
  3. What is $P(X = 1)$?

Memoryless Property of Exponential Distribution

- Probability of an event occurring after $s+t$ minutes condition on event not occurring in the first s minutes, is the probability that the event is going to occur after t minutes

$$P(T \geq s+t | T \geq s) = P(T \geq t)$$

- Intuitively, say a patient visits a doctor every 20 minutes. The probability that a doctor will see a patient in the zero-eth and first minute and the probability that the doctor will see a patient between 22nd and 23rd minutes are the same. Graphical intuition?

Clarifications

- Are the following examples of Poisson distribution?
  - Temperature in a room that varies throughout the day
  - Number of people getting up to adjust the thermostat of a room in a day
- Memoryless vs Independence
  - Independence $P(X > A|X > B) = P(X > A)$
  - Memoryless $P(X > A|X > B) = P(X > A - B)$
- Random Variable vs Probability Distribution

Clarifications

- Are the following examples of Poisson distribution?
  - Temperature in a room that varies throughout the day
  - Number of people getting up to adjust the thermostat of a room in a day
- Memoryless vs Independence
  - Independence $P(X > A|X > B) = P(X > A)$
  - Memoryless $P(X > A|X > B) = P(X > A - B)$
- Random Variable vs Probability Distribution
  - Random variable is a function that maps events to a value on $\mathbb{R}$
  - Probability distribution is a function that maps values of the random variable from $\mathbb{R}$ to a value between $[0, 1]$
- Relationship between Poisson and Exponential distributions. Say $X \sim Pois(\lambda)$. λ is for 1 unit of time, then λt is for t units of time. So the probability of seeing x successes/events in t periods is

$$P(X = x) = e^{-\lambda t}(\lambda t)^x/x! \qquad P(X = 0) = e^{-\lambda t}$$

Variance and Standard Deviation

- How far are the values of the RV X from its mean $E[X]$. The variance is the probability weighted average squared distance from the mean

$$V[X] = E\left[(X - E[X])^2\right]$$

- Properties of the variance operator
  1. $V[c + X] = V[X]$
  2. $V[cX] = c^2 V[X]$
  3. $V[X + Y] = V[X] + V[Y] + 2COV(X, Y)$
- Why not absolute distance from the mean?

Normal Distribution

- A RV X is said to be normally distributed if it's PDF is $X \sim N(\mu, \sigma^2)$. Distribution is centered around μ and σ determines the speread

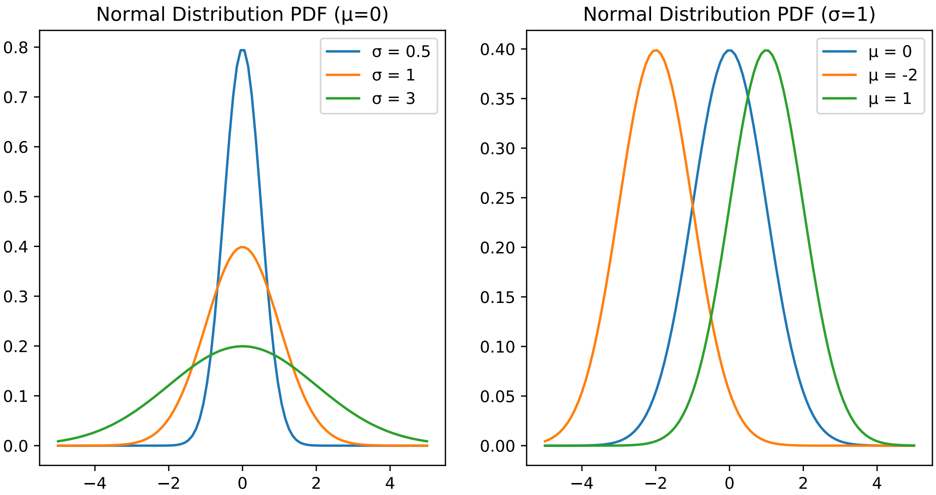
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in (-\infty, \infty), \qquad E[X] = \mu, \qquad V[X] = \sigma^2$$

$$Z \sim N(0,1), \qquad f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad \text{Kernel: } e^{-\frac{x^2}{2}}$$

$$X = \mu + \sigma Z, \qquad Z \sim N(0,1)$$

- Most data we observe in the real world is close to Normal
- $P(|X - \mu| > 6\sigma) < 10^{-8}$

Normal Distribution Example



Moment Generating Function (MGF)

- MGF lets us compute the moments quickly. Moments allow us to summarize the distribution of the RV. We are usually interested in the first few moments of any RV. Moments can either be central of non-central
- Non-central moments are $E[X], E[X^2], E[X^3]$ and so on
- The k-th central moments is $E[(X - E[X])^k]$. How is a RV distributed around a particular value (like the mean)

$$M_X(t) = E[e^{tX}]$$

$$e^{tX} = 1 + tX + \frac{(tX)^2}{2!} + ..$$

$$E[e^{tX}] = 1 + tE[X] + \frac{t^2}{2!}E[X^2] + ..$$

$$\frac{d}{dt}E[e^{tX}]|_{t=0} = 0 + E[X] + tE[X^2] + .. = E[X] \implies E[X^n] = \frac{d^n}{dt^n}M_X(t)|_{t=0}$$

Take Home Questions

1. A checkout counter at a supermarket completes a process according to an exponential distribution with a service rate of 6 per hour. A customer arrives at a checkout counter. Find the probability of
   1.1 Service is completed in less than 5 minutes
   1.2 Customer leaves the checkout counter more than 10 minutes after arriving
   1.3 Service is completed between 20 to 25 minutes
2. You have two normal distributions $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$. What is the distribution of $X + Y$? Also, what is the distribution of $X - Y$?