Probability and Statistics

Omid Safarzadeh

February 10, 2022

Table of contents

*Acknowledgement: This slide is prepared based on Casella and Berger, 2002

Section 1

Discrete Distribution

Discrete Uniform Distribution

- A random variable X has a discrete *uniform*(1,N) distribution if

$$P(X = x|N) = \frac{1}{N}, \qquad x = 1, 2, ..., N,$$

  where N is a specified integer. This distribution puts equal mass on each of the outcomes $1, 2, ..., N$.
- E(X)=(N+1)/2
- Var(X)=(N+1)(N-1)/2

Bernoulli Distribution

- A random variable X has Bernoulli(p) distribution if

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases} \qquad 0 \leq p \leq 1.$$

- $X = 1$ is often termed as "success" and p is, accordingly, the probability of success. Similarly, $X = 0$ is termed a "failure".
- Now,

$$E[X] = 1 * p + 0 * (1-p) = p,$$

$$\text{and} \quad Var(X) = (1-p)^2 p + (0-p)^2 (1-p) = p(1-p).$$

Binomial Distribution

- This is based on a Bernoulli trial which is an experiment with two, and only, two, possible outcomes.
- Assume, we have n trials of a Bernoulli distribution, and we are interested to probability of having y results as success. It means than n-y times we had failure. Also assume that these events are independent of each other. Hence: the distribution of the total number of successes in n trialsis Binomial Distribution
- Examples:
  1. Tossing a coin ($p =$probability of a head, $X = 1$ if heads)
  2. Election polls ($X = 1$ if candidate A gets vote)
  3. Probability of Default Risk ($p =$probability that a person defaults in his loan payments)
  4. in ML we use it to construct Binary Cross-Entropy Loss Function

Binomial Distribution

- Take $Y =$"total number of successes in n trials"
- There are many possible orderings of the events that would lead to this outcome. Any particular such ordering has probability

$$p^y(1-p)^{n-y}.$$

- Since there are $\binom{n}{y}$ such sequences, we have

$$P(Y=y|n,p) = \binom{n}{y} p^y (1-p)^{n-y}, \qquad y = 0, 1, ..., n,$$

  and Y is called a *binomial(n,p)* random variable.
- $E[X] = np$
- $Var(X) = np(1-p)$ (**Proof**: Exercise!)

Example 1.1

suppose it is known that 8% of all emails are spam. If an account receives 30 emails in a given day, use a Binomial Distribution Calculator to find the probability that a certain number of those emails are spam.

- $P(X = 1 \text{ spam emails}) = 0.21382$
- $P(X = 3 \text{ spam emails}) = 0.21881$
- $P(X = 10 \text{ spam emails}) = 0.00006$

And so on.

Poisson Distribution

- In modelling a phenomenon in which we are waiting for an occurrence (such as waiting for a bus), the number of occurrence in a given time interval can be modelled by the Poisson distribution.
- The basic assumption is as follows: for small time intervals, the probability of an arrival is proportional to the length of waiting time.
- If we are waiting for the bus, the probability that a bus will arrive within the next hour is higher than the probability that it will arrive within 5 minutes.
- Other possible applications are distribution of bomb hits in an area or distribution of fish in a lake.
- The only parameter is λ, also sometimes called the "intensity parameter."

Poisson Distribution

- $P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, ...$
- $E[X] = \lambda$
- $Var(X) = \lambda$
- **Proof**: Exercise!

Poisson Distribution

Example 1.2

As an example of a waiting-for-occurrence application, consider a telephone operator who, on average, handles fire calls every 3 minutes. What is the probability that there will be no calls in the next minute? At least two calls? If we let X =number of calls in a minute, then X has a Poisson distribution with $E[X] = \lambda = 5/3$. So,

$$P(\text{no calls in the next minute}) = P(X = 0)$$

$$= \frac{e^{-5/3}(5/3)^0}{0!} = e^{-5/3} = 0.189$$

and

$$P(\text{at least two calls in the next minute}) = P(X \geq 2)$$

$$= 1 - P(X = 0) - P(X = 1)$$

$$= 1 - 0.189 - \frac{e^{-5/3}(5/3)^1}{1!}$$

$$= 0.496.$$

Network Failures/Week

Example 1.3

suppose a company experiences an average of 3 network failure per week. Use Poisson distribution to find the probability that the company experiences a certain number of network failures in a given week:

$E(X) = \lambda = 3$. So

- $P(X = 0 \text{ failures}) = 0.04979$
- $P(X = 1 \text{ failures}) = 0.14936$
- $P(X = 2 \text{ failures}) = 0.22404$ ...

so you have some idea of how many failures are likely to occur each week.

Corporate Banking Default Rate

Example 1.4

A bank has an average of 6 bankruptcies filed by corporate customers each month. Using Poisson distribution, find the probability that the bank receives a specific number of default claims in a given month.

$E(X) = \lambda = 6$. So

- $P(X = 0 \text{ bankruptcies}) = 0.00248$
- $P(X = 1 \text{ bankruptcy}) = 0.01487$
- $P(X = 2 \text{ bankruptcies}) = 0.04462$
- $P(X = 2 \text{ bankruptcies}) = 0.08924$

And so on.
This provides an insight for bank mangers, on how much cash reserve to keep on hand in case a certain number of bankruptcies occur in a given month.

Calls/Hour rate at a Call Center

Example 1.5

A call center receives 12 calls per hour. Use Poisson distribution to find the probability that a call center receives 2, 4, 6, 8 ... calls in a given hour

$E(X) = \lambda = 12$. So

- $P(X = 2 \text{ calls}) = 0.00044$
- $P(X = 4 \text{ calls}) = 0.00531$
- $P(X = 6 \text{ calls}) = 0.02548$
- $P(X = 8 \text{ calls}) = 0.06552$

A CRM managers can understand how many calls most likely the call center will receive per hour.

Section 2

Continuous Distribution

Uniform Distribution

- The continuous uniform distribution is defined by spreading mass uniformly over an interval $[a, b]$. Its pdf is given by

$$f(x|a,b) = \begin{cases} \frac{1}{b-a} & \text{if} \quad x \in [a,b] \\ \\ 0 & \text{if} \quad \text{otherwise} \end{cases}.$$

- One can easily show that

$$\int_a^b f(x)dx = 1,$$

$$E[X] = \frac{b+a}{2},$$

$$Var(X) = \frac{(b-a)^2}{12}.$$

- In many cases, when people say Uniform distribution, they implictly mean $(a, b) = (0, 1)$.

Exponential Distribution

- pdf of Exponential Distribution :

$$f(x|\beta) = \frac{1}{\beta} e^{-e/\beta}, \quad 0 < x < \infty.$$

- we have

$$E[X] = \beta \quad \text{and} \quad Var(X) = \beta^2$$

- this distribution is that it has no memory.

Exponential Distribution

- If $X \sim$exponential(β), then, for $s > t \geq 0$,

$$P(X > s|X > t) = \frac{P(X > s, X > t)}{P(X > t)} = \frac{P(X > s)}{P(X > t)}$$
$$= \frac{\int_s^\infty \frac{1}{\beta} e^{-x/\beta} dx}{\int_t^\infty \frac{1}{\beta} e^{-x/\beta} dx} = \frac{e^{-s/\beta}}{e^{-t/\beta}}$$
$$= e^{-(s-t)/\beta} = P(X > s - t).$$

- This is because,

$$\int_{s-t}^\infty \frac{1}{\beta} e^{-x/\beta} dx = -e^{-x/\beta}|_{s-t}^\infty = e^{-(s-t)/\beta}.$$

- What does this mean? When calculating $P(X > s|X > t)$, what matters is not whether X has passed a threshold or not. What matters is the distance between the threshold and the value to be reached.
- If Mr X has been fired more than 10 times, what is the probability that he will be fired more than 12 times? It is not different from the probability that a person, who has been fired once, will be fired more than two times. History does not matter.

Normal Distribution

- We now consider the normal distribution or the Gaussian distribution.
- Why is this distribution so popular?
  1. Analytical tractability
  2. Bell shaped or symmetric
  3. It is central to Central Limit Theorem; this type of results guarantee that, under (mild) conditions, the normal distribution can be used to approximate a large variety of distribution in large samples.
- The pdf is given by,

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} exp[-\frac{(x-\mu)^2}{2\sigma^2}].$$

Normal Distribution

- This distribution is usually denoted as $N(\mu, \sigma^2)$.
- A very useful result is that for $X \sim N(\mu, \sigma^2)$,

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

- $N(0, 1)$ is known as the standard normal distribution.
- To see this, consider the following:

$$\begin{aligned} P(Z \leq z) &= P(\frac{(X-\mu)}{\sigma} \leq z) \\ &= P(X \leq z\sigma + \mu) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{z\sigma+\mu} e^{-(x-\mu)^2}/2\sigma^2 dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt, \end{aligned}$$

where we substitute $t = (x - \mu)/\mu$. Notice that this implies that $dt/dx = 1/\sigma$. This shows that $P(Z \leq z)$ is the standard normal cdf.

Lognormal Distribution

- Let X be a random variable such that

$$\log X \sim N(\mu, \sigma^2).$$

  Then, X is said to have a lognormal distribution.
- By using a transformation Theorem , the pdf of X is given by,

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} exp[-\frac{(\log x - \mu)^2}{2\sigma^2}],$$

  where $0 < x < \infty$, $-\infty < \mu < \infty$, and $\sigma > 0$.

Example 2.1

Suppose the lifetime of an Engine has a lognormal distribution. What is the probability that the lifetime exceeds 12,000 hours if the mean and variance of the normal random variable are 11 hours and 1.3 hours, respectively?

$$\mu = 11, \quad \sigma = 1.3, \quad P(X > 12000) = ?$$

$$P(X \geq 12000) = 1 - P(X < 12000)$$

$$P(X > 12000) = 1 - \Phi(\frac{ln12000 - 11}{1.3})$$

$$P(X > 12000) = 1 - P(z \leq (\frac{9.393 - 11}{1.3}))$$

$$P(X > 12000) = 1 - P(z < -1.236)$$

$$P(X > 12000) = 1 - (0.10823)$$

$$P(X > 12000) = 0.8918$$

This means that there is about an 89.18% chance that a motor's lifetime will exceed 12,000 hours.

Laplace distribution

- If $X \sim Lap(\mu, b)$,

$$f(x|\mu, b) = \frac{1}{2b} exp(-\frac{|x-\mu|}{b})$$

- then $E[X] = \mu,\ Var(X) = 2b^2$
- The Lasso Regression is sort of a Bayesian regression with a Laplacian prior
- Laplace is applied to extreme events like rainfalls, river discharges

Beta distribution

- The pdf of the beta distribution, for $0 \leq x \leq 1$, and shape parameters $\alpha, \beta > 0$, is a power function of the variable x and of its reflection $(1-x)$ as follows:

$$
\begin{aligned}
f(x;\alpha,\beta) &= \text{constant}.x^{\alpha-1}(1-x)^{\beta-1} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\
&= \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}
\end{aligned}
$$

- $E[X] = \frac{\alpha}{\alpha+\beta}$
- $Var[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Example 2.2

Suppose that DVDs in a certain shipment are defective with a Beta distribution with $\alpha = 2$ and $\beta = 5$. Compute the probability that the shipment has 20% to 30% defective DVDs.

$$P(0.2 \leq X \leq 0.3) = \sum_{x=0.2}^{0.3} \frac{x^{2-1}(1-x)^{5-1}}{B(2,5)} = 0.235185$$

*Ref.

Section 3

Reference

Reference

Casella, G., & Berger, R. (2002). *Statistical inference*. Cengage Learning. https://books.google.fr/books?id=FAUVEAAAQBAJ

Wikipedia contributors. (2022, January 31). Beta distribution. In Wikipedia, The Free Encyclopedia. "Beta distribution"