Probability distributions

l## What is a probability distribution?

Think of a probability distribution as a function that maps an elementary event (i.e. every time an observation happens, one and only one of these events will happen) to the probability that that event happens.

Terms##

Probability mass function (PMF)
probability that a discrete random variable is exactly equal to some value (i.e. the definition of probability distribution that I gave above)

Cumulative density function (CDF)
Probability that a random variable (X) (or distribution function) will take a value less than or equal to x. This should always approach 1. This is the area under the **PMF** from $-\infty$ to x ($\int_{-\infty}^{x} \text{PMF}$).

Probabilty density
For continuous variables, it would not make sense to talk about the probability that X takes on a certain value since that assumes an infinite precision- so the probability of a continuous random variable attaining a specific value is 0. Instead, we use the probability density function- where you take the integral to find the probability that the observation will be within a range of values.

Binomial

Describes discrete events that end in success or failure.

Examples

- Roll a dice, what is the probability that it's 1? (success: dice=1, failure: dice =/= 1)
- flip a coin, what is the probability of getting heads? (success: head, failure: tail)

Binomial distributions have 2 parameters, N (or n) and θ (or p). N is the number of trials, θ is the probability of success.

Formula

$$P(X|\theta, N) = \frac{N!}{X!(N-X)!}\theta^X(1-\theta)^{N-X}$$

Properties

- expected value: $E[X] = N\theta$
- variance: $\mathrm{Var}(X) = N\theta(1-\theta)$

Normal

Bell curve distribution.

Parameters: mean (μ) and standard deviation (σ)

Formula

$$p(X|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(X-\mu)^2}{2\sigma^2})$$

Properties

- expected value: $E[X] = \mu$
- variance: $\mathrm{Var}(X) = \sigma^2$

t

Similar to normal distribution - use this when you think the data is normally distributed but don't know the mean or standard deviation

Chi-Square (χ^2)

Often found in categorical data analysis. This is the result of taking a sum of squares of normally distributed variables

F

Pops up when comparing two different χ^2 distributions

Poisson distribution

Discrete distribution which describes how many times an event (specifically a poisson process) is likely to occur within some period of time

poisson point process - a process that generates discrete events where the average time between events in known.
requirements:

- events are independent

- average rate is constant
- two events cannot occur at the same time

Formula

$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Interpret this as the probability that x events will happen in the time period where λ events happen per time period on average.

properties

- expected value: $E[X] = \lambda$
- variance: $V(X) = \lambda$

Exponential distribution

Continuous probability distribution modeling the time between events in a poisson point process

$$p(x|\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

properties

- expected value: $E[X] = 1/\lambda$
- variance: $V(X) = 1/\lambda^2$

References

- Learning statistic with R chapter 9
- https://en.wikipedia.org/wiki/Binomial_distribution
- https://byjus.com/maths/poisson-distribution/
- https://towardsdatascience.com/the-poisson-distribution-and-poisson-process-explained-4e2cb17d459
- https://en.wikipedia.org/wiki/Exponential_distribution