Data Science 1
Probability
Working with Discrete Distributions
Part 2

Edward L. Boone

Introduction

Recall: $f(x) = P(X = x)$

Probability Distributions we have seen so far:

Bernoulli: $f(x) = \begin{cases} p & \text{x} = 1 \\ 1 - p & \text{x} = 0 \end{cases}$

Binomial: $f(x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0, 1, 2, 3, ..., n$

Geometric: $f(x) = (1-p)^x p,\ x = 0, 1, 2, 3...$

Negative Binomial: $f(x) = \binom{x+k-1}{x} (1-p)^x p^k,\ x = 0, 1, 2, 3...$

Poisson: $f(x) = \frac{\lambda^x e^{-\lambda}}{x!},\ x = 0, 1, 2, 3...$

Notice that the values that the random variable X takes on are discrete (whole numbers).

Calculating Events

Suppose we want to calculate an event that may be difficult to do directly?

A mobile phone case manufacturer wants to know how strong his cases are. One of his experiments is to drop the phone with the case from 3 meters and record whether or not the phone is broken. Since they have limited phones they repeat the experiment until the phone breaks. Hence they are looking for the number of drops from 3m before the phone breaks. Specifically, the manufacturer wants to know what is the probability that the case/phone will withstand more than 10 drops from 3m?

X = Number of drops before the phone breaks. This is a geometric experiment since it is a count until a success.

Hence the probability they want to know is:

$$P(X > 10) = \sum_{x=11}^{\infty} P(X = x) = \sum_{x=11}^{\infty} (1-p)^x p$$

Properties of Probability Distributions

To solve the problem we need to think about probability distributions:

1. $P(X = x) \geq 0$ which means $f(x) \geq 0$.
2. $\sum_x P(X = x) = 1$ which means $\sum_x f(x) = 1$.

These properties will make it easier for us to calculate probabilites.

Back to our example:

Specifically, the manufacturer wants to know what is the probability that the case/phone will withstand more than 10 drops from 3m if the probabilty of a break is 0.08?

X = Number of drops before the phone breaks. This is a geometric experiment since it is a count until a success.

Hence the probability they want to know is:

$$P(X > 10) = \sum_{x=11}^{\infty} P(X = x) = \sum_{x=11}^{\infty} (1-p)^x p$$

We can reframe this as:

$$P(X > 10) = 1 - P(X \leq 10) = 1 - \sum_{x=0}^{10} P(X = x) = 1 - \sum_{x=0}^{10} f(x)$$

Back to our example:

Specifically, the manufacturer wants to know what is the probability that the case/phone will withstand more than 10 drops from 3m if the probabilty of a break is 0.08?

$$
\begin{aligned}
P(X > 10) &= 1 - P(X \leq 10) \\
&= 1 - \left(\sum_{x=0}^{10} (1 - 0.08)^x (0.08) \right) \\
&= 1 - \left(0.92^0 (0.08) + 0.92^1 (0.08) + 0.92^2 (0.08) + ... + 0.92^{10} (0.08) \right) \\
&= 1 - (0.6003) = 0.3996.
\end{aligned}
$$

This is much easier to calculate than an infinite sum!

Another Example:

Suppose Marie is a managing nurse at a small rural hospital. On any given day the average number of patients who come to the emergency room seeking care is about 2 and the staffing reflects this. She is worried about having too many patients come to the emergency room on any day, any more than 5 patients a day will cause a problem. What is the probability she sees more than 5 patients on any given day?

Since X is the number patients per day (count over time) this is a Poisson experiment with mean 2 (given above).

$$P(X > 5) = \sum_{x=6}^{\infty} P(X = x)$$

Too difficult!

Another Example:

Since X is the number patients per day (count over time) this is a Poisson experiment with mean 2 (given above).

$$
\begin{aligned}
P(X > 5) &= \sum_{x=6}^{\infty} P(X = x) \\
&= 1 - P(X \leq 5) \\
&= 1 - \left(\sum_{x=0}^{5} \frac{\lambda^x e^{-\lambda}}{x!} \right) \\
&= 1 - \left(\frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} + \ldots + \frac{2^5 e^{-2}}{5!} \right) \\
&= 1 - (0.1353 + 0.2706 + 0.2706 + 0.1804 + 0.0902 + 0.0361) \\
&= 1 - (0.9832) \\
&= 0.0168
\end{aligned}
$$

Hence the probability that more that 5 patients visit the emergency clinic on any given day is quite small 0.0168.

Summary

Working with probabilities can be simple when using the probabilty distribution.

- Probabilities can just add up.
- If the problem inovolves lots of computation ask "Will the complement work?".
- The knowledge that the densities will add up to 1 is also helpful in looking for errors.