Data Science 1
Probability
Working with Discrete Distributions
Part 1

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

Typically we want to calculate the probability of a specific event.

So far we have only considered a single value for $X = x$.

What if we want the probability for a range of values? Just add up the individual probabilities!

$$
\begin{aligned}
P(a \leq X \leq b) &= \sum_{x=a}^{b} P(X = x) \\
&= \sum_{x=a}^{b} f(x) \\
&= f(a) + f(a+1) + \ldots + f(b-1) + f(b)
\end{aligned}
$$

Example 1:

Suppose Benoit is a lecturer giving a multiple choice test with 10 questions each with four options that are scored as either correct or incorrect. A passing score is 7 or above. If a student simply randomly guesses all of the questions, what is the probability they achieve a passing score?

Example 1:

Suppose Benoit is a lecturer giving a multiple choice test with 10 questions each with four options that are scored as either correct or incorrect. A passing score is 7 or above. If a student simply randomly guesses all of the questions, what is the probability they achieve a passing score?

This is a **Binomial** experiment with $n = 10$ and $p = 1/4$ (randomly pick one from four).

Example 1:

Suppose Benoit is a lecturer giving a multiple choice test with 10 questions each with four options that are scored as either correct or incorrect. A passing score is 7 or above. If a student simply randomly guesses all of the questions, what is the probability they achieve a passing score?

This is a **Binomial** experiment with $n = 10$ and $p = 1/4$ (randomly pick one from four).

We are interested in finding $P(Pass) = P(X \geq 7)$.

Example 1:

Suppose Benoit is a lecturer giving a multiple choice test with 10 questions each with four options that are scored as either correct or incorrect. A passing score is 7 or above. If a student simply randomly guesses all of the questions, what is the probability they achieve a passing score?

This is a **Binomial** experiment with $n = 10$ and $p = 1/4$ (randomly pick one from four).

We are interested in finding $P(Pass) = P(X \geq 7)$.

$$P(X \geq 7) = P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$$

Example 1:

Suppose Benoit is a lecturer giving a multiple choice test with 10 questions each with four options that are scored as either correct or incorrect. A passing score is 7 or above. If a student simply randomly guesses all of the questions, what is the probability they achieve a passing score?

This is a **Binomial** experiment with $n = 10$ and $p = 1/4$ (randomly pick one from four).

We are interested in finding $P(Pass) = P(X \geq 7)$.

$$\begin{aligned} P(X \geq 7) &= P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) \\ &= f(7) + f(8) + f(9) + f(10) \end{aligned}$$

Example 1:

Suppose Benoit is a lecturer giving a multiple choice test with 10 questions each with four options that are scored as either correct or incorrect. A passing score is 7 or above. If a student simply randomly guesses all of the questions, what is the probability they achieve a passing score?

This is a **Binomial** experiment with $n = 10$ and $p = 1/4$ (randomly pick one from four).

$$\begin{aligned} P(X \geq 7) &= f(7) + f(8) + f(9) + f(10) \\ &= \binom{10}{7}\left(\frac{1}{4}\right)^7\left(\frac{3}{4}\right)^3 + \binom{10}{8}\left(\frac{1}{4}\right)^8\left(\frac{3}{4}\right)^2 \\ &+ \binom{10}{9}\left(\frac{1}{4}\right)^9\left(\frac{3}{4}\right)^1 + \binom{10}{10}\left(\frac{1}{4}\right)^{10}\left(\frac{3}{4}\right)^0 \\ &\approx 0.0031 + 0.0004 + 0.0000 + 0.0000 = 0.0035 \end{aligned}$$

Hence it is very unlikely someone just randomly guessing will pass.

Example 2:

Anh is interested in the number of patients who come to the clinic presenting symptoms of malaria per day. From past information the mean number of patients per day is 2.2. She wants to know the probability that less than 4 patients will present symptoms of malaria on a given day.

Example 2:

Anh is interested in the number of patients who come to the clinic presenting symptoms of malaria per day. From past information the mean number of patients per day is 2.2. She wants to know the probability that less than 4 patients will present symptoms of malaria on a given day.

This is a **Poisson** experiment with $\lambda = 2.2$. Notice that it is a count over time.

Example 2:

Anh is interested in the number of patients who come to the clinic presenting symptoms of malaria per day. From past information the mean number of patients per day is 2.2. She wants to know the probability that less than 4 patients will present symptoms of malaria on a given day.

This is a **Poisson** experiment with $\lambda = 2.2$. Notice that it is a count over time.

$$P(X < 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

Example 2:

Anh is interested in the number of patients who come to the clinic presenting symptoms of malaria per day. From past information the mean number of patients per day is 2.2. She wants to know the probability that less than 4 patients will present symptoms of malaria on a given day.

This is a **Poisson** experiment with $\lambda = 2.2$. Notice that it is a count over time.

$$\begin{aligned} P(X < 4) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= f(0) + f(1) + f(2) + f(3) \end{aligned}$$

Example 2:

Anh is interested in the number of patients who come to the clinic presenting symptoms of malaria per day. From past information the mean number of patients per day is 2.2. She wants to know the probability that less than 4 patients will present symptoms of malaria on a given day.

This is a **Poisson** experiment with $\lambda = 2.2$. Notice that it is a count over time.

$$\begin{aligned} P(X < 4) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= f(0) + f(1) + f(2) + f(3) \\ &= \frac{2.2^0 e^{-2.2}}{0!} + \frac{2.2^1 e^{-2.2}}{1!} + \frac{2.2^2 e^{-2.2}}{2!} + \frac{2.2^3 e^{-2.2}}{3!} \\ &\approx 0.11.8 + 0.2458 + 0.2681 + 0.1966 \end{aligned}$$

Example 2:

Anh is interested in the number of patients who come to the clinic presenting symptoms of malaria per day. From past information the mean number of patients per day is 2.2. She wants to know the probability that less than 4 patients will present symptoms of malaria on a given day.

This is a **Poisson** experiment with $\lambda = 2.2$. Notice that it is a count over time.

$$\begin{aligned}
P(X < 4) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\
&= f(0) + f(1) + f(2) + f(3) \\
&= \frac{2.2^0 e^{-2.2}}{0!} + \frac{2.2^1 e^{-2.2}}{1!} + \frac{2.2^2 e^{-2.2}}{2!} + \frac{2.2^3 e^{-2.2}}{3!} \\
&\approx 0.1108 + 0.2458 + 0.2681 + 0.1966 \\
&= 0.8213
\end{aligned}$$

Hence it is quite likely that there will be less than 4 patients will present symptoms of malaria on a given day.

Summary

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

- Probabilities can just add up.
- Be careful about inequalities.
- May need to use other properties to calculate probabilites.... next.