Lecture 10: Bernoulli and Geometric Random Variables

Chapter 3.3-3.5

Goals for Today

Define

- Bernoulli random variables
- Geometric random variables

Mathematical Definition of a Bernoulli Random Variable

A random variable X is a random process or variable with a numerical outcome.

Random variables are described in terms of their distribution.

Bernoulli Distribution

Say we have an experiment where we define each trial (or instance) to have two possible outcomes of interest. Examples

- Coin flips: heads vs tails
- Medical test (for a disease): positive vs negative
- Rolling a die and getting a 6 vs not getting a 6

In each case we can define the outcomes to be success vs failure. No moral judgement; just labels.

Bernoulli Distribution

Say we have trials where we have two outcomes: either a "success" or a "failure". Classic example: coin flips have $p = 0.5$ of heads, if we define heads as the success.

- probability p of a "success." Denote successes with a "1."
- probability $1 - p$ of a "failure." Denote failures with a "0."

Definition of a Bernoulli Random Variable

If X is a random variable that takes value

- 1 with probability of success p
- 0 with probability of failure $1 - p$

then X is a Bernoulli random variable with mean and standard deviation:

$$\begin{aligned} \mu &= p \\ \sigma &= \sqrt{p(1-p)} \end{aligned}$$

Intuition Behind σ

Sample Proportion

Say you repeat n instances of a Bernoulli random variable. You end up with a sample $x_1, \ldots, x_n$

The sample proportion $\widehat{p}$ (p-hat) is the sample mean of these observations. i.e.

$$\widehat{p} = \frac{\#\text{ of successes}}{\#\text{ of trials}} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Example of Bernoulli Distribution

- A success as rolling a 6.
  So $P(X = 1) = P(\text{success}) = p = \frac{1}{6}$.
- A failure as rolling anything else.
  So $P(X = 0) = P(\text{failure}) = 1 - p = \frac{5}{6}$.

Back to Lecture 3.1: Population vs Sample Values

	True Population Value	Sample Value
Mean	μ	$\overline{x}$
Variance	σ^2	s^2
Standard Deviation	σ	s
Proportion	p	$\widehat{p}$

The sample proportion $\widehat{p}$ is a specific kind of sample mean for Bernoulli random variables, which estimates p, a specific kind of population mean.

Scenario

Question: Say

- the San Francisco Giants have equal probability $p = 0.6$ of winning any game
- games are independent

It's the beginning of the season. What is the probability that they don't win their first game until the 5th game of the season?

For this to happen, there must be 4 loses in the first 4 games AND a win in the 5th game:

$$\begin{aligned} P(\text{1st W in 5th game}) &= P(\text{4 loses}) \times P(\text{win}) \\ &= (P(\text{loss}))^4 \times P(\text{win}) \\ &= (1-p)^4 \times p \\ &= 0.4^4 \times 0.6 = 0.01536. \end{aligned}$$

Geometric Random Variables

Geometric Distribution: If the probability of a success in any trial is p, the trials are independent, then the probability of finding the first success on the n^{th} trial is given by

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

Also

$$\begin{aligned} \mu &= \frac{1}{p} \\ \sigma^2 &= \frac{1-p}{p^2} \\ \sigma &= \frac{\sqrt{1-p}}{p} \end{aligned}$$

Intuition Behind μ

Think about μ: $\frac{1}{p}$ is the average number of trials we need until the first success.

So compare:

- Say $p = 0.5$. Then $\mu = \frac{1}{0.5} = 2$
- Say $p = 0.001$. Then $\mu = \frac{1}{0.001} = 1000$

In the first case, the probability of a success is lower, so we expect on average it will take more trials until the first success.