Statistical Inference for Estimation in Data Science

DTSA 5002 offered on Coursera

by the University of Colorado, Boulder

Instructor: J.N. Corcoran

Week One Lesson One

Discrete Distributions

There are several discrete distributions that are important in statistics and data science. So important, in fact, that they get names! Here are a few that we will need in this course.

The Bernoulli Distribution

Flip a, possibly unfair, coin that has some probability p of coming up "heads" and probability $1-p$ of coming up "tails", where p is some fixed number in the interval $[0,1]$.

Define

$$X = \begin{cases} 1 & , \text{ if "heads"} \\ 0 & , \text{ if "tails".} \end{cases}$$

Then X is a random variable that takes the value 1 with probability p and 0 with probability $1-p$.

More generally, think of an experiment you can perform that can result in one of two possible outcomes that you can label as "Success" or "Failure" and suppose that "Success" happens with some probability $0 \le p \le 1$.

Define

$$X = \begin{cases} 1 & , \text{ if the experiment produces a "success"} \\ 0 & , \text{ if the experiment produces a "failure".} \end{cases}$$

The probability mass function (pmf) is

$$f(x) = \begin{cases} p & , \; x=1 \\ 1-p & , \; x=0 \\ 0 & , \; \text{otherwise} \end{cases}$$

This pmf can also be written using indicator notation as

$$f(x) = p^x (1-p)^{x-1} I_{\{0,1\}}(x).$$

This (0/1) type of random variable comes up so often that it gets a name!

X is called a **Bernoulli random variable with parameter** p. Alternatively, we say that X **has a Bernoulli distribution with parameter** p.

We write

$$X \sim Bernoulli(p).$$

Note that most things that we can measure can be thought of as having a Bernoulli distribution when thought of in a certain way. For example, suppose you are at a highway truck weighing station, recording measurements of vehicle weights. Ultimately, you are only interested in whether the trucks stopping at the station weight greater than or equal to 60,000 lbs or less that 60,000 lbs. You could record a 1 or 0 depending on which category each truck falls into.

The Geometric Distribution

Consider an experiment consisting of a sequence of independent trials of something where each trial can have only two possible outcomes, usually labeled as "success" and "failure". We will denote success and failure by S and F, respectively.

Let p be the probability of "success" on any one trial. Here, p must live in the interval $[0, 1]$.

Let

$$X = \text{\# trials until the first success.}$$

Then the possible values for X are $1, 2, 3, \ldots$.

For example, in an idealized world, you might imagine someone shooting baskets in basketball and X as the number of tries they have to take until they make a basket. (In real life though, the assumptions that the success probability is constant for all trials and the trials are independent would probably be violated as the person could be learning from each failed attempt and making adjustments to get better or maybe their arms are tired and they are getting worse!)

Note that

$$P(X = 1) = P(\, S \text{ on first trial}\,) = p.$$

Further, note that

$$P(X = 2) = P(\, F \text{ on 1st trial AND } S \text{ on 2nd trial}\,).$$

By independence of the trials, this is

$$
\begin{aligned}
P(X=2) &\stackrel{indep}{=} P(\,F \text{ on 1st trial}\,) \cdot P(\,S \text{ on second trial}\,) \\
&= (1-p) \cdot p.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
P(X=3) &= P(\,\text{F on 1st trial AND F on 2nd trial AND S on 3rd trial}\,) \\
&\stackrel{indep}{=} P(\,\text{F on 1st trial}\,) \cdot P(\,\text{F on 2nd trial}\,) \cdot P(\,\text{S on 3rd trial}\,) \\
&= (1-p) \cdot (1-p) \cdot p = (1-p)^2 \cdot p.
\end{aligned}
$$

A pattern is forming. Indeed, we have

$$P(X=x) = (1-p)^{x-1} \cdot p$$

for $x = 1, 2, 3, \ldots$.

Since $P(X=x) = 0$ for x not in $\{1, 2, 3, \ldots\}$, we conclude that the pmf is

$$f(x) = \begin{cases} (1-p)^{x-1} \cdot p & , \quad x = 1, 2, 3 \ldots \\ 0 & , \quad \text{otherwise.} \end{cases}$$

This can be written in indicator notation as

$$f(x) = (1-p)^{x-1} p\, I_{\{1,2,3,\ldots\}}(x).$$

X is said to have a **geometric distribution** with parameter p. We write

$$X \sim geom(p).$$

Note: The geometric random variable can also be defined as

$$X = \text{\# failures before the first success.}$$

In this case, X can take on the value 0 and it is easy to verify that the pmf changes slightly to

$$f(x) = \begin{cases} (1-p)^{x} \cdot p & , \quad x = 0, 1, 2 \ldots \\ 0 & , \quad \text{otherwise} \end{cases}$$

which may also be written as

$$f(x) = (1-p)^{x} \cdot p\, I_{\{0,1,2,\ldots\}}(x).$$

The Binomial Distribution

Consider a sequence of n independent trials of an experiment where each trial can result in either "successes" (S) or "failure" (F). Suppose that the probability of success remains the same from trial to trial. Call it p where $0 \leq p \leq 1$.

Let

$$X = \text{\# of successes in } n \text{ trials.}$$

While it looks similar, this is different from the geometric random variable. There, we continued trials until the first success. Here, we will have a n (a fixed number) of trials and count up all the successes.

In this case, X is said to have a **binomial distribution** with parameters n and p.

We write

$$X \sim bin(n,p).$$

As the number of successes in n trials, X can take on values in $\{0,1,2,\ldots,n\}$.

The probability mass function (pmf) is

$$f(x) = P(X = x) = P(SSFSF\ldots F \text{ or } SFSFS\ldots S \text{ or}\ldots)$$

where each listed configuration of outcomes includes exactly x S's and exactly $n-x$ F's. Since the outcomes are disjoint, we get

$$f(x) = P(X = x) = P(SSFSF\ldots F) + P(SFSFS\ldots S) + \cdots \tag{0.1}$$

Since the trials are independent,

$$P(SSFSF\ldots F) = p\cdot p\cdot(1-p)\cdot s\cdot(1-p)\cdots(1-p) = p^x(1-p)^{n-x}.$$

In fact, every term in (0.1) gives that same probability since every term has the same number of S's and F's! So,

$$f(x) = P(X = x) = c\cdot p^x(1-p)^{n-x}$$

where c is the number of terms in (0.1).

How many ways are there to write down sequences of n S's and F's with exactly x S's? We need to choose x slots out of n in which to put the S's. There are

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

different ways to do this. Thus,

$$c = \binom{n}{x}.$$

So, we have seen that $X \sim bin(n,p)$ means that X has pmf

$$f(x) = P(X = x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x} \, I_{\{0,1,\ldots,n\}}(x) \tag{0.2}$$

The Poisson Distribution

When counting the number of occurrences of some event, it is often the case that you will want to use the **Poisson distribution**.

Let X be a random variable taking on values in $\{0,1,2,\ldots\}$. X is said to have a **Poisson distribution** if the probability mass function for X is

$$f(x) = P(X = x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & , \quad x = 0,1,2,\ldots \\ 0 & , \quad \text{otherwise} \end{cases}$$

which may also be written as

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

Here, $\lambda > 0$ is a parameter for the distribution.

We write

$$X \sim Poisson(\lambda).$$

While this pmf might appear to be highly structured, it really is the epitome of randomness. Imagine taking a 20 acre plot of land and dividing it into 1 square foot sections. (There are 871,200 sections!) Suppose you were able to scatter 5 trillion grass seeds on this land in a completely random way that does not favor one section over another. One can show that the number of seeds that fall into any one section follows a Poisson distribution with some parameter λ. More specifically, one can show that the Poisson distribution is a limiting case of the binomial distribution when n gets really large and p get really small. "Success" here is the event that any given seed falls into one particular section. We then want to count the number of successes in 5 trillion trials.

In general, the Poisson distribution is often used to describe the distribution of rare events in a large population.