The Bernoulli, Binomial, and Poisson Distributions

There are a few standard distributions for modeling random variables. In this lecture, we will consider three common ones for discrete random variables. Though it is rare for a population to follow a function exactly, they can be useful in understanding key properties.

The Bernoulli Distribution

The simplest experiment we can think of is one that has two possible outcomes. We can label these "success" and "failure," with the probability of success being p and the probability of failure $1 - p$. Each iteration in such an experiment is called a **Bernoulli trial** with success probability p.

A Bernoulli random variable X is defined to be $X = 1$ in the case of a success, and $X = 0$ otherwise. The probability mass function is

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

In this case, X is said to have the **Bernoulli distribution** with parameter p, which is sometimes written $X \sim \text{Bernoulli}(p)$.

Mean and Variance

It is straightforward to find the mean μ_X and variance σ_X^2 of a Bernoulli random variable:

$$\begin{aligned} \mu_X &= 0 \cdot (1-p) + 1 \cdot p = p \\ \sigma_X^2 &= 0^2 \cdot (1-p) + 1^2 \cdot p - p^2 = p(1-p) \end{aligned}$$

As an example, suppose 20% of the components manufactured by a certain process are defective. With $X = 1$ if a component is defective and $X = 0$ otherwise, $X \sim \text{Bernoulli}(0.2)$. Then $\mu_X = 0.2$ and $\sigma_X^2 = 0.2 \cdot 0.8 = 0.16$.

The Binomial Distribution

In many cases, an experiment boils down to several independent Bernoulli trials, *e.g.*, flipping a coin several times or sampling several components from a defective lot. The number of successes in these experiments (*e.g.*, number of heads or number of defective components) is then a random variable, which is said to have a **binomial distribution**.

Formally, suppose a total of n Bernoulli trials are conducted such that (i) the trials are independent and (ii) each trial has the same success probability p. If X counts the number of successes in the n trials, then X has the binomial distribution with parameters n and p, denoted $X \sim \text{Bin}(n, p)$.

When will this independence assumption hold? Note that when we are sampling from a finite population (*e.g.*, components in a lot, people in the US, resistors in a box), the size of the population will get smaller each time; back in Chapter 1, we talked about how this means the samples are not strictly independent. But if the sample is much smaller than the population, independence is still a reasonable assumption.

In particular, assume that a finite population contains items of two types, successes and failures, and that a simple random sample is drawn from the population. If the sample size is no more than 5% of the population, the binomial distribution may be used to model the number of successes.

Probability Mass Function

The PMF for $X \sim \text{Bin}(n, p)$ is

$$p_X(x) = p(X = x) = \begin{cases} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} & x = 0, 1, ..., n \\ 0 & \text{otherwise} \end{cases}$$

Why? If the probability of success is p, then for any *particular sequence* of outcomes, the chance that x of them are successes (and $n - x$ are not) is $p^x \cdot (1-p)^{n-x}$. But there's several possible ways of choosing x successes from n outcomes, ${}^nC_x$ to be exact. So we multiply by this.

Figure 1 below plots the probability histograms for $\text{Bin}(10, 0.5)$, $\text{Bin}(10, 0.1)$, and $\text{Bin}(10, 0.9)$ side-by-side. Intuitively, the density will shift based on the value of p, because a higher p means more chance for success in each trial and a lower p means less chance.

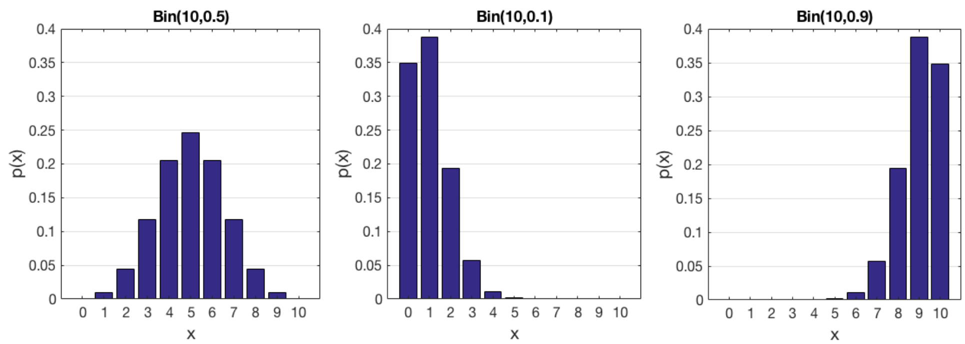

Figure 1: Probability mass functions for the binomial distribution with $n = 10$, and various values of p.

As an example, suppose a six-sided die is rolled ten times. We are interested in the probability that we will roll a 3 or a 4. Letting X represent the number of times the die comes up as this, $p = 2/6 = 1/3$ and $n = 10$, so $X \sim \text{Bin}(10, 1/3)$.

What is the probability that we roll a 3 or 4 exactly twice?

$$p(X = 2) = {}^{10}C_2 \cdot (1/3)^2(2/3)^8 = 0.195$$

What is the probability that we roll it more than twice?

$$p(X > 2) = 1 - \sum_{i=0}^{2} p(X = i) = 0.70$$

What is the probability that we roll it less than 4 times?

$$p(X < 4) = \sum_{i=0}^{3} p(X = i) = 0.56$$

Mean and Variance

How can we compute the mean and variance of X? In theory, we could apply the definitions from scratch, but the math would get messy. It's easier instead to realize that a binomial random variable is just a sum of Bernoulli random variables, *i.e.*, if the Y_i are Bernoulli, then $X_i = \sum_i Y_i$ is binomial. And when random variables are *independent*, their means and variances are additive, *i.e.*,

$$\mu_X = \sum_i \mu_{Y_i} \qquad \sigma_X^2 = \sum_i \sigma_{Y_i}^2$$

Since successive Bernoulli trials are independent, we just multiply by n to get the mean and variance of X, *i.e.*, for $X \sim \mathrm{Bin}(n, p)$,

$$\mu_X = np \qquad \sigma_X^2 = np(1-p)$$

Estimating the Success Probability

In many cases, we won't actually know what p for the population is. To estimate it, we may decide to conduct n independent trials and count the number of successes, which gives us the sample proportion $\hat{p}$:

$$\hat{p} = \frac{\text{number of successes}}{\text{number of trials}} = \frac{X}{n}$$

It is common to use the "hat" in this way to indicate an estimate (*i.e.*, from the sample) of a true value (*i.e.*, from the population).

When using $\hat{p}$ to estimate p, it is desirable to know how accurate we can expect the estimate to be. If the sampling procedure is a simple random sample, then we can show that (i) $\hat{p}$ is unbiased and (ii) the uncertainty in $\hat{p}$ is

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

In practice, when computing $\sigma_{\hat{p}}$, we use p in place of $\hat{p}$, since p will be unknown (otherwise we wouldn't be trying to estimate it in the first place).

As an example, let's say we want to determine how many screws from a factory are out of spec. We take a sample of 50 screws and find 5 of them to be. With X as the random variable for the number of screws that will be out of spec, our estimate of p is then

$$\hat{p} = 5/50 = 0.2$$

And the uncertainty is

$$\sigma_{\hat{p}} = \sqrt{\frac{0.2(1-0.2)}{50}} = 0.0566$$

Approximately how many more samples would we need to draw to reduce our uncertainty to 0.04? We solve for n:

$$\sqrt{\frac{0.2(1-0.2)}{n}} = 0.04$$

which yields $n = 100$. So, we need $100 - 50 - 50$ more.

The Poisson Distribution

The Poisson distribution occurs quite frequently in nature and scientific work.

Formally, if a discrete random variable X has a **Poisson distribution**, we write $X \sim \text{Poisson}(\lambda)$, and the probability mass function is

$$p(x) = \begin{cases} e^{-\lambda}\frac{\lambda^x}{x!} & \text{if } x \text{ is a non-negative integer} \\ 0 & \text{otherwise} \end{cases}$$

λ is called the **rate parameter** for the distribution. It will always be a positive constant. In fact, the Poisson probability mass function is very close to the binomial probability mass function when n is large, p is small, and $\lambda = np$.

Mean and Variance

If $X \sim \text{Poisson}(\lambda)$, then the mean and variance of X are given by

$$\mu_X = \lambda \qquad \sigma_X^2 = \lambda$$

You are encouraged to show this using the definitions of mean and variance (it is easier than in the binomial case). Intuitively, though, we can use the fact that $\mu_X = np$ and $\sigma_X^2 = np(1-p)$ for a binomial random variable; λ can be taken as np, and when p is very small, $1 - p \approx 1$.

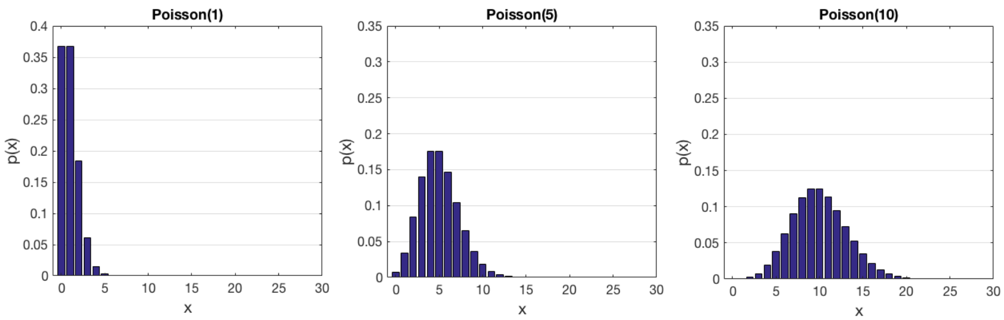

Figure 2: Probability mass functions for the Poisson distribution at various values of λ.

Let's take an example from the book: Someone bakes chocolate chip cookies in batches of 100, and puts 300 chips into the dough. We want the probability mass function of the number of chips in a randomly chosen cookie. Modeling as a Poisson distribution (which is valid because this is a type of particle suspension problem, see the book), we can take the rate parameter to be $\lambda = 300/100 = 3$. So $X \sim \text{Poisson}(3)$. The probability that a cookie contains less than 2 chips is

$$P(X < 2) = P(X = 0) + P(X = 1) = e^{-3}3^0/0! + e^{-3}3^1/1! = 0.1991$$

Estimating the Rate

We often need to estimate the rate parameter λ. If λ is to represent the mean number of events that occur in one unit of time or space, we count the number of occurrences X in t units, and take the average:

$$\hat{\lambda} = X/t$$

As with the binomial estimator for $\hat{p}$, this one is also unbiased. The uncertainty can be shown to be

$$\sigma_{\hat{\lambda}} = \sqrt{\lambda/t}$$

As an example, suppose we want to estimate the distribution of the number of hits on a website per hour. We take measurements over the course of one day (24 hrs), and find that there were roughly 1000 hits. If we believe the random variable of hits per second is Poisson, we can estimate

$$\hat{\lambda} = 1000/24 \approx 42$$

And therefore our Poisson PDF would be

$$p(x) = e^{-42}\frac{42^x}{x!}$$

So, the probability of having *e.g.*, more than 30 hits in one hour is

$$p(X > 10) = 1 - p(X \leq 10) = 1 - \sum_{i=0}^{10} e^{-42}\frac{42^i}{i!} = 0.967$$

The uncertainty in this estimate of $\hat{\lambda}$ is

$$\sigma_{\hat{\lambda}} = \sqrt{42/24} = 1.33$$