3.3 Bernoulli and Binomial

Unit 3: Discrete Random Variables

Adapted from Blitzstein-Hwang Chapter 3.

Some distributions are so ubiquitous in probability and statistics that they have their own names. We will introduce these *named distributions* throughout the course, starting with a very simple but useful case: an r.v. that can take on only two possible values, 0 and 1.

> **Definition 3.3.1 (Bernoulli distribution).**
>
> An r.v. X is said to have the *Bernoulli distribution* with parameter p if $P(X=1)=p$ and $P(X=0)=1-p$, where $0<p<1$. We write this as $X \sim \text{Bern}(p)$. The symbol $\sim$ is read "is distributed as".

Any r.v. whose possible values are 0 and 1 has a Bern(p) distribution, with p the probability of the r.v. equaling 1. This number p in Bern(p) is called the *parameter* of the distribution; it determines which specific Bernoulli distribution we have. Thus there is not just one Bernoulli distribution, but rather a *family* of Bernoulli distributions, indexed by p. For example, if $X \sim \text{Bern}(1/3)$, it would be correct but incomplete to say "X is Bernoulli"; to fully specify the distribution of X, we should both say its name (Bernoulli) and its parameter value ($1/3$), which is the point of the notation $X \sim \text{Bern}(1/3)$.

Any event has a Bernoulli r.v. that is naturally associated with it, equal to 1 if the event happens and 0 otherwise. This is called the *indicator random variable* of the event; we will see that such r.v.s are extremely useful.

> **Definition 3.3.2 (Indicator random variable).**
>
> The *indicator random variable* of an event A is the r.v. which equals 1 if A occurs and 0 otherwise. We will denote the indicator r.v. of A by I_A or $I(A)$. Note that $I_A \sim \text{Bern}(p)$ with $p=P(A)$.

We often imagine Bernoulli r.v.s using coin tosses, but this is just convenient language for discussing the following general story.

Story 3.3.3 (Bernoulli trial).

An experiment that can result in either a "success" or a "failure" (but not both) is called a *Bernoulli trial*. A Bernoulli random variable can be thought of as the *indicator of success* in a Bernoulli trial: it equals 1 if success occurs and 0 if failure occurs in the trial.

Because of this story, the parameter p is often called the *success probability* of the $\mathrm{Bern}(p)$ distribution. Once we start thinking about Bernoulli trials, it's hard not to start thinking about what happens when we have more than one Bernoulli trial.

Story 3.3.4 (Binomial distribution).

Suppose that n *independent* Bernoulli trials are performed, each with the same success probability p. Let X be the number of successes. The distribution of X is called the *Binomial distribution* with parameters n and p. We write $X \sim \mathrm{Bin}(n, p)$ to mean that X has the Binomial distribution with parameters n and p, where n is a positive integer and $0 < p < 1$.

Notice that we define the Binomial distribution not by its PMF, but by a *story* about the type of experiment that could give rise to a random variable with a Binomial distribution. The most famous distributions in statistics all have stories which explain why they are so often used as models for data, or as the building blocks for more complicated distributions.

Thinking about the named distributions first and foremost in terms of their stories has many benefits. It facilitates pattern recognition, allowing us to see when two problems are essentially identical in structure; it often leads to cleaner solutions that avoid PMF calculations altogether; and it helps us understand how the named distributions are connected to one another. Here it is clear that $\mathrm{Bern}(p)$ is the same distribution as $\mathrm{Bin}(1, p)$: the Bernoulli is a special case of the Binomial.

Using the story definition of the Binomial, let's find its PMF.

Theorem 3.3.5 (Binomial PMF).

If $X \sim \mathrm{Bin}(n, p)$, then the PMF of X is

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, \ldots, n$ (and $P(X = k) = 0$ otherwise).

Proof

An experiment consisting of n independent Bernoulli trials produces a sequence of successes and failures. The probability of any specific sequence of k successes and $n-k$ failures is $p^k(1-p)^{n-k}$. There are $\binom{n}{k}$ such sequences, since we just need to select where the successes are. Therefore, letting X be the number of successes,

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, \ldots, n$, and $P(X = k) = 0$ otherwise. This is a valid PMF because it is nonnegative and it sums to 1 by the binomial theorem. Figure 3.3.6 shows plots of the Binomial PMF for various values of n and p. Note that the PMF of the $\text{Bin}(10, 1/2)$ distribution is symmetric about 5, but when the success probability is not $1/2$, the PMF is *skewed*. For a fixed number of trials n, X tends to be larger when the success probability is high and lower when the success probability is low, as we would expect from the story of the Binomial distribution. Also recall that in any PMF plot, the sum of the heights of the vertical bars must be 1.

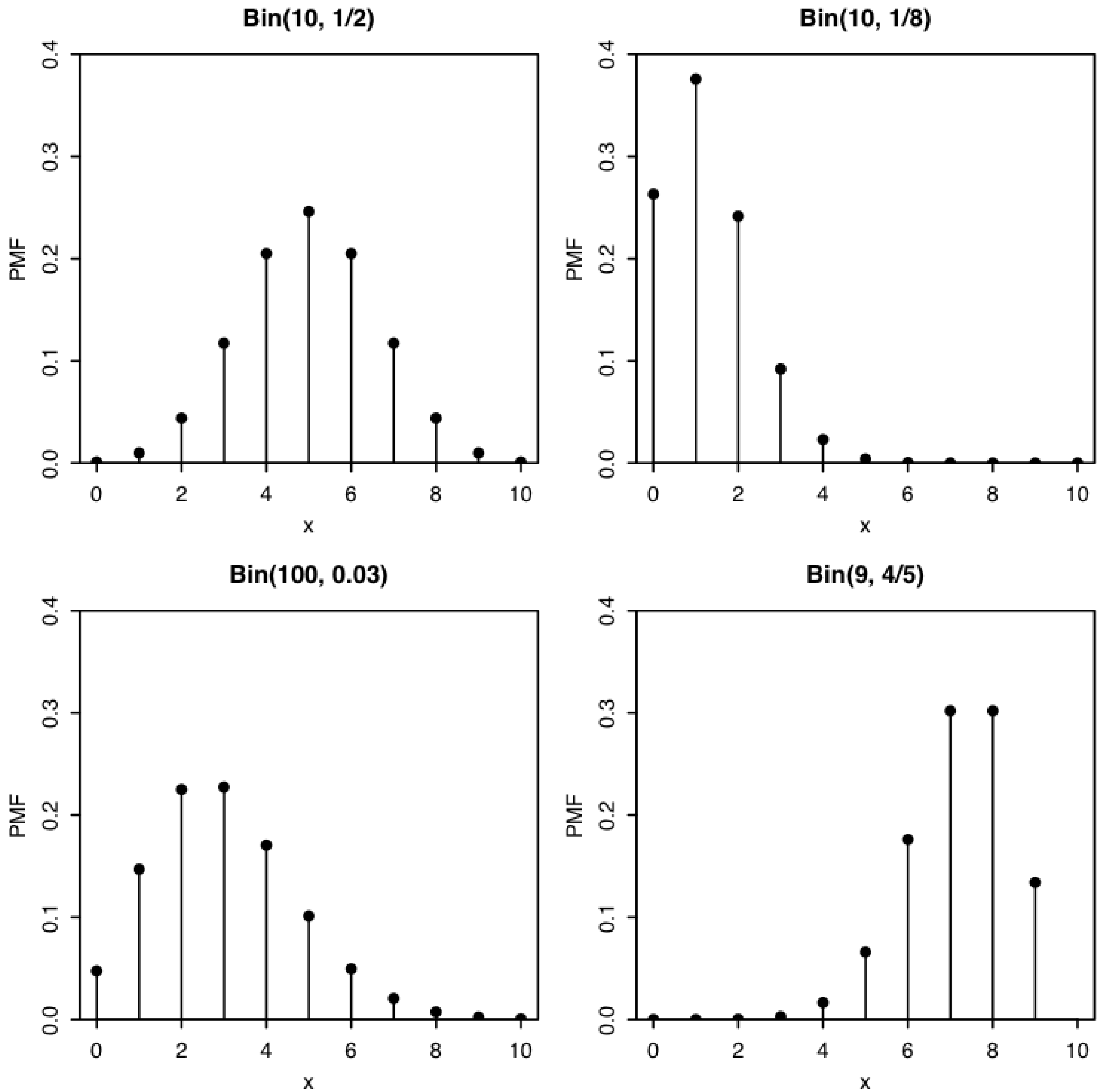

Figure 3.3.6: Some Binomial PMFs. In the lower left, we plot the $\mathrm{Bin}(100, 0.03)$ PMF between 0 and 10 only, as the probability of more than 10 successes is close to 0.

View Larger Image
Image Description

We've used Story 3.3.4 to find the $\mathrm{Bin}(n, p)$ PMF. The story also gives us a straightforward proof of the fact that if X is Binomial, then $n - X$ is also Binomial.

THEOREM 3.3.7.

Let $X \sim \mathrm{Bin}(n, p)$, and $q = 1 - p$ (we often use q to denote the failure probability of a Bernoulli trial). Then $n - X \sim \mathrm{Bin}(n, q)$.

Proof

Using the story of the Binomial, interpret X as the number of successes in n independent Bernoulli trials. Then $n - X$ is the number of failures in those trials. Interchanging the roles of success and failure, we have $n - X \sim \mathrm{Bin}(n, q)$. Alternatively, we can check that $n - X$ has the $\mathrm{Bin}(n, q)$ PMF. Let $Y = n - X$. The PMF of Y is

$$P(Y = k) = P(X = n - k) = \binom{n}{n-k} p^{n-k} q^k = \binom{n}{k} q^k p^{n-k},$$

for $k = 0, 1, \ldots, n$.

Corollary 3.3.8.

Let $X \sim \mathrm{Bin}(n, p)$ with $p = 1/2$ and n even. Then the distribution of X is symmetric about $n/2$, in the sense that $P(X = n/2 + j) = P(X = n/2 - j)$ for all nonnegative integers j.

Proof

By Theorem 3.3.7, $n - X$ is also $\mathrm{Bin}(n, 1/2)$, so

$$P(X = k) = P(n - X = k) = P(X = n - k)$$

for all nonnegative integers k. Letting $k = n/2 + j$, the desired result follows. This explains why the $\mathrm{Bin}(10, 1/2)$ PMF is symmetric about 5 in Figure 3.3.6.