§3.3-3.4 Bernoulli and Binomial distributions

Def A random variable X has the Bernoulli distribution with parameter $p \in (0,1)$ if its range is $\{0,1\}$ and

$$P(X = k) = \begin{cases} p & \text{if } k=1 \\ 1-p & \text{if } k=0 \end{cases}.$$

Shorthand: we write $X \sim \text{Ber}(p)$

"is distributed as"

Remark We often use Bernoulli random variables for modeling a single trial in a "success or failure" experiment.

Def Two random variables are independent if $P(X=x, Y=y) = P(X=x)P(Y=y)$ for all x, y in their respective ranges.

Example Consider the experiment of tossing a biased coin (with heads probability $p \neq \frac{1}{2}$) 15 times.

Express the event of getting exactly

① 0 heads ② 1 heads ③ 2 heads

in terms of Bernoulli random variables and find its probability.

Let $X_k = \begin{cases} 1 & \text{if toss } k \text{ is heads} \\ 0 & \text{if toss } k \text{ is tails.} \end{cases}$

Then $X_1, \ldots, X_{15} \sim \mathrm{Ber}(p)$ are independent and identically distributed (i.i.d.) random variables.

① Our event is $\{X_1 + \cdots + X_{15} = 0\}$ and

$$P(X_1 + \cdots + X_{15} = 0) = P(X_1 = 0, \ldots, X_{15} = 0)$$
$$= (1-p)^{15}$$

② $\{X_1 + \cdots + X_{15} = 1\}$ and

$$\begin{aligned} P(X_1 + \cdots + X_{15} = 1) &= P(X_1 = 1, X_2 = 0, \ldots, X_{15} = 0) \\ &\quad + P(X_1 = 0, X_2 = 1, X_3 = 0, \ldots, X_{15} = 0) \\ &\quad \cdots + P(X_1 = 0, \ldots, X_{14} = 0, X_{15} = 1) \\ &= 15\, p (1-p)^{14} \end{aligned}$$

③ $P(X_1 + \cdots + X_{15} = 2) = \binom{15}{2} p^2 (1-p)^{13}$

Def A random variable X has the binomial distribution with parameters n and p if its range is $\{0, 1, 2, \ldots, n\}$ and

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, \ldots, n.$$

Shorthand: $X \sim \mathrm{Bin}(n, p)$.

Remarks (1) if $X_1, \dots, X_n$ are i.i.d. Ber(p), then

$$X = X_1 + \cdots + X_n \text{ is } \mathrm{Bin}(n,p) \text{ distributed.}$$

(2) $X \sim \mathrm{Bin}(n,p)$ counts the number of successes in n independent trials of a success / failure experiment where the success probability is p.

R commands if $X \sim \mathrm{Bin}(n,p)$, then

- $P(X = k)$ can be computed with dbinom(k,n,p)
- $P(X \le k)$ with pbinom(k,n,p)
- we can get an i.i.d. sample of size k with rbinom(k,n,p)

Example Consider rolling a 4-sided die 13 times. Let X count the number of times 2 comes up.

Find the probability of getting

(1) exactly 6 2's

(2) at least 6 2's.

Notice $X \sim \mathrm{Bin}(13, 1/4)$. Therefore

(1) $P(X = 6) = \binom{13}{6} (1/4)^6 (3/4)^7$

can be compute with R command dbinom$(6, 13, 1/4)$

(2) $P(X \ge 6) = P(X = 6) + P(X = 7) + \cdots + P(X = 13)$

$= 1 - P(X < 6)$

$= 1 - P(X \le 5)$

$= 1 - (P(X = 0) + P(X = 1) + \cdots + P(X = 5))$

can be computed with R command

$1 - $ pbinom$(5, 13, 1/4)$

Problem 1. Consider an urn which contains 12 red, 2 green, and 3 blue balls. We draw from the urn 14 times, sampling with replacement, and let X count the number of times we drew a green ball.

a. The random variable X is binomially distributed. What are the parameters n and p?

b. Express each of the following events in terms of X and compute its probability

1. Exactly 1 draw is green
2. At least 1 draw is green
3. Exactly 4 draws are green
4. At least 4 draws are green
5. At least 3 but no more than 8 draws are green

c. If we change our experiment so that sampling is done without replacement, is it still the case that X is binomially distributed? Why or why not?

(a) $n = 14$, $p = \frac{2}{17}$

(b) ① $P(X=1)$

② $P(X \geq 1)$

③ $P(X=4)$

④ $P(X \geq 4)$

⑤ $P(3 \leq X \leq 8)$

Problem 1

```{r}
dbinom(1, 14, 2/17)
1-dbinom(0, 14, 2/17)
dbinom(4, 14, 2/17)
1-pbinom(3, 14, 2/17)
pbinom(8, 14, 2/17) - pbinom(2, 14, 2/17)
```

```
[1] 0.323638
[1] 0.8266225
[1] 0.05485065
[1] 0.07290558
[1] 0.2224933
```

(c) No, trials no longer independent and no longer have constant success probability

Problem 2. Suppose we have an 8×8 grid of squares. For each square in the grid, we roll a die and color the square black if a prime number is rolled and white if a non-prime is rolled. Let X be the number of black squares in the grid after completing this coloring process.

a. What are the independent trials in this experiment? What corresponds to a *successful* trial?

b. The random variable X has binomial distribution. What are the parameters n and p?

c. Express each of the following events in terms of X and compute its probability

1. Exactly 31 squares are colored black
2. At least 31 squares is colored black
3. Exactly 37 squares are colored black
4. At most 57 squares are colored black

(a) a square is colored black

(b) $n = 64$, $p = \frac{1}{2}$

(c) ① $P(X = 31)$

② $P(X \geq 31)$

③ $P(X = 37)$

④ $P(X \leq 57)$

Problem 2

```
```{r}
knitr::opts_chunk$set(echo = TRUE, options(digits = 15))
dbinom(31, 64, 1/2)
1-pbinom(30, 64, 1/2)
dbinom(37, 64, 1/2)
pbinom(57, 64, 1/2)
```
```

```
[1] 0.0963362460586346
[1] 0.646009622932617
[1] 0.0458962825684751
[1] 0.999999999995486
```

Problem 3. Consider the following checklist to determine whether a random variable X has the binomial distribution.

- Does the experiment involve a predetermined number of trials?
- Does each trial result in two possible outcomes, success or failure?
- Is the success probability the same for each trial?
- Is each trial independent?

Use this checklist to identify whether or not a random variable X has a binomial distribution. If it does, give n and p and explain any assumptions you're making; if not, explain why not.

a. We make 100 tosses of a coin with heads probability 1/3 and let X count the number of tails.

b. Each day Amy goes out for lunch, there is a 25% chance she will choose pizza. Let X be the number of times she chose pizza in the last 10 days.

c. Brenda plays basketball, and there is a 60% she makes a free throw. Let X be the number of successful baskets she makes in a game.

d. A bowl contains 100 red candies and 150 blue candies. Carl reaches in and takes out a sample of 10 candies. Let X be the number of red candies in his sample.

e. Evan is reading a 600-page book. On even-numbered pages, there is a 1% chance of a typo. On odd-numbered pages, there is a 2% chance of a typo. Let X be the number of typos in the book.

(a) yes, $n = 100$, $p = 2/3$

(b) yes, $n = 10$, $p = 0.25$ assuming days are independent

(c) no, no fixed number of trials

(d) no, no fixed success prob. (unless sampling with replacement)

(e) no, no fixed success prob.