Chapter 14, 15, 16, 19, 20: Some Important Discrete RVs

Meike Niederhausen and Nicky Wakim

2023-10-25

Table of contents

Learning Objectives

1. Distinguish between Bernoulli, Binomial, Geometric, Hypergeometric, and Discrete Uniform distributions when reading a story.
2. Identify the variable and the parameters in a story, and state what the variable and parameters mean.
3. Use the formulas for the pmf/CDF, expected value, and variance to answer questions and find probabilities.

Chapter 14: Bernoulli RVs

Properties of Bernoulli RVs

- **Scenario:** One trial, with outcome success or failure
- Shorthand: $X \sim \text{Bernoulli}(p)$

has distribution

"\sim"
LaTeX

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

$$p_X(x) = P(X = x) = \underline{p}^x (1 - p)^{1-x} \text{ for } x = 0, 1$$ pmf

$$E(X) = p$$ → class notes

$$\text{Var}(X) = pq = p(1 - p)$$ → in HW

Bernoulli Example 1

Example 1

- We roll a fair 6-sided die.
- We get \$1 if we roll a 5, and nothing otherwise.
- Let X be how much money we get.
- Find the mean and variance of X.

Let X = money we get

$$X = \begin{cases} 1 & \text{if roll a 5 w/ } p = \frac{1}{6} \\ 0 & \text{else} \quad \text{w/ } 1-p = 5/6 \end{cases}$$

We can say $X \sim \text{Bernoulli}(p = \frac{1}{6})$

$$E(X) = p = \frac{1}{6}$$

$$\text{var}(X) = p(1-p) = \frac{1}{6}\left(\frac{5}{6}\right) = \frac{5}{36}$$

Chapter 15: Binomial RVs

Properties of Binomial RVs

- **Scenario:** There are n independent trials, each resulting in a success or failure, with constant probability, p, in each trial. We are counting the number of successes (or failures).
- Shorthand: $X \sim \text{Binomial}(n, p)$

$$X = \text{Number of successes of } n \text{ independent trials}$$

$\binom{1}{x} = 1$ (if $n=1$ to make Bernoulli)

$$p_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, 2, \ldots, n$$

$$E(X) = np \longrightarrow np \quad \text{Chapter 12}$$

$$Var(X) = npq = np(1-p)$$

$q = 1 - p$

Bernoulli to Binomial

Example 2

- Suppose we roll a fair 6-sided die 50 times.
- We get \$1 every time we roll a 5, and nothing otherwise.
- Let X be how much money we get on the 50 rolls.
- Find the mean and variance of X.

Let X = money after 50 rolls

$n = 50, \quad p = \frac{1}{6}$

$X = \sum_{i=1}^{50} X_i$ if X_i is our bernoulli from ex 1

$E(X) = np = 50\left(\frac{1}{6}\right) = \frac{50}{6}$

Bernoulli varianc

$Var(X) = np(1-p) = 50\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)$

$= 50\left[\frac{5}{36}\right]$

$= \frac{250}{36}$

$E(Y_i) = \sum_{i=1}^{2} x_i P_X(x)$

$= 0\left(\frac{5}{6}\right) + 2\left(\frac{1}{6}\right) = 2/6$

bernoulli

$Y_i = \begin{cases} 2 & \text{if roll a 5} \\ 0 & \text{else} \end{cases}$

$Y_i = 2X_i$

2 or 0   1 or 0

$E(Y_i) = E(2X_i)$

$2E(X_i)$

Chapter 16: Geometric RVs

Geometric RVs

★ can have $x \to \infty$

- **Scenario:** There are repeated independent trials, each resulting in a success or failure, with constant probability of success for each trial. We are counting the number of trials until the first success.
- Shorthand: $X \sim \text{Geo}(p)$ or $X \sim \text{Geometric}(p)$ or $X \sim G(p)$

→ #fail = #trials − 1 success

	★ X = **Number of trials needed for first success** (count x *includes* the success)	X = **Number of failures before first success** (count x *does not include* the success)
pmf	$p_X(x) = P(X = x) = (1-p)^{x-1}p$ for $x = 1, 2, 3, \ldots$	$p_X(x) = P(X = x) = (1-p)^{x}p$ for $x = 0, 1, 2, \ldots$
CDF	$F_X(x) = P(X \le x) = 1-(1-p)^{x}$ → $1-q^x$ for $x = 1, 2, 3, \ldots$	$F_X(x) = P(X \le x) = 1-(1-p)^{x+1}$ for $x = 0, 1, 2, \ldots$
	$E(X) = \cancel{\frac{1-p}{p}}\ \frac{1}{p}$, $(q = 1-p)$ $Var(X) = \frac{1-p}{p^2}$	$E(X) = \frac{1-p}{p} \rightarrow \frac{1}{p} - \frac{p}{p} = \underbrace{\frac{1}{p}}_{\text{Exp val \# trials}} - 1$ $Var(X) = \frac{1-p}{p^2}$

$$\sum_{x=1}^{\infty} x(1-p)^{x-1}p$$

Bullseye (1/6)

Example 3

We throw darts at a dartboard until we hit the bullseye. Assume throws are independent and the probability of hitting the bullseye is 0.01 for each throw.

1. What is the pmf for the number of throws needed to hit the bullseye?
2. What are the mean and variance for the number of throws needed to hit the bullseye?
3. Find the probability that our first bullseye:
   - is on the fourth try
   - is on one of the first four tries
   - is after the fifth try
   - is on one of the first fifty tries
   - is after the 50^{th} try, given that it did not happen on the first 20 tries

~~4. Find the expected number of misses until we hit the bullseye.~~

Bullseye (2/6)

Example 3

We throw darts at a dartboard until we hit the bullseye. Assume throws are independent and the probability of hitting the bullseye is 0.01 for each throw.

1. What is the pmf for the number of throws needed to hit the bullseye?

Let X = # trials until 1st success (bullseye) (including 1st succ.)

key M, H

	x
H	1
M H	2
M M H	3
M M M H	4
....	$\vdots$
	∞

$p_X(1) = P(H) = 0.01$

$p_X(2) = P(M)P(H) = 0.99 \cdot 0.01$

$p_X(3) = P(M)P(M)P(H) = 0.99 \cdot 0.99 \cdot 0.01$

$p_X(4) = P(M)P(M)P(M)P(H) = 0.99^3 \, 0.01$

$p_X(x) = \underbrace{(1-p)}_{P(M)}{}^{x-1} \underbrace{p}_{P(H)}$

$$\boxed{p_X(x) = 0.99^{x-1} \, 0.01 \quad \text{for } x = 1, 2, 3, \ldots}$$

$X \sim Geo(p = 0.01)$

Bullseye (3/6)

Example 3

We throw darts at a dartboard until we hit the bullseye. Assume throws are independent and the probability of hitting the bullseye is 0.01 for each throw.

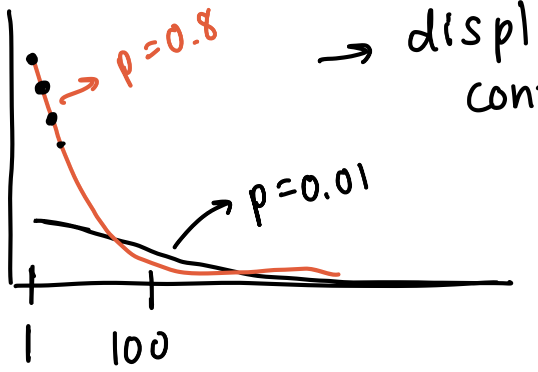
2. What are the mean and variance for the number of throws needed to hit the bullseye?

$$E(X) = \frac{1}{p} = \frac{1}{0.01} = 100 \text{ throws}$$

$$Var(X) = \frac{1-p}{p^2} = \frac{q}{p^2} = \frac{1-0.01}{(0.01)^2}$$

$$= 9{,}900 \text{ throws}^2$$

$$SD(X) = 99.499 \text{ throws}$$



→ displaying pmf as cont, BUT not ~~[illegible]~~ true

Bullseye (4/6)

on 4th try:

$$P(X=4) = p_X(4) = 0.99^3(0.01) = 0.0097$$

Example 3

We throw darts at a dartboard until we hit the bullseye. Assume throws are independent and the probability of hitting the bullseye is 0.01 for each throw.

3. Find the probability that our first bullseye:
   - is on the fourth try
   - is on one of the first four tries
   - is after the fourth try

one of 1st 4 tries:

$$P(X \le 4) = P(X=1) + P(X=2) + P(X=3) + P(X=4)$$

OR

$$= \sum_{x=1}^{4} p_X(x) = \underbrace{\sum_{x=1}^{4} 0.99^{x-1}}(0.01)$$

$$= 0.01\,(\underset{x=1}{1} + \underset{x=2}{0.99} + 0.99^2 + 0.99^3)$$

$$P(X \le 4) = 0.0394$$

after 4th try:

$$P(X > 4) = 1 - P(X \le 4) = 1 - 0.0394 = 0.9606$$

$$P(X \le 4) + P(X > 4) = 1$$

$$P(X > 4) = 0.9606$$

Bullseye (5/6)

Example 3

We throw darts at a dartboard until we hit the bullseye. Assume throws are independent and the probability of hitting the bullseye is 0.01 for each throw.

3. Find the probability that our first bullseye:
    - is on one of the first fifty tries
    - is after the 50^{th} try, given that it did not happen on the first 20 tries

one of 1st 50:

$$P(X \le 50) = F_X(50) = 1 - (\underbrace{1 - 0.01}_{1-p})^{50} = 1 - \underbrace{0.99}_{q}{}^{50} = 0.395$$

after 50th, given NOT in 1st 20:

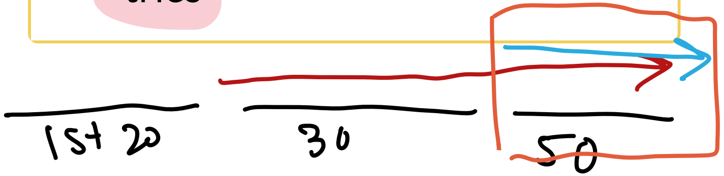
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(X > 50 \mid X > 20) = \frac{P(X > 50 \;\&\; X > 20)}{P(X > 20)}$$

$$= \frac{P(X > 50)}{P(X > 20)} = \frac{1 - P(X \le 50)}{1 - P(X \le 20)}$$

$$= \frac{\boxed{\cancel{1} - (\cancel{1}} + 0.99^{50})}{\boxed{\cancel{1} - (\cancel{1}} + 0.99^{20})} = \frac{0.99^{50}}{0.99^{20}} = 0.99^{30}$$

$$0.99^{30} = P(X > 30)$$



Memoryless property for Geometric RVs

If we know X is greater than some number (aka given $X > j$), then the probability of $X > k + j$ is just the probability that $X > k$.

$$P(X > 50 \mid X > 20) = P(X > 30 + 20 \mid X > 20) = P(X > 30)$$

$$P(X > k + j | X > j) = P(X > k)$$

$$P(X > k + j | X > j) = \frac{P(X > k + j \text{ and } X > j)}{P(X > j)} = \frac{P(X > k + j)}{P(X > j)} = \frac{(1-p)^{k+j}}{(1-p)^{j}} = (1-p)^k$$

$$(1-p)^{(k+j)-j}$$

applies to: $$P(X = k+j \mid X > j) = \frac{P(X = k+j \;\&\; X > j)}{P(X > j)}$$

$\to$ just $P(X = k+j)$

$$= \frac{\overbrace{(1-p)^{k+j}\, p}^{P(X = k+j)}}{(1-p)^{j}} = (1-p)^k p = P(X = k)$$

Chapter 19: Hypergeometric RVs

Hypergeometric RVs

- **Scenario:** There are a fixed number of successes and failures (which are known in advance), from which we make n draws without replacement. We are counting the number of successes from the n trials.
  - There is a finite population of N items
  - Each item in the population is either a success or a failure, and there are M successes total.
  - We randomly select (sample) n items from the population. w/out replacement
- Shorthand: $X \sim \text{Hypergeo}(M, N, n)$

$$X = \text{Number of successes in } n \text{ draws}$$

$$p_X(x) = P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

for x integer-valued

$$\max(0, n - (N - M)) \le x \le \min(n, M)$$

2

$$E(X) = \frac{nM}{N}$$

np

#suc / #total = often how calc p

$$\text{Var}(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Wolf population

Example 4

A wildlife biologist is using mark-recapture to research a wolf population. Suppose a specific study region is known to have 24 wolves, of which 11 have already been tagged. If 5 wolves are randomly captured, what is the probability that 3 of them have already been tagged?

24 wolves $\rightarrow N = 24$

11 tagged ("success") $\rightarrow M = 11$

5 sampled $\rightarrow n = 5$

Let X = # "successes" (aka tagged wolves) in $n = 5$ trials

$$P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$\rightarrow$ size of event choosing 3/5 tagged

$\rightarrow$ size sample

$$p_X(x) = \frac{\binom{11}{x}\binom{13}{5-x}}{\binom{24}{5}}$$

$$P(X=3) = p_X(3) = \frac{\binom{11}{3}\binom{13}{2}}{\binom{24}{5}}$$

Hypergeometric vs. Binomial RVs

Suppose a hypergeometric RV X has the following properties:

- the population size N is really big,
- the number of successes M in the population is relatively large,
  - $\frac{\mathrm{M}}{\mathrm{N}}$ shouldn't be close to 0 or 1
- and the number of items n selected is small.

M not close to N
M not close to 0

Rule of thumb

$\frac{n}{N} < 0.05$

or $N > 20n$

Then, in this case, making n draws from the population doesn't change the $\underbrace{\text{probability of success}}$ much, and the hypergeometric r.v. can be approximated by a binomial r.v.

$p = \frac{M}{N}$ fairly stable as we draw n samples

Wolf population revisited

Example 5

Suppose a specific study region is known to have 2400 wolves, of which 1100 have already been tagged.

n

1. If 50 wolves are randomly captured, what is the probability that 20 of them have already been tagged?
2. Approximate the probability in part (1) using the binomial distribution.

① $X \sim \text{Hypergeo}(N=2400, M=1100, n=50)$

$$P(X=20) = \frac{\binom{1100}{20}\binom{1300}{30}}{\binom{2400}{50}} = 0.08132$$

② check $N > 20n \rightarrow 2400 > 20 \cdot 50 = 1000$ ✓

$X \sim \text{Binomial}\left(n=50, p=\frac{M}{N}=\frac{11\cancel{00}}{24\cancel{00}}\right)$

$$P(X=20) = \binom{50}{20}\left(\frac{11}{24}\right)^{20}\left(\frac{13}{24}\right)^{30} = 0.08107$$

Chapter 20: Discrete Uniform RVs

Discrete Uniform RVs

- **Scenario:** There are N possible outcomes, which are all equally likely.
- Shorthand: $X \sim \text{Uniform}(N)$

$$X = \text{Outcome of interest, with } x = 1, 2, \ldots, N$$

$$p_X(x) = P(X = x) = \frac{1}{N} \text{ for } x = 1, 2, 3, \ldots, N$$

$$E(X) = \frac{N+1}{2}$$

$$\text{Var}(X) = \frac{N^2 - 1}{12}$$

What discrete uniform RVs have we seen already?

Example 6

Examples of discrete uniform RVs