MATH 320: Probability

Lecture 10: Discrete Distributions

Chapters 2: Distributions (2.1, 2.3, 2.4, 2.5, 2.6)
2.7

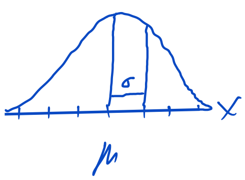
Introduction

- Statistical distributions are used to model populations.

  We usually deal with a **family** of distributions rather than a single distribution (family = type of distribution).

- This family is indexed by one or more parameters, which allows us to vary certain characteristics of the distribution while staying with one functional form.

  The functional form determines the unique features of the distribution.



- For example, if the distribution of a population of symmetric and bell-shaped, then a Normal distribution is a reasonable choice. Then we will specify the parameters μ & σ^2.

- In the previous sections, we saw a number of examples of discrete probability distributions.

  Recall a random variable X is said to have a discrete distribution if $\mathcal{X}$, range of X is countable.

- In the next sections we will study some special distributions that are extremely useful and widely applied, some of which we have already seen before.

Discrete uniform distribution

Definition

- Scenario: If a finite number of values are equally likely to be observed, then a discrete uniform distribution is used to model.

- Definition: A random variable X has a **discrete uniform** (N_0, N_1) distribution if

  min max

$$P(X = x \mid N_0, N_1) = \frac{1}{N_1 - N_0 + 1}, \qquad x = N_0, \ldots, N_1,$$

\# of numbers included in $\mathcal{X}$

where N_0 and N_1 are specified integers ($N_0 \leq N_1$).

- If X follows a discrete uniform (DU) distribution with parameters N_0 and N_1, we can summarize this using the following notation:

$$\text{RV} \sim \text{Distribution (parameters)} \qquad X \sim \text{Discrete uniform}\,(N_0, N_1)$$

  "The random variable X **follows** (is distributed as / assumed to have) a discrete uniform distribution with parameters N_0 and N_1".

  Can also use this more generally with $X \sim F_X(x)$ or $X \sim f_X(x)$.

- Example: Rolling a fair 6-sided die.

$$P(X = x \mid N_0, N_1) = \frac{1}{N_1 - N_0 + 1} = \frac{1}{6-1+1} = \frac{1}{6}, \quad x = 1, 2, \ldots, 6$$

  $f(x \mid N_0, N_1)$

- A note on notation: When we are dealing with parametric distributions, the distribution is dependent on the values of the parameters. In order to emphasize this fact and keep track of the parameters, write them in the pmf preceded by: | (given).

  In general, we have $P(X = x \mid \theta)$, where θ could be a vector of parameters.

  ↳ ex) DU -> $\theta = (N_0, N_1)$

Mean and variance

- One of the advantages of having families of distributions is that the pmfs have the same functional form (i.e. follow a certain pattern). This is also true for their expected value and variances.

- If a random variable X has a discrete uniform distribution (N_0, N_1),

$$E(X) = \frac{N_0 + N_1}{2} \qquad V(X) = \frac{(N_1 - N_0 + 1)^2 - 1}{12} \implies SD(X) = \sqrt{V(X)}$$

- The mean and variance are functions of the parameters.

- Continuing example: Find the mean and variance of rolling a fair 6-sided die.

  Use formulas:

$$E(X) = \frac{N_0 + N_1}{2} = \frac{1+6}{2} = 3.5$$

$$V(X) = \frac{(N_1 - N_0 + 1)^2 - 1}{12} = \frac{(6-1+1)^2 - 1}{12} = \frac{35}{12}$$

  Confirm with definition:

$$E(X) = \sum x f(x) = \frac{1}{6}\sum_{x=1}^{6} x = 3.5 \checkmark$$

Summary

- If we know (or assume) the family of a distribution and the values of the parameters

  $\iff$ know __pmf__ $\iff$ __cdf__

  We also know the expected value and variance as well.

- This statement works backwards too.
- *IMPORTANT STRATEGY* when solving problems:

  If we recognize that we have a pmf where the the range of the random variable and the probabilities match the scenario of a specific distribution, then that random variable must follow that specific distribution.

- Examples: How can we model the following populations?

  1. Suppose $f_X(x) = \frac{1}{4}$, $x = 7, 8, 9, 10$

     equally likely   Discrete range in a sequence

     $\Rightarrow Y \sim DU(N_0 = 7, N_1 = 11)$

     Find $E(X^2)$. $V(X) = E(X^2) - (E(X))^2 \Rightarrow E(X^2) = V(X) + (E(X))^2$

     $= \frac{(10-7+1)^2 - 1}{12} + \left(\frac{7+10}{2}\right)^2$

     $= \frac{16}{12} + (8.5)^2$

  2. Suppose the values 3, 6, 9, 12, 15 are equally likely.

     $X \sim DU(N_0 = 1, N_1 = 5) \longrightarrow$ transform $Y = 3X$

  3. Suppose the values 5, 9, 13, 17 are equally likely.

     $X \sim DU(N_0 = 1, N_1 = 4) \longrightarrow$ transform $Y = 4X + 1$

Bernoulli distribution

Motivation

- The following four distributions are constructed based on Bernoulli experiments: Binomial, Geometric, Hypergeometric, Negative Binomial.
- Example experiments: Success
  1. Produce a product and see if the product is defective. defective
  2. Roll a die and see if a number greater than 4 appears. $\# > 4$
  3. Choose a student in this class at random and see if a female student is chosen. Female
- What is the common characteristic of these experiments? $\approx$ True / false
- The main event of interest is labeled Success and the other is labeled failure.

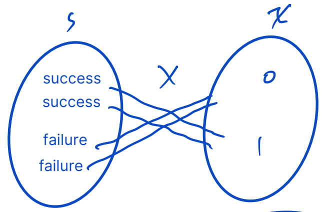
Definition

$X = \{0, 1\}$

- Scenario: The random variable for which 0 and 1 are chosen to describe the two possible values is called a **Bernoulli random variable**.
- The outcomes in Success and Failure are assigned __1__ and __0__ by the Bernoulli random variable, respectively.

exp 2

5
6
1
2
3
4



- Definition: A random variable X has a **Bernoulli** (p) distribution if

probability of failure

$$P(X = x \mid p) = \begin{cases} 1 - p = q & x = 0 \\ p & x = 1 \\ 0 & \text{otherwise} \end{cases}$$

where the parameter p represents the probability of success and $0 < p < 1$.

Equivalently, $f(x \mid p) = p^x \underbrace{(1-p)}_{=q}{}^{1-x}, \quad x = 0, 1.$

- Notation: $X \sim \text{Bernoulli}\,(p)$ (Ber)
- Example: A basketball player shoots a free throw with a 75% probability of success. Let X denote the number of points scored.

↳ $= p$

↳ $x = \{0,1\}$

(a) What is the distribution of X?

$X \sim Ber(p = 0.75)$

(b) Find the pmf of X.

$f_X(x \mid p = 0.75) = 0.75^x (0.25)^{1-x}, \quad x = 0, 1$

(c) Find the expected value of X (by hand using the definition).

$E(X) = \sum_{x=0}^{1} x f(x)$

$= 0(0.25) + 1(0.75)$

$= 0.75$

$\boxed{= p}$

(d) Find the variance of X (by hand using the definition).

$V(X) = E[(X - \underbrace{\mu_X}_{↳ p = 0.75})^2] = \sum_{x=0}^{1} (x - 0.75)^2 f(x)$

$= (0 - 0.75)^2 (0.25) + (1 - 0.75)^2 (0.75)$

$= 0.1875$

$= 0.75(0.25)$

$\boxed{= pq}$

Mean and variance

- If $X \sim$ Bernoulli (p)

$$E(X) = p \qquad V(X) = p(1-p) = pq \implies SD(X) = \sqrt{pq}$$

- Derive $E(X)$. *HINT*: Generalize the calculations from the previous examples.

$$E(X) = \sum_{x=0}^{1} x f(x)$$
$$= 0(1-p) + 1(p)$$
$$= p$$

- Derive $V(X)$.

$$V(X) = E[(X - \underbrace{\mu_X}_{=p})^2]$$
$$= \sum_{x=0}^{1} (x-p)^2 f(x)$$
$$= (0-p)^2 q + (1-p)^2 p$$
$$= p^2 q + q^2 p$$
$$= pq[\underbrace{p + \underbrace{q}_{=1-p}}_{=1}]$$
$$= pq$$

** We can always try to derive the mean and variance of distributions by going back to the definitions.

Binomial distribution

Motivation

- Example experiments:
  1. A basketball player gets 10 3-point shots and has a probability of success (making a basket) of 1/5. Let X be the number of makes out of the 10 attempts.
  2. A factory produces 100 products with the probability of defect 2/100. Let Y be the number of defective products out of 100 products.
  3. Roll a fair die 3 times. Let Z be the number of 6s out of the 3 rolls.
- What is the common characteristic of these experiments?

success / failure multiple times

Binomial experiments and binomial random variables

- Definition: An experiment is called a **binomial experiment** if all of the following hold:

  These are the conditions needed in order to have a binomial experiment.

  1. The experiment consists of a fixed number, n, of identical trials.

     (Note that **identical** $\Longrightarrow$ from the same family with the same parameter values.)

  2. Each trial results in one of two events: success or failure.
  3. There is a constant probability of success, p, for each trial.
  4. The trials are independent.
  5. The outcome is a sequence of successes and failures. e.g. (0, 1, 1, 0, 1)

  Said another way, a binomial experiment is a sequence of n identical Bernoulli trials.

- Definition: If X is the number of successes in a binomial experiment, X is called a **binomial random variable**.

- Examples:

  (a) Check the conditions to determine if example experiment 1 is indeed a binomial experiment.

  1. n=10 trials (shots) ✓
  2. Each trial is success (make) or failure (miss) ✓
  3. constant probability of success ($p = 0.2$) ✓
  4. Each trial is independent ✓
  5. outcome is a sequence of makes or misses ✓ S F S F F S ...

  1 0

  $\Longrightarrow X \sim \text{Binomial}(n = 10, p = 0.2)$

  (b) Are the conditions met for experiments 2 and 3? If so, identify n and p.

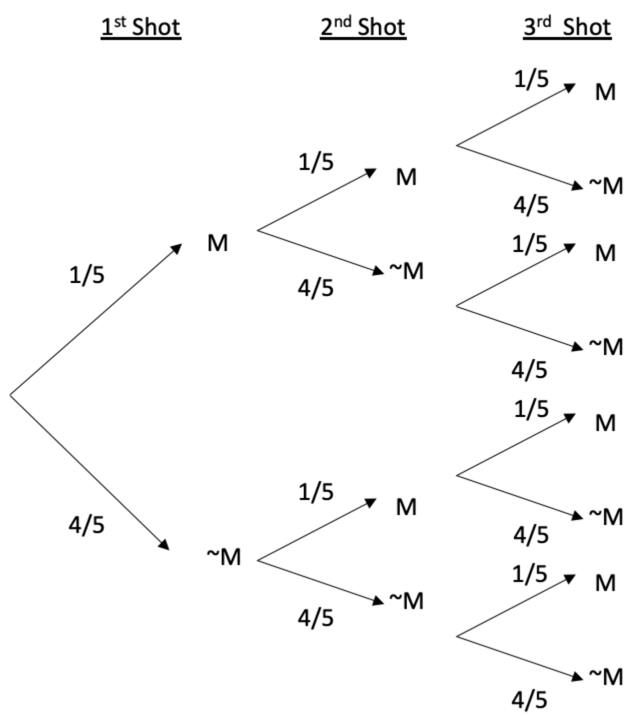
  Exp 2, Y
  Exp 3, Z
  } $\rightarrow$ All conditions are met ✓

  $Y \sim \text{Bin}(n = 100, p = 2/100)$

  $Z \sim \text{Bin}(n = 3, p = 1/6)$

Binomial probabilities

- Probabilities for binomial distribution are just an application of the multiplication rule for independent event.
- Example: Lets say our basketball player has only 3 shot attempts now. Lets visualize the tree diagram illustrating the different outcomes of this binomial experiment (M = make, $\sim M$ = miss).



Outcome	Branch	# Makes, X	Probability
M M M	1	3	$(\frac{1}{5})^3$
M M ~M	2	2	$(\frac{1}{5})^2 \frac{4}{5}$
M ~M M	3	2	$(\frac{1}{5})^2 \frac{4}{5}$
M ~M ~M	4	1	$\frac{1}{5}(\frac{4}{5})^2$
~M M M	5	2	$(\frac{1}{5})^2 \frac{4}{5}$
~M M ~M	6	1	$\frac{1}{5}(\frac{4}{5})^2$
~M ~M M	7	1	$\frac{1}{5}(\frac{4}{5})^2$
~M ~M ~M	8	0	$(\frac{4}{5})^3$

(a) Find $P(X = 3)$. $= 1 \cdot (\frac{1}{5})^3$

$=$ (1 branch) (probability of branch)

(b) Find $P(X = 2)$. $=$ 2 + 3 + 5

$3\,[(\frac{1}{5})^2 \frac{4}{5}]$

$=$ (# branches with exactly 2 makes) $*$ (Probability of a single branch with exactly 2 makes)

Another way to think about the number of branches: 3 sequences that include exactly 2 makes.

1 1 0
1 0 1
0 1 1

This could also be found using combinations.

S S S S F F F F
S S S F S F F F
0
0
0

(c) Now suppose $n = 8$. Find $P(X = 4)$.

$\binom{8}{4}$ = 70 sequences = 70 ways to choose 4 success out of the $n = 8$ trials.

Now just need the probability of 4 success and 4 failures:

$$P(X=4 \mid n=8, p) = \underbrace{\binom{8}{4}}_{\#} \underbrace{p^4 q^4}_{\text{prob}}$$

- We can use these patterns to obtain a general formula for the $P(X = x)$ for a binomial random variable with (n, p).

Definition

- Definition: A random variable X has a **binomial** distribution based on n trials with success probability p if the pmf of X has the form

$$P(X = x \mid n, p) = \underbrace{\binom{n}{x}}_{\#\ \text{sequences}} \underbrace{p^x (1-p)^{n-x}}_{\text{probability of each sequence}}, \qquad x = 0, 1, \ldots, n$$

$\to \dfrac{n!}{x!(n-x)!}$

where $0 < p < 1$ and n is an integer such that $n \geq 1$.

- Notation: $X \sim \text{Binomial}\,(n, p)$ + range $\mathcal{X} = \{0, 1, 2, \ldots, n\}$

- Example: Suppose a student is taking a multiple choice exam with 15 questions (4 choices each). They will be randomly guessing on each question. Let X be the number of questions out of 15 for which the student guesses correctly.

a) What is the distribution of X? conditions met ✓

$X \sim \text{Binomial}(n = 15, p = 0.25)$

b) Find the pmf of X.

★ Always need to include range

$f_X(x \mid p, n) = \binom{15}{x} 0.25^x \, 0.75^{15-x}, \quad x = 0, 1, \ldots, 15$

c) Find $P(X = 7)$.

$P(X=7) = \binom{15}{7} 0.25^7 \, 0.75^8$

Note: on Hw, okay to leave answers in this form

d) Find the probability of at least one correct.

at least 1 none

$P(X \geq 1) = 1 - P(X=0)$

$= 1 - \binom{15}{0} 0.25^0 \, 0.75^{15}$

$= 1 - 0.75^{15}$

e) Find $P(X < 6)$.

$P(X < 6) = P(X \le 5) = \sum_{x=0}^{5} \binom{15}{x} 0.25^x \, 0.75^{15-x}$

True because discrete

f) Find $P(X \ge 6)$.

$P(X \ge 6) = 1 - P(X \le 5) = \sum_{x=6}^{15} \binom{15}{x} 0.25^x \, 0.75^{15-x}$

Right

left

- Binomial cdf: $P(X \le x) = \sum_{t=0}^{x} \binom{n}{t} p^t q^{n-t}$

  Right sided probability: $P(X > x) = 1 - P(X \le x)$

  be careful with $\ge x$ vs $> x$

Mean and variance

- If $X \sim \text{Binomial}\,(n, p)$

$$E(X) = np \qquad V(X) = np(1-p) = npq \quad \Rightarrow \quad SD(X) = \sqrt{npq}$$

- Continuing example:

  (f) Find the expected value and variance of the number of questions guessed correctly.

  $E(X) = np$
  $= 15(0.25)$
  $= 3.75$ correct questions

  $V(X) = npq$
  $= 15(0.25)(0.75)$
  $= 2.8125 \Rightarrow SD(X) = 1.68$ correct questions

  (g) Suppose each question is worth 4 points. Find the expected value and variance of the number of points missed on the exam if it is worth a total of 60 points.

  * X = number of question correct:
  * Y = number of points earned: $Y = 4X$
  * Z = number of points missed: $Z = 60 - Y = 60 - 4X$

$\rightarrow E(Z) = E(60 - 4X) = 60 - 4\underbrace{E(X)}_{= 3.75} = 45$

$\rightarrow V(Z) = V(60 - 4X) = (-4)^2 \underbrace{V(X)}_{= 2.8125} = 45$

More examples

1. A coin is weighted so that the probability of flipping heads is 0.65. The coin is flipped 10 times and each flip is independent of every other flip. Let X be the the number of heads in the 10 flips.

   (a) Give the pmf of X, μ_X and σ_x^2. $X \sim Bin(n=10, p=0.65)$

   $f(x) = \binom{10}{x} 0.65^x \, 0.35^{10-x}, \quad x = 0, 1, \dots, 10 \longrightarrow$ $E(X) = 10(0.65) = 6.5$, $V(X) = \sqrt{10(0.65)(0.35)} \approx 1.508$

   (b) Find: $P(X = 3)$ $\qquad$ $P(X < 3)$ $\qquad$ $P(X \geq 3)$.

   $P(X=3)$ = Binompdf(trials $= 10$, p $= 0.65$, x value $= 3$) ≈ 0.0212

   $= P(X \leq 2)$ = Binomcdf(trials $= 10$, p $= 0.65$, x value $= 2$) ≈ 0.0048

   $= 1 - P(X \leq 2)$ $= 1 -$ Binomcdf$(10, 0.65, 2)$ ≈ 0.9952

   (c) If $Y = 10 - X$ give the distribution of Y. Find $P(Y \leq 1)$.

   Y: # failures $\qquad$ X: # successes

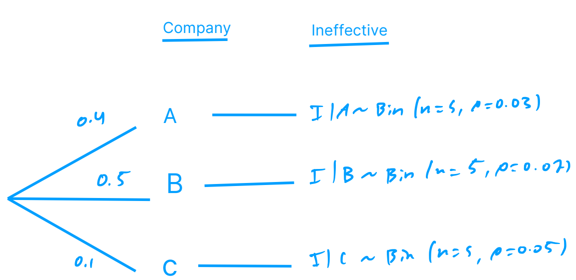
   $Y \sim$ Binomial$(n=10, p=0.35)$ original q

   $\to P(Y \leq 1)$ = Binomcdf$(10, 0.35, 1) = 0.8595$ $\Longleftrightarrow$ $P(X \geq 9) = 1 -$ Binomcdf$(10, 0.65, 8) = 0.08595$ ✓

2. A hospital obtains 40% of its flu vaccine shipments from Company A, 50% from Company B, and 10% from Company C.

   From manufacturing specifications, it is known that 3% of the vials from A are ineffective, 2% from B are ineffective, and 5% from C are ineffective.

   The hospital tests five vials from each large shipment from a company. If at least one of the five is ineffective, find the conditional probability of that shipment having come from C.



$$P(C \mid I \geq 1) = \frac{P(C \cap I \geq 1)}{P(I \geq 1)}$$

$$= \frac{P(C)\, P(I \geq 1 \mid C)}{P(A)\, P(I \geq 1 \mid A) + P(B)\, P(I \geq 1 \mid B) + P(C)\, P(I \geq 1 \mid C)}$$

$P(I \geq 1 \mid \cdot) = 1 - P(I = 0 \mid \cdot) = 1 - p^5$

$$= \frac{0.5(1 - (0.95)^5)}{0.4(1 - (0.97)^5) + 0.5(1 - (0.98)^5) + 0.1(1 - (0.95)^5)}$$

≈ 0.1779

Relationship between Bernoulli and binomial

- We can think of the result of a binomial experiment as a sequence of 0s and 1s with length n, where each individual number is the result of a Bernoulli experiment.
- X is the number of successes. Take $n = 5$ for example. We could have:

  $(0, 0, 1, 0, 0)$ $x = 1$

  $(1, 0, 1, 1, 0)$ $x = 3$ } $X =$ Sum of the vectors

  $(1, 1, 1, 1, 1)$ $x = 5$

- Suppose that $X \sim \text{Bin}(n, p)$ and $Y_1, \ldots, Y_n \overset{\perp\!\!\!\perp}{\sim} \text{Ber}(p) \iff Y_i \overset{iid}{\sim} \text{Ber}(p)$ where $iid =$ independent and identically distributed. $\hookrightarrow i = 1, \ldots, n$

  $$X = \sum_{i=1}^{n} Y_i$$

  Bernoulli distribution is a special case of Binomial when $n = 1 \longrightarrow \text{Ber}(p) = \text{Binomial}(n = 1, p)$

  This is intuitively obvious and can be verified in the distribution formula as well.

- Compare the expected values and variances.

  Note that the expected value of a sum is the sum of expected values.

  $E(X) = E\left[\sum_{i=1}^{n} Y_i\right]$

  $np = \sum_{i=1}^{n} \underbrace{E(Y_i)}_{= p \text{ for all } Y_i}$

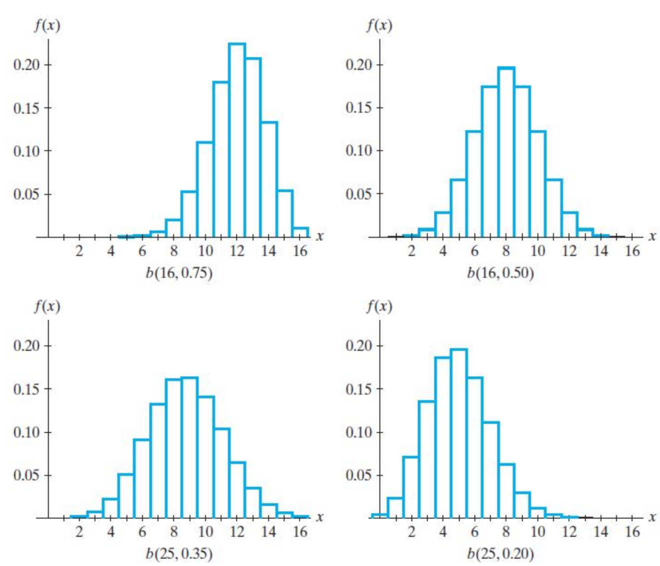
  $\checkmark = np$

  $V(X) = V\left[\sum_{i=1}^{n} Y_i\right]$

  $npq = \sum_{i=1}^{n} \underbrace{V(Y_i)}_{= pq \text{ for all } Y_i}$

  $\checkmark = npq$

Visualizing binomial distributions



Geometric distribution

Motivation

- Example experiments:
  1. A basketball player shoots three pointers (with success probability of 1/5) until they make the first one. Let X be the number of shots it took to make the first one.
  2. An oil prospector will drill a succession of holes in a given area to find a productive well. He stops when he finds a productive well. Let Y be the number wells drilled before the first productive well.
  3. Roll a fair die until the number 6 appears. Let Z be the number of rolls to get the first 6.
- What is the common characteristic of these experiments? success/ failure, go until first success

Geometric experiments and geometric random variables

- The general setting for a geometric distribution problem has many features in common with a binomial distribution problem.
- Characteristics of a **geometric experiment**.

  Same as Binomial
  1. Each trial is a Bernoulli experiment.
  2. The trials are identical and independent.
  3. The sample space is: $S = \{S, FS, FFS, FFFS, \ldots\}$
- Geometric is a waiting time random variable, with the number of success fixed at 1.
- There are TWO forms of a **geometric RV**. We will be using the FIRST.
  - These are just two different ways to interpret outcomes in the sample space (i.e. define the range of the random variable).
  1. The random variable of interest X is the number of trials. → like exp 1
  2. The random variable of interest Y is the number of failures before the first success. → like exp 2
- Relationship between the two formulations.
  - The first way X includes the trial on which the success occurs, whereas the second way Y does not. $\Rightarrow$ $Y = X - 1$
  - This obviously will result in the pmf, expected value, etc. being different between X and Y.

- Example: Check the conditions to determine if example experiment 1 is indeed a geometric experiment.

Each shot follows Independent Bernoulli and want number of shots needed for first make (X)

e.g.) $P(X=4) = 0.8^3(0.2)$

↳ counting trials ☆

Definition

- Informal derivation of the pmf: We can use the multiplication rule of independent events to find the probability of x trials to get the first success.

In terms of X

1. Success on first trial: $P(X=1) = q^0 p \quad = q^{1-1} p$
2. Success on second trial: $P(X=2) = q^1 p \quad = q^{2-1} p$
3. Success on third trial: $P(X=3) = q^2 p \quad = q^{3-1} p$

$\underbrace{q^{x-1}}_{\text{Failure on previous x-1 trials}} \underbrace{p}_{\text{success on last trial}}$

- If $X \sim \text{Geometric}\,(p)$, (Geo)

$$\boxed{P(X = x \mid p) = q^{x-1} p, \quad x = 1, 2, 3, \ldots, \quad 0 < p < 1}$$

Alternate form: $Y =$ # failures before first success

☆ Note different ranges

↳ $P(Y=y|p) = q^y p$, $y = 0, 1, 2, \ldots$

Mean and variance

- If $X \sim \text{Geometric}\,(p)$,

$$E(X) = \frac{1}{p}, \qquad V(X) = \frac{1-p}{p^2} = \frac{q}{p^2}$$

Alternate form: $Y \to E(Y) = \frac{q}{p} \neq E(X)$ $\quad V(Y) = \frac{q}{p^2} \overset{\checkmark}{=} V(X)$

↳ $= X-1$

- Continuing example: Let $X \sim \text{Geometric}(p = 1/5)$. How many trials do you expect to see the first success (make)?

$E(X) = \frac{1}{p} = \frac{1}{1/5} = 5 \to$ this makes sense ✓

Geometric probabilities

- A geometric random variable is a geometric series:

$P(X=1) = pq^0$
$P(X=2) = pq^1$
$P(X=3) = pq^2$
$\vdots$

- **Sums of geometric series**: We will use these properties to derive formulas for geometric probabilities.

  Let q be a real number such that $|q| < 1$, and let m be any positive integer $m \geq 1$.

  - Infinite geometric series:

    (a) $\displaystyle\sum_{i=0}^{\infty} q^i = q^0 + q^1 + q^2 + \cdots = \frac{1}{1-q}$ (b) $\displaystyle\sum_{i=1}^{\infty} q^i = \frac{q}{1-q}$

    PATTERN: Numerator = first term in the series and denominator = $1-q$ ALWAYS when infinite sum.

  - Another sum:

    (c) $\displaystyle\sum_{i=0}^{m} q^i = \underbrace{\sum_{i=0}^{\infty} q^i}_{(a)} - \underbrace{\sum_{i=m+1}^{\infty} q^i}_{(b)} = \frac{1}{1-q} - \frac{q^{m+1}}{1-q} \overset{\checkmark}{=} \frac{1-q^{m+1}}{1-q}$

- By these sums of a geometric series, we can compute the following probabilities. For positive integers x, a, and b such that $a < b$.

  - $\boxed{\text{Total probability: } P(X < \infty) = 1.}$
  - Derive this:

$$P(X < \infty) = \sum_{x=1}^{\infty} q^{x-1} p$$
$$= p \sum_{x=1}^{\infty} q^{x-1}$$

Doesn't match (a) by looks, but if we write it out...

$q^{1-1} + q^{2-1} + q^{3-1}$
$= q^0 + q^1 + q^2$
$\overset{\checkmark}{=}$ (a) $\Rightarrow$ can use result

$$= p\left[\underbrace{\frac{1}{1-q}}_{=1-p}\right]$$
$$= \frac{p}{p}$$
$$\overset{\checkmark}{=} 1$$

  - $\boxed{\text{Cdf: } P(X \leq x) = 1 - q^x}$

$$F(x) = P(X \leq x) = \sum_{t=1}^{x} q^{t-1} p = p \sum_{t=1}^{x} q^{t-1}$$

Transform

let $j = t-1$ $\Rightarrow$ Bounds

$t = x \Rightarrow j = x-1$
$t = 1 \Rightarrow j = 0$

$$= p \underbrace{\sum_{j=0}^{x-1} q^j}_{(c)}$$
$$= p\left[\frac{1-q^{(x-1)+1}}{\underbrace{1-q}_{=p}}\right]$$
$$= 1 - q^x$$

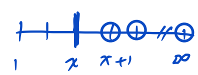
Intuitive derivation of cdf: $F(x) = P(X \le x)$

$= 1 - \underbrace{P(X > x)}_{\text{need more than } x \text{ trials}}$

$= 1 - q^x$

$\Rightarrow$ failed first x times
doesn't matter what happens next

– Right probability (exclusive): $P(X > x) = \sum_{i=x+1}^{\infty} q^{i-1} p$ Directly

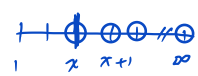


$= 1 - P(X \le x) = 1 - F(x)$

$= 1 - (1 - q^x)$

$\boxed{= q^x}$

– Right probability (inclusive): $\boxed{P(X \ge x) = P(X > x-1) = q^{x-1}}$

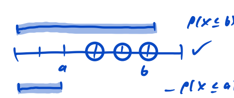


OR $= 1 - (1 - P(X \le x-1))$

$= 1 - (1 - q^{x-1})$

$\overset{\checkmark}{=} q^{x-1}$

– Interval probability: $P(a < X \le b) = P(X \le b) - P(X \le a)$



$= 1 - q^b - (1 - q^a)$

$\boxed{= q^a - q^b}$

– Interval probability (both inclusive): $P(a \le X \le b) = P(a-1 < X \le b)$

$\boxed{= q^{a-1} - q^b}$

- Example: Suppose that the probability of engine malfunction during any one-hour period is $p = 0.02$. Let X be the number of one-hour intervals until the first malfunction.

  a) What is the distribution of X?   Conditions met ✓

  $X \sim \text{Geometric}(p = 0.02)$

  b) Find the pmf of X.

  $f_X(x) = 0.98^{x-1}(0.02), \quad x = 1, 2, 3, \ldots$

c) Find the probability that an engine will survive 3 hours.

we know the first 3 trials (hours) are failure (surviving)
Then we don't care when success (malfunction) happens

$P(X>3) = 0.98^3 \approx 0.941$

$= 1 -$ geometcdf $(p=0.02, x=3) \approx 0.941$ ✓

d) Find the probability that an engine will survive 3 hours, but die before or during the 6^{th} hour.

$P(3 < X \leq 6) = 0.98^3 - 0.98^6 \approx 0.055$

$=$ geometcdf$(p=0.02, x=6)$ $-$ geometcdf$(p=0.02, x=3)$

$= \sum_{x=4}^{6}$ geometpdf $(p=0.02, x=x)$

e) Suppose that an engine survives 3 hours. Find the probability that the engine will survive 6 hours.

conditional

$P(X>6 \mid X>3) = \dfrac{P(X>6)}{P(X>3)}$ $\quad \{X>6\} \subset \{X>3\} \Rightarrow \dfrac{P(A\cap B)}{P(B)} = \dfrac{P(B)}{P(A)}$

$= \dfrac{0.98^6}{0.98^3}$

$= 0.98^3 \Longleftrightarrow = P(X>3) \longrightarrow$ memoryless property

same calculation as

Memoryless property

- The geometric distribution has an interesting property, known as the "memoryless" property. For integers $s > t$, it is the case that

$$P(X > s \mid X > t) = P(X > s - t)$$

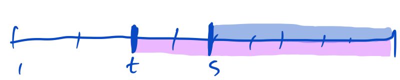
- To help us understand what this means, let's interpret some events for $s > t$.
  - $\{X > s\}$ → First s trials are failures
  - $\{X > t\}$ → First t trials are failures
  - $\{X > s\} \mid \{X > t\}$ → Next s-t trials are failures after the first t trials are failures

- For example, lets say I flip a coin three times and don't get a tails. Do these past failures help to increase the probability of getting a tails next?

No! FFF FFF FFF

- We want to know how (if it all) this past information influences the future event. Recall $P(X > a) = q^a$.

$P(\underbrace{X > s}_{a} \mid \underbrace{X > t}_{b}) = \dfrac{P(\{X>s\}\cap\{X>t\})}{P(X>t)} = \dfrac{P(X>s)}{P(X>t)} = \dfrac{q^s}{q^t} = \boxed{q^{s-t} \overset{\checkmark}{=} P(X>s-t)}$



$P(X>s) \subset P(X>t)$

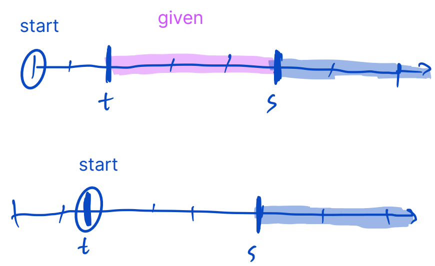
- The probability of getting an additional failures $s-t$, having already observed t failures is the same probability of observing $s-t$ failures at the start of the sequence.

$X > s \mid X > t$

equivalent to

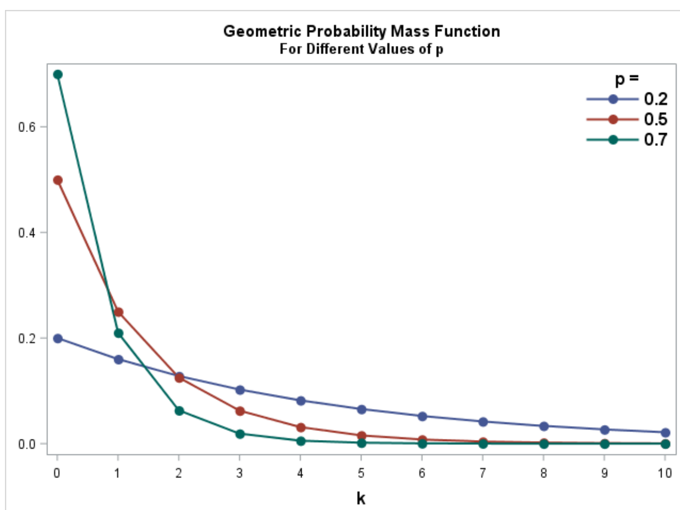
$\Longleftrightarrow$

$X > s-t$



- That's why this property is called the memoryless property. Even though we have some failures before, a geometric experiment is sort of restarting again. Thus, the geometric distribution "forgets" what has occurred.

Visualizing geometric distributions



Negative binomial distribution

Motivation

- Example experiments:
  1. A basketball player shoots three pointers (with success probability of 1/5) until they make the 4 shots. Let X be the number of shots it took to make the 4^{th} one.
  2. An oil prospector will drill a succession of holes in a given area to find a productive well. He stops when he finds three productive wells. Let Y be the number wells drilled before the 3^{rd} productive well.
- What is the common characteristic of these experiments?

  success/ failure
  go until succeed multiple times

Negative binomial experiments and negative binomial random variables

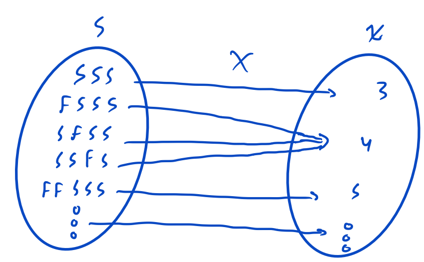
- The negative binomial experiment is a generalized version of the geometric experiment.
- Characteristics of a **negative binomial experiment**.

  same as geometric
  1. Each trial is a Bernoulli experiment.
  2. The trials are identical and independent.
  3. The number of successes to stop a negative binomial experiment is denoted by r.
     - This is a parameter of the distribution. Geometric stopped at a fixed r= 1 successes

       can vary
- **Negative binomial random variable**.

  (Following the logic from the geometric distribution, there is also two ways to define a negative binomial random variable. We will only discuss the one corresponding to the version of the geometric we are using.) # trials form

  - The random variable of interest X is the number of trials until r successes.
  - Simple demonstration of the random variable mapping: When $r = 3$, the sample space and range are:



  - In general, the range of X is

$$\mathcal{X} = \{r, r+1, \dots\}$$

- Example: Check the conditions to determine if example experiment 1 is indeed a negative binomial experiment.

→ Geometric experiment with r=4 successes ✓
→ want total number of shots (x)

Definition

- Informal derivation of the pmf: We can again use the multiplication rule of independent events to find the probability of x trials to get r successes.

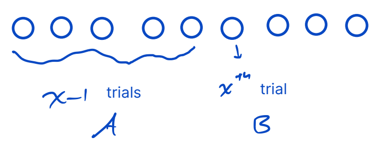
1. The outcomes in the event $\{X = x\}$ must follow these two rules:

   A) The first $x - 1$ trials result in $r - 1$ successes and $\underbrace{x - r}_{\text{total number of failures}}$ failures.

   B) The last x^{th} trial has to result in success.

2. Lets visualize these rules (and think of them as events) to get some probabilities.

   – Consider a sequence of Bernoulli trials with probability of success p, $0 < p < 1$. Let $x = r, r + 1, \ldots$.



   A) $A \sim$ Binomial $(n = x-1, p)$ → want $P(A = r-1)$

   $P(A) = \binom{x-1}{r-1} p^{r-1} q^{x-r}$ → $x-1-(r-1)$

   $A \perp\!\!\!\perp B$

   B) $B \sim$ Bernoulli (p) → want $P(B=1)$

   $P(B) = p$

3. Putting these together:

$$f_X(x \mid r, p) = P(A \cap B) = P(A)\,P(B) = \left[\binom{x-1}{r-1} p^{r-1} q^{x-r}\right] p$$

$$\overset{\checkmark}{=} \binom{x-1}{r-1} p^{r} q^{x-r}$$

- If $X \sim$ Negative binomial (r, p),

$$P(X = x \mid r, p) = \underbrace{\binom{x-1}{r-1}}_{\#\text{ of sequences}} \underbrace{p^r\, q^{x-r}}_{\text{probability of each individual sequence}} \qquad x = r, r+1, \ldots$$

- Note: This is called the [redacted]negative binomial"; distribution because it looks like the binomial pmf, except with minus ones in the combination.

Mean and variance

- If $X \sim$ Negative binomial (r, p),

$$E(X) = \frac{r}{p}, \qquad V(X) = \frac{r(1-p)}{p^2} = \frac{rq}{p^2} \Longrightarrow SD(x) = \sqrt{\frac{rq}{p^2}}$$

Example

- You are playing the slot machine on which the probability of a win on any individual trial is 0.05. You will play until you win twice. Let X denote the number of plays in order to get your second win.

  a) What is the distribution of X?

  Conditions met ✓

  $X \sim NB(r=2, p=0.05)$

  b) Find the pmf of X.

  $f_X(x) = \binom{x-1}{2-1} 0.05^2 \, 0.95^{x-2}, \quad x = 2, 3, \ldots$

  c) Find the probability that the second win occurs on the 8^{th} play.

  $p(X=8) = \binom{7}{1} 0.05^2 \, 0.95^6 \approx 0.0129$

  d) Find $E(X)$ and $V(X)$.

  $E(X) = \frac{r}{p} = \frac{2}{0.05} = 40$ plays

  $V(x) = \frac{rq}{p^2} = \frac{2(0.95)}{0.05^2} = 760 \longrightarrow SD(x) \approx 27.57$ plays

  e) Find the probability that the [redacted] fourth win occurs [redacted] before the 10^{th} play.

  * Be careful with bounds

  $Y \sim NB(r=4, p=0.05) \Longrightarrow p(Y<10) = \sum_{y=4}^{9} \binom{y-1}{4-1} 0.05^4 \, 0.95^{y-4}$

Relationship between geometric and negative binomial

- We stated the negative binomial distribution is a generalized version of the geometric distribution. Here's why.
- The negative binomial is the sum of independent and identical geometric experiments:

  Formally, if $X \sim \text{NB}(r, p)$ and $Y_1, \ldots, Y_r \overset{\perp\!\!\!\perp}{\sim} \text{Geo}(p)$.

  $$X = \sum_{i=1}^{r} Y_i$$

- We can see this for the expected value and variance too.

$E(X) = r \underbrace{\left(\frac{1}{p}\right)}_{\hookrightarrow = E(Y_i)}$ geometric mean

$V(X) = r \underbrace{\frac{q}{p^2}}_{\hookrightarrow = V(Y_i)}$ geometric variance

- Intuitively this makes sense.
  - If $r = 2$. Y_1 is the number of trials to get the first success and Y_2 is the number of subsequent trials to get the second success.
  - If we are waiting for the second success, we wait through Y_1 trials for the first success, and then repeat the process as we go through Y_2 subsequent trials for the second success, for a total of $X = Y_1 + Y_2$ trials.

e.g.) $r = 3$ $\rightarrow$ $\boxed{\text{F S FFS F S}}$ one NB exp $\rightarrow X = 7$

$\boxed{\text{F S}}\boxed{\text{FFS}}\boxed{\text{F S}}$ three Geo exps

$Y_1 = 2$ $Y_2 = 3$ $Y_3 = 2$ $\rightarrow \sum Y_i = 2 + 3 + 2 \overset{\checkmark}{=} 7$

  - Note: Even though Y_1 and Y_2 follow the same kind of geometric distribution, Y_1 and Y_2 can have different values. So $Y_1 + Y_2$ is NOT the same as $2Y_1$.

- The geometric distribution is a special case of the negative binomial distribution when $r = 1$, which counts the number of trials for the first success

  This again is intuitively obvious and can be verified with the distribution formula.

$X \sim NB(r=1, p) \iff \text{Geometric}(p)$

Hypergeometric distribution

Motivation

- Polling problem:
  - Suppose you live in a large city which has 1,000,000 registered voters. The voters will vote on a tax issue next month and you want to estimate the percent of the voters who favor the issue.

    You cannot poll everyone, so you randomly sample 100 voters and ask them if they favor the issue. What are your chances of correctly estimating the true percentage in favor of the issue?

    This is a MATH 321 question! We could build a confidence interval to estimate the unknown p.
  - For now, we can make a simplifying assumption to answer this question. Suppose the true percentage of the voters in favor of the issue is 65%. We don't actually know this number, it is what we are trying to estimate.
  - But with this assumption, when polling voters, we are really doing a binomial experiment.

    $\rightarrow X$ = number in favor of issue

    $\sim$ Binomial ($n = 100$, $p = 0.65$)

    $\rightarrow P(X = 0.65) = ??$

- Checking assumptions for this polling problem.
  - Each voter (trial) is in favor (success) or not in favor (failure).

    Random sampling, so each successive voters are independent.

    Sampling $n = 100$ voters, and get a sequence of S/F.

    Is there a constant probability of success p? No!
  - The usual method of sampling voters is called **sampling without replacement**.

	voter 1		voter 2
options :	1,000,000		999,999
success probability:	p_1	$\neq$	p_2

    $\Rightarrow$ probability in favor changes slightly

  - When there is a very large population N and the sample n is very small in comparison, not replacing changes things very little on each trial, and it is still reasonable to model this scenario with the binomial distribution.

    In intro stats classes, this condition is met if $\frac{1}{10}N \geq n$.
  - The hypergeometric distribution will handle sampling without replacement exactly for any population size.

- Example experiments:
  1. Five cards are dealt from a standard deck. Let X be the number of aces in the hand.
  2. There are three red chips and four blue chips in a bowl. We randomly select four chips without replacement from the bowl. Let Y be the number of blue chips selected.
- What is the common characteristic of these experiments?

number of successes in some amount of trials, without replacement

Example calculations

- Informal derivation of the pmf via an example: We have actually already been doing hypergeometric problems (sampling without replacement) when we did certain counting problems.
  - (Test 1 Q3) A gym teacher is picking students to compete in a pickup basketball game. There are 45 students in the class, which includes 25 boys and 20 girls. The teacher picks 7 students from the 45 at random and without replacement. Find the probability that the team includes exactly 4 girls.

$$P(4 \text{ Girls}, 3 \text{ boys}) = \frac{\binom{20}{4}\binom{25}{3}}{\binom{45}{7}} = \frac{\binom{\#\text{ of ways}}{\text{choose 4 Girls}}\binom{\#\text{ of ways}}{\text{choose 3 Boys}}}{\binom{\#\text{ of ways}}{\text{choose 7 overall}}}$$

≈ 0.246

  - We can easily generalize this to find the probability that the selected team includes any number of girls between 0 and 7. Let X be the number of girls selected.

$$P(X=x) = \frac{\binom{20}{x}\binom{25}{7-x}}{\binom{45}{7}}, \quad x = 0, 1, 2, \ldots, 7$$

  - This leads to the following pmf for X:

x	0	1	2	3	4	5	6	7
$f(x)$	0.011	0.078	0.222	0.318	0.246	0.102	0.021	0.002

  - Again, successive selections are dependent on previous one. So the probabilities change.

→ Direct way

$$\underbrace{P(G_1)}_{\text{Girl 1}} = \frac{20}{45} \approx 0.444 \qquad \underbrace{P(G_2 \mid G_1)}_{\text{Girl 2}} = \frac{19}{44} = 0.433$$

  - Now we can formalize this example and relate it to how we will talk about hypergeometric experiments and random variables.

In Example

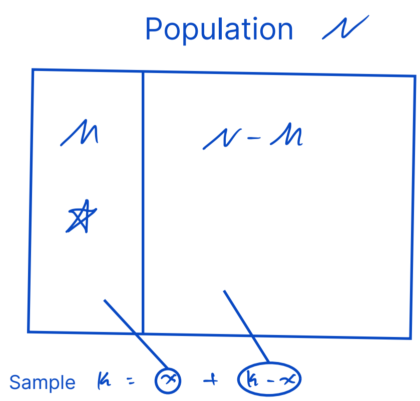
→ K = 7 team members
N=45 total students
M=20 Girls
⇒N-M = 25 Boys

→ x: number of girls selected for team

Hypergeometric experiments and hypergeometric random variables

- Characteristics of a **hypergeometric experiment**.
  1. Each trial is a Bernoulli experiment.
  2. Trials are not identical and not independent.
  3. A sample of size K is being taken from a finite population of size N.
  4. The population has a subgroup of size M that is of interest.

- **Hypergeometric random variable**.
  - The random variable of interest X is the number of members of the subgroup in the sample taken.
  - Three parameters:
    (a) N = Population size
    (b) M = Number of objects of interest
    (c) K = Sample size
  - Lets visualize this scenario.



(order doesn't matter)

Definition

- If $X \sim$ Hypergeometric (N, M, K),

$$P(X = x \mid N, M, K) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}, \qquad x = 0, 1, \ldots, \min(M, K)$$

$$= \frac{\left(\text{choose } x \text{ successes from } M\right)\left(\text{choose } K-x \text{ failures from } N-M\right)}{\left(\text{choose } K \text{ overall from } N\right)}$$

- Two cases for the range of X.
  1. In most applications $M \geq K$, which means the sample size is smaller than the subgroup of interest.
     - This implies: $\mathcal{X} = \{0, 1, \ldots, \min(\underset{20}{M}, \underset{7}{K})\}$.

     ex) $M = 20$ $K = 7$
     $x = 0, 1, 2, \ldots, 7$
  2. But if the sample size K is quite large and we lose this restriction, the formula will still be applicable.
     - But with new range: $\mathcal{X} = \{\max(0, K + M - N), \ldots, \min(\underset{10}{M}, \underset{20}{K})\}$.

       $20 + 10 - 25 = 5$

ex) $N = 25$
$K = 20$
$M = 10$
$N - M = 15$

$x = 5, 6, \ldots, 10$

Examples

1. A lot consisting of 200 fuses is inspected by the following procedure: 20 fuses are chosen at random and tested; if at least 19 blow at the correct amperage, the lot is accepted. If a lot contains 10 defective fuses what is the probability of accepting this lot?

   Solve this problem two different ways.

$X =$ number of defective fuses

$\downarrow \sim HG(N=200, M=10, K=20)$

$\Rightarrow x = 0, 1, 2, \ldots, 10$

$$P(X \le 1) = P(X=0) + P(X=1)$$

$$= \frac{\binom{10}{0}\binom{190}{20} + \binom{10}{1}\binom{190}{19}}{\binom{120}{20}}$$

$$\approx 0.7372$$

$Y =$ Number of Correct fuses

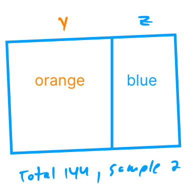
$\downarrow \sim HG(N=200, M=190, K=20)$

$\Rightarrow y = 10, 11, \ldots, 20$

$$P(Y \ge 19) = P(X=19) + P(X=20)$$

$$\frac{\binom{190}{19}\binom{10}{1} + \binom{190}{20}\binom{10}{0}}{\binom{200}{20}}$$

2. A bag contains 144 ping-pong balls. More than half of the balls are painted orange and the rest are painted blue. Two balls are drawn at random without replacement. The probability of drawing two balls of the same color is the same as the probability of drawing two balls of different colors. How many orange balls are in the bag?



$X \sim HG(N=144, M=y, K=2)$

$\Rightarrow 144 - y = z$

$y > z$

$$P(X=0,2) = P(X=1)$$

$$\frac{\cancel{\binom{y}{0}}\binom{z}{2} + \binom{y}{2}\cancel{\binom{z}{0}}}{\cancel{\binom{144}{2}}} = \frac{\binom{y}{1}\binom{z}{1}}{\cancel{\binom{144}{2}}}$$

$$\binom{z}{2} + \binom{y}{2} = yz$$

$$\frac{z!}{2!(z-2)!} + \frac{y!}{2!(y-2)!} =$$

$$\frac{z(z-1)}{2} + \frac{y(y-1)}{2} =$$

$$z(z-1) + y(y-1) = 2yz \quad \rightarrow z = 144 - y$$

$$(144-y)(143-y) + y(y-1) = 2y(144-y)$$

$$y^2 - 287y + 20592 + y^2 - y = -2y^2 + 288y$$

$$4y^2 - 576y + 20592 = 0$$

$$y^2 - 144 + 5148 =$$

$$(y-66)(y-78) =$$

$y = 78$ or $y = 66$

y must be > z

Mean and variance

- If $X \sim$ Hypergeometric (N, M, K),

$$E(X) = K\left(\frac{M}{N}\right), \qquad V(X) = K\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-K}{N-1}\right) \longrightarrow SD(X) = \sqrt{V(X)}$$

  Interpretation of the mean: We want the sample to have the same propoertion of successes as the population.

- Although the mean and variance of the hypergeometric random variable look complicated, it is actually easily understood when we compare it to those of binomial(n, p). $\hookrightarrow Y$

  $E(Y) = np \qquad V(Y) = npq$

  This switch from hypergeometric to binomial means that we are now sampling *with* replacement.

- For hypergeometric, the population size is N and there are M elements of interest.

  The probability of success is $p = \frac{M}{N}$.

  And the sample size K is like the number of trials n.

  Then the mean and variance of HG(N, M, K) can be re-expressed like this:

  In terms of Y

  $$E(X) = np, \quad n \to K,\ p \to \left(\frac{M}{N}\right)$$

  $$V(X) = npq\left(\frac{N-K}{N-1}\right), \quad n \to K,\ p \to \left(\frac{M}{N}\right),\ q \to \left(1 - \frac{M}{N}\right) = \frac{N-M}{N}$$

- Observations when comparing with vs without replacement:

  1. Means: The expected values are the same regardless of the type of sampling.
  2. Variances: The variance is somewhat smaller when sampling without replacement. $\frac{N-K}{N-1} \leq 1$

     This final term in the hypergeometric variance is often called the **finite population correlation factor**, which is an adjustment because the sampling in the hypergeometric experiment is done without replacement and is therefore not independent.

Examples

1. A standard 52 card deck contains 4 different suits, each with 13 cards. Five cards are dealt from a standard deck. Let X be the number of spades in the hand.

   (a) What is the distribution of X?

   $$X \sim HG(N=52, M=13, K=5)$$

   (b) Find the pmf of X.

   $$f(x) = \frac{\binom{13}{x}\binom{39}{5-x}}{\binom{52}{5}}, \quad x = 0, 1, \ldots, 5$$

   (c) Find the probability of two or three spades in the hand dealt.

   $$P(X=2) + P(X=3) = \sum_{x=2}^{3} \frac{\binom{13}{x}\binom{39}{5-x}}{\binom{52}{5}}$$

   (d) Find the probability of at least two spades in the hand dealt.

   $$P(X \geq 2) = \sum_{x=2}^{5} \frac{\binom{13}{x}\binom{39}{5-x}}{\binom{52}{5}}$$

   (e) Find the probability of at most 4 spades in the hand.

   $$P(X \leq 4) = \sum_{x=0}^{4} \frac{\binom{13}{x}\binom{39}{5-x}}{\binom{52}{5}} = 1 - P(X=5) = 1 - \frac{\binom{13}{5}\binom{39}{0}}{\binom{52}{5}}$$

   (f) Find $E(X)$ and $V(X)$.

   $$E(X) = \overset{= np}{5\left(\frac{13}{52}\right)} \downarrow \approx 1.25 \text{ spades}$$

   $$V(X) = \overset{= npq\left(\frac{N-n}{N-1}\right)}{5\left(\frac{13}{52}\right)\left(\frac{39}{52}\right)\left(\frac{47}{51}\right)} \downarrow \approx 0.8639$$

2. A machine shop orders 200 bolts from a supplier. To determine whether to accept the shipment of bolts, the manager of the facility randomly selects 30 bolts without replacement. If of the 30 randomly selected bolts 2 or less are found to be defective, he concludes that the shipment is acceptable.

   (a) If 20% of the bolts in the population are defective, what is the probability that the shipment is acceptable?

   $X \sim HG(N = 200, M = 0.2(200), K = 30)$, $M = 40$ $\longrightarrow$ $P(X \leq 2) = \sum_{x=0}^{2} \frac{\binom{40}{x}\binom{160}{30-x}}{\binom{200}{30}} \approx 0.0331$

   (b) Now assume the 30 bolts are chosen with replacement. If 20% of the bolts in the population are defective, what is the probability that the shipment is acceptable?

   $Y \sim \text{Binomial}(n = 30, p = 0.2)$ $\longrightarrow$ $P(X \leq 2) = \sum_{x=0}^{2} \binom{30}{x} 0.2^x \, 0.8^{30-x} = \text{Binomcdf}(30, 0.2, 2) \approx 0.0442$

Relationship between hypergeometric and binomial

- Hypergeometric experiment (X) is without replacement and binomial (Y) is with replacement.
- As the population size goes to infinity ($N \to \infty$), the mean and variance of hypergeometric(N, M, K) converges to those of binomial(n, p).

  Here's why:

  – We saw the expected values were already equivalent in the finite case. $E(X) = E(Y)$

  – The last term in the variance of HG goes to 1, and we are left with what is equivalent to the binomial variance.

  $$\lim_{N \to \infty} \frac{N-K}{N-1} = 1 \implies \lim_{N \to \infty} V(X) = V(Y) = npq$$

  Just because the means and variances are of hypergeometric converge to those of binomial as we let the sample size go to infinity (i.e. in asymptotics), we cannot say that the *random variables* are identical.

- This turns out to be a true statement, be we need to show convergence in distribution, which is a topic of graduate level probability theory.

Summary of commonly used discrete distributions so far

- We studied four distributions based on Bernoulli experiments:
  - Binomial, Geometric, Negative Binomial, and Hypergeometric.
- Throughout all of these, there were three important aspects:

  1. Number of successes
  2. number of trials
  3. probability of success

- We can organize the four distributions based on what we are interested in (the random variable) and what we are given (as parameters).
  - Distributions counting the number of successes.

    Binomial & Hypergeometric

    Interested in 1.

    2 & 3 are given as parameters.

    Only difference is with vs without replacement.
  - Distributions counting the number of trials.

    Geometric & Negative Binomial

    Interested in 2.

    1 & 3 are given as parameters.

    Only difference is the number of successes.

Poisson Distribution

Motivation

- Example experiments:
  1. The number of accidents at a particular intersection during a time period of one week.
  2. The number of earthquakes in California during a time period of one year.
  3. The number of flaws in 100 feet of wire.
- What is the common characteristic of these experiments?

  Counting the number of occurrences during a given period

Poisson experiments and Poisson random variables

- **Poisson experiments**: The Poisson distribution can be used to model several different types of experiments.

  – A scenario in which we are waiting for an occurrence (such as waiting for a bus, or waiting for customers to arrive in a bank).

  – The number of occurrences in a given time interval (such as experiments 1 and 2) or on physical objects (such as experiment 3).

  – Spatial distributions (such as the distribution of fish in a lake).

  All of these situations are about an event which is said to occur at an average rate λ per given unit (usually time period, could be area, location, etc.).

- **Poisson random variable**.
  – The random variable of interest X is the number of events in a given unit.
  – The range of X is

$$X = \{0, 1, 2, \dots\}$$

Definition

- First we will look at an example to get an idea of the types of problems that we use the Poisson distribution for.
- Example: An analyst studies data on accidents and an intersection and concludes that accidents occur there at an "average rate of $\lambda = 2$ per month".

  The number of accidents X in a month is a random variable. And the Poisson distribution can be used to find probabilities $P(X = x)$ in terms of x and λ the average rate.

- Definition: A random variable X follows the Poisson distribution with parameter (or average rate) λ if

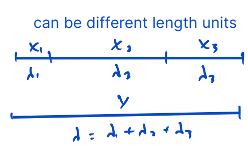
$$P(X = x \mid \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots \quad \text{and} \quad \lambda > 0$$

  Note: λ is the expected value of the number of occurrences per unit.

- Notation: $X \sim \text{Poisson}(\lambda)$
- Theorem (Combining Poisson random variables): If $X_i \overset{\perp\!\!\!\perp}{\sim} \text{Poisson}(\lambda_i)$ (not necessarily identical because λ_is can be different), then

$$Y = \sum_{i=1}^{n} X_i$$

$$\downarrow \sim \text{Poisson}\left(\lambda = \sum_{i=1}^{n} \lambda_i\right)$$



  If X_i's are identical and therefore iid, then the new parameter for Y becomes $\underline{n\lambda}$ and the number of occurrences in n units is

$$Y \sim \text{Poisson}(n\lambda)$$

Mean and variance

- If $X \sim \text{Poisson}(\lambda)$,

$$E(X) = \lambda, \qquad V(X) = \lambda$$

- It should be obvious that the expected value is equal to λ, the average rate at which events occur.

Examples

1. Continuing accidents example: The number of accidents in a month at this intersection can be modeled using the Poisson distribution with an average rate of $\lambda = 2$.

   $X \sim Poisson(\lambda = 2) \Rightarrow f(x|\lambda=2) = \frac{e^{-2} 2^x}{x!}, \ x = 0, 1, 2, \ldots$

   Now we can easily find probabilities and the mean and variance.

   (a) Find the probability there are no accidents this month.

   $P(X=0) = \frac{e^{-2} 2^0}{0!} = e^{-2} = \text{poissonpdf}(\lambda=2, x=0) \simeq 0.135$

   (b) Find the probability there is one accident this month.

   $P(X=1) = \frac{e^{-2} 2^1}{1!} = 2e^{-2} = \text{poissonpdf}(2, 1) = 0.2707$

   (c) Find the probability there are more than two accidents this month.

   $P(X>2) = 1 - P(X \le 2)$
   $= 1 - \sum_{x=0}^{2} \frac{e^{-2} 2^x}{x!}$
   $= 1 - \text{poissoncdf}(\lambda=2, x=2) \simeq 0.3233$

   (d) Suppose each accident[redacted] costs the driver \$350, find the mean and variance for the cost of accidents this month.

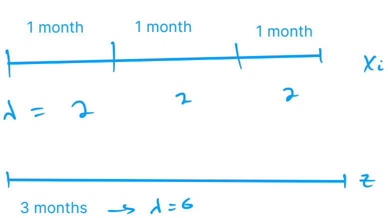
$Y = 350X \Rightarrow E(Y) = E(350X)$
$= 350 \underbrace{E(X)}_{=2}$
$= \$700$

$V(Y) = V(350X)$
$= 350^2 \underbrace{V(X)}_{=2}$
$= 245{,}000$ dollars-squared

$\not\Rightarrow$ Incorrect: $V(X) = 700$ because $Y = 350X$, $y = \{0, 350, 700, \ldots\}$, $\not\sim$ Poisson

Correct: $Z = X_1 + X_2 + \ldots + X_{350}$, $\sim$ Poisson$(\lambda = 700)$, $z = \{0, 1, 2, \ldots\}$

   (e) Find the probability there is four accidents in a three month span.



$\rightarrow$ let z = number of accidents in 3 months
$\sim$ Poisson$(\lambda = 6)$

$\rightarrow P(Z=4) = \frac{e^{-6} 6^4}{4!} = \text{Poissonpdf}(\lambda=6, x=4) \simeq 0.1339$

2. Assume the number of hits, X, per baseball game has a Poisson distribution. If the probability of a no-hit game is 1/10000, find the probability of having 4 or more hits in a particular game.

$\rightarrow X \sim \text{Poisson}(\lambda)$

$\hookrightarrow = ??$ solve for λ using known info

$\rightarrow P(X=0) = \frac{1}{10000}$

$\frac{e^{-\lambda}\cancel{\lambda^0}}{\cancel{0!}} = \frac{1}{10000}$

$\ln(e^{-\lambda}) = \ln\left(\frac{1}{10000}\right)$

$-\lambda = \ln\left(\frac{1}{10000}\right)$

$\Rightarrow \lambda = \ln(10000)$

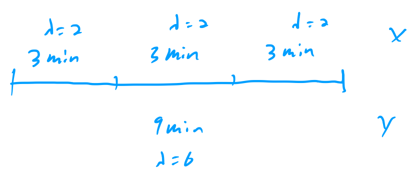
$\rightarrow P(X \geq 4) = 1 - P(X \leq 3)$

$= 1 - \text{Poissoncdf}(\lambda = \ln(10000), x = 3)$

$= 1 - \sum_{x=0}^{3} \frac{e^{-\ln(10000)} (\ln(10000))^x}{x!}$

≈ 0.9817

3. In a large city, telephone calls to 911 come on the average of two every 3 minutes. If the city assumes an approximate Poisson process, what is the probability of five or more calls arriving in a 9-minute period?



$\rightarrow X \sim \text{Poisson}(\lambda = 2)$

$\hookrightarrow$ 3 min

$\rightarrow Y \sim \text{Poisson}(\lambda = 3 \cdot 2 = 6)$

$\hookrightarrow$ 9 min

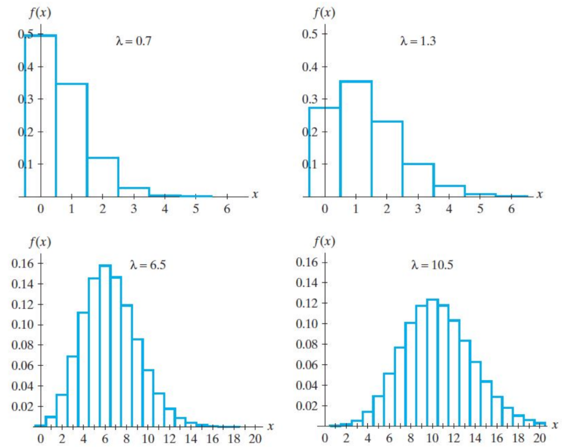
$\rightarrow P(Y \geq 5) = 1 - P(Y \leq 4)$

$= 1 - \text{poissoncdf}(\lambda = 6, x = 4)$

$= 1 - \sum_{y=0}^{4} \frac{e^{-6} 6^y}{y!}$

≈ 0.7149

Visualizing Poisson distributions



Poisson approximation to the binomial distribution

- In grad school :), you will learn that Poisson is the is the limiting distribution of binomial. Said slightly more formally:

$$\lim_{n \to \infty} \text{Binomial}(n, p = \frac{\lambda}{n}) = \text{Poisson}(\lambda = np)$$

- This means that for large n and small p, the Poisson distribution can be used to approximate the binomial distribution.

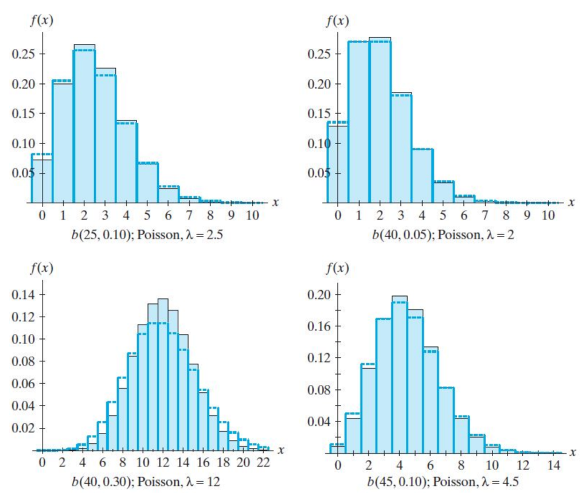


Figure 2.6-2 Binomial (shaded) and Poisson probability histograms