Chapter 4 part 3

Discrete Random Variables

Jingchen (Monika) Hu

Vassar College

MATH 241

Outline

Count the number of ...

In beer brewing, cultures of yeast are kept alive in jars of fluid before being put into the mash.

- It's critical to control the amount of yeast used.
- Number of yeast cells in a fluid sample can be seen under a microscopes.
- Yeast cells are constantly multiplying and dividing.
- A famous statistician, Wiliam Sealy Gosset (aka "Student"), who worked for the Guinness Brewing Compnay in early 1900's, modeled the counts of yeast cells using the *Poisson distribution*.

The Poisson distribution was used in 1898 to count the number of soldiers in the Prussian Army who died accidentally from horse kicks

Poisson distribution

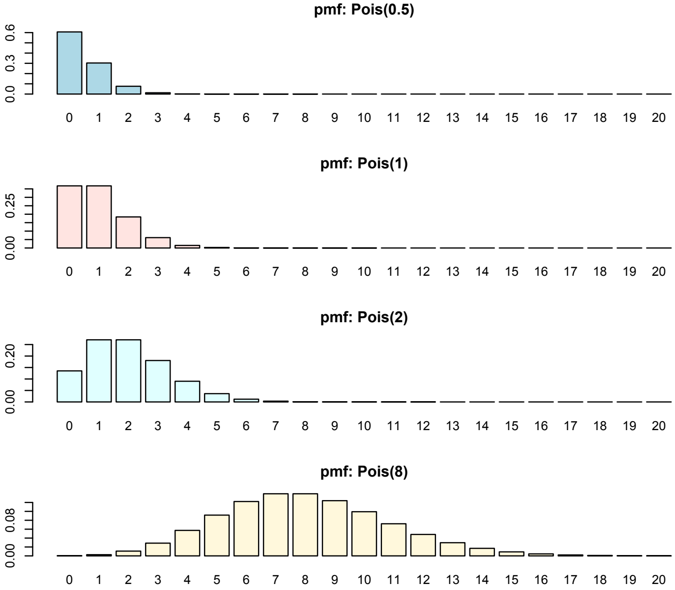
Definition

Denote random variable X that takes value in $\{0, 1, 2, \dots\}$ as having a Poisson distribution with parameter λ if its pmf is

$$X \sim P(\lambda) \iff p(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

- Formulated by French mathematician Siméon Denis Poisson
- Usually is used to model "the number of xxx occur". Hence lower bound is 0, no upper bound.
- Examples
  - Number of rainy days this year
  - Number of mis-placed books in the Main library
  - Number of roses you will receive on the next Valentines Day

Pmf of Poisson distribution



Properties of Poisson distribution $p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$

- Well-defined (validness of pmf): non-negative, and

$$\sum_{k=0}^{\infty} p(k) = 1$$

- Taylor Series

$$f(x) = f(x_0) + \frac{x - x_0}{1!} f'(x_0) + \frac{(x - x_0)^2}{2!} f''(x_0) + \cdots$$

- Use Taylor series to verify that the Poisson distribution is well-defined (required)

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \cdots + \frac{\lambda^k}{k!} + \cdots$$

Properties of Poisson distribution $p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$

- Mean (required)

$$E[X] = \lambda$$

- Variance (required)

$$Var[X] = \lambda$$

Question

Find the probability that a randomly selected Vassar student has odd number of siblings, if the average number of siblings is 1.73.

Question

Find the probability that a randomly selected Vassar student has odd number of siblings, if the average number of siblings is 1.73.

$$X \sim \mathsf{P}(\lambda), \lambda = 1.73$$

$$\begin{aligned} P(X \text{ is an odd number}) &= p(1) + p(3) + p(5) + \cdots \\ &= e^{-\lambda}\frac{\lambda^1}{1!} + e^{-\lambda}\frac{\lambda^3}{3!} + e^{-\lambda}\frac{\lambda^5}{5!} + \cdots \end{aligned}$$

Question

Find the probability that a randomly selected Vassar student has odd number of siblings, if the average number of siblings is 1.73.

$$X \sim \mathsf{P}(\lambda), \lambda = 1.73$$

$$\begin{aligned} P(X \text{ is an odd number}) &= p(1) + p(3) + p(5) + \cdots \\ &= e^{-\lambda}\frac{\lambda^1}{1!} + e^{-\lambda}\frac{\lambda^3}{3!} + e^{-\lambda}\frac{\lambda^5}{5!} + \cdots \end{aligned}$$

Note that

$$\begin{aligned} e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots \\ e^{-x} &= 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} - \cdots \end{aligned}$$

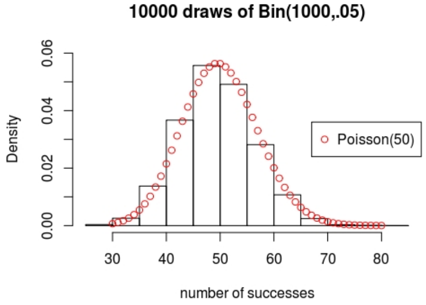
$$P(X \text{ is an odd number}) = e^{-\lambda}\left(\frac{e^{\lambda} - e^{-\lambda}}{2}\right) = \frac{1 - e^{-2\lambda}}{2} = 0.4843$$

Use Poisson to approximate Binomial distribution

Let $X \sim \text{Bin}(n, p)$. If

- p: small
- n: large
- $\lambda = np$: of moderate size

then the distribution of X can be approximated by $\text{P}(\lambda)$.



Question

If you buy a lottery ticket in 50 lotteries, in each of which your chance of winning a prize is 1 / 100, what is the (approximate) probability that you will win a prize (a) at least once? (b) exactly once? (c) at least twice?

Question

If you buy a lottery ticket in 50 lotteries, in each of which your chance of winning a prize is 1 / 100, what is the (approximate) probability that you will win a prize (a) at least once? (b) exactly once? (c) at least twice?

We can use P(λ) to approximate Bin(n, p), if p is small n is large and $\lambda = np$ is of moderate size.

Here we have $X \sim \text{Bin}(50, 0.01)$ with $np = 0.5$. Therefore, we can use P(λ) = P(0.5) for the approximation of Bin(50, 0.01).

- Part (a):
  $P(X >= 1) = 1 - P(X < 1) = 1 - P(X = 0) \approx 1 - e^{-0.5}\frac{0.5^0}{0!} = 1 - e^{-0.5}$
- Part (b):
  $P(X = 1) \approx 1 - e^{-0.5}\frac{0.5^1}{1!} = 1 - e^{-0.5} \times 0.5$
- Part (c):
  $P(X >= 2) = 1 - P(X < 2) = 1 - P(X <= 1) = 1 - P(X = 0) - P(X = 1) \approx 1 - e^{-0.5}\frac{0.5^0}{0!} - e^{-0.5}\frac{0.5^1}{1!} = 1 - e^{-0.5} \times 1.5$

Recap

Poisson distribution $X \sim \mathsf{P}(\lambda)$

$$p(k) = e^{-\lambda}\ \frac{\lambda^k}{k!}$$

- mean $\mu = \lambda$
- variance $\sigma^2 = \lambda$
- Approximate Binomial distribution with small p, large n, moderate np

$$\mathsf{P}(np) \approx \mathsf{Bin}(n, p)$$

Question

$X \sim \mathrm{P}(\lambda)$. Which is the following is FALSE?

a. The mean and standard deviation of X are different.
b. Pmf of X can be a decreasing function.
c. λ can only take values $0, 1, 2, \ldots$.
d. None of the above.

Question

$X \sim \mathrm{P}(\lambda)$. Which is the following is FALSE?

a. The mean and standard deviation of X are different.
b. Pmf of X can be a decreasing function.
c. *λ can only take values $0, 1, 2, \ldots$.*
d. None of the above.

Outline

Geometric distribution

A gambler plays at a roulette table and alway bet on red until he wins...
In each round, his chance of winning is $18/38 = 0.47$.
Let X denote the number of rounds he plays.

Definition

Denote random variable X that takes value in $\{1, 2, \ldots\}$ as having a Geometric distribution with parameter $p \in (0, 1)$ if its pmf is

$$X \sim \textit{Geometric}(p) \iff p(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots$$

- X represents the number of trials performed until we get a success, where p is the probability of success on each trial.
- Note the difference between Geometric distribution and Binomial distribution! Eg: whether the total number of trials is fixed.

Properties of Geometric distribution $p(k) = (1-p)^{k-1}p$

- Well-defined (validness of pmf): non-negative, $\sum_{k=1}^{\infty} p(k) = 1$ (required)

$$\sum_{k=1}^{\infty}(1-p)^{k-1}p = p[1 + (1-p) + (1-p)^2 + \cdots] = \frac{p}{1-(1-p)} = 1$$

Properties of Geometric distribution $p(k) = (1-p)^{k-1}p$

- Well-defined (validness of pmf): non-negative, $\sum_{k=1}^{\infty} p(k) = 1$ (required)

$$\sum_{k=1}^{\infty}(1-p)^{k-1}p = p[1 + (1-p) + (1-p)^2 + \cdots] = \frac{p}{1-(1-p)} = 1$$

- Cdf: $P(X \leq k) = 1 - (1-p)^k$ (required)

$$P(X \geq k+1) = P(\text{The first } k \text{ trials all fail})$$

- Mean (not required) Textbook page 148 for derivation

$$E[X] = \frac{1}{p}$$

- Variance (not required) Textbook page 148 for derivation

$$Var[X] = \frac{1-p}{p^2}$$

Gambler's fallacy

If the gambler loses 5 times in a row, will he more likely to win in the 6th round?

$$P(X > 6 \mid X > 5) \overset{?}{<} P(X > 1)$$

Unfortunately, not.

Definition

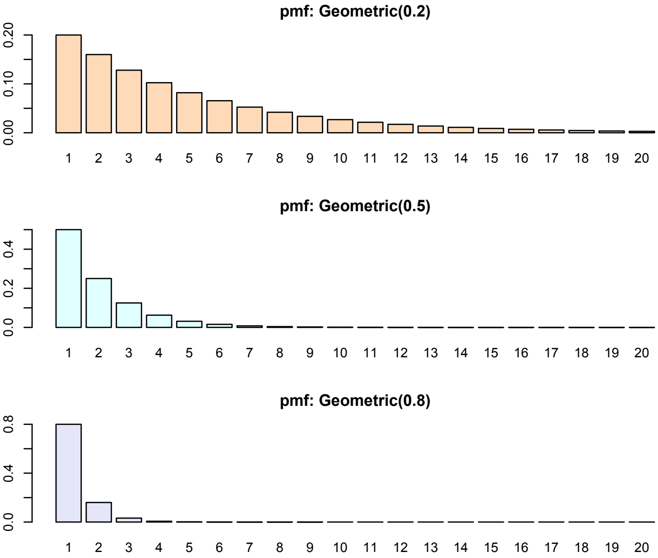
We say a distribution is memoryless, if

$$P(X > n + k \mid X > n) = P(X > k)$$

- Geometric random variable is memoryless.

$$P(X > n + k \mid X > n) = \frac{P(X > n + k)}{P(X > n)}$$
$$= \frac{(1-p)^{n+k}}{(1-p)^n} = (1-p)^k = P(X > k)$$

Pmf of Geometric distribution



Question

X denotes the number of times a die is rolled until 6 is obtained.

1. What are the odds we have to roll it 10 or more times?
2. How many times do we expect to roll?
3. Find $Var[X]$.

Question

X denotes the number of times a die is rolled until 6 is obtained.

1. What are the odds we have to roll it 10 or more times?
2. How many times do we expect to roll?
3. Find $Var[X]$.

1. $X \sim$ Geometric$(p = 1/6)$, and

$$P(X \geq 10) = 1 - P(X \leq 9) = 1 - [1 - (1-p)^9] = (5/6)^9$$

Question

X denotes the number of times a die is rolled until 6 is obtained.

1. What are the odds we have to roll it 10 or more times?
2. How many times do we expect to roll?
3. Find $Var[X]$.

1. $X \sim$ Geometric($p = 1/6$), and

$$P(X \geq 10) = 1 - P(X \leq 9) = 1 - [1 - (1-p)^9] = (5/6)^9$$

2. Mean of Geometric distribution

$$E[X] = 1/p = 6$$

Question

X denotes the number of times a die is rolled until 6 is obtained.

1. What are the odds we have to roll it 10 or more times?
2. How many times do we expect to roll?
3. Find $Var[X]$.

---

1. $X \sim$ Geometric$(p = 1/6)$, and

$$P(X \geq 10) = 1 - P(X \leq 9) = 1 - [1 - (1-p)^9] = (5/6)^9$$

2. Mean of Geometric distribution

$$E[X] = 1/p = 6$$

3. Variance

$$Var[X] = \frac{1-p}{p^2} = \frac{\frac{5}{6}}{\frac{1}{6} \cdot \frac{1}{6}} = 30$$

Negative Binomial distribution

Definition

Denote random variable X that takes value in $\{1, 2, \ldots\}$ as having a Negative Binomial distribution with parameter $p \in (0, 1)$ if its pmf is

$$X \sim NB(r, p) \iff p(k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r, \quad k = r, r+1, \ldots$$

- X represents the number of trials performed until we get r success, where p is the probability of success on each trial.
- Well-defined (validness of pmf). (not required)
- Connection between Negative Binomial and Geometric distributions

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

Note: Proof of $\sum_{k=r}^{\infty} p(k) = 1$ for Negative Binomial distribution is not required.

Properties of Neg Binom $p(k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r$

- Mean (not required)

$$E[X] = \frac{r}{p}$$

  Textbook page 150 for derivation

- Variance (not required)

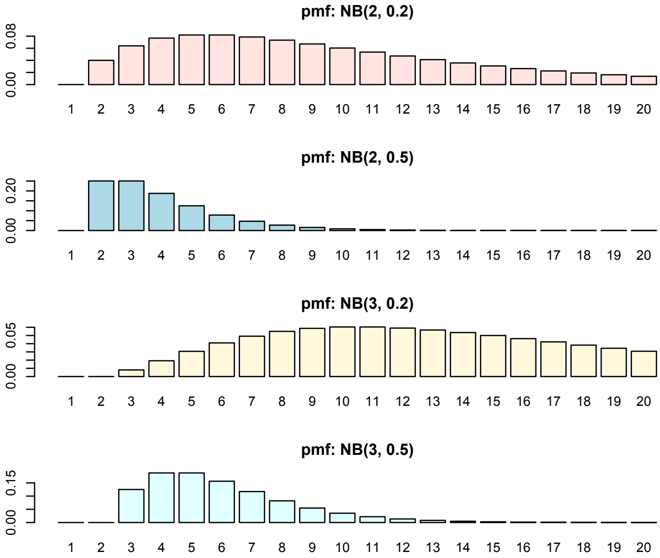
$$Var[X] = \frac{r(1-p)}{p^2}$$

  Textbook page 151 for derivation

- Recall that for Geometric(p),

$$\mu = \frac{1}{p}, \sigma^2 = \frac{1-p}{p^2}$$

Pmf of Negative Binomial distribution



Question

Consider independent trials with success probability p. Let $q = 1 - p$. What's the probability of getting r successes before m failures?

- (a) $p^{r-1} q^m$
- (b) $\binom{r+m-1}{r-1} p^{r-1} q^m$
- (c) $\sum_{k=r}^{r+m} \binom{k-1}{r-1} p^r q^{k-r}$
- (d) $\sum_{k=r}^{r+m-1} \binom{k-1}{r-1} p^r q^{k-r}$

Question

Consider independent trials with success probability p. Let $q = 1 - p$. What's the probability of getting r successes before m failures?

(a) $p^{r-1}q^m$
(b) $\binom{r+m-1}{r-1}p^{r-1}q^m$
(c) $\sum_{k=r}^{r+m} \binom{k-1}{r-1}p^r q^{k-r}$
(d) $\sum_{k=r}^{r+m-1} \binom{k-1}{r-1}p^r q^{k-r}$

Let X denote the number of trials needed to get r successes. Then $X \sim \text{NB}(r, p)$.

$$
\begin{aligned}
&P(r \text{ successes before } m \text{ failures})\\
=&P(r^{\text{th}} \text{ success occurs on trials } r, r+1, \ldots, r+m-1)\\
=&P(X=r) + P(X=r+1) + \cdots + P(X=r+m-1)
\end{aligned}
$$

Recap

X: the number of trials performed until we get r success, where p is the probability of success on each trial.

$$p(k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r, \quad k = r, r+1, \ldots$$

- Negative Binomial distribution $X \sim \text{NB}(r, p)$
- Mean $\mu = \frac{r}{p}$, variance $\sigma^2 = \frac{r(1-p)}{p^2}$.
- If $r = 1$, Geometric distribution $X \sim \text{NB}(1, p) = \text{Geometric}(p)$
- Geometric distribution is memoryless.

Review: discrete distributions

Name	Range	pmf $p(x)$	mean	variance
Ber(p)	$\{0, 1\}$	$p^x(1-p)^{1-x}$	p	$p(1-p)$
Bin(n, p)	$\{0, 1, \ldots, n\}$	$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$
Pois(λ)	$\{0, 1, 2, \ldots\}$	$e^{-\lambda} \frac{\lambda^x}{x!}$	λ	λ
Geometric(p)	$\{1, 2, \ldots\}$	$(1-p)^{x-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
NegBin(r, p)	$\{r, r+1, \ldots\}$	$\binom{x-1}{r-1}(1-p)^{x-r}p^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$