Chapter 1.7 Binomial, Poisson and Geometric Distributions

1.7.1 Binomial Distribution

1.7.2 Geometric Distribution

1.7.3 Poisson Distribution

1.7.4 Summary

17 March 2018

1.7.1 Binomial Distribution

The **Binomial distribution** with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent trials, each of which yields success with probability p.

We write $X \sim \mathrm{Bin}(n, p)$ for a random variable that has Binomial distribution with n trials and probability of success p. The outcome for each trial is success/failure. X takes the value of $0, 1, 2, \cdots, n$.

1.7.1 Binomial Distribution: example

Examples include tossing a fair coin n times. In each trial, we have 2 outcomes: success or failures. The probability of success does not change from one trial to another.

The probability of any combination of 2 heads is $\left(\frac{1}{2}\right)^2\left(1-\frac{1}{2}\right)^{n-2}$

There are $\binom{n}{2}$ possible combinations of 2 heads in n tosses, therefore

$$P(2H|n\ tosses) = \binom{n}{2}\left(\frac{1}{2}\right)^2\left(1-\frac{1}{2}\right)^{n-2} = \binom{n}{2}\left(\frac{1}{2}\right)^n$$

1.7.1 Binomial Distribution

In general if p is the probability of an event (we call this event a *success*) and we repeat n independent trials, the of getting exactly k successes in n trials is

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}; \quad k = 0, 1, \cdots, n$$

where X denotes the number of successes and is called ***Binomial variable***.

1.7.1 Binomial Distribution

If X is Binomial

its mean is

$$\mu = E(X) = np$$

and the variance is

$$\sigma^2 = Var(X) = np(1-p) = (1-p)E(X)$$

Note We usually denote $q = 1 - p$.

1.7.1 Binomial Distribution

Example 1

Compute the probability of obtaining <u>at least</u> two '6' in rolling a fair die 4 times. [*Hint*: $p = \frac{1}{6}, n = 4, k = 2$]

Solution

Let $X \sim \mathrm{Bin}\left(4, \frac{1}{6}\right)$. Required probability is

$P(X \geq 2) = \sum_{k=2}^{4} \binom{4}{k} p^k (1-p)^{n-k} = P(X = 2) + P(X = 3) + P(X = 4) =$

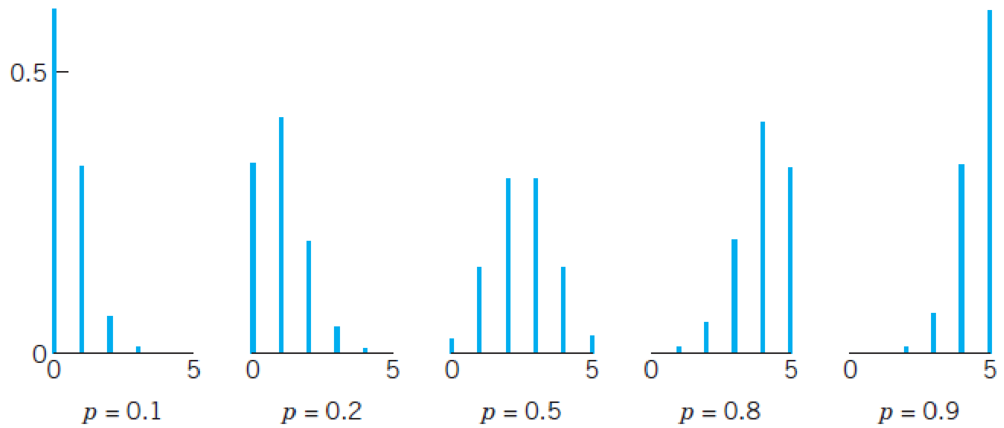
$= \binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2 + \binom{4}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^1 + \binom{4}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^0 = \frac{171}{1296}$

$= 0.1320$ ■

1.7.1 Binomial Distribution

Example 2

If $X \sim \mathrm{Bin}(n, p)$ then as k goes from 0 to n, the probability $P(X = k)$ will first increase monotonically and then decrease monotonically, reaching its largest value when k is the largest integer less than or equal to $(n+1)p$.



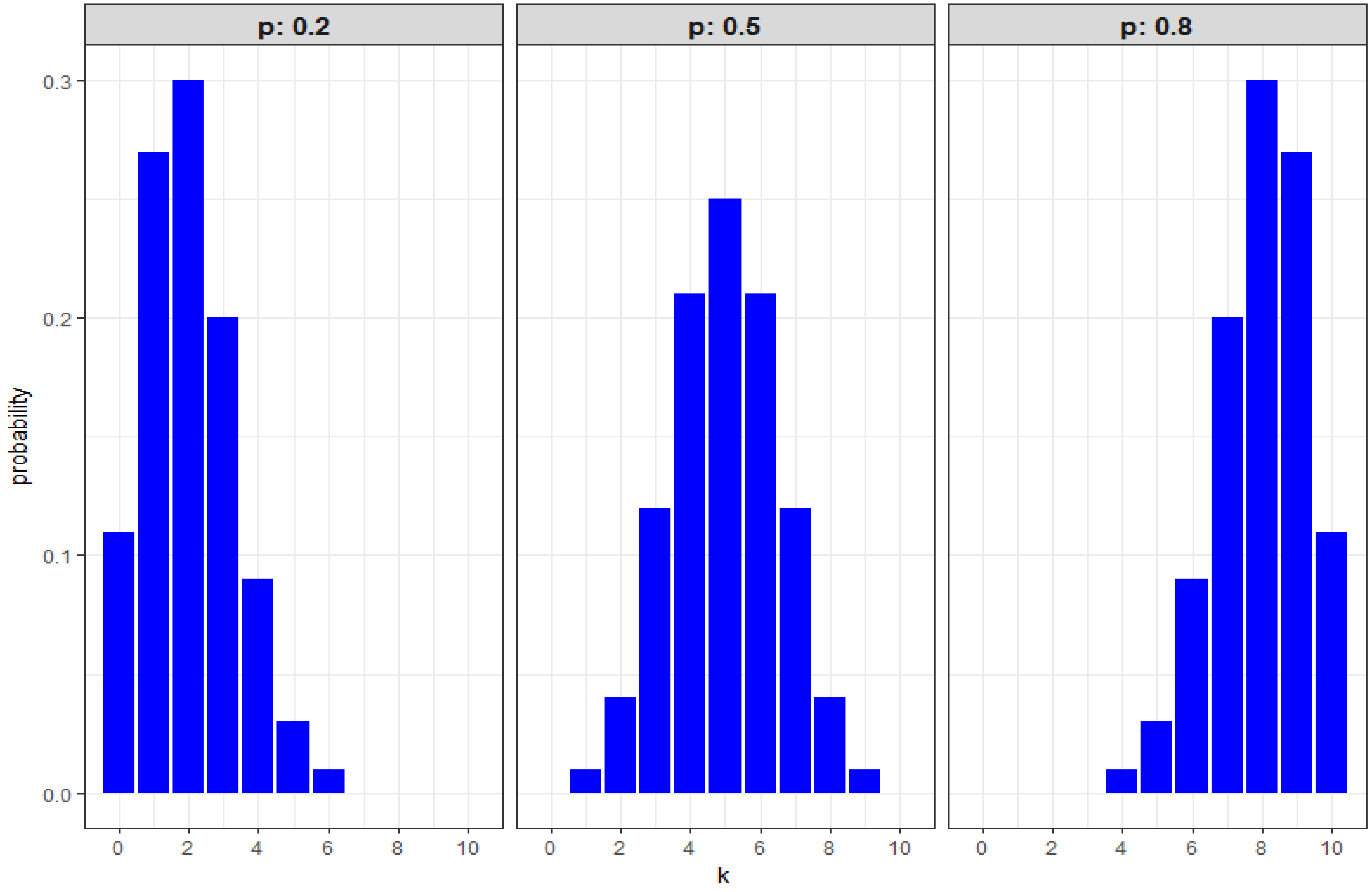
Probability function (2) of the binomial distribution for $n = 5$ and various values of p

1.7.1 Binomial Distribution

The Binomial distribution is symmetric for p= 0.5 and skewed for other values.

Example $n = 10$

$n + 1 = 11$



1.7.1 Binomial Distribution

Example 3

The random variable $X \sim \mathrm{Bin}(n, p)$ where $0 < p < 1$. Given that $\mathrm{Var}(X) = \frac{1}{5}\mathrm{E}(X)$, find the least value of n such that $P(X \geq 1) > 0.95$.

Solution

Solving $\mathrm{Var(X)} = np(1-p) = \frac{1}{5}E(X) = \frac{1}{5}np$, we have $p = \frac{4}{5}$. So, we need n for which $P(X \geq 1) = 1 - P(X = 0) > 0.95$.

1.7.1 Binomial Distribution

$\therefore P(X = 0) = \left(\frac{1}{5}\right)^n < 0.05$

Taking log we obtain $n\ ln\left(\frac{1}{5}\right) > \ln(0.05)$; rearranging, we have

$$n > \frac{ln(0.05)}{\ln\left(\frac{1}{5}\right)} = 1.86$$

Hence minimum n is 2. ■

[for computing, recall that $\ln(0.05) = \ln\left(\frac{1}{20}\right) = -\ln(20)$ and $\ln\left(\frac{1}{5}\right) = -\ln(5)$. So, $\frac{ln(0.05)}{\ln(^1/_5)} = \ln(20)\ /\ln(5)$]

1.7.1 Binomial Distribution: problem

- There are 20 racers. The probability that one finishes the race is 0.9. Find the probability that
  1. 12 racers finish the race
  2. Less than 19 finish the race

1.7.2 Geometric Distribution

The **geometric distribution** is a discrete probability distribution of the number X of independent trials needed in order to get the **first success**. We assume that each trial has constant probability $p \in (0,1)$ of a success.

So, let's denote $S = success$ and $F = failure$, it is the probability of $k - 1$ failures before a success.

For example, if $X = \ number\ of\ failures$ and $p = prob. of\ success,\ q = 1 - p = prob. of\ failure$,

$$P(X = 4) = P(FFFS) = P(F)P(F)P(F)P(S) = q^3 p$$

There is only one possible order: first 3 failures and then 1 success

1.7.2 Geometric Distribution

We write $X \sim \mathrm{Geo}(p)$ for a random variable that has Geometric distribution which is the number of trials before the 1^{st} *success.* X takes the value of $1, 2, \cdots$.

$X = 1$, means success at the first try, $X = 2$ at the second try, and so on.

1.7.2 Geometric Distribution

To have the first success at the k^{th} trial, we need to have first $(k-1)$ failures and then a success. Therefore, letting $q = 1 - p$, the probability of getting the 1^{st} success in k^{th} trials is

$$P(X = k) = p(1-p)^{k-1} = pq^{k-1}; \quad k = 1, 2, \cdots$$

The cdf is found as

$$P(X \le k) = \sum_{i=1}^{k}(1-p)^{i-1}p = p\sum_{i=0}^{k-1} q^i = p\,\frac{1-q^k}{1-q} = \ 1 - q^k$$

Recall that $\sum_{i=0}^{n} r^i = \frac{1-r^{n+1}}{1-r}$

1.7.2 Geometric Distribution

The expected number of trials for the first success is the mean

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

The variance is $\mathrm{Var}(\mathrm{X}) = \sigma^2 = \frac{1-p}{p^2}$.

1.7.2 Geometric Distribution

The geometric distribution is *memoryless*:

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

Proof. We know that $P(X \leq k) = 1 - q^k$, so

$$P(X > k) = q^k$$

Hence,

$$P(X > k + i | X > i) = \frac{q^{k+1}}{q^i} = q^k = P(X > k)$$

1.7.2 Geometric Distribution

Example 7

A box contains N white and M black balls. Balls are selected randomly, one at a time and replaced, until a black one is obtained. If we assume that each ball is replaced before the next one is drawn, what is the probability that:

i. exactly k draws are needed?

ii. at least k draws are needed?

1.7.2 Geometric Distribution

Solution

The probability of a black ball is $p = \frac{M}{M+N}$, for white it is $q = 1 - p = \frac{N}{M+N}$

i. Required probability is

$$P(X = k) = p(1-p)^{k-1} = \frac{MN^{k-1}}{(M+N)^k}.$$

ii. Required probability is

$$P(X \geq k) = 1 - P(X \leq k-1) = 1 - \left(1 - q^{k-1}\right) = q^{k-1}$$

1.7.3 Poisson Distribution

The **Poisson distribution** is a discrete probability distribution that gives the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event.

We write $X \sim \mathrm{Poi}(\lambda)$ for a random variable that has Poisson distribution with rate of occurring in a fixed interval of time and/or space. X takes the value of $0, 1, 2, \cdots$.

1.7.3 Poisson Distribution

Example 4

The Poisson distribution may be useful to model events such as

1. The number of goals scored in a World Cup soccer match
2. The number of meteors greater than 1 meter diameter that strike earth per year
3. The number of occurrences of the DNA sequence "ACGT" in a gene
4. The number of patients arriving in an emergency room between 11 and 12 pm ■

1.7.3 Poisson Distribution

When is the Poisson distribution an appropriate model?

The Poisson distribution is appropriate if the following assumptions are true:

- X is the number of times an event occurs in an interval and X can take values $0,1,2,\cdots$.
- Events occur randomly, independently and singly
- The rate at which events occur is constant. The rate cannot be higher in some intervals and lower in other intervals.
- The probability of an event in an interval is proportional to the length of the interval.

If these conditions are true, then X is a Poisson random variable, and the distribution of X is a Poisson distribution.

1.7.3 Poisson Distribution

Example 5

The Poisson assumptions are violated in the following cases:

- The number of students who arrive at the lecture hall will likely not follow a Poisson distribution, because the **rate is not constant** (low rate during class time, high rate between class times) and the arrivals of individual students are not independent (students tend to come in groups).
- The number of magnitude 5 earthquakes per year in China may not follow a Poisson distribution if one large earthquake increases the probability of aftershocks of similar magnitude.
- Among patients admitted to the intensive care unit of a hospital, the number of days that the patients spend in the ICU is not Poisson distributed because the number of days cannot be zero. ■

1.7.3 Poisson Distribution

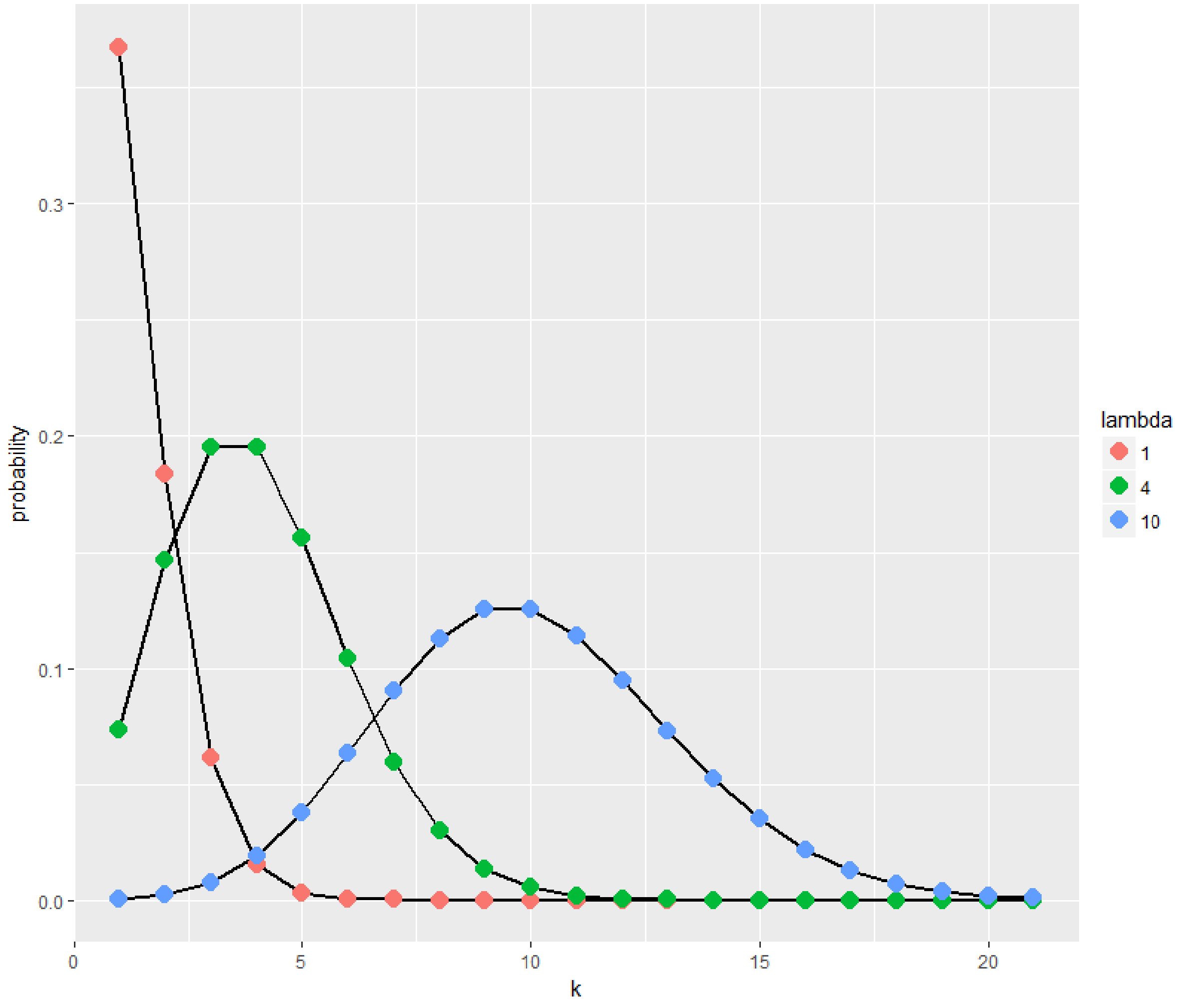
The probability of getting exactly k events in a specified interval with rate λ is

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad ; \quad k = 0,1,2,\cdots$$

with mean $\mathbf{E(X) = \lambda}$ and variance $\mathbf{Var(X) = \lambda}$.

1.7.3 Poisson Distribution

The shape of the Poisson pmf changes with λ , but converges for large λ



1.7.3 Poisson Distribution

The cdf of a Poisson is the sum of the probabilities

$$F(k) = P(X \le k) = \sum_{i=0}^{k} \frac{e^{-\lambda}\lambda^{k}}{k!} \;; \quad k = 0, 1, 2, \cdots$$

$$F(0) = P(X = 0) = e^{-\lambda}, F(1) = e^{-\lambda} + \lambda e^{-\lambda} = e^{-\lambda}(1 + \lambda)$$

So, the probability of at least one event is

$$P(X > 0) = 1 - P(X = 0) = 1 - e^{-\lambda}$$

1.7.3 Poisson Distribution

Example 6

Suppose that earthquakes occur at the rate of 2 per week. By modeling this as a Poisson distribution, find:

i. The probability of 3 earthquakes in the next week;
ii. the probability that at least 2 earthquakes occur during the next week.

1.7.2 Poisson Distribution

Solution

Since the rate is $\lambda = 2$ earthquakes per week

i. $X \sim \mathrm{Poi}(2)$, so $P(X = 3) = \frac{e^{-2}2^3}{3!} = \frac{8e^{-2}}{6} = 0.18045$

ii. The required probability is

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) =$$

$$1 - \frac{e^{-2}2^0}{0!} - \frac{e^{-2}2^1}{1!} = 1 - e^{-2} - 2e^{-2}$$

$$= 1 - 3e^{-2} \approx 0.594 \;\blacksquare$$

1.7.3 Poisson Distribution: problem

Suppose that calls to a switchboard arrive as a Poisson variable with rate 2.5 per minute. Find the probability:

1. that in the next minute there won’t be a call
2. That in the next minute there won’t be more than 1 call.

Also find the average number of calls per minute.

1.7.3 Poisson Distribution: problem

Suppose that calls to a switchboard arrive as a Poisson variable with rate 2.5 per minute. Find the probability:

1. that in the next minute there won't be a call

$$P(x = 0) = e^{-2.5} \approx 0.082$$

2. That in the next minute there won't be more than 1 call.

$$P(x \leq 1) = P(x = 0) + P(x = 1) = e^{-2.5} + 2.5e^{-2.5} \approx 0.288$$

Also find the average number of calls per minute.

The average number of calls is the mean of the variable which is equal to the rate. Hence $E(X) = \lambda = 2.5$.

1.7.4 Summary

1.7.1 Binomial Distribution

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

1.7.2 Geometric Distribution

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

1.7.3 Poisson Distribution

$$P(X = k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$$