cdf: $P(X \le x)$

Chapter 6 Some Probability Models

In this chapter, we will cover the following discrete random variables:

$P(X = a) =$

1) Bernoulli trials
2) Geometric r.v.
3) Binomial r.v.
4) Poisson r.v.

Recall: in the discrete case:

$P(X < x) \neq P(X \le x)$

$*$ $P(a < X \le b) = F(b) - F(a)$ $*$

X = # defective items in 10 items.

$P(\underbrace{2 \le X \le 4}_{\{2,3,4\}}) = P(1 < X \le 4) = F(4) - F(1)$

$= \underbrace{P(X \le 4)}_{\{0,1,\dots,4\}} - \underbrace{P(X \le 1)}_{\{0,1\}}$

Bernoulli Random Variable

A random variable that takes value 1 in case of success and 0 in case of failure.

$$\text{X(Success)} = 1$$
$$\text{X(Failure)} = 0$$

where $P(\text{success}) = p$ and $P(\text{failure}) = 1 - p = q$

$$X \sim Ber(p)$$

pmf: $$P(X = x) = (1-p)^{1-x} p^x, \qquad x = 0, 1$$

where p is the probability of success.

x	0	1
$P(X=x)$	$(1-p)$	p

with mean $E(X) = p$ and $Var(X) = p(1-p)$.

$$E(X) = 0 \cdot (1-p) + 1p = p$$

$$Var(X) = E(X^2) - E(X)^2$$

$$E(X^2) = 0^2(1-p) + 1^2 p$$

$$= p$$

$$Var(X) = p - p^2 = p(1-p)$$

Geometric Random Variables

Suppose we want to model how long it will take to achieve the first success in a series of Bernoulli trials.

A **Geometric random variable** counts the number of independent trials needed until the first success occurs.

X = # days til flood

0 0 0 0 1

$\underbrace{0\ 0\ 0\ 0}_{(x-1)}$ 1 ← xth day

$$X \sim Geo(p)$$

pmf: $P(X = x) = (1-p)^{x-1}p, \qquad x = 1, 2, 3...$

where p is the probability of success.

cdf: $F(x) = 1 - (1-p)^x$

Mean and Variance of Geometric Random Variable

$$E(X) = \frac{1}{p\uparrow} \downarrow \quad \text{(called the } \textbf{return period}\text{)}$$

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

CDF of a Geometric Random Variable

Suppose $X \sim Geo(p)$. We want to show:

$$F(x) = 1 - (1-p)^x$$

We will need to make use of the formula for the sum of a geometric series for the proof:

$$\boxed{[1 + r + r^2 + ...] = \frac{1}{(1-r)}}$$

where $0 < r < 1$.

Proof.

$$F(x) = P(X \le x) = 1 - P(X > x) \qquad X \sim geo(p)$$

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

$$P(X > x) = \sum_{k=x+1}^{\infty} (1-p)^{k-1} p$$

$$= (1-p)^{x} p + (1-p)^{x+1} p + (1-p)^{x+2} p + \dots$$

$$= p(1-p)^{x} \left[1 + (1-p) + (1-p)^2 + \dots \right]$$

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

$$F(x) = 1 - P(X > x)$$

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

Example 1

At a certain plant, experiments were conducted to reduce the proportion of cells being scrapped due to internal shorts. The experiments reduced the percentage of manufactured cells with internal shorts to around 1%. Suppose that testing of cells for shorts begins on a production run in this plant.

(a) What is the probability the second cell inspected will be the first short discovered?

(b) What is the probability that at least 50 cells are tested without finding a short?

(c) How many cells should the plant expect to test until they encounter the first internal short?

X = the # of the test at which the 1st short discovered.

$X \sim geo(p)$ if independent

$= 0.01$ ← prob. any particular test yields shorted cell.

a) $P(X=2) = (1-p)^{2-1} p$

$= \underbrace{(1-0.01)}_{\text{no short}} \, 0.01 = 0.0099.$

b) P(at least 50 cells without finding short)

$= P(X > 50)$

Recall : cdf of geo. r.v.

$F(x) = 1 - (1-p)^x$

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

$P(X > x) = 1 - F(x)$

$= (1-p)^x$

$P(X > 50) = (1-p)^{50} = 0.61$

c) $E(X) = \frac{1}{p} = \frac{1}{0.01} = 100$ batteries.

Expect to wait through 100 tests on average to encounter the 1st internal short.

Binomial Random Variables $X \sim Bin(n, p)$

The Binomial Model is also based on the idea of Bernoulli Trials. A **Binomial random variable** is the number of successes for n independent trials (fixed number of trials). A Binomial model tells us the probability for a random variable that counts the number of successes in a fixed number of Bernoulli trials.

HHTTT
HTHTT

$$X \sim Bin(n, p)$$

$$P(X = x) = \binom{n}{x} \underbrace{p^x(1-p)^{n-x}}_{\text{prob. each of outcomes}}, \qquad x = 0, 1, 2, ..., n$$

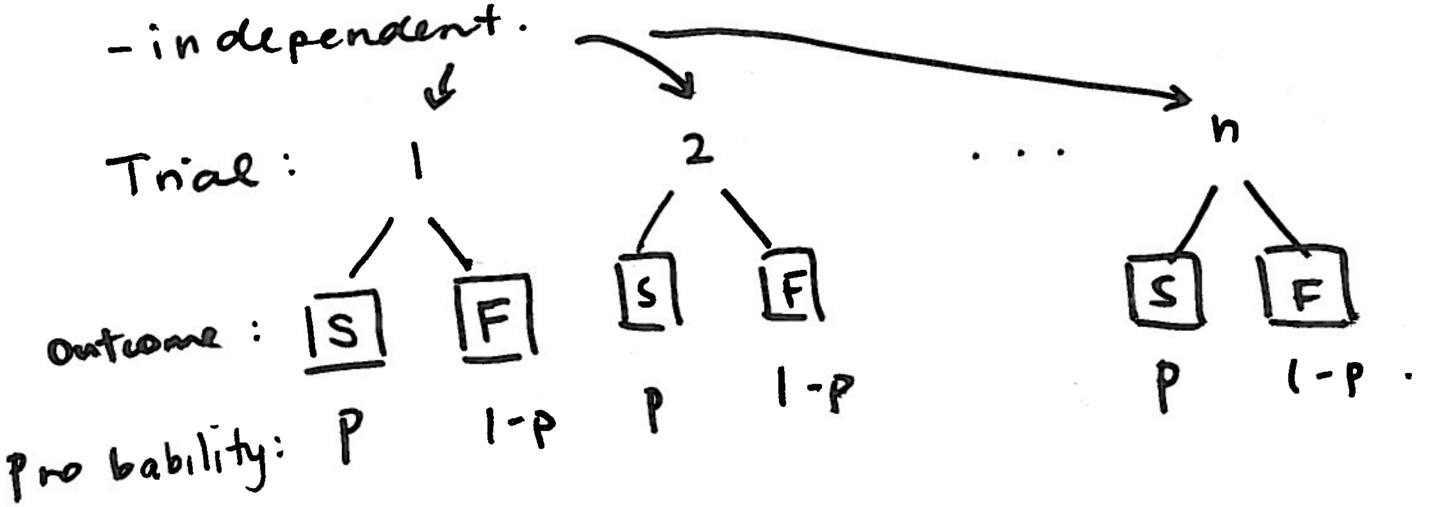
x successes out of n.

where n is the number of trials and p is the probability of success.

Mean and Variance of Binomial Random Variable

$$E(X) = np$$

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


- independent.
Trial: 1 2 . . . n
Outcome: S F S F S F
Probability: p 1-p p 1-p p 1-p.

$X \sim Bin(n, p)$

Oct. 19/16

Last Class:

Bernoulli process: eg. series of coin flips.

- sequence of independent trials
- two outcomes (S/F)
- $P(S) = p$ (constant) on each trial

Bernoulli process

→ (# trials) time until first success? ①

→ For given amount of time (# trials) how many successes have we had? ②

Fix $\underline{\underline{n}}$, how many successes?

① $\underbrace{0 \quad 0 \quad 0 \quad 0 \quad 1}$ ... ← Seq. ind. Bernoulli trials.

X = # trials until 1st success ~ geo(p)

$$P(X=x) = (1-p)^{x-1} p \qquad x = 1, 2, 3 \dots$$

② $\underbrace{0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 0}$ ← ind. Ber. trials.

X = # successes in n trials. ~ Bin(n, p)

(*) $$\boxed{P(X=x) = \binom{n}{x} (1-p)^{n-x} p^{x}} \qquad x = 0, 1, 2, \dots, n.$$

$n = 5$ $\qquad$ X = # sites that have issues in 5 sites.

success = site has an issue

$P(X=2) = ?$

2 Success

3 Failures.

{S and S and F and F and F}

$$P(X=2) \cancel{= p^2} (1-p)^3$$ ↖ in that order.

Motivating Example

In September 2008, Google announced the release of its web browser Chrome. The developers wanted to minimize the probability that their browser would have trouble displaying a website. Before releasing the product, they had to test many sites. The developers sampled websites, recorded whether the browser displayed the website correctly or had an issue.

Suppose Google tests 5 websites. What is the probability exactly 2 of them have problems? Suppose that in this phase of development, 10% of sites exhibited some sort of problem.

There are 10 ways to get two successes in five websites:

- SSFFF FSFSF
- SFSFF FSFFS
- SFFSF FFSSF
- SFFFS FFSFS
- FSSFF FFFSS

$10 \times (0.1)^2 (0.9)^3$

$$P(X=2) = 10 \times p^2(1-p)^3 = 10 \times 0.00729 = 0.0729$$

Each different order in which we can have k successes in n trials is called a **combination**. The total number of ways that can happen is written $\binom{n}{k}$

$5! = 1 \times 2 \times 3 \times 4 \times 5$

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \qquad \text{where } n! = n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$$

Note: $0! = 1$

e.g. $n = 6, k = 2$

$$\binom{6}{2} = \frac{6!}{(6-2)!2!} = \frac{6 \times 5 \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}}{(\cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1})(2 \times 1)} = 15$$

This agrees with our website example, where

$$\binom{5}{2} = \frac{5!}{(5-2)!2!} = \frac{5 \times 4 \times \cancel{3} \times \cancel{2} \times \cancel{1}}{(\cancel{3} \times \cancel{2} \times \cancel{1})(2 \times 1)} = 10$$

0 | 1 | 1 | 1 | 0 | 0

$Y_i \sim Ber(p)$ n

X = # successes in n ind. trials.

$$X = Y_1 + Y_2 + \dots + Y_n$$

binomial bernoulli.

$E(Y_i) = p$

$Var(Y_i) = p(1-p)$

$$E(X) = E(Y_1 + \dots + Y_n) = E(Y_1) + E(Y_2) + \dots + E(Y_n)$$
$$= p + p + \dots + p.$$
$$= np$$

$$Var(X) = Var(Y_1 + Y_2 + \dots + Y_n) = Var(Y_1) + \dots + Var(Y_n)$$
$$= n(1-p)p.$$

$$E(X) = np$$

Interpretation: $E(X)$ long term average.

If perform large # indep. binomial exp. then the avg # of successes from the repeated experiments.

You should be able to recognize a Binomial situation (often on a midterm/exam it will not tell you it is Binomial)

Binomial situation:

1. Fixed number of independent trials (n)
2. Each trial only has 2 possible outcomes
3. The probability of success, p, is the same for each trial

Example 2

A certain city's drainage system is designed for a rainfall intensity that will be exceeded on an average once in 50 years. If we consider the possible outcomes in each year to be flooding or nonflooding then

(a) What is the probability that the city will be flooded only 3 out of 10 years?

(b) What is the probability there will be no floods in 50 years?

design rainfall intensity $\underline{50 \text{ yrs.}}$

avg.

<u>Return period</u> $E(X) = \frac{1}{p}$.

↳ avg length time (in years) for event of a given magnitude is equalled or exceeded.

2) p = probability of flooding in <u>each year</u>.

$p = \frac{1}{50}$

Assume: · flooding / non-flooding in each year.

· Assume years indep.

a) X = # of floods in 10 years. $\sim Bin(n=10, p=0.02)$

$$P(X=3) = \binom{10}{3} 0.02^3 (1-0.02)^{10-3}$$

$$= \frac{10!}{(10-3)!\,3!} 0.02^3\, 0.98^7$$

nr

$$= 0.00083$$

$X = \#$ floods in 50 years $\sim Bin(n=50, p=0.02)$

b) $P(X=0)$

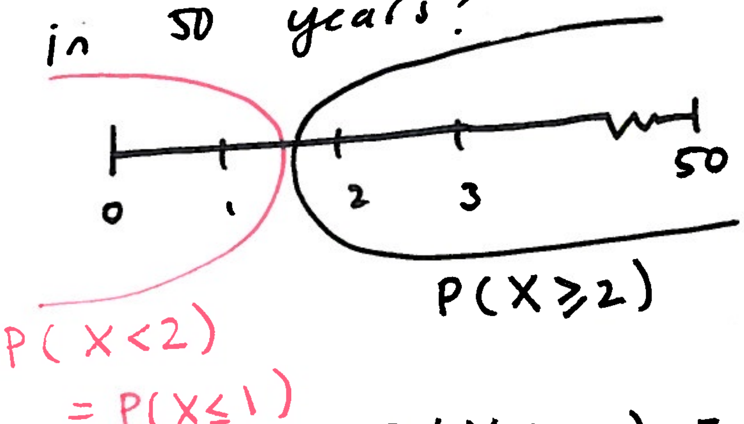
$= \binom{50}{0} 0.02^0\, 0.98^{50}$

$= 0.98^{50} = 0.364$

$X = \#$ of the 1st flood $\sim geo(p = 0.02)$

$$P(X>50) = 1 - P(X \le 50)$$
$$= \cancel{1} - (\cancel{1} - (1-p)^{50})$$
$$= 0.98^{50}$$

c) Probability that there will be 2 or more floods in 50 years?



$$P(X \geq 2) = P(X=2) + P(X=3) + \dots + P(X=50)$$

$$P(X \geq 2) = 1 - P(X \le 1)$$
$$= 1 - [P(X=0) + P(X=1)]$$
$$= 1 - \left[\binom{50}{0} 0.02^0\, 0.98^{50} + \binom{50}{1} 0.02^1\, 0.98^{49}\right]$$

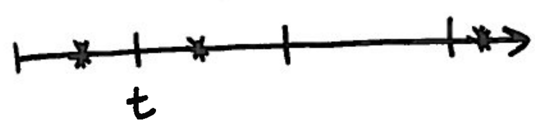
$$= 1 - [\, 0.98^{50} + 50 \times 0.02 \times 0.98^{49} \,]$$

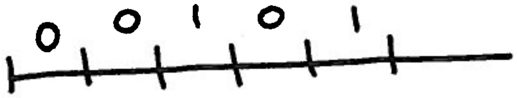
$$= 1 - 0.7356$$

$$= 0.264$$

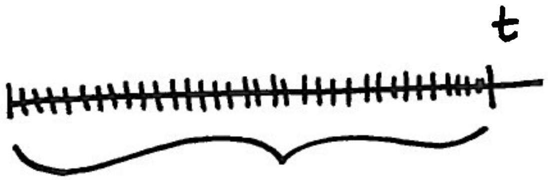
Characteristics of a Poisson Process:

- The number of occurrences of an event in any non-overlapping interval are independent.
- The number of occurrences of the event in an interval is proportional to the size of the interval.
- The probability of an event within a certain interval does not change over different intervals.
- Events cannot occur simultaneously.





bernoulli.



divide t into n intervals

Suppose rate tornadoes is $\underline{\underline{\lambda}}$ times per year.

Period of t years.

λt : tornadoes will occur on avg λt times

p = prob tornado in each interval

$$p = \frac{\lambda t}{n}$$

$$P(x \text{ arrivals}) = \binom{n}{x}\left(\frac{\lambda t}{n}\right)^{x}\left(1 - \frac{\lambda t}{n}\right)^{n-x}$$

$n \to \underline{\underline{\infty}}$

Poisson Process

Suppose we want to model the number of traffic accidents occurring on a highway in a year or the number of customers joining a line in an hour.

The **Poisson Process** is used to model a count of occurrences of events per unit of time/space. (e.g. number of defects on a 1×1m surface). The poisson process gives rise to a discrete random variable called a **Poisson random variable** and a continuous random variable called the **Exponential random variable**.

Rate of the process (denoted by λ) is the average number of occurrences of a certain event per unit time (or per unit of space).

Poisson Random Variable

If the above conditions are satisfied, the random variable $\underline{X}$ = the number of occurrences in a given interval of time/space, has a Poisson distribution.

rate parameter

time interval.

$$X \sim Pois(\lambda t)$$

$$\boxed{P(X = x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!},} \qquad x = 0, 1, 2, 3...$$

where λ is the rate of occurrences of A per unit time and t is the number of unit of time we are looking at.

<u>Mean and Variance of Poisson Random Variable</u>

$$E(X) = \lambda t \qquad Var(X) = \lambda t$$

Example 3

Suppose that the average number of earthquakes with a reading over 8.0 on the Richter scale is 1 per year. What is the probability that there are no earthquakes over 8.0 in the next year? In the next 3 years?

X = # earthquakes in the next year.

$X \sim \text{Pois}(\lambda t = 1 \times 1)$

$$P(X=0) = \frac{e^{-1}(1)^0}{0!} = e^{-1} = 0.368$$

Y = # earthquakes in next 3 years.

$Y \sim \text{Pois}(1 \times 3)$

$$P(Y=0) = \frac{e^{-3}(3)^0}{0!} = 0.0498$$

$\lambda t = 3 \times 2$

$$P(X=5) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} = \frac{e^{-6}\,6^5}{5!}$$

Exponential Random Variable

Recall exponential random variables from Ch. 4. $T =$ the time between consecutive occurrences of the event. T, called the waiting time, is a continuous random variable.

$$\boxed{\begin{array}{ll} T \sim Exp(\lambda) & \\ \text{pdf: } f(t) = \lambda e^{-\lambda t}, & t > 0 \\ \text{cdf: } F(t) = 1 - e^{-\lambda t} & \end{array}}$$

Mean and Variance of Exponential Random Variable

$$E(T) = \frac{1}{\lambda} \qquad Var(T) = \frac{1}{\lambda^2}$$

Important relationship between exponential and Poisson:

Poisson: # occurrences per unit time — discrete.

exp: describe <u>length time</u> between occurrences. — continuous

X = # occurrences ~ Pois

$$P(X=x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$ — # phone calls from 0 to time t.

P(no phone calls in $(0,t]$) $= P(X=0) = \dfrac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$.

Let T = time until 1st occurrence.

$P(T>t)$ = P(wait at least t min before the 1st call)

= P(no calls occur in 1st t min) $= e^{-\lambda t}$

show exp. $F_T(t) = P(T \le t) = 1 - P(T>t) = 1 - e^{-\lambda t}$ // cdf of exponential.

$\dfrac{d F_T(t)}{dt} = \lambda e^{-\lambda t}$

Example 4

Continuing on from example 4 above: Suppose that the average number of earthquakes with a reading over 8.0 on the Richter scale is 1 per year. What is the probability that you wait less than 2 years until the next occurrence of an earthquake over 8.0?

$T \sim \exp(\lambda = 1)$ T = time until next earthquake

1

$P(T < 2) = F_T(2) = 1 - e^{-1 \cdot 2} = 0.865$

match.

$X = \#$ earthquakes in 2 years. $\sim \text{Pois}(1 \cdot 2)$

P(at least 1 earthquake in next 2 years)

2

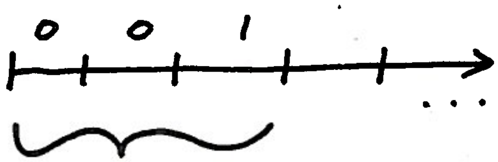
$= P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0)$

$= 1 - \dfrac{e^{-2} 2^0}{0!} = 0.865$

Last Class:

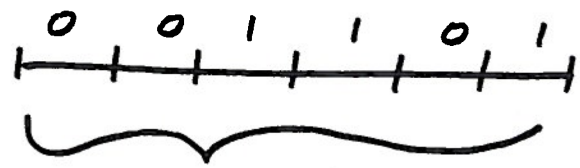
← discrete →

Geometric r.v.



X = # trials until 1st success. ~geo(p)

Binomial r.v



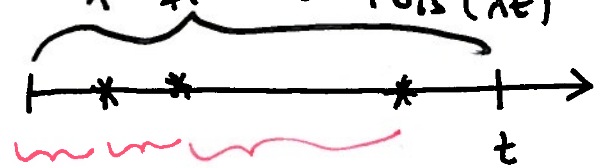
X = # successes in n trials ~ Bin(n, p)

Exponential r.v. (cont.)

X = time until 1st success

Poisson r.v. (discrete)

X = # ~ Pois(λt)



T ~ exp(λ)

6.12 course text

a) X = # of killer whales arriving in next hour.

$X \sim Pois(\lambda t = 4)$ $E(X) = \lambda t = 4$ $Var(X) = \lambda t = 4$

b) T = waiting time (hours) between arrivals $\sim exp(\lambda)$ $= 4$

① $P(T > \frac{1}{2}) = 1 - P(T \leq \frac{1}{2})$

$= 1 - F_T(\frac{1}{2})$

$= 1 - (1 - e^{-4(\frac{1}{2})}) = 0.135 \leftarrow$

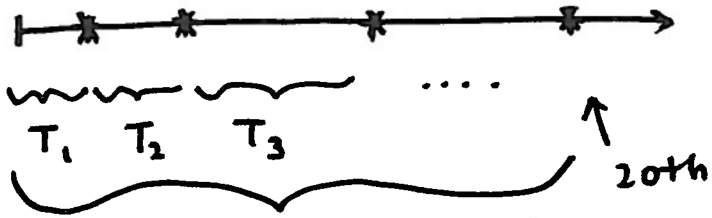
Recall: $F(t) = 1 - e^{-\lambda t}$

② X = # killer whales arriving in next $\frac{1}{2}$ hour

$\sim Pois(\lambda t = 4 \cdot \frac{1}{2} = 2)$

$P(X=0) = \frac{e^{-2}(2)^0}{0!} = 0.135 \leftarrow$

c)



T_i = wait time between $(i-1)$th and ith whale $\sim \exp(\lambda = 4)$

W = time (hour) until 20th.

$E(W) = ?$ $\quad Var(W) = ?$

$$W = \sum_{i=1}^{20} T_i$$

$$E(W) = E(T_1 + \dots + T_{20}) = E(T_1) + \dots + E(T_{20}) = 20 \times \underline{\underline{E(T)}}$$

$$= 20 \times \frac{1}{4} = 5$$

$$Var(W) = Var(T_1 + \dots + T_{20}) = Var(T_1) + \dots + Var(T_{20})$$

T_i indep.

$$= 20\, Var(T) = 20 \times \frac{1}{4^2} = 1.25$$

$$P(X=x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

$P(X=x)$ $\quad x = 0, 1, 2, \dots, n.$

$\lambda t = \underline{\underline{np}}.$ $\quad (0, \infty)$

* large n, small p.

$$\underline{\underline{P(X=x)}} = \frac{e^{-np}(np)^x}{x!}$$

Poisson Approximation to the Binomial

If $X \sim Bin(n, p)$ with large n $(n \geq 20)$ and small p $(np < 5)$, then we can use a Poisson random variable with $\lambda t = np$ to approximate Binomial.

Example 5

If 1% of the output from a machine is defective, then what is the probability that 4 or more are defective in a random sample of 200?

X = # defective items in 200 $\sim Bin(n=200, p=0.01)$

$P(X \geq 4) = 1 - P(X < 4) = 1 - [P(X=0) + P(X=1) + \ldots + P(X=3)]$

$= 1 - \left[\binom{200}{0} 0.01^{0}\, 0.99^{200} + \ldots + \binom{200}{3} 0.01^{3}\, 0.99^{197}\right]$

$= 1 - 0.8580 = 0.142.$

Alternate:

n large: $n \geq 20$ ✓

$np < 5$: $n \times p = 200 \times 0.01 = 2 < 5$ ✓

We can approximate using Poisson.

$$X \sim Pois(\lambda t = np = 2)$$

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

$$= 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3)]$$

$$= 1 - \left[\frac{e^{-2} 2^0}{0!} + \frac{e^{-2} 2^1}{1!} + \frac{e^{-2} 2^2}{2!} + \frac{e^{-2} 2^3}{3!} \right]$$

$$= 0.143$$