



Geometric and Poisson Distributions

Statistics I · Lecture 10

Today

- Geometric distribution and its lack of memory
- Poisson distribution, and when a Binomial becomes a Poisson
- Which model is this?
- Practice, from  to 

Geometric Distribution

Geometric Distribution (Pascal's)

Consider a succession of Bernoulli trials, and let X the r.v. how many trials you need until observing the first success. This is a r.v. because there might be situations when you get success at your first trial, or second, etc. You cannot anticipate this.

$X \sim Geo(p)$ if its probability distribution function $P(X = x)$ is given by:

$$f_X(x) = p(1 - p)^{x-1}, \quad x \in \{1, 2, \dots\}, \quad p \in [0, 1]$$

Geometric Distribution (Pascal's)

The first two moments of the distribution are found as:

- $E[X] = \mu_X = \frac{1}{p}$
- $V[X] = \frac{1-p}{p^2}$
- $F_X(x) = \begin{cases} 0 & x < 1 \\ 1 - (1 - p)^k & k \leq x < k + 1, k \in \mathbb{N} \end{cases}$

Geometric Distribution (Pascal's)

There is a special feature of this distribution, it is said it “*lacks memory*” in the following sense. Let $s > t > 0$

$$\begin{aligned}
 P(X > s | X > t) &= \frac{P(X > s \wedge X > t)}{P(X > t)} = \frac{P(X > s)}{P(X > t)} \\
 &= \frac{1 - F(s)}{1 - F(t)} = \frac{(1 - p)^s}{(1 - p)^t} = (1 - p)^{s-t} \\
 P(X > s - t) &= 1 - P(X \leq s - t) = 1 - F(s - t) \\
 &= (1 - p)^{s-t}
 \end{aligned}$$

Geometric and Binomial

Distribution	r.v. X	Parameter
Binomial	# Success	# Trials
Geometric	# Trials	1 st Success

Example

A company specialized in sport footwear, imports a share of the material that is packed in boxes of 1000 units. To avoid counterfeiting, in each box 5 products are randomly selected, and the whole box is returned if one of these 5 products raises suspicions.

1. What is the probability that a box with 10 suspicious units is returned?

[Sol](#)

2. Check if this probability changes too much if you disregard that you are not replacing the items. [Sol](#)

3. What is the probability that we need to inspect 15 units, independently between them, until finding a suspicious item? [Sol](#)

4. How many units you inspect until finding the first suspicious unit? [Sol](#)

? Geometric · Question 1

$X \sim \text{Geo}(p)$ counts the number of trials up to and including the first success. Then $E[X]$ equals:

A. p

B. $\frac{1-p}{p^2}$

C. $\frac{1}{p}$

D. np

☒ C. If a success happens one time in five, you wait five trials on average. Note $\frac{1-p}{p^2}$ is the **variance**.

? Geometric · Question 2

The Geometric distribution is memoryless, which means that for $s > t$, $P(X > s | X > t)$ equals:

- A. $P(X > s)$
- B. $P(X > s + t)$
- C. $P(X > s - t)$
- D. 0

✓ **C.** We showed $P(X > s | X > t) = (1 - p)^{s-t}$, which is exactly $P(X > s - t)$. The failures already accumulated are forgotten.



Geometric · Question 3

Each call you make is answered with probability $p = 0.2$, independently. Let X be the number of calls until the first one is answered.

Compute $P(X = 3)$, $P(X > 4)$ and $E[X]$.



Geometric · Solution

$$P(X = 3) = (1 - p)^2 p = 0.8^2 \times 0.2 = 0.128$$

$$P(X > 4) = (1 - p)^4 = 0.8^4 = 0.4096$$

$$E[X] = \frac{1}{0.2} = 5 \text{ calls}$$

Poisson Distribution

Poisson Distribution

This distribution is associated with a process of counting, a Poisson process.

Examples:

- Count the number of patients arriving every day to the E.R. in a hospital.
- Count how many calls does a call-center receive in an hour.
- Count how many typos does a book have.

You can count something in a time frame, or a particular region.

Poisson process

The Poisson process has the following features:

1. It must be homogeneous in time/space. It only depends on the length that you consider to measure, not where, not when.
2. Events happening in disjoint regions/time slots must be independent.
3. The probability of having an event in the exact same space or at the exact same time is negligible (no simultaneous events).

Poisson Distribution

The r.v. X follows a Poisson distribution if $X \sim Poi(\lambda)$ if its probability distribution function $P(X = x)$ is given by:

$$f_X(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & x = 0, 1, \dots \\ 0 & otherwise \end{cases}$$

with $\lambda > 0$, which represents the average number of events in a given time slot or region.

Poisson Distribution

The first two moments of the distribution are found as:

- $E[X] = \lambda$
- $V[X] = \lambda$

Additivity Theorem of the Poisson Distribution

Let k r.v.s X_i with $i = 1, 2, \dots, k$, independent, where $X_i \sim Poi(\lambda_i)$. Let

$$S_k = X_1 + \dots + X_k = \sum_{i=1}^k X_i$$

Then $S_k \sim Poi \left(\sum_{i=1}^k \lambda_i \right)$

Binomial $\rightarrow$ Poisson

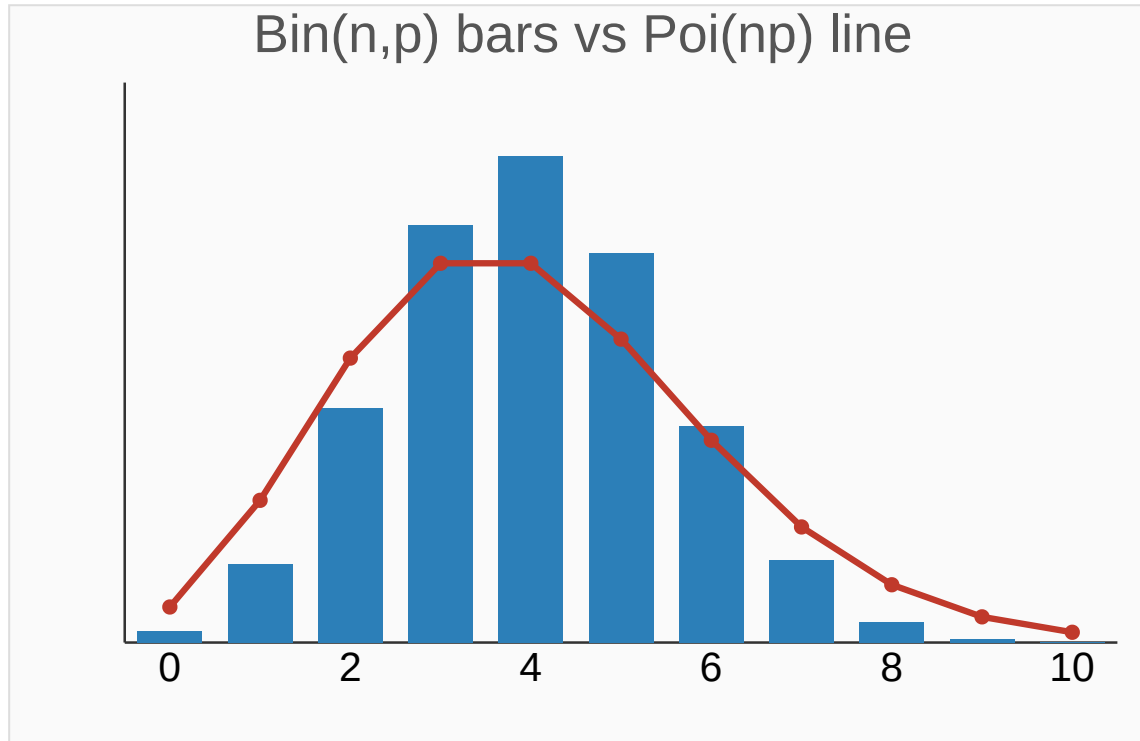
Let $X \sim \text{Bin}(n, p)$. If $n \gg 1 \gg p$, i.e. if this is a very rare event (very low success probability) in a very large sample, you can approximate this with a Poisson distribution, where $\lambda = np$.

$$X \sim \text{Bin}(x, n, p) \Leftrightarrow X \approx \text{Poi}(x, \lambda = np)$$

As a rule of thumb, do not approximate if $p \in [0.1, 0.9]$ or if $n \leq 20$.

When is the approximation safe?

Bars are the exact $\text{Bin}(n, p)$. The red outline is $\text{Poi}(\lambda = np)$. Slide n up and p down keeping λ near 4, and watch them merge.



$n = 10$



$p = 0.40$



$\lambda = np = 4.00$

biggest gap = 0.0555

Rule of thumb: do NOT approximate

needs $p < 0.1$ or $p > 0.9$, and $n > 20$

What the chart showed

At $n = 10$, $p = 0.4$ the red line misses the bars badly. Drag p down to 0.02 and n up to 200: same $\lambda = 4$, and the two become almost indistinguishable.

The approximation is not about λ being right, it is about **the event being rare in a large sample**. Keeping λ fixed while raising n and lowering p is exactly the limit the theorem describes.

Watch the “biggest gap” readout against the rule of thumb: it is what the $p \notin [0.1, 0.9]$ and $n > 20$ conditions are protecting you from.

Example

Find the value of the following probabilities, using the Poisson [table](#)

1. $P(X \leq 5)$ if $X \sim Poi(10)$
2. $P(4 \leq X \leq 8)$ if $X \sim Poi(5)$ [Sol](#)
3. $P(X = 7)$ if $X \sim Poi(10)$ [Sol](#)

►  Answer

Example

The number of patients arriving daily to the ICU in a hospital, follows a **Poisson** process with mean 4. The ICU has capacity of 6 patients, the others, are derived to the nearest hospital. Assess the validity of the following sentences:

1. The probability that, on a given day, there is no need to transfer any patient is 0.8893. **Sol**
2. The most likely number of patients arriving daily to the ICU is 6. **Sol**
3. The probability that, on a given day, arrive 5 patients, given that in the previous day only 2 patients arrived, is 0.1563. **Sol**
4. The probability that, in 5 days, at least 15 patients arrive to the ICU is 0.8435. **Sol**
5. To ensure that approx. 97% of the time (days) there are no transfers, it is necessary to increase the capacity in 4 more beds. **Sol**

? Poisson · Question 1

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

A. $E[X] = \lambda, V[X] = \lambda^2$

B. $E[X] = \frac{1}{\lambda}$

C. $E[X] = \lambda, V[X] = 1$

D. $E[X] = V[X] = \lambda$

 **D.** The Poisson is the one distribution in this chapter whose mean and variance coincide.

? Poisson · Question 2

The Binomial is well approximated by a Poisson when:

- A. n is large, p is small, and np stays moderate
 - B. n is small and p large
 - C. $n = p$
 - D. always
-  **A.** Many trials, each very unlikely, with a stable expected count $\lambda = np$: rare events over a fixed interval.



Poisson · Question 3

A call centre receives on average 3 calls per minute, following a Poisson process.

Compute $P(X = 0)$, $P(X \leq 2)$ and $P(X > 2)$ for a given minute.



Poisson · Solution

With $\lambda = 3$ and $f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$:

$$P(X = 0) = e^{-3} \approx 0.0498 \cdot P(X = 1) = 3e^{-3} \approx 0.1494 \cdot P(X = 2) = \frac{9}{2}e^{-3} \approx 0.2240$$

$$P(X \leq 2) \approx 0.4232, \text{ so } P(X > 2) = 1 - P(X \leq 2) \approx 0.5768$$

Practice

Practice

The exercises go from easy to hard:

-  warm-up: one formula, one step
-  standard: two or three steps
-  exam level: this is what you will find in the midterm
-  stretch: you need to combine several things

In the midterm, nobody will tell you which distribution to use. So we start with that.

Practice · Which model is this?

Name the distribution of X , and its parameters:

1. You pick 20 items with replacement, each is defective with probability 0.03. X is the number of defective items.
2. You draw 5 cards  from a deck of 52. X is the number of aces.
3. You receive on average 12 emails  per hour. X is the number of emails in the next hour.
4. You call clients  until one accepts the offer, each accepts with probability 0.15. X is the number of calls.

 1. $\text{Bin}(20, 0.03)$ 2. $\text{Hypergeometric}(52, 4, 5)$ 3. $\text{Poi}(12)$ 4. $\text{Geo}(0.15)$

Practice · Which model is this?

Name the distribution of X , and its parameters:

5. A class has 30 students, 12 of them Erasmus. You pick 4 for a group work. X is the number of Erasmus students in the group.
6. A report has 10 pages, with on average 0.5 typos per page. X is the number of typos in the whole report.
7. A bank processes 500 transactions, each is fraudulent with probability 0.002. X is the number of frauds.

 5. *Hypergeometric*(30, 12, 4) 6. *Poi*(5), by additivity 7. *Bin*(500, 0.002), which is approximately *Poi*(1)

Practice · Exercise 1

Let $X \sim Poi(2.5)$. Using the [table](#), find:

(a) $P(X \leq 1)$

(b) $P(X = 4)$

(c) $P(X > 5)$

(a) $F(1) = 0.2873$

(b) $P(X = 4) = F(4) - F(3) = 0.8912 - 0.7576 = 0.1336$

(c) $P(X > 5) = 1 - F(5) = 1 - 0.9580 = 0.0420$

Practice · Exercise 2

A salesperson  closes a sale in each call with probability 0.2, independently between calls. Let X be the number of calls until the first sale.

- (a)** Find the probability that the first sale happens in the 3rd call.
- (b)** Find the probability that more than 5 calls are needed. How many calls are expected?
- (c)** The first 4 calls failed. What is the probability that more than 3 additional calls are needed?



Practice · Exercise 2 · Solution

$$X \sim \text{Geo}(0.2)$$

(a) $P(X = 3) = 0.2 \times 0.8^2 = 0.128$

(b) $P(X > 5) = 0.8^5 \approx 0.3277$, and $E[X] = \frac{1}{0.2} = 5$ calls.

(c) By the lack of memory,

$P(X > 7 | X > 4) = P(X > 3) = 0.8^3 = 0.512$. The 4 failed calls do not change anything, the phone does not remember.

Practice · Exercise 3

A website  has, on average, 2 errors per hour, following a Poisson process.

- (a) Find the probability of no errors in the next 30 minutes.
- (b) Find the probability of at least 3 errors in the next 2 hours.

The parameter follows the time window: in 30 minutes $X \sim Poi(1)$, and in 2 hours $Y \sim Poi(4)$.

(a) $P(X = 0) = e^{-1} \approx 0.3679$

(b) $P(Y \geq 3) = 1 - F(2) = 1 - 0.2381 = 0.7619$

Practice · Exercise 4

The probability that a subject is allergic to the active ingredient of drug M  is 0.05. In a laboratory 80 subjects, randomly chosen, are tested.

Find the probability that exactly 2 subjects are allergic.

$X \sim \text{Bin}(80, 0.05)$, and $n = 80$ is not in the table.

n is large and p small, so $X \approx \text{Poi}(80 \times 0.05) = \text{Poi}(4)$:

$$P(X = 2) \approx F(2) - F(1) = 0.2381 - 0.0916 = 0.1465$$

The exact Binomial value is 0.1446.

Practice · Exercise 5

In Branch 17 of the Bank of Coins , the clients asking about financial services follow a Poisson process. The probability that no client shows up in one hour is 0.3679. The branch opens from 8:30 until 14:30, and 25% of the clients are female.

- (a)** How many clients are expected between 12:00 and 13:00?
- (b)** Find the probability that at most 6 clients show up in a whole day.
- (c)** Out of 10 clients, find the probability that 6 or fewer are female.
- (d)** Find the probability that at least 4 clients are needed to see the first female client.



Practice · Exercise 5 · Solution

- (a) $P(X = 0) = e^{-\lambda} = 0.3679 \Rightarrow \lambda = 1$ client per hour, so we expect 1 client.
- (b) The day has 6 hours, so $Y \sim Poi(6)$ and $P(Y \leq 6) = 0.6063$
- (c) $W \sim Bin(10, 0.25)$ and $P(W \leq 6) = 0.9965$
- (d) $G \sim Geo(0.25)$. At least 4 means the first 3 were all male: $P(G \geq 4) = P(G > 3) = 0.75^3 \approx 0.4219$

● Practice · Exercise 6

Let X and Y be independent, with the same expected value, 3:

$$X \sim \text{Bin}(10, 0.3) \quad \text{and} \quad Y \sim \text{Poi}(\lambda)$$

(a) Find $E[X - 2Y]$.

(b) Find $P(X + Y = 5 | X < 2)$.

(a) $E[Y] = \lambda = 3$, so $E[X - 2Y] = 3 - 2 \times 3 = -3$

(b) If $X < 2$, then either $X = 0$ and $Y = 5$, or $X = 1$ and $Y = 4$:

$$\begin{aligned} P(X + Y = 5 | X < 2) &= \frac{P(X = 0)P(Y = 5) + P(X = 1)P(Y = 4)}{P(X \leq 1)} \\ &= \frac{0.0282 \times 0.1008 + 0.1211 \times 0.1680}{0.1493} \approx 0.1553 \end{aligned}$$

Practice · Exercise 7

True or false?

(a) If $X \sim \text{Poi}(2)$ and $Y \sim \text{Poi}(3)$ are independent, then $X + Y \sim \text{Poi}(5)$.

(b) If $X \sim \text{Poi}(2)$, then $2X \sim \text{Poi}(4)$.

(a)  True, this is the additivity theorem.

(b)  False. $2X$ only takes even values, $0, 2, 4, \dots$, and a Poisson can take any value in $\mathbb{N}_0$. It has the right mean, $E[2X] = 4$, but $V[2X] = 4V[X] = 8 \neq 4$.



Homework

Problem set **3.1**, Questions 6 to 11.

Bibliography

- Murteira, B.; Silva Ribeiro, C.; Andrade e Silva, J. & Pimenta, C., Introdução à Estatística, Escolar Editora, McGraw-Hill, 2010
- Paulino, C. D. & Branco, J. A. (2005). Exercícios de Probabilidade e Estatística. Escolar Editora
- Pimenta, F., Andrade e Silva, J.; Silva Ribeiro, C. & Murteira, B., Introdução à Estatística, 3ª Edição, Escolar Editora, 2015

Appendix

Binomial Table

Distribution Function

n = 1

x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
0	0.9500	0.9000	0.8500	0.8000	0.7500	0.7000	0.6500	0.6000	0.5500	0.5000
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n = 2

x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

n = 3

x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
---	------	------	------	------	------	------	------	------	------	------

Hypergeometric Example

Back to the exercise

First, note that $X \sim \text{Hypergeometric}(1000, 10, n)$

$$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) = 1 - P(X \leq 0) \\ &= 1 - P(X = 0) = 1 - \frac{\binom{10}{0} \binom{990}{5}}{\binom{1000}{5}} \\ &= 1 - 0.9509 \approx 0.0491 \end{aligned}$$

Hypergeometric Example

Back to the exercise

With reposition we could approximate with the Binomial (note $n/N = 0.005 < 0.1$), $X \approx \text{Bin}(n = 5, M/N = 0.01)$

$$P(X \geq 1) = 1 - P(X = 0) \approx 1 - 0.95099 \approx 0.04901$$

Both values are very close.

Hypergeometric Example

Back to the exercise

$$Y \sim \text{Geo}(p = 0.01),$$

$$\begin{aligned} P(Y = 15) &= p(1 - p)^{y-1} = 0.01(1 - 0.01)^{14} \\ &= 0.01 \times 0.8687458 \approx 0.0087 \end{aligned}$$

Hypergeometric Example

Back to the exercise

$Y \sim \text{Geo}(p = 0.01)$, then

$$E[Y] = \frac{1}{p} = \frac{1}{0.01} = 100$$

You need to inspect 100 units on average.

Poisson Table

[Back to exercise](#)

Distribution Function

$x \setminus \lambda$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725	0.7358
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371	0.9197
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865	0.9810
4	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977	0.9963
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9994
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$x \setminus \lambda$	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0

Poisson Example

[Back to the exercise](#)

If X is the r.v. of the number of patients arriving every day at the ICU, $X \sim \text{Poi}(4)$. We are interested in $P(X \leq 6)$, looking at the table we get $F(6) = 0.8893$.  True.

Poisson Example

Back to the exercise

Note that from the table we can see that

$P(X = 3) = P(X = 4) = 0.1954$, and also $\lambda = 4$. Given that $\lambda = E[X]$, the most likely value must be λ or very close. In sum, the sentence is  false.

Poisson Example

Back to the exercise

We need to remember two important facts about the Poisson distribution. In two disjoint periods, the random variables are independent. So the number of patients one day, and the number of patients in the next days are independent. On the other hand, remember that the distribution must be the same for the same time-window (they are independently but identically distributed):

$$P(X_{day2} = 5 | X_{day1} = 2) = P(X_{day2} = 5) = 0.1563$$

The sentence is  true.

Poisson Example

Back to the exercise

Let $Y \sim Poi(20)$ (because of the additivity theorem).

$$\begin{aligned} P(Y \geq 15) &= 1 - P(Y < 15) \\ &= 1 - P(Y \leq 14) = 1 - F(14) \\ &= 1 - 0.1049 = 0.8951 \end{aligned}$$

The sentence is **✗** false.

Poisson Example

Back to the exercise

In this exercise, we need to find, given that $\lambda = 4$, the x that makes $F(x)$ greater or equal than 0.97. By inspection in the table, we see that it is $x = 8$, but we have know 6 beds, and therefore we would need 2 more beds to satisfy our requirement.

The sentence is  false.