Useful Distributions (pt 2)

Negative binomial distribution

Practice

A college student working at a psychology lab is asked to recruit 10 couples to participate in a study.

She decides to stand outside the student center and ask every person leaving the building whether they are in a relationship and, if so, whether they would like to participate in the study with their significant other.

Suppose the probability of finding such a person is 10%. What is the probability that she will need to ask 30 people before she hits her goal?

Practice

A college student working at a psychology lab is asked to recruit 10 couples to participate in a study.

She decides to stand outside the student center and ask every person leaving the building whether they are in a relationship and, if so, whether they would like to participate in the study with their significant other.

Suppose the probability of finding such a person is 10%. What is the probability that she will need to ask 30 people before she hits her goal?

Negative binomial distribution

- The negative binomial distribution describes the probability of observing the k-th success on the nth trial.
- The following four conditions are useful for identifying a negative binomial case:
  1. The trials are independent.
  2. Each trial outcome can be classified as a success or failure.
  3. The probability of success (p) is the same for each trial.
  4. The last trial must be a success.

  Note that the first three conditions are common to the binomial distribution.

Negative binomial distribution

$$P(k^{th} \text{ success on the } n^{th} \text{ trial}) = \binom{n-1}{k-1} p^k (1-p)^{n-k},$$

where p is the probability that an individual trial is a success. All trials are assumed to be independent.

Negative binomial distribution

	X is counting...	Probability mass function	Formula	Alternate formula (using equivalent binomial)	Alternate formula (simplified using: $n = k + r$)	Support
1	k failures, given r successes	$f(k; r, p) \equiv \Pr(X = k) =$	$\binom{k+r-1}{k} p^r (1-p)^k$ [7][5][8]	$\binom{k+r-1}{r-1} p^r (1-p)^k$ [2] [9][10][11]	$\binom{n-1}{k} p^r (1-p)^k$	for $k = 0, 1, 2, \ldots$
2	n trials, given r successes	$f(n; r, p) \equiv \Pr(X = n) =$	$\binom{n-1}{r-1} p^r (1-p)^{n-r}$ [5][11][12][13][14]	$\binom{n-1}{n-r} p^r (1-p)^{n-r}$		for $n = r, r+1, r+2, \ldots$
3	n trials, given r failures	$f(n; r, p) \equiv \Pr(X = n) =$	$\binom{n-1}{r-1} p^{n-r} (1-p)^r$	$\binom{n-1}{n-r} p^{n-r} (1-p)^r$	$\binom{n-1}{k} p^k (1-p)^r$	
4	r successes, given n trials	$f(r; n, p) \equiv \Pr(X = r) =$	This is the binomial distribution: $\binom{n}{r} p^r (1-p)^{n-r}$			for $r = 0, 1, 2, \ldots, n$

Negative binomial distribution

An RV X has the Negative Binomial distribution with parameters r and p if it has the pf

$$f(x \mid p, r) = \begin{cases} \binom{r+x-1}{x} p^r (1-p)^x & x = 0,1,2,\dots, \\ 0 & \text{otherwise.} \end{cases}$$

where $0 < p < 1$ and r is a positive integer.

Say we have an infinite sequence of Bernoulli trials with parameter p, and $X =$ number of "failures" before the r th "success". Then $X \sim$ Neg Binomial (r, p).

- Parameter space: r positive integer, $p \in (0,1)$.

- $E(X) = \frac{r(1-p)}{p} \quad \text{Var}(X) = \frac{r(1-p)}{p^2}$

Properties

Theorem (Sum of Geometric is Negative Binomial)

If $X_1, \ldots, X_r$ are i.i.d. and each $X_i \sim$ Geometric(p) then $X = X_1 + \cdots + X_r \sim$ NegBinomial (r, p).

Practice

A college student working at a psychology lab is asked to recruit 10 couples to participate in a study.

She decides to stand outside the student center and ask every person leaving the building whether they are in a relationship and, if so, whether they would like to participate in the study with their significant other.

Suppose the probability of finding such a person is 10%. What is the probability that she will need to ask 30 people before she hits her goal?

$P(10^{th}$ success on the 30^{th} trial)

Practice

A college student working at a psychology lab is asked to recruit 10 couples to participate in a study.

She decides to stand outside the student center and ask every person leaving the building whether they are in a relationship and, if so, whether they would like to participate in the study with their significant other.

Suppose the probability of finding such a person is 10%. What is the probability that she will need to ask 30 people before she hits her goal?

$$P(10^{th} \text{ success on the } 30^{th} \text{ trial}) = \binom{29}{9} \times 0.10^{10} \times 0.90^{20}$$

Practice

A college student working at a psychology lab is asked to recruit 10 couples to participate in a study.

She decides to stand outside the student center and ask every person leaving the building whether they are in a relationship and, if so, whether they would like to participate in the study with their significant other.

Suppose the probability of finding such a person is 10%. What is the probability that she will need to ask 30 people before she hits her goal?

$$\begin{aligned} P(10^{th} \text{ success on the } 30^{th} \text{ trial}) &= \binom{29}{9} \times 0.10^{10} \times 0.90^{20} \\ &= 10,015,005 \times 0.10^{10} \times 0.90^{20} \end{aligned}$$

Practice

A college student working at a psychology lab is asked to recruit 10 couples to participate in a study.

She decides to stand outside the student center and ask every person leaving the building whether they are in a relationship and, if so, whether they would like to participate in the study with their significant other.

Suppose the probability of finding such a person is 10%. What is the probability that she will need to ask 30 people before she hits her goal?

$$
\begin{aligned}
P(10^{th} \text{ success on the } 30^{th} \text{ trial}) &= \binom{29}{9} \times 0.10^{10} \times 0.90^{20} \\
&= 10,015,005 \times 0.10^{10} \times 0.90^{20} \\
&= 0.00012
\end{aligned}
$$

Binomial vs negative binomial

How is the negative binomial distribution different from the binomial distribution?

Binomial vs negative binomial

How is the negative binomial distribution different from the binomial distribution?

- In the binomial case, we typically have a fixed number of trials and instead consider the number of successes.

- In the negative binomial case, we examine how many trials it takes to observe a fixed number of successes and require that the last observation be a success.

Practice

Which of the following describes a case where we would use the negative binomial distribution to calculate the desired probability?

(a) Probability that a 5 year old boy is taller than 42 inches.

(b) Probability that 3 out of 10 softball throws are successful.

(c) Probability of being dealt a straight flush hand in poker.

(d) Probability of missing 8 shots before the first hit.

(e) Probability of hitting the ball for the 3rd time on the 8th try.

Practice

Which of the following describes a case where we would use the negative binomial distribution to calculate the desired probability?

(a) Probability that a 5 year old boy is taller than 42 inches.

(b) Probability that 3 out of 10 softball throws are successful.

(c) Probability of being dealt a straight flush hand in poker.

(d) Probability of missing 8 shots before the first hit.

(e) Probability of hitting the ball for the 3rd time on the 8th try.

Poisson Distribution

Poisson Distribution

Wight et al (2004) looked at the variation in cadaveric heart beating organ donor rates in the UK.

There were 1330 organ donors, aged 15-69, across the UK for the two years 1999 and 2000 combined.

X: Number of donors per day

P(X)?

The mean number of organ donors per day over the two-year period is calculated as:

$\lambda = \frac{1330}{365+365} = \frac{1330}{730} = 1.82$ organ donations per day

82 million person years, each person has a very small probability of becoming an organ donor.

Poisson distribution

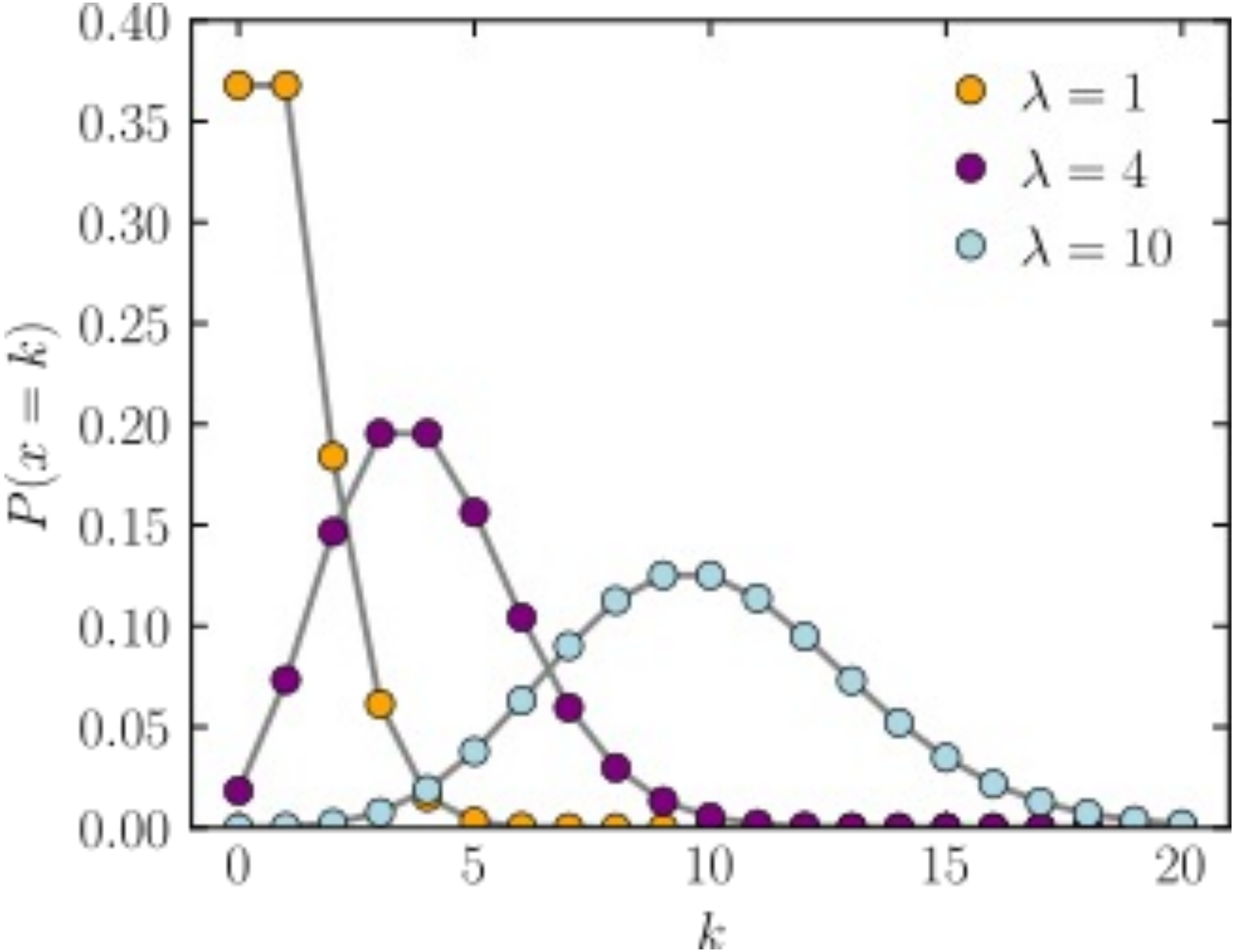
- The Poisson distribution is often useful for estimating the number of rare events in a large population over a short unit of time for a fixed population if the individuals within the population are independent.

- The **rate** for a Poisson distribution is the average number of occurrences in a mostly-fixed population per unit of time, and is typically denoted by λ.

- Using the rate, we can describe the probability of observing exactly *k* events in a single unit of time.

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

Poisson distribution



Poisson or Binomial

Let $X \sim \text{Binomial}\left(n, p = \frac{\lambda}{n}\right)$, where $\lambda > 0$ is fixed. Then for any $k \in \{0,1,2,\dots\}$, we have

$$\lim_{n\to\infty} P_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Proof
We have

$$\lim_{n\to\infty} P_X(k) = \lim_{n\to\infty} \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \lambda^k \lim_{n\to\infty} \frac{n!}{k!\,(n-k)!}\left(\frac{1}{n^k}\right)\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \cdot \lim_{n\to\infty} \left(\left[\frac{n(n-1)(n-2)\dots(n-k+1)}{n^k}\right]\left[\left(1-\frac{\lambda}{n}\right)^n\right]\left[\left(1-\frac{\lambda}{n}\right)^{-k}\right]\right).$$

Poisson or Binomial

Let $X \sim \text{Binomial}\left(n, p = \frac{\lambda}{n}\right)$, where $\lambda > 0$ is fixed. Then for any $k \in \{0,1,2,\dots\}$, we have

$$\lim_{n\to\infty} P_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Proof
We have

$$\lim_{n\to\infty} P_X(k) = \lim_{n\to\infty} \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \lambda^k \lim_{n\to\infty} \frac{n!}{k!\,(n-k)!}\left(\frac{1}{n^k}\right)\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!}\cdot \lim_{n\to\infty}\left(\left[\frac{n(n-1)(n-2)\dots(n-k+1)}{n^k}\right]\left[\left(1-\frac{\lambda}{n}\right)^n\right]\left[\left(1-\frac{\lambda}{n}\right)^{-k}\right]\right).$$

1 $\qquad e^{-\lambda} \qquad$ 1

Example #1

Suppose that in a rural region of a developing country electricity power failures occur following a Poisson distribution with an average of 2 failures every week. Calculate the probability that in a given week the electricity fails only once.

Example #1

Suppose that in a rural region of a developing country electricity power failures occur following a Poisson distribution with an average of 2 failures every week. Calculate the probability that in a given week the electricity fails only once.

Given $\lambda = 2$.

Example #1

Suppose that in a rural region of a developing country electricity power failures occur following a Poisson distribution with an average of 2 failures every week. Calculate the probability that in a given week the electricity fails only once.

Given λ = 2.

$$P(\text{only 1 failure in a week}) = \frac{2^1 \times e^{-2}}{1!}$$

Example #1

Suppose that in a rural region of a developing country electricity power failures occur following a Poisson distribution with an average of 2 failures every week. Calculate the probability that in a given week the electricity fails only once.

Given λ = 2.

$$P(\text{only 1 failure in a week}) = \frac{2^1 \times e^{-2}}{1!}$$

$$= \frac{2 \times e^{-2}}{1}$$

Example #1

Suppose that in a rural region of a developing country electricity power failures occur following a Poisson distribution with an average of 2 failures every week. Calculate the probability that in a given week the electricity fails only once.

Given $\lambda = 2$.

$$
\begin{aligned}
P(\text{only 1 failure in a week}) &= \frac{2^1 \times e^{-2}}{1!} \\
&= \frac{2 \times e^{-2}}{1} \\
&= 0.27
\end{aligned}
$$

Example #2

Suppose that in a rural region, electricity power failures occur following a Poisson distribution with an average of 2 failures every week. Calculate the probability that on a given day the electricity fails three times.

Example #2

Suppose that in a rural region, electricity power failures occur following a Poisson distribution with an average of 2 failures every week. Calculate the probability that on a given day the electricity fails three times.

We are given the weekly failure rate, but to answer this question we need to first calculate the average rate of failure on a given day: λ_{day} = 2/7 = 0.2857. Note that we are assuming that the probability 7 of power failure is the same on any day of the week, i.e. we assume independence.

Example #2

Suppose that in a rural region, electricity power failures occur following a Poisson distribution with an average of 2 failures every week. Calculate the probability that on a given day the electricity fails three times.

We are given the weekly failure rate, but to answer this question we need to first calculate the average rate of failure on a given day: λ_{day} = 2/7 = 0.2857. Note that we are assuming that the probability 7 of power failure is the same on any day of the week, i.e. we assume independence.

$$P(\text{3 failures on a given day}) = \frac{0.2857^1 \times e^{-0.2857}}{3!}$$

Example #2

Suppose that in a rural region, electricity power failures occur following a Poisson distribution with an average of 2 failures every week. Calculate the probability that on a given day the electricity fails three times.

We are given the weekly failure rate, but to answer this question we need to first calculate the average rate of failure on a given day: λ_{day} = 2/7 = 0.2857. Note that we are assuming that the probability 7 of power failure is the same on any day of the week, i.e. we assume independence.

$$
\begin{aligned}
P(\text{3 failures on a given day}) &= \frac{0.2857^1 \times e^{-0.2857}}{3!} \\
&= \frac{0.2857 \times e^{-0.2857}}{6}
\end{aligned}
$$

Example #2

Suppose that in a rural region, electricity power failures occur following a Poisson distribution with an average of 2 failures every week. Calculate the probability that on a given day the electricity fails three times.

We are given the weekly failure rate, but to answer this question we need to first calculate the average rate of failure on a given day: λ_{day} = 2/7 = 0.2857. Note that we are assuming that the probability 7 of power failure is the same on any day of the week, i.e. we assume independence.

$$
\begin{aligned}
P(\text{3 failures on a given day}) &= \frac{0.2857^1 \times e^{-0.2857}}{3!} \\
&= \frac{0.2857 \times e^{-0.2857}}{6} \\
&= 0.0358
\end{aligned}
$$

Is it Poisson?

- A random variable may follow a Poisson distribution if the event being considered is rare, the population is large, and the events occur independently of each other

- However we can think of situations where the events are not really independent. For example, if we are interested in the probability of a certain number of weddings over one summer, we should take into consideration that weekends are more popular for weddings.

- In this case, a Poisson model may sometimes still be reasonable if we allow it to have a different rate for different times; we could model the rate as higher on weekends than on weekdays.

- The idea of modeling rates for a Poisson distribution against a second variable (day of the week) forms the foundation of some more advanced methods called *generalized linear models*.

Exponential Distribution

Suppose that we just had an electricity power failure. What is the expected time until the next failure?

Exponential Distribution

Suppose that we just had an electricity power failure. What is the expected time until the next failure?

Let $\lambda > 0$. A random variable X follows the exponential distribution with parameter λ if it has a continuous distribution with pf:

$$f(x \mid \lambda) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Parameter space: $\lambda \in [0, \infty)$.

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

Exponential Distribution

Relationship of the Poisson and Exponential distribution

X: Time between two consecutive arrivals

N_t: Number of arrivals during time t

$$P(X > x) = P(N_t = N_{t+x}) = P(N_x = 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda}$$

$$P(X \leq x) = 1 - P(X < x) = 1 - e^{-\lambda}$$

Memorylessness of the Exponential

$$
\begin{aligned}
P(X > x + a \mid X > a) &= \frac{P(X > x + a, X > a)}{P(X > a)} \\
&= \frac{P(X > x + a)}{P(X > a)} \\
&= \frac{1 - F_X(x + a)}{1 - F_X(a)} \\
&= \frac{e^{-\lambda(x+a)}}{e^{-\lambda a}} \\
&= e^{-\lambda x} \\
&= P(X > x).
\end{aligned}
$$

Poisson Example

The number of emails that I get in a weekday can be modelled by a Poisson distribution with an average of 0.2 emails per minute.

1. What is the probability that I get no emails in an interval of length 5 minutes?
2. What is the probability that I get more than 3 emails in an interval of length 10 minutes?
3. I just got an email. What is the probability that I will wait more than 3 minutes until the next email?