Andrew Gard - equitable.equations@gmail.com

The Poisson Distribution

Suppose we're interested in the number of times a certain event occurs over a specified period.

Suppose we're interested in the number of times a certain event occurs over a specified period. For instance,

- The number of calls to a customer-service line in a ten-minute interval

Suppose we're interested in the number of times a certain event occurs over a specified period. For instance,

- The number of calls to a customer-service line in a ten-minute interval
- The number of accidents along a busy stretch of road in an hour

Suppose we're interested in the number of times a certain event occurs over a specified period. For instance,

- The number of calls to a customer-service line in a ten-minute interval
- The number of accidents along a busy stretch of road in an hour
- The number of mistakes made by an amateur typist in a 100-word paragraph

Suppose we're interested in the number of times a certain event occurs over a specified period. For instance,

- The number of calls to a customer-service line in a ten-minute interval
- The number of accidents along a busy stretch of road in an hour
- The number of mistakes made by an amateur typist in a 100-word paragraph

The number of occurrences of an event in a specified period can be viewed as a discrete random variable X whose support (list of possible values) is the non-negative integers.

Suppose we're interested in the number of times a certain event occurs over a specified period. For instance,

- The number of calls to a customer-service line in a ten-minute interval
- The number of accidents along a busy stretch of road in an hour
- The number of mistakes made by an amateur typist in a 100-word paragraph

The number of occurrences of an event in a specified period can be viewed as a discrete random variable X whose support (list of possible values) is the non-negative integers. If the events are independent, occur at a constant average rate, and satisfy a few technical axioms which I'll list at the end of the vid, then X is said to have a **Poisson distribution.**

The probability mass function (or pmf) of a random variable X with a Poisson distribution is

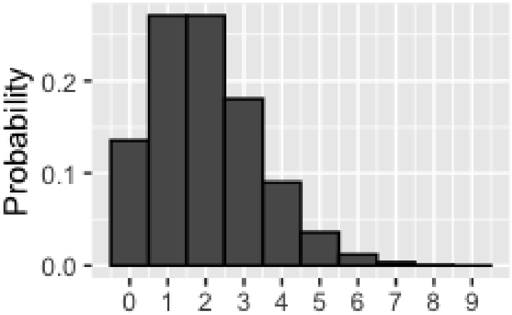
$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where λ is the mean number of occurrences per interval.

The probability mass function (or pmf) of a random variable X with a Poisson distribution is

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where λ is the mean number of occurrences per interval. For instance, if $\lambda = 2$, then the distribution of X looks like this:



Notice the strong right skew and the mode centered just to the left of the mean $x = 2$.

Example. Around dinnertime, calls to a pizza shop come at an average rate of 1 every 5 minutes. What is the probability there will be exactly 4 calls between 6:40 and 7:00pm? That there will be less than 4? More than 4?

Example. Around dinnertime, calls to a pizza shop come at an average rate of 1 every 5 minutes. What is the probability there will be exactly 4 calls between 6:40 and 7:00pm? That there will be less than 4? More than 4?

First, we need to find λ, the average number of calls in the interval of time in question, which in this case is 20 minutes.

Example. Around dinnertime, calls to a pizza shop come at an average rate of 1 every 5 minutes. What is the probability there will be exactly 4 calls between 6:40 and 7:00pm? That there will be less than 4? More than 4?

First, we need to find λ, the average number of calls in the interval of time in question, which in this case is 20 minutes. Since calls come at an average rate of 1 every 5 minutes, this average must be $\lambda = 4$.

Example. Around dinnertime, calls to a pizza shop come at an average rate of 1 every 5 minutes. What is the probability there will be exactly 4 calls between 6:40 and 7:00pm? That there will be less than 4? More than 4?

First, we need to find λ, the average number of calls in the interval of time in question, which in this case is 20 minutes. Since calls come at an average rate of 1 every 5 minutes, this average must be $\lambda = 4$. Thus

$$P(x=4) \quad = \quad \frac{\lambda^x e^{-\lambda}}{x!} \quad = \quad \frac{4^4 e^{-4}}{4!}$$

Example. Around dinnertime, calls to a pizza shop come at an average rate of 1 every 5 minutes. What is the probability there will be exactly 4 calls between 6:40 and 7:00pm? That there will be less than 4? More than 4?

First, we need to find λ, the average number of calls in the interval of time in question, which in this case is 20 minutes. Since calls come at an average rate of 1 every 5 minutes, this average must be $\lambda = 4$. Thus

$$P(x = 4) \;=\; \frac{\lambda^x e^{-\lambda}}{x!} \;=\; \frac{4^4 e^{-4}}{4!} \;\approx\; .195$$

Example. Around dinnertime, calls to a pizza shop come at an average rate of 1 every 5 minutes. What is the probability there will be exactly 4 calls between 6:40 and 7:00pm? That there will be less than 4? More than 4?

First, we need to find λ, the average number of calls in the interval of time in question, which in this case is 20 minutes. Since calls come at an average rate of 1 every 5 minutes, this average must be $\lambda = 4$. Thus

$$
\begin{aligned}
P(x = 4) &= \frac{\lambda^x e^{-\lambda}}{x!} = \frac{4^4 e^{-4}}{4!} \approx .195 \\
P(x < 4) &= \sum_{x=0}^{3} \frac{\lambda^x e^{-\lambda}}{x!}
\end{aligned}
$$

Example. Around dinnertime, calls to a pizza shop come at an average rate of 1 every 5 minutes. What is the probability there will be exactly 4 calls between 6:40 and 7:00pm? That there will be less than 4? More than 4?

First, we need to find λ, the average number of calls in the interval of time in question, which in this case is 20 minutes. Since calls come at an average rate of 1 every 5 minutes, this average must be $\lambda = 4$. Thus

$$P(x=4) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{4^4 e^{-4}}{4!} \approx .195$$

$$P(x<4) = \sum_{x=0}^{3} \frac{\lambda^x e^{-\lambda}}{x!} = \frac{4^0 e^{-4}}{0!} + \frac{4^1 e^{-4}}{1!} + \frac{4^2 e^{-4}}{2!} + \frac{4^3 e^{-4}}{3!}$$

Example. Around dinnertime, calls to a pizza shop come at an average rate of 1 every 5 minutes. What is the probability there will be exactly 4 calls between 6:40 and 7:00pm? That there will be less than 4? More than 4?

First, we need to find λ, the average number of calls in the interval of time in question, which in this case is 20 minutes. Since calls come at an average rate of 1 every 5 minutes, this average must be $\lambda = 4$. Thus

$$
\begin{aligned}
P(x = 4) &= \frac{\lambda^x e^{-\lambda}}{x!} = \frac{4^4 e^{-4}}{4!} \approx .195 \\
P(x < 4) &= \sum_{x=0}^{3} \frac{\lambda^x e^{-\lambda}}{x!} = \frac{4^0 e^{-4}}{0!} + \frac{4^1 e^{-4}}{1!} + \frac{4^2 e^{-4}}{2!} + \frac{4^3 e^{-4}}{3!} \\
&\approx .018 + .073 + .147 + .195 \approx .433
\end{aligned}
$$

Example. Around dinnertime, calls to a pizza shop come at an average rate of 1 every 5 minutes. What is the probability there will be exactly 4 calls between 6:40 and 7:00pm? That there will be less than 4? More than 4?

First, we need to find λ, the average number of calls in the interval of time in question, which in this case is 20 minutes. Since calls come at an average rate of 1 every 5 minutes, this average must be $\lambda = 4$. Thus

$$
\begin{aligned}
P(x=4) &= \frac{\lambda^x e^{-\lambda}}{x!} = \frac{4^4 e^{-4}}{4!} \approx .195 \\
P(x<4) &= \sum_{x=0}^{3} \frac{\lambda^x e^{-\lambda}}{x!} = \frac{4^0 e^{-4}}{0!} + \frac{4^1 e^{-4}}{1!} + \frac{4^2 e^{-4}}{2!} + \frac{4^3 e^{-4}}{3!} \\
&\approx .018 + .073 + .147 + .195 \approx .433 \\
P(x>4) &= 1 - P(x \leq 4)
\end{aligned}
$$

Example. Around dinnertime, calls to a pizza shop come at an average rate of 1 every 5 minutes. What is the probability there will be exactly 4 calls between 6:40 and 7:00pm? That there will be less than 4? More than 4?

First, we need to find λ, the average number of calls in the interval of time in question, which in this case is 20 minutes. Since calls come at an average rate of 1 every 5 minutes, this average must be $\lambda = 4$. Thus

$$P(x = 4) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{4^4 e^{-4}}{4!} \approx .195$$

$$P(x < 4) = \sum_{x=0}^{3} \frac{\lambda^x e^{-\lambda}}{x!} = \frac{4^0 e^{-4}}{0!} + \frac{4^1 e^{-4}}{1!} + \frac{4^2 e^{-4}}{2!} + \frac{4^3 e^{-4}}{3!}$$

$$\approx .018 + .073 + .147 + .195 \approx .433$$

$$P(x > 4) = 1 - P(x \leq 4) \approx 1 - .195 - .433$$

Example. Around dinnertime, calls to a pizza shop come at an average rate of 1 every 5 minutes. What is the probability there will be exactly 4 calls between 6:40 and 7:00pm? That there will be less than 4? More than 4?

First, we need to find λ, the average number of calls in the interval of time in question, which in this case is 20 minutes. Since calls come at an average rate of 1 every 5 minutes, this average must be $\lambda = 4$. Thus

$$
\begin{aligned}
P(x = 4) &= \frac{\lambda^x e^{-\lambda}}{x!} = \frac{4^4 e^{-4}}{4!} \approx .195 \\
P(x < 4) &= \sum_{x=0}^{3} \frac{\lambda^x e^{-\lambda}}{x!} = \frac{4^0 e^{-4}}{0!} + \frac{4^1 e^{-4}}{1!} + \frac{4^2 e^{-4}}{2!} + \frac{4^3 e^{-4}}{3!} \\
&\approx .018 + .073 + .147 + .195 \approx .433 \\
P(x > 4) &= 1 - P(x \leq 4) \approx 1 - .195 - .433 \approx .372
\end{aligned}
$$

Example. A bookstore keeps 5 copies of a popular cookbook in stock. Suppose the number of sales in a day follows a Poisson distribution with mean 4. What is the probability that the store can't meet demand on a given day?

Example. A bookstore keeps 5 copies of a popular cookbook in stock. Suppose the number of sales in a day follows a Poisson distribution with mean 4. What is the probability that the store can't meet demand on a given day?

$$P(X > 5) \; = \; 1 - P(X \leq 5)$$

Example. A bookstore keeps 5 copies of a popular cookbook in stock. Suppose the number of sales in a day follows a Poisson distribution with mean 4. What is the probability that the store can't meet demand on a given day?

$$P(X > 5) \;=\; 1 - P(X \le 5) \;=\; 1 - \sum_{x=0}^{5} \frac{4^x e^{-4}}{x!}$$

Example. A bookstore keeps 5 copies of a popular cookbook in stock. Suppose the number of sales in a day follows a Poisson distribution with mean 4. What is the probability that the store can't meet demand on a given day?

$$P(X > 5) \;=\; 1 - P(X \leq 5) \;=\; 1 - \sum_{x=0}^{5} \frac{4^x e^{-4}}{x!} \;\approx\; .215$$

Example. A bookstore keeps 5 copies of a popular cookbook in stock. Suppose the number of sales in a day follows a Poisson distribution with mean 4. What is the probability that the store can't meet demand on a given day?

$$P(X > 5) \;=\; 1 - P(X \leq 5) \;=\; 1 - \sum_{x=0}^{5} \frac{4^x e^{-4}}{x!} \;\approx\; .215$$

How many copies should they keep in stock so that the probability of turning a customer away is below 10%?

Example. A bookstore keeps 5 copies of a popular cookbook in stock. Suppose the number of sales in a day follows a Poisson distribution with mean 4. What is the probability that the store can't meet demand on a given day?

$$P(X > 5) \;=\; 1 - P(X \leq 5) \;=\; 1 - \sum_{x=0}^{5} \frac{4^x e^{-4}}{x!} \;\approx\; .215$$

How many copies should they keep in stock so that the probability of turning a customer away is below 10%?

We need to find a value x such that $P(X > x) < .10$, or equivalently, such that $P(X \leq x) \geq .90$.

Example. A bookstore keeps 5 copies of a popular cookbook in stock. Suppose the number of sales in a day follows a Poisson distribution with mean 4. What is the probability that the store can't meet demand on a given day?

$$P(X > 5) \;=\; 1 - P(X \leq 5) \;=\; 1 - \sum_{x=0}^{5} \frac{4^x e^{-4}}{x!} \;\approx\; .215$$

How many copies should they keep in stock so that the probability of turning a customer away is below 10%?

We need to find a value x such that $P(X > x) < .10$, or equivalently, such that $P(X \leq x) \geq .90$. We can do this *inverse Poisson calculation* using technology or by manually constructing the cumulative probability distribution for X.

When $\lambda = 4$, we have the following probabilities:

x	0	1	2	3	4	5	6	7	8
$P(X \leq x)$	.018	.092	.238	.433	.629	.785	.889	.949	.979

When $\lambda = 4$, we have the following probabilities:

x	0	1	2	3	4	5	6	7	8
$P(X \leq x)$	.018	.092	.238	.433	.629	.785	.889	.949	.979

The probability of selling at most 6 copies is about 89%, while the probability of selling at most 7 copies is about 95%. If the store keeps 7 copies in stock, the probability of turning a customer away is well under 10%.

When $\lambda = 4$, we have the following probabilities:

x	0	1	2	3	4	5	6	7	8
$P(X \leq x)$	.018	.092	.238	.433	.629	.785	.889	.949	.979

The probability of selling at most 6 copies is about 89%, while the probability of selling at most 7 copies is about 95%. If the store keeps 7 copies in stock, the probability of turning a customer away is well under 10%.

In *R*, we would solve this problem with *qpois()* command.

```
qpois(.9, 4)
```

```
## [1] 7
```

Most theoretical derivations involving Poisson distributions rely on the MacLaurin series for the exponential function.

$$e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!}$$

Most theoretical derivations involving Poisson distributions rely on the MacLaurin series for the exponential function.

$$e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!}$$

For instance, we compute the moment-generating function (mgf) of a random variable with a Poisson distribution as follows:

$$M(t) \;=\; E\left(e^{tX}\right) \;=\; \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!}$$

Most theoretical derivations involving Poisson distributions rely on the MacLaurin series for the exponential function.

$$e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!}$$

For instance, we compute the moment-generating function (mgf) of a random variable with a Poisson distribution as follows:

$$\begin{aligned} M(t) &= E\left(e^{tX}\right) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \end{aligned}$$

Most theoretical derivations involving Poisson distributions rely on the MacLaurin series for the exponential function.

$$e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!}$$

For instance, we compute the moment-generating function (mgf) of a random variable with a Poisson distribution as follows:

$$\begin{aligned} M(t) &= E\left(e^{tX}\right) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} \end{aligned}$$

Most theoretical derivations involving Poisson distributions rely on the MacLaurin series for the exponential function.

$$e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!}$$

For instance, we compute the moment-generating function (mgf) of a random variable with a Poisson distribution as follows:

$$\begin{aligned} M(t) &= E\left(e^{tX}\right) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} \\ &= e^{\lambda(e^t - 1)} \end{aligned}$$

The mgf $M(t)$ immediately gives us the expected value and variance of the Poisson distribution with parameter λ:

$$\mu = M'(0) = \lambda, \qquad \sigma^2 = M''(0) = \lambda^2$$

Both the mean and standard deviation are equal to the average number of occurrences per unit of time.

Example. A machine producing cardboard boxes averages 2 flawed boxes per hour.

1. What are the expected value and standard deviation of the number of flawed boxes produced in an 8-hour shift?

Example. A machine producing cardboard boxes averages 2 flawed boxes per hour.

1. What are the expected value and standard deviation of the number of flawed boxes produced in an 8-hour shift?

$$\begin{aligned} \mu &= \lambda = 8 \cdot 2 = 16 \\ \sigma &= \lambda = 16 \end{aligned}$$

Example. A machine producing cardboard boxes averages 2 flawed boxes per hour.

1. What are the expected value and standard deviation of the number of flawed boxes produced in an 8-hour shift?

$$\begin{aligned} \mu &= \lambda = 8 \cdot 2 = 16 \\ \sigma &= \lambda = 16 \end{aligned}$$

2. What is the probability the machine produces between 12 and 20 flawed boxes in an 8-hour shift?

Example. A machine producing cardboard boxes averages 2 flawed boxes per hour.

1. What are the expected value and standard deviation of the number of flawed boxes produced in an 8-hour shift?

$$\begin{aligned} \mu &= \lambda = 8 \cdot 2 = 16 \\ \sigma &= \lambda = 16 \end{aligned}$$

2. What is the probability the machine produces between 12 and 20 flawed boxes in an 8-hour shift?

$$P(12 < X < 20) = \sum_{x=13}^{19} \frac{16^x e^{-16}}{x!} = .619$$

The Poisson distribution is closely related to several other distributions. Two are particularly worth noting.

The Poisson distribution is closely related to several other distributions. Two are particularly worth noting.

- The **gamma distribution** models the waiting time until a given total number of occurrences of an event under the same assumptions as the Poisson distribution. Notice that waiting time is a *continuous* random variable.

The Poisson distribution is closely related to several other distributions. Two are particularly worth noting.

- The **gamma distribution** models the waiting time until a given total number of occurrences of an event under the same assumptions as the Poisson distribution. Notice that waiting time is a *continuous* random variable.
- The **binomial distribution** models the the number of occurrences of an event over a finite number of trials when the event has a fixed probability in each trial.

The Poisson distribution is closely related to several other distributions. Two are particularly worth noting.

- The **gamma distribution** models the waiting time until a given total number of occurrences of an event under the same assumptions as the Poisson distribution. Notice that waiting time is a *continuous* random variable.
- The **binomial distribution** models the the number of occurrences of an event over a finite number of trials when the event has a fixed probability in each trial. By contrast, there are no individual trials for random variables with Poisson distributions.

Let's conclude with a quick look at the technical requirements for a random variable to have a Poisson distribution with parameter λ.

Let's conclude with a quick look at the technical requirements for a random variable to have a Poisson distribution with parameter λ.

- The numbers of occurrences in disjoint subintervals are independent.

Let's conclude with a quick look at the technical requirements for a random variable to have a Poisson distribution with parameter λ.

- The numbers of occurrences in disjoint subintervals are independent.
- The probability of exactly one occurrence in a sufficiently small subinterval of width Δt is approximately $\lambda \Delta t$.

Let's conclude with a quick look at the technical requirements for a random variable to have a Poisson distribution with parameter λ.

- The numbers of occurrences in disjoint subintervals are independent.
- The probability of exactly one occurrence in a sufficiently small subinterval of width Δt is approximately $\lambda \Delta t$.
- The probability of more than one occurrence in a sufficiently small subinterval of width Δt is negligible.

Let's conclude with a quick look at the technical requirements for a random variable to have a Poisson distribution with parameter λ.

- The numbers of occurrences in disjoint subintervals are independent.
- The probability of exactly one occurrence in a sufficiently small subinterval of width Δt is approximately $\lambda \Delta t$.
- The probability of more than one occurrence in a sufficiently small subinterval of width Δt is negligible.

In practice, these assumptions are never quite fully met. The phrase "approximate Poisson process" is commonly used to describe situations in which they are reasonable simplifying assumptions.