

Poisson Distribution

The Poisson Distribution is used to model the number of rare events that occur in space, time, volume, or any other dimension. For this distribution,

- The average number of events in a given time period or space is λ .
- The number of events in non-overlapping time periods or spaces are independent.
- The probability of one event in a short time period or space h is λh .
- The probability of more than one event in a short time period or space h is 0.
- The random variable Y = number of events in a given time period or space.
- The parameter for the Poisson random variable Y is λ .
- The probability distribution function for the Poisson random variable Y is

$$P(Y = y) = p(y) = \frac{\lambda^y e^{-\lambda}}{y!} \quad y = 0, 1, 2, \dots$$

- The theoretical mean of the Poisson random variable Y is

$$E(Y) = \lambda$$

- The theoretical variance of the Poisson random variable Y is

$$V(Y) = \lambda$$

A Poisson distribution can be used to model the number of accidents that occur within a week at a given intersection, the number of telephone calls handled by a switchboard in a given time interval, the number of radioactive particles that decay in a particular time period, the number of errors a typist makes in typing a page, and the number of automobiles using a freeway access ramp in a 10-minute interval.

Unlike the other discrete distributions we've studied, we must either assume the random variable Y has a Poisson distribution (usually based on the information given above) or we must check that the Poisson distribution would seem a reasonable model to use based on an analysis of collected data.

Working with Poisson random variables in R.

To find a probability $P(Y = y) = p(y)$ for a single value y , the command in R is

```
dpois(y,lambda)
```

To find the probability $P(Y \leq y)$, use the sum command to add up all $p(y)$ values for y between and including 0 and y .

```
sum(dpois(0:y,lambda))
```

To find the probability $P(y_1 \leq Y \leq y_2)$, use the sum command to add up all $p(y)$ values for y between and including y_1 and y_2 .

```
sum(dpois(y1:y2,lambda))
```

To find the probability $P(Y \geq y) = 1 - P(Y < y) = 1 - P(Y \leq y - 1)$, use the sum command to find $P(Y \leq y - 1)$ and subtract this value from 1.

```
1 - sum(dpois(0:y-1,lambda))
```

To generate values from a Poisson random variable Y with mean λ , the command in R is

```
rpois(numobs,lambda)
```

where **numobs** is the number of observations of the random variable Y you would like to generate.

Problems.

1. Show that $p(y)$ has the two properties of a probability distribution function.
2. Derive the expected value of a Poisson random variable.
3. Use R to generate 10,000 observations from a Poisson random variable with $\lambda = 3$. Make a histogram of your data and calculate the mean, variance, and five number summary of your observations. Use this information to describe the distribution of your data.
4. Now use R to generate 10,000 observations from a Poisson random variable with $\lambda = 10$. Make a histogram of your data and calculate the mean, variance, and five number summary of your observations. How does this distribution compare with the one from problem 3.
5. A police officer visits a location $y = 0, 1, 2, 3, \dots$ times every half-hour. On average, the police officer visits a location once every half-hour. For the next half-hour time period, find the probability the police officer will
 - (a) miss a particular location.
 - (b) visit the location once.
 - (c) visit the location twice.
 - (d) visit the location at least once.
6. Tree seedlings are randomly dispersed in a particular area with an average dispersment of 5 seedlings per square yard.
 - (a) In a one square yard area, what is the probability that no seedlings are found?
 - (b) In ten one square yard areas, what is the probability that no seedlings are found?
 - (c) In ten one square yard areas, what is the probability that at least one out of the ten areas contains seedlings?
 - (d) In ten one square yard area, what is the probability that exactly three out of the ten areas contains seedlings?