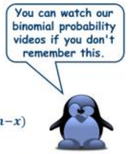
Introduction to Poisson probability

$$p(x) = \frac{e^{-\bar{x}}\bar{x}^x}{x!}$$

THE POISSON COMES FROM THE BINOMIAL



The **binomial probability** of seeing x successes in n trials when the success probability is p is:

$$p(x) = \binom{n}{x} p^x (1-p)^{(n-x)} = \frac{n!}{n!\,(n-x)!} p^x (1-p)^{(n-x)}$$

When n gets large and p gets small (i.e., n>100, np<10), this equation simplifies into the **Poisson probability**:

$$p(x) = \frac{\mu^{-x} e^{-\mu}}{x!} = \frac{\bar{x}^{-x} e^{-\bar{x}}}{x!}$$

THE POISSON PROBABILITY

$$p(x) = \frac{\mu^{-x} e^{-\mu}}{x!} = \frac{\bar{x}^{-x} e^{-\bar{x}}}{x!}$$

Unlike the binomial which requires a set number of trials and individual probabilities known, the Poisson does not.

Used for the probability of seeing X events (i.e., successes) in an area or over a set time period when we know the mean.

The potential number can be unknowable, so binomial is not appropriate, but the mean is easier to determine.

POISSON PROBABILITY SCENARIOS

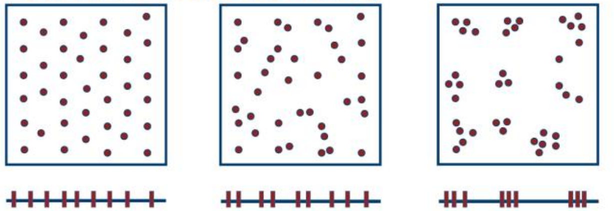
The probability of seeing X events in an area:

If we know that the average number of snails per square meter is 3, what are the odds of seeing none?

The probability of seeing X events in a set time period:

e.g., If I know there is an average of 5 deaths in a retirement home per month, what are the chances of seeing 10?

COEFFICIENT OF DISPERSION (CD)



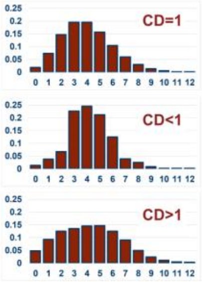
Over-dispersed or Uniform (CD<1)

Random or Poisson (CD=1)

Under-dispersed or Clumped (CD>1)

COEFFICIENT OF DISPERSION (CD)

$$CD = \frac{s^2}{\bar{x}} = \frac{\sigma^2}{\mu}$$



Poisson distribution has a CD=1.

CD =1, distribution is *probably* Poisson.

CD < 1, distribution is **under-dispersed** or **uniform**.

CD > 1, distribution is **over-dispersed** or **clumped**.

POISSON PROBABILITY, USEFUL PROPERTY

MEAN = VARIANCE

Entire distribution can therefore be specified with one value.

This can be used to test hypotheses about whether a distribution is due to Poisson (i.e., random) processes.

$$\bar{x} = s^2\ ?$$

This is usually tested using the **coefficient of dispersion**:

$$CD = \frac{s^2}{\bar{x}} = 1?$$

POISSON PROBABILITY ASSUMPTIONS

Assumptions:

- ▶ Events occur randomly with respect to one another (from binomial independence assumption).
- ▶ Events are relatively rare.
- ▶ Probability of occurrence doesn't change over time (from binomial assumption of constant probability).

$$\lim_{p \to 0, n \to \infty} (binomial) = Poisson$$

ANOTHER USEFUL POISSON PROPERTY

Consecutive Poisson probabilities are related to each other:

$$p(x) = \frac{\mu^x e^{-\mu}}{x!} = \frac{\mu}{x} \times \frac{\mu^{(x-1)} e^{-\mu}}{(x-1)!} = \frac{\mu}{x} \times p(x-1)$$

e.g, if μ=3

$$p(1) = \frac{3^1 e^{-3}}{1!} = \frac{3}{(1)e^3} = 0.1494$$

$$p(2) = \frac{3^2 e^{-3}}{2!} = \frac{9}{(2 \times 1)e^2} = 0.2240$$

$$p(2) = \frac{\mu}{x} \times p(1) = \frac{3}{2} p(1) = \frac{3}{2}(0.1494) = 0.2240$$

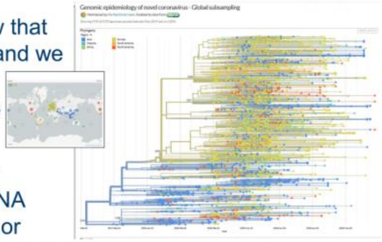
USING THE POISSON PROBABILITY

Application: if we know that a process is random and we have a mean, then we can predict probabilities and proportions of numbers of observations.

Can determine whether individuals are located randomly or due to non-random factors (e.g., soil type, depth).

USING THE POISSON PROBABILITY

https://nextstrain.org/ncov/global



Application: if we know that the process is random and we have a mean, then we can predict probabilities of numbers of events.

Can determine whether observed changes in DNA sequences are random or due to non-random process (e.g., natural selection)

USING THE POISSON PROBABILITY

Application: if we think a process might be random (or non-random), we can measure the distribution and compare it to the predictions from the Poisson model (i.e., is the CD=1?).

When we see unusually large values or apparent patterns, are they genuine or just random?

StatsExamples.com