Recap :

- Binomial distribution : $X \sim \text{binomial}(n, p)$

n: # of experiments

p: probability of success.

$X :=$ The number of successes out of "n" Bernoulli trials. $\boxed{x = 0, 1, 2, \ldots, n}$

- each trial is independent of the other trials

- The probability of success, p, is the same over all n trials.

- No closed form CDF. (e.g $P(X \le 4) = P(x=0$ or $x=1$ or $x=2$ or $x=3$ or $x=4)$)

- $EX = n \cdot p$

- $\text{var}X = np(1-p)$

- $SD(X) = \sqrt{\text{var}X} = \sqrt{np(1-p)}$

- Geometric distribution : $X \sim \text{Geom}(p)$

$X :=$ The number of trials until observing the first success. $\boxed{x = 1, 2, 3, \ldots}$

- Each trial is independent of others.

- The prob. of success, p, is the same for all trials.

(e.g $P(X \leq 20) = 1 - (1-p)^{20}$) ← - Closed form CDF : $F_X(x) = 1 - (1-p)^x$

- $EX = 1/p$

- $\text{Var}\, X = \dfrac{1-p}{p^2}$

- $SD(X) = \sqrt{\text{Var}\, X} = \sqrt{\dfrac{1-p}{p^2}}$

Mean

and

Variance

of Geometric Distrbution

The Geometric Distribution

$X \sim Geom.(p)$

Expected value:

$$E(X) = \frac{1}{p}$$

Variance:

$$Var(X) = \frac{1-p}{p^2}$$

Common Distributions

Background

Bernoulli

Binomial

Geometric

Example

NiCad batteries: An experimental program was successful in reducing the percentage of manufactured NiCad cells with internal shorts to around 1%. Let T be the test number at which the first short is discovered. Then, $T \sim \text{Geom}(p)$.

Calculate

$T \sim \text{Geom}(0.01)$, $f_X(x) = p(1-p)^{x-1}$, $x = 1, 2, 3, \ldots$

- $P(\text{1st or 2nd cell tested has the 1st short})$

$$P(T=1 \text{ or } T=2) = P(T=1) + P(T=2)$$

$$= f(1) + f(2) = p(1-p)^{1-1} + p(1-p)^{2-1} = (0.01)(0.99)^0 + (0.01)(0.99)^1 = 0.02$$

- $P(\text{at least 50 cells tested without finding a short})$

$$P(T > 50) = 1 - P(T \leq 50)$$

$$= 1 - F_X(50) = 1 - \left[1 - (1-0.01)^{50}\right]$$

$$= (1-0.01)^{50} = (0.99)^{50} = 0.61$$

Common Distributions

Background

Bernoulli

Binomial

Geometric

Example

NiCad batteries: $T \sim \text{Geom}(p=0.01)$

Calculate the expected test number at which the first short is discovered and the variance in test numbers at which the first short is discovered.

$$E(T) = \frac{1}{p} = \frac{1}{0.01} = 100$$

(on average, you need to test 100 batteries until you observe the first short)

$$\text{Var}\,T = \frac{1-p}{p^2} = \frac{1-0.01}{(0.01)^2} = \frac{0.99}{(0.01)^2} = 9900$$

$$SD(T) = \sqrt{\text{Var}\,T} = \sqrt{9900}$$

Example

A shipment of 200 widgets arrives from a new widget distributor. The distributor has claimed that the widgets there is only a 10% defective rate on the widgets. Let X be the random variable asociated with the number of trials untill finding the first defective widgets.

- What is the probability distribution associated with this random variable X? Precisely specify the parameter(s).

  $X \sim \text{Geom}(p = 0.1)$

- How many widgets would you expect to test before finding the first defective widget?

$$EX = \frac{1}{p} = \frac{1}{0.1} = 10$$

So, we expect to test 10 widgets (on average) until finding the first defective one.

Common Distributions

Background

Bernoulli

Binomial

Geometric

Example

You find your first defective widget while testing the thrid widget.

- What is the probability that a the first defective widget would be found **on** the third test if there are only 10% defective widgets from in the shipment?

$$P(x = 3) = p(1-p)^{x-1}$$

$$= 0.1(1-0.1)^{3-1}$$

$$= 0.1(0.9)^2 = \boxed{0.081}$$

Common Distributions

Background

Bernoulli

Binomial

Geometric

Example

- What is the probability that a the first defective widget would be found **by** the third test if there are only 10% defective widgets from in the shipment?

$$P(\cancel{x} \leq 3) = F_X(3) = 1 - (1 - p)^3$$

$$= 1 - (1 - .1)^3$$

$$= 1 - (0.9)^3 = \boxed{0.271}$$

in Geom: $F_X(x) = 1 - (1-p)^x$

in Geom: pmf: $f_X(x) := P_X(x) = p(1-p)^{x-1}$; $x = 1, 2, 3, \ldots$

The Poisson Distribution

Common Distributions

Background

Bernoulli

Binomial

Geometric

Poisson

The Poisson Distribution

Origin: A rare occurance is watched for over a specified interval of time or space.

It's often important to keep track of the total number of occurrences of some relatively rare phenomenon.

Definition

Consider a variable

> X : the count of occurences of a phenomenon across a specified interval of time or space

or

> X: the number of times the rare occurance is observed

Common Distributions

Background

Bernoulli

Binomial

Geometric

Poisson

The Poisson Distribution

Parameter

Poisson(λ)

probability function:

The **Poisson**~~(λ)~~ (λ) distribution is a discrete probability distribution with pmf

$$f(x) = \begin{cases} \boxed{\dfrac{e^{-\lambda}\lambda^x}{x!}} & x = 0, 1, \ldots \\ 0 & o.w. \end{cases}$$

For $\boxed{\lambda > 0}$

The Poisson Distribution

These occurrences must:

- be independent
- be sequential in time (no two occurances at once)
- occur at the same constant rate λ

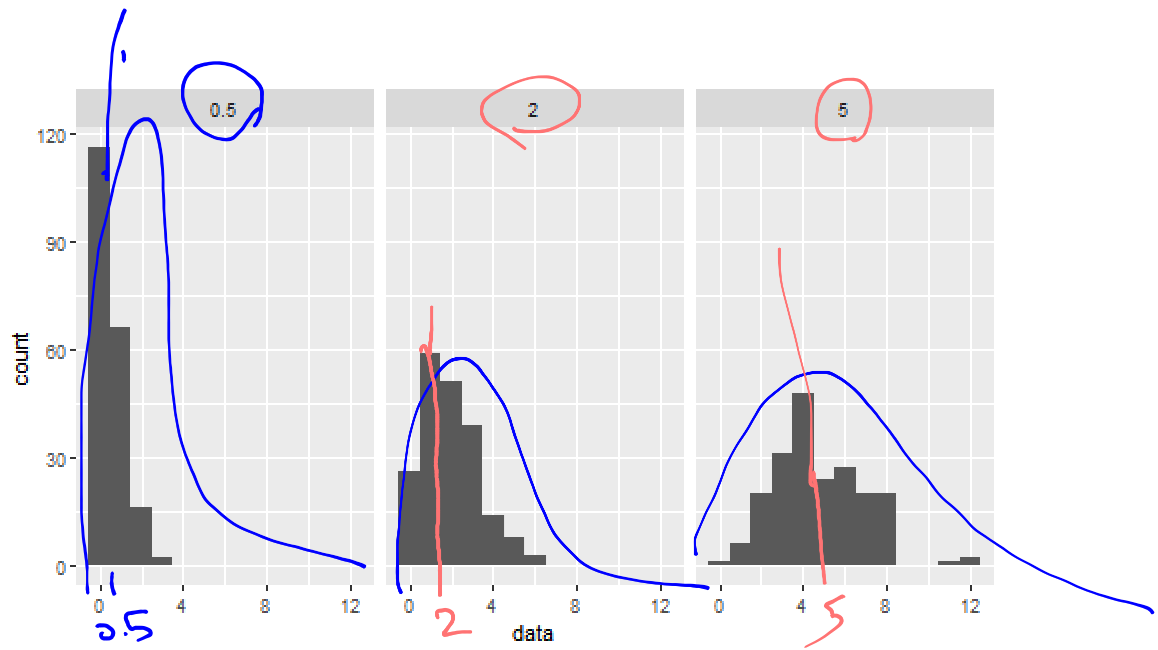
λ the $\boxed{\textit{rate parameter}}$, is the expected number of occurances in **the specified interval of time or space** (i.e $\mathrm{E(X)} = \lambda$)

The Poisson Distribution

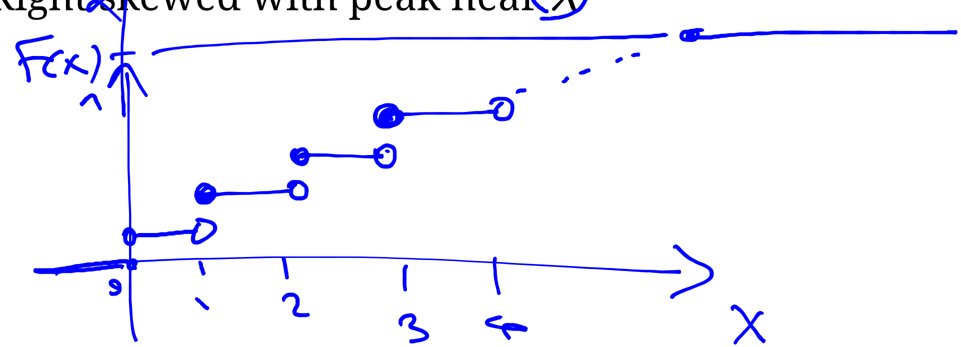
Examples that could follow a Poisson(λ) distribution :

> Y is the number of shark attacks off the coast of CA next **year**, $\lambda = 100$ attacks per year
>
> Z is the number of shark attacks off the coast of CA next **month**, $\lambda = 100/12$ attacks per month
>
> N is the number of α-particles emitted from a small bar of polonium, registered by a counter in a minute, $\lambda = 459.21$ particles per **minute**
>
> J is the number of particles per hour, $\lambda = 459.21 * 60 = 27,552.6$ particles per **hour**.

The Poisson Distribution



> Right skewed with peak near λ



The Poisson Distribution

For X a Poisson$\cancel{(\lambda)}$ (λ) random variable,

$$\mu = \mathrm{E}X = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^{x}}{x!} = \lambda$$

$$\sigma^2 = \mathrm{Var}X = \sum_{x=0}^{\infty} (x-\lambda)^2 \frac{e^{-\lambda}\lambda^{x}}{x!} = \lambda$$

Common Distributions

Background

Bernoulli

Binomial

Geometric

Poisson

Example

Arrivals at the library

Some students' data indicate that between 12:00 and (one min.) 12:10pm on Monday through Wednesday, an average of around 125 students entered Parks Library at ISU. Consider modeling

> M : the number of students entering the ISU library between 12:00 and 12:01pm next Tuesday

Model $M \sim \text{Poisson}(\lambda)$. What would a reasonable choice of λ be?

$$\lambda = \frac{125}{10}$$

Common Distributions

Background

Bernoulli

Binomial

Geometric

Poisson

Example

Arrivals at the library

Under this model, the probability that between $\boxed{\text{10 and 15}}$ students arrive at the library between 12:00 and 12:01 PM is:

$$M \sim \text{Poisson}\left(\lambda = \frac{125}{10}\right)$$

$$p_M(m) = \frac{e^{-\lambda}\lambda^m}{m!}, \quad m = 0, 1, 2, \ldots$$

$$P(10 \le M \le 15) = p(10) + p(11) + \cdots + p(15)$$

$$= \frac{e^{-12.5}(12.5)^{10}}{10!} + \frac{e^{-12.5}(12.5)^{11}}{11!} + \cdots \frac{e^{-12.5}(12.5)^{15}}{15}$$

$= 0.6$

Common Distributions

Background

Bernoulli

Binomial

Geometric

Poisson

Shark attacks

No Closed Form for CDF in Poisson!

Let X be the number of unprovoked shark attacks that will occur off the coast of Florida $\boxed{\text{next year}}$. Model

$$X \sim \text{Poisson}(\lambda).$$

From the shark data at http://www.flmnh.ufl.edu/fish/sharks/statistics/FLactivity.html, 246 unprovoked shark attacks occurred from 2000 to 2009.

What would a reasonable choice of λ be?

246 attacks per 10 years.

The rate (λ) for next year (only one year)

$$\lambda = \frac{246}{10} = 24.6$$

Shark attacks

Under this model, calculate the following:

- P(no attacks next year)

$$P(X=0) = P(0) = \frac{e^{-24.6}(24.6)^0}{0!} = e^{-24.6} \approx 2.07 \times 10^{-11}$$

- P(at least 5 attacks)

$$P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4)$$

$$= 1 - [P(X=0 \text{ or } X=1 \text{ or } X=2 \text{ or } X=3 \text{ or } X=4]$$

$$= 1 - [P(X=0) + P(X=1) + \cdots + P(X=4)]$$

- P(more than 10 attacks)

$$P(X > 10) = 1 - P(X \leq 10)$$

$$1 - \left[\frac{e^{-24.6}(24.6)^0}{0!} + \frac{e^{-24.6}(24.6)^1}{1!} + \cdots + \frac{e^{-24.6}(24.6)^4}{4!} \right]$$

$$= 1 - e^{-24.6}(3626 6812)$$

$$= 0.999249$$