Binomial Distribution

Expected Value and Variance

Common Distributions

Background

Bernoulli

Binomial

The Binomial Distribution

Expected value:

$$\boxed{E(X) = n \cdot p}$$ ✓

Variance:

$$Var(X) = n \cdot (1-p) \cdot p$$ ✓

- Recall: Bernoulli distribution:

$EX = p$ ✓ (Binomial is "n" independent Bernoulli trials): $EX = np$

$var\ x = p(1-p)$ = " " " " ": $Var(x) = np(1-p)$

Common Distributions

Background

Bernoulli

Binomial

The Binomial Distribution

Example [10 component machine]

Calculate the expected number of components to succeed and the variance.

$Y \sim \text{Binomial}(n=10, p=.95)$

$EY = n.p = 10(.95) = 9.5$

$var(Y) = n.p(1-p)$

$= 10(0.95)(0.05)$

$= .475$

standard deviation $\leftarrow SD(Y) = \sqrt{var(Y)} = \sqrt{0.475} = 0.689$

Common Distributions

Background

Bernoulli

Binomial

The Binomial Distribution

A few useful notes:

- In order to say that " X has a binomial distribution with n trials and success probability p" we write $\boxed{X \sim Binomial(n, p)}$
- If $X_1, X_2, \ldots, X_n$ are "n" independent Bernoulli random variables with the same p then $X = X_1 + X_2 + \ldots + X_n$ is a binomial random variable with n trials and success probability p.
- Again, n and p are referred to as "parameters" for the Binomial distribution. Both are considered fixed.
- Don't focus on the actual way we got the expected value - focus on the trick of trying to get part of your complicated summation to "go away" by turning it into the sum of a probability function.

$X \sim binomial(10, 0.7)$

$P(X \leq 5)$

Note: No closed form of CDF of Binomial

The Geometric Distribution

The Geometric Distribution

another generic discrete r.v:

Origin: A series of independent random experiments, or trials, are performed. Each trial results in one of two possible outcomes: successful or failure. The probability of a successful outcome, p, is the same across all trials. The trials are performed until a successful outcome is observed.

Definition: X is the trial upon which the first successful outcome is observed. X can take values $1, 2, \ldots$.

probability function:

There's only one parameter.

With $\boxed{0 < p < 1,}$

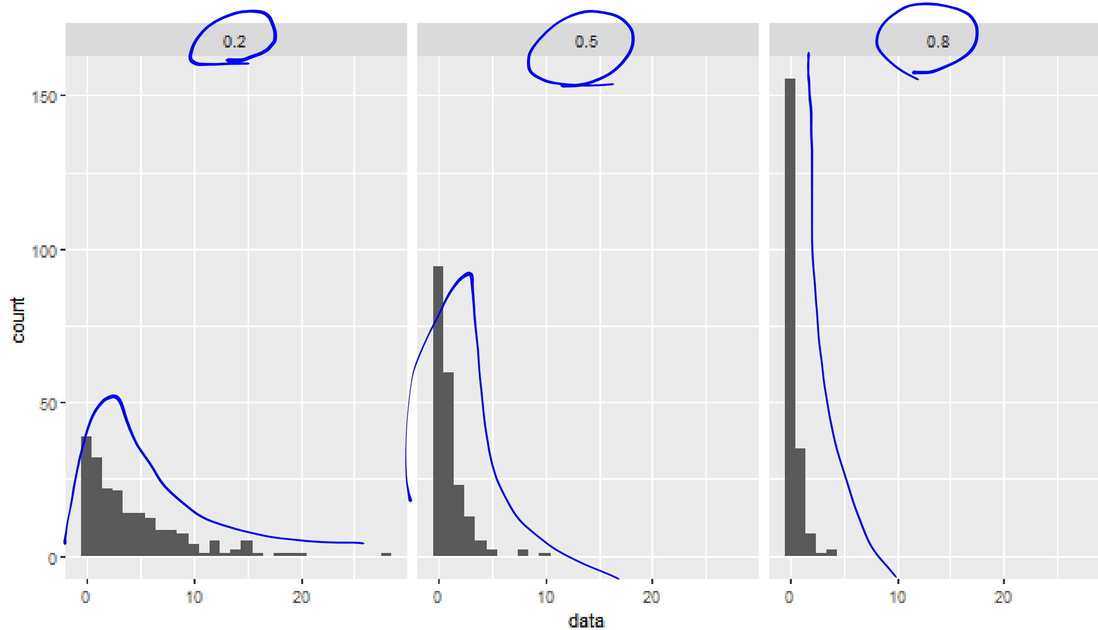
$$f(x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, \ldots, \\ 0 & o.w. \end{cases}$$

at least one trial to observe the first success.

Examples of Geometric Distribution

- Number of rolls of a fair die until you land a 5
- Number of shipments of raw materials you get until you get a defective one (**success** does not need to have positive meaning)
- Number of car engine starts untill the battery dies.

Shape of Geometric Distribution



> The probability of observing the first success decreases as the number of trials increases(even at a faster rate as p increases)

Common Distributions

Background

Bernoulli

Binomial

Geometric

The Geometric Distribution

closed form CDF

Cumulative probability function: $\boxed{F(x) = 1 - (1 - p)^x}$

Here's how we get that cumulative probability function:

optional reading

- The probability of a failed trial is $1 - p$.
- The probability the first trial fails is also just $1 - p$.
- The probability that the first two trials both fail is $(1 - p) \cdot (1 - p) = (1 - p)^2$.
- The probability that the first x trials all fail is $(1 - p)^x$
.
- This gets us to this math:

$$F(x) = P(X \leq x)$$

$$= 1 - P(X > x)$$

$$= 1 - (1 - p)^x$$

Mean

and

Variance

of Geometric Distrbution

Common Distributions

Background

Bernoulli

Binomial

Geometric

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

$T \sim \text{Geom}(0.01)$

- P(1st or 2nd cell tested has the 1st short)

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

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

- P(at least 50 cells tested without finding a short)

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

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

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

Example

NiCad batteries:

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.

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).

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

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?

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

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

$$= 0.1(0.9)^2 = 0.081$$

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?

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

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

$$= 1 - (0.9)^3 = 0.271$$

The Poisson Distribution

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

The Poisson Distribution

probability function:

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

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

For $\lambda > 0$

The Poisson Distribution

These occurrences must:

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

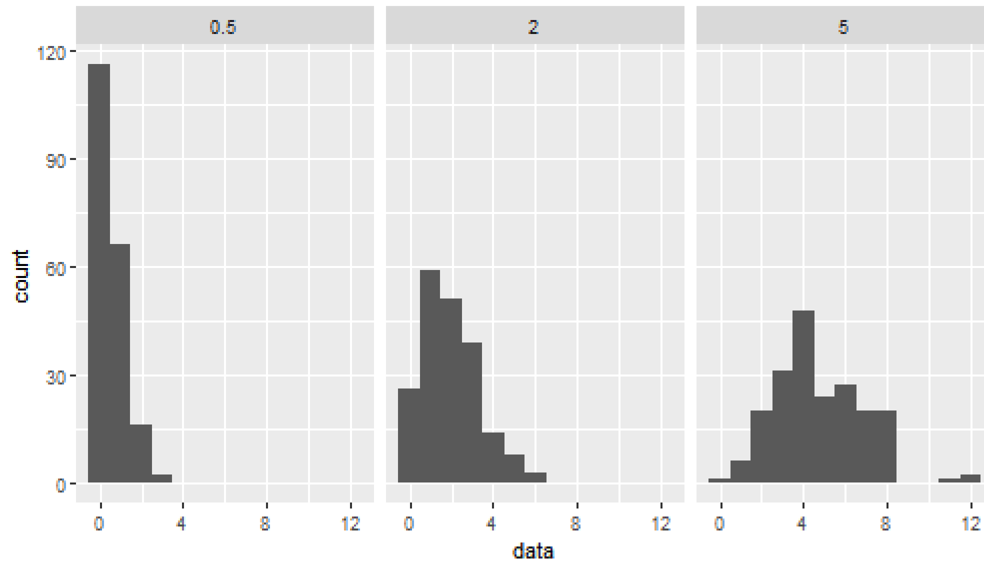
λ the *rate parameter*, is the expected number of occurances in **the specified interval of time or space** (i.e $\mathrm{E}(\mathrm{X}) = \lambda$)

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(λ) random variable,

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

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

Example

Arrivals at the library

Some students' data indicate that between 12:00 and 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?

Example

Arrivals at the library

Under this model, the probability that between 10 and 15 students arrive at the library between 12:00 and 12:01 PM is:

Shark attacks

Let X be the number of unprovoked shark attacks that will occur off the coast of Florida next year. Model

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

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

What would a reasonable choice of λ be?

Shark attacks

Under this model, calculate the following:

- $P(\text{no attacks next year})$

- $P(\text{at least 5 attacks})$

- $P(\text{more than 10 attacks})$