Examples of Binomial Distribution

- Number of hexamine pallets in a batch of $\boxed{n = 50}$ total pallets made from a palletizing machine that conform to some standard.
- Number of runs of the same chemical process with percent yield above 80 given that you run the process 1000 times.
- Number of winning lottery tickets when you buy 10 tickets of the same kind.

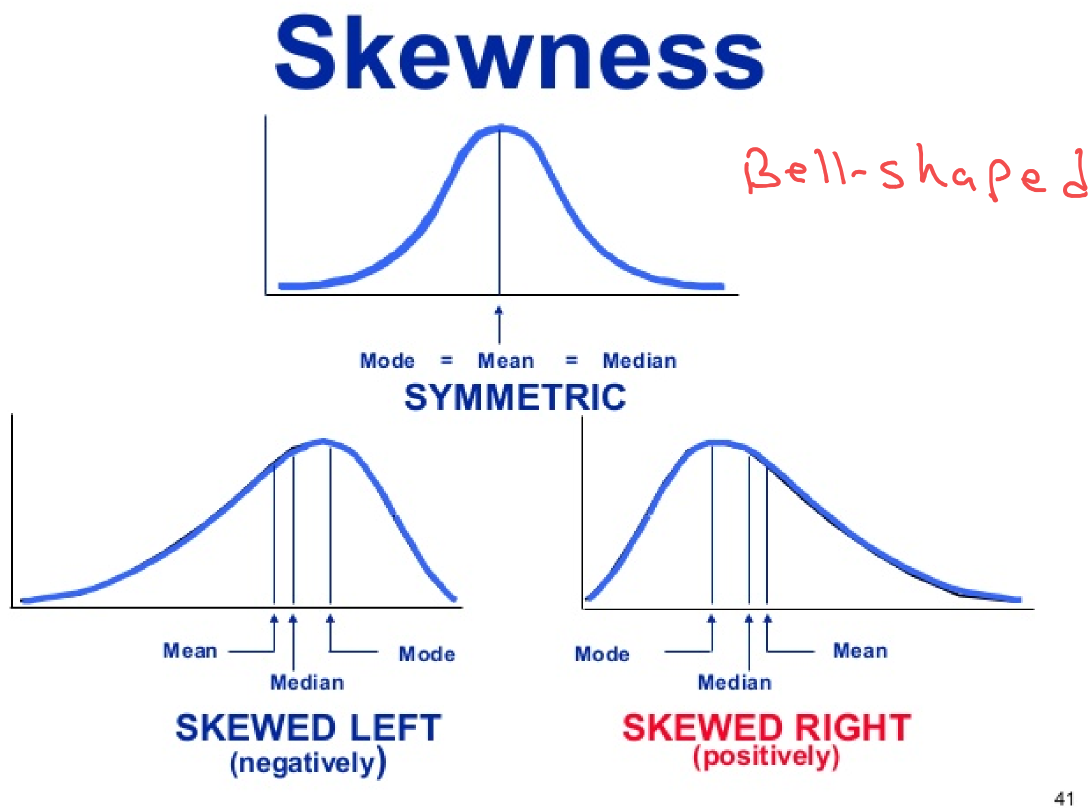
Recall: shape of distributions.

Common Distributions

Background

Bernoulli

Binomial



for different values of n, p, binomial distribution has different shapes.

Common Distributions

Background

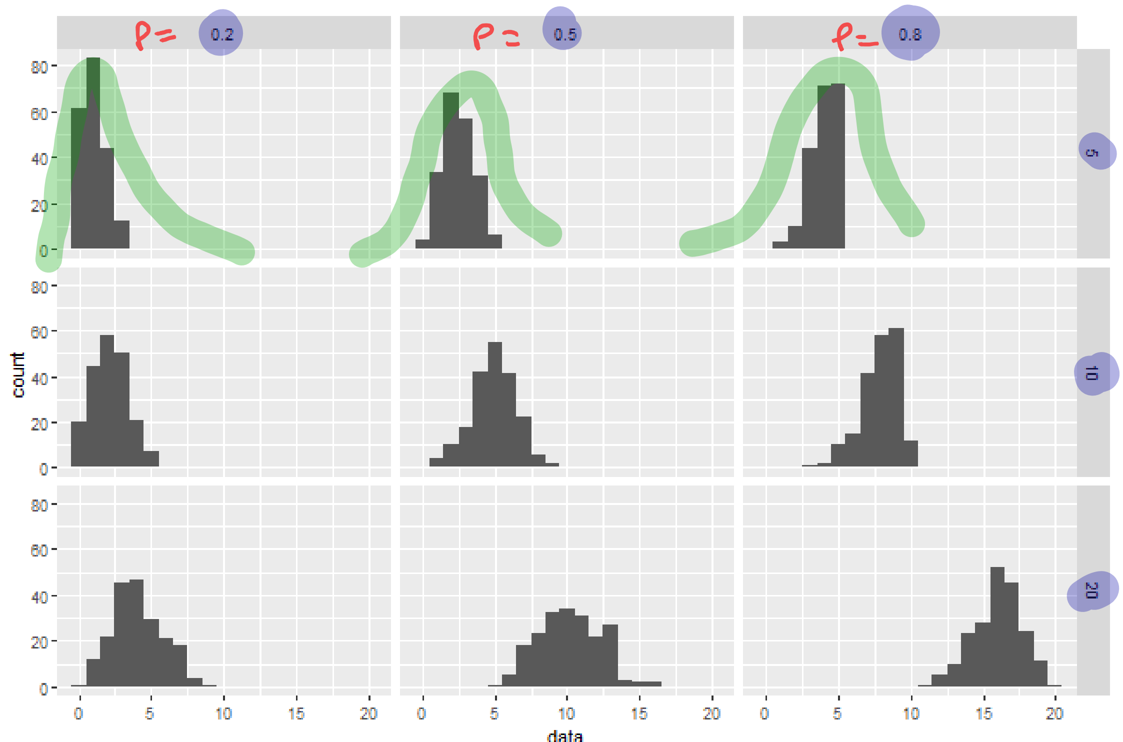
Bernoulli

Binomial

as n↑, skewness decreases

The Binomial Distribution

Plots of Binomial distribution based on different success probabilities and sample sizes.



p<0.5 => right-skewed

P= 0.5 Bell-shaped

P>0.5 left-skewed

The Binomial Distribution

Example [10 component machine]

Suppose you have a machine with 10 independent components in series. The machine only works if all the components work. Each component succeeds with probability $p = 0.95$ and fails with probability $1 - p = 0.05$.

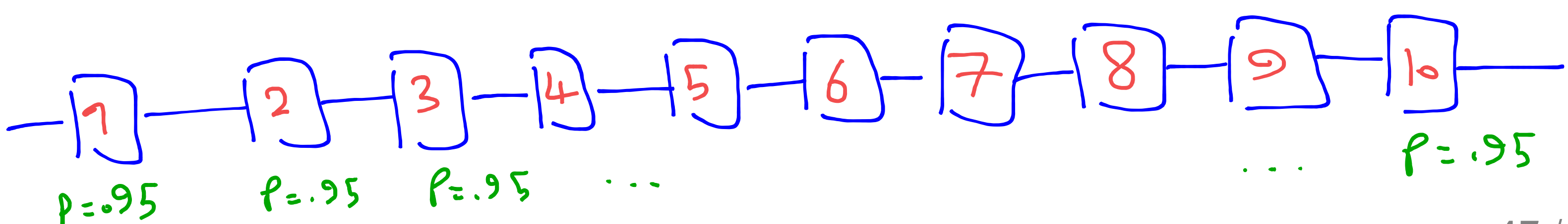
Let Y be the number of components that succeed in a given run of the machine. Then

Fully specified distribution

$$Y \sim \text{Binomial}(n = 10, p = 0.95)$$

Question: what is the probability of the machine working properly?

series component system:



Each part works with probability $p = .95$

a series system works properly if all components work.

So,

P("machine working") = P("all components work")

$$= P(Y = 10)$$

$$= f(10)$$

Recall: $f(y) = \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}$

$(n = 10,\ p = .95)$

$$= \frac{10!}{10!\underbrace{(10-10)!}_{0! = 1}} (0.95)^{10} (1-0.95)^{10-10}$$

$$= (0.95)^{10} = 0.5987$$

Not a very reliable system

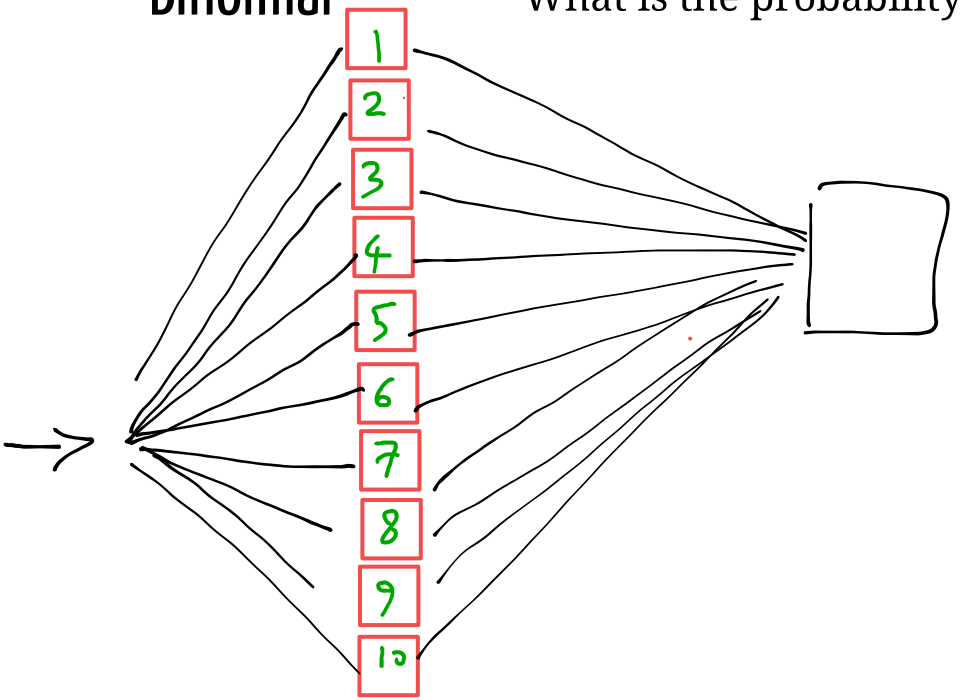
The Binomial Distribution

Example [10 component machine]

$$Y \sim \text{Binomial}(n = 10, p = 0.95)$$

What if I arrange these 10 components in parallel? This machine succeeds if at least 1 of the components succeeds.

What is the probability that the new machine succeeds?



P("machine works") = P("at least one component works")

$$= P(Y \geq 1)$$

(compliment) $$= 1 - P(Y < 1)$$

$$= 1 - P(Y = 0)$$

$$= 1 - p(0)$$

$$= 1 - \frac{10!}{0!(10-0)!}(0.95)^0(1-0.95)^{10-0}$$

$$= 1 - (0.05)^{10}$$

$$\approx 1$$ (a very reliable system)

Binomial Distribution

Expected Value and Variance

Common Distributions

Background

Bernoulli

Binomial

The Binomial Distribution

> **Expected value:**
>
> $$E(X) = n \cdot p$$

> **Variance:**
>
> $$Var(X) = n \cdot (1-p) \cdot p$$

Recall: Bernoulli distribution ≡ Bernoulli (P)

E(X) = P ⟶ Binomial is "n" independent Bernoulli trials

So, E(Binomial) = n . P

Var(x) = P(1-P) ⟶ Similarly, Var(Binomial) = n P(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.

$$Y \sim \text{Binomial}(n=10, p=0.95)$$

$$E(Y) = n \cdot p = 10(0.95) = 9.5$$

(we expect 9.5 components to succeed working in the machine)

$$Var(Y) = np(1-p) = 10(0.95)(1-0.95) = 0.475$$

Standard Deviation

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

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 $X \sim Binomial(n, p)$
- 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.

Note: There is No Close Form For CDF of Binomial.

The Geometric Distribution

another generic discrete r.v.

The Geometric Distribution

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:

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

There's only one parameter.

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

at least one trial to observe the first success.

Common Distributions

Background

Bernoulli

Binomial

Geometric

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.

Common Distributions

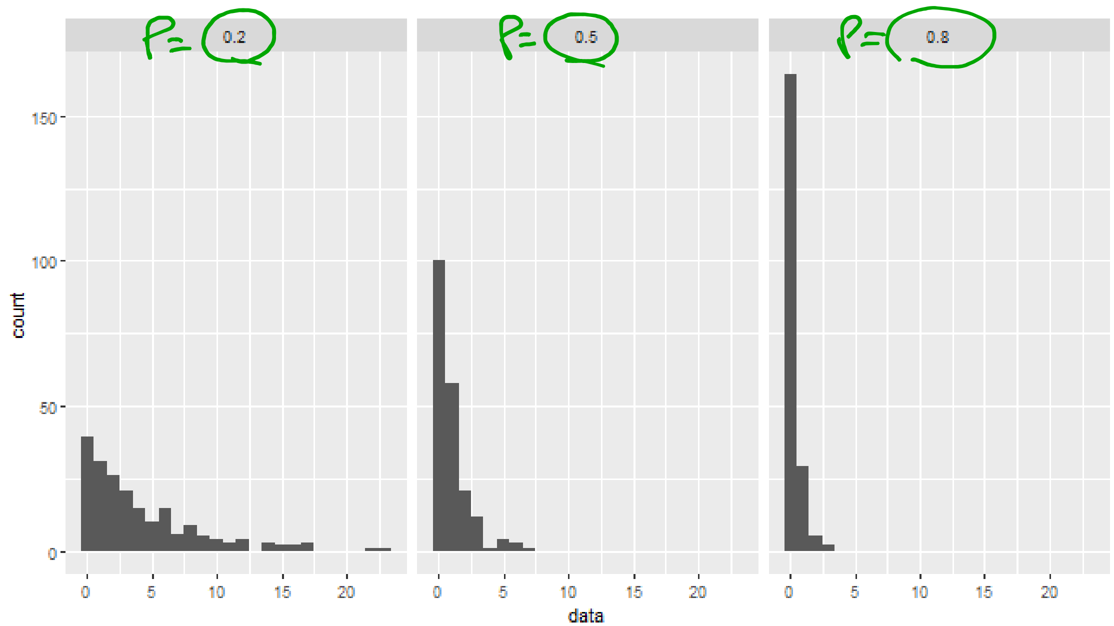
Background

Bernoulli

Binomial

Geometric

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

close form CDF

Cumulative probability function: $F(x) = 1 - (1-p)^x$

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:

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

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

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

CDF of Geometric distribution

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, $T \sim \text{Geom}(p)$.

$\to f(t) = P(T=t) = p(1-p)^{t-1}, \quad t = 1, 2, \cdots$

Calculate

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

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

$f(t) = (0.01)(0.99)^{t-1} = (0.01)(0.99)^{1-1} + (0.01)(0.99)^{2-1} = 0.0199$

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

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

$= 1 - [1 - (1-p)^{50}]$

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

(Geometric dist. has closed form CDF)

So, with 61% probability, the first 50 Nicad have no short.

Common Distributions

Background

Bernoulli

Binomial

Geometric

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.

$$T \sim Geom(p) \Rightarrow E(T) = 1/p$$

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

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

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

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

Common Distributions

Background

Bernoulli

Binomial

Geometric

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.

$p = 0.1$

(success here is finding defective widget)

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

$X \sim Geom(p = 0.1), \; x = 1, 2, 3, \ldots$

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

$$E(X) = \frac{1}{p} = \frac{1}{0.1} = 10$$

i.e we need to test on avg. 10 widgets to see the first defective one.

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?

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

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

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

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?

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

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

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

Recall: in Geometric distribution:

CDF : $F(x) = 1 - (1-p)^x$

pmf : $f(x) = p(1-p)^{x-1}$, $x = 1, 2, 3, \ldots$

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

This count/number of times the rare occurance is observed can be associated to a well-known pmf.

The Poisson Distribution

another generic pmf.

probability function:

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

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

$\boxed{\text{For } \lambda > 0}$

$X \sim \text{poisson}(\lambda)$

parameter of Poisson dist.

The Poisson Distribution

These occurrences must:

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

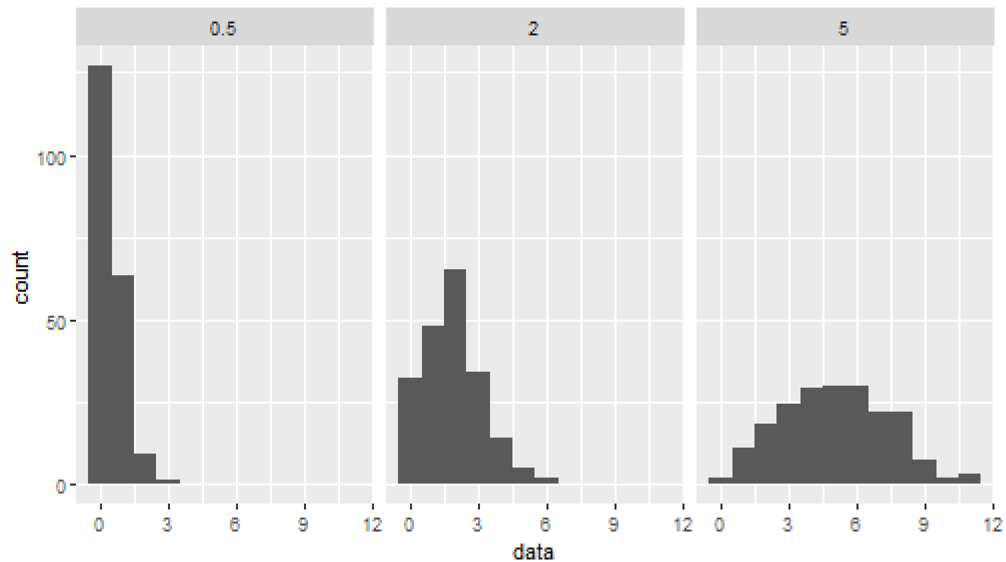
λ the *rate parameter*, is the expected number of occurances in **the specified interval of time or space** (i.e $\mathrm{E}(\mathrm{X}) = \lambda$)

The Poisson Distribution

Examples that could follow a Poisson(λ) distribution :

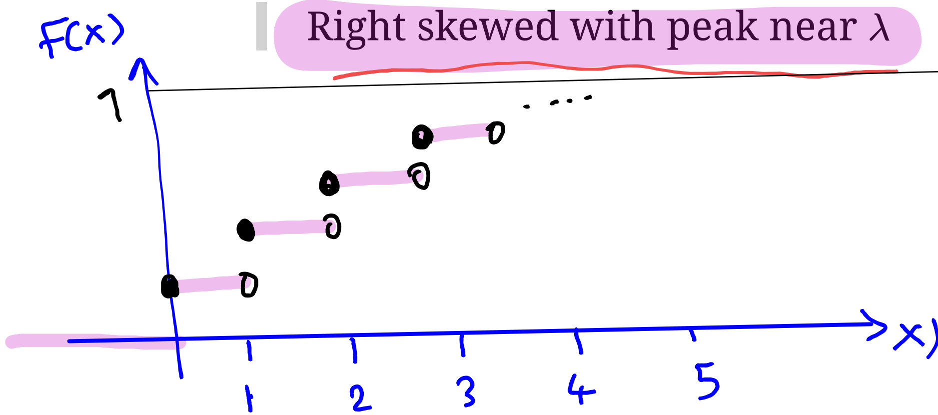
> Y is the number of shark attacks off the coast of CA next **year**, $\lambda = 100$ attacks per year
>
> Z is the number of shark attacks off the coast of CA next **month**, $\lambda = 100/12$ attacks per month
>
> N is the number of α-particles emitted from a small bar of polonium, registered by a 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

CDF =



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?

$$\begin{cases} 125 \text{ students in } 10' \\ \lambda \quad\quad\quad\;\; \text{ in } 1' \end{cases} \Rightarrow \lambda = \frac{125}{10}$$

Common Distributions

Background

Bernoulli

Binomial

Geometric

Poisson

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:

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

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

$$P(10 \le M \le 15) = f(10) + f(11) + f(12) + f(13) + f(14) + f(15)$$

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

$$= 0.6$$

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?

246 attacks in 10 years

λ attacks in next year.

(only a year)

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

Common Distributions

Background

Bernoulli

Binomial

Geometric

Poisson

Shark attacks

Under this model, calculate the following:

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

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

(so unlikely to have no attacks ☹)

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

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

$$= 1 - [f(0) + f(1) + f(2) + f(3) + f(4)]$$

$$= \ldots = 0.999249$$

(so probable to have at least 5 attacks)

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

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

wrap up:

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

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 $X=4)$)

- $EX = n \cdot p$

- $Var X = np(1-p)$

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

- Geometric distribution ; $X \sim Geom(p)$

$X :=$ 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 : $F_X(x) = 1-(1-p)^x$

- $EX = 1/p$

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

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

Poisson distribution:

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

$X :=$ The count of occurances over a specific period of time or space.

$X = 0, 1, 2, \ldots$

- independent occurances
- no two occurances at the same time
- occur at the same constant rate λ.

λ is the rate parameter & is the expected number of occurances in the specific interval of time or space