USC School of Business

Data Driven Decision Making: Exponential & Poisson Distributions

GSBA 545, Fall 2021

Professor Dawn Porter

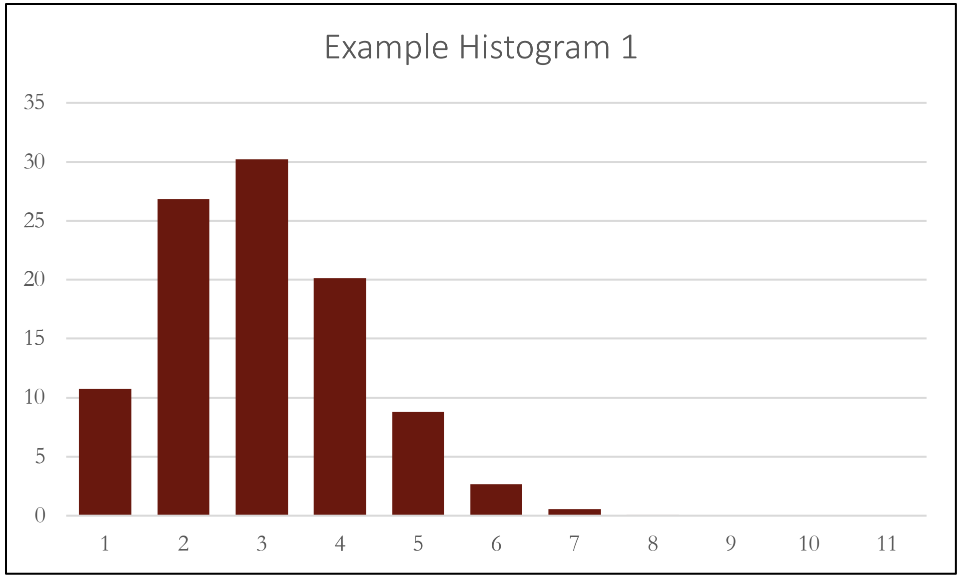
Exponential & Poisson Variables

- Exponential Distribution Basics
  - Probability calculations for Exponential RVs
- Poisson Distribution Basics
  - Probability calculations for Poisson RVs
- Appendix: Other Useful Distributions
  - Bernoulli
  - Binomial
  - Negative Binomial
  - Triangular
  - Gamma

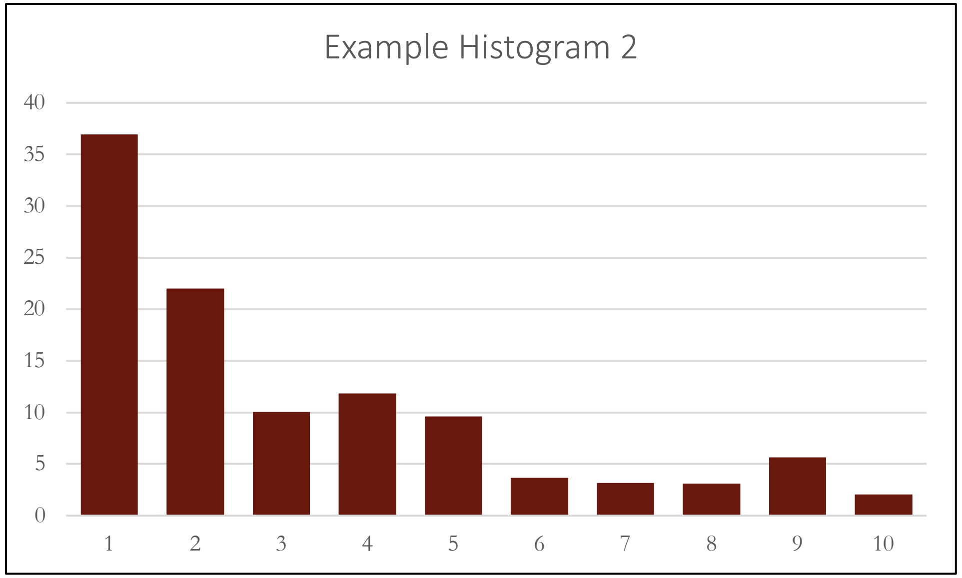
Common Distributions

Discrete	*Continuous*
Uniform	Uniform
Bernoulli Binomial Geometric Negative Binomial	Normal Triangular
Poisson	Exponential Gamma

Motivation



What are these distributions and how can we use them to calculate probabilities?



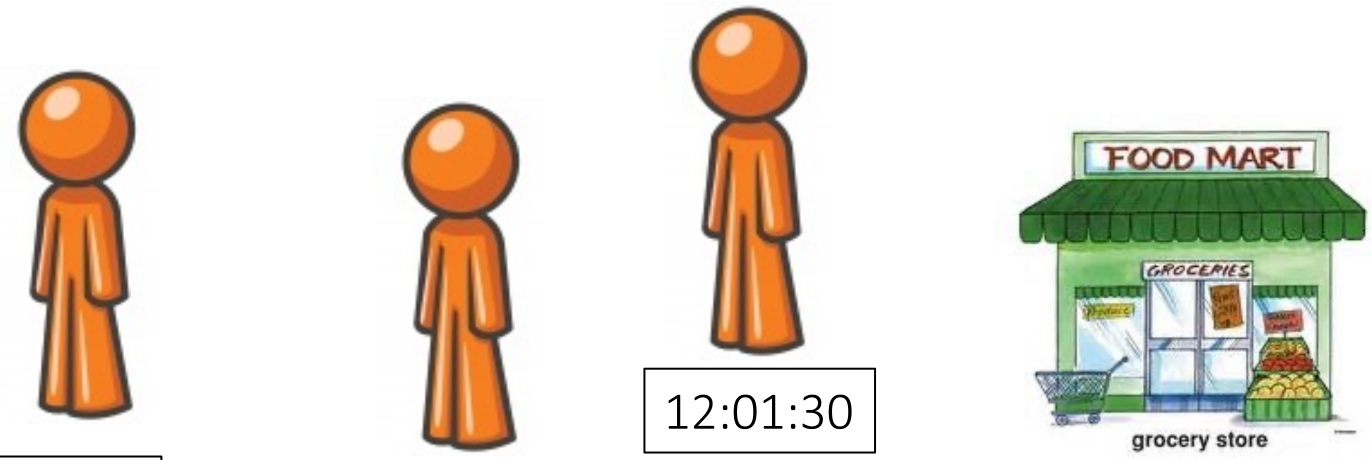
Example: Retail Store Demand Estimation

- You run a large grocery store, that also serves quick lunches, in a heavily trafficked location.
- Your busiest times are the weekdays, between noon and 2 pm.
- You would like to better understand the arrival pattern of your lunchtime customers during your busiest times.
- You have one day's worth of data that shows the arrival times of customers to your store during busy hours.

Graphically

What are the random variables of interest?

And so on …



12:10:00

12:03:30

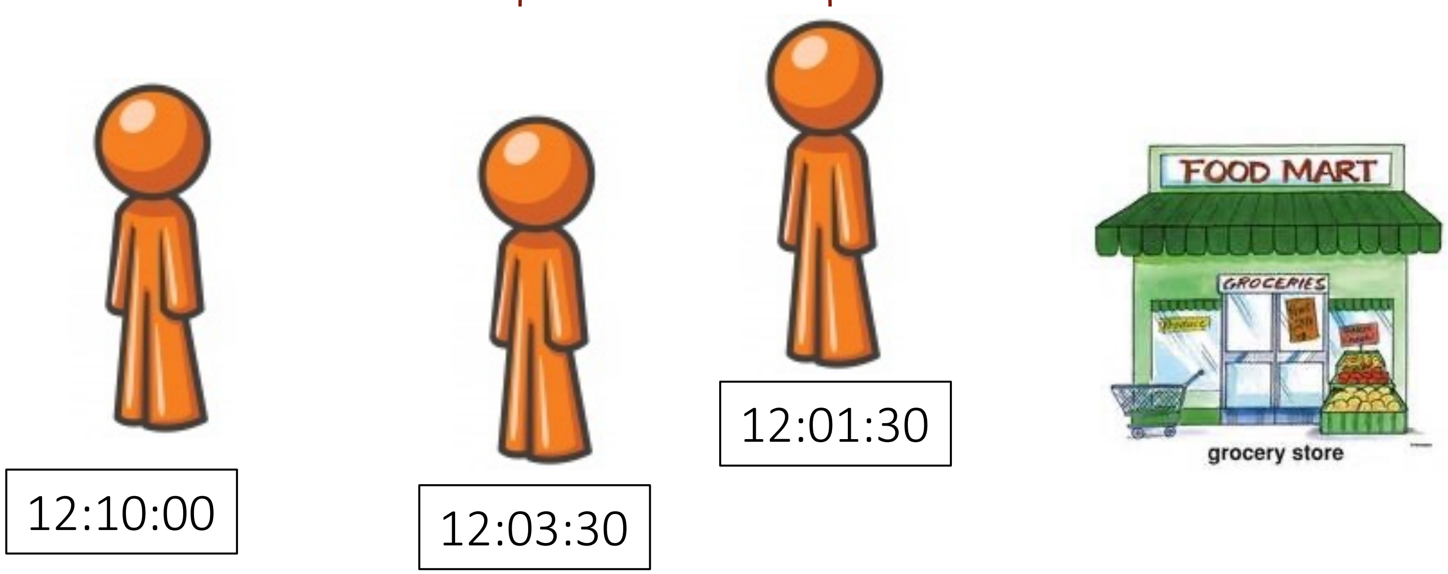
12:01:30

Graphically: Interarrival Time

What is the random variable of interest in this version?

Interarrival time = 2 min.

And so on …



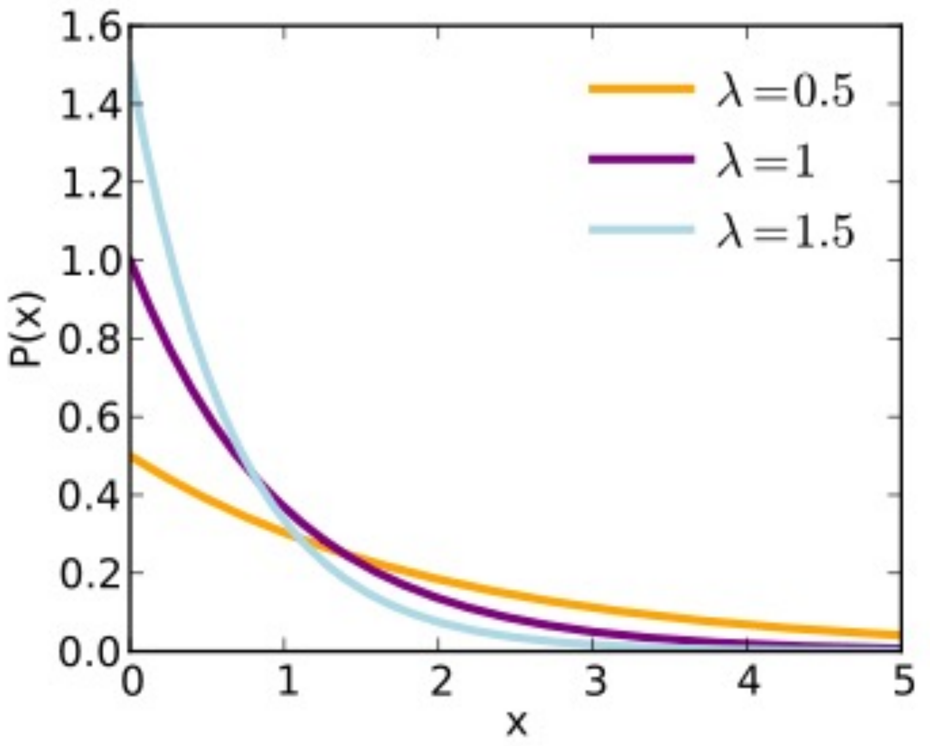
Interarrival time = 6 min., 30 sec.

Exponential Distribution

Your goal is to estimate the distribution of inter-arrival times.

The parameter λ determines the mean ($1/\lambda$) and variance ($1/\lambda^2$).

> The *Exponential Distribution* is often used to model inter-arrival times.



Exponential RV Summary Statistics

For an *Exp*(λ) random variable:

- Mean. $\mu = \frac{1}{\lambda}$
- Median. $M_d = \frac{1}{\lambda} \ln 2$
- Variance. $\sigma^2 = \frac{1}{\lambda^2}$

Example: Post Office Wait Times

Let X = amount of time (in mins) a postal clerk spends with a customer. The time is exponentially distributed with the average amount of 4 mins = $1/\lambda$.

a) Find the probability that a clerk spends between four and five minutes with a randomly selected customer.

b) Half of all customers are finished in how long?

---

a) **Soln:** We are looking for $P(4 \le X \le 5)$.

In Excel, we need a value for x, λ, and whether we want cumulative results.

For $P(X \le 5)$: EXPON.DIST(X,lambda,Cum.) = EXPON.DIST(5,0.25,1) = 0.7135

For $P(X \le 4)$: EXPON.DIST(X,lambda,Cum.) = EXPON.DIST(4,0.25,1) = 0.6321

So $P(4 \le X \le 5) = 0.7135 - 0.6321 = \boxed{0.0814}$

b) **Soln:** The **median** for an exponential RV is $M_d = \frac{1}{\lambda}\ln 2 = (4)\times\ln 2 = \boxed{2.77 \text{ mins}}$

Example: Airline Ticket Purchasing

USC School of Business

The number of days ahead travelers purchase their airline tickets can be modeled by an exponential distribution with the average amount of time equal to 15 days = $1/\lambda$.

a) Find the probability that a traveler will purchase a ticket fewer than 10 days in advance.

b) How many days do half of all travelers wait to purchase a ticket?

---

a) **Soln:** We are looking for P(X < 10).

In Excel, we need a value for x, λ, and whether we want cumulative results.

For P(X < 10): EXPON.DIST(X,lambda,Cum.) = EXPON.DIST(10,0.0667,1) = $\boxed{0.4868}$

b) **Soln:** The **median** for an exponential RV is $M_d = \frac{1}{\lambda}\ln 2$ = (15)×ln 2 = $\boxed{10.397 \text{ days}}$

Example: Computer Part Lifetime

USC School of Business

On average, a certain computer part lasts 10 years and so then follows an exponential distribution with $\lambda = 1/10 = 0.10$.

a) What is the probability that a computer part lasts more than 7 years?

b) On average, how long would 5 computer parts last if they are used one after another?

c) Fifty percent of computer parts last at most how long?

d) What is the probability that a computer part lasts between 9 and 11 yrs?

---

a) We are looking for $P(X > 7) = 1 - P(X \leq 7)$.

Soln: $P(X \leq 7)$ = EXPON.DIST(7,0.10,1) = 0.5034 → $P(X > 7) = 1 - 0.5034 = \boxed{0.4966}$

b) **Soln:** This is simply $5 \times \mu = \boxed{50 \text{ yrs}}$

c) **Soln:** The **median** for an exponential RV is $M_d = \frac{1}{\lambda} \ln 2 = (10) \times \ln 2 = \boxed{6.931 \text{ yrs}}$

d) **Soln:** We are looking for $P(9 \leq X \leq 11) = 0.6671 - 0.5934 = \boxed{0.0737}$

Exponential & Poisson Relationship

Suppose that instead of determining the inter-arrival time distribution, you preferred to determine the number of customers to arrive to your store every minute.

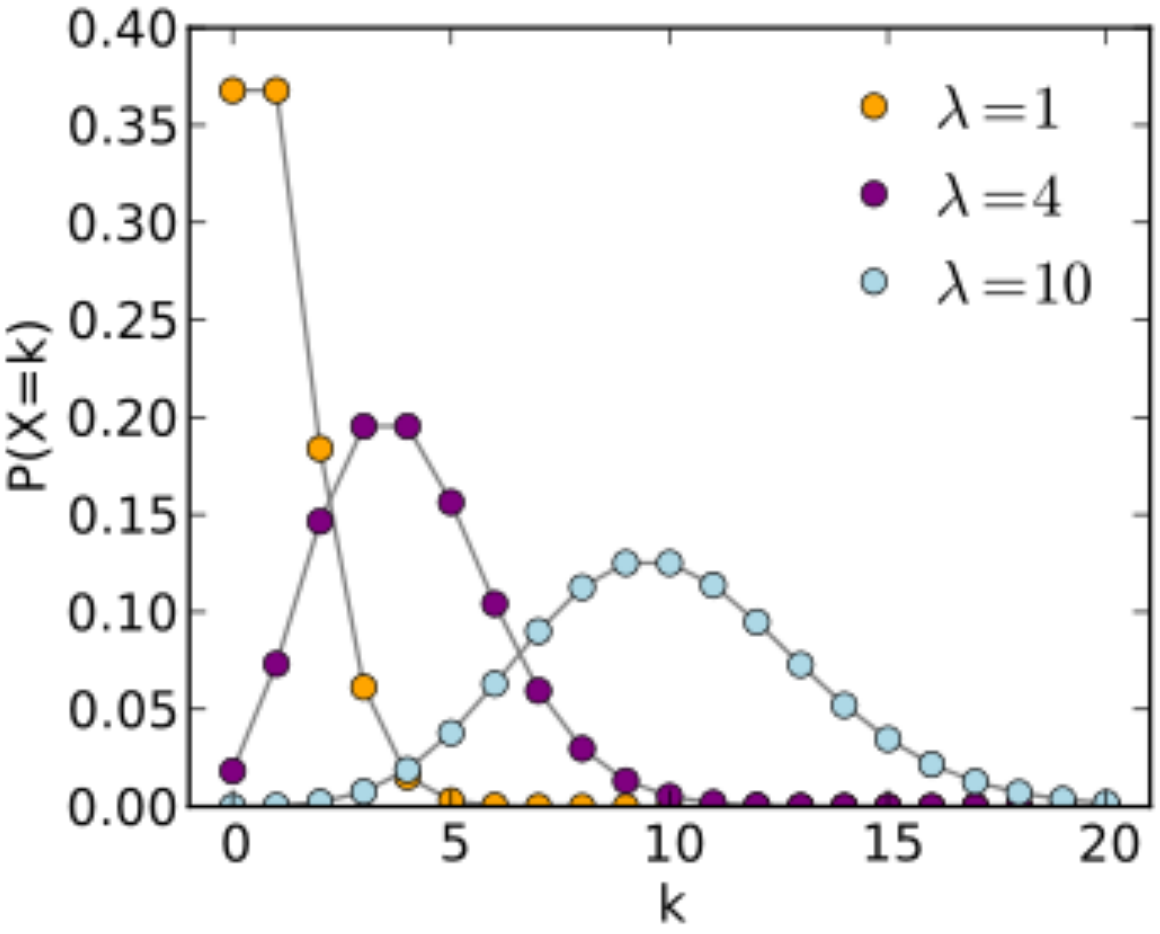
- What do you expect to be the mean of this distribution?
- How can you estimate the distribution from the dataset?

Poisson Distribution

Now you want to estimate the number of customers (k) arriving in a particular time window (one minute).

The parameter λ determines the mean (λ) and variance (λ).

The *Poisson Distribution* is often used to model the number of people arriving per unit of time.



Poisson RV Summary Statistics

For a *Poisson*(λ) random variable:

- Mean. $\mu = \lambda$
- Median. $\lambda - ln2 \leq M_d < \lambda + \frac{1}{3}$
- Variance. $\sigma^2 = \lambda$

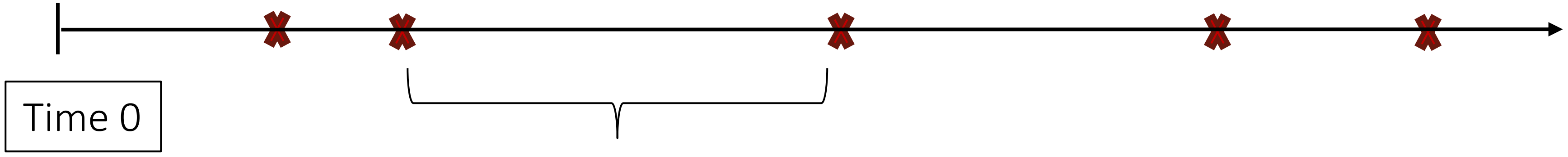
Exponential & Poisson Relationship

Suppose the inter-arrival time of "events" (for example, customers) follows an exponential distribution with mean $1/\lambda$.

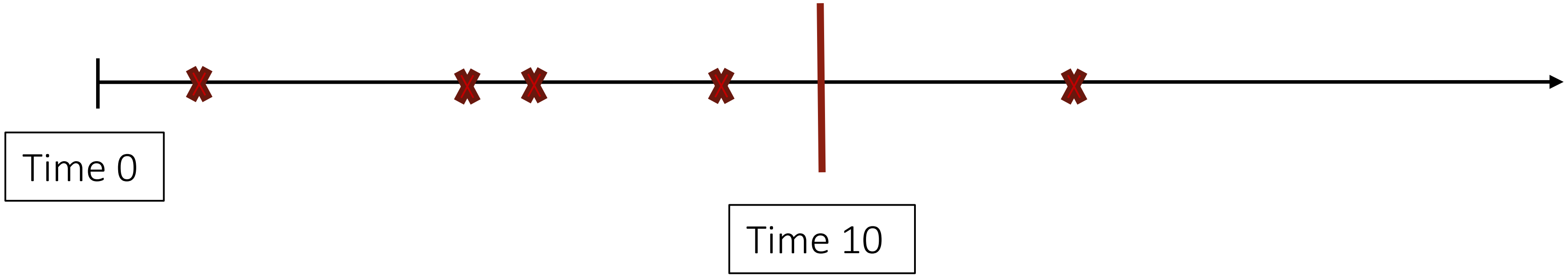
- We can think of λ as the rate at which events occur (per time unit).
- The average number of events in time interval $[0, t]$ is λt.
- The distribution of the number of events that occur in the time interval $[0, t]$ is Poisson, with mean λt.
- (In today's example, the time between customer arrivals was on average 30 seconds, and on average 1/30 customers arrive every second, or 2 customers arrive every minute.)

Graphical Depiction

Suppose you know the distribution of time between events is exponential with mean 3 minutes.



Suppose *X* is a random variable that represents the number of events that will occur in the next 10 minutes. What is the distribution of *X*?



Example: Bakery Loaves

The average number of loaves of bread put on a shelf in a bakery in a half-hour period is 12. Of interest is the number of loaves of bread put on the shelf in five minutes. The time interval of interest is 5 minutes.

a) What is the *average* number of loaves put on the shelf in 5 mins?

b) What is the probability the number of loaves put on the shelf in 5 mins equals three?

---

a) **Soln:** The *average* in 30 mins is 12, so we should expect the *average* in 5 mins is

$$\lambda = \left(\frac{5}{30}\right) \times (12) = \boxed{\text{2 loaves}}$$

b) **Soln:** We want P(*X* = 3). In Excel, we need a value for *x,* the mean (λ), and whether we want cumulative results.

For P(*X* = 3): POISSON.DIST(*X*,mean,Cum.) = POISSON.DIST(3,2,0) = $\boxed{0.1804}$

Example: Answering Machine Calls

Leah's answering machine receives about 6 telephone calls between 8 a.m. and 10 a.m. What is the probability that Leah receives more than one call **in the next 15 minutes?**

a) What is the *average* number of calls received in a 15-minute window?

b) What is the probability that Leah receives more than one call **in the next 15 minutes?**

---

a) **Soln:** The *average* in 120 mins is 6, so we should expect the *average* in 15 mins is

$$\lambda = \left(\frac{15}{120}\right) \times (6) = \boxed{0.75 \text{ calls}}$$

b) **Soln:** We want P(X > 1). In Excel, we need a value for *x,* the mean (λ), and whether we want cumulative results.

For P(X > 1): 1 – POISSON.DIST(1,0.75,1) = 1 – 0.8266 = $\boxed{0.1734}$

Example: Restaurant Arrivals

Students arrive at a local bar and restaurant according to an approximate Poisson process at a mean rate of 30 students per hour.

What is the probability the bouncer waits more than 3 mins to card the next student?

Soln: This is asking about the *time* it takes, so this should follow an **Exponential Distribution.**

- If the *time window* is 1 hr, λ = 30 students. We need a *time window* of 3 mins, though, so we need to convert our estimate to students per *minute:* 30/60 = ½ student per min = λ.

$$P(t \geq 3) = 1 - P(t < 3).$$

In Excel, this is P(*t* ≥ 3) = 1 – EXPON.DIST(3,0.50,1) = 1 – 0.7769 → P(*t* ≥ 3) = $\boxed{0.2231}$

- Using a *Poisson* perspective: Waiting more than 3 mins for the next student means 0 students arrive in the 3-min time window. For *that* timeframe, λ = 1.5.

In Excel: P(*X* = 0) = POISSON.DIST(0,1.50,0) = $\boxed{0.2231.}$

Appendix: Other Distributions

- Bernoulli Distribution
- Binomial Distribution
- Geometric Distribution
- Negative Binomial Distribution
- Triangular Distribution
- Gamma Distribution

Bernoulli Distribution

Flip a (possibly biased) coin. The random variable X is 1 if the coin lands heads and 0 if the coin lands tails.

- The parameter is: p = the probability of a head.
- The probability mass function is: $P(X = 1) = p, P(X = 0) = 1 - p$.
- What are examples of situations in which Bernoulli random variables arise?
- What are the mean and variance?

$$\boxed{E[X] = p} \qquad \boxed{Var(X) = p(1-p)}$$

- How do you produce a sample from a Bernoulli in Excel?

IF(RAND()<p,1,0)

Binomial Distribution

A coin lands heads with probability *p* and tails with probability $q = 1 - p$. If you flip:

- 2 coins, and *X* is the number of heads, how do you calculate probabilities?
- 3 coins, and *X* is the number of heads, how do you calculate probabilities?

The binomial: Suppose you flip *n* coins, and let *X* count the number of heads.

- What are examples of situations that follow the binomial distribution?

$$p(i) = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, i = 0,1,2,\ldots,n \text{ where } \binom{n}{i} = \frac{n!}{i!\times(n-i)!}$$

What are the mean and variance? $\boxed{E[X] = np, Var(X) = np(1-p)}$

Geometric Distribution

A coin lands heads with probability *p*, tails with probability $q = 1 - p$.

- Let *X* be the number of flips until you see the first head.
- What are examples of situations in which the geometric distribution arises?
- What is the pmf?

$$p(i) = P(X = i) = p(1-p)^{i-1}, i \geq 1$$

What are the mean and variance?

$$\boxed{E[X] = \frac{1}{p}, Var\ (X) = \frac{1-p}{p^2}}$$

Negative Binomial Distribution

A coin lands heads with probability *p,* tails with probability *q* = 1 – *p*.

- Let *X* be the number of flips until you see the k^{th} head.
- What are examples of situations in which the negative binomial distribution arises?
- What is the pmf?

$$p(i) = P(X = i) = \binom{i-1}{k-1} p(1-p)^{i-1}, i \geq k$$

What are the mean and variance?

$$\boxed{E[X] = \frac{k}{p}, Var\ (X) = \frac{k(1-p)}{p^2}}$$

Triangular Distribution

The triangular distribution arises when there are natural minimum and maximum values, and you believe some values occur more frequently than other values.

- The distribution is continuous, because any value between the minimum and maximum can occur (in fact, an infinite number of values can occur).
- This is written as:

Minimum value

Maximum value

Triangular(a,b,c).

Most likely value (mode)

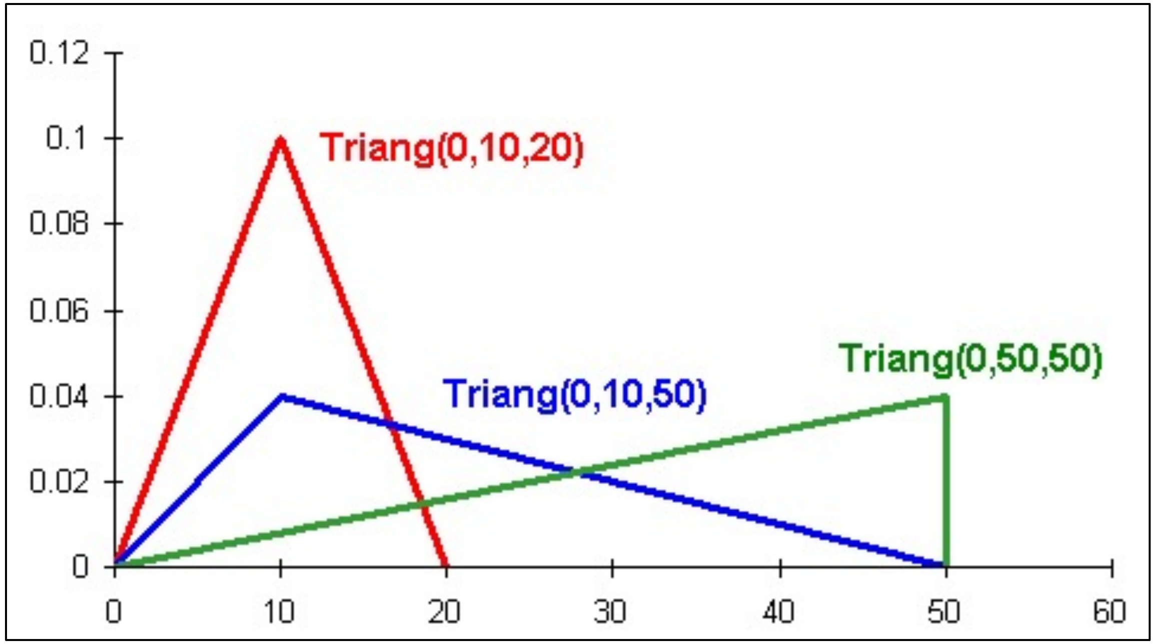
Triangular Distribution

Triangular(a,b,c):

Mean = $\frac{a+b+c}{3}$

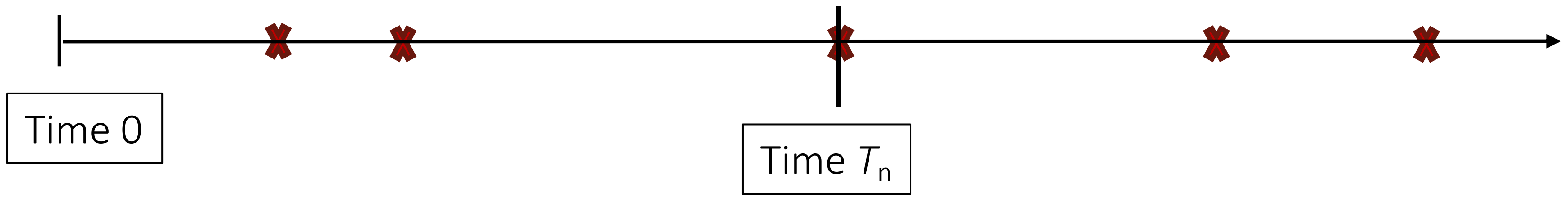
Std Dev = $\sqrt{\frac{a^2+b^2+c^2-ab-ac-bc}{18}}$

Median = $\begin{cases} a+\sqrt{\frac{(b-a)(c-a)}{2}} & \boxed{\text{if } c \geq \frac{a+b}{2}} \\ b-\sqrt{\frac{(b-a)(b-c)}{2}} & \boxed{\text{if } c \leq \frac{a+b}{2}} \end{cases}$



Gamma Distribution

Suppose you would like to know the time T_n until the occurrence of the n^{th} event in a Poisson process with rate λ.



The time T_n has a gamma distribution, with parameters (n, λ), and pdf

$$f(t) = \lambda e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}, t > 0.$$

What are the mean and variance?

$$\boxed{E[t] = \frac{n}{\lambda}, \qquad Var(t) = \frac{n}{\lambda^2}}$$