1: Definitions

Law of Large Numbers

For a small number of trials, *anything* can happen.
As number of trials increases, the ***experimental probability*** approaches the ***theoretical probability.***

Definitions

Trial: One complete 'occurrence' of a situation .
Outcome: One possible result that can occur when a trial is conducted.
Sample space: Set of all possible outcomes that can occur in a trial.
Event: Any subset of the sample space (the "desired" outcomes).
Union: $A \cup B$ (A ***OR*** B)
Intersection: $A \cap B$ (A ***AND*** B)
Parameter: Number describing a *population (or model)*, e.g. μ, σ
Statistic: Number describing a *sample*, e.g. $\bar{x}, s$

2: Equally-likely outcomes

If all outcomes are equally likely:

$$P(E) = \frac{\text{number of outcomes in event } E}{\text{number of outcomes in the sample space } S}$$

$$= \frac{\text{number of 'desired' outcomes}}{\text{total number of outcomes}}$$

2: Counting Strategies

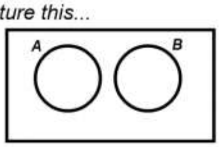
Strategy	When to Use
1. List out all the cases and just count them up. (can also use tree diagrams, grids for 2 dice, methodical listing system to help)	1. Best strategy, but only good for small numbers.
2. Multiplication Principle (one box per choice, fill in with number of ways to make that choice, multiply)	2. Multiple choice to make, every possible choice in each box can be paired with every other choice.
3. Permutations (use calculator) (special case: 'choose all' = n! ways)	3. A set of distinct objects (no repeats), choosing some or all, and objects are 'used up' as you choose them, and order matters.
4. Combinations (use calculator)	4. A set of distinct objects (no repeats), choosing some or all, and objects are 'used up' as you choose them, and order does not matter.
5. Multiplication Principle w/Combinations	5. Multiple choices to make, but each is a choice of a number of items out of a set.
6. Distinguishable permutations $\# \text{ distinguishable permutations} = \frac{n!}{n_1!n_2!n_3!\ldots}$	6. Number of ways to arrange all items in a set if there are repeats.

4: Compound Events (OR)

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Special Case: If two events have no overlap, they are called 'mutually exclusive' or 'disjoint' events.

Picture this...



So the OR formula is simplified...

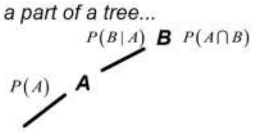
$$P(A \cup B) = P(A) + P(B) - \cancel{P(A \cap B)}$$

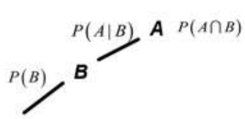
$$\left(P(A \cap B) = 0\right)$$

5: Compound Events (AND)

$$P(A \cap B) = P(A) \bullet P(B \mid A) \qquad P(A \cap B) = P(B) \bullet P(A \mid B)$$

Picture a part of a tree...





Special Case: If the probability of an event does not change regardless of whether or not another event happens, then the events are <u>independent events.</u>

For independent events: $P(B) = P(B \mid A)$

So... $P(A \cap B) = P(A) \bullet P(B \mid A)$ ...simplifies to... $P(A \cap B) = P(A) \bullet P(B)$

3: Conditional Probability

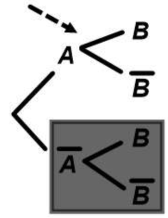
$$P(\text{video games} \mid \text{girl}) = \frac{12}{60} = .20$$

event — The event is always contained within the conditional sample space.

condition — The condition is always just a portion of the sample space (the *conditional sample space*).

The **conditional sample space** is a portion of the **sample space**.
The **event** is a portion of the **conditional sample space**.

	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100



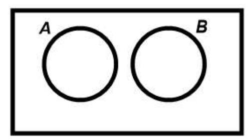
3: Conditional Probability

The event goes in the numerator of the fraction.

$$P(\text{video games} \mid \text{girl}) = \frac{12}{60} = .20$$

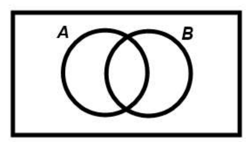
The condition goes in the denominator of the fraction.

4: Disjoint Events



A and B are mutually-exclusive
A and B are disjoint events

$$P(A \cap B) = 0$$



A and B are non mutually-exclusive
A and B are not disjoint events
A and B are joint events

$$P(A \cap B) \neq 0$$

5: Independent Events

Test for independent events:

Two events are independent if:

$$P(B) = P(B \mid A) = P(B \mid \overline{A})$$

(check any two)

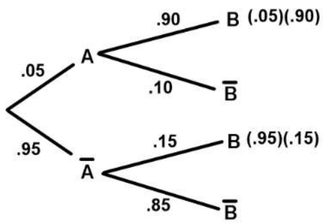
Note: Some books also use the simplified version of the AND formula as a 'test for independence'...

If $P(A \cap B) = P(A) \bullet P(B)$
then A and B are independent

...but this is more a consequence of independence, not the reason.

6: AND/OR together

We often need to use the AND and OR rules together:



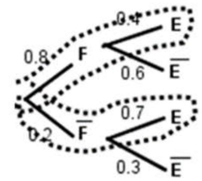
$$B = (A \text{ and } B) \text{ or } (\overline{A} \text{ and } B)$$
$$P(B) = P(A \text{ and } B) \text{ or } P(\overline{A} \text{ and } B)$$
$$P(B) = P(A \text{ and } B) + P(\overline{A} \text{ and } B)$$
$$P(B) = (P(A) \cdot P(B \mid A)) + (P(\overline{A}) \cdot P(B \mid \overline{A}))$$
$$P(B) = (.05)(.90) + (.95)(.15)$$
$$P(B) = .1875$$

6: Bayes' Formula

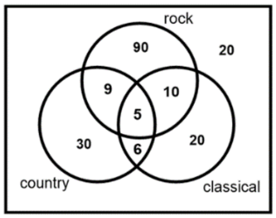
$$P(A \mid E) = \frac{P(A) \cdot P(E \mid A)}{P(E)}$$

But use probability of paths on a tree diagram:

$$P(F \mid E) = \frac{(0.8)(0.4)}{(0.8)(0.4) + (0.2)(0.7)}$$



7: Venn Diagrams



Venn diagrams are great for word problems with lots of information.

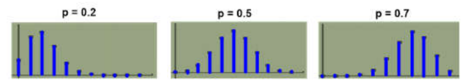
Always start with most overlapped region, and don't forget to subtract what has already been accounted for.

You can fill with either counts or probabilities (but be consistent).

8: Discrete Probability Models

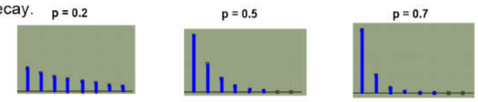
Binomial Shape depends upon p.

$\mu = np \qquad \sigma = \sqrt{npq}$



Geometric Shape is always exponential decay.

$\mu = \frac{1}{p} \qquad \sigma = \frac{\sqrt{1-p}}{p}$



General Discrete Models

$\mu = \text{'expected value'} = \sum X \cdot P(X)$

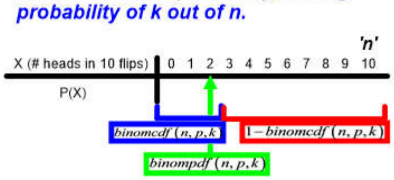
$\sigma \text{ (and } \mu) \text{ found using } L1(data), L2(freqList), 1\text{-}Var\ Stats$

8: Discrete Probability Models

Binomial

- Only 2 outcomes
- Probabilities must be the same each trial.
- Probabilities of trials must be independent.
- Must have fixed number of trials, n

***Best for**: independent trials, fixed number of trials (known n), finding probability of k out of n.*

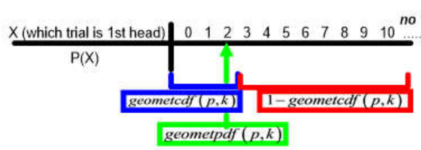


$$P(\text{exactly } k \text{ successes out of } n \text{ trials}) = {}_nC_k (p)^k (q)^{n-k}$$

Geometric

- Only 2 outcomes
- Probabilities must be the same each trial.
- Probabilities of trials must be independent.
- May or may not have fixed number of trials, n

***Best for**: independent trials, non-fixed number of trials (unknown n), finding probability of 'when' the 1st success occurs.*

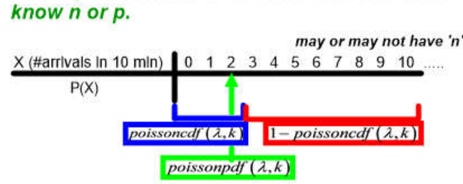


$$P(\text{success on the } k^{th} \text{ trial}) = (q)^{k-1}(p)$$

Poisson

- Only 2 outcomes
- Probabilities of trials **do not** need to be independent.
- Need a 'typical/expected' value (λ) for number of successes but don't need to know p or n explicitly.

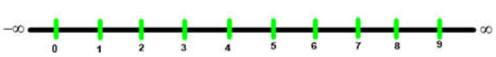
***Best for**: trials not independent or when we can find expected number of successes but don't know n or p.*



$$P(\text{exactly } k \text{ successes when } \lambda \text{ are expected}) = \frac{\lambda^k e^{-\lambda}}{k!}$$

9: Discrete vs. Continuous

Discrete

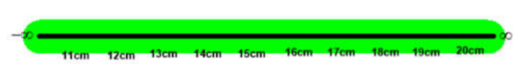


Discrete variables take on only specific values

Discrete situations often involve 'counting' the number of items in specific categories so discrete variables are sometimes referred to as 'categorical' or 'qualitative'

We can use Binomial or Geometric models to analyze probability, and the discrete expected value formula to find the mean of a discrete distribution

Continuous



Continuous variables can take on any value (sometimes within specified limits)

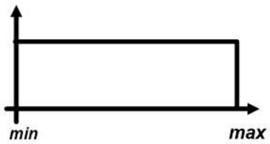
Continuous situations always involve a variable that is <u>numerical</u> (and usually includes units). Continuous variables are sometimes referred to as 'numerical' or 'quantitative'

We use an integral to find the area under the curve between boundaries to find probability.

9: Continuous Probability Models

Uniform

$f(x) = constant$



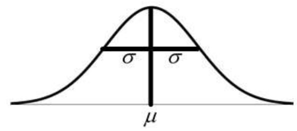
Exponential

$f(x) = \lambda e^{-\lambda x}$

$\lambda =$ 'rate parameter'
(would be given as part of the problem)

Normal

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



$\mu = mean$
$\sigma = standard\ deviation$

$$P(c < x < d) = \int_c^d f(x)\,dx$$

where $f(x)$ is a probability density function

calculator functions:

$P(LB < x < UB) = normalcdf(LB, UB, \mu, \sigma)$

$UB = invNorm(leftarea, \mu, \sigma)$

9: Expected Value vs. Average Value

(discrete) $Expected\ Value = \sum X \cdot P(X)$

(continuous) $Expected\ Value = \int_{lowest\ x}^{highest\ x} x \cdot f(x)\,dx$

(discrete) $Average\ Value = \dfrac{7+18+27+31+120}{5}$

(continuous) $Average\ Value = \dfrac{\int_a^b f(x)\,dx}{b-a}$

Expected Value

$$Expected\ Value = \int_a^b x \cdot f(x)\,dx$$

The 'value' we are finding the typical value of is the ***x***

f(x) gives the probability of the ***x*** values

Average Value

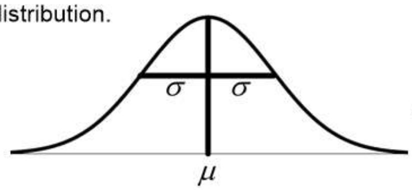
$$Average\ Value = \frac{\int_a^b f(x)\,dx}{b-a}$$

f(x) gives the values

(There is no PDF, so all the f(x) values must be equally-likely)

10: The Normal Model

Many values which have a continuous, infinite number of possible outcomes, especially quantities found in natural systems, can be modeled with a Normal distribution.

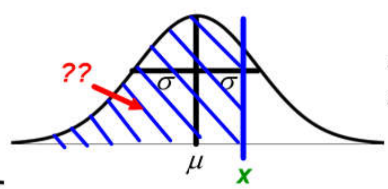


Normal distributions are symmetrical, centered at a mean μ

The average distance data is from this mean (on both sides) is called the standard deviation σ

2 calculator functions for use with a Normal distribution:

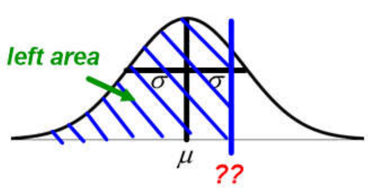
Have boundaries → Need area



$area = normalcdf(left\ boundary, right\ boundary, \mu, \sigma)$

$area = normalcdf(-999, x, \mu, \sigma)$

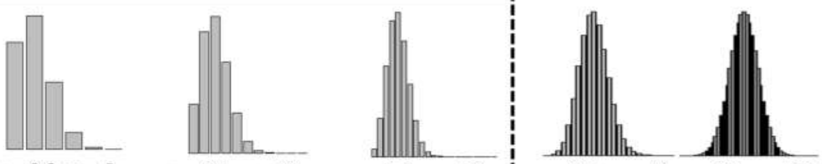
Have area → Need boundary



$upper\ boundary = invNorm(left\ area, \mu, \sigma)$

$x = invNorm(left\ area, \mu, \sigma)$

10: Normal Approximation of Binomial Model



$p = 0.2, n = 5$ $(np = 1)$ | $p = 0.2, n = 10$ $(np = 2)$ | $p = 0.2, n = 20$ $(np = 4)$ | $p = 0.2, n = 50$ $(np = 10)$ | $p = 0.2, n = 100$ $(np = 20)$

Can use Normal approximation for the Binomial distribution.

If $np \geq 10$ *and* $nq \geq 10$

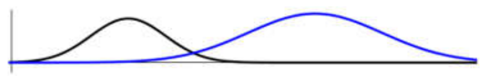
a Binomial distribution can be approximated with a Normal distribution with:

$$\mu = np$$

$$\sigma = \sqrt{npq}$$

11: Transforming a Single Distribution

Multiplying/dividing affects both center and spread...



Adding/Subtracting affects only center...

$$If\ Y = aX \pm b \qquad \mu_Y = a\mu_X \pm b \qquad \sigma_Y = a\sigma_X$$

11: Combining Multiple Distributions

Define an algebraic expression for how the source distributions are used to build the new distribution:

$$E = A + B - C - D$$

The means are always determined by <u>the defining algebraic expression</u>:

$$\mu_E = \mu_A + \mu_B - \mu_C - \mu_D$$

But because each source of variability increases <u>overall variation, the variances always add</u>:

$$\sigma_E^{\ 2} = \sigma_A^{\ 2} + \sigma_B^{\ 2} + \sigma_C^{\ 2} + \sigma_D^{\ 2}$$

However, we must know for certain that the variables are all varying independently of one another. (If not independent, we can find mean but not standard deviation).