Handy Transformations -

$$\mathbb{E}[a+bX] = a + b\mathbb{E}[X]$$
$$\text{Var}(a+bX) = b^2\text{Var}(X)$$
$$\text{SD}(a+bX) = |b|\text{SD}(X)$$
$$\text{Cov}(a+bX, c+dY) = b\cdot d\cdot \text{Cov}(X,Y)$$

Set Algebra:

- **Union -** A or B; $A\cup B$.
- **Intersection -** A and B; $A\cap B$.
- **Complement -** not A; A^C.
- **Difference -** A but not B; $A\backslash B$
- **Disjoint Events aka. mutually exclusive -** events A and B are disjoint if they don't share any outcomes in common (i.e., A and $B=\varnothing$).
- **Subset -** $A\subseteq B$

Trial - a repetition of a random experiment/process. Trials're independent: none gives information about the others; are stable: reuslts could have appeared in any order.
Outcome - a possible result of a trial.
Sample space - the set of all possible outcomes. Often denoted as S.
Event - a set of outcomes of an experiment (i.e., a subset of the sample space).
Probability - is a long run proportion of an outcome in repeated trials.

- Probabilities act as "targets" of estimation
- Proportions based on data "estimate" probabilities. Would approach probabilities if observe infinite trials.

Formally, $\boxed{A\mapsto \mathbb{P}(A), \mathbb{P}(A)\in[0,1]}$ A probability $\mathbb{P}(\cdot)$ on a sample space S is a function that assigns a snumber between 0 and 1 to all events, A in the sample space (i.e., any possible subset of the sample space) and subject to three requirements (axioms):

1. $\mathbb{P}(S)=1$: probability of *something* in the sample space happening is 1
2. $\mathbb{P}(A)\geq 0, \forall A$
3. A and $B=\varnothing$ (A, B disjoint) $\Rightarrow \mathbb{P}(A\cup B) = \mathbb{P}(A)+\mathbb{P}(B)$

More takeaways

- $A\subseteq B \Rightarrow \mathbb{P}(A)\leq \mathbb{P}(B)$
- A, B, C are pairwise disjoint $\Rightarrow \mathbb{P}(A\cup B\cup C) = \mathbb{P}(A)+\mathbb{P}(B)+\mathbb{P}(C)$
- $\mathbb{P}(A\cup B) = \mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A \text{ and } B)$

Joint Probability - $\mathbb{P}(A \text{ and } B)$ is the joint probability that events A and B occur.
Conditional Probability -
$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \text{ and } B)}{\mathbb{P}(B)}, \mathbb{P}(B)>0$ the probability of observing event A if (given that) one has observed B. Bear in mind: $\mathbb{P}(A|B)\neq \mathbb{P}(B|A)$
Product Rule - $\mathbb{P}(A \text{ and } B) = \mathbb{P}(B)\cdot\mathbb{P}(A|B)$.
Independent Events - A and B are independent if $\mathbb{P}(A \text{ and } B) = \mathbb{P}(A)\cdot\mathbb{P}(B)$; one event happening doesn't affect the probability of the other event happening. Can easily deduce that $\mathbb{P}(A|B)=\mathbb{P}(A)$ and $\mathbb{P}(B|A)=\mathbb{P}(B)$.
Independence and Disjointness are **NOT** synonyms.

- Independent $\Rightarrow \mathbb{P}(A \text{ and } B) = \mathbb{P}(A)\cdot\mathbb{P}(B)$
- Disjoint $\Rightarrow \mathbb{P}(A \text{ and } B) = 0$. Disjoint events are extremely dependent: If one event occurs, the other cannot.

Random variable - a numerical function on a sample space with probabilities. (Think as a scoring mechanism.)

- Input: an outcome in the sample space
- Output: a number

Discrete RVs - only countably many values are possible
Continous RVs - can take on uncountably infinitely many values
Probability Distribution Function (PDF) -
$\boxed{\mathbf{p}_X(x) = \mathbb{P}(X=x)}$ is the probability that the random variable X takes on the value x. I really hate $\mathbf{p}_X(x)$ this styling, so only $\mathbb{P}(X=x)$ moving forward.
Properties of PDFs - any function that satisfies the following conditions is a probability distribution function of a Discrete random variable:

1. $\mathbb{P}(X=x)\geq 0, \forall x\in\mathbb{R}$ (for any real number)
2. $\mathbb{P}(X=x)>0$ for values that the random variable X can actually take on
3. $\mathbb{P}(X=x)=0$ for values that aren't possible for the random variable X
4. $\sum_x \mathbb{P}(X=x)=1$

Expected Value - $\boxed{\mathbb{E}[X] = \sum_x x\cdot\mathbb{P}(X=x)}$
Variance - a probability weighted mean of the possible squared deviations.

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}\left[(X-\mathbb{E}[X])^2\right]\\ &= \sum_x (x-\mathbb{E}[X])^2\cdot\mathbb{P}(X=x)\\ &= \mathbb{E}\left[X^2\right]-(\mathbb{E}[X])^2\end{aligned}$$

Standard Deviation - $\boxed{\text{SD}(X)=\sqrt{\text{Var}(X)}}$
Given $Y=g(X)$ and X's PDF -

$$\mathbb{E}[Y] = \sum_x g(x)\cdot\mathbb{P}(X=x)$$
$$\text{Var}(Y) = \sum_x (g(x)-\mathbb{E}[g(X)])^2\cdot\mathbb{P}(X=x)$$

RVs with only 2 outcomes - (not necessarily Bernoullis yet) Suppose RV X's PDF is:

$$\begin{cases}\mathbb{P}(X=a) & =p\\ \mathbb{P}(X=b) & =1-p\text{, then:}\\ \mathbb{P}(X=\text{all other values}) & =0\end{cases}$$

$$\mathbb{E}[X] = ap+b(1-p)$$
$$\text{Var}(X) = (a-b)^2p(1-p)$$
$$\text{SD}(X) = |a-b|\sqrt{p(1-p)}$$

Bernoulli Random Variable - aforementioned when $\begin{cases}a & =1\\ b & =0\end{cases}$. If $\boxed{X\sim\text{Bern(p)}}$, then:

$$\mathbb{E}[X]=p$$
$$\text{Var}(X) = p(1-p)$$
$$\text{SD}(X) = \sqrt{p(1-p)}$$

- Variance maximized when $p=0.5$
- Variance minimized when $p=0$ or 1

Useful for tracking how many successes happen in n independent trials.
Binomial Problems - following must hold:

1. Constant success probability p and failure probability $(1-p)$.
2. Fixed total number of trials: n
3. trials are **independent**
4. Only two outcomes of interest (success or failure) on each trial
5. Want to find the probability of observing k successes among the total number of n trials. (Order doesn't matter.)

$\boxed{\mathbb{P}(k \text{ successes in } n \text{ trials}) = \binom{n}{k}p^k(1-p)^{n-k}}$
Combination - how many ways to choose k out of n: $\boxed{\binom{n}{k} = \frac{n!}{k!(n-k!)}}$.
Binomial Distribution - $\boxed{X\sim\text{Binom}(n,p)}$ where X is an RV tracking the number of successes in n independent trials with success probability p. X's PDF:
$\boxed{\forall k\in\{0,\ldots,n\}, \mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}}$.
Attn: X here is not for one single trial!.

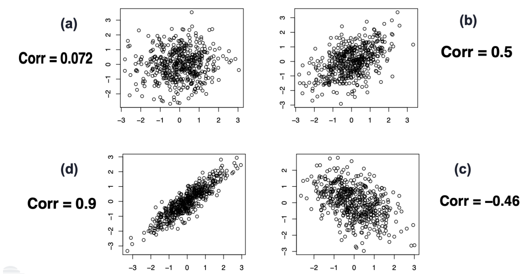
- A **Bernoulli RV**: useful for one trial's success/failure.
- A **Binomial RV**: useful for total number of successes.

Binomial as the Sum of Bernoullis - n independent Bernoulli RVs each with the same success probability p: $\forall i\in 1,\ldots,n, X_i\sim\text{Bern}(p)$.
Define $\boxed{S_n=\sum_{i=1}^n X_i}$, then denote $\boxed{S_n\sim\text{Binom}(n,p)}$. $\text{Binom}(1,p)=\text{Bern}(p)$
Joint Distribution of 2 RVs - the probability that 2 RVs simultaneously take on 2 values.
$\forall x\in X, \forall y\in Y$ $\boxed{\mathbb{P}(X=x, Y=y)}$.
Marginal probability distribution - can be found given with the joint PDF:
$\boxed{\mathbb{P}(X=x)=\sum_y \mathbb{P}(X=x,Y=y)}$
Z-Score of a Random Variable X -
$\boxed{Z(X)=\frac{X-\mathbb{E}[X]}{SD(X)}}$
$\mathbb{E}[Z(X)]=0$ and $SD(Z(X))=1$
Correlation Between RVs X and Y - "average of the product of z-scores"

$$\begin{aligned}\text{Corr}(X,Y) &= \mathbb{E}[Z(X)\cdot Z(Y)]\\ &= \frac{\text{Cov(X, Y)}}{SD(X)\cdot SD(Y)}\end{aligned}$$

- $\text{Corr}(X,Y)$ is unit-free.
- $\text{Corr}(X,Y)$ doesn't exist if either $SD(X)=0$ or $SD(Y)=0$ (can't divide by 0!).
- Correlation is guaranteed to lie between $+1$ (perfect positive correlation) and -1 (perfect negative correlation). Hence Corr is more commonly used than Covariance.

Correlation Examples



Covariance Between RVs X and Y - "average of the product of the centered variables". Necessary for assessing variability of sums of RVs (e.g. portfolios).

$$\begin{aligned}\text{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]\\ &= \text{Corr}(X,Y)\cdot SD(X)\cdot SD(Y)\end{aligned}$$

- $\text{Cov}(X,Y)$ has funny units: product of the X and Y units.
- $\text{Cov}(X,Y)$ always exists. If SDs are 0, $\text{Cov}(X,Y)=0$
- $\text{Cov}(X,X)=V(X)$
- If $SD(X)>0$ and $SD(Y)>0$, then $\text{Corr}(X,Y)$ and $\text{Cov}(X,Y)$ have the same sign.

Expected Value of RVs summed - is regardless of RVs' joint distribution:

$$\mathbb{E}[X+Y] = \mathbb{E}[X]+\mathbb{E}[Y]$$
$$\mathbb{E}[X+Y+W] = \mathbb{E}[X]+\mathbb{E}[Y]+\mathbb{E}[W]$$
$$\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = \mathbb{E}[X_i]+\cdots+\mathbb{E}[X_n]$$

Variance of of RVs summed -

$$\text{Var}(X+Y) = \text{Var}(X)+\text{Var}(Y)+2\text{Cov}(X,Y)$$
$$\text{Var}(X+Y+W) = \text{Var}(X)+\text{Var}(Y)+\text{Var}(W)$$
$$+2\text{Cov}(X,Y)+2\text{Cov}(X,W)+2\text{Cov}(Y,W)$$
$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)+2\sum_{i<j}\text{Cov}(X_i,X_j)$$

Have to consider the covariance of all possible pairs: X_i and X_j.

- If $\text{Corr}(X,Y)$ increases, then $\text{Var}(X+Y)$ increases.
- If $V(X)=V(Y)$, then $\text{Var}(X+Y)$ is maximized when $\text{Cov}(X,Y)$ is maximized.

Uncorrelated RVs - if $\mathrm{Corr}(X,Y) = 0$. Equivalently, they are uncorrelated if $\mathrm{Cov}(X,Y) = 0, SD(X) > 0, SD(Y) > 0$

Variance of of Uncorrelated or Independent RVs summed -

$$\boxed{\begin{aligned} \mathrm{Var}(X+Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) \\ \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) \end{aligned}}$$

No change in expected value's formula.

Independent RVs -

$$\boxed{\forall x, y, \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)}$$

Independence implies uncorrelatedness: if two RVs X and Y are independent, then they are uncorrelated.

$$\boxed{\text{Independence} \Rightarrow \mathrm{Corr}(X,Y) = 0 = \mathrm{Cov}(X,Y)}$$

But uncorrelated RVs can be dependent!

***(iid)* Independent and Identically Distributed RVs -** for a collection of *iid* RVs $\{X_1, \ldots, X_n\}$: $\forall i \in \{1, \ldots, n\}$

$$\boxed{\begin{aligned} \mathbb{E}[X_i] &= \mu \\ \mathrm{Var}(X_i) &= \sigma^2 \\ \mathrm{SD}(X_i) &= \sigma \end{aligned}}$$

Sum of *iid* RVs - $S_n = X_1 + \cdots + X_n$:

$$\boxed{\begin{aligned} \mathbb{E}[S_n] &= n \cdot \mu \\ \mathrm{Var}(S_n) &= n \cdot \sigma^2 \\ \mathrm{SD}(X_i) &= \sqrt{n} \cdot \sigma \end{aligned}}$$

Central Limit Theorem (CLT)

If $\{X_1, \ldots, X_n\}$ are *iid* with expected value $\mathbb{E}[X_i] = \mu$ and variance $\mathrm{Var}(X_i) = \sigma^2 < \infty$, then as $n \to \infty$:

- $S_n \sim \mathcal{N}(n \cdot \mu, \sqrt{n} \cdot \sigma)$
- $\mathrm{Mean}_n \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$

If n is large enough (heuristic: $n > 30$), we can calculate probabilities for the sum and mean of RVs by using the normal distribution.

Emperical Rules under CLT -

- 50% of the time,
  - S_n will fall within $n\mu \pm \frac{2}{3}\sqrt{n}\sigma$
  - M_n will fall within $\mu \pm \frac{2}{3}\frac{\sigma}{\sqrt{n}}$
- 68% of the time,
  - S_n will fall within $n\mu \pm \sqrt{n}\sigma$
  - M_n will fall within $\mu \pm \frac{\sigma}{\sqrt{n}}$
- 95% of the time,
  - S_n will fall within $n\mu \pm 2\sqrt{n}\sigma$
  - M_n will fall within $\mu \pm 2\frac{\sigma}{\sqrt{n}}$
- 99.7% of the time,
  - S_n will fall within $n\mu \pm 3\sqrt{n}\sigma$
  - M_n will fall within $\mu \pm 3\frac{\sigma}{\sqrt{n}}$

Sampling and Confidence Intervals

Confidence Interval - contains an unknown (population) quantity at some specified sampling frequency.

- Confidence intervals do not depend on population size, but only on sample size.
- For a given sample size, can be very precise with low confidence or very imprecise with high confidence.
- 2x the precision requires 4x sample size; 3x the precision requires 9x sample size.

Wording matters...

- OK: "I am 95% confident the interval [a,b] contains the true population proportion"
- OK: "There is a 95% probability the interval [a,b] contains the true population proportion"
- Not OK: "There is a 95% probability the true population proportion lies in the interval [a,b]"

Confidence Level (L) to c -

$\boxed{c = \texttt{qnorm}(p = \frac{(1+L)}{2}, \mu = 0, \sigma = 1)}$ (find the value c such that the area under $\mathcal{N}(\mu, \sigma)$ is p)

confidence level	L	c
90%	0.9	1.65
95%	0.95	1.96
99%	0.99	2.58

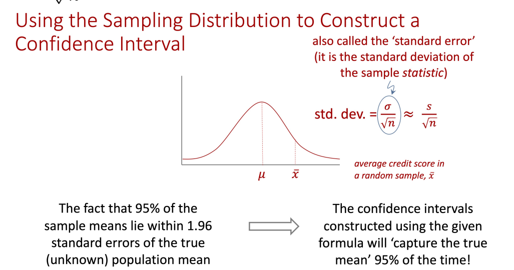
For a population mean - Given sample size n, sample average $\bar{x}$, and sample standard deviation s, we are X% confident the true population mean lies in the interval: $\boxed{\bar{x} \pm \left(c\frac{s}{\sqrt{n}}\right)}$ "MOE": $\left(c\frac{s}{\sqrt{n}}\right)$

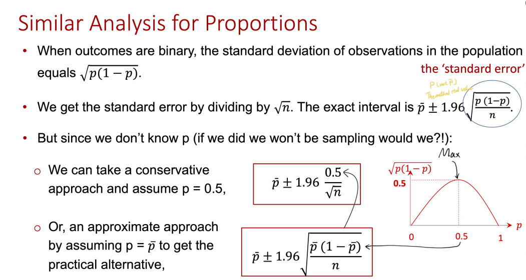
For a population proportion - Given sample size n, sample proportion $\bar{p}$, and standard deviation in the population to be 0.5, we are X% confident the true population mean lies in the interval: $\boxed{\bar{p} \pm \left(c\frac{0.5}{\sqrt{n}}\right)}$ "MOE": $\left(c\frac{0.5}{\sqrt{n}}\right)$

Important assumptions:

- Sample is random
- Sample is large enough ($n > 30$) for CLT
- The worst possible standard deviation in the population to be 0.5

Sampling Distribution - is well approximated by $\mathcal{N}(\mu = \text{true population parameter}, \sigma = \frac{\text{population SD}}{\sqrt{n}})$, based on CLT.





Sampling Errors - the sample-to-sample variations due to pure chance. MOE and confidence intervals quantify this uncertainty well.

Non-Sampling Errors - (some examples)

- Selection Bias: happens when each member of the population does not have the same chance of being selected.
- Response/Non-response Bias: happens when some fraction of the individuals surveyed don't respond for reasons related to what's being asked in the survey

R Distribution Functions

- `p` ("probability"): cumulative distribution function ("what is the probability above or below a cutoff?")
- `q` ("quantile"): inverse CDF ("what value do we find at, say, 80% of the way to the maximal value?")
- `d` ("density"): density function (gives us the "height" or y-value of distribution for a particular z-score - mainly useful in plotting)

```
# Exactly k successes in n trials
    given success probability p
dbinom(k, size=n, p=p)
# k or more successes in n trials
    given success probability p
sum(dbinom(k:n, size=n, p=p))

# Probability of this value or less
pnorm(value, mean, sd)
# Probability of this value or
    greater
pnorm(value, mean, sd, lower.tail=
    FALSE)
# Highest value associated with a
    given percentile
qnorm(percentile, mean, sd)
```