\[ X \sim \text{Bin}(m, p) \]
\[ Y \sim \text{Bin}(n, p) \]
\[ X + Y \sim \text{Bin}(m + n, p) \]

Binomial: trials must be iid.
- independent
- same \( p \).

---

Random 5 card hand

Joint distribution of the number of Aces:

\[ X = \# \text{ aces}. \]
\[ \{0, 1, 2, 3, 4, 5\} \]

\[ P(X = k) = 0 \] unless \( k \in \{0, 1, 2, 3, 4, 5\} \)

- trials are not independent.

\[ P(X = k) = \frac{\binom{4}{k} \cdot \binom{48}{5-k}}{\binom{52}{5}} = \frac{\text{4 aces. choose } k \text{ cards}}{\text{5 cards choose } 5 - k}. \]

number of 5 card hands.
Generalize.

Deer on campus,
- some have been tagged
- some have not been tagged.

Let tagged deer = $t$
Let untagged deer = $d$

Pick a simple random sample of size $n$
- all subsets of size $n$ are equally likely.

Find the distribution of the number of tagged deer in the sample.

$$P(X = k) = \frac{\binom{t}{k} \times \binom{d}{n-k}}{\binom{t+d}{n}} \quad 0 \leq k \leq t \quad 0 \leq n-k \leq b$$

Hypergeometric Distribution
- sampling without replacement.

$$\left( \frac{\binom{w}{k}/\binom{b}{n-k}}{\binom{w+b}{n}} \right)$$

(sampling with replacement.

if $t+d > n$
Is this a valid PMF?

\[ P(x = k) \geq 0 \]

\[ \sum_{k} P(x = k) = 1. \]

\[ \sum_{k=0}^{\infty} \frac{(e)(d)}{(e+d)^n} \cdot \frac{(n-k)!}{(n)!} \cdot \frac{1}{(e+d)^n} \cdot \frac{(e)(d)}{(n-k)!} = 1 \]

CDF:

\[ F(x) = P(x \leq x) \]

- Real value
- RV: may be discrete or continuous
- PFM: jump sizes
- Jumps are non-negative and add to 1.
\[ P(1 < x \leq 3) \]

\[ P(x \leq 1) + P(1 < x \leq 3) = P(x \leq 3) \]

\[ P(1 < x \leq 3) = F(3) - F(1) \]

\[ P(a < x \leq b) = F(b) - F(a) \]

- **Discrete RV** - be careful with \( < \) \( \leq \)
- **Continuous RV** - doesn't matter.

**Important Properties**

1) **Increasing**  \( \text{(not strictly increasing)} \)

2) **Right continuous**
   - as you approach any point from the right, the value of the function converges to the value at that point.

3) \[ F(x) \rightarrow 0 \quad \text{as} \quad x \rightarrow -\infty \]
   \[ F(x) \rightarrow 1 \quad \text{as} \quad x \rightarrow +\infty \]

**iff** \( \forall x \) \( F(x) \) satisfies these 3 conditions, it corresponds to a valid CDF.
Independence of RVs.

\[ P(x \leq x, y \leq y) = P(x \leq x) P(y \leq y) \quad \text{for all } x, y \]
\[ \text{event } x \leq x \quad \text{event } y \leq y. \]
\[ \text{these 2 events are independent.} \]

Discrete.

\[ P(x = x, y = y) = P(x = x) P(y = y) \]
Averages.

\[ 1, 2, 3, 4, 5, 6 \]
\[ \text{mean} = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3.5 \]

\[ 1, 1, 1, 1, 3, 4, 6, 6 \]

2 ways to find average

\[ \frac{1 + 1 + 1 + 3 + 4 + 6 + 6}{8} = \frac{23}{8} \]

\[ \frac{4}{8} \times 1 + \frac{1}{8} \times 3 + \frac{1}{8} \times 4 + \frac{2}{8} \times 6 \]

Find the mean as weighted sum.

Weights are relative frequencies of each value.

\[ E(X) = \sum_{\infty} x \cdot P(X=x) \quad \text{Average (Expectation) of a discrete RV, X.} \]

\[ X \sim \text{Bernoulli} (p) \]

\[ E(X) = 1 \cdot P(X=1) + 0 \cdot P(X=0) \]

\[ = 1 \cdot p + 0 \]

\[ = p. \]
Indicator RV.

\[ X = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases} \quad \text{for event } A. \]

\[ E(X) = P(A) \]

The probability that you are interested in is equivalent to the expected value of a suitably chosen indicator RV.

\[ X \sim \text{Binomial} \left(n, p\right) \]

\[ E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \]

\[ = \sum_{k=0}^{n} n \binom{n-1}{k-1} p^k q^{n-k} \]

\[ = np \sum_{k=0}^{n} \binom{n-1}{k-1} p^k q^{n-k} \]

\[ = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k}. \]

let \( j = k-1 \)

\[ \binom{n}{k} = 0 \quad \text{for } k < n. \]
\[
= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} \\
\text{Binomial Theorem} \quad (p+q)^{n-1} = 1.
\]

\[
= np
\]

\[
E(X + Y) = E(X) + E(Y) \quad \text{Linearity of Expectation}
\]

\[
E(cX) = cE(X) \quad \text{even if } X \text{ and } Y \text{ are dependent}
\]

\[
X \sim \text{Binomial}(n, p) \\
\text{Sum of } n \text{ Bernoulli}(p) \text{ random variables}
\]

\[
X = X_1 + X_2 + \ldots + X_n
\]

\[
E(X) = E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n) = np
\]

\^ This is true even when the \(X_i\)'s are not independent.
Hypergeometric

\[ E(X) = \sum_{k=0}^{t} \binom{k}{x} \left( \frac{t-k}{t} \right) \cdot \binom{n-k}{d-x} \cdot \binom{n-d}{d-k} \]

- unpleasant.

5 cards. \( X = \# \text{aces} \).

\( X_j \) \( j \)th card is an ace

\[ X = \sum_{j=1}^{5} X_j \]

\[ E(X) = E(X_1 + X_2 + \ldots + X_5) \]

\[ = E(X_1) + E(X_2) + \ldots + E(X_5) \]

\[ = 5 \cdot E(X_1) \]

\[ = 5 \times \frac{4}{52} \]

\[ E(X) = \frac{5}{13} \]

\( \uparrow \) even though \( X_j \)'s are dependent.

Generalize:

Expected value of \( \frac{X}{n} \) = \( \frac{np}{n} \) of a Hypergeometric
**Geometric Distribution.**

$X \sim \text{Geometric}(p)$

Independent Bernoulli ($p$) trials. Count the number of failures before the first success.

(don't count the success; sometimes $\text{Geom}(p)$ is defined)

(careful: sometimes $\text{Geom}(p)$ is defined)

including the success

PMF. $P(X = k)$

$\begin{align*}
\text{H.} & \quad P(X = 5) = q^5 p \\
\text{H.} & \quad P(X = k) = q^k p \\
& \quad k \in \{0, 1, 2, 3, \ldots \}
\end{align*}$

Is this a valid PMF?

1) $P(X = k) > 0$

2) $\sum_{k=0}^{\infty} q^k p = \frac{p}{1-q}$

---

**Geometric Series.**

$\frac{1}{1-q}$
Expected value.

\[ E(X) = \sum_{k=0}^{\infty} k \cdot p \cdot q^k \]

\[ = p \sum_{k=0}^{\infty} kq^k \]

\[ = p \frac{q}{(1-q)^2} \]

\[ = \frac{pq^2}{p^2} = \frac{q}{p} \]

\[ \sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \]

\[ \sum_{k=0}^{\infty} kq^k = \frac{q}{(1-q)^2} \]

Take derivative wrt \( q \).

\[ \sum_{k=0}^{\infty} kq^{k-1} = \frac{q}{(1-q)^2} \]

Multiply by \( q \).

\[ \sum_{k=0}^{\infty} kq^k = \frac{q}{(1-q)^2} \]

Alternative way.

Consider the first flip.

If \( H \), then no failures \( X=0 \) with prob. \( p \).

\[ E(X) = p \cdot 0 + q \left( 1 + E(X) \right) \]

\[ \rightarrow \text{after the failure, we have} \]

1 failure, and the problem re-starts.

\[ E(X) = q + q \cdot E(X) \]

\[ E(X) = \frac{q}{1-q} = \frac{q}{p} \]
$$E(x) = 9 + 9 \delta E(x).$$

$$E(x)(1-\delta) = 9.$$ 

$$E(x) = \frac{9}{1-\delta} = \frac{9}{\delta}.$$