8

Random Variable Function

what is the average of 3 coin flips ?
- you can NOT average coin flips - HHT

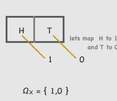
how random are the 3 coin flips (measure of randomness) ?
- you can't perform computations on HHT (arbitrary sets)

$\Omega = \{ H,T \}$
$\omega_1 = H$
$\omega_2 = H$ n=3
$\omega_3 = T$

what if I create a function (mapping) e.g.

indicator function returns bool

$\mathbb{1}_{\omega=H} = \begin{cases} 1 & \text{if } \omega=H \\ 0 & \text{otherwise} \end{cases}$ alternative possibility $\mathbb{1}_{\omega=T} = \begin{cases} 0 & \text{if } \omega=T \\ 1 & \text{otherwise} \end{cases}$



lets map H to 1 and T to 0

once represented numerically can I take the average ?

$\Omega_X = \{ 1,0 \}$
$\omega_1 = 1$
$\omega_2 = 1$ n=3
$\omega_3 = 0$

avg r.v. $\overline{X} = \frac{1+1+0}{3} = \frac{2}{3}$
*future insight

expected outcome values you care about in your random model

what did we do ?
X is a function called a **random variable** (r.v.) -
$X(\omega)$
$X(H) = 1$
$X(T) = 0$

Def
a **random variable** ('r.v.') is a function (usually denoted X) UPPER CASE

difficult to compute — easy to compute
$X: \underset{\text{sample space}}{\Omega} \rightarrow \underset{\text{value of rv.}}{\mathbb{R}}$
X is r.v. function s/t if its input/domain is Ω (sample space), then its output/range are Real Numbers

what is the probability that X=1 P(X=1) ? —same as→ Def: $P: 2^{\Omega} \rightarrow [0,1]$
technically illegal 'abuse of notation' - overloaded operator
the probability of all the possible events is between 0 and 1 (inclusively)

X only returns 1 when given H

$P(\{\omega: X(\omega)=1\}) = P(\{H\}) = \frac{|\{H\}|}{|\Omega|} = 1/2 \rightarrow P(X=1) = 1/2$, $P(X=0) = 1/2$

thus the probability that X=1 is the same as of H

```
#include <iostream>
using namespace std ;

int  X( string coin_Side ) {
     if ( coin_Side=="Heads") return 1 ;
     else return 0 ;
}
int main() {
     int outcome = X( "Heads" ) ;

     cout << outcome << endl ;
     return 0 ;
}
```

Program Output:
1
Program ended with exit code: 0

Support of X

Def
the **support of X** is the range of X (set of all possible values)

$\text{Supp}[X] := \{ x: \underbrace{P(X=x)}_{\text{cashing in on the previous 'abuse of notation'}} > 0 \} \subseteq \mathbb{R}$
x : an arbitrary value in the range of the function X

self-note:
the probability that x is in the range of X must obviously be > 0 since x is in the range of X. It is like asking 'who is buried in Grants tomb?'.

thus in the coin flip experiment $\text{Supp}[X] = \{ 1,0 \}$

Discrete X

Def
a **discrete X** is an X s/t $|\text{Supp}[X]| \leq |\mathcal{N}|$ (no decimals)

$\sum_{x \in \text{Supp}[X]} P(X=x) = 1$
we are adding the probabilities of all of the values in the range/support of X
recall that: $P(\{H\}\cup\{T\}) = 1$ OR simply $P(\Omega) = 1$

$|\mathcal{N}| = \aleph$ — aleph - countable infinity
infinity that you can count to
anything that can happen will happen
Murphy's Law: "Anything that can go wrong, will go wrong."

thus in the coin flip experiment $|\text{Supp}[X]| = 2$

$\bigcup_{x \in \text{Supp}[X]} \{ \omega \text{ s.t. } X(\omega)=x \} = \Omega$?
ω outcomes are collectively exhaustive

if not $\exists\omega$ s/t $X(\omega) \notin \text{Supp}[X] \rightarrow P(\{\omega\}) = 0 \rightarrow$ collectively exhaustive
if an outcome ω is not in the 'support of X', then its probability of occurrence must be 0

$\{ \omega: X(\omega)=x_1 \} \cap \{ \omega: X(\omega)=x_2 \} = \varnothing$?

YES otherwise $X(\omega_1) = x_1$ and x_2 which violates the definition of a function

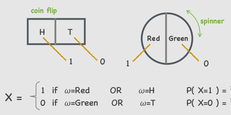
$P(\Omega) = P(\{\omega: X(\omega)=x_1\}) + P(\{\omega: X(\omega)=x_2\}) + \dots$
$1 = \sum_{x \in \text{Supp}[X]} P(X=x)$ QED

something has to happen
probability of everything happening is 1

Deg - In mathematics, a **degenerate distribution** or **deterministic distribution** is the probability distribution of a random variable which **only takes a single value**. Examples include a two-headed coin and rolling a die whose sides all show the same number. This distribution satisfies the definition of "random variable" even though it does not appear random in the everyday sense of the word; hence it is considered degenerate.

Bernoulli

Are these two experiments the same ?



self-note:
if two disparate experiments are only observed from the perspective of their outcomes and their outcomes happen to be homogenous, then the source of origin looses its individuality.

$X = \begin{cases} 1 \text{ if } \omega=\text{Red} & \text{OR} \quad \omega=H \quad P(X=1) = 1/2 \\ 0 \text{ if } \omega=\text{Green} & \text{OR} \quad \omega=T \quad P(X=0) = 1/2 \end{cases}$

Technically $X: \{H,T\} \rightarrow \{1,0\}$ and $X: \{R,G\} \rightarrow \{1,0\}$ are different b/c their formal parameters are different (LHS accepts apples and RHS accepts oranges) but if we were to ONLY compare their outcomes then they are identical (both LHS and RHS returns either 1 or 0).

$X \underset{\text{distributive as}}{\sim} \begin{cases} 1 & \text{wp } 1/2 \\ 0 & \text{wp } 1/2 \end{cases}$ with probability
'random variable' is distributive as 0 or 1 with probability of $1/2$ for each

We can map as many things to 1 and 0 as we want, therefore there are many Ωs that can produce this X. Thus we do NOT care about Ω anymore. We know that Ω is there. We know there is some underlying experiment and sample space, but we do NOT need to know what it is.

$X \sim \text{Bernoulli}(1/2) := \begin{cases} 1 & \text{wp } 1/2 \\ 0 & \text{wp } 1/2 \end{cases}$ 'Standard Bernoulli'
'random variable' is distributive as Bernoulli($1/2$) which is defined as 0 or 1 with probability of $1/2$ for each

or more generally $\text{Supp}[X] = \{ 1,0 \}$
remember! if Supp is NOT 0 and 1, NOT a Bernoulli

$X \sim \text{Bernoulli}(p) := \begin{cases} 1 & \text{wp } p \\ 0 & \text{wp } 1-p \end{cases}$

X is distributed Bernoulli with '**parameter p**' (p is a probability). A parameter is a choice which defines a model. Space in which p lives is called the **parameter space** (in parallel to event space).

what are the possible values of p ?

$p \in [0,1]$
probability of occurring
what if p=1? $X = \{ 1 \text{ wp } 1 \quad 1,1,1,1,1,1 \}$ — there is 100% chance of something happening
what if p=0? $X = \{ 0 \text{ wp } 0 \quad 0,0,0,0,0,0 \}$ — there is 0% chance of something happening
probability of occurring
→ **degenerate X**

if $X \sim \text{Deg}(c) := \{ c \text{ wp } 1 \}$, $\text{Supp}[X] = \{c\}$
here X is a r.v. by definition but it is trivial and unnecessary. Why? B/c it just spits out the same value each time. Why do we need to measure the probability of something that is 100% bound to happen ? - boring

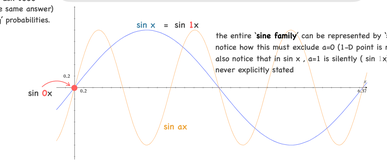
Just like $f(x)= \sin(0x) = 0$ is unnecessary and a=0 was not included in the set of values defining the sine curve family, so too p=0 and p=1 are not included in the parameter space.
thus
$p \in (0,1)$ and NOT $p \in [0,1]$

sample space — event space (power set of Ω)
$\Omega := \{H,T\} \rightarrow 2^{\Omega} = \{ \varnothing, \{H\}, \{T\}, \{H,T\} \}$

Power Set (by definition of a non-empty set) must also include Ø and itself. note that: $P(\Omega)=1$ and $P(\varnothing)=0$ is non-interesting (you can ask 1000 times about the probability of either and you will always get the same answer) Thus the **parameter space** excludes evaluating these two 'boring' probabilities.

$2^{\Omega} = \{ \varnothing, \{H\}, \{T\}, \{H,T\} \}$
$\varnothing$: p=0, $\{H\}$: $p=1/2$, $\{T\}$: $p=1/2$, $\{H,T\}$: p=1
$p \in (0,1)$

MNEMONIC: 1,0
X~Bernoulli( )



the entire '**sine family**' can be represented by '**sin ax**'
notice how this must exclude a=0 (1-0 point is not a curve)
also notice that in sin x , a=1 is silently (sin 1x) included and never explicitly stated

PMF

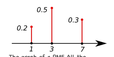
- if a value=x is in the support of X its probability of occurring must be greater than 0
- if a value=x is NOT in the support of X its probability of occurring must be 0

Def
the **Probability Mass Function - PMF is p(x)** $:= P(X=x)$

$p: \mathbb{R} \rightarrow [0,1]$ defined everywhere
if $x \in \text{Supp}[X]$, $p(x) > 0$
if $x \notin \text{Supp}[X]$, $p(x) = 0$

$\sum_{x \in \text{Supp}[X]} p(x) = 1$ — monotonically increase
the probabilities of all the values in the support of X add up to 1

X or r.v
discrete
Probability Mass Function (PMF) is a function that gives the probability that a discrete random variable is exactly equal to some value.



The graph of a PMF. All the values of this function must be non-negative and sum up to 1.

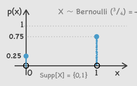
real-valued (continuous)
Cumulative Distribution Function (CDF) is a function whose value is the probability that a corresponding continuous random variable has a value less than or equal to the argument of the function.

In probability theory and statistics, the cumulative distribution function (CDF) of a real-valued random variable X, or just distribution function of X, evaluated at x, is the probability that X will take a value less than or equal to x.

similar to asking : what is the probability that the coin will be H ?

it is useful to plot PMF

$X \sim \text{Bernoulli}(1/2) := \begin{cases} 1 & \text{wp } 1/2 \\ 0 & \text{wp } 1/2 \end{cases}$



p(x) takes the x_i from the Supp[X] and maps them to their probabilities

is p(x) a cont. function ? - No
# of discontinuities ? = Supp[X]
Supp[X] = {0,1}

$X \sim \text{Rademacher} = \begin{cases} 1 & \text{wp } 1/2 \\ -1 & \text{wp } 1/2 \end{cases}$ $\text{Supp}[X] = \{-1,1\}$

random walk

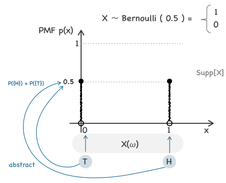
the Rademacher Distribution (which is named after Hans Rademacher) is a discrete probability distribution where a random variable X has a 50% chance of being either +1 or -1.

$X \sim \text{Uniform}(\{1,10,100\}) := \begin{cases} 1 & \text{wp } 1/3 \\ 10 & \text{wp } 1/3 \\ 100 & \text{wp } 1/3 \end{cases}$

6 sided dice is:
$X \sim \text{Uniform}(\{1,2,3,4,5,6\}) = \begin{cases} 1 & \text{wp } 1/6 \\ 2 & \text{wp } 1/6 \\ 3 & \text{wp } 1/6 \\ 4 & \text{wp } 1/6 \\ 5 & \text{wp } 1/6 \\ 6 & \text{wp } 1/6 \end{cases}$

$X \sim \text{Uniform}(A)$ $\text{Supp}[X]=A$
Parameter Space: $A \subset \mathbb{R}$
$|A| \in \mathcal{N} \setminus \{1\}$ if $|A|=1$, Deg!
you need at least 2 values in the support of X

$X \sim \text{Bernoulli}(0.5) = \begin{cases} 1 & \text{wp } 1/2 \\ 0 & \text{wp } 1/2 \end{cases}$



$\text{Supp}[X] = \{0,1\}$ $\{H,T\} \in \Omega$
X

CDF

Def
the **Cumulative Distribution Function - CDF is F(x)** $:= P(X \leq x)$

properties (derivable form the above definition)
1 $F(x) \in [0,1]$ — monotonically increasing
2 $\lim_{x \to \infty} F(x) = 1$
3 $\lim_{x \to -\infty} F(x) = 0$
4 $a \leq b \rightarrow F(a) \leq F(b)$
$a < b \rightarrow F(a) < F(b)$ False

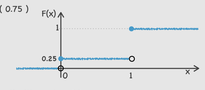
self-note:
can you ask: "what is the probability that the coin will be between H and T" or "be less than or equal to H?"

NO - this question is only valid in a discrete world.

Imagine if you could freeze time as the coin is rotating through the air. You take snapshots of each H and T and once the coin has landed you analyze the captured times. Your measuring device is extremely accurate and gives you precision to 20 decimal places. Using CDF function you can now express your data.

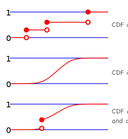
$X \sim \text{smthg}(p) = \begin{cases} 0 & : x < 0 \\ .5 & : 0 \leq x < 1 \\ 1 & : 1 \leq x \end{cases}$

$X \sim \text{Bernoulli}(0.75)$



P(-37) = 0
P(.89) = .25
P(5) = .25
P(1001) = 1

# of discontinuities = Supp[X]



$X \sim \text{Bernoulli}(p) := \begin{cases} 1 & \text{wp } p \\ 0 & \text{wp } 1-p \end{cases} = p(x) = p^x(1-p)^{1-x}$
nice !
piecewise function (ugly difficult)

$X \sim \text{Bernoulli}(p) := p^x(1-p)^{1-x}$

$X_1 \sim \text{Bernoulli}(p)$ $p_1(x) = p^x(1-p)^{1-x}$
$X_2 \sim \text{Bernoulli}(p)$ $p_2(x) = p^x(1-p)^{1-x}$
share the same PMF
equal

Def
X_1 and X_2 are "**equal in distribution**" $X_1 \overset{d}{=} X_2$
if $p_1(x) = p_2(x)$ or $F_1(x) = F_2(x)$

verify: $\sum_{x \in \text{Supp}[X]} p(x) = 1$

$\sum_{x \in \{0,1\}} p^x(1-p)^{1-x} = p^0(1-p)^{1-0} + p^1(1-p)^{1-1}$
$= 1(1-p)^1 + p^1(1-p)^0$
$= 1-p + p$
$= 1$ ✔

9

Hypergeometric

Drawing cards without replacement.
10 Cards
4 Red
6 Blue

$P(2R \text{ in } 3 \text{ draws}) = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}}$

$P(xR \text{ in } 3 \text{ draws}) = \frac{\binom{4}{x}\binom{6}{3-x}}{\binom{10}{3}}$

$P(xR \text{ in } n \text{ draws}) = \frac{\binom{4}{x}\binom{6}{n-x}}{\binom{10}{n}}$

In probability theory and statistics, the **hypergeometric distribution** is a **discrete probability distribution** that describes the probability of **x** successes in **n** draws, without replacement, from a finite population of size **N** that contains exactly **K** successes, wherein each draw is either a success or a failure.

10 Cards
K Red
10-K Blue

$P(xR \text{ in } n \text{ draws}) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$

N Cards — **population size**
K Red — **# of success**
N-K Blue — failure
n — **sample size**
3 knobs you can turn !

$P(xR \text{ in } n \text{ draws}) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$
P(Success)

N=7
K=3
n - up to you

$X \sim \text{Hypergeometric}(n,K,N) := p(x) :=$
parameter space with 3 parameters - Bernoulli had only 1

e.g. We have a bag with 118 balls , 37 are success. We draw 23. What is the probability of getting 16 successes?

N=118
K=37 $X \sim \text{Hypergeometric}(23,37,118)$
n=23

$P(X=16) = \frac{\binom{37}{16}\binom{81}{7}}{\binom{118}{23}}$ to calculate this use Sterlings approximation

e.g. We have a 100 soldiers , 53 are female. Pick 8 at random. What is the probability of 6 being female?

N=100
K=53 $X \sim \text{Hypergeometric}(8,53,100)$
n=8

$P(X=6) = \frac{\binom{53}{6}\binom{47}{2}}{\binom{100}{8}}$

Parameter Space , Support

what can these knobs be? (figuring out the parameter space)

N=0 ? (bag with no balls in it - NOT interesting)

N=1 ?
K=0 or 1 (ball must be success or failure)
n=0 (we draw 0 balls - NOT interesting)
n=1 (we draw 1 ball - NOT interesting)
Degenerate Distribution
X~Deg(0) or (1)

N=2 ? not interesting
K=0 , 1 or 2
whether n=1 or n=2 always failure — whether n=1 or n=2 always success
K=0 → X~Deg(0)
K=2 → X~Deg(1)
n=1
draw 1 of 2 is it a success? hmm... interesting
$K=1 \rightarrow X \sim \text{Hyper}(1,1,2) = \frac{\binom{1}{x}\binom{1}{1-x}}{\binom{2}{1}}$

special case
$X \sim \text{HyperGeometric}(1,1,2) = X \sim \text{Bernoulli}(1/2)$
$P(X=0) = 1/2$
$P(X=1) = 1/2$ = Bernoulli($1/2$)

N=3 ?
K=1,2 0,3→Deg
n=1,2 n=3 ? → no mystery
1 ball - NOT interesting - start with at least 2
having all successes - NOT interesting
make successes < total size
$N \in \mathcal{N} \setminus \{1\}$
$K \in \{1,2,\dots,N-1\}$
$n \in \{1,2,\dots,N-1\}$
Parameter Space

$X \sim \text{HyperGeometric}(1,K,N) := \begin{cases} 1 & \text{wp } K/N \\ 0 & \text{wp } 1-K/N \end{cases}$

$X \sim \text{Bernoulli}(p) := \begin{cases} 1 & \text{wp } p \\ 0 & \text{wp } 1-p \end{cases}$

$X \sim \text{Hypergeometric}(1,K,N) = \text{Bernoulli}(\frac{K}{N}) = \frac{\binom{K}{x}\binom{N-K}{1-x}}{\binom{N}{1}}$
always the case for Bernoulli
"p"
Supp[X] = {0,1}

$P(X=0) = \frac{\binom{K}{0}\binom{N-K}{1}}{\binom{N}{1}} = \frac{N-K}{N} = \frac{N}{N} - \frac{K}{N} = 1 - \frac{K}{N}$ "1-p"

$P(X=1) = \frac{\binom{K}{1}\binom{N-K}{1-1}}{\binom{N}{1}} = \frac{K}{N}$ "p"
probability of success

N=7
K=3
n=1
$P(X=1) = \frac{K}{N} = \frac{3}{7}$
probability of success

4 cases of $X \sim \text{Hypergeometric}(n,K,N)$

	draws (n)	success (K)	total (N)	Supp[X]	
1	2	4	10	{0,1,2}	
2	5	4	10	{0,1,2,3,4}	5 draws
3	8	4	10	{2,3,4}	8 draws
4	5	7	10	{2,3,4,5}	5 draws

- 0: both draws can be failures
- 2: both draws can be successes — 2 draws
- 4: at most 4 possible successes
- 0: all 5 draws can be failures — 5 draws
- 4: at most 4 possible successes
- 2: at most 4 possible failures — 8 draws
- 2: at most 3 possible failures
- 5: can't have 6 successes in 5 draws — 5 draws

import Foundation

```
let n=8          // sample size
let N=10         // population size
let K=4          // # of success

let x = max( 0, n-(N-K) )
let y = min( n,K )

print(x)
print(y)
```

Program Output:
2
4
Program ended with exit code: 0

General Model
$X \sim \text{Hypergeometric}(n,K,N)$ Supp[X] ?

1 $n < K$, $n < N-K$ — $\text{Supp}[X] = \{0,1,\dots,n\}$ — choose less then success & less then failure
2 $n \geq K$, $n < N-K$ — $\text{Supp}[X] = \{0,1,\dots,K\}$ — choose more/equal then/to success & less then failure
3 $n \geq K$, $n \geq N-K$ — $\text{Supp}[X] = \{n-(N-K),\dots,K\}$ — choose more/equal then/to success & more/equal then/to failure
4 $n < K$, $n \geq N-K$ — $\text{Supp}[X] = \{n-(N-K),\dots,n\}$ — choose less then success & more/equal then/to failure

	n < K	n ≥ K
n < N-K	{0,1,...,n}	{0,1,...,K}
n ≥ N-K	{n-(N-K),...,n}	{n-(N-K),...,K}

$\text{Supp}[X] = \{ \max(0, n-(N-K)), \dots, \min(n,K) \}$

$\sum_{x \in \text{Supp}[X]} p(x) = 1$?
use Vandermonde's Identity to prove this

Equivalent Parametrization

n - sample size
N - population size
K - # of success
p - proportion of success

let $p = \frac{K}{N}$, then $K = pN$ — plug in for K — **equivalent parametrization**

$X \sim \text{Hypergeometric}(n,K,N) := \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$
N-pN = N(1-p)

$X \sim \text{Hypergeometric}(n,p,N) := \frac{\binom{pN}{x}\binom{(1-p)N}{n-x}}{\binom{N}{n}}$

Parameter Space
$N \in \mathcal{N} \setminus \{1\}$
$p \in \{ 1/N, 2/N, \dots, N-1/N \}$
$n \in \{1,2,\dots,N-1\}$

N=9
p=0.7 ? - NOT Legal
b/c K=6.3 makes not sense

7 failure balls
- 1st draw: the probability of picking a successful ball is $1/3$. Note that $1/3$ is a fraction aka ratio which is a proportion of success vs failure.
- 2nd draw: since there is no replacement the ratio can be either $2/5$ or $1/5$
- 3rd draw: $1/5$ or $2/6$ or $3/6$
- 4th draw: $2/5$ or $1/6$ or $3/6$ or $2/6$
- 5th draw: $2/6$ or $1/6$ or $3/6$ or $2/6$
- 6th draw: $2/6$ or $1/6$ or $3/6$ or $2/6$
- 7th draw: $2/6$ or $1/6$ or $3/6$ or $1/6$
- 8th draw: $2/6$ or $1/6$ or $2/6$ or $1/6$
- 9th draw: $2/6$ or $1/6$ or $2/6$
- 10th draw: $2/6$ or $1/6$

self-note:
10 balls total
3 success balls

consider p=0.5 , n=6 , N=100
$P(X=3) = \frac{\binom{50}{3}\binom{50}{3}}{\binom{100}{6}} = .3223$

consider p=0.5 , n=6 , N=1000
$P(X=3) = \frac{\binom{500}{3}\binom{500}{3}}{\binom{1000}{6}} = .3134$

consider p=0.5 , n=6 , N=10000
$P(X=3) = \frac{\binom{5000}{3}\binom{5000}{3}}{\binom{10000}{6}} = .3126$

converge $\lim_{N \to \infty} P(X=3)$?

$\lim_{N \to \infty} \text{PMF} = \lim_{N \to \infty} \frac{\binom{pN}{x}\binom{(1-p)N}{n-x}}{\binom{N}{n}} = \lim_{N \to \infty} \frac{\frac{(pN)!}{x!(pN-x)!} \cdot \frac{((1-p)N)!}{(n-x)!((1-p)N-(n-x))!}}{\frac{N!}{n!(N-n)!}}$

pull the constant
$\lim_{N \to \infty} a\, f(x) = a \lim_{N \to \infty} f(x)$

$= \frac{\frac{1}{x!(n-x)!}}{\frac{1}{n!}} \lim_{N \to \infty} \frac{\frac{(pN)!}{(pN-x)!} \cdot \frac{((1-p)N)!}{((1-p)N-(n-x))!}}{\frac{N!}{(N-n)!}}$

$\frac{7!}{(7-3)!} = \frac{7!}{4!} = \frac{7 \cdot 6 \cdot 5 \cdot \cancel{4!}}{\cancel{4!}} = 7 \cdot 6 \cdot 5$

$\frac{pN\,(pN-1)\,(pN-2)\dots(pN-(x-1))\,(pN-x)!}{(pN-x)!} = pN\,(pN-1)\,(pN-2)\dots(pN-(x-1))$ x

$\frac{(1-p)N\,((1-p)N-1)\,((1-p)N-2)\dots((1-p)N-(n-x-1))\,((1-p)N-(n-x))!}{((1-p)N-(n-x))!} = (1-p)N\,((1-p)N-1)\,((1-p)N-2)\dots((1-p)N-n+x+1)$ n-x

$\frac{N\,(N-1)\,(N-2)\dots(N-(n-1))\,(N-n)!}{(N-n)!} = N\,(N-1)\,(N-2)\dots(N-(n-1))$ n

$\frac{n!}{x!(n-x)!} = \binom{n}{x}$ x

$= \binom{n}{x} \lim_{N \to \infty} \frac{pN\,(pN-1)\,(pN-2)\cdots(pN-(x-1))}{N\cdot(N-1)\cdot(N-2)\cdots(N-(n-1))} \cdot \frac{((1-p)N)\,((1-p)N-1)\,((1-p)N-2)\cdots((1-p)N-n+x+1)}{}$ n-x

split the limit
$\lim_{N \to \infty} f(x)g(x) = \lim_{N \to \infty} f(x) \lim_{N \to \infty} g(x)$

$= \binom{n}{x} \underbrace{\lim_{N \to \infty} \frac{pN}{N}}_{p} \cdot \underbrace{\lim_{N \to \infty} \frac{pN-1}{N-1}}_{p} \cdot \underbrace{\lim_{N \to \infty} \frac{pN-2}{N-2}}_{p} \cdots \underbrace{\lim_{N \to \infty} \frac{pN-(x-1)}{N-(x-1)}}_{p} \cdot \underbrace{\lim_{N \to \infty} \frac{(1-p)N}{N-x}}_{1-p} \cdot \underbrace{\lim_{N \to \infty} \frac{(1-p)N-1}{N-(x+1)}}_{1-p} \cdots \underbrace{\lim_{N \to \infty} \frac{(1-p)N-(n-x-1)}{N-(n-1)}}_{1-p}$ x, n-x

pull the constants
$\lim_{N \to \infty} \frac{pN}{N} = p \lim_{N \to \infty} \frac{N}{N} = p \cdot 1 = p$
$\lim_{N \to \infty} \frac{(1-p)N}{N-x} = (1-p) \lim_{N \to \infty} \frac{N}{N-x} = (1-p) \cdot 1 = (1-p)$
$m \to x \approx \infty$

remember ?
$X \sim \text{Bernoulli}(p) := p^x(1-p)^{1-x}$ $= \binom{n}{x} p^x (1-p)^{n-x}$

Binomial distribution

In probability theory and statistics, the **binomial distribution** with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p. A success/failure experiment is also called a Bernoulli experiment or Bernoulli trial; when n = 1, the binomial distribution is a Bernoulli distribution. The binomial distribution is the basis for the popular binomial test of statistical significance.
The binomial distribution is frequently used to model the number of successes in a sample of size n drawn **with replacement** from a population of size N. If the sampling is carried out without replacement, the draws are not independent and so the resulting distribution is a hypergeometric distribution, not a binomial one. However, for N much larger than n, the binomial distribution remains a good approximation, and widely used.

huge N

probability of success
sample size
$X \sim \text{Binomial}(n,p) := \binom{n}{x} p^x(1-p)^{n-x}$ ← $X \sim \text{Hypergeometric}(n,p,N)$
# of successes

Parameter Space
amount that we can sample : n
$n \in \mathcal{N}$
$p \in (0,1)$

so $\text{Supp}[X] = \{ 0, 1, \dots, n \}$
each individual trial could be a failure or all coins from a capsized cup are Tails
each individual trial could be a success or all coins from a capsized cup are Heads

why not $p \in [0,1]$?
$p=0 \rightarrow X \sim \text{Binomial}(n,0) = \binom{n}{0} 0^x(1-0)^{n-x} \rightarrow P(X=0) = \binom{n}{0} 0^0\, 1^{1-0} = 1 \rightarrow \text{Deg}(0)$
$p=1 \rightarrow X \sim \text{Binomial}(n,1) = \binom{n}{x} 1^x(1-1)^{n-x} \rightarrow P(X=1) = \binom{n}{1} 1^1\, 0^{1-1} = n \rightarrow \text{Deg}(n)$

$0^0=1$ $1^{any\ power}=1$
$\binom{n}{0}$
$\binom{n}{n}$
$\binom{1}{0} = 1$
$\binom{1}{1} = 1$

$X \sim \text{Binomial}(1,p) = \binom{1}{x} p^x(1-p)^{1-x} = p^x(1-p)^{1-x} = X \sim \text{Bernoulli}(p)$
$\binom{1}{1} = 1$
$\binom{1}{0} = 1$
so $\text{Supp}[X] = \{ 0,1 \}$

Binomial distribution Proof

special case
$X \sim \text{Binomial}(1,p) = X \sim \text{Bernoulli}(p) = X \sim \text{HyperGeometric}(1,1,2)$

$X \sim \text{Binomial}(n,p) = X \sim \text{Bernoulli}(p) + X \sim \text{Bernoulli}(p) + \dots + X \sim \text{Bernoulli}(p)$ — n
all is the same

We have a cup with 1 coin. We draw the coin out. Record whether H or T and put the coin back in the cup. We repeat the experiment n times then see # of Hs.
Ex:

We have a cup with n coins. We draw them ALL out and record how many Hs.

static — static by default b/c replacement
$X \sim \text{HyperGeometric}(n,p,N) = X \sim \text{Binomial}(n,p)$

We are on a beach made of black and white grains of sand. We collect a handful of sand and examine how many are black. Beneath our feet there is still the entire beach of sand left. If we were to collect another handful of sand with the other hand, the ratio of black to white is still the same. The ratio of black to white can be viewed as static.

property of PMF: $\sum_{x \in \text{Supp}[X]} p(x) = 1$

verify $\sum_{x \in \{0,1,\dots,n\}} \binom{n}{x} p^x(1-p)^{n-x} = 1$ if the Binomial it truly a r.v. we must prove = 1 how ?
Binomial Supp[X] = { 0,1, ... , n } ⟶ x∈{0,...,n}

recall the **Binomial Theorem** $(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}$ let a=p, b=1−p, i=x
reason why we call this distribution 'binomial'

plug in
$(p+(1-p))^n =$
$(p-p+1)^n =$
$(1)^n =$
$1 =$
$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x}$
QED