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)

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

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

e.g. return 1(true) of the outcome is H — indicator function returns bool

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

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

let's map H to 1 and T to 0

once represented numerically can I take the average ?

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

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

X: function as a black box; input: abstract; output: number

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

$X: \Omega \rightarrow \mathcal{R}$ (difficult to compute: sample space → easy to compute: value of r.v.)

X is r.v. function s/t if its input/domain is Ω (sample space), then its output/range are Real Numbers called **realizations**

what is the probability that X=1 $P(X=1)$? — same as → Def: $P: 2^{\Omega} \rightarrow [0,1]$

unrealized r.v. before the experiment takes place: 'what is the prob. of getting H?'

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

Support of X

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

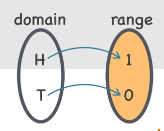
$\textbf{Supp[X]} := \{x: P(X=x) > 0\} \subseteq \mathcal{R}$

x : an arbitrary value in the range of the function X

cashing in on the previous 'abuse of notation'

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 Supp[X] = { 1,0 }



Discrete X

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

$\sum_{x \in Supp[X]} P(X=x) = 1$

recall that: $P(\{H\} \cup \{T\}) = 1$ OR simply $P(\Omega) = 1$

we are adding the probabilities of all of the values in the range/support of X

thus in the coin flip experiment |Supp[X]| = 2

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

if an outcome ω is not in the 'support of X', then its probability of occurrence must be 0

if not $\exists\omega$ s/t $X(\omega) \notin Supp[X] \rightarrow P(\{\omega\}) = 0 \rightarrow$ collectively exhaustive

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

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

$P(\Omega) = P(\{\omega: X(\omega)=x_1\}) + P(\{\omega: X(\omega)=x_2\}) + \ldots$

$1 = \sum_{x \in Supp[X]} P(X=x)$ 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.

Deg is a r.v. that has only one value

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

Power Set (by definition of a non-empty set) must also include ∅ 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.

$2^{\Omega} = \{\varnothing, \{H\}, \{T\}, \{H,T\}\}$ — p=0, p=1/2, p=1/2, p=1

$p \in (0,1)$

MNEMONIC: X~Bernoulli( ) — 1,0

Bernoulli

Are these two experiments the same ?

coin flip: H, T → 1, 0 — spinner: Red, Green → 1, 0

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

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

$X \sim \begin{cases} 1 & wp \ 1/2 \\ 0 & wp \ 1/2 \end{cases}$ (~ distributed as; wp with probability)

'random variable' is distributed 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. (think a generic function)

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

or more generally — Supp[X] = { 1,0 } (remember! if Supp is NOT 0 and 1, NOT a Bernoulli)

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

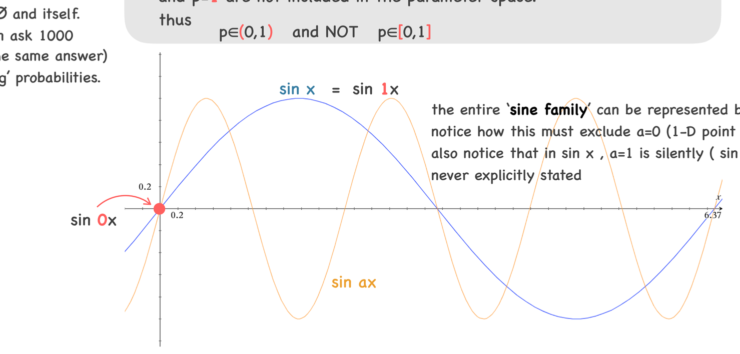
X is distributed Bernoulli with '**parameter p**' (p is a probability). A parameter is a choice which defines a model. Space in which p lives is called the **parameter space** (in parallel to event space).

what are the possible values of p ?

$p \in [0,1]$ (includes 0&1)
what if p=1? (probability of everything = 1) $X = \{1 \ wp \ 1$ 1,1,1,1,1,1 (there is 100% chance of something happening)
what if p=0? (probability of nothing = 0) $X = \{0 \ wp \ 0$ 0,0,0,0,0,0,0 (there is 0% chance of something happening)
→ degenerate X, Deg

if $X \sim Deg(c) := \{c \ wp \ 1$, 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 a trivial sine curve 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 $p \in [0,1]$

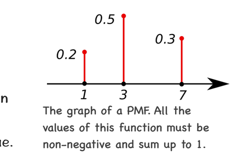


the entire 'sine family' can be represented by 'sin ax'
notice how this must exclude a=0 (1-D point is not a curve)
also notice that in sin x , a=1 is silently (sin 1x) included and never explicitly stated

- 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

PMF

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.

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

$p: \mathcal{R} \rightarrow [0,1]$ defined everywhere
if $x \in Supp[X]$, $p(x) > 0$
if $x \notin Supp[X]$, $p(x) = 0$
$\sum_{x \in Supp[X]} p(x) = 1$ (as previously proved) — the probabilities of all the values in the support of X add up to 1

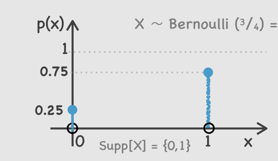
cashing in on the previous 'abuse of notation'; prob. that an unrealized r.v. X will be equal to a value in its support

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

it is useful to plot PMF

$X \sim Bernoulli(0.75)$

$X \sim Bernoulli(3/4) = \begin{cases} 1 & wp \ 3/4 \\ 0 & wp \ 1/4 \end{cases}$

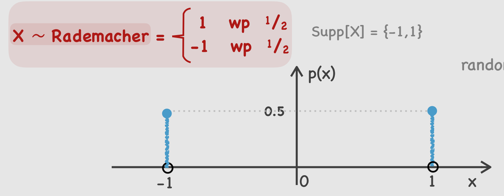


p(x) takes the x_s from the Supp[X] and maps them to their probabilities

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

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

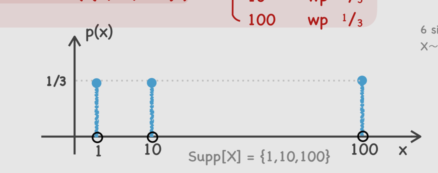
random walk



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.

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

6 sided dice is: X~Uniform({1,2,3,4,5,6}) = 1 wp 1/6, 2 wp 1/6, 3 wp 1/6, 4 wp 1/6, 5 wp 1/6, 6 wp 1/6

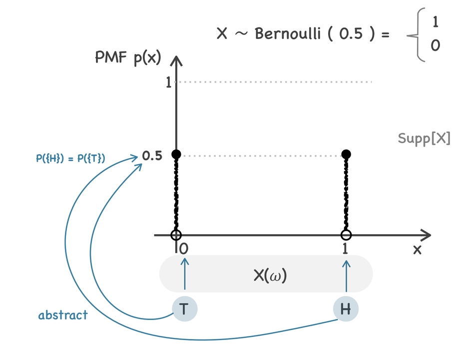


X ~ Uniform (A) Supp[X]=A (set A contains all of the members of the support of X)
Parameter Space: $A \subset \mathcal{R}$, $|A| \in \mathcal{N} \setminus \{1\}$

you need at least 2 values in the support of X — if |A|=1, Deg! if A contains only 1 member, degenerate. e.g. X~Uniform({5}) is Deg

cardinality of A is the set of all Natural numbers except 1

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



Supp[X] = {0,1} {H,T} =: Ω

CDF

Cumulative Distribution Function (CDF) is a function whose value is the probability that a corresponding 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.

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

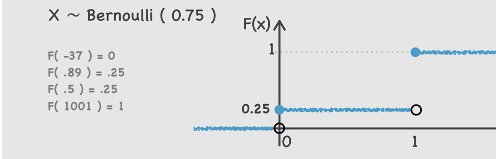
prob. that an unrealized r.v. X will be less than or equal to a value in its support

properties (derivable form the above definition)

1. $F(x) \in [0,1]$ — monotonically increasing (Calc-1 slope ≥ 0)
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

X ~ Bernoulli (0.75)

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



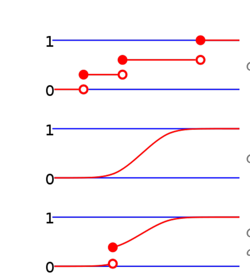
# of discontinuities = Supp[X]

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 all the way to the 'plank time'. Using CDF function you can now express your data.

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



CDF of a discrete probability distribution,

CDF of continuous probability distribution,

CDF a distribution which has both a continuous part and a discrete part.

Bernoulli PMF

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

piecewise function (ugly difficult)

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

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

Def
X_1 and X_2 are "**equal in distribution**" $X_1 \overset{d}{=} X_2$
if $p_1(x) = p_2(x)$ or $F_1(x) = F_2(x)$
X_1 & X_2 have the same PMF p(x) or CDF F(x)

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

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

end-game scoop

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

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

$X \sim Binomial(n,p) = X \sim Bernoulli(p), X \sim Bernoulli(p), \ldots, X \sim Bernoulli(p)$ — n (all p_i the same) $X_1,\ldots,X_n \overset{iid}{\sim} Bernoulli(p)$

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

$X \sim HyperGeometric(n,p,N) = X \sim Binomial(n,p)$ (static; N → ∞; static by default b/c replacement)

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.

9

Hypergeometric PMF

Drawing cards without replacement. 10 Cards, 4 Red, 6 Blue

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

$P(A) = \frac{|A|}{|\Omega|}$ working def for prob. of A

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

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

10 Cards, K Red, 10-K Blue

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

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.

Hypergeometric only 2 types e.g. {Tails,Heads} {Red,Blue} {True,False} — MNEMONIC — if more than 2 types, NOT Hypergeometric (something that denotes a success and something that denotes a failure)

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(X=success)

$X \sim Hypergeometric(n,K,N) := p(K) :=$

N=7, K=3, n - up to you

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, n=23 — $X \sim Hypergeometric(23,37,118)$

$P(X=16) = \frac{\binom{37}{16}\binom{81}{7}}{\binom{118}{23}}$ — 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, n=8 — $X \sim Hypergeometric(8,53,100)$

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

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 ? (bag with 1 ball in it)
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) — X~Deg(0) or (1)
→ Degenerate Distribution

N=2 ? K=0 , 1 or 2 (not interesting)
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 → $X \sim Hyper(1,1,2) = \frac{\binom{1}{x}\binom{1}{1-x}}{\binom{2}{1}}$

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

N=3 ? K=1,2 (0,3→Deg), n=1,2 (n=3 ? → no mystery)

$N \in \mathcal{N} \setminus \{1\}$ (1 ball - NOT interesting - start with at least 2)
$K \in \{1,2,\ldots,N-1\}$ (having all successes - NOT interesting; make successes < total size)
$n \in \{1,2,\ldots,N-1\}$
Parameter Space

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

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

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

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

4 cases of X~Hypergeometric(n,K,N) (draws, success, total)

1	X~Hypergeometric(2,4,10)	Supp[X] = {0,1,2}	- 0: both draws can be failures - 2: both draws can be successes — 2 draws
2	X~Hypergeometric(5,4,10)	Supp[X] = {0,1,2,3,4}	- 4: at most 4 possible successes - 0: all 5 draws can be failures — 5 draws
3	X~Hypergeometric(8,4,10)	Supp[X] = {2,3,4}	- 4: at most 4 possible successes - 2: at most 6 possible failures — 8 draws
4	X~Hypergeometric(5,7,10)	Supp[X] = {2,3,4,5}	- 2: at most 3 possible failures - 5: can't have 6 successes in 5 draws — 5 draws

General Model
X~Hypergeometric(n,K,N) Supp[X] ?

1	n < K , n < N-K	Supp[X] = {0,1,...,n}	- choose less then success & less then failure
2	n ≥ K , n < N-K	Supp[X] = {0,1,...,K}	- choose more/equal then/to success & less then failure
3	n ≥ K , n ≥ N-K	Supp[X] = {n-(N-K),...,K}	- choose more/equal then/to success & more/equal then/to failure
4	n < K , n ≥ N-K	Supp[X] = {n-(N-K),...,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}

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

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

use Vandermonde's Identity to prove this

```
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

self-note:
10 balls total
3 success balls
7 failure balls

- 1st draw: the probability of picking a successful ball is 3/7 . Note that 3/7 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 2/7 or 3/6
- 3rd draw: 1/7 or 2/6 or 3/5
- 4th draw: 0/7 or 1/6 or 2/5 or 3/4
- 5th draw: 0/6 or 1/5 or 2/4 or 3/3
- 6th draw: 0/5 or 1/4 or 2/3 or 3/2
- 7th draw: 0/4 or 1/3 or 2/2 or 3/1
- 8th draw: 0/3 or 1/2 or 2/1 or 3/0
- 9th draw: 0/2 or 1/1 or 2/0
- 10th draw: 0/1 or 1/0

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

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

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

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

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

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

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

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

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

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

$\frac{pN(pN-1)(pN-2)\ldots(pN-(x-1))(pN-x)!}{(pN-x)!} = pN(pN-1)(pN-2)\ldots(pN-(x-1))$ — x

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

$\frac{N(N-1)(N-2)\ldots(N-(n-1))(N-n)!}{(N-n)!} = N(N-1)(N-2)\ldots(N-(n-1))$ — n

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

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

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

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

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

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

$= \binom{n}{x} p^x (1-p)^{n-x}$

Binomial distribution PMF

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

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

so $Supp[X] = \{0,1,\ldots,n\}$
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 $p \in [0,1]$?
$p=0 \rightarrow X \sim Binomial(n,0) = \binom{n}{x} 0^x(1-0)^{n-x} \rightarrow P(X=0) = \binom{n}{0} 0^0 1^{1-0} = 1 \rightarrow Deg(0)$
$p=1 \rightarrow X \sim Binomial(n,1) = \binom{n}{x} 1^x(1-1)^{n-x} \rightarrow P(X=1) = \binom{n}{1} 1^1 0^{1-1} = n \rightarrow Deg(n)$
$0^0=1$, $1^{\text{any power}}=1$

$\binom{1}{0} = 1$
$\binom{1}{1} = 1$

$X \sim Binomial(1,p) = \binom{1}{x} p^x(1-p)^{1-x} = p^x(1-p)^{1-x} = X \sim Bernoulli(p)$
$\binom{1}{1} = 1$, $\binom{1}{0} = 1$

so Supp[X] = { 0,1 }

In probability theory and statistics, the **binomial distribution** with parameters n and p is the probability mass function (PMF) 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 →

Binomial distribution Proof

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

verify $\sum_{x \in [0,\ldots,n]} \binom{n}{x} p^x(1-p)^{n-x} = 1$ — if the Binomial is truly a r.v. we must prove = 1 how ?

Binomial Supp[X] = { 0,1, ... , 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
$(p+(1-p))^n = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x}$
$(p-p+1)^n =$
$(1)^n =$
$1 =$ QED