Probability Theory

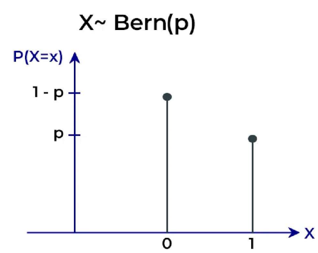
Textbook: *Introduction to Probability* by Blitzstein and Hwang

Previous Lecture

- Random Vars (rvs)
- Discrete Rvs
- Prob Mass Functions (PFs)
- Valid PFs

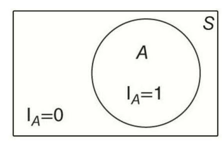
§3.3 - Bernoulli and Binomial

Extremely Common/Useful Distributions

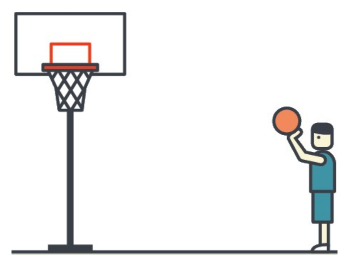


Def (**Bernoulli Distr**): Rv X has Bernoulli distr w/parameter p if $P(X = 1) = p$ and $P(X = 0) = 1 - p$, where $0 < p < 1$. We write $X \sim Bern(p)$.
Symbol $\sim$ is read: "is distributed as." [describes distr]

Def (**Indicator Rv**): The indicator rv of event A (I_A or $I(A)$) equals 1 if A occurs, and 0 otherwise. Note that $I_A \sim Bern(p)$ with $p = P(A)$. [generic indicator rv]

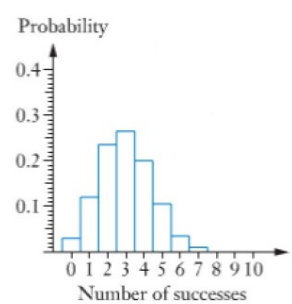


Story (**Bernoulli Trial**): An experiment that can result in either "success" event or "failure" (but not both) is a **Bernoulli trial**. A **Bernoulli Rv** X is an indicator of success in a Bernoulli trial: it equals 1 for success and 0 for failure.
[indicator rv specific to a Bernouli Trial]



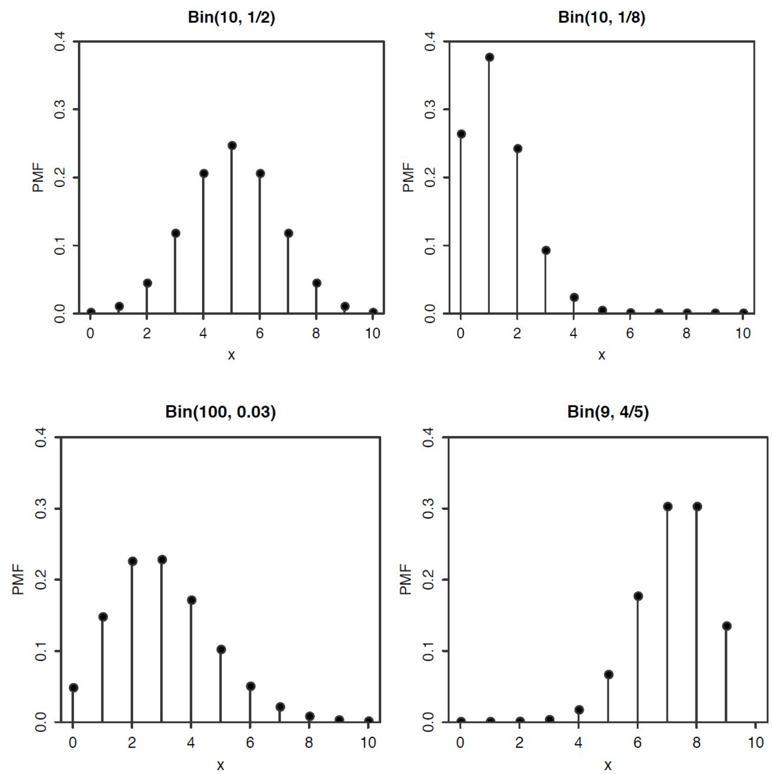
Making a free throw is a Bernoulli trial

Story (**Binomial Distr**): Suppose that n indep Bernoulli trials are performed, each with the same *success prob* p. Let rv N be # of successes. The distr of N is called the Binomial distr w/parameters n and p. We write $N \sim Bin(n,p)$ to mean N has Binomial distr w/params n and p, where n is a positive integer and $0 < p < 1$.



Thm (**Binomial PF**): If $N \sim Bin(n,p)$, then the PF of N is $P(N = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, \dots, n$ (and $P(N = k) = 0$ otherwise). Here we see why it's call Binomial.

Proof. An experiment consisting of n indep Bernoulli trials produces a sequence of successes and failures. The prob of any specific sequence of k successes and $n - k$ failures is $p^k(1-p)^{n-k}$. There are $\binom{n}{k}$ such sequences, since we just need to select where the successes are. So, letting X be # of successes, $P(X = k) = p^k(1-p)^{n-k}$ for $k = 0, 1, \ldots, n$ and $P(X = k) = 0$ otherwise. This is a valid PF because it is nonnegative and it sums to 1 by the binomial thm.



Ex (Binomial Basketball): On a basketball court, you throw 10 free throws. Assume you make 8 of every 16 free throws you attempt. Let X be the number of free throws you make.

a) What is the name of the distribution of X?

X is distributed as a Binomial, with $n = 10$ and $p = 8/16$.

b) What is the PF $f(x)$ of X?

$$f(x) = \begin{cases} \binom{10}{x}\left(\frac{8}{16}\right)^x\left(1 - \frac{8}{16}\right)^{10-x} & \text{if } x = 0, 1, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

c) What is the prob of you making 6 free throws?

$$f(6) = \binom{10}{6}\left(\frac{8}{16}\right)^6\left(1 - \frac{8}{16}\right)^4 \approx 0.2466.$$

Thm (Failure Bernoulli): Let $X \sim Bin(n,p)$, and $q = 1 - p$ (q denotes failure prob of a Bernoulli trial). Then $n - X \sim Bin(n,q)$.

Proof. Using the story of Binomial, interpret X as # of successes in n indep Bernoulli trials. Then $n - X$ is # of failures in those trials. Interchanging roles of success and failure, we have $n - X \sim Bin(n,q)$.

Alternatively, we can check that $n - X$ has a $Bin(n,q)$ PF. Let $Y = n - X$. The PF of Y is

$P(Y = k) = P(X = n - k) = \binom{n}{n-k} p^{n-k} q^k = \binom{n}{k} q^k p^{n-k}$; for $k = 0, 1, \ldots, n$. ■

Corollary. Let $X \sim Bin(n,p)$ w/$p = \frac{1}{2}$ and n even. Then the distr of X is symmetric about $n = 2$, in the sense that $P(X = \frac{n}{2+j}) = P(X = \frac{n}{2-j})$ for all nonnegative integers j.

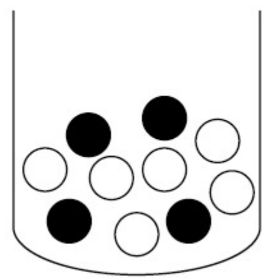
Proof. By Failure Bernoulli Thm, $n - X$ is also $Bin(n, \frac{1}{2})$, so $P(X = k) = P(n - X = k) = P(X = n - k)$ for all nonnegative integers k. Letting $k = n = 2 + j$, the desired result follows.
This explains why the $Bin(10, \frac{1}{2})$ PF is symmetric about 5 in the figure above.

Ex (**Coin Tosses Cont'd**). Going back to the Coin Toss example in the §3.1 lecture, we now know that $X \sim Bin(2, \frac{1}{2})$, $Y \sim Bin(2, \frac{1}{2})$, and $I \sim Bern(\frac{1}{2})$. Consistent w/Failure Bernoulli thm, X and $Y = 2 - X$ have the same distr, and consistent w/the above corollary, the distr of X (and of Y) is symmetric about 1.

Harvard Video (middle part): youtube.com/watch?v=LX2q356N2rU&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=9

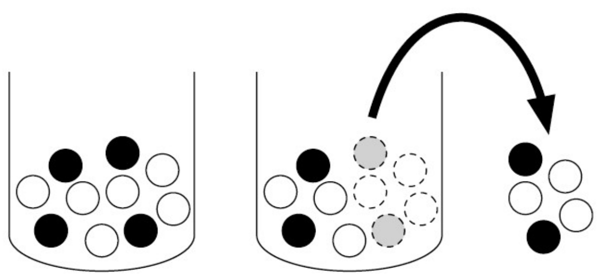
---

§3.4 - Hypergeometric



Let's say we have an urn filled with w white and b black balls, and we want to know how many times white balls are drawn when drawing n balls w/replacement.

From §3.3, we have: $X \sim Bin(n, \frac{w}{w+b})$, where X is # of white balls.
But what if we drew **without** replacement?



Story (**Hypergeometric Distr**). Consider an urn with w white balls and b black. We draw n balls at random without replacement, such that all $\binom{w+b}{n}$ samples are equally likely. Let X be # of white balls in the sample. Then X is a **Hypergeometric distr** w/parameters w, b, and n. We denote this by $X \sim HGeom(w, b, n)$.

Thm (**Hypergeometric PF**). If $X \sim HGeom(w, b, n)$, then the PF of X is $P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$, for integers k satisfying $0 \le k \le w$ and $0 \le n - k \le b$, and $P(X = k) = 0$ otherwise.

Proof. To get $P(X = k)$, we first count the # of possible ways to draw exactly k white balls and $n - k$ black balls (without distinguishing between different orderings for getting the same set of balls).

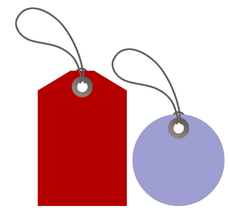
If $k > w$ or $n - k > b$, then the draw is impossible. Otherwise, there are $\binom{w}{k}\binom{b}{n-k}$ ways to draw k white and $n - k$ black balls (multiplication rule), and $\binom{w+b}{n}$ total ways to draw n balls.

Since all samples are equally likely, the naive definition of probability gives $P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$ for integers k satisfying $0 \leq k \leq w$ and $0 \leq n - k \leq b$.

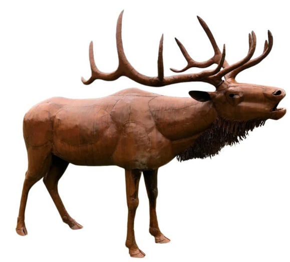
Side Note: $\sum_{k=1}^{w} \binom{w}{k}\binom{b}{n-k} = \binom{w+b}{n}$ by "Vandermonde's Identity" (see example 1.5.3 in book).

So this PF is valid because it sums to 1. ■

Hypergeometric distr's describe the # of items in a population that have two sets of tags. In the urn example, it was "white" and "sampled" balls. In the next example, it is "captured", and "recaptured" elk.



Ex (**Elk Capture-Recapture**): A forest has 400 elk. Yesterday, 40 elk were captured, tagged, and released back into the forest. Today, 50 are captured at random. Assume the 50 newly captured are equally likely to be any of the original 400. What's the prob there are two tagged elk in the newly captured elk?



Solution: By Hypergeometric distr, # of tagged elk in the newly captured is $HGeom(w, b, n)$ or $HGeom(40, 360, 50)$.

The 40 tagged elk correspond to the white balls and the 360 untagged elk correspond to the black balls. Instead of sampling n balls from the urn, we capture 50 elk from the forest. So, $P(X = 2) =$ ...

$$P(X = 2) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}} = \frac{\binom{40}{2}\binom{360}{50-2}}{\binom{40+360}{50}} = 0.07. \qquad \text{(7\% chance)}$$

Thm (**Hypergeometric Equality**). The $HGeom(w, b, n)$ and $HGeom(n, w + b - n, w)$ distr's are identical. That is, if $X \sim HGeom(w, b, n)$ and $Y \sim HGeom(n, w + b - n, w)$, then X and Y have the same distr.

Proof. Using the story of the Hypergeometric, imagine an urn with w white balls, b black balls, and a sample size n made without replacement.

Let $X \sim HGeom(w, b, n)$ be the # of white balls in the sample, thinking of white/black as the first set of tags and sampled/not sampled as the second set of tags.

Let $Y \sim HGeom(n, w + b - n, w)$ be the # of sampled balls among the white balls, thinking of sampled/not sampled as the first set of tags and white/black as the second set of tags.

Both X and Y count the # of white sampled balls, so they have the same distr. Alternatively, we can check algebraically that X and Y have the same PF:

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}} = \frac{w!b!n!(w+b-n)!}{k!(w+b)!(w-k)!(n-k)!(b-n+k)!},$$

$$P(Y = k) = \frac{\binom{n}{k}\binom{w+b-n}{w-k}}{\binom{w+b}{w}} = \frac{w!b!n!(w+b-n)!}{k!(w+b)!(w-k)!(n-k)!(b-n+k)!}. \quad \blacksquare$$

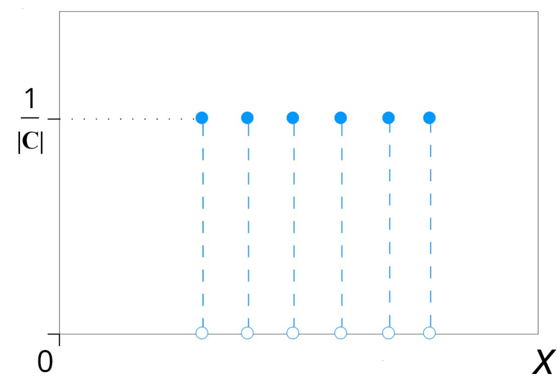
We prefer the story proof because it is less tedious and more memorable.

(**Binomial vs. Hypergeometric**). The Binomial and Hypergeometric distr are often confused. Both are discrete w/integer values between 0 and n for some n. Both can be interpreted as # of successes in n Bernoulli trials (for Hypergeometric, each tagged elk in the second captured sample can be considered a success and each untagged a failure).

However, a crucial part of the Binomial story is that the Bernoulli trials involved are *indep*.
In the Hypergeometric story they are *dependent*, since sampling is done without replacement.
(knowing that one elk in our sample is tagged decreases the prob that the second elk will also be tagged.)

---

§3.5 - Discrete Uniform



PMF of a discrete uniform rv.

$P(X = x) = \frac{1}{6}$

Story (**Discrete Uniform Distr**): Let C be a finite set of numbers. Choose a # uniformly at random (values in C are chosen equally likely). Call the chosen number X. Then X has the discrete uniform distr w/parameter C. We denote this: $X \sim DUnif(C)$. What's the PF?

The PF of $X \sim DUnif(C)$ is $P(X = x) = \frac{1}{|C|}$ for $x \in C$ (0 otherwise), since a PF must sum to 1.

(for any set T, $|T|$ denotes the # of elements in T)

For $A \subseteq C$, we have $P(X \in A) = \frac{|A|}{|C|}$. (If $A = \{2,4,6\}$ - evens, then $P(X \in A) = \frac{|A|}{|C|} = \frac{3}{6}$)

Ex (**Random Slips of Paper**): There are 100 slips of paper in a hat, each of which has one of $1, 2, \ldots, 100$ written on it. Five slips are drawn, one at a time.

First consider random sampling **with** replacement (equal probs).

(a) What's the distr of the # of drawn slips having a value of at least 80?

Solution: By Binomial, the distr is $Bin(5, 0.21)$.

$P(N = k) = \binom{5}{k}(0.21)^k(1-0.21)^{5-k}$, for $1 \leq k \leq 5$ (0 otherwise)

(b) What's the distr of the value of the jth draw (for $1 \leq j \leq 5$)?

Solution: Let X_j be the value of the jth draw. By symmetry, $X_j \sim DUnif(\{1, 2, \ldots, 100\})$.
There aren't certain slips that love being chosen on the jth draw and others that avoid being chosen then; all are equally likely ($P(X_j = k) = \frac{1}{100}$).

(c) What's the prob that 100 is drawn at least once?

Solution: $P(X_j = 100 \text{ for at least one } j) = 1 - P(X_1 \neq 100, \ldots, X_5 \neq 100)$.
Since the $X_i = 100$ are independent: $1 - P(X_1 \neq 100) \ldots P(X_5 \neq 100) = 1 - (\frac{99}{100})^5 \approx 0.049$.

Now consider random sampling **without** replacement (all sets of five slips equally likely to be chosen).

(d) What's the distr of the # of drawn slips having a value of at least 80 written on them?

Solution: By Hypergeometric, distr is $HGeom(21, 79, 5)$.

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}} = \frac{\binom{21}{k}\binom{79}{5-k}}{\binom{21+79}{50}}.$$

(e) What's the distr of the value of the jth draw (for $1 \leq j \leq 5$)?

Solution: Let Y_j be the value of the jth draw. By symmetry, $Y_j \sim DUnif(1, 2, \ldots, 100)$.

Learning any Y_i gives info about the other values ($Y_1, \ldots, Y_5$ are not indep),

but symmetry holds since, unconditionally, the jth slip drawn is equally likely to be any slip. This is the

unconditional distr of Y_j: we are working from a vantage point **before drawing any of the slips**.

Imagine that instead of one person drawing five slips, one at a time. Five people who simultaneously draw one slip each, with all possibilities equally likely for who gets which slip.

(f) What's the prob that 100 is drawn in the sample?

Solution: The events $Y_1 = 100, \ldots, Y_5 = 100$ are disjoint since we're now sampling without replacement,

so $P(Y_j = 100 \text{ for some } j) = P(Y_1 = 100) + \ldots + P(Y_5 = 100) = \frac{5}{100}$.

Activity 6

What did we learn?

- Bernoulli Distr
- Binomial PF
- Hypergeometric Distr/PF
- Discrete Uniform Distr