Probability Theory

Textbook: *Introduction to Probability* by Blitzstein and Hwang

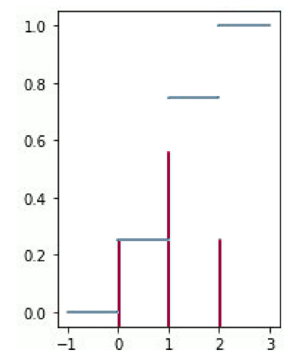
Previous Lecture

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

t

[Last time I taught, these notes ran a bit long, didn't have time for the activity]

§3.6 - Cumulative Distribution Functions (CDFs)



PF in red, CDF in blue

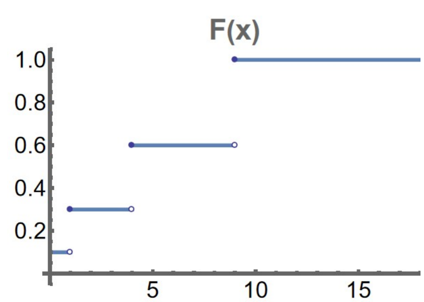
What's the prob that X is less than some value x?

Def (Cumulative Distr Function, CDF): A CDF of X, denoted F_X is given by $F_X(x) := P(X \leq x)$. Or just $F(x) := P(X \leq x)$.

Ex (CDF): Let X have the following PF: $f(x) = \begin{cases} \frac{\sqrt{x}}{10} & \text{for } x = 1, 4, 9, 16 \\ 0 & \text{otherwise.} \end{cases}$

Find the CDF $F(x)$ of X. And plot it.

$$F(x) = \begin{cases} 0.1 & x < 1 \\ 0.3 & 1 \leq x < 4 \\ 0.6 & 4 \leq x < 9 \\ 1 & 9 \leq x < 16 \end{cases}$$



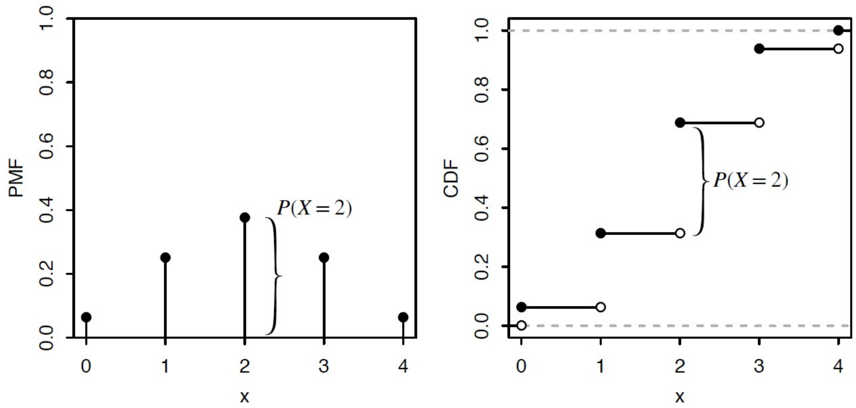
From PF → CDF: To find $P(X \leq 4.5)$, which is the CDF evaluated at 4.5, we sum the PF over all values of support that are less than or equal to 4.5. So, $P(X \leq 4.5) = P(X = 1) + P(X = 4) = \frac{\sqrt{1}}{10} + \frac{\sqrt{4}}{10} = \frac{3}{10}$.

More generally, the value of the CDF at an arbitrary point x (so, $P(X \leq x)$) is the sum of the heights of the vertical bars of the PF at values less than or equal to x.

From CDF → PF: The height of a jump in the CDF at x is equal to the value of PF at x.

In the plot below, the height of the jump at 2 is same as height of corresponding vertical bar in the PF. Flat regions of CDF correspond to values outside the support, so PF is equal to 0 there.

Ex:



$X \sim Bin(4, \frac{1}{2})$

Thm (**Valid CDFs**): Any CDF F is:

- ♦ **Increasing**: If $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$. (since probs are positive)

- ♦ **Right-Continuous**: As in above figure, CDF is continuous except possibly having some jumps.
  At jumps, the CDF is continuous from the right: for any a, we have: $F(a) = \lim_{x \to a^+} F(x)$.

- ♦ Convergence to 0 and 1 in the limits: $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

[Proofs in Book]

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

§3.7 - Functions of Random Variables

What if we add two rvs $X + Y$?

Just add the result! How about X^2 , or $\frac{\sqrt{X}}{\ln Y^2}$? And are the result rvs ?!?

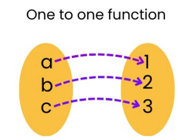
Yes! A function $f(X)$ of a rv X **IS** a rv. For example: X^2, e^X, $\sin(X)$, etc.

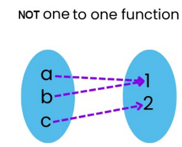
Def (**Function of a Rv**): For an experiment w/sample space S, a rv X, and $g : \mathbb{R} \to \mathbb{R}$, we define $Y := g(X)$. Y is a rv that maps s to $g(X(s))$ for all $s \in S$.

This is a composition of functions. We are saying, "first apply $X(s)$, then apply $g(x)$."

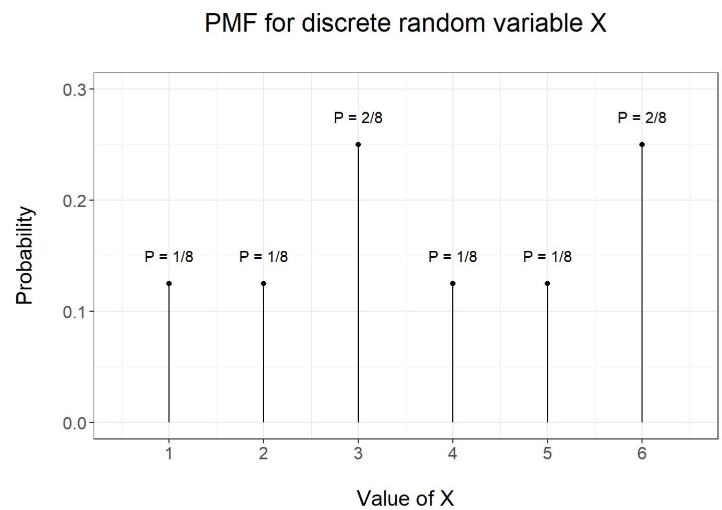
If we know the PF for X, can we find the PF for $Y = g(X)$? (yes, let's see how)

If g is a **one-to-one** function, the support of Y is the set of all $g(x)$ with x in the support of X.





The graph below depicts the PFs for two rvs. For the first, X has support $\{1,2,3\}$. For the second, $g(X) = X+3$, has support $\{4,5,6\}$.



Ex (One to One): Define X w/PF: $f_X(x) = \begin{cases} \frac{x}{6} & \text{for } x = 1,2,3 \\ 0 & \text{otherwise.} \end{cases}$

Let $Y = 3X$ (one to one!), what is the PF of Y?

Solution: Notice our Y values are $\{3,6,9\}$.

So $P(Y=3) = P(3X=3) = P(X=1) = \frac{1}{6}$,

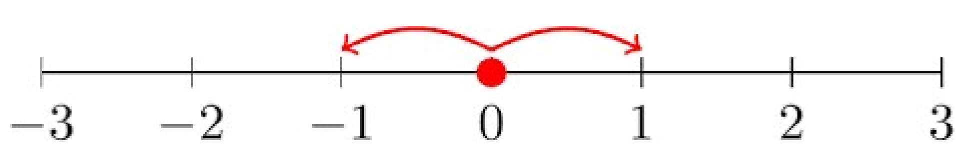
$P(Y=6) = P(X=2) = \frac{2}{6}$, and $P(Y=9) = P(X=3) = \frac{3}{6}$.

So, $f_Y(y) = \begin{cases} \frac{y}{3\cdot 6} & \text{for } y = 3,6,9 \\ 0 & \text{otherwise.} \end{cases}$

Sanity check: $\frac{3}{18} + \frac{6}{18} + \frac{9}{18} = 1$. $\square$

Strategy: To find the PF of a rv w/unfamiliar distr: express the rv as a one-to-one function of a known distr.

Ex (Random Walk). A particle moves n steps on a number line. It starts at 0, and at each step it moves 1 to the right or left, with equal prob's. Assume all steps are indep. Let Y be it's position after n steps. Find the PF of Y.



Solution: Consider each step to be a Bernoulli trial, where right is a "success" and left a "failure."

The # of steps the particle takes to the right is a $Bin(n, \frac{1}{2})$ rv, which we name X.

If $X = j$, then we have j steps to the right and $n - j$ to the left, giving a final position: $j - (n - j) = 2j - n$.

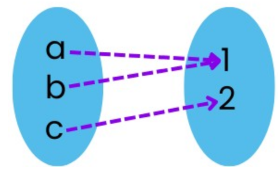
So we can express Y as a one-to-one function of X: $Y = 2X - n$.
Since X takes values in $\{0, 1, 2, \ldots, n\}$, Y takes values in $\{-n, 2-n, 4-n, \ldots, n\}$.

The PF of Y can then be found from the PF of X: $P(Y = k) = P(2X - n = k) = P(X = \frac{n+k}{2}) = \binom{n}{\frac{n+k}{2}}\left(\frac{1}{2}\right)^n$
if k is an integer between $-n$ and n (inclusive) such that $n + k$ is an even number. ■

If g is NOT one-to-one?

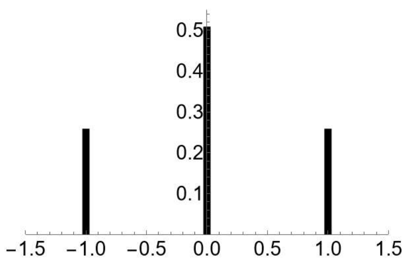
Then, if $g(x_1) = y$, there may be another x_2 such that $g(x_2) = y$ (!!)

$S \to X(s) \in \{x_1, x_2, x_3\}$. Then:

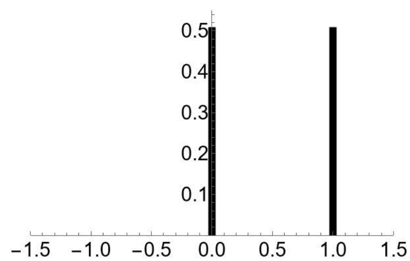


$\{x_1, x_2, x_3\} \to g(x_i)$

Thm (**PF of** $g(X)$). Let X be discrete and $g : \mathbb{R} \to \mathbb{R}$. Then the support of $g(X)$ is the set of all y such that $g(x) = y$ for at least one x in the support of X. The PF of $g(X)$ is:
$P(g(X) = y) = \sum_{x:g(x)=y} P(X = x)$, for all y in the support of $g(X)$.



$\to \quad g(X) = X^2 \quad \to$



$P(X = 0) = \frac{1}{2}$, $P(X = \pm 1) = \frac{1}{4}$

$P(g(X) = 0) = \frac{1}{2}$, $P(g(X) = 1) = \frac{1}{2}$

Ex (**Not One to One**): Define X w/PF: $f_X(x) = \begin{cases} \frac{|x|}{32} & \text{for } x = -10, -5, 1, 0, 1, 5, 10 \\ 0 & \text{otherwise.} \end{cases}$

Let $Y = X^2$, what is the PF of Y?

Solution: Notice our Y values are $\{100, 25, 1, 0\}$ with 100, 25 and 1 getting double probabilities.

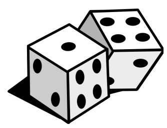
So $P(Y = 100) = 2 \cdot \frac{|10|}{32} = \frac{10}{16}$, $P(Y = 25) = 2 \cdot \frac{|5|}{32} = \frac{5}{16}$, and $P(Y = 1) = 2 \cdot \frac{|1|}{32} = \frac{1}{16}$

So, $f_Y(y) = \begin{cases} \frac{\sqrt{y}}{16} & \text{for } y = 100, 25, 1, 0 \\ 0 & \text{otherwise.} \end{cases}$

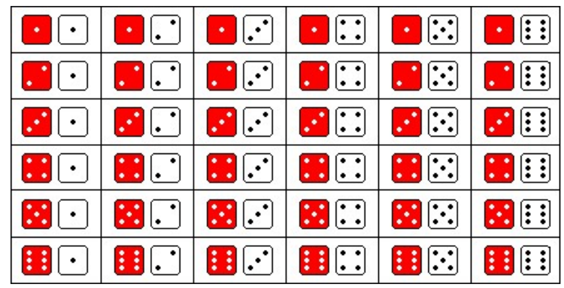
Sanity check: $\frac{\sqrt{100}}{16} + \frac{\sqrt{25}}{16} + \frac{\sqrt{1}}{16} + \frac{\sqrt{0}}{16} = 1.$ $\square$

Def (**Function of Two Rvs**). Given an experiment with sample space S, if X and Y map $s \in S$ to $X(s)$ and $Y(s)$ respectively, then $g(X, Y)$ is the rv that maps s to $g(X(s), Y(s))$.

Ex (**Maximum of Two Die Rolls**).



s	X	Y	$\max(X,Y)$
$(1,2)$	1	2	2
$(1,6)$	1	6	6
$(2,5)$	2	5	5
$(3,1)$	3	1	3
$(4,3)$	4	3	4
$(5,4)$	5	4	5
$(6,6)$	6	6	6



$P(\max(X, Y) = 1) = \frac{1}{36}.$

$P(\max(X, Y) = 2) = \sum_{s:\max(s)=2} P(s)$

$= P(X = 2, Y = 2) + P(X = 1, Y = 2) + P(X = 2, Y = 1)$

$= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{3}{36}.$

$P(\max(X, Y) = 3) = \frac{5}{36}.$

$P(\max(X, Y) = 4) = \frac{7}{36}.$

$P(\max(X, Y) = 5) = \frac{9}{36}.$

$P(\max(X, Y) = 6) = \frac{11}{36}.$

Note: $P(\max(X, Y) = 5) = P(X = 5, Y \leq 4) + P(X \leq 4, Y = 5) + P(X = 5, Y = 5)$

$= 2P(X = 5, Y \leq 4) + \frac{1}{36}$ (symmetry)

$= 2\left(\frac{4}{36}\right) + \frac{1}{36} = \frac{9}{36}.$

Common error: to confuse a rv w/its distr.

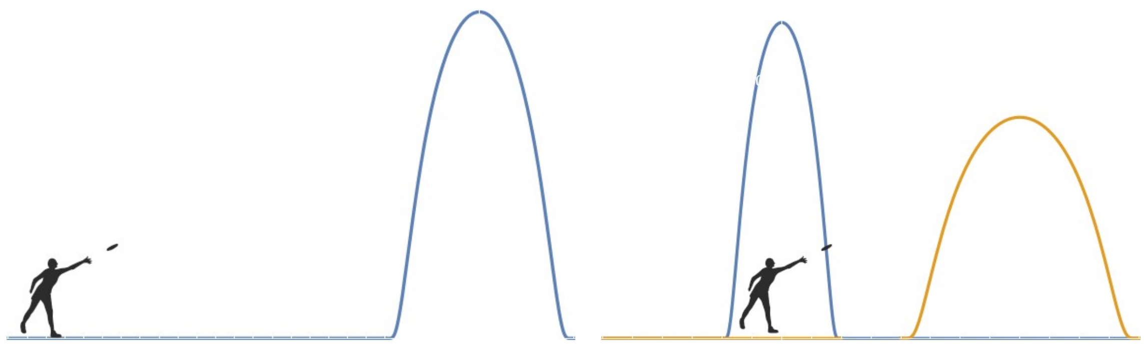
- The PF of $2X$ cannot be obtained by multiplying the PF of X by 2.
- If X, Y have the same distr, it's not (necessarily) true that $X = Y$.

§3.8 - Independence of Rvs

Similar to indep of *events* examined earlier. Intuitively, indep *rvs* X and Y means: if you know the value of X, this gives you no info about the value of Y.

For example, seeing the result of flipping a (fair) coin ($X = 1$, heads or $X = 0$, tails) gives you no info about the next flip (Y).

Def (**Indep of Two Cont**. **Rvs**): Continuous rvs X and Y are indep if $P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$, for **ALL** $x, y \in \mathbb{R}$.
(Recall $P(A,B) = P(A)P(B)$ for indep events)

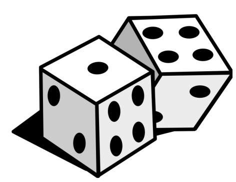


Repeating from same spot, indep. Throw from previous landing spot, not indep.

Def (**Indep of Two Discrete Rvs**): Discrete rvs X and Y are indep if $P(X = x, Y = y) = P(X = x)P(Y = y)$, for all x, y with x in the support of X and y in the support of Y.

Def (**Indep of Many Rvs**): Continuous rvs $X_1, \ldots, X_n$ are indep if $P(X_1 \le x_1, \ldots, X_n \le x_n) = P(X_1 \le x_1) \ldots P(X_n \le x_n)$, for *ALL* $x_1, \ldots, x_n \in \mathbb{R}$. (hence only one equation, as opposed to many eqs for events/discrete rvs)
For infinitely many cont. rvs, we say that they are indep if every finite subset of the rvs is indep.

☣ If $X_1, \ldots, X_n$ are indep, then they are pairwise indep, i.e., X_i is indep of X_j for $i \ne j$.
The idea behind proving that X_i and X_j are indep is to let all the x_k (other than x_i, x_j) go to ∞ in the definition of indep, since we already know $X_k < \infty$ is true (though it takes some work to give a complete justification for the limit).
But pairwise indep does not imply indep in general, as we saw in Chapter 2 for events.



Ex (**Dice Roll Indep**). In a roll of two fair dice, if X is the # on the first die and Y is the # on the second die, is $X + Y$ independent of $X - Y$??

Solution: No. Note that:

$$0 = P(X + Y = 12, X - Y = 1)$$

$$\ne P(X + Y = 12)P(X - Y = 1) = \tfrac{1}{36}\tfrac{5}{36}.$$

Def (**iid**): We'll often work with rvs that are indep and have the same distr.
We call such rvs indep and identically distributed, or iid for short.

Repeated throws are iid

Thm (**Rv Function Indep**): If X and Y are indep rvs, then any function of X is indep of any function of Y. For example X^2 would be indep from $\sqrt{\ln Y}$.

Thm (**Binomial via Bernoulli**): If $X \sim Bin(n,p)$, viewed as # of successes in n indep Bernoulli trials w/success prob p, we can write $X = X_1 + \ldots + X_n$ where the X_i are iid $Bern(p)$.

Proof. Let $X_i = 1$ if the ith trial was a success, and 0 if the ith trial was a failure.

It's like we have a person assigned to each trial, and ask each to raise their hand if their trial was a success.

If we count the raised hands (which is the same as adding up the X_i), we get the total # of successes. ■

Activity 7

Thm (**Indep Binomial Addition**). If $X \sim Bin(n,p)$, $Y \sim Bin(m,p)$, and X is indep of Y, then $X + Y \sim Bin(n+m,p)$.

Proofs (**1 more in book**)

Representation: Represent both X&Y as sum of i.i.d. $Bern(p)$ rvs:
$X = X_1 + \ldots + X_n$ and $Y = Y_1 + \ldots + Y_m$, where X_i and Y_j are all iid $Bern(p)$.

Then $X + Y$ is the sum of $n + m$ iid $Bern(p)$ rvs, so its distr, by previous thm, is $Bin(n+m,p)$. ■

Story: By the Binomial story, X is # of successes in n indep trials and Y is # of successes in m additional indep trials, all w/same success probability.

So $X + Y$ is total # of successes in the $n + m$ trials, which is the story of the $Bin(n+m,p)$ distr. ■

Def (**Conditional Indep of Rvs**). Rvs X and Y are conditionally indep given Z if for all $x, y \in \mathbb{R}$ and all z in the support of Z, $P(X \le x, Y \le y | Z = z) = P(X \le x | Z = z) P(Y \le y | Z = z)$.

For discrete rvs, an equivalent definition is to require: $P(X = x, Y = y | Z = z) = P(X = x | Z = z) P(Y = y | Z = z)$.

Def (**Conditional PF**). For any discrete rvs X and Z, the function $P(X = x | Z = z)$, when considered as a function of x for fixed z, is the conditional PF of X given $Z = z$.

❗ As with events, indep of rvs does not imply conditional indep (or vice versa).

Ex (**Matching pennies**). Consider a game called matching pennies. Each of two players, A and B, has a fair penny. They flip their pennies independently. If the pennies match, A wins; otherwise, B wins. Let X be 1 if A's penny lands Heads and -1 otherwise, and define Y similarly for B.

Let $Z = XY$, which is 1 if A wins and -1 if B wins. Then X and Y are unconditionally indep, but given $Z = 1$, we know $X = Y$ (the pennies match). So X and Y are conditionally dependent given Z. ☐

What did we learn?

- CDFs and Valid CDFs
- Functions of rvs
- Indep of rvs
