Math 340/640 Lec 1

A discrete random variable (rv) X is defined by its prob. mass function (PMF) $p(x) := P(X = x)$ and denote $X \sim p(x)$ and read "X is distr. as $p(x)$". x is the realized value

and cumulative distribution function CDF

$F(x) := P(X \le x)$ and CDF-complement or survival function

$S(x) := 1 - F(x) = P(X > x)$

X: r.v.

x: realization or "draw"

multiple realizations: "data"

The discrete rv has "support", i.e. the set of possible unique realized values

$$\mathrm{Supp}(X) := S_X := \{x : p(x) > 0, x \in \mathbb{R}\} \text{ and } |S_X| \le |\mathbb{N}|$$

The size is at most ctbl. ∞ i.e. the # of elements in the set $\{1, 2, 3, \ldots\}$.

The support and PMF are related by the following identity:

$$\sum_{x \in S_X} p(x) = 1$$

which I call the Humpty-Dumpty identity i.e. all the prob. "pieces" must be able to be put back together into the whole.

The first non-trivial support set is $S_X = \{0, 1\}$, with a PMF which must look like:

$$p(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases} = p^x(1-p)^{1-x}$$ the "Bernoulli" r.v.

piecewise function

The value p is known as a "parameter." We'll address this later. e.g. how do you find its derivative?

Piecewise functions are difficult to work with, let's express as non-piecewise

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

Silly question.... $P(X=7) = 0$ so...

$p(7) = p^7(1-p)^{1-7} = \frac{p^7}{(1-p)^6} \neq 0$. What happened? $p(x)$ is only valid for $x \in S_X$!

This is as annoying as a piecewise function. We have to check each time if we're permitted to use the PMF. Let's modify the PMF so its domain is all $\mathbb{R}$. To do so, let's define:

$$\mathbb{1}_C := \begin{cases} 1 & \text{if condition } C \text{ is true} \\ 0 & \text{if condition } C \text{ is false} \end{cases}$$ "indicator is 'on'" "indicator is 'off'"

PMF old-style

page 27

$$X \sim \text{Bern}(p) := p(x) = \underbrace{p^x(1-p)^{1-x}}_{p_{old}(x)} \mathbb{1}_{x \in \{0,1\}}$$

We will see both old and new versions of PMFs.

"Brand name" r.v.'s: famous and popular distributions. They're defined by their PMF's.

In general $p(x) = p_{old}(x) \cdot \mathbb{1}_{x \in S_X}$

Now we can say $\sum_{x \in \mathbb{R}} p(x) = 1$ since for $x \in \mathbb{R} \setminus S_X$, $p(x) = 0$.

Some practice with indicator functions...

We will be using them frequently for "selecting sets" e.g.

$$\sum_{x \in \{1,2,\dots,7\}} \mathbb{1}_{x \in \{1,5,6\}} = 1 + 0 + 0 + 0 + 1 + 1 + 0 = 3$$

$$\sum_{x \in \mathbb{N}} \mathbb{1}_{x \in \{1,5,6\}} = 1 + 0 + 0 + 0 + 1 + 1 + 0 + 0 + \dots = 3$$

$$\sum_{x \in \mathbb{R}} \mathbb{1}_{x \in \{1,5,6\}} = 3$$

$$\sum_{x \in \{1,5,6\}} \mathbb{1}_{x \in \{1,2,\dots,7\}} = 1 + 1 + 1 = 3$$

$$\sum_{x \in \{1,5,6\}} \mathbb{1}_{x \in \mathbb{N}} = 1 + 1 + 1 = 3$$

$$\sum_{x \in \{1,5,6\}} \overset{1}{\cancel{\mathbb{1}_{x \in \mathbb{R}}}} = 1 + 1 + 1 = 3$$

$$\sum_{x \in \mathbb{R}} f(x) \mathbb{1}_{x \in \{1,5,6\}} = f(1) + f(5) + f(6)$$

$$\sum_{x \in \mathbb{R}} f(x) \mathbb{1}_{x \in A} = \sum_{x \in A} f(x)$$

$$\sum_{x \in B} f(x) \mathbb{1}_{x \in A} = \sum_{x \in A \cap B} f(x)$$

We ignored the "p": $P(X=1)=p$. Higher the p, more likely $X=1$. What if $p=1$?

$$X \sim \text{Bern}(1) := 1^x (1-1)^{1-x} \mathbb{1}_{x \in \{0,1\}}\ ; \quad p(1) = 1^1 (1-1)^{1-1} \cdot 1 = 1 \cdot \underset{\text{if } 0^0}{0^0} \cdot 1 = 1$$

$$p(0) = 0^0 (1-1)^{1-0} \cdot 1 = 0^0 \cdot 0 \cdot 1 = 0$$

What is 0^0? Discussed for 100's of years! It's actually indeterminate, (undefined). So we define it to be one. If we didn't, a whole lot of stuff would break e.g.

$$4 = 2^2 = (2+0)^2 \underset{\text{binom thm.}}{=} \sum_{i=0}^{2} \binom{2}{i} 2^i 0^{2-i} = \cancel{\binom{2}{0} 2^0 0^2} + \cancel{\binom{2}{1} 2^1 0^1} + \binom{2}{2} 2^2 0^0 \underset{\text{if } 0^0=1}{=} 4$$

$$e^x \underset{\text{Taylor series}}{=} \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

$$1 = e^0 = \frac{0^0}{0!} + \cancel{\frac{0^1}{1!} + \frac{0^2}{2!} + \cdots} \underset{\text{if } 0^0=1}{=} 1$$

There are some other examples. Ours is equally compelling!

Anyways... If $p=1$, this rv always realizes to $x=1$!

If $p=0$, $X \sim \text{Bern}(0) := 0^x (1-0)^{1-x} \mathbb{1}_{x \in \{0,1\}}$

$$p(1) = 0^1 (1-0)^{1-1} \cdot 1 = 0 \cdot 0^0 \cdot 1 = 0$$

$$p(0) = 0^0 (1-0)^{1-0} \cdot 1 = 0^0 \cdot 1^1 \cdot 1 = 1$$

If $p=0$, this rv always realizes to $x=0$!

Oxymoron to call this a "random" variable. So we call it "degenerate", our second brand name r.v.

$$X \sim \text{Deg}(c) := \begin{cases} 1 \text{ if } x=c \\ 0 \text{ if } x \neq c \end{cases} = \mathbb{1}_{x=c}$$

$\text{Bern}(1) = \text{Deg}(1)$, $\text{Bern}(0) = \text{Deg}(0)$

Original question. What could values of p be?

Could p be 2?

$$p(x) = 2^x(1-2)^{1-x} \mathbb{1}_{x \in \{0,1\}}$$
$$p(1) = 2^1 \cdot (-1)^{-1} \cdot 1 = -2 \ \ldots \text{ not a prob.}$$

$\Rightarrow$ these are illegal param values!

Same thing for all $p > 1$ and $p < 0 \Rightarrow p \in [0,1]$, the set of legal param values resulting in a valid rv

But we just saw $p = 0$ or $p = 1$ results in a degeneracy.

So the convention is to omit parameter values resulting in a degeneracy.

$\Rightarrow p \in (0,1)$, the "parameter space". So what is p really?

A tuning knob to control how often you get 1's or 0's.

If we have $X_1, X_2, \ldots, X_n$, n r.v.'s, we can group them into a vector rv, $\vec{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ and define the joint mass function (jmf)

as $p(\vec{x}) = p_{\vec{X}}(\vec{x}) = p_{X_1, \ldots, X_n}(x_1, \ldots x_n)$ valid for all $\vec{x} \in \mathbb{R}^n$ plus $\sum_{\vec{x} \in \mathbb{R}^n} p_{\vec{X}}(\vec{x}) = 1$

If $X_1, \ldots X_n \overset{ind}{\sim}$ (are independent), $p_{\vec{X}}(\vec{x}) = p_{X_1}(x_1) p_{X_2}(x_2) \cdots p_{X_n}(x_n) = \prod_{i=1}^n p_{X_i}(x_i)$

If $X_1, \ldots X_n$ are equally distr.

$p(x) = p_{X_1}(x_1) = p_{X_2}(x_2) = \ldots = p_{X_n}(x_n)$ but there is no simplification of $p_{\vec{X}}(\vec{x})$.

If $X_1, \ldots X_n \overset{iid}{\sim}$ independent and identically distr.,

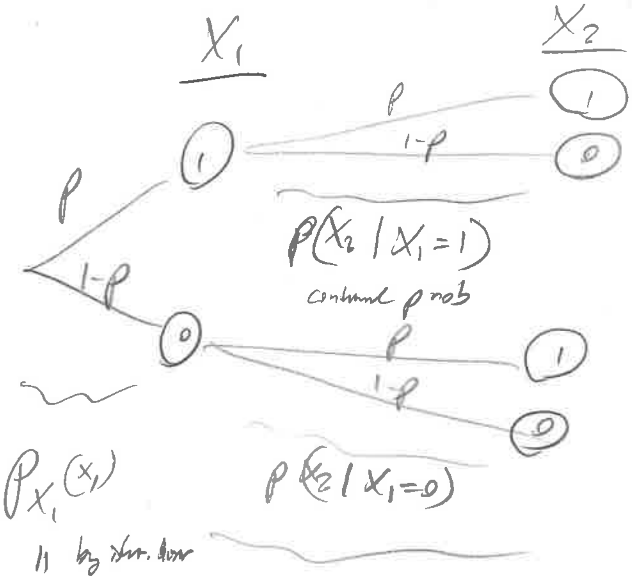
$$p_{\vec{X}}(\vec{x}) = \prod_{i=1}^n p(x_i)$$

no need for subscript as all $p(x)$ are the same.

Consider $X_1, X_2 \overset{iid}{\sim} \text{Bern}(p)$. Let $T := f(X_1, X_2) = X_1 + X_2 \sim p_T(t)$, a new r.v. built as a function of two r.v. inputs. Can we figure out $p_T(t)$? First $S_T \overset{?}{=} \{0,1,2\} = S_{X_1} + S_{X_2}$ set addition:

$A + B := \{a + b : a \in A, b \in B\}$

Now we need to examine all possible input configurations. Use a tree:



$p_{\vec{X}}(\vec{x})$	T
$p^2 = p([1, 1]^T)$	2
$p(1-p) = p([1\ 0]^T)$	1
$(1-p)\,p = p([0\ 1]^T)$	1
$(1-p)^2 = p([0\ 0]^T)$	0

by ident. distr.

$p(x_1)$ $\quad p_{X_2|X_1}(x_2, x_1) = p_{X_2}(x_2) = p(x_2)$

Conditional PMF

Before we derive the PMF of T_2, we first review joint, cond., marginal PMFs. Consider this table

$$p(X_2 \mid X_1 = 0) = \begin{cases} \frac{p(1-p)}{1-p} & \text{if } x_2 = 1 \\ \frac{(1-p)^2}{1-p} & \text{if } x_2 = 0 \end{cases}$$

$p_{X_1,X_2}(x_1, x_2)$ JMF

$$(1-p)^2 + p(1-p) = 1 - 2p + \cancel{p^2} + p - \cancel{p^2} = 1 - p$$

	X_2: 0	1	
X_1: 0	$(1-p)^2$	$p(1-p)$	$1-p$
1	$p(1-p)$	p^2	p
$p_{X_2}(x_2)$	p	$1-p$	1

$p_{X_1}(x_1)$

marginal PMF "in margin" of table

2 ways to get $T=1$, added (6

$$T \sim \begin{cases} 2 & \text{wp } p^2 \\ 1 & \text{wp } p(1-p)+p(1-p) = 2p(1-p) \\ 0 & \text{wp } (1-p)^2 \end{cases}$$

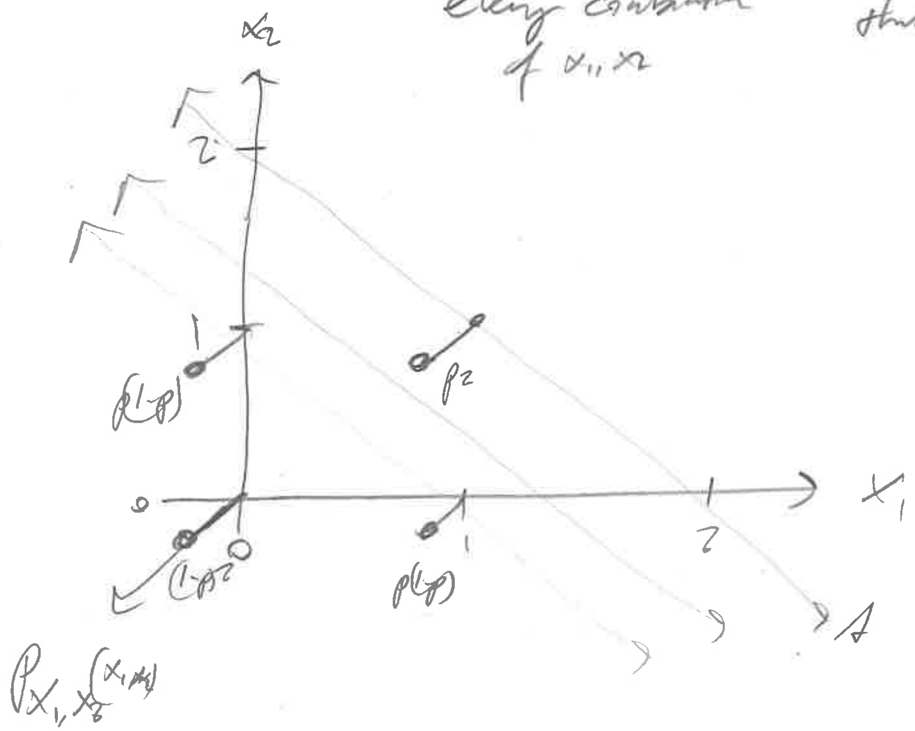
Generally ...

finally selects the combination of x_1, x_2 that yield the desired total t.

$$p_{X_1}(x_1) * p_{X_2}(x_2) := P(T=t) = \underbrace{\sum_{x_1 \in \mathbb{R}} \sum_{x_2 \in \mathbb{R}}}_{\text{search through every combination of } x_1, x_2} \underbrace{p_{X_1,X_2}(x_1,x_2)}_{\text{calc the prob of that combination}} \mathbb{1}_{x_1+x_2=t}$$

convolution operator

but this formula isn't so useful ...



If $t = x_1 + x_2$

$\Rightarrow x_2 = t - x_1$

$$\Rightarrow P(T=t) = \sum_{x_1 \in \mathbb{R}} p_{X_1,X_2}(x_1, t-x_1)$$

- the indicator $\mathbb{1}_{x_1+x_2=t}$ is omitted since it's always "on"
  $x_1 + (t-x_1) = t$ ✓
- the sum over $\sum_{x_2 \in \mathbb{R}}$ is omitted since once given a value of x_1, x_2 is fixed at $t-x_1$

Using this formula, let's graph

$P(T=2)$ = sum of probs on line $x_2 = 2 - x_1$

$\Rightarrow P(T=2) = p^2$ $\qquad P(T=1) = p(1-p) + p(1-p)$

$P(T=1.5) = 0$

Can we get a better convolution formula?