Math 180A: Introduction to Probability

Lecture B00 (Nemish)

math.ucsd.edu/~ynemish/teaching/180a

Lecture C00 (Au)

math.ucsd.edu/~bau/f20.180a

Today: ASV 6.1, 6.2

Video: Prof. Todd Kemp, Fall 2019

Next: ASV 6.2, 6.3

Week 8: Homework 7 (due Friday, Dec 4)

Regrades for Homework 5 (Nov 30 - Dec 1)

Definition: Given (discrete) random variables $X_1, X_2, \dots, X_n$ all defined on the same sample space, their <u>joint distribution</u> is the collection of all

$$\left\{ \begin{array}{l} \mathbb{P}(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n) \\ \text{all possible values } k_1 \text{ of } X_1,\ k_2 \text{ of } X_2, \dots, k_n \text{ of } X_n \end{array} \right\}$$

Eg. $X, Y \sim \mathrm{Ber}(p)$,

(1) $X = Y$

(2) X, Y independent

$$\mathbb{P}(X = k_1, Y = k_2)$$
$$= \mathbb{P}(X = k_1)\,\mathbb{P}(Y = k_2)$$

$Y \backslash X$	0	1
0	$1-p$ $(1-p)^2$	0 $p(1-p)$
1	0 $(1-p)p$	p p^2

$$p_{X,Y}(k_1, k_2) = \mathbb{P}(X = k_1, Y = k_2)$$

Recovering X_j from $\underline{X} = (X_1, X_2, \ldots, X_n)$: <u>Marginals</u>

Suppose we know $p_{\underline{X}}(\underline{k})$ for all $\underline{k} = (k_1, k_2, \ldots, k_n)$.

How can we find $p_{X_1}(t)$?

Eg. Toss a fair coin twice. X_1, X_2. $\in \{0,1\}$

$$\frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

$$\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 0, X_2 = 0) + \mathbb{P}(X_1 = 0, X_2 = 1)$$

$$\{X_1 = 0\} = \{X_1 = 0, X_2 = 0\} \cup \{X_1 = 0, X_2 = 1\}$$

In general,

$$\mathbb{P}(X_1 = t) = \sum_{k_2, k_3, \ldots, k_n} \mathbb{P}(X_1 = t, X_2 = k_2, \ldots, X_n = k_n) = \sum_{k_2, \ldots, k_n} p_{\underline{X}}(t, k_2, k_3, \ldots, k_n)$$

$$\{X_1 = t\} = \bigcup_{k_2, \ldots, k_n} \{X_1 = t, X_2 = k_2, X_3 = k_3, \ldots, X_n = k_n\}$$

$$\mathbb{P}(X = t) = \sum_k p_{(X,Y)}(t, k) \qquad \mathbb{P}(Y = t) = \sum_k p_{(X,Y)}(k, t)$$

(X, Y)

Eg. Toss a fair coin 3 times. X = # tails in first toss (0 or 1)

Y = total # tails in all 3 (0, 1, 2, 3)

$\text{Ber}(\tfrac{1}{2})$

	$Y=0$	1	2	3	
$X=0$	$1/8$	$1/4$	$1/8$	0	$\to 1/2$
1	0	$1/8$	$1/4$	$1/8$	$\to 1/2$
	$1/8$	$3/8$	$3/8$	$1/8$	$\leftarrow \text{Bin}(3,\tfrac{1}{2})$

$\mathbb{P}(X=k, Y=l)$

k	l
0	0
1	1
	2
	3

Outcome	X	Y
HHH	0	0
HHT	0	1
HTH	0	1
HTT	0	2
THH	1	1
THT	1	2
TTH	1	2
TTT	1	3

each has $p = \frac{1}{8}$.

Question: Are X, Y independent?

$\mathbb{P}(X=1, Y=0) = 0$

$\mathbb{P}(X=1)\mathbb{P}(Y=0) = \frac{1}{2} \cdot \frac{1}{8} = \frac{1}{16} \neq 0.$

No!

Joint distributions are just distributions of random vectors.

$X_1, X_2, \dots, X_n \rightsquigarrow \underline{X} = (X_1, X_2, \dots, X_n)$

possible values for $\underline{X}$ are vectors $(k_1, \dots, k_n)$.

Eg. Multinomial Distribution.

Often, trials have more than 2 outcomes.

Consider a trial w r possible outcomes, w probabilities $p_1, p_2, \dots, p_r$

$p_1 + p_2 + \dots + p_r = 1$

Perform n trials.

For $1 \le j \le r$, X_j = # times we get outcome j.

Possible values for $\underline{X} = (X_1, \dots, X_r)$: $(k_1, \dots, k_r) \leftarrow k_j \in \{0, 1, \dots, n\}$

$k_1 + \dots + k_r = n$

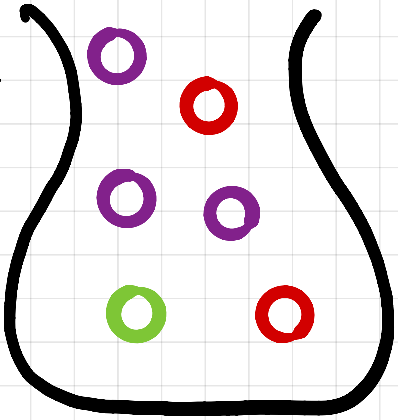
$$\mathbb{P}(\underline{X} = \underline{k}) = \left(\begin{matrix}\#\text{arrangements} \\ \text{w } k_1 \text{ 1s, } k_2 \text{ 2s}, \dots, k_r \text{ r's}\end{matrix}\right) p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$$

$(r=2)$: $\left(\frac{n!}{k_1! \, k_2!} = \frac{n!}{k_1!(n-k_1)!} = \binom{n}{k_1} \right)$

pmf of $\mathrm{Mult}(n; p_1, \dots, p_n)$

$$\binom{n}{k_1, k_2, \dots, k_r} = \frac{n!}{k_1! \, k_2! \cdots k_r!}$$

Eg.



Sample 10 times with replacement.

$\mathbb{P}(3 \text{ green}, 2 \text{ red}, 5 \text{ blue})$

$G = \#\text{green}, \quad R = \#\text{reds}, \quad B = \#\text{blues}$

$\underline{X} = (G, R, B) \sim \text{Mult}(10; \tfrac{1}{6}, \tfrac{1}{3}, \tfrac{1}{2})$

$\mathbb{P}(\underline{X} = (3, 2, 5)) = \frac{10!}{3!\,2!\,5!} \left(\frac{1}{6}\right)^3 \left(\frac{1}{3}\right)^2 \left(\frac{1}{2}\right)^5 = 4.05\%$

Note: if $r = 2$, $\text{Mult}(n; p, q)$

Eg. Suppose $\underline{X} \sim \text{Mult}(n; p_1, p_2, \ldots, p_r)$. Find the marginal distribution of X_1.

$$\mathbb{P}(X_1 = t) = \sum_{\substack{k_2, k_3, \ldots, k_r \geq 0 \\ t + k_2 + \cdots + k_r = n}} \frac{n!}{t!\,k_2! \cdots k_r!} p_1^t p_2^{k_2} \cdots p_r^{k_r}$$

Complicated! Instead observe:

X_1 = # successes in n indep. trials

where success = outcome 1 ($\mathbb{P} = p_1$)

failure = outcomes $2 - r$ ($\mathbb{P} = p_2 + \cdots + p_r = 1 - p$)

$X_1 \sim \text{Bin}(n, p)$

Expectations

Let $\underline{X} = (X_1, \dots, X_n)$ be a (discrete) random vector with joint probability mass function $p_{\underline{X}}$.

If $g: \mathbb{R}^n \to \mathbb{R}$ is a function, $Y = g(\underline{X})$ is a random variable (still discrete).

$$\mathbb{E}(Y) = \sum_t t \cdot \mathbb{P}(Y = t)$$

$$= \sum_{\underline{k}} g(\underline{k}) \, \mathbb{P}(\underline{X} = \underline{k})$$

$$= \sum_{\underline{k}} g(\underline{k}) \, p_{\underline{X}}(\underline{k})$$

Eg. Toss a fair coin twice, $X_1, X_2 \in \{0, 1\}$.

$$\mathbb{E}(X_1 X_2) = \sum_{k_1, k_2} \underbrace{g(k_1, k_2)}_{k_1 k_2} \, \underbrace{\mathbb{P}(X_1 = k_1, X_2 = k_2)}_{1/4}$$

$$= 0 \cdot 0 \cdot \tfrac{1}{4} + \cancel{0 \cdot 1 \cdot \tfrac{1}{4}} + \cancel{1 \cdot 0 \cdot \tfrac{1}{4}} + 1 \cdot 1 \cdot \tfrac{1}{4}$$

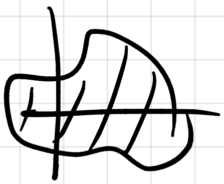
$$= 1/4 \leftarrow \mathbb{E}(X)\mathbb{E}(Y)$$

$\mathbb{E}(X) = 1/2$

$\mathbb{E}(Y) = 1/2$

Jointly Continuous Random Vectors

A random vector $\underline{X} = (X_1, \dots, X_n)$ has a pdf $f_{\underline{X}} : \mathbb{R}^n \to \mathbb{R}_+$ if, for "nice" subsets $B \subseteq \mathbb{R}^n$



$$\mathbb{P}(\underline{X} \in B) = \int_B f_{\underline{X}}(x_1, \dots, x_n)\, dx_1 \cdots dx_n .$$

(we say $X_1, X_2, \dots, X_n$ are jointly continuous.)

Properties: (1) $f_{\underline{X}} \geq 0$

(2) $\int_{\mathbb{R}^n} f_{\underline{X}} = 1$.

Eg. Standard Multivariate Normal $f(x_1, \dots, x_n) = (2\pi)^{-n/2} \underbrace{e^{-\frac{(x_1^2 + \dots + x_n^2)}{2}}}_{e^{-x_1^2/2}\, e^{-x_2^2/2} \cdots e^{-x_n^2/2}}$

$$\int_{\mathbb{R}^n} f(x_1, \dots, x_n)\, dx_1 \cdots dx_n = (2\pi)^{-n/2} \int_{-\infty}^{\infty} e^{-x_1^2/2}\, dx_1 \cdot \int_{-\infty}^{\infty} e^{-x_2^2/2}\, dx_2$$

$$1 = \cdots \int_{-\infty}^{\infty} e^{-x_n^2/2}\, dx_n = (2\pi)^{-n/2} (\sqrt{2\pi})(\sqrt{2\pi}) \cdots (\sqrt{2\pi})$$