MATH 180A: INTRO TO PROBABILITY (FOR DATA SCIENCE)

www.math.ucsd.edu/~tkemp/180A

Today: § 3.1-3.2

Next: § 2.4-2.5

HW2 due TONIGHT!

Lab 2.2 due Monday, 10/14.

* Regrade requests for HW1: window Tuesday 10/15 8am – 11pm
  - ↳ separate request for each problem
  - ↳ detailed, polite responses please.
* Numerical answers for HW in rubric on Gradescope
* Lab 1 solutions posted on datahub.ucsd.edu.

Random Variables

3.1

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ a random variable is a function

$$X : \Omega \to \mathbb{R}$$

This is a bad, old-fashioned name. Would be better to call it a random function or random measurement.

Eg. Toss a fair coin 4 times. Let X = number of tails.

Eg. Shoot an arrow at a circular target. Y = distance from center.

Eg. Your car is in a minor accident; the damage repair cost is a random number between \$100 and \$1500. Your insurace deductible is \$500. Z = your out of pocket expenses.

In all these examples, think about what you observe. You can't really see a formula for the function. By repeating the experiment over and over, all you can learn is the

Probability Distribution

3.2

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable

$$X : \Omega \to \mathbb{R}$$

the probability distribution or law of X is a probability measure μ_X ON $\mathbb{R}$.

$$A \subseteq \mathbb{R} \rightsquigarrow \mu_X(A) =$$

[Caution: for this to make sense, we need to have a designated set of allowed "events" in $\mathbb{R}$; call this collection $\mathcal{B}(\mathbb{R})$

Then, we must have

$$\boxed{}$$

This is a condition on X; we call such functions measurable.

We will ignore these technicalities in this course; all our random variables are indeed measurable.]

E.g. $\underbrace{\text{Toss a fair coin 4 times}}$. Let X = number of tails.

$$\Omega = \{ (x_1, x_2, x_3, x_4) \in \{H, T\}^4 \}$$

$\mathbb{P}$ = uniform on Ω;

$$\mathbb{P}\{(x_1, x_2, x_3, x_4)\} = \frac{1}{2^4} = \frac{1}{16}$$

$$X \in \{0, 1, 2, 3, 4\} \subset \mathbb{R}$$

if $A \subseteq \mathbb{R}$ does not contain one of these numbers, $\mu_X(A) = 0$.

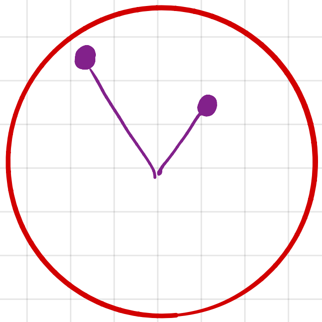
By additivity of probability measures, to understand μ_X, just need to know

$$\mu_X(\{k\}) = ? \quad 0 \le k \le 4$$

$$\mathbb{P}(X = k) = p_X(k)$$

k	0	1	2	3	4
$p_X(k)$					

Eg. Shoot an arrow at a circular target of radius 1.



Y = distance from center.

Last lecture, you calculated that:

- If $r \leq 0$, $\mathbb{P}(Y \leq r) = 0$
- If $r \geq 1$, $\mathbb{P}(Y \leq r) = 1$.
- If $r \in [0,1]$, $\mathbb{P}(Y \leq r) = \dfrac{\text{Area}(\mathbb{D}_r)}{\text{Area}(\mathbb{D}_1)} = r^2$.

$$\mathbb{P}(Y \leq r) = \mathbb{P}(\{Y \in (-\infty, r]\})$$

$\leftarrow$ Can get information about other sets $A \subseteq \mathbb{R}$ using the properties of $\mathbb{P}$.

What can we say about $\mathbb{P}(Y = 0.4)$?

We will focus mostly on two kinds of random variables:

<u>discrete</u>: There are finitely (or countably) many possible values $\{k_1, k_2, k_3, \ldots\}$ for X.

↳ μ_X is described by the probability mass function $p_X(k) = \mathbb{P}(X = k)$, $k \in \{k_1, k_2, k_3, \ldots\}$

In this case, by the laws of probability,

<u>continuous</u>: For any real number $t \in \mathbb{R}$, $\mathbb{P}(X = t) = 0$.

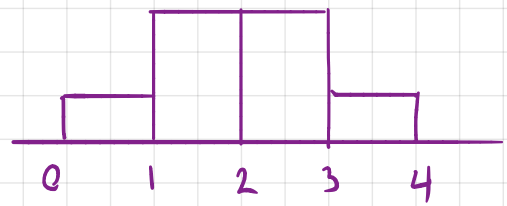
↳ μ_X is captured by understanding $\mathbb{P}(X \leq r)$ as a function of r.

Cumulative Distribution Function (CDF)

For any random variable X, $F_X(r) = \mathbb{P}(X \leq r)$. ← Actually works in all cases – including discrete.

E.g. $\mu_X = \mathrm{Bin}(3, \frac{1}{2})$

k	0	1	2	3
$p_X(k)$	$1/8$	$3/8$	$3/8$	$1/8$



- $r < 0$, $\{X \leq r\} =$
- $0 \leq r < 1$, $\{X \leq r\} =$
- $1 \leq r < 2$, $\{X \leq r\} =$
- $2 \leq r < 3$, $\{X \leq r\} =$
- $r \geq 3$, $\{X \leq r\} =$

Properties of the CDF $F_X(r) = \mathbb{P}(X \le r)$

(1) Monotone increasing: $s \le t \Rightarrow F_X(s) \le F_X(t)$

(2) $\lim_{r \to -\infty} F_X(r) = 0$, $\lim_{r \to +\infty} F_X(r) = 1$.

(3) The function F_X is right-continuous: $\lim_{t \to r+} F_X(t) = F_X(r)$.

<u>Corollary</u>: If X is a continuous random variable, F_X is a continuous function.

Densities

Some continuous random variables have probability densities. This is an infinitesimal version of a probability mass function.

X discrete, $\in \{k_1, k_2, k_3, \ldots\}$	X continuous
$p_X(k) = \mathbb{P}(X = k)$	$\mathbb{P}(X = t) = 0$ for all $t \in \mathbb{R}$.
$\mathbb{P}(X \in A) = \sum_{k \in A} \mathbb{P}(X = k) = \sum_{k \in A} p_X(k)$	
$p_X(k) \geq 0, \quad \sum_{k} p_X(k) = 1.$	

Eg. Shoot an arrow at a circular target of radius 1.

Y = distance from center.

$$\int_{-\infty}^{r} f(t)\,dt \overset{?}{=} \mathbb{P}(Y \in (-\infty, r]) = F_Y(r) = \begin{cases} 0, & r \leq 0 \\ r^2, & 0 \leq r \leq 1 \\ 1, & r \geq 1 \end{cases}$$

Theorem: If F_X is continuous and piecewise differentiable, then X has a density $f_X = F_X'$.

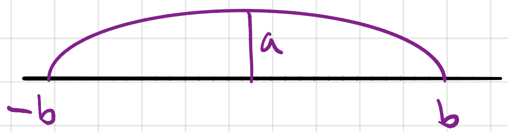
Proof:

Eg. Let X = a uniformly random number in $[0,1]$.

As we discussed in lecture 2, this means

$$\mathbb{P}(X \in [s,t]) = t - s \quad \text{if} \quad 0 \leq s < t \leq 1.$$

Eg. Let $f(t) = c\sqrt{b^2 - t^2}$ for $|t| \le b$, 0 otherwise
(for some positive constants $b, c > 0$).



Is f a probability density?

Eg. Your car is in a minor accident; the damage repair cost is a random number between \$100 and \$1500. Your insurace deductible is \$500. Z = your out of pocket expenses.

The random variable Z is

(a) Continuous

(b) discrete

(c) neither

(d) both