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

OH Tomorrow 9-11am APM 6402

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

Today: § 3.3-3.4

Lab 3.2 due TONIGHT!

Next: Review (practice exams)

* Regrade requests for HW2: Tuesday, 10/22 } 8am – 11pm
for Lab2: Thursday, 10/24 }

* Midterm 1: Wednesday, 10/23, 8-10pm, PCYNH 109
Assigned seats (see TritonEd)
1 double sided sheet of (hand-written) notes allowed
Practice Midterms posted on webpage

Expectation

Definition: Let X be a discrete random variable with possible values $t_1, t_2, t_3, \ldots$ The expectation or expected value of X is

$$E(X) := \sum_j t_j P(X = t_j)$$

It is also called the mean of X, and is often denoted μ.

E.g. Let X be a discrete random variable with probability mass fn.

$$p_X(k) = \frac{c}{k(k+1)}, \quad k = 1, 2, 3, \ldots \qquad \sum_{k=1}^{\infty} p_X(k) = 1$$

Find c, and compute $E(X)$.

$$\frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}$$

$$1 = \sum_{k=1}^{\infty} \frac{c}{k(k+1)} = c \sum_{k=1}^{\infty} \frac{1}{k(k+1)}$$

$$= c\left(\left(1 - \cancel{\tfrac{1}{2}}\right) + \left(\cancel{\tfrac{1}{2}} - \cancel{\tfrac{1}{3}}\right) + \left(\cancel{\tfrac{1}{3}} - \cancel{\tfrac{1}{4}}\right) + \left(\cancel{\tfrac{1}{4}} - \cancel{\tfrac{1}{5}}\right) + \cancel{\cdots} \cdots\right) = c$$

$$E(X) = \sum_{k=1}^{\infty} k \cdot p_X(k) = \sum_{k=1}^{\infty} \cancel{k} \cdot \frac{1}{\cancel{k}(k+1)} = \infty.$$

Expectation of Continuous Random Variable (with density)

X discrete, $X \in \{t_1, t_2, t_3, \dots\}$

$$\mathbb{E}(X) = \sum_j \underbrace{t_j \, \mathbb{P}(X = t_j)}_{} = \sum_t t \cdot p_X(t)$$

X continuous ($\mathbb{P}(X=t)=0$ for each $t \in \mathbb{R}$)

$\uparrow$

propability density $f_X(t)$

<u>Def:</u> $$\boxed{\mathbb{E}(X) := \int_{-\infty}^{\infty} t \, f_X(t) \, dt}$$

Eg. Let $U \sim \mathrm{Unif}([a,b])$. $f_U(t) = \begin{cases} \frac{1}{b-a} & a \le t \le b \\ 0 & \text{otherwise} \end{cases}$ $\leftarrow$ 0 except on $[a,b]$

$$\mathbb{E}(U) = \int_{-\infty}^{\infty} t \, f_U(t) \, dt = \int_a^b t \cdot \frac{1}{b-a} \, dt = \frac{1}{b-a} \cdot \frac{1}{2} t^2 \Big|_{t=a}^{t=b}$$

$$= \frac{1}{2(b-a)} (b^2 - a^2) = \frac{a+b}{2}.$$

Question:

Shoot an arrow at a circular target of radius 1. What is the expected distance of the arrow from the center?

(a) 1

(b) 2/3

(c) 1/2

(d) 1/4

(e) 0

X = dist. from center

$$F_X(r) = \begin{cases} 0 & r \leq 0 \\ r^2 & 0 \leq r \leq 1 \\ 1 & r \geq 1 \end{cases}$$

$$f_X(r) = \begin{cases} 0 & r < 0 \\ 2r & 0 \leq r \leq 1 \\ 0 & r > 1 \end{cases}$$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} r f_X(r)\, dr$$

$$= \int_0^1 r \cdot 2r\, dr = \frac{2}{3} r^3 \Big|_0^1$$

$$= 2/3$$

$$\left(\frac{\text{Area}(\mathbb{D}_{2/3})}{\text{Area}(\mathbb{D}_1)} = \frac{\pi (2/3)^2}{\pi 1^2} = \frac{4}{9} \right)$$

E.g. $f(t) = \begin{cases} 0 & t \leq 1 \\ 1/t^2 & t > 1 \end{cases}$ is a $\underbrace{\text{probability density}}_{\geq 0 \checkmark}$.

$$\int_{-\infty}^{\infty} f(t)\,dt = \int_{1}^{\infty} \frac{1}{t^2}\,dt = -\frac{1}{t}\Big|_1^{\infty} = \left(-\frac{1}{\infty} - \left(-\frac{1}{1}\right)\right)$$

$$= \underline{1} \quad \checkmark$$

If X is a continuous random variable with $f_X = f$, what is $\mathbb{E}(X)$?

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} t \cdot f(t)\,dt = \int_{1}^{\infty} t \cdot \frac{1}{t^2}\,dt = \int_{1}^{\infty} \frac{1}{t}\,dt$$

$$= \ln|t| \Big|_1^{\infty}$$

$$= \ln|\infty| - \cancel{\ln|1|}_{0}$$

$$= \infty.$$

Expectations of Functions of Random Variables

$$\Omega \xrightarrow[\text{random variable}]{X} \mathbb{R} \xrightarrow[\text{function}]{g} \mathbb{R}$$

$g(X) = g \circ X$ is a new random variable.

Eg. $X \sim Bin(n,p)$ is the number of successes in n trials

$g(X) = X/n$ is the proportion of successful trials.

$g(X) \in \{0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \dots, \frac{n}{n}=1\}$

$$\mathbb{E}(g(X)) = \sum_t t \cdot \mathbb{P}(g(X)=t) = \sum_{k=0}^{n} \frac{k}{n} \mathbb{P}(g(X)=\frac{k}{n}) = \sum_{k=0}^{n} \frac{k}{n} \cdot \mathbb{P}(\frac{X}{n} = \frac{k}{n}) = \sum_{k=0}^{n} \frac{k}{n} \mathbb{P}(X=k) = \frac{1}{n}\mathbb{E}(X) = \frac{np}{n} = p.$$

In general: for a discrete random variable

Proposition: $\mathbb{E}(g(X)) = \sum_s g(s)\, \mathbb{P}(X=s)$.

Note, by definition, $\mathbb{E}(g(X)) = \sum_t t \cdot \mathbb{P}(g(X)=t)$.

Proposition: for a discrete random variable

$$E(g(X)) = \sum_s g(s)\, \mathbb{P}(X=s).$$

Pf. We know

$$\sum_t t \cdot \mathbb{P}(g(X)=t) \qquad \{g(X)=t\} = \bigcup_{s:\, g(s)=t} \{X=s\}$$

(disjoint)

$$= \sum_t t \sum_{s:\, g(s)=t} \mathbb{P}(X=s)$$

$$= \sum_t \sum_{s:\, g(s)=t} t \cdot \mathbb{P}(X=s)$$

$$= \underbrace{\sum_t \sum_{s:\, g(s)=t}}_{\sum_s} g(s)\, \mathbb{P}(X=s)$$

$\square$

Proposition: For a continuous random variable X with probability density f_X,

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(t) f_X(t)\, dt.$$

Eg. Let U be a uniform random variable on $[a,b]$. Then

$$\mathbb{E}(U^2) = \int_a^b t^2 f_U(t)\, dt = \int_a^b \frac{1}{b-a} t^2\, dt = \frac{1}{b-a} \cdot \frac{1}{3} t^3 \Big|_a^b$$

$$f_U(t) = \begin{cases} \frac{1}{b-a} & a \le t \le b \\ 0 & \text{otherwise} \end{cases}$$

$$= \frac{b^3 - a^3}{3(b-a)}$$

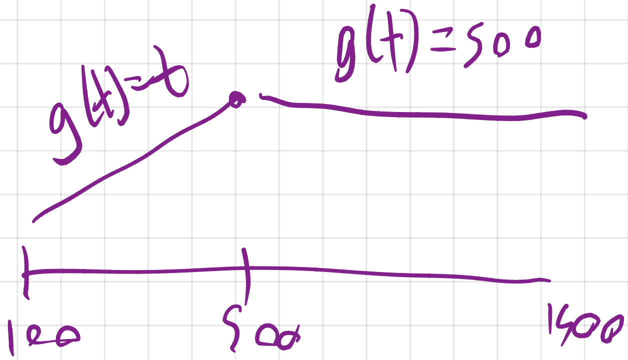
$$= \frac{a^2 + ab + b^2}{3}.$$

$$\left(a = 0,\ b = 1,\ \mathbb{E}(U_{[0,1]}^2) = \frac{1}{3}. \right)$$

Eg. Recall the car accident / insurance example.
An accident causes $\$Y$ of damage to your car, where $Y \sim \text{Unif}([100, 1500])$.
Your insurance deductible is \$500.
What is the expected amount you pay?

X = amount you pay $= \min(Y, 500)$
(neither continuous nor discrete r.v.) $= g(Y)$

$$\mathbb{E}(X) = \mathbb{E}(g(Y)) = \int_{100}^{1500} \min(t, 500) \frac{1}{1400} \, dt$$



$$= \left.\begin{aligned} &\int_{100}^{500} t \cdot \frac{1}{1400} \, dt \\ &+ \int_{500}^{1500} 500 \cdot \frac{1}{1400} \, dt \end{aligned}\right\} = \frac{3100}{7} \doteq 442.86.$$