Math 180A: Intro to Probability (for Data Science)

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
  for Lab 2: 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

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

It is also called the mean of $X$, and is often denoted $\mu$.

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

$$P_X(k) = \frac{c}{k(k+1)}, \quad k = 1, 2, 3, \ldots$$

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

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

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

$$E(X) = \sum_{k=1}^{\infty} k \cdot P_X(k) = \sum_{k=1}^{\infty} k \cdot \frac{1}{k(k+1)} = \infty.$$
Expectation of Continuous Random Variable (with density)

Let $X$ be a discrete random variable with $X \in \{x_1, x_2, x_3, \ldots \}$.

- **Discrete Case**: 
  \[ E(X) = \sum_j t_j P(X = t_j) \]
  \[ \sum t \cdot P_X(t) \]

Let $X$ be a continuous random variable (with probability density $f_X(t)$).

- **Continuous Case**: 
  \[ \text{Def: } E(X) = \int_{-\infty}^{\infty} t \cdot f_X(t) \, dt \]

Example: Let $U \sim \text{Unif}(a, b)$.

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

- \[ E(U) = \int_{-\infty}^{\infty} t \cdot f_U(t) \, dt = \int_{a}^{b} t \cdot \frac{1}{b-a} \, dt = \frac{1}{b-a} \int_{a}^{b} t \, dt = \frac{1}{b-a} \left[ \frac{t^2}{2} \right]_{t=a}^{t=b} = \frac{1}{b-a} \left( b^2 - a^2 \right) = \frac{a+b}{2} \]
Question:
Shoot an arrow at a circular target of radius 1. What is the expected distance of the arrow from the center?

(a) $\frac{1}{2}$
(b) $\frac{2}{3}$
(c) $\frac{1}{4}$
(d) $0$

$X = \text{dist. from center}$

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

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

$$E(X) = \int_0^\infty r f_x(r) \, dr$$

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

$$\left( \frac{\text{Area (ID2/3)}}{\text{Area (ID)}} \right) = \frac{\pi (2/3)^2}{\pi (1)^2} = \frac{4}{9}$$
E.g. \( f(t) = \begin{cases} 0 & t \leq 1 \\ \frac{1}{t^2} & t > 1 \end{cases} \) is a probability density.

\[
\int_{-\infty}^{\infty} f(t) \, dt = \int_{1}^{\infty} \frac{1}{t^2} \, dt = \left. -\frac{1}{t} \right|_{1}^{\infty} = \left( \frac{1}{\infty} - \frac{1}{1} \right) = 1 \checkmark
\]

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

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

\[
= \ln |t| \bigg|_{1}^{\infty} = \ln |\infty| - \ln |1| = \infty
\]
Expectations of Functions of Random Variables

\[ \Omega \rightarrow \mathbb{R} \rightarrow \mathbb{R} \]

- random variable \( X \)
- function \( g \)

\[ g(X) = g \circ X \text{ is a new random variable.} \]

E.g., \( X \sim \text{Bin}(n, p) \) is the number of successes in \( n \) trials.

\[ g(X) = \frac{X}{n} \text{ is the proportion of successful trials.} \]

\[ g(x) \in \{0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots, \frac{n}{n} = 1\} \]

\[
E(g(X)) = \sum_t t \cdot P(g(X) = t) = \sum_{k=0}^{n} \frac{k}{n} \cdot P\left(\frac{X}{n} = \frac{k}{n}\right)
\]

In general: for a discrete random variable \( X \),

Proposition: \( E(g(X)) = \sum_s g(s) \cdot P(X = s) \).

Note, by definition, \( E(g(X)) = \sum_t t \cdot P(g(X) = t) \).
Proposition: for a discrete random variable

$$E(g(X)) = \sum_s g(s) P(X=s).$$

Pf. We know

$$\sum_t \{ g(X)=t \} = \bigcup \{ X=s \}_{s: g(s)=t}$$

(disjoint)

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

$$= \sum_t \sum_{s: g(s)=t} g(s) P(X=s)$$
**Proposition:** For a continuous random variable $X$ with probability density $f_X$, \[ E(g(X)) = \int_{-\infty}^{\infty} g(t) f_X(t) \, dt. \]

**Eg.** Let $U$ be a uniform random variable on $[a, b]$. Then \[ E(U^2) = \int_{a}^{b} t^2 f_U(t) \, dt = \int_{a}^{b} t^2 \frac{1}{b-a} \, dt = \frac{1}{b-a} \cdot \frac{1}{3} b^3 - \frac{1}{3} a^3 \]

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

\[ E(U^2) = \frac{b^3 - a^3}{3(b-a)}. \]

(a = 0, b = 1, \( E(U_{[0,1]}^2) = \frac{1}{3} \)).
Eg. Recall the car accident/insurance example. An accident causes $Y$ of damage to your car, where your insurance deductible is $500. What is the expected amount you pay? $Y \sim \text{Unif}([100,1500])$.

What is the expected amount you pay?

$$X = \text{amount you pay}$$

(If $Y$ is a continuous r.v., $X$ is not a continuous r.v. $X$ is neither continuous nor discrete r.v.)

$$\mathbb{E}(X) = \mathbb{E}(g(Y)) = \min(Y, 500) = g(Y)$$

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

$$= \int_{100}^{500} t \cdot \frac{1}{1400} dt \quad \text{and} \quad \int_{500}^{1500} 500 \cdot \frac{1}{1400} dt$$

$$= \frac{3100}{7} \approx 442.86.$$