SDSC2102
Statistical Methods and Data Analysis

Topic 1. Basic Probability and Statistics Theory

Expectation

Examples

- Example 1: ATM takes 1000 photos

  520 pictures have no one in it

  294 pictures have one person at ATM

  186 pictures have two people at ATM

Let X be the # of people at ATM. Find $E[X]$.

$$E(X) = 0 \times P(X = 0) + 1 \times P(X = 1) + 2 \times P(X = 2)$$

$$= 0 \times \frac{520}{1000} + 1 \times \frac{294}{1000} + 2 \times \frac{186}{1000}$$

$$= 0.666$$

Examples

- Example 2: There is a game in which with probability $p(x_i)$ we can win x_i dollars, for $i = 1,2,\ldots,k$. How many dollars can we win per game on average?

Let X be the dollars we can win in a game.

$$E(X) = x_1 \times p(x_1) + x_2 \times p(x_2) \ldots. + x_k \times p(x_k)$$

$$= \sum_{i=1}^{k} x_i p(x_i)$$

Expected Value

- For a discrete r.v. X with PMF $p(x)$

$$\mu = E(X) = \sum_{x} x p(x)$$

- For a continuous r.v. X with PDF $f(x)$

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Class Problems

1. Let X be a discrete random variable with pmf:

p(0) = P{X=0} = 0.2, p(1) = P{X=1} = 0.1, p(2) = P{X=2} = 0.3, p(4) = P{X=4} = 0.2, p(5) = P{X=5} = 0.1, p(6) = P{X=6} = 0.1

Find E(X)

$$\mu = E(X) = \sum_x xp(x)$$

$$= 0 \times 0.2 + 1 \times 0.1 + 2 \times 0.3 + 4 \times 0.2 + 5 \times 0.1 + 6 \times 0.1$$

$$= 2.6$$

Class Problems

2. A total of 4 trucks carrying 148 products from the same factory arrive at a warehouse. The trucks carry, respectively, 40, 33, 25, and 50 products. One of the parts is randomly selected after the products are dumped into the warehouse. Let X denote the number of products that were on the truck carrying this randomly selected product. What is E(X)?

$$P(X = 40) = \frac{40}{148} \quad P(X = 33) = \frac{33}{148}$$

$$P(X = 25) = \frac{25}{148} \quad P(X = 50) = \frac{50}{148}$$

$$E[X] = 40 \times \frac{40}{148} + 33 \times \frac{33}{148} + 25 \times \frac{25}{148} + 50 \times \frac{50}{148}$$

Class Problems

3. For problem 2, pick a truck driver at random. Let Y denote the number of products in his/her truck. Compute E(Y).

$$P(Y = 40) = \frac{1}{4} \qquad P(Y = 33) = \frac{1}{4}$$

$$P(Y = 25) = \frac{1}{4} \qquad P(Y = 50) = \frac{1}{4}$$

$$E[Y] = 40 \times \frac{1}{4} + 33 \times \frac{1}{4} + 25 \times \frac{1}{4} + 50 \times \frac{1}{4}$$

$$= \frac{(40 + 33 + 25 + 50)}{4} = 37$$

Class Problems

4. Compute E(X) if X is a continuous r.v. with pdf:

$$f(x) = \begin{cases} \frac{3}{4}(1-x^2) & -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$= \int_{-\infty}^{-1} x \cdot 0dx + \int_{-1}^{1} x \cdot \frac{3}{4}(1 - x^2)dx + \int_{1}^{+\infty} x \cdot 0dx$$

$$= 0$$

Class Problems

5. If the CDF of a continuous r.v. is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ ax^2+bx+c & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

If E(X) = 3/5, find a, b and c.

Three unknowns: a, b, c

$$F(0) = a \times 0^2 + b \times 0 + c = 0 \Rightarrow c = 0$$

$$F(1) = a \times 1^2 + b \times 1 + c = 1 \Rightarrow a + b = 1$$

Class Problems

Find PDF:

$$f(x) = \begin{cases} 0, & x < 0 \\ 2ax + b, & 0 \le x \le 1 \\ 0, & x > 1 \end{cases}$$

$$\Rightarrow a = \frac{3}{5}, b = \frac{2}{5}$$

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

$$= \int_{-\infty}^{0} x \cdot 0 dx + \int_{0}^{1} x \cdot (2ax + b) dx + \int_{1}^{\infty} x \cdot 0 dx$$

$$= \left(\frac{2}{3} a x^3 + \frac{b}{2} x^2 \right) \Bigg|_0^1 = \frac{2}{3} a + \frac{b}{2} = \frac{3}{5}$$

Expected Value of A Function of X

- For a discrete r.v. X with PMF $p(x)$, the expected value for a function $g(X)$ is

$$E[g(X)] = \sum_{x} g(x)p(x)$$

- For a continuous r.v. X with PDF $f(x)$, the expected value for a function $g(X)$ is

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

- The distribution for $g(X)$ is not needed.

Examples

- Discrete Example: Flip 2 coins, X = #heads

$$p(0) = 1/4, \quad p(1) = 1/2, \quad p(2) = 1/4$$

$$E\left[\frac{3}{X+1}\right] = \left(\frac{3}{0+1}\right)\left(\frac{1}{4}\right) + \left(\frac{3}{1+1}\right)\left(\frac{1}{2}\right) + \left(\frac{3}{2+1}\right)\left(\frac{1}{4}\right) = \frac{7}{4}$$

- Continuous Example:

$$f(x) = 2(x-1) \quad \text{for} \quad 1 < x < 2; \quad \text{o/w} \quad f_X(x) = 0$$

$$E[X^2] = \int_1^2 \left(x^2\right) 2(x-1)\,dx = 2\int_1^2 x^3 - x^2\,dx = \frac{17}{6}$$

Variance

- For any r.v. X, the variance is the expected squared deviation about its mean:

$$\sigma^2 = Var(X) = E\left[(X-\mu)^2\right] = E\left[(X - E[X])^2\right]$$

  - Note that variance cannot be negative.

$$\begin{aligned} E\left[(X - E[X])^2\right] &= E[X^2 - 2XE[X] + (E[X])^2] \\ &= E[X^2] - 2E[X] \cdot E[X] + (E[X])^2] \\ &= E[X^2] - (E[X])^2] \end{aligned}$$

$$Var(X) = E[X^2] - (E[X])^2]$$

Linear Combinations

- a, b are constant
- Basic rules of expected value
  - $E[a] = a$
  - $E[aX] = a\,E[X]$
  - $E[aX + b] = a\,E[X] + b$

- Basic rules of variance
  - $V(a) = 0$
  - $V(aX) = a^2\,V(X)$
  - $V(aX + b) = a^2\,V(X)$

Class Problems on Expected Value of a Function

1. A newspaper boy buys newspapers each worth \$0.35 from a distributor. He sells each newspaper for \$0.50 to customers. Let X be a r.v. denoting the demand for newspapers in a day, with pmf: $p(0) = P\{X=0\} = 0.1$, $p(1) = P\{X=1\} = 0.2$, $p(2) = P\{X=2\} = 0.3$, $p(3) = P\{X=3\} = 0.4$,
   The newspaper boy buys 3 newspapers at the beginning of each day. Assume that people do not buy old newspapers.

   a. Write down the daily profit as a function of X; g(X)
   b. Compute the expected profit, E[g(X)]

X = Demand for newspapers

(a) Daily lost: $\$0.35 \times 3 = \$1.05$

Daily income: $\$0.50 * X$

Daily profit: $g(X) = 0.5X - 1.05$

(b) $E[g(X)] = E[0.5X - 1.05]$

Class Problems

Method 1:

$$E[g(X)] = \sum_{x=0}^{3} (0.5x - 1.05)p(x)$$

$$= (0 - 1.05) * 0.1 + (0.5 - 1.05) * 0.2 + (1 - 1.05) * 0.3 + (1.5 - 1.05) * 0.4$$

$$= -0.05$$

Method 2:

$$E[g(X)] = E[0.5X - 1.05] = 0.5E(X) - 1.05$$

$$E[X] = 0 * 0.1 + 1 * 0.2 + 2 * 0.3 + 3 * 0.4 = 2$$

$$E[g(X)] = 0.5 * 2 - 1.05 = -0.05$$

Class Problems

2. In the above problem, what is the variance of the demand for newspapers?

$$Var(X) = E[X^2] - (E[X])^2]$$

$$E\left[X^2\right] = 0^2 * 0.1 + 1^2 * 0.2 + 2^2 * 0.3 + 3^2 * 0.4 = 5$$

$$Var(X) = 5 - 2^2 = 1$$

Class Problems

3. If X is a continuous r.v. denoting the size of a document at a web server with pdf

$$f(x) = \begin{cases} 24/x^4 & x \geq 2 \\ 0 & \text{otherwise} \end{cases}$$

Compute the variance of X.

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_{2}^{\infty} x \cdot \frac{24}{x^4} dx = 3$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_{2}^{\infty} x^2 \cdot \frac{24}{x^4} dx = 12$$

$$Var(X) = E[X^2] - (E[X])^2] = 12 - 3^2 = 3$$

Class Problems

4. If the time to download a document, T, is a r.v. such that T = aX + b (√X) + c for some constants a, b, and c, compute the average time to download a document, E(T). Use the pdf of X from problem 3 above.

$$E[T] = E\left[aX + b\sqrt{X} + c\right] = aE[X] + bE\left[\sqrt{X}\right] + c$$

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_{2}^{\infty} x \cdot \frac{24}{x^4} dx = 3$$

$$E\left[\sqrt{X}\right] = \int_{2}^{\infty} \sqrt{x} \cdot \frac{24}{x^4} dx = 9.6 \cdot 2^{-2.5}$$

$$E[T] = 3a + 9.6b \cdot 2^{-2.5} + c$$

Class Problems

5. If X and Y are random variables with X = 3Y + 16 and E(Y) = 12 and Var(Y) = 4, the compute E(X) and Var(X)

$$E[X] = E[3Y + 16] = 3E[Y] + 16 = 3 * 12 + 16 = 52$$

$$Var[X] = Var[3Y + 16] = 3^2 Var(Y) = 9 * 4 = 36$$

Binomial Distribution

- PMF

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0,1,2,\dots,n$$

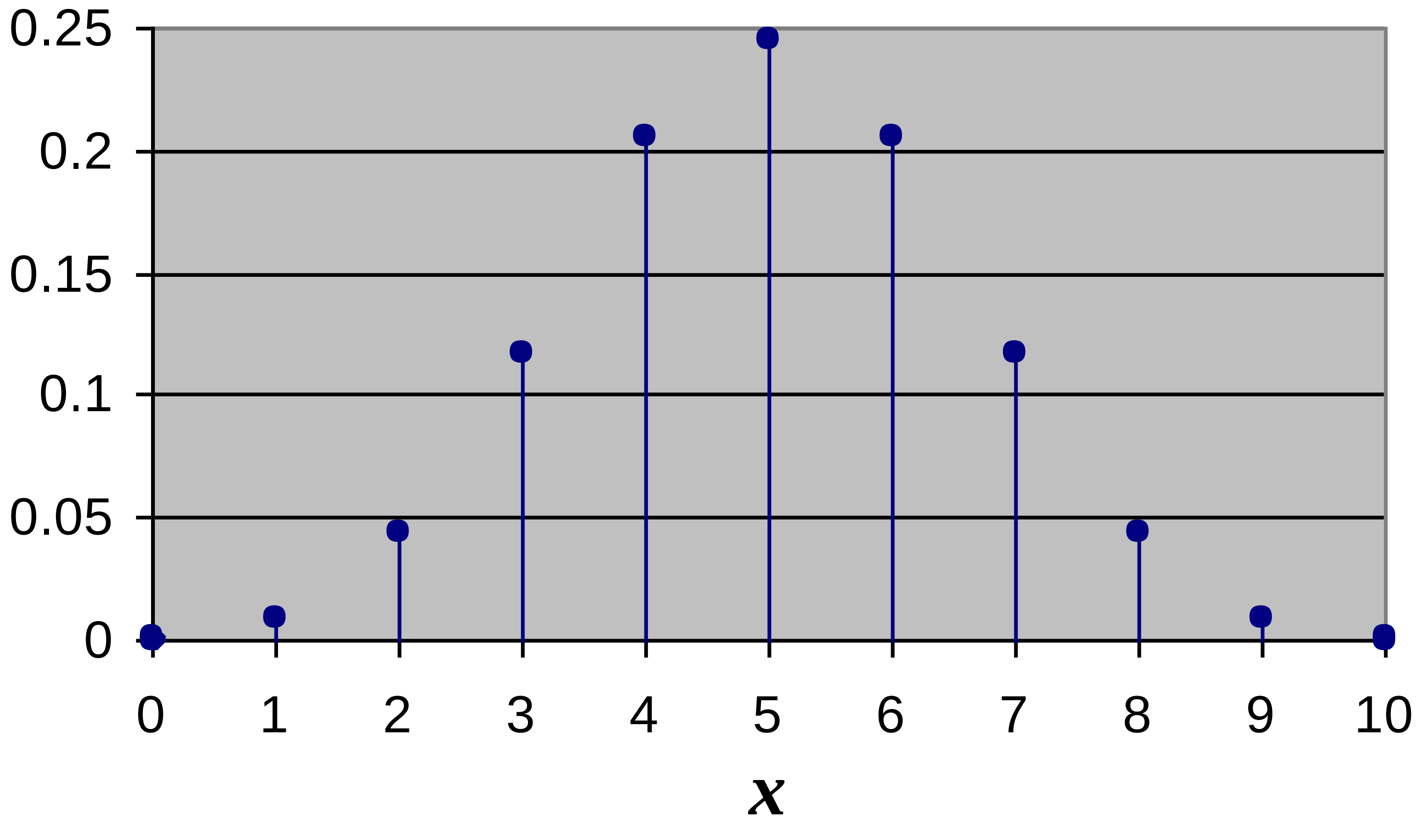
- $E(X) = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x} = np$

$$E\left[X^2\right] = \sum_{x=0}^{n} x^2 \cdot \binom{n}{x} p^x (1-p)^{n-x} = n^2p^2 - np^2 + np$$

- $Var(X) = E[X^2] - [E(X)]^2 = np(1-p)$

Binomial Distribution

- Example: $n = 10$, $p = 0.5$



$$E[X] = np = 10 * 0.5 = 5$$

$$Var(X) = np(1 - p) = 10 * 0.5 * 0.5 = 2.5$$

Geometric Distribution

- PMF

$$P(X = x) = (1-p)^{x-1}p \quad \text{for } x = 1,2,\dots,$$

- $E(X) = \sum_{x=1}^{\infty} x \cdot (1-p)^{x-1}p = \frac{1}{p}$

$$E\left(X^2\right) = \sum_{x=1}^{\infty} x^2 \cdot (1-p)^{x-1}p = \frac{2-p}{p^2}$$

- $Var(X) = E\left(X^2\right) - [E(X)]^2 = \frac{1-p}{p^2}$

Geometric Distribution

- Example: Roll a die until we see a "6"
  - 1 = "6", 0 = "not a 6"
  - $p = P(6) = 1/6$

$$E[X] = \frac{1}{1/6} = 6 \text{ trials}$$

$$Var[X] = \frac{\frac{5}{6}}{\left(\frac{1}{6}\right)^2} = 30$$

Poisson Distribution

- PMF

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{for } x = 0,1,2,\ldots$$

- $E(X) = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^x}{x!} = \lambda$

$$E(X^2) = \sum_{x=0}^{\infty} x^2 \cdot \frac{e^{-\lambda}\lambda^x}{x!} = \lambda^2 + \lambda$$

- $Var(X) = E(X^2) - [E(X)]^2 = \lambda$

Exponential Distribution

- PDF

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

- $E(X) = \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{\lambda}$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \frac{2}{\lambda^2}$$

- $Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2}$

Exponential Distribution

➢ Example: A battery is expected to last for 500 hours on average. Its lifetime follows an exponential distribution.

Let X be the lifetime of the battery ~ Exp (λ)

$$\frac{1}{\lambda} = 500 \Rightarrow \lambda = \frac{1}{500}$$

$$X \sim Exp\left(\frac{1}{500}\right)$$

Uniform Distribution

- PDF

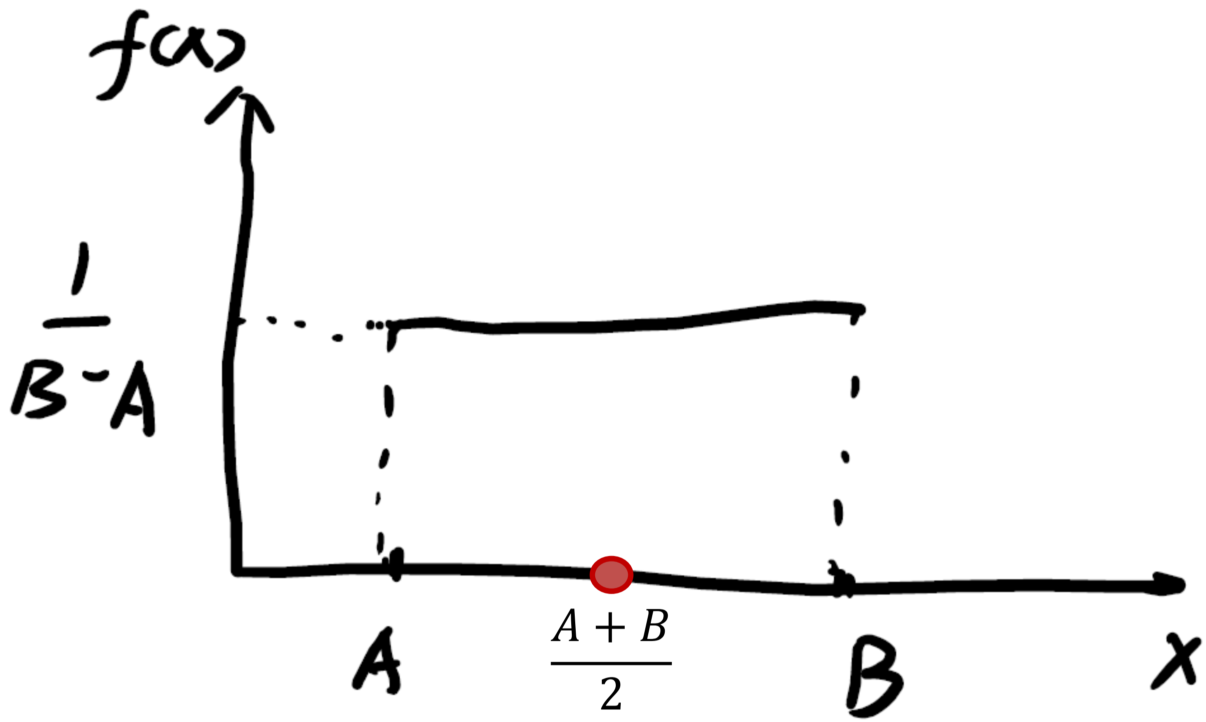
$$f(x) = \begin{cases} 0, & x < A \\ \dfrac{1}{B - A}, & A \leq x \leq B \\ 0, & x > B \end{cases}$$

- $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \frac{A+B}{2}$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \frac{A^2+AB+B^2}{3}$$

- $Var(X) = E(X^2) - [E(X)]^2 = \frac{(B-A)^2}{12}$

Uniform Distribution



Standard Normal Distribution

- PDF

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

- $E(Z) = \int_{-\infty}^{\infty} z f(z) dz = 0$

$$E\left(Z^2\right) = \int_{-\infty}^{\infty} z^2 f(z) dz = 1$$

- $Var(Z) = E\left(Z^2\right) - [E(Z)]^2 = 1$

Normal Distribution

- PDF

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$Z = \frac{X - \mu}{\sigma} \Rightarrow X = \mu + Z\sigma$$

- $E(X) = E[\mu + Z\sigma] = \mu + \sigma E[Z] = \mu$
- $Var(X) = Var[\mu + Z\sigma] = \sigma^2$

Gamma Distribution

- PDF

$$f(x) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

- $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \alpha\beta$
- $Var(X) = \alpha\beta^2$