Math 3070, Applied Statistics

Section 1

September 6, 2019

Lecture Outline, 9/6

Section 3.3

- Bernoulli Random Variable
- Expected Value of X
- Variance of X
- Linear Transformations and Examples

Bernoulli Random Variable, Definition

A Bernoulli Random Variable X with a PMF of

$$f(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \\ 0, & \text{otherwise} \end{cases}$$

is denoted as $X \sim \text{bern}(p)$, a Bernoulli random variable with parameter p.

Note, parameters are variables that determine a random variable.

Expected Value of X, Definitions

The **expected value** or **mean value** of a discrete random variable X is

$$E(X) = \mu_x = \sum_{x \in D} x \cdot p(x)$$

where $p(x)$ is the PMF of X and D is the all possible values of X.

The **expected value** or **mean value** of a function of a random variable, also a random variable, $h(X)$ is

$$E[h(x)] = \sum_{x \in D} h(x) \cdot p(x).$$

- Often used to describe distributions of random variables. Their estimation is key to generating probabilistic models.
- Notice that that the mean value is a fixed number from a fixed distribution while the sample mean changes with the sample.

Expected Value of X, Examples

Consider X, a random variable that takes the values of $\pi/2$ one fourth of the time and π the remaining time. Compute the PMF. Compute the expected value of X and $\sin(X+\pi)$.

$$f(x) = \begin{cases} 1/4, & x = \pi/2 \\ 3/4, & x = \pi \\ 0, & \text{otherwise} \end{cases}$$

$$E(X) = \frac{\pi}{2}\frac{1}{4} + \pi\frac{3}{4} = \pi\frac{7}{8}$$

$$\begin{aligned} E[\sin(X+\pi)] &= \sin(\pi/2+\pi)\frac{1}{4} + \sin(\pi+\pi)\frac{3}{4} \\ &= (-1)\frac{1}{4} + 0\frac{3}{4} = -\frac{1}{4} \end{aligned}$$

Expected Value of X, Examples

Consider X to be a bernoulli random variable, takes either 0 or 1. Given that $E(X) = 0.4$, find the PMF.

$$f(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \\ 0, & \text{otherwise} \end{cases}$$

$$0.4 = E(X) = 1 \cdot p + 0 \cdot (1 - p) = p$$

$$p = 0.4$$

$$f(x) = \begin{cases} 0.4, & x = 1 \\ 0.6, & x = 0 \\ 0, & \text{otherwise} \end{cases}$$

Expected Value of X, Comments and Questions

- $E(X) = \mu_x = \sum_{x \in D} x \cdot p(x)$
- $E[h(x)] = \sum_{x \in D} h(x) \cdot p(x)$
- $h(X)$ is a random variable.
- Parameters of random variables can be estimated using techniques based on expected values of functions of random variables. For example, sample mean estimates expected value and sample variance estimates variance.
- Estimators are random. Parameters and expectations are not.

Variance of X, Definition

The **variance** of a discrete random variable X is

$$V(X) = \sigma_x^2 = \sum_{x \in D} (x - \mu)^2 \cdot p(x)$$

where $p(x)$ is the PMF of X, $\mu = E(X)$, and D is the all possible values of X.

The **standard deviation** of X is

$$\sigma_x = \sqrt{\sigma_x^2}$$

Note,

$$\sigma_x^2 = E[(X - \mu)^2]$$

Since $V(X)$, sums over nonnegative quantities, $V(X) \geq 0$.

Variance of X, Shortcut

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

proof:

$$\begin{aligned}
V(X) &= E[(X-\mu)^2] \\
&= E[X^2 - 2\mu X + \mu^2] \\
&= \sum_{x \in D} X^2 p(x) - \sum_{x \in D} 2\mu X p(x) + \sum_{x \in D} \mu^2 p(x) \\
&= \sum_{x \in D} X^2 p(x) - 2\mu \sum_{x \in D} X p(x) + \mu^2 \sum_{x \in D} p(x) \\
&= E(X^2) - 2\mu\mu + \mu^2 = E(X^2) - 2E(X)^2 + E(X)^2 \\
&= E(X^2) - [E(X)]^2
\end{aligned}$$

Variance of X, Example

Consider X, the value of a fair die roll. Compute the variance and standard deviation.

$$E(X) = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{21}{6}$$

$$E(X^2) = 1^2\frac{1}{6} + 2^2\frac{1}{6} + 3^2\frac{1}{6} + 4^2\frac{1}{6} + 5^2\frac{1}{6} + 6^2\frac{1}{6} = \frac{91}{6}$$

$$\sigma^2 = E(X^2) - E(X)^2 = \frac{91}{6} - \left(\frac{21}{6}\right)^2 \approx 2.9166$$

$$\sigma = \sqrt{\sigma^2} \approx 1.707$$

You can check for yourself that the direct calculation is longer.

Variance of X, Comments and Questions

- $\sigma^2 = Var(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2$
- $\sigma = \sqrt{\sigma^2}$, standard deviation
- $V(X) \geq 0$ and $\sigma_X \geq 0$

Questions?

Linear Transformations, Explanation

Linear transformation of X scale μ_X, σ_x^2 and σ_x as follows.

$$E(aX + b) = a \cdot E(X) + b = a\mu_x + b$$

$$Var(aX + b) = a^2 Var(X) = a^2\sigma_x^2$$

$$\sigma_{aX+b} = |a|\sigma_X$$

Same scaling laws as their sample versions.

Linear Transformations, Mean Explanation

$$E(aX + b) = a \cdot E(X) + b$$

proof

$$\begin{aligned} E(aX + b) &= \sum_{x \in D} (ax + b) p(x) \\ &= a \sum_{x \in D} x p(x) + b \sum_{x \in D} p(x) \\ &= a \cdot E(X) + b \end{aligned}$$

Linearity of Expected Value, Explanation

Expected values are linear.

$$E[h(X) + ag(X) + b] = E[h(X)] + a \cdot E[h(Y)] + b$$

proof

$$\begin{aligned} E[h(X) + ag(X) + b] &= \sum_{x \in D} (h(X) + ag(X) + b)p(x) \\ &= \sum_{x \in D} h(X)p(x) + a \sum_{x \in D} g(X)p(x) + b \sum_{x \in D} p(x) \\ &= E[h(x)] + a \cdot E[g(x)] + b \end{aligned}$$

Can be useful but not in book.

This also work with two random variables, but we will get this this case later.

Linear Transformations, Variance Explanation

$$Var(aX + b) = a^2 Var(X)$$

proof

$$\begin{aligned} Var(aX + b) &= E[(aX + b)^2] - (E[(aX - b)])^2 \\ &= E[a^2X^2 - 2abX + b^2] - (aE(X) - b)^2 \\ &= a^2E[X^2] - 2abE[X] + b^2 - (a^2E(X)^2 - 2abE(X) - b^2) \\ &= a^2E[X^2] - a^2E(X)^2 \\ &= a^2 Var(X) \end{aligned}$$

Consequently,

$$\sigma_{aX+b} = |a|\sigma_X$$

Linear Transformations, Example

Consider X to be the price of a used car. The expected value is $E[X] = 4,000$ and $V(X) = 200$. As function of the inital price X, yearly maintaince cost is $h(X) = 0.1X + 100$. Determine the expected value and standard deviation of the yearly maintaince.

Use the linear transformation formulas.

$$E[h(X)] = E[0.1X + 100] = 0.1E[X] + 100 = 500$$

$$V(h(X)) = V(0.1X + 100) = 0.1^2 V(X) = 2$$

Linear Transformations, Non-example

Consider X to be the price of a used car. The expected value is $E[X] = 4,000$ and $E(X^2) = 2,000$. Why is the last sentence false?

$$V(X) = E[X^2] - (E[X])^2 = 2000 - 4000^2 = -15998000 < 0$$

Linear Transformations, $E(X^2)$ from $V(X)$ example

Consider X to be the price of a used car. The expected value is $E[X] = 4,000$ and $V(X) = 200$. The yearly insurance cost is modeled as $g(X) = X - 0.0002X^2$ Determine the expected value of the yearly insurance costs in this model.

Use linearity of expected value.

$$E[g(X)] = E[0.3X - 0.01X^2] = E[X] - 0.0002E[X^2]$$

$$V(X) = E[X^2] - E[X]^2 \rightarrow E[X^2] = V(X) + E[X]^2$$
$$E[X^2] = 200 + 4000^2 = 16000200$$

$$E[g(X)] = 4000 - 0.0002 \cdot 16000200 = 799.96$$

Summary

- $E[aX + b] = a \cdot E[X] + b$
- $V(aX + b) = a^2 \cdot V(X)$ and $\sigma_{aX+b} = |a|\sigma_X$
- $E[g(X) + ah(X) + b] = E[g(X)] + a \cdot E[h(X)] + b$

Questions?