Lecture Notes EMSE 4765:
DATA ANALYSIS - Probability Review

Chapter 7: Expectation and Variance

Version: 01/18/2021

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

Text Book: A Modern Introduction to Probability and Statistics, Understanding Why and How

By: F.M. Dekking. C. Kraaikamp, H.P.Lopuhaä and L.E. Meester

7 Expectation and Variance

7.1 Expected Values - Discrete Random Variables . . .

- **Random variables are complicated objects**, **containing a lot of information** on **the uncertainty** that are modeled by them. Typically, random variables are **summarized** by two numbers:

The expected value: also called **the expectation or mean**, gives the center - **in the sense of average value** - of the distribution of the random variable.

The variance: a measure of spread of the distribution of the random variable.

Definition: The expectation of a discrete random variable X taking **the outcomes** $a_1, a_2, \ldots$ and with probability mass function p is the number :

$$E[X] = \sum_i a_i Pr(X = a_i) = \sum_i a_i p(a_i)$$

$E[X]$ is called the expected value or mean of X, or its distribution

7 Expectation or The Mean

7.1 Expected Values - Discrete Random Variables ...

Die Experiment: $X \equiv$ "Outome of the Die"

$$E[X] = \sum_{x=1}^{6} x Pr(X = x) = \sum_{x=1}^{6} \frac{x}{6} = 3.5$$

Coin Toss experiment:

$X \equiv$ "Number of Times Heads in a Series or Sequence of 10 coin tosses"

$$E[X] = \sum_{x=0}^{10} x Pr(X = x) = \sum_{x=1}^{10} x \binom{10}{x} \left(\frac{1}{2}\right)^{x} \left(\frac{1}{2}\right)^{10-x} = 5$$

Definition: The expectation of **a Binomial distribution**.

$$X \sim Bin(n, p) \Rightarrow E[X] = \sum_{k=0}^{n} k Pr(X = k) = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np$$

7 Expectation or The Mean

7.1 Expected Values - Discrete Random Variables ...

Lottery experiment:

$$X \equiv \text{The number of weeks until success} \sim Geo(p), \text{ where } p = 10^{-4}.$$

Definition: The expectation of **a geometric distribution**.

$$X \sim Geo(p) \Rightarrow E[X] = \sum_{k=1}^{\infty} k \times Pr(X = k) = \sum_{k=1}^{\infty} k(1-p)^{k-1} p = \frac{1}{p}$$

Conclusion: If you buy a lottery ticket every week and you have a chance of **1 in 10,000 of winning the jackpot**, what is the **expected number of weeks** you have to buy tickets before you get the jackpot? **Answer: 10,000 weeks** **(≈ Two centuries).**

$$\sum_{k=1}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2} \Rightarrow p \times \sum_{k=1}^{\infty} k(1-p)^{k-1} = p \times \frac{1}{p^2} = \frac{1}{p}$$

7 Expectation or The Mean

7.1 Expected Values - Continuous Random Variables . . .

Definition: The **expectation**, **expected value** or **mean** of a continuous random variable X is the number :

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

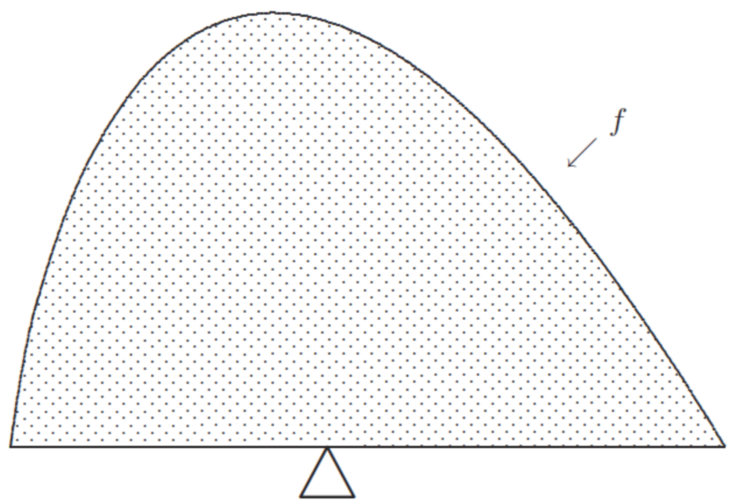

Fig. 7.2. Expected value as center of gravity, continuous case.

7 Expectation or The Mean

7.1 Expected Values - Continuous Random Variables . . .

Excercise: Compute the expectation of a random variable U that is uniformly distributed over $[2, 5]$.

Answer: $f(u) = \frac{1}{3}$, $u \in [2, 5]$ and 0 elsewhere. Hence,

$$E[U] = \int_2^5 u \cdot \frac{1}{3} du = \left[\frac{1}{3} \cdot \frac{1}{2} u^2\right]_2^5 = \frac{1}{6}\left[u^2\right]_2^5 = \frac{25}{6} - \frac{4}{6} = \frac{21}{6} = 3\frac{1}{2}.$$

which is **the balancing point $(2 + 5)/2$** !

Definition: The expectation of **a uniform distribution on $[a, b]$**.

$$X \sim U[a, b] \Rightarrow E[X] = \int_{x=a}^{b} \frac{1}{b-a} dx = \frac{a+b}{2}$$

Definition: The expectation of **an exponential distribution**.

$$X \sim Exp(\lambda) \Rightarrow E[X] = \int_{x=0}^{\infty} x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

Definition: The expectation of **a normal distribution**.

$$X \sim N(\mu, \sigma^2) \Rightarrow E[X] = \int_{x=-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = \mu$$

Example: $X \sim U[0, 10] \Rightarrow f_X(x) = 1/10 \,\text{if}\, X \in [0, 10],\ Y = X^2 \in [0, 100]$

$$E[Y] = E[X^2] = \int_0^{10} x^2 f_X(x) dx = \int_{x=0}^{10} x^2 \frac{1}{10} dx = \Big[\frac{1}{10}\frac{1}{3}x^3\Big]_0^{10} = 33\frac{1}{3}$$

Definition: The change of variable formula. Let X be a rv and $g: \mathbb{R} \to \mathbb{R}$.

X discrete on $a_1, \ldots, a_n \Rightarrow E[g(X)] = \sum_i g(a_i) Pr(X = a_i) = \sum_i g(a_i) p(a_i)$

X continuous, $X \sim f(\cdot) \Rightarrow E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$

Suppose $g(x) = ax + b,\ a, b \in \mathbb{R}$ and X is a continuous RV.

$$
\begin{aligned}
E[g(X)] = E[aX + b] &= \int_{-\infty}^{\infty} g(x) f(x) dx = \int_{-\infty}^{\infty} \boldsymbol{(ax + b)} f(x) dx \\
&= \int_{-\infty}^{\infty} \boldsymbol{ax} f(x) + \boldsymbol{b} f(x) dx = \boldsymbol{a} \int_{-\infty}^{\infty} \boldsymbol{x} f(x) dx + \boldsymbol{b} \int_{-\infty}^{\infty} f(x) dx \\
&= \boldsymbol{a} E[X] + \boldsymbol{b} \times 1 = \boldsymbol{a} E[X] + \boldsymbol{b}
\end{aligned}
$$

Linearity of Expectation! Same applies when X is a discrete RV!

7.4 Variance and standard deviation...

- Suppose you are offered an opportunity for an investment **at the cost of \$500** whose **expected return or payoff** is \$500. Seems an OK opportunity!

What if we have for payoff $Y_1 : Pr(Y_1 = \$450) = 50\%, Pr(Y_1 = \$550) = 50\%$?
What if we have for payoff $Y_2 : Pr(Y_2 = \$0) = 50\%, Pr(Y_2 = \$1000) = 50\%$?

Clearly, **the spread (around the mean)** makes you feel different about them. Usually this is measured by **the expected squared distance from the mean**.

Definition: The variance $\boldsymbol{Var(X)}$ of a random variable X is the number **:**

$$Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2 \geq 0$$

- **Definition: Standard Deviation $\boldsymbol{X \equiv \sqrt{Var(X)}}$ (same units as $\boldsymbol{E[X]}$)**

Exercise: Calculate the mean, variance and standard deviation for Y_1 and Y_2

$$\text{Payoff } Y_1 : Pr(Y_1 = \$450) = 50\%, Pr(Y_1 = \$550) = 50\% \text{ ?}$$
$$\text{Payoff } Y_2 : Pr(Y_2 = \$0) = 50\%, Pr(Y_2 = \$1000) = 50\% \text{ ?}$$

Answer:

$$E[Y_1] = 450 \cdot Pr(Y_1 = \$450) + 550 \cdot Pr(Y_1 = \$550) = \frac{450 + 550}{2} = \$500$$
$$E[Y_2] = 0 \cdot Pr(Y_1 = \$0) + 1000 \cdot Pr(Y_1 = \$1000) = \frac{0 + 1000}{2} = \$500$$

$$Var(Y_1) = (450 - 500)^2 \cdot Pr(Y_1 = \$450) + (550 - 500)^2 \cdot Pr(Y_1 = \$550)$$
$$= \frac{2500 + 2500}{2} = 2500 \Rightarrow St.Dev(X) = \$50.$$
$$Var(Y_2) = (0 - 500)^2 \cdot Pr(Y_1 = \$0) + (1000 - 500)^2 \cdot Pr(Y_1 = \$1000)$$
$$= \frac{250000 + 250000}{2} = 250000 \Rightarrow St.Dev(X) = \$500.$$

7 Variance

7.4 Variance and standard deviation. . .

Definition: The variance of **a Bernouilli distribution**.

$$X \sim Bern(p) \Rightarrow Var(X) = p(1-p)$$

Definition: The variance of **a Binomial distribution**.

$$X \sim B(n,p) \Rightarrow Var(X) = np(1-p)$$

Definition: The variance of **a Geometric distribution**.

$$X \sim G(p) \Rightarrow Var(X) = (1-p)/p^2$$

7 Variance

7.4 Variance and standard deviation...

Definition: The variance of **a uniform distribution**.

$$X \sim U(a,b) \Rightarrow Var(X) = (b-a)^2/12$$

Definition: The variance of **an exponential distribution**.

$$X \sim E(\lambda) \Rightarrow Var(X) = 1/\lambda^2$$

Definition: The variance of **a normal distribution**.

$$X \sim N(\mu,\sigma^2) \Rightarrow Var(X) = \sigma^2$$

Expectation and Variance under a change of units: For any random random variable X and any real numbers a and b **:**

$$E[aX+b] = aE[X]+b \qquad Var(aX+b) = a^2 Var(X)$$

Without calculation, why is $Var(X)$ not affected above by b?

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