STAT 230 Chapter 7: Expected Value and Variance.

7.1 Summarizing Data on Random Variables

Defn The median of a sample is a value such that half the results are below it and half above it, when the results are arranged in numerical order.

Defn The mode of the sample is the value which occurs most often. there is no guarantee there will be only a single mode.

7.2 Expectation of a Random Variable

Defn Let $X$ be a discrete random variable with range $(X) = A$ and pf $f(x)$. The Expected value (also called the mean or the expectation) of $X$ is given by $\boxed{E(x) = \sum_{x \in A} x f(x)}$

thm Let $X$ be a discrete random variable with range $(X) = A$ and pf $f(x)$. The expected value of some function $g(X)$ of $X$ is given by

$$\boxed{E[g(X)] = \sum_{x \in A} g(x) f(x)}$$

notes: ① you can interpret $E[g(X)]$ as the average value of $g(X)$ in an infinite series of repetitions of the process where $X$ is defined.

② the average value of $g(X)$ may be a value which $g(X)$ never takes

③ $\mu$ and $E[X]$ mean the same thing. $\mu$ makes it visually clearer that the expected value is NOT a random variable like $X$ but a non-random const

④ when calculating expectations. look at your answers to be sure it makes sense

Linearity Property of Expectation

- For constants $a$ and $b$.

  $\boxed{E[ag(X)+b] = aE[g(X)] + b}$

- Similarly for constants $a$ and $b$ and two functions $g_1$ and $g_2$:

  $\boxed{E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(X)]}$

- for $g(X)$ a non-linear function it is not generally true that $E[g(X)] = g(E[X])$

## 7.3 Some Applications of Expectation

ex. Expected Winning in a Lottery

A small lottery sells 1000 tickets numbered 000, 001 ⋯ 999

The tickets cost \$10 each. When all the tickets have been sold the draw takes place: this consists of a single ticket from 000 to 999 being chosen at random. For ticket holders the prize structure is as follows:

- your ticket is drawn - win \$5000
- your ticket has the same first 2 number as the winning ticket, but the third is different - win \$100
- your ticket has the same first number as the winning ticket, but the second number is different - win \$10
- all other cases - win nothing

Let the r.v. $X$ represents the winning from a given ticket, find $E(X)$

Soln The possible values for $X$ are $0, 10, 100, 5000$ (dollars)

First, we need to find the pf for $X$.

We find that $f(x) = P(X=x)$ has values:

$$f(0)=0.9 \quad f(10)=0.09 \quad f(100)=0.009 \quad f(5000)=0.001$$

Then $E[X] = \sum x f(x) = \$6.80$, net winning is negative $(-\$3.20)$

remark consider $X=3$ all digits match

$X=2$ $1^{st}$ 2 digits match

$X=1$ $1^{st}$ match

$X=0$ no match

Now, define the function $g(X)$ as the winnings when the outcome $X=x$ occurs.

$$g(0)=0 \quad g(1)=10 \quad g(2)=100 \quad g(3)=5000$$

Then $E[g(X)] = \sum g(x) f(x) = \$6.80$ the same as before.

ex. Diagnostic Medical Tests: Suppose we have two cheap tests and one expensive test.

Let a person be chosen at random, and let $D = \{\text{person has the condition}\}$

The three tests are: • $P(\text{positive test} \mid \bar{D}) = 0.05$ \$5.00

$P(\text{positive test} \mid \bar{D}) = 0.03$ \$8.00

$P(\text{positive test} \mid \bar{D}) = 0$ \$40.00

we want to check a large number of people for the condition, and have to choose among three testing strategies:

i) Use test 1, followed by Test 3 if positive.

ii) Use test 2, followed by Test 3 if positive

iii) Use test 3.

Determine the expected cost per person under each strategies i ii and iii

We will then choose the strategy with the lowest expected cost.

It is known that about 0.001 of the population have the condition. $P(D) = 0.001$

Sol: Define r.v. $X$ :

$X = 1$ if the initial test is negative.

$X = 2$ if the initial test is positive.

Also, let $g(x)$ be the total cost of testing the person.

Then, $E[g(X)] = \sum_{1}^{2} g(x) f(x)$

The pf $f(x)$ for $X$ and $g(x)$ differ from strategies.

i) $P(X=2) = P(\text{init test is neg})$

$= P(D) + P(\text{positive test} \mid \bar{D}) \cdot P(\bar{D})$

$= 0.0510 \quad \therefore f(2) = 0.0510$

$f(1) = P(X=1) = 1 - f(2) = 1 - 0.0510 = 0.949$

$g(1) = 5 \quad g(2) = 45 \quad \therefore E[g(X)] = 5(0.949) + 45(0.0510) = \$7.04$

ii) $f(2) = 0.001 + (0.03)(0.999) = 0.03097$

$f(1) = 1 - f(2) = 0.96903$

$g(1) = 8 \quad g(2) = 48$

$E[g(X)] = 8(0.96903) + 48(0.03097) = \$9.2388$

iii) $f(2) = 0.001 \quad f(1) = 0.999$

$g(0) = g(1) = 40$

$E[g(X)] = \$40.00$

$\therefore$ it is cheapest to use strategy (i)

## 7.4 Means and Variances of Distributions

ex. Expected value of Bin rv

Let $X \sim Bin(n,p)$ Find $E[X]$

soln $\boxed{\mu = E(X) = np}$

ex. Expected value of Poi rv

Let $X \sim Poi(\lambda t)$, find $\mu = E(X)$

soln $\boxed{\mu = E(X) = \lambda t}$

| | |
|---|---|
| for Hyper Geo | $\mu = E(X) = \frac{nr}{N}$ |
| for NB | $\mu = E(X) = \frac{k(1-p)}{p}$ |

Defn The variance of a r.v. $X$ is $E[(X-\mu)^2]$, and is denoted by $Var(X)$ or by $\sigma^2$

In words, the variance is the average square of the distance from the mean.

property:
- $\boxed{Var(X) = E(X^2) - \mu^2}$
- $\boxed{Var(X) = E[X(X-1)] + \mu - \mu^2}$

defn: $\boxed{\text{The standard deviation of r.v. } X \text{ is } \sigma = sd(X) = \sqrt{Var(X)} = \sqrt{E[(X-\mu)^2]}}$

ex. Suppose $X$ is a r.v. with prob.

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $f(x)$ | 0.07 | 0.1 | 0.12 | 0.13 | 0.16 | 0.13 | 0.12 | 0.1 | 0.07 |

Find $E[X]$ and $Var(X)$

soln: $\mu = E(X) = 1(0.7) + 2(0.1) + 3(0.12) + \cdots + (9)(0.07) = 5$

Without doing any calculations we also know that $Var(X) = \sigma^2 \leq 16$

This is because the possible values of $X$ are $\{1, 2 \cdots 9\}$ and so the max possible value for $(X-\mu)^2$ is $(9-5)^2 = 16 = (1-5)^2$

$E(X^2) = (\sum f(x) \cdot x^2 = 1^2 \cdot 0.07 + 2^2 \times 0.1 + \cdots \; 9^2 \times 0.07$

$= 30.26$

$Var(X) = E(X^2) - \mu^2 = 30.26 - 5^2 = 5.25$.

and $sd(X) = \sqrt{Var(X)} = 2.2935$

ex. Var of Bin r.v.

Let $X \sim Bin(n,p)$, find $Var(X)$

soln $\boxed{Var(X) = np(1-p)}$

ex. Var of Poi r.v.

Let $X \sim Poi(\mu)$

soln $\boxed{Var(X) = \mu}$

Properties if a and b are constants and $Y = aX + b$ then

$$\boxed{\mu_Y = a\mu_X + b \text{ and } \sigma^2_Y = a^2\sigma^2_X}$$

in other words:

$$E(Y) = E(aX+b) = aE(X) + b$$

$$Var[Y] = E[(Y - E(Y))^2] = E[(aX+b-(aE[X]+b))^2]$$

$$= E[(aX - aE[X])^2]$$

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

$$= a^2 Var(X)$$

$$\boxed{SD(Y) = \sqrt{Var[Y]} = \sqrt{a^2 Var(X)} = |a| SD(X)}$$