Statistical Distributions with Discrete Random Variables

A short reminder from December

Recall from last semester that the expectation of a random variable can be found using the following:

$$E(X)=\sum_{i=1}^{n} p_i\, x_i$$

A person plays a game with a pair of coins. If two heads appear then \$10 is won. If a head and a tail appear then \$3 is won. If two tails appear then \$5 is lost.

a How much would a person expect to win playing this game once?

b If the organiser of the game is allowed to make an average of \$1 per game, how much should be charged to play the game once?

Result	Win

Mean and Standard Deviation of a Discrete Random Variable

Suppose x_i are the possible values of the random variable X, and f_i are the frequencies with which these values occur.

We calculate the **population mean** as $\mu = \dfrac{\sum f_i x_i}{\sum f_i}$,

the **population standard deviation** as $\sigma = \sqrt{\dfrac{\sum f_i (x_i - \mu)^2}{\sum f_i}}$,

and the **population variance** is σ^2.

If a discrete random variable has n possible values $x_1, x_2, x_3, \ldots.., x_n$ with probabilities $p_1, p_2, p_3, \ldots.., p_n$ of occurring,

then
- the population **mean** is $\mu = \sum x_i p_i$
- the population **variance** is $\sigma^2 = \sum (x_i - \mu)^2 p_i$
- the population **standard deviation** is $\sigma = \sqrt{\sum (x_i - \mu)^2 p_i}$.

The population mean of a discrete random variable is often referred to as the '*expected value of* x' or sometimes as the '*average value of* x *in the long run*'.

In practice, we can define:

- the **mean** as $\mathrm{E}(X) = \mu = \sum x_i p_i$ and
- the **variance** as $\mathrm{Var}(X) = \sigma^2 = \sum (x_i - \mu)^2 p_i = \mathrm{E}(X - \mu)^2$.

Example 2

A magazine store recorded the number of magazines purchased by its customers in one week. 23% purchased one magazine, 38% purchased two, 21% purchased three, 13% purchased four, and 5% purchased five.

a What is the random variable?

b Make a probability table for the random variable.

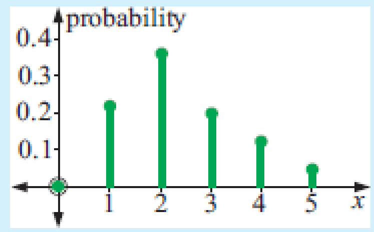
c Graph the probability distribution using a spike graph.

a The random variable X is the number of magazines sold.
So, $x = 0, 1, 2, 3, 4$ or 5.

b

x_i	0	1	2	3	4	5
p_i	0.00	0.23	0.38	0.21	0.13	0.05

c



Find the mean and standard deviation of the data of **Example 2**.

The probability table is:

x_i	0	1	2	3	4	5
p_i	0.00	0.23	0.38	0.21	0.13	0.05

Now $\mu = \sum x_i p_i$

$= 0(0.00) + (0.23) + 2(0.38) + 3(0.21) + 4(0.13) + 5(0.05)$

$= 2.39$

so in the long run, the average number of magazines purchased per customer is 2.39.

The standard deviation

$$\sigma = \sqrt{(x_i - \mu)^2 p_i}$$

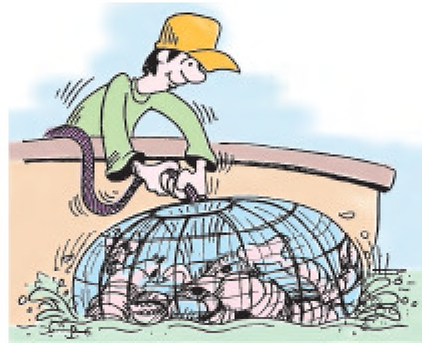
$$= \sqrt{(1-2.39)^2 \times 0.23 + (2-2.39)^2 \times 0.38 + ... + (5-2.39)^2 \times 0.05}$$

$$\approx 1.122$$

A country exports crayfish to overseas markets. The buyers are prepared to pay high prices when the crayfish arrive still alive.

If X is the number of deaths per dozen crayfish, the probability distribution for X is given by:

x_i	0	1	2	3	4	5	> 5
$P(x_i)$	0.54	0.26	0.15	k	0.01	0.01	0.00



a Find k.

b Over a long period, what is the mean number of deaths per dozen crayfish?

c Find σ, the standard deviation for the probability distribution.

A random variable X has probability distribution given by

$P(x) = \dfrac{x^2 + x}{20}$ for $x = 1, 2, 3$. Calculate μ and σ for this distribution.

A random variable X has probability distribution given by

$$P(x) = C_x^3\,(0.4)^x(0.6)^{3-x} \quad \text{for} \quad x = 0,\ 1,\ 2,\ 3.$$

a Find $P(x)$ for $x = 0, 1, 2$ and 3 and display the results in table form.

b Find the mean and standard deviation for the distribution.

An insurance policy covers a \$20 000 sapphire ring against theft and loss. If it is stolen then the insurance company will pay the policy owner in full. If it is lost then they will pay the owner \$8000. From past experience, the insurance company knows that the probability of theft is 0.0025 and of being lost is 0.03. How much should the company charge to cover the ring if they want a \$100 expected return?

PROPERTIES OF $\mathrm{E}(X)$

If $\mathrm{E}(X)$ is the expected value of random variable X then:

- $\mathrm{E}(k) = k$ for any constant k
- $\mathrm{E}(kX) = k\mathrm{E}(X)$ for any constant k
- $\mathrm{E}(A(X) + B(X)) = \mathrm{E}(A(X)) + \mathrm{E}(B(X))$ for functions A and B
  i.e., the expectation of a sum is the sum of the individual expectations.

These properties enable us to deduce that:

$\mathrm{E}(5) = 5$, $\mathrm{E}(3X) = 3\mathrm{E}(X)$ and $\mathrm{E}(X^2 + 2X + 3) = \mathrm{E}(X^2) + 2\mathrm{E}(X) + 3$

PROPERTY OF $\mathrm{Var}(X)$

$$\mathbf{Var}(X) = \mathbf{E}(X^2) - \{\mathbf{E}(X)\}^2 \quad \text{or} \quad \mathbf{Var}(X) = \mathbf{E}(X^2) - \mu^2$$

Proof:

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}(X-\mu)^2 \\ &= \mathrm{E}(X^2 - 2\mu X + \mu^2) \\ &= \mathrm{E}(X^2) - 2\mu\mathrm{E}(X) + \mu^2 \quad \{\text{properties of } \mathrm{E}(X)\} \\ &= \mathrm{E}(X^2) - 2\mu^2 + \mu^2 \\ &= \mathrm{E}(X^2) - \mu^2\end{aligned}$$

Example 9

X has probability distribution

x	1	2	3	4
p_x	0.1	0.3	0.4	0.2

Find:

a the mean of X **b** the variance of X **c** the standard deviation of X.

a $\mathrm{E}(X) = \sum x p_x = 1(0.1) + 2(0.3) + 3(0.4) + 4(0.2)$

$\therefore \mathrm{E}(X) = 2.7$ so $\mu = 2.7$

b $\mathrm{E}(X^2) = \sum x^2 p_x = 1^2(0.1) + 2^2(0.3) + 3^2(0.4) + 4^2(0.2) = 8.1$

$$\begin{aligned}\therefore \mathrm{Var}(X) &= \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2 \\ &= 8.1 - 2.7^2 \\ &= 0.81\end{aligned}$$

c $\sigma = \sqrt{\mathrm{Var}(X)} = 0.9$

X has probability distribution:

x	5	6	7	8
p_x	0.2	k	0.4	0.1

Find:

a the value of k **b** the mean of X **c** the variance of X.

(a) $\sum p_x = 1$ $\quad .2 + k + .4 + .1 = 1$

$k = .3$

(b) $\mu = E(x) = \sum x p_x = 5(.2) + 6(.3) + 7(.4) + 8(.1)$

$E(x) = 6.4$

(c) $VAR(x) = E(x^2) - \{E(x)\}^2$

$= [5^2(.2) + 6^2(.3) + 7^2(.4) + 8^2(.1)] - 6.4^2$

$= .84$

X has probability distribution:

x	1	2	3	4
p_x	0.4	0.3	0.2	0.1

Find:

a $E(X)$ **b** $E(X^2)$ **c** $Var(X)$ **d** σ

e $E(X+1)$ **f** $Var(X+1)$ **g** $E(2X^2+3X-7)$

(a) $E(X) = \sum x p_x = 1(.4) + 2(.3) + 3(.2) + 4(.1)$

$= 2$

(b) $E(x^2) = \sum x^2 p_x = 1^2(.4) + 2^2(.3) + 3^2(.2) + 4^2(.1)$

$= 5$

(c) $Var(x) = E(x^2) - \{E(x)\}^2$

$= 5 - 2^2$

$= 1$

(d) $\sigma = \sqrt{VAR(x)}$

$= 1$

(e) $E(x+1) = E(x) + E(1)$

$= 2 + 1$

$= 3$

(f) $VAR(X+1) = E(X+1)^2 - \{E(X+1)\}^2$

$= E(X^2+2X+1) - 3^2$

$= E(X^2) + 2E(X) + E(1) - 9$

$= 5 + 2(2) + 1 - 9$

$= 1$

(g) $E(2X^2+3X-7)$

$= 2E(X^2) + 3E(X) - E(7)$

$= 2(5) + 3(2) - 7$

$= 9$

X has probability distribution:

- **a** Find a and b given that $E(X) = 2.8$.
- **b** Hence show that $Var(X) = 1.26$.

x	1	2	3	4
p_x	0.2	a	0.3	b

$E(X) = 2.8$

$2.8 = \sum x p_x$

$2.8 = 1(.2) + 2a + 3(.3) + 4b$

[1] $1.7 = 2a + 4b$

$\sum p_x = .2 + a + .3 + b$

$1 = .5 + a + b$

[2] $.5 = a + b$

$a = .15$
$b = .35$

$VAR(X) = E(X^2) - \{E(X)\}^2$

$1.26 = 1^2(.2) + 2^2 a + 3^2(.3) + 4^2 b - 2.8^2$

$1.26 = .2 + 4(.15) + 2.7 + 16(.35) - 2.8^2$

Suppose X is the number of marsupials entering a park at night. It is suspected that X has a probability distribution of the form

$P(X = x) = a(x^2 - 8x)$ where $X = 0, 1, 2, 3, \ldots.\ 8.$

a Find the constant a.

b Find the expected number of marsupials entering the park on a given night.

c Find the standard deviation of X.

need $\sum P(x) = 1$

IN ORDER TO FIND a

'Kay so AFTER A LOT OF ARITHMETIC

$a = -\frac{1}{84}$

$E(X) = \sum x p_x = 1\left(\frac{7}{84}\right) + 2\left(\frac{12}{84}\right) + 3\left(\frac{15}{84}\right) + 4\left(\frac{16}{84}\right)$

$+ 5\left(\frac{15}{84}\right) + 6\left(\frac{12}{84}\right) + 7\left(\frac{7}{84}\right)$

$E(x) = 4$

$VAR(x) = E(X^2) - \{E(x)\}^2$

$= 19 - 4^2$

$= 3$

so $\sigma = \sqrt{VAR(x)} = \sqrt{3}$