1

Stat 104: Quantitative Methods

Class 11: Discrete Random Variables

2

Review Question

If events B and C are independent, which of the following is not always true?

a) P(B and C) = P(C)P(B)
b) P(B | C) = P(C)
c) P(B or C) = P(B) + P(C) – P(B)P(C)
d) P(C | B) = P(C)

3

Turning Letters into Numbers

- We've been studying statements like P(A) or P(B|A). That is, probability of different events happening.
- But many things in the real world are easier explained using numerical outcomes.
- The concept of random variables enables us to talk about the probability of numerical outcomes.

4

Random Variables

- A random variable gives us a way to numerically describe the possible outcomes in a probability experiment.
- More specifically, a random variable is a rule or function that translates the outcomes of a probability experiment into numbers.
- *Discrete random variables* can assume only a finite or limited set of values.
- *Continuous random variables* can assume any one of an infinite set of values.

5

Examples

EXPERIMENT	OUTCOME	RANDOM VARIABLES	RANGE OF RANDOM VARIABLES
Stock 50 Christmas trees	Number of Christmas trees sold	X	0, 1, 2,…, 50
Inspect 600 items	Number of acceptable items	Y	0, 1, 2,…, 600
Send out 5,000 sales letters	Number of people responding to the letters	Z	0, 1, 2,…, 5,000
Build an apartment building	Percent of building completed after 4 months	R	$0 \le R \le 100$
Test the lifetime of a lightbulb (minutes)	Length of time the bulb lasts up to 80,000 minutes	S	$0 \le S \le 80{,}000$

6

Examples-Create Random Variables

EXPERIMENT	OUTCOME	RANDOM VARIABLES	RANGE OF RANDOM VARIABLES
Students respond to a questionnaire	Strongly agree (SA) Agree (A) Neutral (N) Disagree (D) Strongly disagree (SD)	X = 5 if SA 4 if A 3 if N 2 if D 1 if SD	1, 2, 3, 4, 5
One machine is inspected	Defective Not defective	Y = 0 if defective 1 if not defective	0, 1
Consumers respond to how they like a product	Good Average Poor	Z = 3 if good 2 if average 1 if poor	1, 2, 3

A Little Notation

- We may get sloppy from time to time, but usually in the rest of this course, we will stress the distinction between a random variable and the values it can take on by following the convention of using **capital letters** such as X to denote random variables and **lowercase letters** such as x to denote their values.
- For example, say we throw a die and let X be the outcome. The random variable X can take the specific values x=1,x=2,....,x=6.

A Summary Before we Begin

Sample (Data)	X Random Variable
Histogram	Distribution
Mean $\bar{x}$	Expectation E(X)
Variance s^2	Variance Var(X)

The Probability Function

The probability mass function $P_X(x)$ of a discrete random variable expresses the probability that X takes the value x:

$$P_X(x) = P(X = x)$$

Realize that outcome of probability mass function is 0->1

Example: X = outcome when we roll a fair die.

$$P_X(1) = 1/6, P_X(2) = 1/6, etc\ldots$$

What is $P_X(17)$??

0, not undefined. It's a function so any input is possible.
For most, output will be zero. For specific inputs, some value between 0 and 1

We assign probabilities (chances) to all the values that the random variable can take on.

There are certain requirements about the values we assign:

If a random variable X can take values x_i, then the following must be true:

$$(1)\ 0 \le P_X(x_i) \le 1$$

$$(2)\ \sum_{\text{all } x_i} P_X(x_i) = 1 \quad \text{(exhaustive)}$$

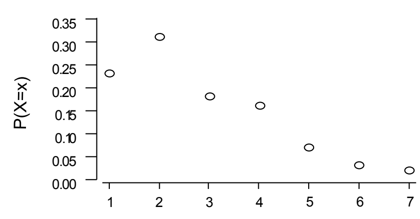
$P_X(x)$ is sometimes called the probability distribution function

Example: Let X denote the number of people in an American household. According to the Statistical Abstract of the United States (1988), the probability distribution of X is as follows (rounded to two decimal places):

x	1	2	3	4	5	6	7
P(X=x)	0.23	0.31	0.18	0.16	0.07	0.03	0.02

- What is the most likely number of people in a household ? 2
- What is the probability of a household having fewer than four people ? 0.23+0.31+0.18
- More than seven people ? 0

Here is a graph of the probability mass function.

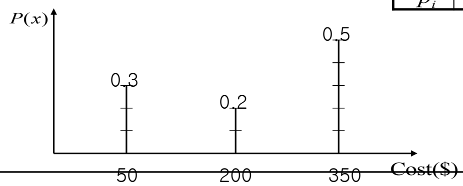


Realize that here we can not connect dots. This is not a continuous distribution. Probability is non zero only for certain values

Example

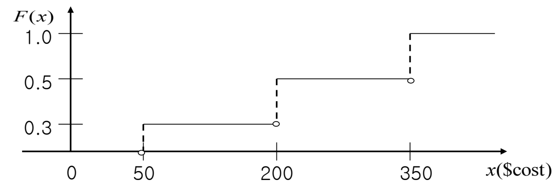
- Machine Breakdowns
  - P (cost=50)=0.3, P (cost=200)=0.2, P (cost=350)=0.5
  - 0.3 + 0.2 + 0.5 =1

x_i	50	200	350
p_i	0.3	0.2	0.5



Cumulative Distribution Function

- Cumulative Distribution Function
  - Function : $F(x) = P(X \le x)$ $F(x) = \sum_{y: y \le x} P(X = y)$
  - Abbreviation : c.d.f



Practice

A random variable Y has the following probability distribution:

Y	-1	0	1	2
P(Y)	3C	7C	0.4	0.1

The value of the constant C is:

A. .10
B. .15
C. .20
D. .05

3c+ 7c + 0.4 + 0.1 = 1
10c = 0.5
c = 0.05

Practice

- Which of the following are legal probability distributions?

A)

Value of X	1	2	3	4	5
Probability	0.1	0.2	0.3	0.2	0.1

x : total not 1

B)

Value of X	1	2	3	4	5
Probability	−0.1	0.2	0.3	0.2	0.4

x : -ve probability

C)

Value of X	1	2	3	4	5
Probability	0.1	0.2	0.3	0.4	0.5

x : total >1

D)

Value of X	1	2	3	4	5
Probability	0.1	0.2	0.3	0.2	0.2

Looks good!

Practice

Can the function $f(x) = \frac{x+6}{24}$, for x = 1, 2, and 3, be the probability distribution for some random variable?

(A) Yes.
(B) No, because probabilities cannot be negative.
(C) No, because probabilities cannot be greater than 1.
(D) No, because the probabilities do not sum to 1.
(E) Not enough information is given to answer the question.

Example: Pick Three Lottery

Description of possible 3-digit plays	Example	You win if any of these combinations are drawn	Payout 50¢	Payout $1.00 play
STRAIGHT: Play 3 digits in exact order. 1 way to win (Odds-- 1:1,000)	123	123	$250	$500

- P(WIN) = 1/1000
- You pick one number out of 1000……

Example: Pick 3 Lottery

- W = net gain
- Costs $1 to play so net gain is 500-1=499 or, if you lose, net gain = -$1.

w	P(W=x)
-1	999/1000
499	1/1000

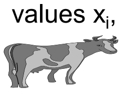
Earlier in the course, we saw that to fully understand a data set, one doesn't just examine the histogram but also calculates summary statistics such as the **mean and variance**.

The probability distribution of a numerical random variable is very much like a histogram based on a huge sample size, and the ideas presented in this section work like the summary statistics we saw for data sets, enabling us to summarize a random variable.

Expectation and its Applications

For a random variable, the analogy to the sample mean is called the expectation or expected value. The letter E usually denotes an expected value, and this symbol is usually followed by brackets enclosing the random variable of interest. This obviously assumes discrete random variables

Definition: Given a random variable X with values x_i, the expected value of X is



$$\mu_X = E(X) = \sum_{all\ x_i} x_i P(X = x_i)$$
$$= \sum_{all\ x_i} x_i P(x_i)$$

The expected value is simply a *weighted average* of the possible values X can assume. where the weights are the probabilities of occurrence of those values.

Formulas kinda similar

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \frac{1}{n} x_i$$

Different weights

$$\mu_X = E(X) = \sum_{all\ x_i} x_i P_X(x_i)$$

GEORGIA LOTTERY CA$H 3

Description of possible 3-digit plays	Example	You win if any of these combinations are drawn	Payout	
			50¢	$1.00 play
STRAIGHT: Play 3 digits in exact order. 1 way to win (Odds-- 1:1,000)	123	123	$250	$500

Let W = winnings. W = 499 (cost $1 to play) or -1

$$E(W) = 499 * \frac{1}{1000} + (-1) * \frac{999}{1000} = -0.50$$

Interpretation

- So in the long run, we lose 50 cents each time we play.
- Note that on any one play we never lose 50 cents (we either win 499 or lose 1); rather, this is saying that if you play the game 10000 times, you can expect to be roughly down $5000 at the end.
- An even better way to look at it is that if 10 million people play the game every day, the state can expect to only have to give back about $5 million, a daily profit of a cool $5 million (this is why states run lotteries!)

Example

- What is the expected value when you roll a fair 6-sided die?

1(1/6)+2(1/6)+3(1/6)+4(1/6)+5(1/6)+6(1/6) = 3.5

Insurance Example

Expected values are useful to find the expected outcome of betting games, or equivalently insurance policies. Suppose a certain make of automobile worth $15,000 is insured by the Metter's Insurance Company. In their experience, the following losses occur with the following probabilities:

loss (X)	Probability
$0	0.8
$1,000	0.1
$5,000	0.05
$10,000	0.03
$15,000	0.02

What should they charge for a premium if they want to make on average $100 per policy ?

Insurance Example (cont)

loss (X)	Probability
$0	0.8
$1,000	0.1
$5,000	0.05
$10,000	0.03
$15,000	0.02

E(X)=0(.8)+1000(0.1)+5000(0.05)+10000(0.03)+15000(0.02)=950

Expected loss is 950/policy so they should charge 1050/policy

Former Exam Question

Consider the following probability distribution:

k	$Pr(X=k)$
c	1/6
9	1/3
16	1/2

If the mean for this probability distribution is 12, what is c?

c(1/6)+9(1/3)+16(1/2) = 12

c =

Example

- Tom pays 5 dollars to play the following game. First, he rolls a 4-sided die. The number he rolls gives the number of dollars he wins. If Tom rolls a 2, 3, or 4, the game is over. If he rolls a 1, he gets to flip a coin. If the coin flip results in heads, he gets 10 dollars more, and the game ends. If the coin flip results in tails, he wins no additional money, and the game ends.
- Let K be Tom's net earnings and find E(K).

Example (cont)

k	P(X=k)
-4	1/8
-3	1/4
-2	1/4
-1	1/4
6	1/8

E(k)=(-4)(1/8)+(-3)(1/4)+(-2)(1/4)+(-1)(1/4)+6(1/8)= -10/8

Some Expectation Rules

- $E(cX) = cE(X)$ where c is any constant
- $E(X+c) = E(X)+c$
- $E(X+Y) = E(X)+E(Y)$ for any two random variables
- Note-in general combining random variables can lead to some issues, but expectations always behave nicely.

Think of this on same lines as mean

Dropping Coins

- Suppose a game involves dropping 3 coins on the table—a nickel, a dime, and a quarter. Each coin that lands "heads up" you are allowed to keep, so that the possible reward R ranges from 0 to 40c.
- Find the mean of R.
- This can be done by brute force by first finding the distribution of R, then the mean.
- But using the idea of sums of random variables, there is an easier approach.

Breaking down the problem

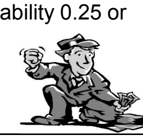
- Define the following random variables
  - $X1$ = the nickel's contribution to the reward
  - $X2$ = the dime's contribution to the reward
  - $X3$ = the quarter's contribution to the reward
- Then $R = X1 + X2 + X3$
- $E(X1) = .5(.05)+.5(0)$ [why??]
- $E(X2) = .5(.1)+.5(0)$
- $E(X3) = .5(.25)+.5(0)$
- $E(R) = E(X1)+E(X2)+E(X3) = 0.20$

One more betting example

Find the expected value of each of the following bets:

a) you get \$5 with probability 1.0.
b) you get \$10 with probability 0.5, or \$0 with probability 0.5.
c) you get \$5 with probability 0.5, \$10 with probability 0.25 and \$0 with probability 0.25.
d) you get \$5 with probability 0.5, \$105 with probability 0.25 or lose \$95 with probability 0.25.

Which bet is the best bet ?



Variance of a Random Variable

- The mean of a random variable is a very useful and informative quantity, but we are often interested in other measures of a distribution.
- The variance and standard deviation are measures of the dispersion of a random variable around its mean.

As we saw a few slides ago, all of the following bets have the same expected value.

a) you get \$5 with probability 1.0.
b) you get \$10 with probability 0.5, or \$0 with probability 0.5.
c) you get \$5 with probability 0.5, \$10 with probability 0.25 and \$0 with probability 0.25.
d) you get \$5 with probability 0.5, \$105 with probability 0.25 or lose \$95 with probability 0.25.

So how do we determine how risky the different bets are ? The variance and standard deviation reflect this degree of diversity.

Which bet do you think is the safest ? The riskiest ? See next week's homework ☺

Technically, the variance of a random variable is the *expected value of the squared deviation of a random variable from its mean*. **Phew**.

The general mathematical formula is

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

For discrete random variables, this simplifies to

$$\sigma_X^2 = Var(X) = E[(X-\mu_X)^2] = \sum_{all\ x_i}(x_i-\mu)^2 P(X=x_i)$$

The standard deviation is of course $\sigma_X = \sqrt{\sigma_X^2}$

weighted average of deviations from the mean

Formulas kinda similar

Sample mean

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 = \sum_{i=1}^{n}\frac{1}{n-1}(x_i-\overline{x})^2$$

Different weights

$$\sigma_X^2 = \sum_{all\ x_i}(x_i-\mu)^2 P(X=x_i)$$

Population mean

Example : Pick 3 Game

$$\sigma_X^2 = Var(X) = E[(X-\mu_X)^2] = \sum_{all\ x_i}(x_i-\mu)^2 P(X=x_i)$$

(Val - mean)2 * probab

$$Var = [499-(-.5)]^2 x \frac{1}{1000} + [-1-(-.5)]^2 x \frac{999}{1000} = 249.75$$

The standard deviation of this is sqrt(249.75)=$15.8

What does it imply that the std deviation is so much larger than the mean ?

Insurance Example

- Calculating the variance requires a bit of work-its not difficult but be careful with the math:

$$\begin{aligned} Var(X) &= \sum(x-E(X))^2 P(X=x) \\ &= (0-950)^2(.8) + (1000-950)^2(.1) \\ &+ (5000-950)^2(.05) + (10000-950)^2(.03) \\ &+ (15000-950)^2(.02) = \$^2 7{,}947{,}500 \end{aligned}$$

Std Dev = sqrt(7,947,500)=$2819

Alternative Variance Formula

- It can be shown that another formula for variance is

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

- This formula is sometimes easier to compute than the previous variance formula.

Example

- Consider the following game:
- You pay $6 to pick a card from a standard 52-card deck.
- If the card is a diamond (♦), you get $22; if the card is a heart (♥), you get $6; otherwise, you get nothing.
- What is the expected value and variance of this game?

Example (cont)

Outcome	W
♠	-6
♡	0
♣	-6
♢	16

w	-6	0	16
$p(w)$	0.50	0.25	0.25

Example (cont)

Outcome	W
♠	-6
♡	0
♣	-6
♢	16

w	-6	0	16
$p(w)$	0.50	0.25	0.25

Basic Rules for Variance

- Var(X+c) = Var(X) for any constant c
- Var(cX)=c^{2}Var(X)
- In general, Var(X+Y) ≠Var(X)+Var(Y) [more on this later]

Interpreting the Variance

- The standard deviation of a random variable is important.
- We already know that the random variable X can assume different values $x_1, x_2, \ldots, x_n$.
- The expected value of X gives the central location, but it is quite possible that the actual outcome will differ from E(X).
- However, Chebyshev's rule also holds for random variables, so we know that at least 75% of the time, the random variable outcome will be within 2 standard deviations of the mean value.
- More info on interpreting in a few classes from now.

Chebyshev's rule for random vars.

	Chebyshev's Rule Applies to any probability Distribution	Empirical Rule Applies to probability Distributions that are mound Shaped and symmetric
$P(\mu-\sigma<x<\mu+\sigma)$	≥ 0	$\approx 68\%$
$P(\mu-2\sigma<x<\mu+2\sigma)$	$\geq 75\%$	$\approx 95\%$
$P(\mu-3\sigma<x<\mu+3\sigma)$	$\geq 89\%$	$\approx 100\%$

Rules for Expectation and Variance

Let X be a random variable with mean μ and variance σ^2. Let a and b be any constant fixed numbers. Define the random variable W=a+bX.

Then

$$\mu_W = E(W) = E(a+bX) = a + bE(X) = a + b\mu_X$$

$$Var(W) = Var(a+bX) = b^2\sigma_X^2$$

Example

- An author receives from a publisher a contract, according to which she is to be paid a fixed sum of $10,000 plus $1.50 for each copy of her book sold.
- Her uncertainty about total sales of the book can be represented by a random variable with mean 30,000 and standard deviation 8,000.
- Find the mean and standard deviation of the total payments she will receive

Solution

- Payment =10000+1.5*X, where X = # of books sold
- This is the a+bX framework with a=10000 and b=1.5
- E(Payment) = 10000+1.5E(X), E(X)= 30000
- Var(Payment) = $(1.5)^2$ Var(X), Var(X)= 8000^2

Things you should know

- ❑What is a random variable
- ❑Probability function
- ❑Expectation
- ❑Variance
- ❑Expectation Algebra (a+bX rule)