Lecture 9

How Can We Describe the Distribution of a Discrete Random Variable?

In describing a distribution, we always discuss the center (measure of central tendency) and spread. The discrete probability distribution is no exception.

Mean (aka Expected Value, Expectation, or Average)

For a discrete distribution, this is just the weighted average of the possible values with corresponding probabilities as the weights.

Formulaically:

Suppose that X is a discrete random variable with the following probability distribution:

x	x_1	x_2	x_3	…	x_n
P(x)	p_1	p_2	p_3	…	p_n

Then:

Example – Defective Products
Consider a company with 2 production lines. The following probability distribution function of the random variables X and Y represents the daily number of defective products coming out of each production line.

x,y	0	1	2	3	4
P(x)	$.15$	$.30$	$.25$	$.20$	$.10$
P(y)	$.05$	$.05$	$.10$	$.75$	$.05$

a) Find the expectation of X.

b) Just using your intuition, which of the 2 production lines do you think will have a higher average daily number of defective products?

c) Find the expected value of Y and confirm your intuition.

Example – Life-Insurance

Suppose that you work for an insurance company, and you sell a $150,000 whole-life insurance policy at an annual premium of $1250. Actuarial tables show that the probability of death during the next year for a person of your customer's age, sex, health, etc. is 0.005.

a) Let the RV X be the company's gain (i.e. amount of money made by the company) for a policy of this type. What is the probability distribution function of X in table form?

b) What is the expectation with regard to company profit (gain) for a policy of this type?

c) Suppose that a decade has passed and your actuarial tables indicate that the probability of death during the next year for a person of your customer's **current** age is .025. This change in probability will obviously be reflected in the annual premium paid. What should the annual premium be (instead of $1250) if the company intends to keep the same expected profit (gain)?

Standard Deviation

Formulaically:

Suppose that X is a discrete random variable with the following probability distribution:

x	x_1	x_2	x_3	...	x_n
P(x)	p_1	p_2	p_3	...	p_n

and mean μ_x.

The standard deviation of X (σ_x) is again a weighted average:

Similar to when we discussed descriptive statistics, we have the following qualities about the standard deviation of a discrete probability distribution:

i.

ii.

iii.

iv.

Example – Defective Products (cont.)
Note the probability distributions of the two random variables we discussed earlier regarding defective products:

x,y	0	1	2	3	4
P(x)	$.15$	$.30$	$.25$	$.20$	$.10$
P(y)	$.05$	$.05$	$.10$	$.75$	$.05$

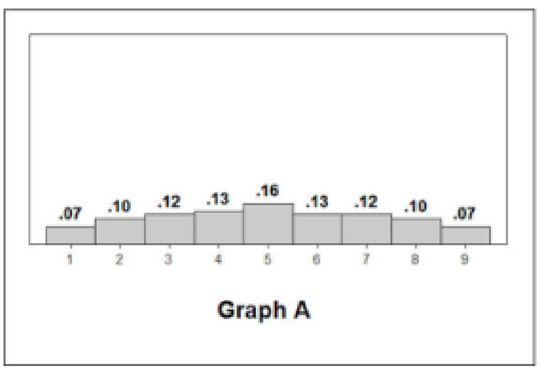
a) Find the standard deviation of the random variable X.

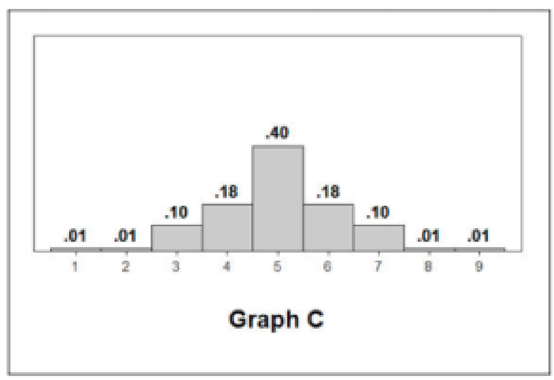
b) Using only your intuition, which RV do you think has the larger standard deviation?

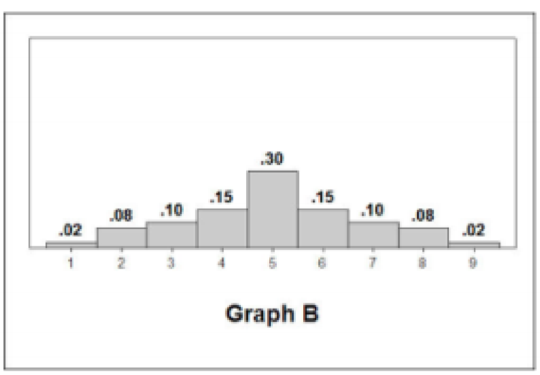
c) Confirm your intuition by calculating σ_Y.

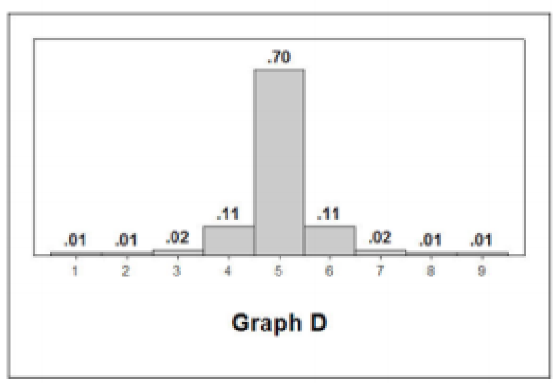
So, we can conclude that, on average, the daily number of defective products from production line X is ______, give or take _______.

Let's Get a Better Understanding of Standard Deviation:









Rank the standard deviations of the 4 graphs above from smallest to largest:

A special case of a discrete random variable is the Binomial Random Variable.

The Binomial Random Variable

Before we begin, we'll need some mathematical background:

i. Factorial

Example – Factorial Calculation
Calculate 5! & 3!

ii. Combinations

The number of ways you can choose k objects out of n total objects (i.e. k is the number of combinations) is:

Notice that

How?

Example – Combinations
Suppose Austin is randomly selecting a team of 5 students from his class of 12. In how many different ways can this be done?

The Binomial Random Variable

The binomial random variable is defined on the following experiment:

F –

I –

T –

S –

So, if the random variable, X for example, follows a binomial distribution, we define X by the number of trials (n) and the probability of success:

Example – Determining Binomial Random Variables
For each of the following, decide whether the random variable of interest is indeed a Binomial Random Variable. If not, identify why not.

a) A die is rolled 4 times with the RV X reflecting the number of “3’s”

b) The random variable X represents the number of tosses until we get 7 “Heads”

c) A student who has no clue uses an independent random guess to answer each question on a test consisting of 25 multiple choice questions (each with 5 options) and 4 T/F. The random variable X represents the number of questions correct.

d) A student who has no clue uses an independent random guess to answer each question on a test consisting of 29 multiple choice questions (each with 5 options). The random variable X represents the number of questions correct.

e) Suppose that among 20 donors in a blood drive, 8 have blood type A. Three of the 20 donors are chosen at random. Let X be the number of donors with blood type A that are chosen (out of 3).

Note: Remember that when we choose a sample of subjects from a population, we can assume independence and use the Binomial distribution if:

i. **The sample is *random***
ii. **The population size is large (at least 20 times larger) compared to the sample size**

If $X \sim Bin(n, p)$, then the probability distribution of X is

NOTICE THAT ____ IS ALWAYS A POSSIBLE VALUE OF X!!!

Example – Coin Tosses

Suppose that we conduct a random experiment by tossing the same coin 20 times. Let the random variable X be the number of heads tossed in the experiment.

a) Is X a discrete or continuous random variable?

b) What are the possible values of X? Is this a subset of the sample space?

c) Does X follow a binomial distribution, i.e. is X a Binomial Random Variable? If so, properly label the distribution.

d) What is the probability distribution function of X?

e) What is the probability that exactly 8 tosses are heads?

f) What is the probability that less than 3 tosses are heads?

g) What is the probability that at most 2 tosses are heads?

h) What is the probability that at least 3 tosses are heads?

Describing A Binomial Random Variable

If $X \sim Bin(n, p)$, then we have the following:

i.

ii.

So, to continue the Coin Tosses example:

i) Find the expected number or heads to come up.

j) How much are we “off” by, on average, when we use the mean of X to estimate the number of heads that will come up?

k) What is the probability that the number of heads that are tossed is within one standard deviation from the mean?