Statistics Ch 4

Ch 4.1 Discrete Random Variable

Random Variable

Random Variable: is a variable (X) that has a single numerical value, determined by chance, for each outcome of a procedure.

Probability distribution: is a table, formula or graph that gives the probability of each value of the random variable.

A **Discrete Random Variable** has a collection of values that is finite or countable similar to discrete data.
A **Continuous Random Variable** as infinitely many values and the collection of values is not countable.

Ex : X = the number of times "four" shows up after tossing a die 10 times is a discrete random variable.
X = weight of a student randomly selected from a class. X is a continuous random variable
X = the method a friend contacts you online. X is not a random variable. (X is not numerical)

A Probability Distribution (PDF) for a Discrete Random Variable is a table, graph or formula that gives Probability of each value of X.

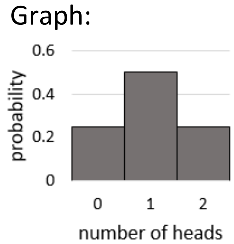
A Probability Distribution Function (PDF or PD) satisfies the following requirements:
1. The value X is numerical, not categorical and each P(x) is associated with the corresponding probability.
2 ΣP(X) = 1 . A ΣP(X) = 0.999 or 1.01 is acceptable as a result of rounding.
3. 0 ≤P(x) ≤ 1 for all P(x) in the PDF.

Ex1: PDF for number of heads in a two-coin toss are given as a table and a graph.

Table:

x	P(x)
0	0.25
1	0.5
2	0.25

Graph:



Both are valid PDF because Σ P(X) = 1
and each value of P(X) is between 0 and 1.

Ex2: The number of medical tests a patient receives after entering a hospital is given by the PDF below

x	1	2	3	4
p(x)	0.02	0.18	0.3	0.4

a) Is the table a valid PDF?

The table is not a prob. distribution.
because Σ P(x) = 0.02+0.18+0.3+0.4 = 0.9 is not 1
b) Define the random variable x.
x = no. of medical tests a patient receives after entering a hospital.
c) Explain why the x = 0 is not in the PDF?
A patient always receives at least one medical test in the hospital.

Parameters of a Probability distribution:

Mean μ for a probability distribution:
$E(x) = \mu = \sum xP(x)$
Variance σ^2 for a probability distribution:
$\sigma^2 = \sum(x-\mu)^2 P(x)$
Standard deviation for a probability distribution:
$\sigma = \sqrt{\sum(x-\mu)^2 P(x)}$
Google "easycalculation.com discrete random variable calculator " link or use
https://www.easycalculation.com/statistics/discrete-random-variable.php:
Enter Number of outcomes, each X and P(X), calculate.
round off rule: one more decimal place than for E(x)
Two decimal places for σ and σ^2.

Expected value = the long-term outcome of average of x when the procedure is repeated infinitely many times. Round to one decimal place.

Non-significant values of X.
1. The range of X from $\mu - 2\sigma\ to\ \mu + 2\sigma$ is non-significant. (Range of rule of Thumb)
2. X that are outside of X $\mu - 2\sigma\ to\ \mu + 2\sigma$ are significant that is unlikely to occur.

Ex1: X = no. of year a new hire will stay with the company. P(x) = Prob. that a new hire with stay for x year.

x	P(X)
0	0.12
1	0.18
2	0.3
3	0.15
4	0.1
5	0.1
6	0.05

a) Find mean, variance, st. deviation and determine the Expected number of years a new hire will stay.

Use easycalculation.com statistics discrete random variable calculator, Enter number of outcomes = 7. Mean = 2.4, σ^2 = 2.73, σ =1.65
The Expected no. of year a new hire will stay is 2.4 years.

b) Find prob that a new hire will stay for 4 years or more.
P(4 or more) = 0.1 + 0.1 + 0.05 = 0.25

c) Find prob that a new hire will stay for between 3 or 5 years inclusive.
P(3 to 5 inclusive) = 0.15 + 0.1+0.1 = 0.35

d) Find the prob that a new hire will stay for 2 years or fewer.
P(2 or fewer) = 0.12 + 0.18 + 0.30 = 0.6

e) Find the range of non-significant year of stay.
2.4 -2(1.65) to 2.4 + 2(1.65) is -0.9 to 5.7

Ex2: Given x = of number of textbooks a student buy per semester. What is the expected number of textbooks?

x	P(x)
1	0.02
2	0.03
3	0.45
4	0.45
5	0.03
6	0.02

a) Find mean, variance and standard deviation.
Use easycalculation.com statistics discrete random variable calculator,
Enter number of outcomes = 6
E(x) = μ = 3.5, σ^2 = 0.61, σ = 0.78
Expected number of textbook is 3.5 books.
b) Find Probability that a student buys at least 5 textbook.
P(at least 5 = 5 or more) = 0.03 + 0.02 = 0.05,

c) Find probability that x is at most 2.
P(at most 2 = 2 or fewer) = 0.02 + 0.03 = 0.05

c) Find the range non-significant.
Range of non-significant is 3.5 – 2(0.78) to 2.5 + 2(0.78) is 1.94 to 5.06.

Ch 4.2 Application of Probability Distribution

Application of expected value. The expected average gain or profit for a game or business in the long run.

X = Long term average net gain after a bet or Profit for a business, E(x) is expected gain for a game or a business after constructing a probability distribution.

Ex1. Suppose you play lottery where the chance of winning is 0.0001. You need to pay $3 to play. If you win, you will collect $10,000. What is the expected value of the game?
P(lose) = 1- P(win) = 1-0.0001 = 0.9999

Net win = 10,000 – 3 = 9997
Build a PD as below:

Event	X	P(X)
Win	9,997	0.0001
lose	-3	0.9999

E(X) = $\sum xP(x) = 9997(0.0001) + (-3)(0.9999)$
= -2 (negative means loss)
In the long run, the expected loss per game is $2.

Ex1. Bet $5 on number 7 in roulette can be summarized below:

Event	x	P(X)	x · P(X)
lose	-5	37/38	-4.868421
win	175	1/38	4.605263

EV = -0.26 or 26 cents.
For every $5 bet, you can expect to lose an average of 26 cents.

Ex 3. The probability a 25- year- old male passes away within the year is .0.0005. He pays $275 for a one year $160000 life insurance policy. What is the expected value of the policy for the insurance company? Round your answer to the nearest cent.

From the insurance company's point of view,
P(live) = 1 – P(die) = 1 – 0.0005 = 0.9995
If policy holder die, net loss = 275 - 160000 = -159725
Use the information to build the PD as follow

Event	x	P(X)
live	275	0.9995
die	-159725	0.0005

E(X) = 275(0.9995) + (-159725)(0.0005) = $195
In the long run, the company will gain $195 per policy.

Ch 4.3 Binomial Probability distribution $\boldsymbol{B(n,p)}$

Requirements for Binomial Distribution:

X can be modelled by binomial distribution if it satisfies four requirements:

1. The procedure has a fixed number of trials. (n)
2. The trials must be independent.
3. Each trial has exactly two outcomes, success and failure, where x = number of success in n trials.
4. The probability of a success remains the same in all trials. P(success in one trial) = p.
   P(failure in one trial) = 1 – p = q
   P(X) = x number of success in n trials.

Note: for sampling, use 5% guideline for independent.

Ex1: Determine if the following X is binomial or not
a. X = number of adults out of 5 who use iPhone.
b. X = number of times a student raises his/her hand in a class.
c. X = number of one after tossing a die 7 times.
d. X = number of tosses until the "one" shows up.
e. X = the way student commute to school.

a and c are binomial. a = B(5, p), c = B(7, 1/6)
b,d does not have a fixed number of trials.
e : X is not a count of success.

Find P(X) or P(range of X) when X is binomial:
n = number of trials, p= P(success in one trial)
q = P(failure in one trial), X = number of success.
Method 1: use formula:
$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$
Method 2: Use Statdisk /Analysis/ Probability Distribution/Binomial distribution
Enter n, p, x.. output in sample editor under P(x), P(x or fewer) or P(x or greater).
Optional: use OnlineStatbook binomial calculator:
http://onlinestatbook.com/2/calculators/binomial_dist.html
input n and p (to N and ∏), select above, below or between.

Parameters of binomial distribution:
mean μ = np
variance: σ^2 = npq
standard deviation σ = $\sqrt{npq}$

Range rule of thumb:
Values not significant: Between (μ - 2σ) and (μ + 2σ)

Find parameters of binomial distribution
Use Statdisk /Analysis/ Probability Distribution/ Binomial distribution, enter n, p, x, evaluate.
Mean, standard deviation and variances are under the sample editor.

Ex1. In a college, 35% of all students are full-time students. If 11 students are randomly chosen.
a) Can probability of X = number of full time students out of 11 be modeled by binomial distribution?

b) what is the probability that there are 4 full-time students out of 11?
Use statdisk binomial distribution n = 11, p = 0.35
x = 4. Use P(x)

c) What is the probability that there are less than 5 full-time students?
Use statdisk binomial distribution n = 11, p = 0.35,
x = 4 use P(x or fewer)

d) What is the probability of there are more than 3 full-time students?
Use Statdisk binomial distribution n= 11, p = 0.35,
x = 4 use P(x or greater)

Ex2: A bookstore manager estimates that 9.5% of all customers coming in the store will buy a book or magazine. If 24 customers visit the store on a certain business hour,
a) Can x = number of customers out of 24 who buy a book or magazine be modelled by binomial distribution?

Use Statdisk n = 24, p =0.095
b) Find the probability that exactly 3 customers will buy a book or magazine.
Use Statdisk binomial distribution n = 24, p =0.095,
x = 3, use P(x)

c) Find the probability that at least 5 customers will buy a book or magazine.
Use Statdisk binomial distribution n = 24, p =0.095,
x = 5, use P(x or greater)

d) Find the probability that at most 2 customers will buy a book or magazine.
Use Statdisk binomial distribution n = 24, p =0.095,
x = 2, use P(x or fewer)

e) Find the non-significant range of customer who will buy a book or magazine out of 24 customers.
Find mean and standard deviation from statdisk
Mean = 2.3, sd = 1.44
Non-significant range = 2.3 – 2(1.44) = -0.58 to
2.3 + 2(1.44) = 5.18

Ex3. A small airline has a policy of booking as many as 60 persons on an airplane that can seat only 53. (Past studies have revealed that only 78% of the booked passengers actually arrive for the flight.)

a) Find the probability that if the airline books 60 persons, not enough seats will be available.

b) Find the non-significant range of passengers who will arrive out of 60 passengers.

Answer:
a) Assume the arrival of each passenger are independent, we can model x passenger arrival by binomial distribution.

b) P(not enough seats) = P(more than 53 arrival)
Use statdisk binomial distribution, n = 60, p = 0.78
x = 54, use P(X or greater)

c) From Statdisk binomial distribution n = 60, p = 0.78
Mean = 46.8,
standard deviation = 3.21
Non-significant range is
= 46.8 -2(3.21) to 46.8 + 2(3.21) = 40.4 to 53.22