

Examples in Probability

Mehdi Hakim-Hashemi

Sample Spaces

Example

Two dice with different colors are rolled. Find the sample space and the number of outcomes if:

1. The numbers on the dice are observed.
2. The sum of the numbers on the dice is observed.

Solution:

1. The sample space is:

$$A = \{(i, j) : i, j = 1, \dots, 6\}$$

The number of outcomes is $|A| = 6 * 6 = 36$.

The following code generates the sample space in matrix form. Each row is an outcome.

```
# Initialize. A is an empty matrix with one row.
```

```
A <- matrix(, nrow = 1, ncol = 2)
```

```
for (die_1 in 1:6){
```

```
  for (die_2 in 1:6){
```

```
    A <- rbind(A, c(die_1, die_2))
```

```
  }
```

```
}
```

```
# The first row is NA NA. Remove it.
```

```
# A[-m,] removes the m-th row.
```

```
# A[, -m] removes the m-th column.
```

```
A <- A[-1,]
```

```
# First six rows of matrix A
```

```
head(A, n = 6)
```

```
##      [,1] [,2]
```

```
## [1,]    1    1
```

```
## [2,]    1    2
```

```
## [3,]    1    3
```

```
## [4,]    1    4
```

```
## [5,]    1    5
```

```
## [6,]    1    6
```

```
# Last six rows of matrix A
```

```
tail(A, n = 6)
```

```
##      [,1] [,2]
```

```
## [31,]    6    1
```

```
## [32,]    6    2
```

```
## [33,]    6    3
```

```
## [34,]    6    4
```

```
## [35,]    6    5
```

```
## [36,]    6    6
```

```
# Number of outcomes is number of rows of A
outcomes_number <- nrow(A)
show(outcomes_number)
```

```
## [1] 36
```

2. It is obvious that the sample space, B , is any number between $1 + 1 = 2$ and $6 + 6 = 12$. So

$$B = \{2, 3, \dots, 12\}$$

It is instructive to write a program that generates this sample space.

```
# Initialize B as an empty set.
B = {}
for (die_1 in 1:6){
  for (die_2 in 1:6){
    sum = die_1 + die_2
    B <- union(B, sum)
  }
}
# Sample space is
show(B)
```

```
## [1] 2 3 4 5 6 7 8 9 10 11 12
```

```
# Number of outcomes is number of elements of B
# First we turn B into a matrix
B <- as.matrix(B)
# Number of outcomes is
show(nrow(B))
```

```
## [1] 11
```

Events

An Event is a subset of a sample space. Here are some examples:

Example

Two dice with different colors are rolled. In how many ways can we get a sum of 7.

Solution: We have to find the number of outcomes in the following event:

$$E = \{(i, j) : i + j = 7\}$$

This happens in 6 different ways

$$7 = 1 + 6 = 2 + 5 = 3 + 4 = 4 + 3 = 5 + 2 = 6 + 1$$

The following code finds E and calculates its number of outcomes (elements).

```

# Initialize E as a matrix
E <- matrix( , nrow = 1, ncol = 2)
for (die_1 in 1:6){
  for (die_2 in 1:6){
    if (die_1 + die_2 == 7){
      E <- rbind(E,c(die_1,die_2))
    }
  }
}
# Remove the first row
E <- E[-1,]
# Show E
show(E)

```

```

##      [,1] [,2]
## [1,]    1    6
## [2,]    2    5
## [3,]    3    4
## [4,]    4    3
## [5,]    5    2
## [6,]    6    1

```

```

# Number of outcomes is number of rows of E
show(nrow(E))

```

```

## [1] 6

```

Quiz

Run the following code to answer the two questions. You have to make some changes to the program.

```

# E_3 is the event of rolling three dice and obtaining a sum of SUM
SUM <- 5
# Initialize E_3 as a matrix
E_3 <- matrix( , nrow = 1, ncol = 3)
#
for (die_1 in 1:6){
  for (die_2 in 1:6){
    for (die_3 in 1:6){
      if (die_1 + die_2 + die_3 == SUM){
        E_3 <- rbind(E_3,c(die_1, die_2, die_3))
      }
    }
  }
}
# Remove the NA NA NA row (the first row)
E_3 <- E_3[-1,]
# Show E_3
show(E_3)

```

```

##      [,1] [,2] [,3]
## [1,]    1    1    3
## [2,]    1    2    2
## [3,]    1    3    1
## [4,]    2    1    2

```

```
## [5,] 2 2 1
## [6,] 3 1 1
```

```
# Show number of rows of E_3, i.e., number of outcomes
show(nrow(E_3))
```

```
## [1] 6
```

Quiz

- 1) Three dice with different colors are rolled. In how many ways can we get a sum of 9?
- 2) Three dice with different colors are rolled. In how many ways can we get a sum of 7?

Note

Analytical solution to the above two exercises is not simple. For instance the first one is equivalent to writing 9 as the sum of three positive integers i, j and k with $1 \leq i, j, k \leq 6$. The problem of writing a number, say n , as the sum of m nonzero integers is easy and in fact the answer is $\binom{n}{m}$.

Probability

Probability of an event E is defined as

$$P(E) = \frac{|E|}{|S|} \quad \text{where } |A| = \text{number of elements of } A$$

Probability of the complement of E is

$$P(E^c) = 1 - P(E)$$

Probability of the union of two events E and F is

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Written differently

$$P(E \cap F) = P(E) + P(F) - P(E \cup F)$$

Another useful formula comes from the fact that any event E can be written in the form

$$E = (E \cap F) \cup (E \cap F^c) \quad \text{for any event } F$$

Therefore

$$P(E) = P(E \cap F) + P(E \cap F^c) - P[(E \cap F) \cap (E \cap F^c)]$$

But

$$(E \cap F) \cap (E \cap F^c) = E \cap (F \cap F^c) = E \cap \emptyset = \emptyset \implies P[(E \cap F) \cap (E \cap (F^c))] = 0$$

This gives

$$P(E) = P(E \cap F) + P(E \cap F^c)$$

This identity is useful in finding $P(E \cap F^c)$, i.e., probability of E but not F happening.

Example

Sixty percent of the students at a certain school neither have a tattoo nor a necklace. Twenty percent have tattoo and 30 percent wear a necklace. If one of the students is chosen randomly, what is the probability that this student

1. has a tattoo or wears a necklace?
2. has a tattoo and wears a necklace?
3. has a tattoo but not a necklace?

Solution: We are given

$$\text{Let } \begin{cases} T & \text{the event that a student has a tattoo} \\ N & \text{the event that a student has a necklace} \end{cases} \text{ then we are given } \begin{cases} P(T^c \cap N^c) = 0.6 \\ P(T) = 0.2 \\ P(N) = 0.3 \end{cases}$$

We have to calculate $P(T \cup N)$, $P(T \cap N)$ and $P(T \cap N^c)$.

1. Calculating $P(T \cup N)$:

$$P(T \cup N) = 1 - P[(T \cup N)^c] = 1 - P(T^c \cap N^c) = 1 - 0.6 = 0.4$$

2. Calculating $P(T \cap N)$:

$$P(T \cap N) = P(T) + P(N) - P(T \cup N) = 0.2 + 0.3 - 0.4 = 0.1$$

3. Calculating $P(T \cap N^c)$:

$$P(T) = P(T \cap N) + P(T \cap N^c) \implies P(T \cap N^c) = P(T) - P(T \cap N) = 0.2 - 0.1 = 0.1$$

Quiz

A total of 30 percent of Americans drink red wine, 8 percent drink white wine, and 5 percent drink both red and white wine.

- 1) What percent of Americans drink either red or white wine?
- 2) What percent of Americans drink neither red nor white wine?
- 3) What percent of Americans drink white wine but not red wine?
- 4) What percent of Americans drink red wine but not white wine?

Conditional Probability

Probability of event E given event F is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

This implies

$$P(E \cap F) = P(E)(E|F) = P(F)P(F|E)$$

Law of Total Probability

If sample space S is a disjoint union of events F_1, F_2, \dots, F_n then

$$P(E) = P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + \dots + P(E|F_n)P(F_n) = \sum_{k=1}^n P(E|F_k)P(F_k)$$

This gives **Bayes'** formulas

$$P(F_i|E) = \frac{P(E|F_i)P(F_i)}{P(E)} = \frac{P(E|F_i)P(F_i)}{\sum_{k=1}^n P(E|F_k)P(F_k)}$$

Example

Suppose that 5 percent of men and 0.25 percent of women are color blind. A color-blind person is chosen at random.

1. Assuming that there are an equal number of males and females, what is the probability of this person being male?
2. What if the population consisted of twice as many males as females?

Solution: Let us define events

$$\begin{cases} M & \text{males} \\ F & \text{females} \\ C & \text{color blinds} \end{cases}$$

1. Equal number of males and females: $P(M) = P(F) = 0.5$.

$$P(C) = P(C|M)P(M) + P(C|F)P(F) = (0.05)(0.5) + (0.0025)(0.5) = 0.02625$$

$$P(M|C) = \frac{P(C|M)P(M)}{P(C)} = \frac{(0.05)(0.5)}{0.02625} = 0.9523 = \frac{20}{21}$$

2. Twice as many males as females: $P(M) = 2P(F) = \frac{2}{3}$

$$P(C) = (0.05)\frac{2}{3} + (0.0025)\frac{1}{3} = 0.034167$$

$$P(M|C) = \frac{P(C|M)P(M)}{P(C)} = \frac{(0.05)\frac{2}{3}}{0.034167} = \frac{40}{41}$$

Example

Prostate cancer is the most common type of cancer found in males. As an indicator of whether a male has prostate cancer, doctors often perform a test that measures the level of the prostate specific antigen (PSA) that is produced only by the prostate gland. Although PSA levels are indicative of cancer, the test is notoriously unreliable. Indeed, the probability that a noncancerous man will have an elevated PSA level is approximately 0.135, increasing to approximately 0.268 if the man does have cancer. If, on the basis of other factors, a physician is 70 percent certain that a male has prostate cancer, what is the conditional probability that he has the cancer given that

1. the test indicated an elevated PSA level?
2. the test did not indicate an elevated PSA level?

Repeat the preceding calculation, this time assuming that the physician initially believes that there is a 30 percent chance that the man has prostate cancer.

Solution

Define the events

$$\begin{cases} E & \text{a randomly selected person has cancer} \\ F & \text{a randomly selected person tests positive} \end{cases}$$

1.

$$P(E|F) = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^c)P(E^c)} = \frac{(0.268)(0.7)}{(0.268)(0.7) + (0.135)(1 - 0.7)} = 0.8224$$

2.

$$\begin{aligned} P(E|F^c) &= \frac{P(F^c|E)P(E)}{P(F^c|E)P(E) + P(F^c|E^c)P(E^c)} = \\ &= \frac{(1 - 0.268)(0.7)}{(1 - 0.268)(0.7) + (1 - 0.135)(1 - 0.7)} = 0.6638 \end{aligned}$$

Example

Suppose that an insurance company classifies people into one of three classes: good risks, average risks, and bad risks. The company's records indicate that the probabilities that good-, average-, and bad-risk persons will be involved in an accident over a 1-year span are, respectively, 0.05, 0.15, and 0.30. Suppose 20 percent of the population is a good risk, 50 percent an average risk, and 30 percent a bad risk. Define the following events

$$\begin{cases} GR & \text{a randomly selected person is good risk} \\ AR & \text{a randomly selected person is average risk} \\ BR & \text{a randomly selected person is bad risk} \\ AC & \text{a randomly selected person has an accident in a year} \end{cases}$$

- 1) What proportion of people have accidents in a fixed year? I.e., find $P(AC)$.
- 2) If policyholder A had no accidents in 1997, what is the probability that he or she is a good risk? I.e., find $P(GR|AC^c)$.
- 3) If policyholder A had no accidents in 1997, what is the probability that he or she is a average risk? I.e., find $P(AR | AC^c)$.

Solution

$$P(AC) = P(AC|GR)P(GR) + P(AC|AR)P(AR) + P(AC|BR)P(BR) = \\ (0.05)(0.2) + (0.15)(0.5) + (0.3)(0.3) = 0.175$$

$$P(GR|AC^c) = \frac{P(AC^c|GR)P(GR)}{P(AC^c)} = \frac{(1 - 0.05)(0.2)}{1 - 0.175} = \frac{38}{165} \approx 0.2303$$

$$P(AR|AC^c) = \frac{P(AC^c|AR)P(AR)}{P(AC^c)} = \frac{(1 - 0.05)(0.5)}{1 - 0.175} = \frac{17}{33} \approx 0.5151$$

Random Variables, Probability Distributions, Expected Value, and Variance

A random variable X is a function from a sample space into \mathbb{R} , i.e., $X : S \rightarrow \mathbb{R}$ with the property that for every $a \in \mathbf{R}$ the set $X^{-1}(a)$ is an event in S . Most often this condition is satisfied and one should not worry too much about it. We give several examples to clarify this definition. Before that we introduce the notation $\{X = a\}$ which means $X^{-1}(a)$, i.e., those values x in S for which $X(x) = a$.

$$\{X = a\} = X^{-1}(a) = \{x : X(x) = a\} \subseteq S$$

Probability Distributions

A random variable is **discrete** if it takes values in a countable set. Such random variables are usually counts. The probability mass function (**pmf**), f , of a discrete random variable is

$$f(a) = P(\{X = a\}) = P(X^{-1}(a))$$

A random variable is **continuous** if it takes values in an infinite (uncountable) set. Such random variables are usually measurements. The cumulative distribution function (**cdf**), F , of a continuous random variable is

$$F(x) = P(\{X \leq x\}) = P[X^{-1}((-\infty, x])]$$

A **cdf** is right continuous and non decreasing. If continuous its derivative $f(x) = F'(x)$ is called the probability density function (**pdf**) of X . A discontinuous **cdf** gives rise to a **pdf** that is not differentiable at points where the **cdf** is discontinuous. Such **cdf**'s are common in **Actuarial Mathematics**.

Expected Value

For a discrete random variable X with **pmf** f the expected value is defined

$$E[X] = \sum_x xP(\{X = x\}) = \sum_x xf(x)$$

Expected value of a function of X , $g(X)$ is defined

$$E[g(X)] = \sum_x g(x)f(x)$$

For a continuous random variable with continuous **pdf** f we have

$$E[X] = \int_{-\infty}^{\infty} xf(x) dx$$

Similarly the expected value of a function of X , $g(X)$ is defined

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x) dx$$

Variance

In both cases variance can be defined in the following way

$$\text{Var}[X] = E[X^2] - (E[X])^2$$

Example of a Discrete Probability Distribution

Suppose you roll three dice with different colors then add the numbers and find the non trivial prime factors of it. Define a random variable X as the sum of these prime factors. For example if you roll $(6, 6, 4)$ then $6 + 6 + 4 = 16 = 2 * 2 * 2 * 2$ which gives $X((6, 6, 4)) = 2 + 2 + 2 + 2 = 8$. We also have $X((6, 6, 6)) = 8$.

- Find the **pmf** of X , i.e., $f(x)$.
- Find $E[X]$.
- Find $\text{Var}[X]$.

The following program calculates number of times each sum occurs. Use it to answer the above three questions.

```
# Initialize E_3 as a matrix
E_3 <- matrix( , nrow = 1, ncol = 3)
#
for (die_1 in 1:6){
  for (die_2 in 1:6){
    for (die_3 in 1:6){
      E_3 <- rbind(E_3,c(die_1, die_2, die_3))
    }
  }
}
#
```

```

# Remove the NA NA NA row (the first row)
#
E_3 <- E_3[-1,]
#
# Find sum of rows
#
E_3_row_sum <- as.matrix(rowSums(E_3))
#
# Define a function that finds the sum of the prime factors (excluding 1)
#
sum_prime <- function(x){sum(primeFactors(x))}
#
# Apply this function to E_3_row_sum
#
prime_factors_sum <- apply(E_3_row_sum,1,sum_prime)
#
# Find the max and min of prime_factors_sum
#
sum_max <- max(prime_factors_sum)
sum_min <- min(prime_factors_sum)
#
# Find number of occurrences for each number between sum_min and sum_max
# Initialize a row vector that keeps the number of occurrences
#
w <- matrix(0, nrow = 2, ncol = sum_max - sum_min + 1)
#
# Populate w
#
for (j in 1:sum_max - sum_min + 1){
  w[1,j] <- j + (sum_min - 1)
  w[2,j] <- sum(prime_factors_sum == j + (sum_min - 1))
}
show(w)

##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
## [1,]    3    4    5    6    7    8    9   10   11   12   13   14   15
## [2,]    1    3   16   46   67   17   15    0   27    0   21    0    0
##      [,14] [,15]
## [1,]     16     17
## [2,]      0      3

#
# Find Probabilities
#
w[2,] <- as.matrix(w[2,])/(6^3)
#
# UNCOMMENT TO SEE THE PROBABILITY TABLE
#
# show(w)
#
# Find expected value
#
# Find the product of the columns of the matrix w
#

```

```

x_times_fx <- apply(w,2,prod)
#
# Expected value in the sum of x_times_fx
#
expected_value_of_X <- sum(x_times_fx)
#
# UNCOMMENT TO SEE THE EXPECTED VALUE OF X
#
# show(expected_value_of_X)
#
# Calculate variance
# Square the first row of w
#
x_squared <- function(x){x^2}
w[1,]<- apply(as.matrix(w[1,]),1, x_squared)
#
# Find x^2 f(x)
x_squared_times_fx <- apply(w,2,prod)
#
# Variance
#
var_X <- sum(x_squared_times_fx) - (expected_value_of_X)^2
#
# UNCOMMENT TO SEE THE VARIANCE OF X
#
# show(var_X)
#

```

The first row of the output shows the sums that occur. The minimum is 3 which is $X((1, 1, 1))$ and the maximum is 17, which occurs three times, i.e.,

$$X((6, 6, 5)) = X((6, 5, 6)) = X((5, 6, 6)) = 17$$

As another example 8 occurs 17 times, two of them are $X((6, 6, 6))$ and $X((6, 6, 4))$. Use this table to calculate probabilities. Remember there are $6^3 = 216$ cases, i.e., the sample space has 216 outcomes. Uncomment certain lines and run the program to check your answers.

Answer: The pmf is:

X	3	4	5	6	7	8	9	11	13	17
$f(x)$	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{16}{216}$	$\frac{46}{216}$	$\frac{67}{216}$	$\frac{17}{216}$	$\frac{15}{216}$	$\frac{27}{216}$	$\frac{21}{216}$	$\frac{3}{216}$

Example of a Continuous Probability Distribution

n random numbers are generated using four distributions. By looking at the histograms guess the distribution and use R to find the expected value and variance of each set of outcomes.

```

#
# Number of sample points
#
n <- 100000
#
# Bin width

```

```

#
m <- 0.1
#
# Generate Samples
sample_1 <- as.matrix(rnorm(n, 5, 0.25))
sample_2 <- as.matrix(rexp(n,5))
sample_3 <- as.matrix(rgamma(n, 2, 5))
sample_4 <- as.matrix(rbeta(n, 2,3))

#
# Plot the histograms
#
require(gridExtra)

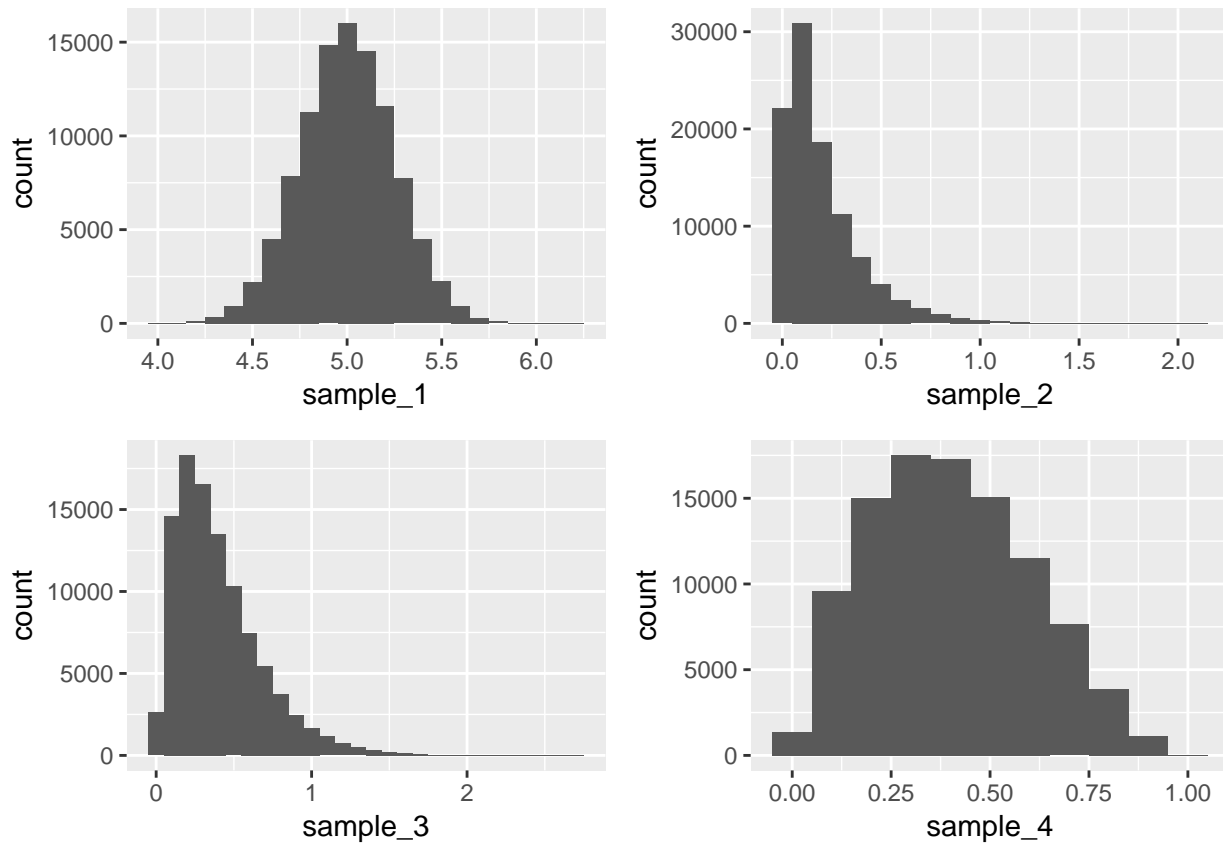
## Loading required package: gridExtra

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##      combine

plot1 <- qplot(sample_1, geom="histogram", binwidth=m)
plot2 <- qplot(sample_2, geom="histogram", binwidth=m)
plot3 <- qplot(sample_3, geom="histogram", binwidth=m)
plot4 <- qplot(sample_4, geom="histogram", binwidth=m)
#
# Plot all four in a grid
#
grid.arrange(plot1, plot2, plot3, plot4, ncol = 2, nrow = 2)

```



Expected value and variance of the above samples are

```
#
# Number of sample points. Change it as you like.
#
n <- 100000
#
# Bin width. Change it to see the effect on the plots.
#
m <- 0.1
#
# Generate Samples
sample_1 <- as.matrix(rnorm(n, 5, 0.25))
sample_2 <- as.matrix(rexp(n,5))
sample_3 <- as.matrix(rgamma(n, 2, 5))
sample_4 <- as.matrix(rbeta(n, 2,3))
#
# Expected values, aka Means
#
mean_1 <- mean(sample_1)
mean_2 <- mean(sample_2)
mean_3 <- mean(sample_3)
mean_4 <- mean(sample_4)
show(c("Expected Values", mean_1,mean_2,mean_3,mean_4))

## [1] "Expected Values"    "5.00004585497432"  "0.199663924372984"
## [4] "0.401518898248772"  "0.400996352097969"
```

```

#
# standard deviation = sqrt(Variance)
#
var_1 <- (sd(sample_1))^2
var_2 <- (sd(sample_2))^2
var_3 <- (sd(sample_3))^2
var_4 <- (sd(sample_4))^2
show(c("Variances", var_1,var_2,var_3,var_4))

## [1] "Variances"          "0.0618604372902574" "0.0401456155731336"
## [4] "0.0800690024080002" "0.0399319398410465"

```

In the above two codes there are four commands that generate random values from four distributions. The “r” stands for random and the rest is the distribution’s name (or part of it). Study these four lines and connect the expected values and variances with the numbers in paranthesis for each distribution.