STATISTICS AND DATA ANALYSIS

TA Session: Probability Practice

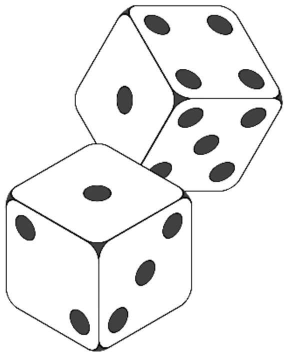
October 6, 2014

Practice 1

A fare dice is rolled and a fair coin is tossed. Find the probability that the dice shows an odd number and the coin shows a head.

- The probability of getting an **odd number** after rolling a dice is 3/6.
- The probability of showing a **head** after tossing a coin is 1/2.
- Because they are **independent events**, all we need to do is to **time** the two probabilities together:

3/6 * 1/2 = 1/4



Practice 2

Suppose A and B are independent events, B and C are mutually exclusive, and A and C are independent events. Moreover, we have $\Pr(A)$=0.4, $\Pr(B)$=0.9 and $\Pr(C)$=0.1. Find the following probabilities:

a) $\Pr(A \cap B)$
b) $\Pr(B \cap C)$
c) $\Pr(A \cap \mathrm{B} \cap C')$
d) $\Pr((A \cap C) \cup B)$
e) $\Pr((A \cap C') \cup B)$

Practice 2

Suppose A and B are independent events, B and C are mutually exclusive, and A and C are independent events. Moreover, we have $\Pr(A)$=0.4, $\Pr(B)$=0.9 and $\Pr(C)$=0.1. Find the following probabilities:

a) $\mathbf{Pr}(\boldsymbol{A} \cap \boldsymbol{B})$
b) $\Pr(B \cap C)$
c) $\Pr(A \cap B \cap C')$
d) $\Pr((A \cap C) \cup B)$
e) $\Pr((A \cap C') \cup B)$

Solution:

Because A and B are independent events.

$\Pr(\boldsymbol{A}\cap\boldsymbol{B}) = \Pr(\boldsymbol{A}) \times \Pr(\boldsymbol{B}) = 0.36$

Practice 2

Suppose *A* and *B* are independent events, *B* and *C* are mutually exclusive, and *A* and *C* are independent events. Moreover, we have $\Pr(A)$=0.4, $\Pr(B)$=0.9 and $\Pr(C)$=0.1. Find the following probabilities:

a) $\Pr(A \cap B)$
b) $\boldsymbol{\Pr(B \cap C)}$
c) $\Pr(A \cap B \cap C')$
d) $\Pr((A \cap C) \cup B)$
e) $\Pr((A \cap C') \cup B)$

Solution:

Because *B* and *C* are mutually exclusive,
They have no intersection!
$\Pr(B \cap C) = 0$

Practice 2

Suppose *A* and *B* are independent events, *B* and *C* are mutually exclusive, and *A* and *C* are independent events. Moreover, we have $\Pr(A)$=0.4, $\Pr(B)$=0.9 and $\Pr(C)$=0.1. Find the following probabilities:

a) $\Pr(A \cap B)$

b) $\Pr(B \cap C)$

c) $\mathbf{Pr}(\boldsymbol{A} \cap \boldsymbol{B} \cap \boldsymbol{C'})$

d) $\Pr((A \cap C) \cup B)$

e) $\Pr((A \cap C') \cup B)$

Solution:

$\Pr(A \cap B \cap C')$

$= \Pr(A \cap B \cap B)$

$= \Pr(A \cap B)$

$= 0.36$

Practice 2

Suppose *A* and *B* are independent events, *B* and *C* are mutually exclusive, and *A* and *C* are independent events. Moreover, we have $\Pr(A)$=0.4, $\Pr(B)$=0.9 and $\Pr(C)$=0.1. Find the following probabilities:

a) $\Pr(A \cap B)$
b) $\Pr(B \cap C)$
c) $\Pr(A \cap B \cap C')$
d) $\boldsymbol{\Pr((A \cap C) \cup B)}$
e) $\Pr((A \cap C') \cup B)$

Solution:

$\Pr((A \cap C) \cup B)$
$= \Pr(A \cap C) + \Pr(B)$
$= 0.94$

Practice 2

Suppose *A* and *B* are independent events, *B* and *C* are mutually exclusive, and *A* and *C* are independent events. Moreover, we have $\Pr(A)$=0.4, $\Pr(B)$=0.9 and $\Pr(C)$=0.1. Find the following probabilities:

a) $\Pr(A \cap B)$
b) $\Pr(B \cap C)$
c) $\Pr(A \cap B \cap C')$
d) $\Pr((A \cap C) \cup B)$
e) $\mathbf{Pr}((\boldsymbol{A} \cap \boldsymbol{C}') \cup \boldsymbol{B})$

Solution:

$\Pr\big((A \cap C') \cup B\big)$
$= \Pr((A \cap B) \cup B)$
$= \Pr(B)$
$= 0.9$

Practice 3

You decide you're only going to buy a lottery ticket if your expected winning is larger than the ticket price. Suppose a ticket costs $10:

With probability 0.01, you win $1000.

With probability 0.05, you win $100.

With probability 0.1, you win $10.

Should you buy a ticket for this lottery? Why?

Practice 3

- **Four** kinds of possible outcomes:
  1) Your win \$1000 with probability 0.01.
  2) You win \$100 with probability 0.05.
  3) You win \$10 with probability 0.1.
  4) You win **nothing** with probability 1-0.01-0.05-0.1=0.84.

- **The expected value** of winning the lottery:

$$1000\times0.01+100\times0.05+10\times0.1+0\times0.84=10+5+1=16(\$)$$

- 16 > 10. We **should** buy the ticket.

Practice 4

Define X = number of **heads** after **3 flips** of an **unfair coin** with the following distribution:

$$\Pr(X = \text{Head}) = 0.3 \text{ and } \Pr(X = \text{Tail}) = 0.7.$$

a) List all the possible outcomes of X.
b) What are the probabilities of all the outcomes of X? (You may use R as a calculator.)

Practice 4

Define X = number of **heads** after **3 flips** of an **unfair coin** with the following distribution:

$$\Pr(X = \text{Head}) = 0.3 \text{ and } \Pr(X = \text{Tail}) = 0.7.$$

a) List all the possible outcomes of X.

X	Combinations
$X = 0$	(TTT)
$X = 1$	(HTT), (THT), (TTH)
$X = 2$	(HHT), (HTH), (THH)
$X = 3$	(HHH)

Practice 4

Define X = number of **heads** after **3 flips** of an **unfair coin** with the following distribution:

$$\Pr(X = \text{Head}) = 0.3 \text{ and } \Pr(X = \text{Tail}) = 0.7.$$

b) What are the probabilities of all the outcomes of X? (You may use R as a calculator.)

X	**Combinations**	**Probabilities of** X
$X = 0$	(TTT)	**0.7×0.7×0.7** = 0.343
$X = 1$	(HTT), (THT), (TTH)	**0.3×0.7×0.7** + **0.7×0.3×0.7** + **0.7×0.7×0.3** = 0.441
$X = 2$	(HHT), (HTH), (THH)	**0.3×0.3×0.7** + **0.3×0.7×0.3** + **0.7×0.3×0.3** = 0.189
$X = 3$	(HHH)	**0.3×0.3×0.3** = 0.027

Practice 4

Define X = number of **heads** after **3 flips** of an **unfair coin** with the following distribution:

$$\Pr(X = \text{Head}) = 0.3 \text{ and } \Pr(X = \text{Tail}) = 0.7.$$

c) Using the R code mentioned in the video, find the expected value of X.

d) Using the R code mentioned in the video, find the variance and the standard deviation of X.

The **expected value** (or mean) of X is

$$\mu \equiv \mathbb{E}[X] = \sum_{i \in S} x_i \Pr(x_i).$$

The **variance** of X is

$$\sigma^2 \equiv \text{Var}(X) \equiv \mathbb{E}\left[(X-\mu)^2\right] = \sum_{i \in S} (x_i - \mu)^2 \Pr(x_i).$$

The **standard deviation** of X is $\sigma \equiv \sqrt{\sigma^2}$.

Practice 4

Define X = number of **heads** after **3 flips** of an **unfair coin** with the following distribution:

$$\Pr(X = \text{Head}) = 0.3 \text{ and } \Pr(X = \text{Tail}) = 0.7.$$

c) Using the R code mentioned in the video, find the expected value of X.

X	$\Pr(X)$
$X = 0$	0.343
$X = 1$	0.441
$X = 2$	0.189
$X = 3$	0.027

The expected value (or mean) of X is

$$\mu \equiv \mathbb{E}[X] = \sum_{i \in S} x_i \Pr(x_i).$$

R code:

```
> x <- 0:3
> p <- c(0.343, 0.441, 0.189, 0.027)
> m <- sum(x * p)
```

Practice 4

Define X = number of **heads** after **3 flips** of an **unfair coin** with the following distribution:

$$\Pr(X = \text{Head}) = 0.3 \text{ and } \Pr(X = \text{Tail}) = 0.7.$$

d) Using the R code mentioned in the video, find the variance and the standard deviation of X.

The variance of X is

$$\sigma^2 \equiv \text{Var}(X) \equiv \mathbb{E}[(X-\mu)^2] = \sum_{i \in S} (x_i - \mu)^2 \Pr(x_i).$$

The standard deviation of X is $\sigma \equiv \sqrt{\sigma^2}$.

R code:

```
> v <- sum((x - m) ^ 2 * p)
> s <- sqrt(v)
```

Practice 5

Consider the wholesale data set:

Load the data set "SDA-Fa14 data wholesale.txt" by executing this Statement.

```
W <- read.table("SDA-Fa14_data_wholesale.txt", header=TRUE)
```

Practice 5

a) Extract sales data collected from channel 1 and region 1, with 4 columns: Channel, Region, Fresh and Milk.

Hint:

Use "which" and "data.frame" function.

R code:

```
index <- which(W$Channel == 1 & W$Region == 1)
ws <- W[ index,  ]
ews <- data.frame( Channel = ws$Channel,
                   Region = ws$Region,
                   Fresh = ws$Fresh,
                   Milk = ws$Milk )
```

Practice 5

b) For sales data collected from channel 1 and region 1, calculate the means, medians, and sample variances for milk sales.

Hint:
Use "mean", "median" and "var" function.
You can also use "summary" function to see what happened.

R code:

```
mean( ews$Milk )
median( ews$Milk )
var( ews$Milk )
summary( ews$Milk )
```

Practice 5

c) For sales data collected from channel 1 and region 1, draw a histogram for milk sales data with the default number of classes and class intervals.

Hint:
Use “hist” function.

R code:

```
hist( ews$Milk )
```

Practice 5

d) For each of the six channel-region combination, calculate the sample correlation coefficient between fresh food sales and milk sales.

R code:

```
C1R1 <- W[ which(W$Channel==1 & W$Region==1), ]
C1R2 <- W[ which(W$Channel==1 & W$Region==2), ]
C1R3 <- W[ which(W$Channel==1 & W$Region==3), ]
C2R1 <- W[ which(W$Channel==2 & W$Region==1), ]
C2R2 <- W[ which(W$Channel==2 & W$Region==2), ]
C2R3 <- W[ which(W$Channel==2 & W$Region==3), ]
```

R code:

```
cor(C1R1$Fresh, C1R1$Milk)      # -0.03010351
cor(C1R2$Fresh, C1R2$Milk)      # 0.5380095
cor(C1R3$Fresh, C1R3$Milk)      # 0.2912192
cor(C2R1$Fresh, C2R1$Milk)      # 0.1291734
cor(C2R2$Fresh, C2R2$Milk)      # -0.2059987
cor(C2R3$Fresh, C2R3$Milk)      # 0.2761697
```

Practice 5

e) Draw scatter plots for the channel-region combinations with the highest and lowest correlation coefficients.

Hint:
Use "plot" function.

R code:

```
plot(C1R2$Fresh, C1R2$Milk) # Pay attention to the outlier!
plot(C2R2$Fresh, C2R2$Milk)
```