CS544 Module2

Suresh Kalathur

Module2

- Probability
- Conditional Probability
- R Programming Constructs

Probability

- Random Experiment
- Sample Space
  - Set of all possible outcomes
- “prob” package of R
  - Common sample spaces
    - Tossing coins, rolling dice, cards, etc.
- Sampling from an Urn
- Event
  - Subset of sample space

Probability using R

Package prob

```
> library(prob)
> S <- tosscoin(3, makespace = TRUE)
> S
  toss1 toss2 toss3 probs
1     H     H     H 0.125
2     T     H     H 0.125
3     H     T     H 0.125
4     T     T     H 0.125
5     H     H     T 0.125
6     T     H     T 0.125
7     H     T     T 0.125
8     T     T     T 0.125
```

…Probability using R

```
> S <- rolldie(2, makespace = TRUE)
> head(S, n = 3)
  X1 X2      probs
1  1  1 0.02777778
2  2  1 0.02777778
3  3  1 0.02777778
> tail(S, n = 3)
   X1 X2      probs
34  4  6 0.02777778
35  5  6 0.02777778
36  6  6 0.02777778
```

Prob function

- Probability of the event

```
> S <- cards(makespace = TRUE)
```

```
> A <- subset(S, rank == "Q")
> A
   rank    suit      probs
11    Q    Club 0.01923077
24    Q Diamond 0.01923077
37    Q   Heart 0.01923077
50    Q   Spade 0.01923077
```

```
> Prob(A)
[1] 0.07692308
>
> Prob(S, rank == "Q")
[1] 0.07692308
```

...Probability using R

Add random variable

```
> S <- rolldie(2, makespace = TRUE)
> S <- addrv(S, U = X1 + X2)
> head(S, n = 2)
  X1 X2 U      probs
1  1  1 2 0.02777778
2  2  1 3 0.02777778
> tail(S, n = 2)
   X1 X2  U      probs
35  5  6 11 0.02777778
36  6  6 12 0.02777778
```

```
> S <- rolldie(2, makespace = TRUE)
> S <- addrv(S, FUN = sum, name = "U")
> head(S, n = 2)
  X1 X2 U      probs
1  1  1 2 0.02777778
2  2  1 3 0.02777778
> tail(S, n = 2)
   X1 X2  U      probs
35  5  6 11 0.02777778
36  6  6 12 0.02777778
```

```
> S <- rolldie(2, makespace = TRUE)
> mySum <- function(data) { data[1] + data[2] }
> S <- addrv(S, FUN = mySum, name = "U")
> head(S, n = 2)
  X1 X2 U      probs
1  1  1 2 0.02777778
2  2  1 3 0.02777778
> tail(S, n = 2)
   X1 X2  U      probs
35  5  6 11 0.02777778
36  6  6 12 0.02777778
```

...Probability using R

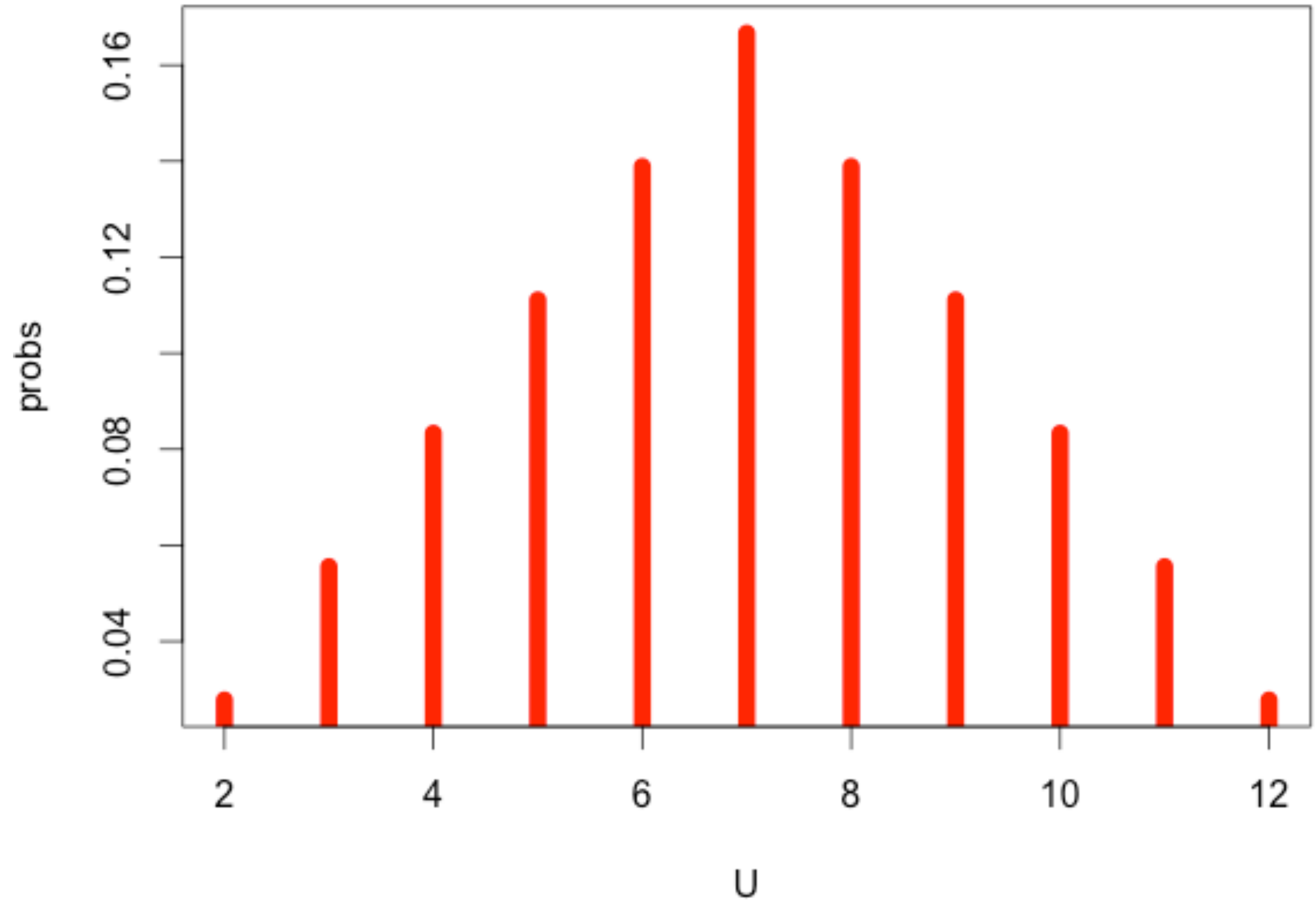
Marginal distribution

```
> S <- marginal(S, vars = "U")
> S
    U      probs
1   2 0.02777778
2   3 0.05555556
3   4 0.08333333
4   5 0.11111111
5   6 0.13888889
6   7 0.16666667
7   8 0.13888889
8   9 0.11111111
9  10 0.08333333
10 11 0.05555556
11 12 0.02777778
```

...R

Plot

```
> plot(probs ~ U, S, type='h', col = "red", lwd=10)
```



Conditional Probability

- P(B|A)

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

- Multiplication Rule

$$P(A \cap B) = P(A \text{ and } B) = P(A) \cdot P(B|A)$$

- Independent Events

$$P(A \cap B) = P(A) \cdot P(B)$$

Bayes Theorem

- Developed by Reverend Bayes
  - To infer the existence of God
- Historical
  - Cracking the infamous Nazi Enigma code in WWII (Alan Turing)
- Finance & Business
  - Evaluating interest rates
  - Managing net income streams
  - Lending Credit
- Insurance Companies
  - Risk of flooding in coastal areas
- Health
  - Probability of having disease X given that test Y is positive
- AI - Driverless vehicles
  - Improving decision making using probabilities on road conditions
- AI – Robots
  - Robot’s next step given the steps it already has executed
- Others
  - Sort spam from e-mail

Bayes Theorem

$$P(A|B) = \frac{P(A)\,P(B|A)}{P(B)}$$

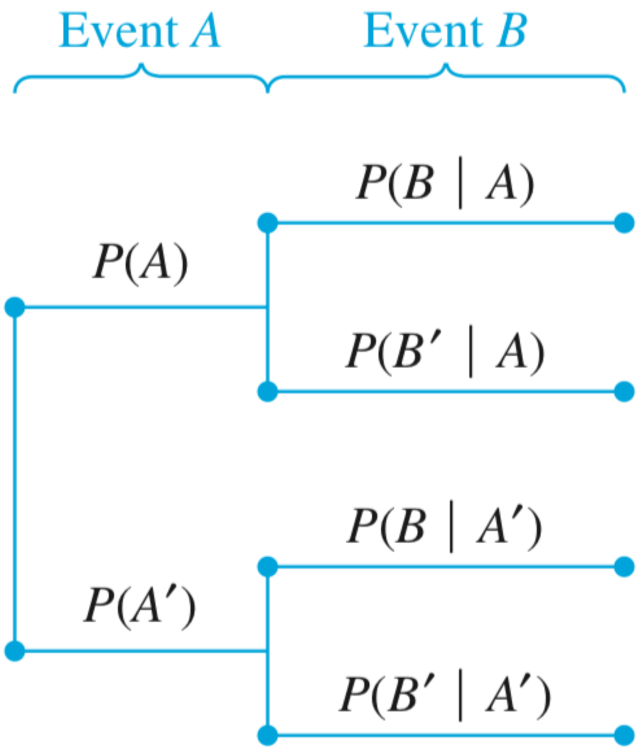
- Do a search for
  - Automatic shoe laces movie
- Result
  - Back to the future

- What we know
  - P(A) – how likely A is on its own
  - P(B) – how likely B is on its own
  - P(B|A) – how often B happens given that A happens
- What the theorem tells us?
  - How often A happens given that B happens , **P(A|B)**

Example (Fire and Smoke)

- P(Fire) – how often there is a fire
- P(Smoke) – how often we see smoke
- P(Smoke|Fire) – how often we can see smoke given there is fire
- P(Fire|Smoke) – how often there is fire given we can see smoke
- Given that dangerous fires are rare (1%), smoke is fairly common (10%), and that 90% of dangerous fires make smoke
  - P(Fire) = 0.01, P(Smoke) = 0.10, P(Smoke|Fire) = 0.90
- What is the probability of a dangerous fire given that we see a smoke?
  - P(Fire|Smoke) = $\frac{P(Fire) * P(Smoke|Fire)}{P(Smoke)} = \frac{0.01 * 0.90}{0.10}$ = 0.09
- Answer: 9% probability of a dangerous fire given we sighted smoke

More Examples: https://www.mathsisfun.com/data/bayes-theorem.html



Bayes Theorem…

- Forward looking probability
  - Probability that event B will occur given event A occurred
  - Given for us
- Backward looking probability
  - Probability that event A has occurred given event B has occurred

Rule of Total Probability

Rule of Total Probability

Suppose the events A_1, A_2, …, A_k are **mutually exclusive** and **exhaus**tive, i.e., exactly one of these events will occur and they cover the entire sample space.

For any event B, the events (A_1 and B), (A_2 and B), …, (A_k and B) are mutually exclusive, and hence P(B) =

P(A_1 and B) + P(A_2 and B) + … + P(A_k and B)

Using the multiplication rule,

$$P(B) = P(B|A_1)*P(A_1) + P(B|A_2)*P(A_2) + \ldots + P(B|A_k)*P(A_k)$$

$$P(B) = \sum_{j=1}^{k} P(B|A_j) * P(A_j)$$

Bayes' Theorem

Bayes' Theorem:

Suppose the events A_1, A_2, …, A_n are mutually exclusive and exhaustive. Let B be any event.

Given

Prior probabilities: $P(A_1)$, $P(A_2)$, …, $P(A_n)$, and

Conditional probabilities: $P(B|A_1)$, $P(B|A_2)$, …, $P(B|A_n)$

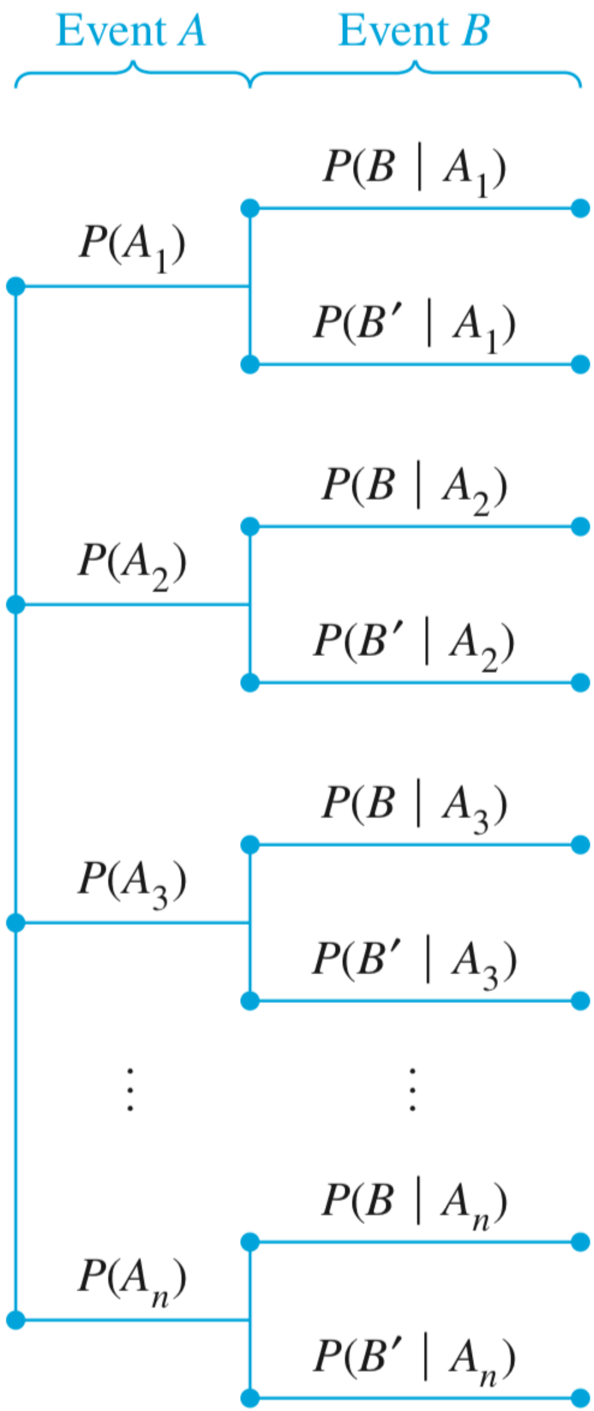
Determine

Posterior probabilities: $P(A_1|B)$, $P(A_2|B)$, …, $P(A_n|B)$

$$P(A_i|B) = \frac{P(A_i \; and \; B)}{P(B)} = \frac{P(B|A_i) * P(A_i)}{P(B)}$$

$$P(A_i|B) = \frac{P(B|A_i) * P(A_i)}{\sum_{j=1}^{n} P(B|A_j) * P(A_j)}$$

Bayes Theorem…



$$P(B) = P(A_1)*P(B|A_1) + P(A_2)*P(B|A_2) + \ldots + P(A_n)*P(B|A_n)$$

$$P(A_1|B) = \frac{P(A_1)*P(B|A_1)}{P(B)}$$

…

$$P(A_n|B) = \frac{P(A_n)*P(B|An)}{P(B)}$$

Example1 – Rule of Total Probability

Example: In an university, 60% are undergraduate students, 35% are graduate students, and 5% are postdocs. 55% of undergraduates are female, 15% of graduate students are female, and 10% of postdocs are female.
What is the probability that a randomly selected student is female?

Event B = Selected student is a female

Event A1 = Selected student is an undergraduate Event A2 = Selec
Event A3 = Selected student is a postdoc

A1, A2, and A3 are mutually exclusive and exhaustive P(B) = P(A1

P(B) = P(B|A1)*P(A1) + P(B|A2)*P(A2) + P(B|A3)*P(A3)
= 0.55*0.60 + 0.15*0.35 + 0.10*0.05
= 0.3875

With a probability of 0.3875, a randomly selected student is a female

Type	Percentage of college students	Percentage females
Undergraduate	60	55
Graduate	35	15
Postdoc	5	10
	100%	

P(A1) = 0.60	P(B|A1) = 0.55
P(A2) = 0.35	P(B|A2) = 0.15
P(A3) = 0.05	P(B|A3) = 0.10

Example1 - Bayes' Theorem

Example: In an university, 60% are undergraduate students, 35% are graduate students, and 5% are postdocs. 55% of undergraduates are female, 15% of graduate students are female, and 10% of postdocs are female. What is the probability that a randomly selected female student is:

an undergraduate? a graduate? A postdoc?

Type	Percentage of college students	Percentage females
Undergraduate	60	55
Graduate	35	15
Postdoc	5	10
	100%	

Event B = Selected student is a female

Event A1 = Selected student is an undergraduate Event A2 = Se

Event A3 = Selected student is a postdoc

P(B) = P(B|A1)*P(A1) + P(B|A2)*P(A2) + P(B|A3)*P(A3)

= 0.55*0.60 + 0.15*0.35 + 0.10*0.05

= 0.3875

P(A1) = 0.60	P(B|A1) = 0.55
P(A2) = 0.35	P(B|A2) = 0.15
P(A3) = 0.05	P(B|A3) = 0.10

P(A1|B) = P(B|A1)*P(A1)/P(B) = 0.55*0.60/0.3875 = 0.85 P(A2|B) = P(

= 0.14 P(A3|B) = P(B|A3)*P(A3)/P(B) = 0.10*0.05/0.3875 = 0.01

With a probability of 0.85, a randomly selected female student is an Un

Example2 – Rule of Total Probability

Example: A company orders parts from three different suppliers, *Supplier1*, *Supplier2*, and *Supplier3*. From historical records, 3% of parts provided by *Supplier1* are defective, 5% of parts provided by *Supplier2* are defective, and 4% of parts provided by *Supplier3* are defective. The current inventory consists of 5000 units from *Suppler1*, 3500 units from *Supplier2*, and 2000 units from *Supplier3*.

What is the probability that a randomly selected part is defective?

Type	Inventory	Percentage Defective
Supplier1	5000	3
Supplier2	3500	5
Supplier3	2000	4
	10500	

Event D = Selected part is a defective one

Event S1 = Selected part is from Supplier1

Event S2 = Selected part is from Supplier2

Event S3 = Selected part is from Supplier3

S1, S2, and S3 are mutually exclusive and exhaustive

P(D) = P(S1 and D) + P(S2 and D) + P(S3 and D)

P(D) = P(D|S1)*P(S1) + P(D|S2)*P(S2) + P(D|S3)*P(S3)

= 0.03*0.48 + 0.05*0.33 + 0.04*0.19

= 0.039

So, there is a 4% chance that a randomly selected part is a defective

$P(S1) = \frac{50}{105} = 0.48$	$P(D\|S1) = 0.03$
$P(S2) = \frac{35}{105} = 0.33$	$P(D\|S2) = 0.05$
$P(S3) = \frac{20}{105} = 0.19$	$P(D\|S3) = 0.04$

Example2 – Bayes Theorem

Example: A company orders parts from three different suppliers, *Supplier1*, *Supplier2*, and *Supplier3*. From historical records, 3% of parts provided by *Supplier1* are defective, 5% of parts provided by *Supplier2* are defective, and 4% of parts provided by *Supplier3* are defective. The current inventory consists of 5000 units from *Suppler1*, 3500 units from *Supplier2*, and 2000 units from *Supplier3*.

What is the probability that a randomly selected defective part: came from *Supplier1*? Came from *Supplier2*? Came from *Supplier3*?

Type	Inventory	Percentage Defective
Supplier1	5000	3
Supplier2	3500	5
Supplier3	2000	4
	10500	

Event D = Selected part is a defective one

Event S1 = Selected part is from *Supplier1*

Event S2 = Selected part is from *Supplier2*

Event S3 = Selected part is from *Supplier3*

$P(S1) = \frac{50}{105} = 0.48$	$P(D\|S1) = 0.03$
$P(S2) = \frac{35}{105} = 0.33$	$P(D\|S2) = 0.05$
$P(S3) = \frac{20}{105} = 0.19$	$P(D\|S3) = 0.04$

$P(D) = P(D|S1)*P(S1) + P(D|S2)*P(S2) + P(D|S3)*P(S3)$
$= 0.03*0.48 + 0.05*0.33 + 0.04*0.19 = 0.039$

$P(S1|D) = P(D|S1)*P(S1)/P(D) = 0.03*0.48/0.039 = 0.37$
$P(S2|D) = P(D|S2)*P(S2)/P(D) = 0.05*0.33/0.039 = 0.43$
$P(S3|D) = P(D|S3)*P(S3)/P(D) = 0.04*0.19/0.039 = 0.20$

So, there is a 37% chance that a randomly selected defective part came from *Supplier1*.

R Programming Constructs

- Functions
- Scope of variables
- Control structures
  - if-else, for, while, repeat
- Reading and Writing Data