ECO239 Statistics I

Week 6

Probability

Introduction to Probability

Important terms

1. **Random Experiment** – a process leading to an uncertain outcome

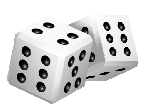
e.g. coin toss

die rolling

2. **Basic Outcomes-** a possible outcome of a random experiment

e.g. coin toss: Head or Tail

die rolling: 1, 2, 3, 4, 5, 6



3. **Sample space** – the collection of all possible outcomes of a random experiment.

e.g. coin toss S = [H, T]

die rolling S = [1,2,3,4,5,6]

4. **Event**- any subset of basic outcomes from the sample space.

e.g. die rolling: odd number { 1, 3, 5} ,even number {2, 4, 6}, less than 3 {1, 2}…

*Null event = absence of a basic outcome. $\emptyset$

5. **Intersection of Event**

If A and B are two events in a sample space S, then the intersection, A∩B, is the set of all outcomes in S that belong to both A and B.

e.g. event A: {1, 3, 5}, event B: {1, 2, 3}

=> A∩B= {1, 3}

6. **Mutually Exclusive Events**

Events A and B are mutually exclusive events if they have no basic outcomes in common.

A∩B =$\emptyset$

e.g. A={1, 3, 5} , B={2, 4, 6}

A∩B = $\emptyset$ => A and B are mutually exclusive.

7. **Union** of Events

Union A U B is the set of all outcomes in S that belong to either A or B.

A={1, 3, 5}

B={3, 4, 5}

AUB={1, 3, 4, 5}

8. Events E1, E2,… Ek are Collectively Exhaustive events if E1 U E2 U…U Ek = S.

(Union of all events = sample space itself)

e.g. Rolling a die case.

E1 = {1, 2}, E2 = {3, 4}, E3 = {5, 6}.

E1 U E2 U E3 = S.

e.g. Rolling a die case.

E1= {1, 3, 5}, E2 = {2, 4, 6}, E3 = {3, 4, 5}

E1 U E2 U E3 = S.

9. Complement of an event A is the set of all basic outcome in the sample space that do not belong to A.

$$A \cup \bar{A} = S$$

e.g. Rolling a die

A = {1, 3, 5}

$\bar{A}$ = {2, 4, 6}

Practice

- Rolling a die
- A= {2, 3, 6}, B = {4, 5, 6}

A∩B =

AUB=

$A \cup \bar{B}$

$\bar{A} \cup B$

$\bar{A} \cap B$

$A \cap \bar{B}$

Are A and B Mutually Exclusive?

Are A and B Collectively Exhaustive?

Important Result

1. $(A \cap B) \cup (\bar{A} \cap B) = B$

2. $A \cup (\bar{A} \cap B) = A \cup B$

3. E1, E2, … Ek: k mutually exclusive and collectively exhaustive events.

"A" is some other event.

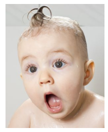
$$(E_1 \cap A) \cup (E_2 \cap A) \cup \ldots \cup (E_k \cap A) = A$$

Probability

- The chance that an uncertain event will occur.
- P(A) = Probability of an event A
- P(A) = NA/N = (# of outcomes that satisfying the event) / (total # of outcome in the sample space).
- 0≤ P(A) ≤ 1

Practice

Which of the following events would you be most surprised by?



(a) exactly 3 heads in 10 coin flips
(b) exactly 3 heads in 100 coin flips
(c) exactly 3 heads in 1000 coin flips

Practice

Which of the following events would you be most surprised by?

(a) exactly 3 heads in 10 coin flips
(b) exactly 3 heads in 100 coin flips
(c) exactly 3 heads in 1000 coin flips

Practice

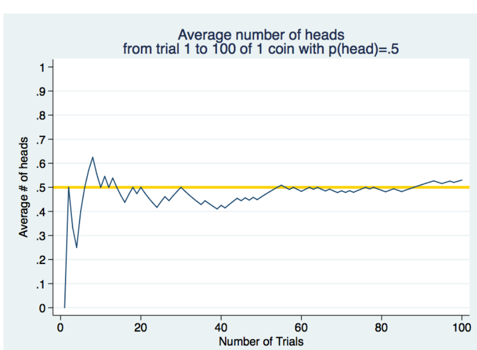
Which of the following events would you be most surprised by?

(a) exactly 3 heads in 10 coin flips
P(A) = 3/10 = 0.3.
(b) exactly 3 heads in 100 coin flips
P(A)=3/100=0.03.
(c) exactly 3 heads in 1000 coin flips
P(A)=3/1000=0.003.

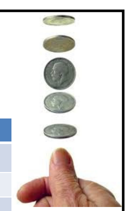
Law of large numbers

Law of large numbers states that as more observations are collected, the proportion of occurrences with a particular outcome, $\hat{p}n$, converges to the probability of that outcome, p.

LLN: Example "Coin Toss"



Let's Try



	H or T		H or T		H or T
1	H	11	T	21	
2	H	12	T	22	
3	H	13	H	23	
4	H	14	T	24	
5	T	15	T	25	
6	T	16	H	26	
7	T	17	H	27	
8	T	18	H	28	
9	T	19	H	29	
10	T	20	T	30	

Law of large numbers (cont.)

When tossing a fair coin, if heads comes up on each of the first 10 tosses, what do you think the chance is that another head will come up on the next toss?

(a) 0.5

(b) less than 0.5

(c) more than 0.5

H H H H H H H H H H ?

Law of large numbers (cont.)

When tossing a fair coin, if heads comes up on each of the first 10 tosses, what do you think the chance is that another head will come up on the next toss? 0.5, less than 0.5, or more than 0.5?

H H H H H H H H H H ?

- The probability is still 0.5, or there is still a 50% chance that another head will come up on the next toss.

$P(H \text{ on } 11^{th} \text{ toss}) = P(T \text{ on } 11^{th} \text{ toss}) = 0.5$

Practice

What is the probability of drawing **a jack or a red card** from a well shuffled full deck?

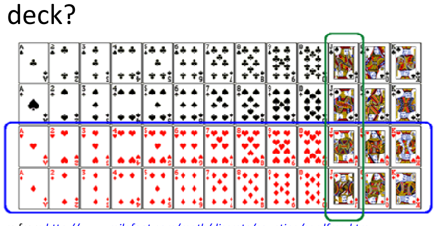
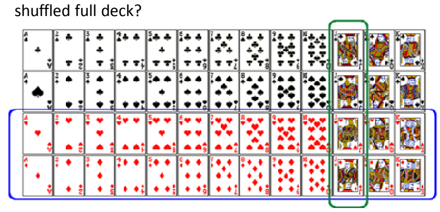


Figure from http://www.milefoot.com/math/discrete/counting/cardfreq.htm

P(AUB)=P(A)+P(B)-P(A∩B)

What is the probability of drawing a jack or a red card from a well shuffled full deck?



$$P(\text{jack or red}) = P(\text{jack}) + P(\text{red}) - P(\text{jack and red}) = \frac{4}{52} + \frac{26}{52} - \frac{2}{52} = \frac{28}{52}$$

Figure from http://www.milefoot.com/math/discrete/counting/cardfreq.htm

Practice

What is the probability that a randomly sampled student thinks marijuana should be legalized or they agree with their parents' political views?

	Share Parents' Politics		
Legalize MJ	No	Yes	Total
No	11	40	51
Yes	36	78	114
Total	47	118	165

(a) (40 + 36 - 78) / 165

(b) (114 + 118 - 78) / 165

(c) 78 / 165

(d) 78 / 188

(e) 11 / 47

Practice

What is the probability that a randomly sampled student thinks marijuana should be legalized or they agree with their parents' political views?

	Share Parents' Politics		
Legalize MJ	No	Yes	Total
No	11	40	51
Yes	36	78	114
Total	47	118	165

(a) (40 + 36 - 78) / 165

(b) (114 + 118 - 78) / 165

(c) 78 / 165

(d) 78 / 188

(e) 11 / 47

Recap

General addition rule

P(A or B) = P(A) + P(B) - P(A and B)

Note: For mutually exclusive (disjoint) events *P(A and B)* = 0, so the above formula simplifies to *P(A or B) = P(A) + P(B)*

Practice

- 75% of customers use mustard
- 80% of customers use ketchup
- 65% of customers use both.

Q: What is the probability that a customer uses at least one?

Q: What is the probability that a customer uses none of them?

Practice

- 75% of customers use mustard
- 80% of customers use ketchup
- 65% of customers use both.

Q: What is the probability that a customer uses at least one?
P(use at least one) = P(mustard)+P(ketchup)-P(mustard∩ketchup) = 0.75+0.8-0.65=0.9

Q: What is the probability that a customer uses none of them?
P(use none)= 1- P(use at least one) = 1 – 0.9 = 0.1

Combination Formula

- # of combination to pick K out of n

$$C_k^n = \frac{n!}{k!\,(n-k)!}$$

Practice $C_k^n = \frac{n!}{k!\,(n-k)!}$

- Ayse will choose 3 laptops out of 20. What is the probability of choosing 2 HP and 1 SONY out of 10 HP, 5 SONY and 5 DELL?

$$N = C_3^{20} = \frac{20!}{3!\,17!} = \frac{20*19*18}{3*2*1} = 1140$$

$$N_A = C_2^{10} C_1^5 = \frac{10!}{2!\,8!}\frac{5!}{1!\,4!} = \frac{10*9}{2*1}*\frac{5}{1} = 225$$

$$\frac{N_A}{N} = \frac{225}{1140} = 0.197$$

Probability distributions

A *probability distribution* lists all possible events and the probabilities with which they occur.

- The probability distribution for the gender of one kid:

Event	Male	Female
Probability	0.5	0.5

Probability distributions

A *probability distribution* lists all possible events and the probabilities with which they occur.

- The probability distribution for the gender of one kid:

Event	Male	Female
Probability	0.5	0.5

- Rules for probability distributions:
  1. The events listed must be disjoint
  2. Each probability must be between 0 and 1
  3. The probabilities must total 1

Probability distributions

A *probability distribution* lists all possible events and the probabilities with which they occur.

- The probability distribution for the gender of one kid:

Event	Male	Female
Probability	0.5	0.5

- Rules for probability distributions:
  1. The events listed must be disjoint
  2. Each probability must be between 0 and 1
  3. The probabilities must total 1
- The probability distribution for the genders of two kids:

Event	MM	FF	MF	FM
Probability	0.25	0.25	0.25	0.25

Practice

In a survey, 52% of respondents said they are Democrats. What is the probability that a randomly selected respondent from this sample is a Republican?

(a) 0.48
(b) more than 0.48
(c) less than 0.48
(d) cannot calculate using only the information given

Practice

In a survey, 52% of respondents said they are Democrats. What is the probability that a randomly selected respondent from this sample is a Republican?

(a) 0.48
(b) more than 0.48
(c) less than 0.48
(d) cannot calculate using only the information given

If the only two political parties are Republican and Democrat, then (a) is possible. However it is also possible that some people do not affiliate with a political party or affiliate with a party other than these two. Then (c) is also possible. However (b) is definitely not possible since it would result in the total probability for the sample space being above 1.

Conditional Probability

Relapse

Researchers randomly assigned 72 chronic users of cocaine into three groups: desipramine (antidepressant), lithium (standard treatment for cocaine) and placebo. Results of the study are summarized below.

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

http://www.oswego.edu/~srp/stats/2_way_tbl_1.htm

Marginal probability

What is the probability that a patient relapsed?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

Marginal probability

What is the probability that a patient relapsed?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

P(relapsed) = 48 / 72 ~ 0.67

Joint probability

What is the probability that a patient received the antidepressant (desipramine) and relapsed?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

Joint probability

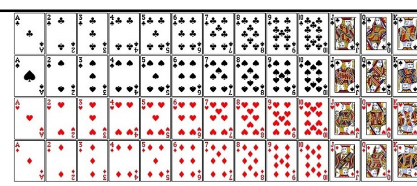
What is the probability that a patient received the antidepressant (desipramine) and relapsed?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

P(relapsed and desipramine) = 10 / 72 ~ 0.14

Probability Table

	A	A_complement $\bar{A}$	
B	$A \cap B$	$\bar{A} \cap B$	P(B)
B_completment $\bar{B}$	$A \cap \bar{B}$	$\bar{A} \cap \bar{B}$	$P(\bar{B})$
	P(A)	$P(\bar{A})$	1



Q: Complete the Probability Table

- Let event A = cards is an Ace {H1, D1, C1, S1}
- Let event B = cards is a red suit {H1~13, D1~13}

	A	A_complement $\bar{A}$	
B	$A \cap B$	$\bar{A} \cap B$	P(B)
B_completment $\bar{B}$	$A \cap \bar{B}$	$\bar{A} \cap \bar{B}$	$P(\bar{B})$
	P(A)	$P(\bar{A})$	1

	A (Ace)	A_complement (Not Ace) $\bar{A}$	
B (Red)	$A \cap B$ ={H1, D1} =2/52	$\bar{A} \cap B$ ={H2~13, D2~13} =24/52	$P(B)$ =26/52
B_completment (Not Red = Black) $\bar{B}$	$A \cap \bar{B}$ ={C1, S1} =2/52	$\bar{A} \cap \bar{B}$ ={C2~13, S2~13} =24/52	$P(\bar{B})$ =26/52
	$P(A)$ =4/52	$P(\bar{A})$ =48/52	1

Conditional probability

The conditional probability of the outcome of interest A given condition B is calculated as

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

The conditional probability of the outcome of interest B given condition A is calculated as

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

Conditional probability

The conditional probability of the outcome of interest A given condition B is calculated as

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

$$P(relapse|desipramine) = \frac{P(relapse\ and\ desipramine)}{P(desipramine)}$$

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

Conditional probability

The conditional probability of the outcome of interest A given condition B is calculated as

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

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

$$P(relapse|desipramine) = \frac{P(relapse\ and\ desipramine)}{P(desipramine)} = \frac{10/72}{24/72} = \frac{10}{24} = 0.42$$

Conditional probability (cont.)

If we know that a patient received the antidepressant (desipramine), what is the probability that they relapsed?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

Conditional probability (cont.)

If we know that a patient received the antidepressant (desipramine), what is the probability that they relapsed?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

P(relapse | desipramine) = 10 / 24 ~ 0.42

Conditional probability (cont.)

If we know that a patient received the antidepressant (desipramine), what is the probability that they relapsed?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

P(relapse | desipramine) = 10 / 24 ~ 0.42

P(relapse | lithium) = 18 / 24 ~ 0.75

P(relapse | placebo) = 20 / 24 ~ 0.83

Conditional probability (cont.)

If we know that a patient relapsed, what is the probability that they received the antidepressant (desipramine)?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

Conditional probability (cont.)

If we know that a patient relapsed, what is the probability that they received the antidepressant (desipramine)?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

P(desipramine | relapse) = 10 / 48 ~ 0.21

Conditional probability (cont.)

If we know that a patient relapsed, what is the probability that they received the antidepressant (desipramine)?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

P(desipramine | relapse) = 10 / 48 ~ 0.21

P(lithium | relapse) = 18 / 48 ~ 0.38

P(placebo | relapse) = 20 / 48 ~ 0.42

Practice

- Of the PCs in the computer lab,
- 40% is touch screen (A)
- 70% has camera (B)
- 20% has both.

Q1: What's the probability that a PC is touch screen given that it has camera?

Q2. What's the probability that the PC has camera given that it is touch screen?

Q3. What's the probability that a PC does not have either?

Q1: What's the probability that a PC is touch screen (A) given that it has camera (B)?
$P(A|B)= P(A \cap B)/P(B)=0.2/0.7 = 0.286$

Q2. What's the probability that the PC has camera (B)given that it is touch screen(A)?
$P(B|A)= P(B \cap A)/P(A)=0.2/0.4 = 0.5$

Q3. What's the probability that a PC does not have either?
$1-[P(A)+P(B)-P(A \cap B)]=1-[0.4+0.7-0.2]=0.1$

Multiplication Rule

- P(A∩B)=P(A|B)P(B) =P(B|A)P(A)

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

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

P(A∩B): joint probability
P(A|B): conditional probability
P(A): marginal probability

Practice: P(A∩B)=P(A|B)P(B) =P(B|A)P(A)

- Confirm the rule by using previous example
- 40% touch screen (A)
- 70% camera (B)
- 20% both

P(A∩B)=P(A|B)P(B)=0.286*0.7=0.2

P(A∩B)=P(B|A)P(A)=0.5*0.4=0.2

Independence

Two processes are *independent* if knowing the outcome of one provides no useful information about the outcome of the other.

$$\boxed{P(A \cap B) = P(A)P(B)}$$

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

Independence

Two processes are *independent* if knowing the outcome of one provides no useful information about the outcome of the other.

- Knowing that the coin landed on a head on the first toss does not provide any useful information for determining what the coin will land on in the second toss.
  \>\> Outcomes of two tosses of a coin are independent.

Independence

Two processes are *independent* if knowing the outcome of one provides no useful information about the outcome of the other.

- Knowing that the coin landed on a head on the first toss does not provide any useful information for determining what the coin will land on in the second toss.
  \>\> Outcomes of two tosses of a coin are independent.
- Knowing that the first card drawn from a deck is an ace does provide useful information for determining the probability of drawing an ace in the second draw.
  \>\> Outcomes of two draws from a deck of cards (without replacement) are dependent.

Practice

- Of the PCs in the computer lab,
- 40% is touch screen. (A)
- 70% has camera. (B)
- 20% has both.

Are event A and B statistically independent?

P(A ∩ B) = P(A)P(B) ???
P(A) = 0.4, P(B) = 0.7, P(A)P(B)=0.28
$\neq$ 0.2=P(A∩B) => Not independent.

Practice

Between January 9-12, 2013, SurveyUSA interviewed a random sample of 500 NC residents asking them whether they think widespread gun ownership protects law abiding citizens from crime, or makes society more dangerous. 58% of all respondents said it protects citizens. 67% of White respondents, 28% of Black respondents, and 64% of Hispanic respondents shared this view (Consider these are conditional probabilities (P(protect|Race)). Which of the below is true?

Opinion on gun ownership and race ethnicity are most likely
(a) complementary
(b) mutually exclusive
(c) independent
(d) dependent

http://www.surveyusa.com/client/PollReport.aspx?g=a5f460ef-bba9-484b-8579-1101ea26421b

Hint

- P(protect) = 0.58
- P(protect|White) = 0.67
- P(protect|Black) = 0.28
- P(protect|Hispanic)=0.64

P(protect|White) = P(protect) ???
P(protect|Black) = P(protect)???
P(protect|Hispanic)=P(protect)???

Practice

Between January 9-12, 2013, SurveyUSA interviewed a random sample of 500 NC residents asking them whether they think widespread gun ownership protects law abiding citizens from crime, or makes society more dangerous. 58% of all respondents said it protects citizens. 67% of White respondents, 28% of Black respondents, and 64% of Hispanic respondents shared this view. Which of the below is true?

Opinion on gun ownership and race ethnicity are most likely
(a) complementary
(b) mutually exclusive
(c) independent
(d) dependent

http://www.surveyusa.com/client/PollReport.aspx?g=a5f460ef-bba9-484b-8579-1101ea26421b

Checking for independence

If P(A occurs, given that B is true) = *P(A | B) = P(A)*, then A and B are independent.

Checking for independence

If P(A occurs, given that B is true) = *P(A | B) = P(A)*, then A and B are independent.

P(protects citizens) = 0.58