W3203
Discrete Mathematics

Probability

Spring 2015
Instructor: Ilia Vovsha

http://www.cs.columbia.edu/~vovsha/w3203

Outline

- Probability spaces, events
- Probability Measures
- Distributions
- Conditional Probability
- Random variables
- Bayes' Theorem
- Expected value, variance
- Text: Rosen 7
- Text: Lehman 16-18

Monty Hall Problem

- The Problem:
  - Given 3 doors, behind one door is a car, behind the others, goats.
  - The car is equally likely to be hidden behind each of the three doors.
  - The player is equally likely to pick each of the three doors, regardless of the car's location.
  - After the player picks a door, the host must open a different door with a goat behind it and offer the player the choice of staying with the original door or switching.
  - If the host has a choice of which door to open, then he is equally likely to select each of them.
  - Is it to your advantage to switch your choice of doors?
- The Challenge:
  - How do we model the situation mathematically?
  - How do we solve the resulting mathematical problem?

The Four Step Method

- Goal: model the situation mathematically
  1. Find the sample space: identify quantities and outcomes that determine the experiment
  2. Define events of interest: translate questions to precise set of outcomes
  3. Determine outcome probabilities: assess likelihood and assign probability to each outcome
  4. Compute event probabilities: use outcome probabilities to determine probability of event

Probability Jargon

- Definition: a *sample space S* is a nonempty set
- Definition: an element $a \in S$ of a sample space is called an *outcome*
- Definition: a subset of a sample space S is called an *event*
- Definition: the power set of a sample space S is called the *event space*
- Definition: an *experiment* is a process that produces an outcome (point in sample space)
- Typical basic examples: coins, dice, cards

Coins

- Experiment 1: toss a coin, get heads (H) or tails (T)
  - Sample Space: S = {H, T}
  - Event Space: P(S) = { Ø, {H}, {T}, {H,T} }
- Experiment 2: toss two coins
  - Sample Space: S = {HH, HT, TH, TT}
  - Event Space: P(S) = { Ø, {HH}, {HT}, ... , {HH, TT}, ... }
  - Named events:
    - "match": {HH, TT}
    - "At least one head": {HT, TH, HH}

Dice

- Experiment 1: roll a die
  - Sample Space: S = {1, 2, 3, 4, 5, 6}
  - Named events:
    - "odd": {1, 3 , 5}   "even": {2 , 4, 6}
    - "2 mod 3": {2 , 5}
- Experiment 2: roll two dice
  - Sample Space: S = {11, ... , 16, ... , 61, ... , 66}
  - Named events:
    - "doubles": {11, 22, ... , 66}
    - "sum is nine": {36, 45, 54, 63}

Standard Deck of Cards

- Experiment 1: deal one card from deck
  - Sample Space: S = 52-card deck (4 suits)
  - Named events:
    - “heart”: {2H, ... , AH}
    - “seven”: {7C , 7D, 7H, 7S}
- Experiment 2: deal five cards from deck
  - Sample Space: S = all possible 5-card hands
  - Named events:
    - “4-of-a-kind”, “full-house”, “straight flush”

Probability Functions

- A *probability function (measure)* on a sample space is a total function $(\mathcal{Pr} : S \rightarrow \boldsymbol{R})$ such that,
  - $\forall \mathrm{a} \in \mathrm{S},\ \mathcal{Pr}(\mathrm{a}) \geq 0$
  - $\Sigma_{\mathrm{a} \in \mathrm{S}} \mathcal{Pr}(\mathrm{a}) = 1$
- Definition: a sample space S with a probability function $(\mathcal{Pr} : S \rightarrow \boldsymbol{R})$ is called a *probability space*
- The probability of an event E is defined as the sum of the probabilities of the outcomes in E:

$$\mathcal{Pr}(\mathrm{E}) = \Sigma_{\mathrm{a} \in \mathrm{E}} \mathcal{Pr}(\mathrm{a})$$

Probability Spaces

- A *discrete probability function (measure)* is defined on a *finite or countably infinite* sample space (set)
- How to assign probabilities to infinite spaces? Events rather than outcomes
- What about uncountable sets (real numbers)? Use integrals instead of sums
  - $\Sigma_{a \in S} \mathcal{Pr}(a) = 1 \qquad \mathcal{Pr}(E) = \Sigma_{a \in E} \mathcal{Pr}(a)$

Coin Probabilities

- Experiment 1: toss a coin
  - Sample Space: S = {H, T}
  - $\mathcal{Pr}(\emptyset) = 0$, $\mathcal{Pr}$(H) = $\mathcal{Pr}(\mathrm{T})$ = ½, $\mathcal{Pr}$(H,T) = 1
- Experiment 2: toss two coins
  - Sample Space: S = {HH, HT, TH, TT}
  - Named events:
    - E1: “match” {HH, TT}
    - E2: “At least one head” {HT, TH, HH}
  - $\mathcal{Pr}(\mathrm{E}_1) = ½$, $\mathcal{Pr}(\mathrm{E}_2)$ = ¾

Dice Probabilities

- Experiment 1: roll a die
  - Sample Space: S = {1, 2, 3, 4, 5, 6}
  - Named events:
    - E1: “odd” {1, 3 , 5}   E2: “even” {2 , 4, 6}
  - $\mathcal{Pr}(\mathrm{E}_1) = 3/6 = ½$ $\mathcal{Pr}(\mathrm{E}_2) = 3/6 = ½$
- Experiment 2: roll two dice
  - Sample Space: S = {11, ... , 16, ... , 61, ... , 66}
  - Named events:
    - E1: “doubles”   {11, 22, ... , 66}
    - E2: “sum is nine”   {36, 45, 54, 63}
  - $\mathcal{Pr}(\mathrm{E}_1) = 6/36 = 1/6$ $\mathcal{Pr}(\mathrm{E}_2) = 4/36 = 1/9$

Monty Hall (sample space)

A. Find the sample space:

1. The door concealing the car {A,B,C}
2. The door initially chosen by the player {A,B,C}
3. The door that the host opens to reveal a goat {A,B,C}

$$S = \left\{ \begin{array}{l} (A,A,B),\ (A,A,C),\ (A,B,C),\ (A,C,B),\ (B,A,C),\ (B,B,A), \\ (B,B,C),\ (B,C,A),\ (C,A,B),\ (C,B,A),\ (C,C,A),\ (C,C,B) \end{array} \right\}$$

Monty Hall (events)

B. Define events of interest:

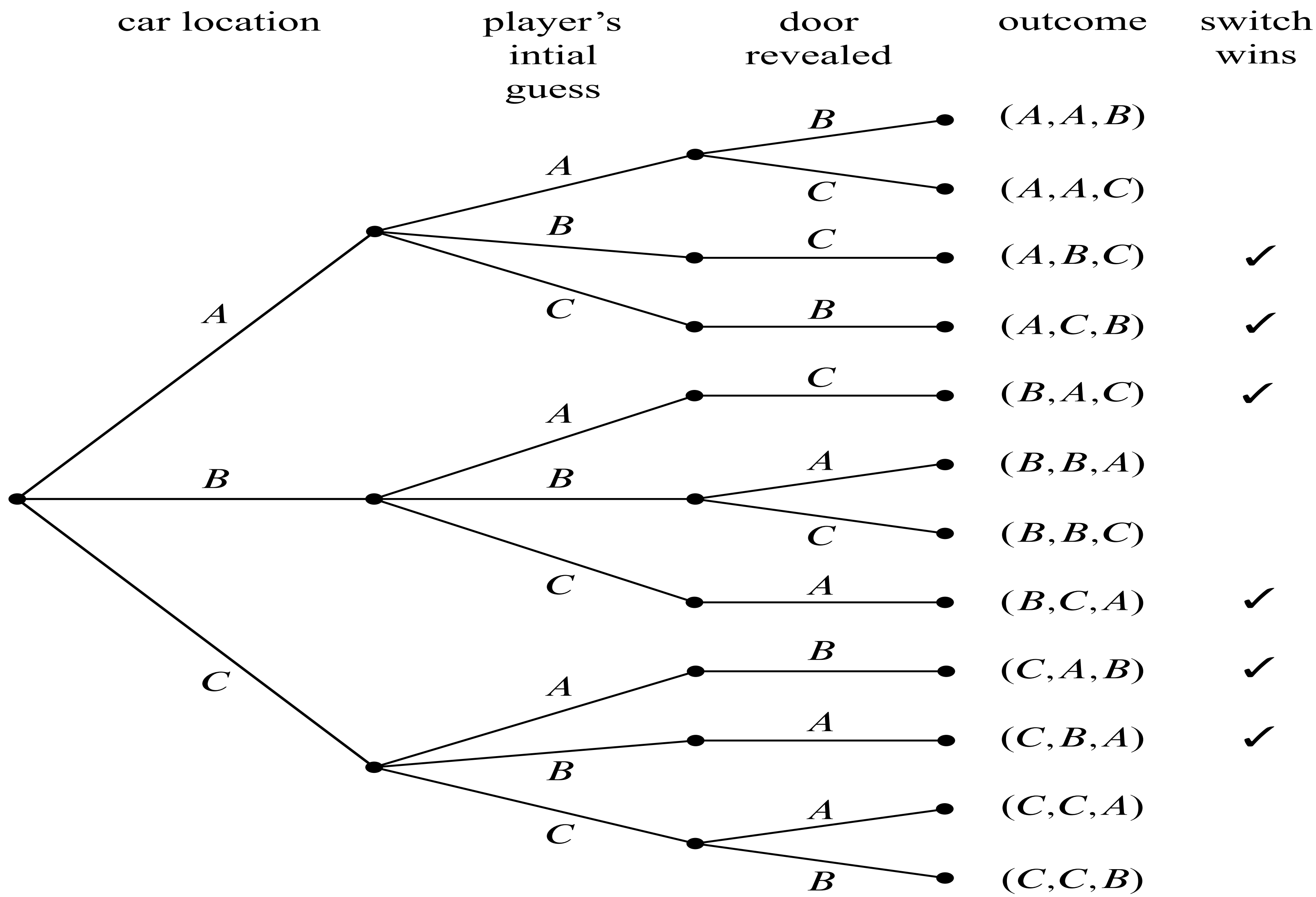
- "Prize behind door C": { (C,A,B), (C,B,A), (C,C,A), (C,C,B) }
- "Prize behind door first picked by player":

  { (A,A,?), (B,B,?), (C,C,?) }
- "Player wins by switching":

  { (A,B,C), (A,C,B), (B,A,C), (B,C,A), (C,A,B), (C,B,A) }

$$S = \left\{ \begin{array}{l} (A,A,B), (A,A,C), (A,B,C), (A,C,B), (B,A,C), (B,B,A), \\ (B,B,C), (B,C,A), (C,A,B), (C,B,A), (C,C,A), (C,C,B) \end{array} \right\}$$

Monty Hall (tree diagram)



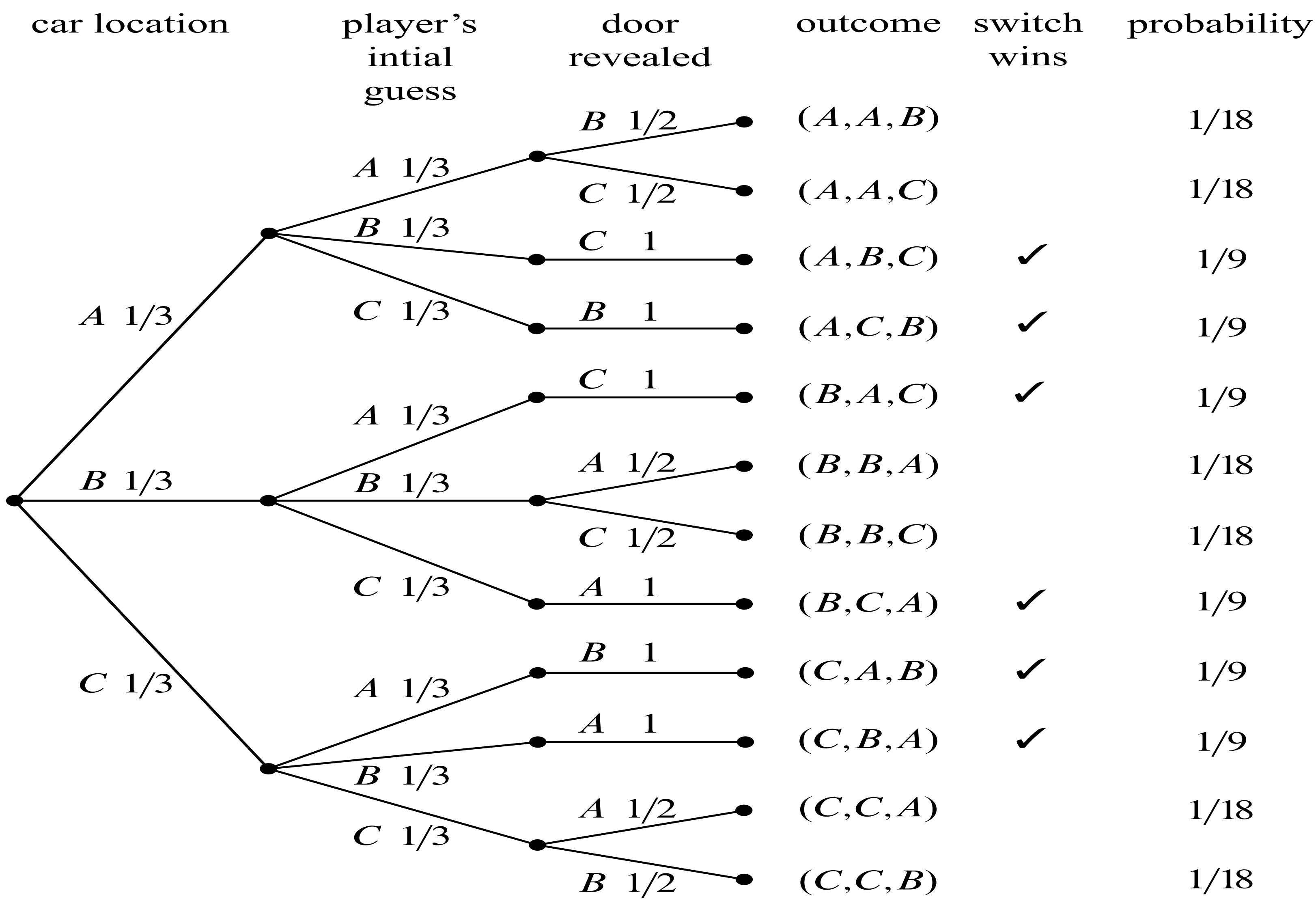
Monty Hall (outcome probabilities)

C. Determine outcome probabilities:

   1. Record probability on each edge of tree, make use of problem assumptions
   2. Multiply edge probabilities on the path from the root to the outcome (justification?)
   3. Probability of each outcome is the product from step (2)

   $\mathcal{P}r[(\mathrm{A,A,B})] = 1/18$ $\qquad$ $\mathcal{P}r[(\mathrm{A,B,C})] = 1/9$

Monty Hall (outcome prob.)



Monty Hall (event probabilities)

D. Compute event probabilities:

- Use the rule: $\mathcal{Pr}(\mathrm{E}) = \Sigma_{\mathrm{a} \in \mathrm{E}} \mathcal{Pr}(\mathrm{a})$
- Event of interest: "player wins by switching"
- Relevant outcomes:

  a ∈ { (A,B,C), (A,C,B), (B,A,C), (B,C,A), (C,A,B), (C,B,A) }
- $\forall \mathrm{a} \in \mathrm{E}, \mathcal{Pr}(\mathrm{a}) = 1/9$
- $\mathcal{Pr}(\mathrm{E}) = 6 * (1/9) = 6/9 = 2/3$

Probability Rules

- The probability of an event E is defined as the sum of the probabilities of the outcomes in E:

$$\mathcal{Pr}(\mathrm{E}) = \Sigma_{\mathrm{a} \in \mathrm{E}} \mathcal{Pr}(\mathrm{a})$$

- *Sum Rule*: if $\{E_1, E_2, \ldots, E_n\}$ is a collection of disjoint events, then

$$\mathcal{Pr}[\cup \mathrm{E_i}] = \Sigma_{\mathrm{i}} \mathcal{Pr}(\mathrm{E_i})$$

- *Complement Rule*: if $\bar{A}$ denotes the complement of the event A, then

$$\mathcal{Pr}[\bar{A}] + \mathcal{Pr}[A] = 1$$
$$\mathcal{Pr}[\bar{A}] = 1 - \mathcal{Pr}[A]$$

Probability Rules (from set theory)

- Basic facts about probability parallel facts about cardinalities of finite sets:

1. $\mathcal{P}r[\, A \cup B \,] \;=\; \mathcal{P}r[\, A \,] + \mathcal{P}r[\, B \,] - \mathcal{P}r[\, A \cap B \,]$

2. $\mathcal{P}r[\, A \cup B \,] \;\leq\; \mathcal{P}r[\, A \,] + \mathcal{P}r[\, B \,]$

3. $A \subseteq B \rightarrow \mathcal{P}r[\, A \,] \;\leq\; \mathcal{P}r[\, B \,]$

4. $\mathcal{P}r[\, \cup\, E_i \,] \;\leq\; \Sigma_i\, \mathcal{P}r[\, E_i \,]$

Probability Rules (examples)

- Example 1: toss a coin 10 times. Find probability of at least one tail
  - Event: “at least one tail”
  - Complement of event: “10 heads”
  - $\mathcal{Pr}[\,A\,] = 1 - \mathcal{Pr}[\,\bar{A}\,] = 1 - (1/1024) = 1023/1024$
- Example 2: choose integer from the interval [1, ... , 100]. Find probability it is divisible either by 6 or by 15
  - Event 1: “divisible by 6” Event 2: “divisible by 15”
  - $\mathcal{Pr}[\,E_1\,] = 16/100$ $\mathcal{Pr}[\,E_2\,] = 6/100$
  - $\mathcal{Pr}[\,E_1 \cap E_2] = 3/100$
  - $\mathcal{Pr}[\,E_1 \cup E_2] = \mathcal{Pr}[\,E_1\,] + \mathcal{Pr}[\,E_2\,] - \mathcal{Pr}[\,E_1 \cap E_2] =$
    $= 16/100 + 6/100 - 3/100 = 19/100$

Uniform Probability Space

- Definition: a sample space *S* with a probability function ($\mathcal{Pr}$: $S \rightarrow R$) is called a *probability space*
- A finite probability space *S* is *uniform* if $\forall a \in S$, $\mathcal{Pr}(a)$ is the same
- For events:

$$\mathcal{Pr}(E) = |E| / |S|$$

Conditional Probability

- Definition: the *conditional probability* of an event X given that event Y has occurred is $\Pr[X \mid Y] = \dfrac{\Pr[X \cap Y]}{\Pr[Y]}$

- Example 1: toss two coins

$$\Pr[HH \mid \neg(TT)] = \frac{\Pr[HH \cap \neg(TT)]}{\Pr[\neg(TT)]} = \frac{1/4}{3/4}$$

- Example 2: roll two fair dice
  - At least one is 4, what is the prob. that both are fours?

$$\Pr[44 \mid "4?-or-?4"] = \frac{\Pr[44 \cap "..."]}{\Pr["..."]} = \frac{1/36}{11/36} = \frac{1}{11}$$

Conditional Probability (product rule)

- What about multiple events? Generalize idea:

$$\Pr[X \mid Y] = \frac{\Pr[X \cap Y]}{\Pr[Y]} \quad \rightarrow \quad \Pr[X \cap Y] = \Pr[X \mid Y]\Pr[Y]$$

$$\Pr[E_1 \cap E_2 \cap E_3] = \Pr[E_1 \mid E_2 E_3]\Pr[E_2 \mid E_3]\Pr[E_3]$$

$$\mathrm{P}(E_1 E_2 E_3) = \mathrm{P}(E_1 \mid E_2 E_3)\mathrm{P}(E_2 \mid E_3)\mathrm{P}(E_3)$$

Independence

- Definition: event A is (probabilistically) *independent* of event B f knowing that B happens does not alter the probability that A happens

$$\Pr[A \mid B] = \Pr[A]$$

$$\Pr[A \cap B] = \Pr[A \mid B]\Pr[B] = \Pr[A]\Pr[B]$$

- Note: disjoint events are never independent!
- Mutual independence:

$$\mathrm{P}(E_1 E_2 E_3) = \mathrm{P}(E_1 \mid E_2 E_3)\mathrm{P}(E_2 \mid E_3)\mathrm{P}(E_3)$$
$$= \mathrm{P}(E_1)\mathrm{P}(E_2)\mathrm{P}(E_3)$$

Independence (example)

- Example: roll two fair dice. Suppose A is the event "first die is 1", and B is the event "sum of rolls is odd". Are the events independent?

$$P(A) = \frac{1}{6} \qquad P(B) = 2 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

$$P(AB) = P("12","14","16") = \frac{3}{36}$$

$$P(B \mid A) = \frac{P(AB)}{P(A)} = \frac{3/36}{1/6} = \frac{3}{6} = \frac{1}{2}$$

Email Spam (example)

- How can we detect spam email messages?
- Suppose we try to identify characteristic keywords.
- Example: 12,600 of 100,000 spam messages contained the word "Rolex". However, in another 20,000 non-spam messages the word "Rolex" appeared 100 times.
- Estimate the probability that a new message with the word "Rolex" is spam.
- Define A as the event "message contains word"
- Define B as the event "new message is spam"
- Goal: estimate $P(B \mid A)$

Cond. Prob. (law of total prob.)

- Idea: breaking probability calculation into cases
- Calculate the probability of an event A by splitting into two cases based on whether or not another event E occurs, and then using sum rule:

$$\Pr[A] = \Pr[A \mid E]\Pr[E] + \Pr[A \mid \bar{E}]\Pr[\bar{E}]$$

Email Spam (notation)

- Define A as the event "message contains word"
- Define B as the event "new message is spam"
- Goal: estimate P(B | A)

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(AB)}{\mathrm{P}(A)}$$

$$\mathrm{P}(B) \approx \frac{2}{3}$$

$$\mathrm{P}(A \mid B) \approx \frac{12600}{100000} = 0.126$$

$$\mathrm{P}(A \mid \overline{B}) \approx \frac{120}{20000} = 0.006$$

Email Spam (solution)

- **A**: "message contains word" **B**: "new message is spam"

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(AB)}{\mathrm{P}(A)}$$

$$\mathrm{P}(B) \approx \frac{2}{3} \quad \mathrm{P}(A \mid B) \approx 0.126 \quad \mathrm{P}(A \mid \overline{B}) \approx 0.006$$

$$\mathrm{P}(AB) = \mathrm{P}(BA) = \mathrm{P}(A \mid B)\mathrm{P}(B) \approx 0.126 \cdot \frac{2}{3} = 0.084$$

$$\mathrm{P}(A) = \mathrm{P}(AB) + P(A\overline{B}) = \mathrm{P}(A \mid B)\mathrm{P}(B) + \mathrm{P}(A \mid \overline{B})P(\overline{B})$$

$$\approx 0.084 + 0.006 \cdot \frac{1}{3} = 0.086$$

$$\mathrm{P}(B \mid A) \approx \frac{0.084}{0.086} = 0.9767$$

Bayes' Theorem

- *Bayes' Theorem* (or rule) relates a pair of probabilities

$$\Pr[B \mid A] = \frac{\Pr[A \mid B]\Pr[B]}{\Pr[A]}$$

- We can use the law of total probability to rewrite the denominator:

$$\Pr[B \mid A] = \frac{\Pr[A \mid B]\Pr[B]}{\Pr[A \mid B]\Pr[B] + \Pr[A \mid \overline{B}]\Pr[\overline{B}]}$$

Random Variables

- A *random variable* is a total function whose domain is the sample space
- It is not a variable and it is not random!
- Example: toss a coin 3 times, let C(outcome) be the random variable denoting number of heads that appear

C(HHH) = 3 C(HHT) = 2 C(HTH) = 2

C(HTT) = 1 C(THH) = 2 C(THT) = 1

C(TTH) = 1 C(TTT) = 0

Probability Density Function

- The *probability density function (pdf)* of a random variable R measures the probability that R = x.
- The *cumulative distribution function (cdf)* of a random variable R measures the probability that R ≤ x.

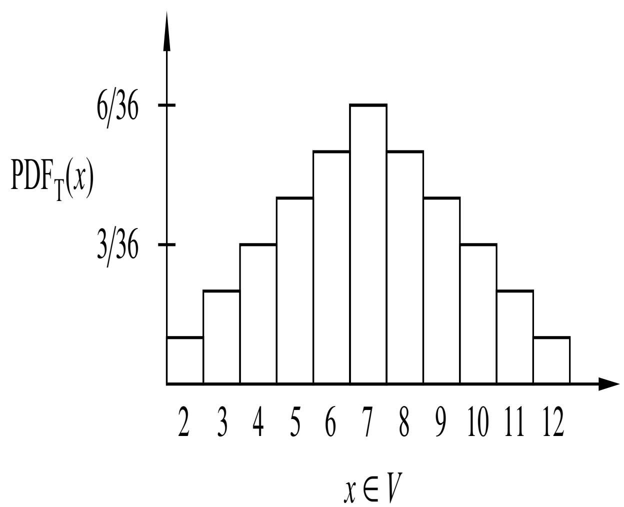


Figure 18.1 The probability density function for the sum of two 6-sided dice.

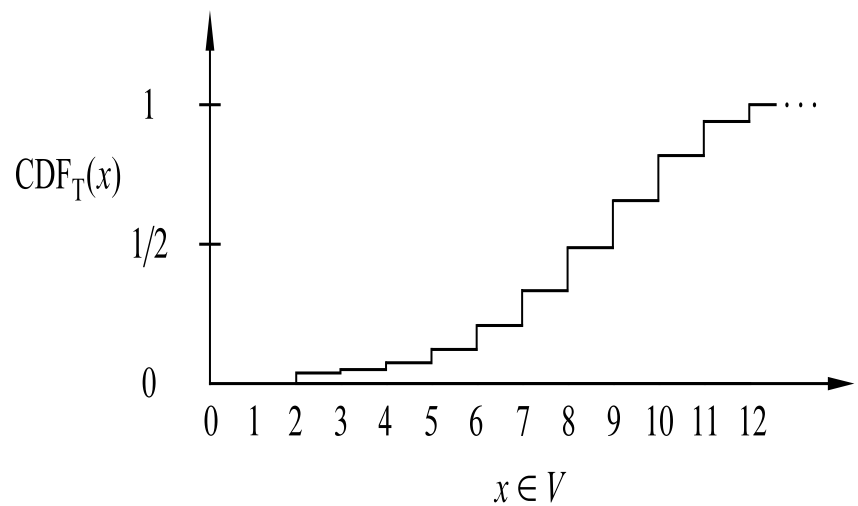


Figure 18.2 The cumulative distribution function for the sum of two 6-sided dice.

Bernoulli Distribution

- A *Bernoulli distribution* is the distribution function for a (two-point) Bernoulli variable
- Probability density function:

  $f_p: \{0,1\} \rightarrow [0,1]$

  $f_p(0) = p$

  $f_p(1) = 1-p$
- Example: toss a biased (loaded) coin
  - Sample Space: S = {H, T}
  - $\mathcal{Pr}$({H}) = p = 4/5 $\mathcal{Pr}$({T}) = 1 - p = 1/5
- Example: roll a fair die
  - Sample Space: S = {1, 2, 3, 4, 5, 6}
  - $\mathcal{Pr}$("even") = p = ½ $\mathcal{Pr}$ ("odd") = 1 – p = ½

Binomial Distribution

- A *Binomial distribution* extends the Bernoulli variable to n independent repetitions of the (two-point) experiment
- Probability density function:

  $f_{n,p}: \{0,1,2,\ldots,n\} \rightarrow [0,1]$

  $f_{n,p}(k) = C(n,k)p^k(1-p)^{n-k}$

- Example: toss a biased (loaded) coin 10 times
  - $\mathcal{Pr}(\{H\}) = p = 4/5$ $\mathcal{Pr}(\{T\}) = 1 - p = 1/5$
  - $\mathcal{Pr}$(“k heads”) = $C(10,k)(0.8)^k (0.2)^{10-k}$