ECE 448 Lecture 12: Probability

Slides by Svetlana Lazebnik, 9/2016

Modified by Mark Hasegawa-Johnson, 10/2017

Probability: Review of main concepts (Chapter 13)

Outline

- Motivation: Why use probability?
  - Laziness, Ignorance, and Randomness
  - Rational Bettor Theorem
- Review of Key Concepts
  - Outcomes, Events
  - Random Variables; probability mass function (pmf)
  - Jointly random variables: Joint, Marginal, and Conditional pmf
  - Independent vs. Conditionally Independent events

Outline

- Motivation: Why use probability?
  - Laziness, Ignorance, and Randomness
  - Rational Bettor Theorem
- Review of Key Concepts
  - Outcomes, Events, and Random Variables
  - Joint, Marginal, and Conditional
  - Independence and Conditional Independence

Motivation: Planning under uncertainty

- Recall: representation for planning
- **States** are specified as conjunctions of predicates
  - Start state: At(P1, CMI) $\wedge$ Plane(P1) $\wedge$ Airport(CMI) $\wedge$ Airport(ORD)
  - Goal state: At(P1, ORD)
- **Actions** are described in terms of preconditions and effects:
  - Fly(p, source, dest)
    - **Precond:** At(p, source) $\wedge$ Plane(p) $\wedge$ Airport(source) $\wedge$ Airport(dest)
    - **Effect:** ¬At(p, source) $\wedge$ At(p, dest)

Motivation: Planning under uncertainty

- Let action A_t = leave for airport t minutes before flight
  - Will A_t succeed, i.e., get me to the airport in time for the flight?
- Problems:
  - Partial observability (road state, other drivers' plans, etc.)
  - Noisy sensors (traffic reports)
  - Uncertainty in action outcomes (flat tire, etc.)
  - Complexity of modeling and predicting traffic
- Hence a purely logical approach either
  - Risks falsehood: "A_{25} will get me there on time," or
  - Leads to conclusions that are too weak for decision making:
    - A_{25} will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact, etc., etc.
    - A_{1440} will get me there on time but I'll have to stay overnight in the airport

Probability

Probabilistic assertions summarize effects of

- Laziness: reluctance to enumerate exceptions, qualifications, etc.
- Ignorance: lack of explicit theories, relevant facts, initial conditions, etc.
- Intrinsically random phenomena

When does it make sense to use probability?

- When should an outcome be considered to be random?
  - ... List some examples or reasons....
- When should an outcome _not_ be considered to be random?
  - … list some examples or reasons…

Outline

- Motivation: Why use probability?
  - Laziness, Ignorance, and Randomness
  - Rational Bettor Theorem
- Review of Key Concepts
  - Outcomes, Events, and Random Variables
  - Joint, Marginal, and Conditional
  - Independence and Conditional Independence

Making decisions under uncertainty

- Suppose the agent believes the following:

  P(A_{25} gets me there on time) = 0.04

  P(A_{90} gets me there on time) = 0.70

  P(A_{120} gets me there on time) = 0.95

  P(A_{1440} gets me there on time) = 0.9999
- Which action should the agent choose?
  - Depends on preferences for missing flight vs. time spent waiting
  - Encapsulated by a *utility function*
- The agent should choose the action that maximizes the *expected utility*:

  P(A_t succeeds) * U(A_t succeeds) + P(A_t fails) * U(A_t fails)

Making decisions under uncertainty

- More generally: the expected utility of an action is defined as:

$$EU(\text{action}) = \Sigma_{\text{outcomes of action}}\, P(\text{outcome} \mid \text{action})\, U(\text{outcome})$$

- **Utility theory** is used to represent and infer preferences
- **Decision theory** = probability theory + utility theory

Monty Hall problem

- You’re a contestant on a game show. You see three closed doors, and behind one of them is a prize. You choose one door, and the host opens one of the other doors and reveals that there is no prize behind it. Then he offers you a chance to switch to the remaining door. Should you take it?

http://en.wikipedia.org/wiki/Monty_Hall_problem

Monty Hall problem

- With probability 1/3, you picked the correct door, and with probability 2/3, picked the wrong door. If you picked the correct door and then you switch, you lose. If you picked the wrong door and then you switch, you win the prize.

- Expected utility of switching:

$$\mathbf{EU(Switch) = (1/3) * 0 + (2/3) * Prize}$$

- Expected utility of not switching:

$$\mathbf{EU(Not\ switch) = (1/3) * Prize + (2/3) * 0}$$

Where do probabilities come from?

- **Frequentism**
  - Probabilities are relative frequencies
  - For example, if we toss a coin many times, P(heads) is the proportion of the time the coin will come up heads
  - But what if we're dealing with events that only happen once?
    - E.g., what is the probability that Team X will win the Superbowl this year?
    - "Reference class" problem
- **Subjectivism**
  - Probabilities are degrees of belief
  - But then, how do we assign belief values to statements?
  - What would constrain agents to hold consistent beliefs?

The Rational Bettor Theorem

- Why should a rational agent hold beliefs that are consistent with axioms of probability?
  - For example, $P(A) + P(\neg A) = 1$

- If an agent has some degree of belief in proposition A, he/she should be able to decide whether or not to accept a bet for/against A (De Finetti, 1931):
  - If the agent believes that $P(A) = 0.4$, should he/she agree to bet \$4 that A will occur against \$6 that A will not occur?

- **Theorem:** An agent who holds beliefs inconsistent with axioms of probability can be convinced to accept a combination of bets that is guaranteed to lose them money

Are humans "rational bettors"?

- Humans are pretty good at estimating some probabilities, and pretty bad at estimating others. What might cause humans to mis-estimate the probability of an event?
  - … list some examples …
- What are some of the ways in which a "rational bettor" might take advantage of humans who mis-estimate probabilities?
  - … list some examples …

Outline

- Motivation: Why use probability?
    - Laziness, Ignorance, and Randomness
    - Rational Bettor Theorem
- Review of Key Concepts
    - Outcomes, Events, and Random Variables
    - Joint, Marginal, and Conditional
    - Independence and Conditional Independence

Outcomes of an Experiment

The SET OF POSSIBLE OUTCOMES (a.k.a. the "sample space") is a listing of all of the things that might happen:

1. Mutually exclusive. It's not possible that two different outcomes might both happen.
2. Collectively exhaustive. Every outcome that could possibly happen is one of the items in the list.
3. Finest grain. After the experiment occurs, somebody tells you the outcome, and there is nothing else you need to know.

Example experiment: Alice, Bob, Carol and Duane run a 10km race to decide who will buy pizza tonight.

Outcome = a listing of the exact finishing times of each participant.

Events

- Probabilistic statements are defined over *events*, or sets of world states
  - *A = "It is raining"*
  - *B = "The weather is either cloudy or snowy"*
  - *C = "The sum of the two dice rolls is 11"*
  - *D = "My car is going between 30 and 50 miles per hour"*
- An EVENT is a SET of OUTCOMES
  - B = { outcomes : cloudy OR snowy }
  - C = { outcomes : d1+d2 = 11 }
- Notation: p(A) or P(A) is the probability of the set of world states (outcomes) in which proposition A holds

Kolmogorov's axioms of probability

- For any propositions (events) A, B
  - $0 \leq P(A) \leq 1$
  - $P(True) = 1$ and $P(False) = 0$
  - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
    - Subtraction accounts for double-counting
- Based on these axioms, what is $P(\neg A)$?
- These axioms are sufficient to completely specify probability theory for *discrete* random variables
  - For continuous variables, need *density functions*

Outcomes = Atomic events

- ***OUTCOME or ATOMIC EVENT:*** is a complete specification of the state of the world, or a complete assignment of domain values to all random variables
  - Atomic events are mutually exclusive and exhaustive

- E.g., if the world consists of only two Boolean variables *Cavity* and *Toothache*, then there are four outcomes:

$\neg Cavity \wedge \neg Toothache$
$\neg Cavity \wedge Toothache$
$Cavity \wedge \neg Toothache$
$Cavity \wedge Toothache$

Random variables

- We describe the (uncertain) state of the world using ***random variables***
  - Denoted by capital letters
  - **R**: *Is it raining?*
  - **W**: *What's the weather?*
  - **D**: *What is the outcome of rolling two dice?*
  - **S**: *What is the speed of my car (in MPH)?*
- Just like variables in CSPs, random variables take on values in a *domain*
  - Domain values must be *mutually exclusive* and *exhaustive*
  - **R** in {True, False}
  - **W** in {Sunny, Cloudy, Rainy, Snow}
  - **D** in {(1,1), (1,2), … (6,6)}
  - **S** in [0, 200]

Random variables

- A random variable can be viewed as a function that maps from outcomes to real numbers (or integers, or strings)
- For example: the event “Speed=45mph” is the set of all outcomes for which the speed of my car is 45mph

Probability Mass Function (pmf)

- We use a capital letter for a random variables (RV=the function that maps from outcomes to values), and a small letters for the actual value that it takes after any particular experiment.
- $X_1 = x_1$ is the event "random variable X_1 takes the value x_1"
- $p(X_1 = x_1)$ is a **number**: the probability that this event occurs.
  - We call this number the "probability mass" of the event $X_1 = x_1$
  - The function is called the "probability mass function" or pmf
  - Shorthand: $p(x_1)$ using a small letter x_1
  - Subscript notation, which we won't use in this class: $\boldsymbol{p_{X_1}(x_1)}$
- $p(X_1)$ using a capital letter X_1 is a **function**: the entire table of the probabilities $X_1 = x_1$ for every possible x_1

Events and Outcomes

- An OUTCOME (ATOMIC EVENT) is a particular setting of all of the random variables
  - *Outcome = (die 1 shows 5 dots, die 2 shows 6 dots)*
- An EVENT is a SET of OUTCOMES
  - *"The sum of the two dice rolls is 11"* = { set of all outcomes such that D1+D2 = 11 }
  - "D1=5" = {set of all outcomes such that D1=5, regardless of what D2 is }
- $\mathrm{P}(EVENT) = \sum_{outcomes \,\epsilon\, EVENT} P(outcome)$

Functions of Random Variables

- Suppose we are not really interested in any given random variable, instead we're only interested in a function of the random variables
- Example: the game of craps. We're only interested in the sum of the two dice, e.g., what is the probability that the sum of the two dice is greater than 10.
- Define S=D1+D2. How can we calculate the pmf for S?

Outline

- Motivation: Why use probability?
  - Laziness, Ignorance, and Randomness
  - Rational Bettor Theorem
- Review of Key Concepts
  - Outcomes, Events, and Random Variables
  - Joint, Marginal, and Conditional
  - Independence and Conditional Independence

Joint probability distributions

- A ***joint distribution*** is an assignment of probabilities to every possible atomic event

Atomic event	P
$\neg Cavity \wedge \neg Toothache$	0.8
$\neg Cavity \wedge Toothache$	0.1
$Cavity \wedge \neg Toothache$	0.05
$Cavity \wedge Toothache$	0.05

  - Why does it follow from the axioms of probability that the probabilities of all possible atomic events must sum to 1?

Joint probability distributions

- A ***joint distribution*** is an assignment of probabilities to every possible atomic event
- Suppose we have a joint distribution of *N* random variables, each of which takes values from a domain of size *D*
  - What is the size of the probability table?
  - Impossible to write out completely for all but the smallest distributions

Notation

- $p(X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N)$ refers to a single entry (atomic event) in the joint probability distribution table
  - Shorthand: $p(x_1, x_2, \ldots, x_N)$
  - Subscript notation, which we won't use in this class: $\boldsymbol{p_{X_1,X_2,\ldots,X_N}(x_1, x_2,\ldots, x_N)}$
- $p(X_1, X_2, \ldots, X_N)$ refers to the entire joint probability distribution table
- $P(A)$ can also refer to the probability of an event
  - E.g., $X_1 = x_1$ is an event

Marginal probability distributions

- From the joint distribution p(X,Y) we can find the ***marginal distributions*** p(X) and p(Y)

P(Cavity, Toothache)	
¬Cavity ∧ ¬Toothache	0.8
¬Cavity ∧ Toothache	0.1
Cavity ∧ ¬Toothache	0.05
Cavity ∧ Toothache	0.05

P(Cavity)	
¬Cavity	?
Cavity	?

P(Toothache)	
¬Toothache	?
Toochache	?

Marginal probability distributions

- From the joint distribution p(X,Y) we can find the ***marginal distributions*** p(X) and p(Y)
- To find p(X = x), sum the probabilities of all atomic events where X = x:

$$P(X = x) = P\big((X = x \wedge Y = y_1) \vee \ldots \vee (X = x \wedge Y = y_n)\big)$$

$$= P\big((x, y_1) \vee \ldots \vee (x, y_n)\big) = \sum_{i=1}^{n} P(x, y_i)$$

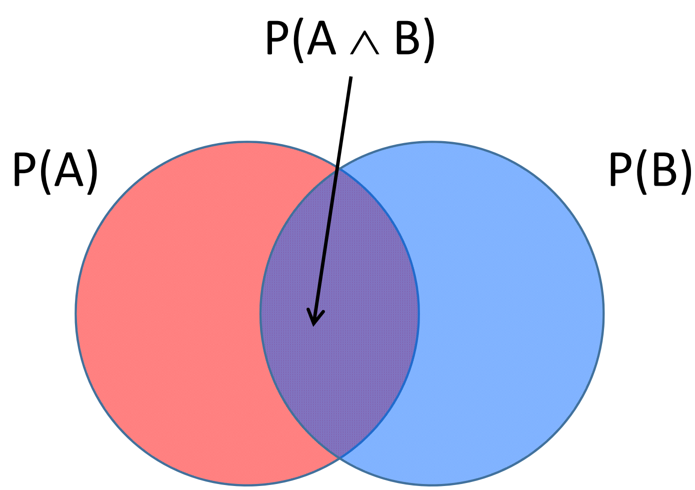
- This is called ***marginalization*** (we are *marginalizing out* all the variables except X)

Conditional probability

- Probability of cavity given toothache:
  P(*Cavity = true* | *Toothache = true*)

- For any two events A and B,

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



Conditional probability

P(Cavity, Toothache)	
$\neg Cavity \wedge \neg Toothache$	0.8
$\neg Cavity \wedge Toothache$	0.1
$Cavity \wedge \neg Toothache$	0.05
$Cavity \wedge Toothache$	0.05

P(Cavity)	
$\neg Cavity$	0.9
Cavity	0.1

P(Toothache)	
$\neg Toothache$	0.85
Toochache	0.15

- What is $p(Cavity = true \mid Toothache = false)$?
  $p(Cavity|\neg Toothache)$ = ?
- What is $p(Cavity = false \mid Toothache = true)$?
  $p(\neg Cavity|Toothache)$ = ?

Conditional distributions

- A conditional distribution is a distribution over the values of one variable given fixed values of other variables

P(Cavity, Toothache)	
¬Cavity ∧ ¬Toothache	0.8
¬Cavity ∧ Toothache	0.1
Cavity ∧ ¬Toothache	0.05
Cavity ∧ Toothache	0.05

P(Cavity | Toothache = true)	
¬Cavity	0.667
Cavity	0.333

P(Cavity|Toothache = false)	
¬Cavity	0.941
Cavity	0.059

P(Toothache | Cavity = true)	
¬Toothache	0.5
Toochache	0.5

P(Toothache | Cavity = false)	
¬Toothache	0.889
Toochache	0.111

Normalization trick

- To get the whole conditional distribution $p(X \mid Y = y)$ at once, select all entries in the joint distribution table matching $Y = y$ and renormalize them to sum to one

P(Cavity, Toothache)	
¬Cavity ∧ ¬Toothache	0.8
¬Cavity ∧ Toothache	0.1
Cavity ∧ ¬Toothache	0.05
Cavity ∧ Toothache	0.05

Select

Toothache, Cavity = false	
¬Toothache	0.8
Toochache	0.1

Renormalize

P(Toothache | Cavity = false)	
¬Toothache	0.889
Toochache	0.111

Normalization trick

- To get the whole conditional distribution p(X | Y = y) at once, select all entries in the joint distribution table matching Y = y and renormalize them to sum to one
- Why does it work?

$$\frac{P(x,y)}{\sum_{x'} P(x',y)} = \frac{P(x,y)}{P(y)}$$

by marginalization

Product rule

- Definition of conditional probability: $$P(A \mid B) = \frac{P(A,B)}{P(B)}$$
- Sometimes we have the conditional probability and want to obtain the joint:

$$P(A,B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

Product rule

- Definition of conditional probability: $P(A \mid B) = \dfrac{P(A,B)}{P(B)}$

- Sometimes we have the conditional probability and want to obtain the joint:

$$P(A,B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

- The chain rule:

$$P(A_1,\ldots,A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1,A_2)\ldots P(A_n \mid A_1,\ldots,A_{n-1})$$
$$= \prod_{i=1}^{n} P(A_i \mid A_1,\ldots,A_{i-1})$$

The Birthday problem

- We have a set of n people. What is the probability that two of them share the same birthday?
- Easier to calculate the probability that n people *do not* share the same birthday

$$
\begin{aligned}
&P(B_1,\dots B_n \text{ distinct })\\
&= P(B_n \text{ distinct from } B_1,\dots B_{n-1} \mid B_1,\dots B_{n-1} \text{ distinct })\\
&\quad P(B_1,\dots B_{n-1} \text{ distinct })\\
&= \prod_{i=1}^{n} P(B_i \text{ distinct from } B_1,\dots B_{i-1} \mid B_1,\dots B_{i-1} \text{ distinct })
\end{aligned}
$$

The Birthday problem

$$P(B_1, \ldots B_n \text{ distinct })$$

$$= \prod_{i=1}^{n} P(B_i \text{ distinct from } B_1, \ldots B_{i-1} \mid B_1, \ldots B_{i-1} \text{ distinct })$$

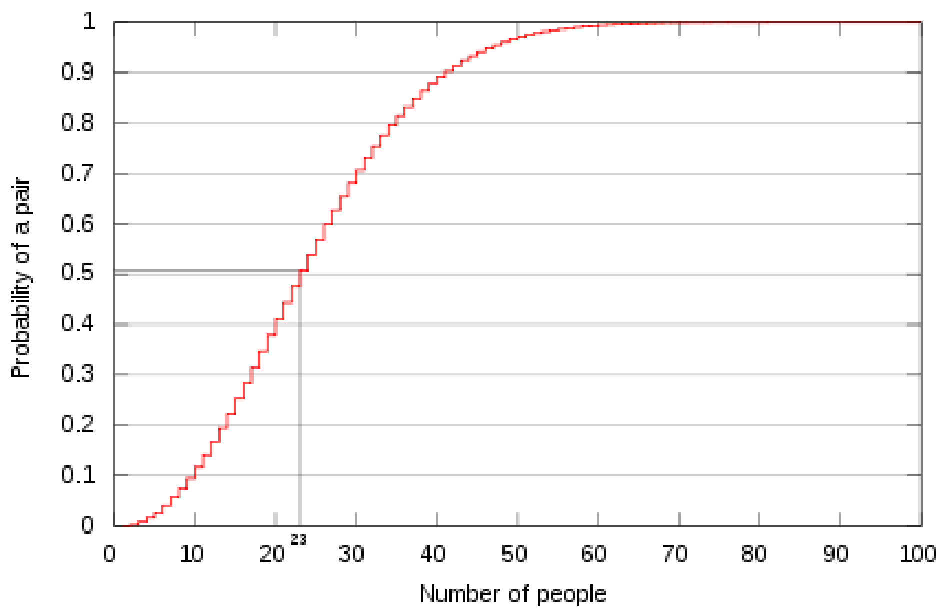
$$P(B_i \text{ distinct from } B_1, \ldots, B_{i-1} \mid B_1, \ldots, B_{i-1} \text{ distinct}) = \frac{365 - i + 1}{365}$$

$$P(B_1, \ldots, B_n \text{ distinct}) = \frac{365}{365} \times \frac{364}{365} \times \ldots \times \frac{365 - n + 1}{365}$$

$$P(B_1, \ldots, B_n \text{ not distinct}) = 1 - \frac{365}{365} \times \frac{364}{365} \times \ldots \times \frac{365 - n + 1}{365}$$

The Birthday problem

- For 23 people, the probability of sharing a birthday is above 0.5!



http://en.wikipedia.org/wiki/Birthday_problem

Outline

- Motivation: Why use probability?
  - Laziness, Ignorance, and Randomness
  - Rational Bettor Theorem
- Review of Key Concepts
  - Outcomes, Events, and Random Variables
  - Joint, Marginal, and Conditional
  - Independence and Conditional Independence

Independence

- Two events A and B are *independent* if and only if $p(A \wedge B) = p(A, B) = p(A)\ p(B)$
  - In other words, $p(A \mid B) = p(A)$ and $p(B \mid A) = p(B)$
  - This is an important simplifying assumption for modeling, e.g., *Toothache* and *Weather* can be assumed to be independent?
- Are two *mutually exclusive* events independent?
  - No, but for mutually exclusive events we have $p(A \vee B) = p(A) + p(B)$

Independence

- Two events A and B are *independent* if and only if
  $p(A \wedge B) = p(A)\, p(B)$
  - In other words, $p(A \mid B) = p(A)$ and $p(B \mid A) = p(B)$
  - This is an important simplifying assumption for modeling, e.g., *Toothache* and *Weather* can be assumed to be independent
- **Conditional independence**: A and B are *conditionally independent* given C iff
  $p(A \wedge B \mid C) = p(A \mid C)\, p(B \mid C)$
  - Equivalent:
    $p(A \mid B, C) = p(A \mid C)$
  - Equivalent:
    $p(B \mid A, C) = p(B \mid C)$

Random Audience Participation Slide

- List some pairs of events that are independent
  - … here is a pair of events ….
- List some pairs of events that are mutually exclusive
  - …. here is some different pair of events ….
- List some pairs of events that are conditionally independent given knowledge of some third event
  - … whoa, now we need event triples. …

Conditional independence: Example

- *Toothache*: boolean variable indicating whether the patient has a toothache
- *Cavity*: boolean variable indicating whether the patient has a cavity
- *Catch*: whether the dentist's probe catches in the cavity

- If the patient has a cavity, the probability that the probe catches in it doesn't depend on whether he/she has a toothache

  p(*Catch*|*Toothache, Cavity*) = p(*Catch*|*Cavity*)
- Therefore, *Catch* is conditionally independent of *Toothache* given *Cavity*
- Likewise, *Toothache* is conditionally independent of *Catch* given *Cavity*

  p(*Toothache*|*Catch, Cavity*) = p(*Toothache* | *Cavity*)
- Equivalent statement:

  p(*Toothache, Catch*|*Cavity*) = p(*Toothache*|*Cavity*) p(*Catch*|*Cavity*)

Conditional independence: Example

- How many numbers do we need to represent the joint probability table p(*Toothache, Cavity, Catch*)?

  $2^3 - 1 = 7$ independent entries

- Write out the joint distribution using chain rule:

  p(*Toothache, Catch, Cavity*)

  = p(*Cavity*) p(*Catch*|*Cavity*) p(*Toothache*|*Catch, Cavity*)

  = p(*Cavity*) p(*Catch*|*Cavity*) p(*Toothache*|*Cavity*)

- How many numbers do we need to represent these distributions?

  1 + 2 + 2 = 5 independent numbers

- In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in *n* to linear in *n*