STA 371G: Statistics and Modeling

Decision Making Under Uncertainty: Probability, Betting Odds and Bayes' Theorem

Mingyuan Zhou
McCombs School of Business
The University of Texas at Austin

http://mingyuanzhou.github.io/STA371G

Probability: Measuring Uncertainty

Probability can be thought of as the mathematical language of uncertainty. The axioms of probability can be stated as

- The probability of an event E is nonnegative, $P(E) \geq 0$
- If S denotes the set of all possible events, then $P(S) = 1$
- If two events E_1 and E_2 are mutually exclusive, then $P(E_1 \ or \ E_2) = P(E_1) + P(E_2)$

Interpretation of Probability

Frequency interpretation:

$$P(E) = \frac{\text{number of elementary events comprising } E}{\text{total number of elementary events in } S}$$

- Examples:
  - How often would one expect to obtain "heads" if a coin is tossed repeatedly?
  - Drawing a marble at random with replacement from a box with 30 red and 70 blue marbles, how often would one expect to obtain a red marble?
- The probability of an event is interpreted as the relative frequency of occurrence of an event to be expected in the long run
- Law of large numbers: if an experiment is repeated many times under identical conditions, the relative frequency of occurrence for any event is likely to be close to the probability of that event

Interpretation of Probability

Subjective interpretation: a probability is interpreted as a measure of degree of belief, or as the quantified judgement of a particular individual.

- Information in the form of observed frequencies might not be available regarding repeated trials under identical conditions.
- Examples:
  - I am likely to get a score above 90 in the test
  - It will probably rain tomorrow
  - UT football might have a 10% chance to win the national championship next year

Probabilities, Lotteries, and Betting Odds

Subjective probabilities may vary from person to person for the same event, as different individuals may have different degrees of belief, or judgments.

A formal definition of subjective probability can be given in terms of lotteries. Your subjective probability $P(E)$ of an event E is the number $P(E)$ that makes you indifferent between the following two lotteries:

Lottery A:	You obtain X with probability $P(E)$ You obtain Y with probability $1 - P(E)$
Lottery B:	You obtain X if E occurs You obtain Y if E does not occur

Probabilities, Lotteries, and Betting Odds

Lottery A:	You win \$100 with probability 40%
	You win \$0 with probability 60%
Lottery B:	You obtain \$100 if it rains tomorrow
	You obtain \$0 if it does not rain tomorrow

Lottery A:	You loose \$10 with probability 70%
	You loose \$20 with probability 30%
Lottery B:	You loose \$10 if it rains tomorrow
	You loose \$20 if it does not rain tomorrow

Probabilities, Lotteries, and Betting Odds

If the probability of an event is p, then the odds in favor of that event are p to $(1 - p)$.

If the odds in favor of an event are a to b, then the probability of that event is equal to $a/(a + b)$. If the odds against an event are a to b, then the probability of that event is equal to $b/(a + b)$.

Two persons make the following bet:

Person A bets \$3 that IBM stock price goes up tomorrow
Person B bets \$2 that IBM stock price does not go up tomorrow

If you think this is a "fair" bet, what are the odds in favor of IBM stock price going up tomorrow? What are the odds against IBM stock price going up tomorrow? What's your probability that IBM stock price goes up tomorrow?

Payoff Tables

The consequence of the decision can be expressed in a Payoff Table.

- For example, depending on whether or not it rains, the consequence of your decision on whether or not to carry an umbrella can be summarized in a Payoff Table as:

	Rain	*No Rain*
Carry umbrella	$-\$3$	$-\$3$
Do not carry umbrella	$-\$90$	$\$0$

- Your uncertainty: $P(Rain)$ and $P(No\ Rain)$
- Your decision: *Carry umbrella* or *Do not carry umbrella*
- The payoffs for different decisions may vary from person to person
- Sources of information: weather report, historical data, yesterday's weather, look out the window before you leave home...

Example: Oil Drilling

The decision maker must decide whether or not to drill for oil at a particular location.

- The cost of drilling is \$300,000.
- If the decision is to drill and Oil is found, then the Oil will be worth \$1,300,000.
- If the decision is not to drill, then the drilling rights will be sold for a fixed sum of \$100,000 plus another \$100,000 contingent upon the presence of oil.

Construct the payoff table and discuss how to model the decision maker's uncertainty about the presence of oil.

- The payoff table can be expressed as:

	Oil	*No Oil*
Drill	\$1,000,000	–\$300,000
Do not drill	\$200,000	\$100,000

- Decision maker's uncertainty: $P(Oil)$ and $P(No\ Oil)$
- Decision maker's decision: *Drill* or *Do not drill*
- Sources of information: experience on similar oil-drilling locations, geological information, a geologist's assessment
- What would be your recommendation to the decision maker if the geologist assesses that the odds against oil is 7 to 3?

Example: Advertising

A decision maker has to decide whether to launch ad campaign.

- If no ad campaign is launched, then the revenue is \$700,000 if the competitor does not introduce a new product and decreases to \$100,000 otherwise.
- If a minor ad campaign at the cost of \$150,000 is launched, then the revenue is \$750,000 if the competitor does not introduce a new product and decreases to \$450,000 otherwise.
- If a major ad campaign at the cost of \$300,000 is launched, then the revenue is \$800,000 if the competitor does not introduce a new product and decreases to \$700,000 otherwise.

Construct the payoff table and discuss how to model the decision maker's uncertainty.

- The payoff table can be expressed as:

	Competitor introduces new product	*Competitor introduces no new product*
No advertising	\$100,000	\$700,000
Minor ad campaign	\$300,000	\$600,000
Major ad campaign	\$400,000	\$500,000

- Decision maker's uncertainty: ?
- Sources of information: ?

The Conditional Nature of Probability

Conditional probability is an important concept in the theory of probability. Often one is interested in the probability that one event will occur, given that a particular second event has occurred or will occur.

$P(E_2|E_1)$: the conditional probability of event E_2 given event E_1.

Conditional Probability

Your (subjective) conditional probability of event E_2 given event E_1 is the number $P(E_2|E_1)$ that makes you indifferent between the following two lotteries:

Lottery A:	If E_1 occurs, you obtain X with probability $P(E_2\|E_1)$ and you obtain Y with probability $1 - P(E_2\|E_1)$.
Lottery B:	If E_1 occurs, you obtain X if E_2 occurs and you obtain Y if E_2 does not occur.

Lottery A:	If it rains today, you win \$100 with probability $1/2$ and you win \$0 with probability $1/2$.
Lottery B:	If it rains today, you win \$100 if it rains tomorrow and you win \$0 if it does not rain tomorrow.

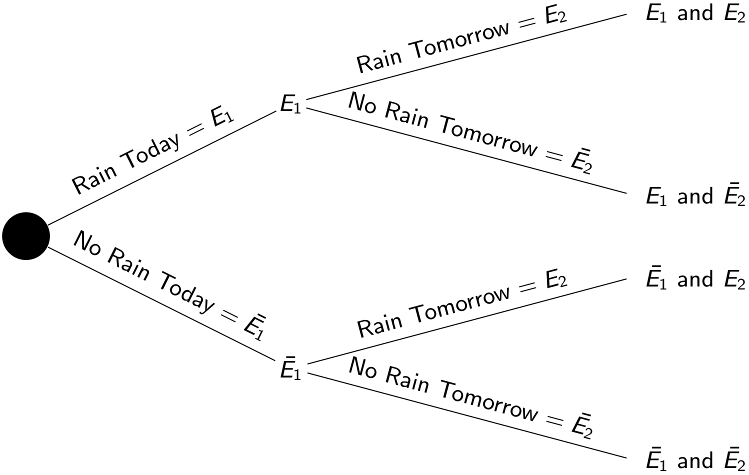
Conditional Probability

It is also possible to consider conditional probability in terms of conditional bets or conditional reference contracts.

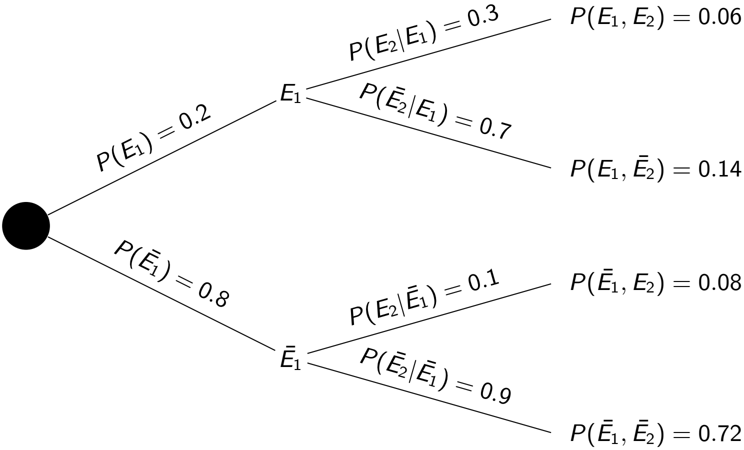
- Person A bets \$50 that the Oakland A's wins over half of its games in the coming season.
- Person B bets \$20 that the team does not win over half of its games, with the proviso that the bet is called off if any of a stipulated list of key players of Oakland A's is injured during the course of the season.

If you think this is a fair conditional bet, then what's your conditional probability $P(\text{Oakland A's wins over half of its games} \mid \text{no injuries of key players}) = ?$

Tree Diagram



Tree Diagram



Joint Probability Table

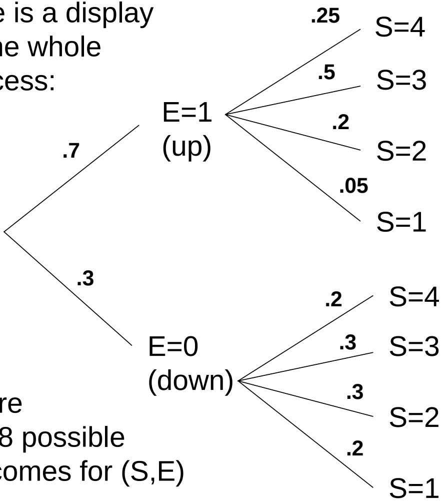
$$P(E_1, E_2) = P(E_2|E_1)P(E_1)$$

	E_2	$\bar{E}_2$	
E_1	$P(E_1, E_2)$	$P(E_1, \bar{E}_2)$	$P(E_1)$
$\bar{E}_1$	$P(\bar{E}_1, E_2)$	$P(\bar{E}_1, \bar{E}_2)$	$P(\bar{E}_1)$
	$P(E_2)$	$P(\bar{E}_2)$	1

	Rain tomorrow	No rain tomorrow	
Rain today	0.14	0.06	0.20
No rain today	0.08	0.72	0.80
	0.22	0.78	1

Tree Diagram

here is a display of the whole process:



p(S=4 and E=1) = .7*.25 = .175

p(S=3 and E=1) = .7*.5 = .35

p(S=2 and E=1) = .7*.2 = .14

p(S=1 and E=1) = .7*.05 = .035

p(S=4 and E=0) = .3*.2 = .06

p(S=3 and E=0) = .3*.3 = .09

p(S=2 and E=0) = .3*.3 = .09

p(S=1 and E=0) = .3*.2 = .06

There are 8 possible outcomes for (S,E)

Joint Probability Table

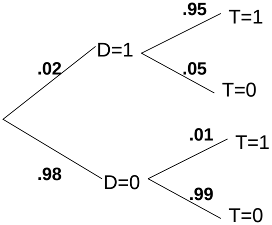
Joint Probability Table for Sales (S) and Economy (E)

E \ S	1	2	3	4	
0	.06	.09	.09	.06	.3
1	.035	.14	.35	.175	.7
	.095	.23	.44	.235	1

Bayes' Theorem

Disease testing example...

Let $D = 1$ indicate you have a disease
Let $T = 1$ indicate that you test positive for it



$\Longrightarrow$

		D	
		0	1
T	0	.9702=.98*.99	.001
	1	.0098	.019

If you take the test and the result is positive, you are really interested in the question: What's the chance for a person with a positive test to have the disease?

Bayes' Theorem

		D	
		0	1
T	0	.9702	.001
	1	.0098	.019

$$P(D=1|T=1) = \frac{0.019}{(0.019+0.0098)} = 0.66$$

$$P(D=1|T=0) = ?$$

Bayes' Theorem

A convenient formula that gives the relationship among various conditional probabilities is *Bayes' Theorem*:

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

- $P(A)$: prior, initial degree of belief in A
- $P(B|A)$: likelihood
- $P(B) = \sum_j P(B|A = A_j)P(A = A_j)$
- $P(A|B)$: posterior, degree of belief in A after accounting for evidence B

Bayes' theorem describes the computation of the posterior probabilities given the prior probabilities and likelihoods.

Bayes' Theorem

In the disease testing example:

$P(D=1|T=1) = \frac{P(T=1|D=1)P(D=1)}{P(T=1)}$

- $P(D=1) = 0.02$: the initial degree of belief in $D=1$ is 2%
- $P(T=1|D=1) = 0.95$: likelihood
- $P(T=1) = P(T=1|D=1)P(D=1) + P(T=1|D=0)P(D=0) = 0.0288$
- $P(D=1|T=1) = \frac{0.95*0.02}{0.0288} = 0.66$:
  given $T=1$, the degree of belief in $D=1$ becomes 66%

$$P(D=0|T=0) = ?$$

Frequentist Interpretation

In Bayesian interpretation, probability measures a degree of belief.
In frequentist interpretation, probability measures a proportion of outcomes.

- Try to think about this intuitively... imagine you are about to test 100,000 people.
- we assume that about 2,000 of those have the disease.
- we also expect 1% of the disease-free people to test positive, ie, 980, and 95% of the sick people to test positive, ie 1,900. So, we expect a total of 2,880 positive tests.
- Choose one of the 2,880 people at random... what is the probability that he/she has the disease?

$$P(D = 1|T = 1) = 1,900/2,880 = 0.66$$

- isn't that the same?!

Bayes' Rule

Bayes' theorem is also commonly referred as Bayes' rule.

Let $A_1, A_2, \ldots, A_p$ be the possible outcomes of a random event... Also, let B be the outcome of an event that happened before we get to see A. By Bayes' Rule:

$$p(A_i|B) = \frac{p(B|A_i)p(A_i)}{p(B|A_1)p(A_1) + \cdots + p(B|A_p)p(A_p)}$$

Example: Drug Testing

Imagine we are trying to test football players for the possible use of steroids.

The result of the test is either *positive* or *negative*. In general, tests are not perfect and some athletes who are drug free test positive (false positive) while some other drug users test negative (false negatives).

Assume that in general you believe that 5% of players use steroids and the probability of a false positive is 3% and of a false negative is 7%.

Assume we decide to perform this test on Colt McCoy and he tests positive! Are we sure he is indeed a drug user?

Example: Drug Testing

- We are looking for $p(S|+)$, ie, the probability of steroids usage (S) given a positive test result ($+$)
- What do we have? $p(+|S)$, $p(+|NS)$, $p(-|S)$, $p(-|NS)$, $p(S)$ and $p(NS)$... right?

By Bayes' rule:

$$p(S|+) = \frac{p(+|S)p(S)}{p(+|S)p(S) + p(+|NS)p(NS)} = 0.620$$

Why?

Example: Drug Testing

- Without knowing whether or not one is a drug user what is the marginal probability of a positive test? $p(+)$

$$p(+) = p(+|S)p(S) + p(+|NS)p(NS) = 0.075$$

- and so $p(-) = 0.925$

Why?

Example: Drug Testing

- Try to think about this intuitively... imagine you are about to test 100,000 players.
- You assume that 5,000 of those use steroids.
- We also expect 3% of the non-users to test positive, ie, 2,850, and 93% of the users to test positive, ie 4,650. So, we expect a total of 7,500 positive tests.
- Choose one of the 7,500 players at random... what is the probability that he is a drug user?

$$p(S|+) = 4,650/7,500 = 0.62$$

- isn't that the same?!

Simpson's Paradox: Berkeley Gender Bias Case

UC-Berkeley was one of the first universities to be sued for sexual discrimination. The graduate admission rates in the fall of 1973 were 44% for the male applicants but only 35% for the female applicants. However, examining the admission rates for individual departments did not reveal discrimination against female applicants.

Table : Graduate admission

Department	A	B	C	D	E	F	A-F
Male	62%	63%	37%	33%	28%	6%	52%
Female	82%	68%	34%	35%	24%	7%	42%

Berkeley Gender Bias Case

Let $A =$"admitted to Berkeley". In 1973 it was noted that $P(A \mid \text{male}) = 0.44$ while $P(A \mid \text{female}) = 0.35$. Meanwhile, individual departments showed no signs of discrimination. Consider that the chemistry department and the psychology department are the only two options and the following information:

	Chemistry	Psychology
$P(A \mid \text{female})$	0.6	0.3
$P(A \mid \text{male})$	0.5	0.25

What is going on?

Berkeley Gender Bias Case

	Chemistry	Psychology
$P(A \mid \text{F})$	0.6	0.3
$P(A \mid \text{M})$	0.5	0.25

$$P(A \mid \text{F}) = P(A \mid \text{F, chem})P(\text{chem} \mid \text{F}) + P(A \mid \text{F, psych})P(\text{psych} \mid \text{F})$$
$$= 0.6q_f + 0.3(1 - q_f) = 0.35 \quad \text{(hence)} \quad q_f = 0.167$$

$$P(A \mid \text{M}) = P(A \mid \text{M, chem})P(\text{chem} \mid \text{M}) + P(A \mid \text{M, psych})P(\text{psych} \mid \text{M})$$
$$= 0.5q_m + 0.25(1 - q_m) = 0.44 \quad \text{(hence)} \quad q_m = 0.76$$

Berkeley Gender Bias Case

The explanation for the apparent overall bias was that women have a higher probability of applying to Psychology than to Chemistry (assuming for simplicity that these are the only two options) and overall Psychology has a lower admissions rate!

This is a cautionary tale! Before we can act on a apparent association between two variables (for example, sue Berkeley) we need to account for potential lurking variables that are the real cause of the relationship. Again, association is NOT causation!

Simpson's Paradox: MBA Promotion

Table : Promotion Rate

	MBA Yes (B)	MBA No ($\bar{B}$)
First Employer Yes (C)	0.10	0.05
First Employer No ($\bar{C}$)	0.35	0.20

$$P(Promote|B, C) > P(Promote|\bar{B}, C)$$

$$P(Promote|B, \bar{C}) > P(Promote|\bar{B}, \bar{C})$$

Employees with MBA degrees are MORE likely to be promoted.

Table : Promotion Rate

MBA Yes (B)	MBA No ($\bar{B}$)
0.125	0.155

Employees with MBA degrees are LESS likely to be promoted!

What's going on?

Simpson's Paradox: MBA Promotion

$$
\begin{aligned}
P(Promote|B) &= P(Promote, C|B) + P(Promote, \bar{C}|B) \\
&= P(Promote|B, C)P(C|B) + P(Promote|B, \bar{C})P(\bar{C}|B) \\
&= 0.10P(C|B) + 0.35(1 - P(C|B)) \\
&= 0.125
\end{aligned}
$$

$=>$ $P(C|B) = 0.90$

$$
\begin{aligned}
P(Promote|\bar{B}) &= P(Promote, C|\bar{B}) + P(Promote, \bar{C}|\bar{B}) \\
&= P(Promote|\bar{B}, C)P(C|\bar{B}) + P(Promote|\bar{B}, \bar{C})P(\bar{C}|\bar{B}) \\
&= 0.05P(C|\bar{B}) + 0.20(1 - P(C|\bar{B})) \\
&= 0.155
\end{aligned}
$$

$=>$ $P(C|\bar{B}) = 0.30$

Simpson's Paradox: MBA Promotion

Assuming $P(B) = P(\bar{B}) = 0.5$ and $Promotion = A$, can we find $P(B, C|A)$, $P(\bar{B}, C|A)$, $P(B, \bar{C}|A)$ and $P(\bar{B}, \bar{C}|A)$?
Try the Bayes Rule:

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

where

$$\begin{aligned} P(A) = {} & P(A|B, C)P(C|B)P(B) \\ & + P(A|\bar{B}, C)P(C|\bar{B})P(\bar{B}) \\ & + P(A|B, \bar{C})P(\bar{C}|B)P(B) \\ & + P(A|\bar{B}, \bar{C})P(\bar{C}|\bar{B})P(\bar{B}) \end{aligned}$$

How about $P(C|A)$ and $P(B|A)$?
How about $P(A|C)$ and $P(A|\bar{C})$?

Simpson's Paradox

Drug A has a higher overall cure rate than Drug B, but has lower cure rates for different disease types.

Player A has a higher batting average than Player B for each of the past five years, but has a lower batting average when these five years are combined.

Additional examples?