Decision Theory

1.6 & 1.7 Probability

George Mason University, Spring 2016

We've got a working understanding of expected utility theory, now we need an account of probability. We'll start with a development of the probability calculus, which is what we need to apply probabilities in decision theory and which is the touchstone that is agreed upon by any philosophical account of probability. We'll then turn to attempts to say just what claims about probability really are saying.

Some terminology:

absolute probabilty $Pr(x)$, e.g., $Pr(\text{ace of spades}) = 1/52$. Read "the probability of x is..."

conditional probability $Pr(x|y)$, e.g., $Pr(\text{ace}|\text{spade}) = \frac{1}{13}$. Read, "the probability of x given y is..."

independent x is independent of y iff $Pr(x) = P(x|y)$. For example, $Pr(ace) = \frac{4}{52}$ and $Pr(\text{ace on draw 2}|\text{ace on draw 1}) = \frac{3}{51}$ if we do not replace the card after the first draw. In this case, the statements are not independent. However, $Pr(\text{ace on draw 2}|\text{ace on draw 1}) = \frac{4}{52}$ if we *do replace* the card after the first draw, so the statements are independent. Things like coin flips, rolls of a die, and spins of a roulette wheel are independent events. The outcome of one doesn't effect the outcome of another.

mutually exclusive x and y are mutually exclusive iff it is impossible for both to be true, i.e., $Pr(x|y) = Pr(y|x) = 0$. For example, "the coin lands heads" and "the coin lands tails" are mutually exclusive statements.

The Kolmogorov Probability Axioms (Set Theoretical Notation):

Ax1. $0 \leq Pr(X) \leq 1$.

Ax2. $Pr(\Omega) = 1$.

Ax3. If $X \cap Y = \emptyset$, then $Pr(X \cup Y) = Pr(X) + Pr(Y)$.

The Kolmogorov Probability Axioms (Logical Notation):

Ax1. $0 \leq Pr(X) \leq 1$.

Ax2. $Pr(\Omega) = 1$.

Ax3. If $X \wedge Y$ are mutually exclusive, then $Pr(X \vee Y) = Pr(X) + Pr(Y)$.

Thm 1.6.1. $Pr(X) + Pr(\neg X) = 1$.

Proof.

Since $X \wedge \neg X$ are mutually exclusive, $Pr(X \vee \neg X) = Pr(X) + Pr(\neg X)$ by Ax3.

$Pr(X \vee \neg X) = Pr(\Omega) = 1$ by Ax2.

Thus, $Pr(X) + Pr(\neg X) = 1$. $\square$

This gives us a rule for calculating negated statements. If we know $Pr(x)$, then we can calculated $Pr(\neg x)$ by subtracting $Pr(x)$ from 1.

Ex. What's the probability of taking more than one roll of a fair die to get a 6? We know it's the same as the probability of not getting a 6 on the first roll, so that's what we need to calculate.

Thm 1.6.2. If X and Y are logically equivalent, then $Pr(X) = Pr(Y)$.

If X and Y are logically equivalent, then X and $\neg Y$ are mutually exclusive.

$Pr(X \vee \neg Y) = Pr(x) + Pr(\neg Y)$ by Ax3.

Since $X \vee \neg Y$ is true (since it is equivalent to $X \vee \neg X$), we have $Pr(X \vee \ \neg Y) = 1$ by Ax2.

So, (1) $Pr(X) + Pr(\neg Y) = 1$.

(2) $Pr(Y) + Pr(\neg Y) = 1$, Thm 1.6.1.

Since $Pr(Y)$ has the same value in (1) and (2), $Pr(X) = Pr(Y)$

Thm 1.6.3. $Pr(X \vee Y) = Pr(X) + Pr(Y) - Pr(X \wedge Y)$.

Proof.

$Pr(X \vee Y) = Pr((X \wedge Y) \vee (\neg X \wedge Y) \vee (X \wedge \neg Y))$.

$Pr(X \vee Y) = Pr((X \wedge Y) \vee (\neg X \wedge Y)) + Pr(X \wedge \neg Y)$ by Ax3.

$Pr(X \vee Y) = Pr(Y) + Pr(X \wedge \neg Y)$.

$Pr(X \wedge Y) + Pr(X \vee Y) = Pr(Y) + Pr(X \wedge \neg Y) + Pr(X \wedge Y)$.

$Pr(X \wedge Y) + Pr(X \vee Y) = Pr(Y) + Pr(x)$.

$Pr(X \vee Y) = Pr((X \wedge Y) \vee (\neg X \wedge Y) \vee (X \wedge \neg Y))$. $\square$

Def 1.6.1. The *conditional probability of X given Y* is $Pr(X|Y) = \frac{Pr(X \wedge Y)}{Pr(Y)}$ given that $Pr(Y) \neq 0$.

Ex. You roll a fair die twice. Given than the first roll is a 5, what is the probability that the sum of the rolls will exceed 9?

$$Pr(T > 9|5) = \frac{Pr(T>9 \wedge 5)}{Pr(5)} = \frac{Pr(T>9) \times Pr(5)}{Pr(5)} = \frac{\frac{1}{3} \times \frac{1}{6}}{\frac{1}{6}} = \frac{1}{3}$$

Def 1.6.2. X is *probabilistically independent* of Y if and only if $Pr(X) = Pr(X|Y)$.

Thm 1.6.4. If X is probabilistically independent of Y, then $Pr(X \wedge Y) = Pr(X) \times Pr(Y)$.

Proof.

This follows directly from Definitions 1.6.1 and 1.6.2.

Ex. What's the probability of flipping 4 heads in a row?

Thm 1.6.5 (Inverse Probability Law). $Pr(X|Y) = \frac{Pr(X) \times Pr(Y|X)}{Pr(Y)}$ given that $Pr(Y) \neq 0$.

Ex. Lung Spot.

You are a physician and have just observed a spot on an X-ray of your patient's lung. You think that the patient might have tuberculosis and you are considering a treatment for this disease with slightly harmful side effects. If the patient has tuberculosis and you treat it, then you will cure the disease. If you do not treat it, then the disease will get worse and the patient will end up suffering considerably. You know that the probability of observing a lung spot given that a patient has tuberculosis is 20%. You also know that the unconditional probability of observing a lung spot is 10% and the incidence of tuberculosis in the general population is 5%. Do you administer the treatment?

	[?]	[?]
	tuberculosis	no tuberculosis
treatment	mild harm	mild harm
no treatment	extreme harm	no harm

$$Pr(LS|TB) = 0.2$$

$$Pr(LS) = 0.1$$

$$Pr(TB) = 0.05$$

$$Pr(TB|LS) = \frac{Pr(TB) \times Pr(LS|TB)}{Pr(LS)} = \frac{0.05 \times 0.2}{0.1} = 0.1$$

	$[\frac{1}{10}]$ tuberculosis	$[\frac{9}{10}]$ no tuberculosis
treatment	mild harm	mild harm
no treatment	extreme harm	no harm

Thm 1.6.6 (Bayes' Theorem).

$Pr(X|Y) = \dfrac{Pr(X) \times Pr(Y|X)}{Pr(X) \times Pr(Y|X) + Pr(\neg X) \times Pr(Y|\neg X)}$ given that $Pr(Y) \neq 0$.

$Pr(X)$ is called the *prior probability*. This is the unconditional probability of some event taking place.

$Pr(X|Y)$ is called the *posterior probability*. This is what we learn when we apply Bayes' or the Inverse Probability Law

$Pr(X|Y)$, i.e., the *posterior probability* can be used as a *new prior* in further applications of the Inverse Probability Law or Bayes' Theorem. That is, as we gather new evidence, we can use this as our starting point for updating our beliefs.

What if you know little about the probability of tuberculosis at the onset?

‘Washing of the priors’: Different prior probabilities will converge to the same value after repeated application of the Inverse Probability Law or Bayes’ Theorem on new data. So, Bayesians just say use your best hunch to get started, then start collecting as much evidence as you can.

Ex. What Do You Learn From a Mammogram

Suppose you are a fifty year old woman who has just had her first mammogram. Your test, unfortunately, came back positive. What does this tell you about the likelihood that you have cancer?

Ex. What Do You Learn From a Mammogram

Suppose you are a fifty year old woman who has just had her first mammogram. Your test, unfortunately, came back positive. What does this tell you about the likelihood that you have cancer?

Here are some things we know. The sensitivity of mammograms is 87%, i.e., in 87% of cases in which a person has breast cancer and undergoes mammography, the test will spot it. Because mammograms are so sensitive, they lack specificity. The false positive rate for mammography for a first mammogram is about 7-12% (let's call it 10%). We also know that the incidence of breast cancer in women 50-54 years old in the US is about 225 in 100,000.

Let's summarize what we know:

Here are some things we know. The sensitivity of mammograms is 87%, i.e., in 87% of cases in which a person has breast cancer and undergoes mammography, the test will spot it. Because mammograms are so sensitive, they lack specificity. The false positive rate for mammography for a first mammogram is about 7-12% (let's call it 10%). We also know that the incidence of breast cancer in women 50-54 years old in the US is about 225 in 100,000.

$Pr(cancer) = ??$ - This is the *prior probability* that you have breast cancer.

$Pr(\neg cancer) = ??$

$Pr(positive|cancer) = ??$

$Pr(positive|\neg cancer) = ??$

$Pr(cancer|positive) = ??$ - This is the *posterior probability*, i.e., the likelihood that you have breast cancer given a positive test result. It's the bit we really want to know.

Let's summarize what we know:

Here are some things we know. The sensitivity of mammograms is 87%, i.e., in 87% of cases in which a person has breast cancer and undergoes mammography, the test will spot it. Because mammograms are so sensitive, they lack specificity. The false positive rate for mammography for a first mammogram is about 7-12% (let's call it 10%). We also know that the incidence of breast cancer in women 50-54 years old in the US is about 225 in 100,000.

$Pr(cancer) = .002$

$Pr(\neg cancer) = .998$

$Pr(positive|cancer) = .87$

$Pr(positive|\neg cancer) = .10$

$Pr(cancer|positive) = ??$ - This is the *posterior probability*, i.e., the likelihood that you have breast cancer given a positive test result. It's the bit we really want to know.

$$Pr(cancer) = .002$$

$$Pr(\neg cancer) = .998$$

$$Pr(positive|cancer) = .87$$

$$Pr(positive|\neg cancer) = .10$$

$$Pr(cancer|positive) =$$

$$\frac{Pr(cancer) \times Pr(positive|cancer)}{Pr(cancer) \times Pr(positive|cancer) + Pr(\neg cancer) \times Pr(positive|\neg cancer)} =$$

$$Pr(cancer) = .002$$

$$Pr(\neg cancer) = .998$$

$$Pr(positive|cancer) = .87$$

$$Pr(positive|\neg cancer) = .10$$

$$Pr(cancer|positive) =$$

$$\frac{Pr(cancer) \times Pr(positive|cancer)}{Pr(cancer) \times Pr(positive|cancer) + Pr(\neg cancer) \times Pr(positive|\neg cancer)} =$$

$$\frac{.002 \times .87}{.002 \times .87 + .998 \times .10} = .02$$

Let's look at this graphically:

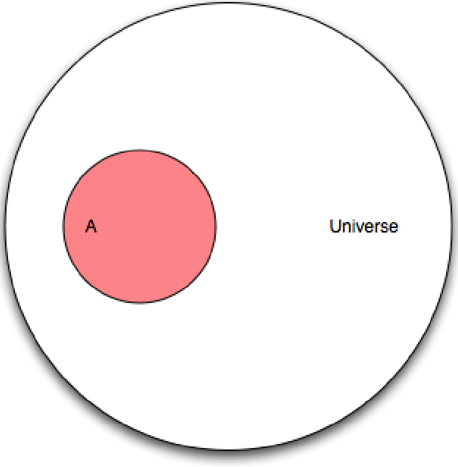
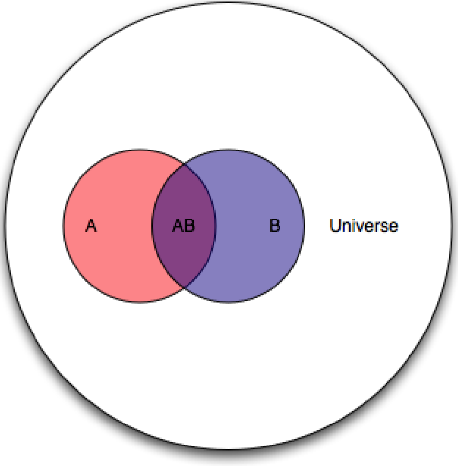


Figure: $P(A)$

Now we add another event:

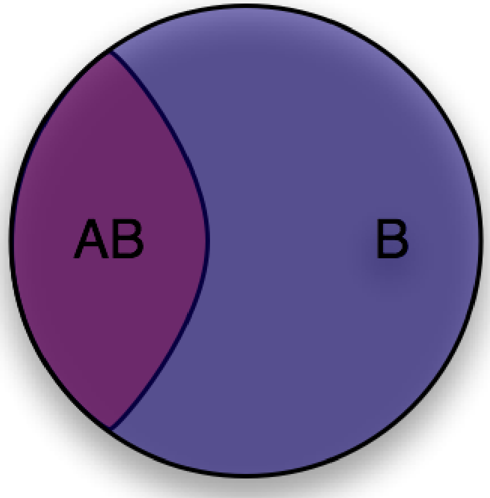
$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

$P(AB) = P(A \wedge B) = P(A) \times P(B)$

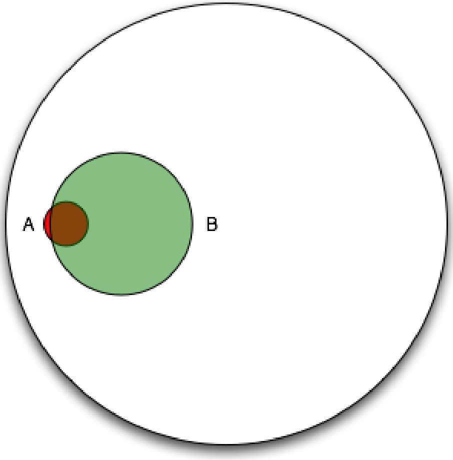


What are we really looking for?

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



Now let's look at our case:
$Pr(A) = Pr(cancer) = .002$ (the circle should actually be smaller)
$Pr(B) = Pr(positive) = .087$ (we didn't know this before, but we can calculate it...how?)
$Pr(A|B) = Pr(cancer|positive) = .02$



$$Pr(B) = Pr(positive) = .087 =$$

$$\frac{Pr(cancer) \times Pr(positive|cancer)}{Pr(cancer|positive)}$$

Inverse Probability Law vs. Bayes' Theorem

When is each to be used?

Both help you calculate the probability that some statement is true (or that some event is the case) given some new information that you have (the result of a test, experiment, observation, testimony from someone, etc.).

Bayes' uses information that is often more readily available to you. In our example, since we didn't know the prior probability of getting a positive mammogram result, we couldn't use the Inverse Probability Law. However, since we knew the false positive rate, i.e., the probability of getting a positive given that you don't have cancer, we could use Bayes'.

We often find ourselves in similar situations: we know a conditional probability but not an absolute probability, in those cases, Bayes' works. If we know the absolute probability in question, then we use the Inverse Probability Law.

Calculating the Value of New Information:

Clark is deciding whether to invest \$50,000 in the Daltex Oil Company. The company is a small one owned by some acquaintances of his, and Clark has heard a rumor that Daltex will sell shares of stock publicly within the year. If that happens he will double his money; otherwise he will earn only the unattractive return of 5% for the year and would be better off taking his other choice—buying a 10% savings certificate. He believes there is about an even chance that Daltex will go public. (Resnik 57)

Let's construct Clark's decision matrix.

Clark's Decision Matrix for Daltex Investment

	[.5] Daltex Goes Public	[.5] Does Not	EMV
Invest in Daltex	\$100,000	\$52,500	=\$76,250
Savings Certificate	\$55,000	\$55,000	=\$55,000

Now suppose that Clark has an absolutely reliable source who can tell him how with certainty whether Daltex is going public—for a price. How much should he pay for this information?

Well, we don't need to do any fancy math.

We can see that if Clark learns Daltex is going public, his EMV=\$100,000, otherwise, he'd choose the savings certificate and his EMV=\$55,000. Since he thinks there are even odds that he could learn either bit of information, we can calculate his EMV of getting the information: $\$100,000(.5) + \$55,000(.5) = \$77,500$.

Since this is \$1,250 more than his EMV w/o the information, that's the most he should pay for it.

Now suppose that Clark knows Daltex is preparing a confidential annual report and he also knows that if they are going public there is a chance of .9 that they will says so in the report and only a .1 chance that they will deny it. On the other hand, if they are not going public there is a chance of .5 that they will say they are not and .5 chance that they will lie and say they are.

Clark knows someone who will show him the report—for a price. How much should Clark pay to see it?

What we want to figure out is the probability that Daltex will (or will not) go public given that they affirm (or deny) that they will in the report.

Let:
P=going public
Y=affirm they will
D=deny they will

We use Bayes':

$$Pr(P|Y) = \frac{Pr(P) \times Pr(Y|P)}{Pr(P) \times Pr(Y|P) + Pr(\neg P) \times Pr(Y|\neg P)}$$

$$= \frac{.5 \times .9}{.5 \times .9 + .5 \times .5} = .64$$

We can also calculate:

$$Pr(P|D) = \frac{Pr(P) \times Pr(D|P)}{Pr(P) \times Pr(D|P) + Pr(\neg P) \times Pr(D|\neg P)}$$

$$= \frac{.5 \times .1}{.5 \times .1 + .5 \times .5} = .17$$

Now what do we do with this information? What's it mean?

We construct two new decision tables:

If Clark finds out they affirm:

	[.64] Daltex Goes Public	[.36] Does Not	EMV
Invest in Daltex	\$100,000	\$52,500	=\$82,900
Savings Certificate	\$55,000	\$55,000	=\$55,000

If Clark finds out they deny:

	[.17] Daltex Goes Public	[.83] Does Not	EMV
Invest in Daltex	\$100,000	\$52,500	=\$60,575
Savings Certificate	\$55,000	\$55,000	=\$55,000

In this case, we see that no matter what Clark learns about the report, he's not going to change his behavior. As such, he shouldn't be willing to pay anything at all for it.

Now let's consider the same scenario with a few tweaks. In this case if they are going public there is a chance of .95 that they will says so in the report and only a .05 chance that they will deny it. On the other hand, if they are not going public there is a chance of .95 that they will say they are not and .05 chance that they will lie and say they are.

Clark knows someone who will show him the report—for a price. How much should Clark pay to see it?

$$Pr(P|Y) = = \frac{.5 \times .95}{.5 \times .95 + .5 \times .05} = .95$$

$$Pr(P|D) = = \frac{.5 \times .05}{.5 \times .05 + .5 \times .95} = .05$$

We again construct two new decision tables:

If Clark finds out they affirm:

	[.95] Daltex Goes Public	[.05] Does Not	EMV
Invest in Daltex	\$100,000	\$52,500	=\$97,625
Savings Certificate	\$55,000	\$55,000	=\$55,000

If Clark finds out they deny:

	[.05] Daltex Goes Public	[.95] Does Not	EMV
Invest in Daltex	\$100,000	\$52,500	=\$54,875
Savings Certificate	\$55,000	\$55,000	=\$55,000

This time, it does affect the decision, so we need to find the EMV of gaining access to the report. Clark thinks there's a .5 chance that the report will say they're going public:

$.5(\$97,625) + .5(\$55,000) = \$76,312.50$

So, $\$76,312.50 - \$76,250 = \$62.50$ is the most he should pay.

Where do the probabilities come from that enter into expected utility calculations? Or, to put it another way, what do the statements of the probability calculus mean?

We'll look at four possibilities:

- Classical or Laplacean View
- Frequency View
- Propensity View
- Subjective View

Each of these interpretations can be shown to satisfy the probability axioms, but each also has it's problems.

Classical Laplacean View: The probability of X is the ratio of the number of X-cases to the total number of relevant cases.

Ex. An American roulette wheel has 38 spaces numbered 00-36. The odds of the winning number being 7 is just $\frac{7}{38}$, i.e., the number of winning cases to the total number of relevant cases.

This is an *objective logical* view of probability.

Problem 1. It is assumed that each of the relevant cases associated with X is equally possible. Is this a reasonable assumption? What justifies this assumption?

Problem 2. What if there are an infinite number of equally possible cases? Ex., I ask you to choose a real number between 0 and 1.

Long Run Frequency View: The probability of X is the frequency of X-cases in repeated trials in the limit.

Ex. A fair coin is flipped 1000 times and comes up heads 482 times, $Pr(H) = .482$.

This is an *objective empirical* view of probability.

Problem 1. Observed frequencies can be far from long run frequencies. We need to identify the appropriate *reference class*.

Problem 2. This view does not apply nicely, if at all, to single events that are not amenable to trials.

- Venn's response: distinguish *observed* and *limiting* frequencies.

Propensity View: Probabilities are to be understood as claims about the disposition of things in the world to bring about certain effects. They are claims about the nature of the objects.

Ex. A fair die is disposed to land with six facing up once every six rolls.

This is an *objective empirical* view of probability.

Problem 1. We can't observe propensities, only events. Any evidence for claims about dispositions must come from observed events, but then this collapses into the frequency view.

Problem 2. Humphrey's Paradox: we can state inverted probabilities, but because dispositions track causal relationships, we can't invert propensities. Propensities have a temporal direction, probabilities do not.

Subjective View: The probability of X is an agent's degree of belief, or *credence*, in X's occurrence.

Problem. How to measure credences?

Problem. An agent's credences needn't satisfy the probability axioms.

Ramsey and De Finetti:

Credences are reflected in willingness to bet. An agent has credence $Cr(X)$ in X's occurrence just in case this agent is willing to take either side of a bet B where for any stake \$$S$, the loser pays the winner $\$(1 - Cr(X)) \times S$ if X and $\$Cr(X) \times S$ if $\neg X$:

Bet B	
X	$\$(1 - Cr(X)) \times S$
$\neg X$	$\$Cr(X) \times S$

Problem. Placing a bet on an event can sometimes affect whether this event occurs.

Problem. Betting can have collateral benefits or costs.

Problem. Utility needn't be linear in dollars.

Problem. A bet on an event only becomes winning when its occurrence becomes known. But this can be difficult or even impossible to verify.

A *rational* agent's credences obey the probability axioms.

Many arguments have been offered for this thesis. Here is the most famous of them:

Thm 1.6.6 (Dutch Book Theorem). An agent's credence function $Cr : 2^{\Omega} \to \mathbb{R}[0,1]$ is a probability measure just in case there doesn't exist a set of bets, a *Dutch Book*, each of which the agent is indifferent between purchasing and selling that collectively guarantee a monetary loss.

A rational agent is not Dutch Bookable. Thus, by the Dutch Book Theorem, a rational agent has probabilistically coherent credences.

Partial Proof. Assume that an agent's credence function Cr violates Ax3 of the probability calculus; specifically, $X \wedge Y = \emptyset$ but $Cr(X) + Cr(Y) > Cr(X \vee Y)$. We show that this agent is Dutch Bookable.

A bookie can bet against X and Y but bet for $X \vee Y$:

Bet B_1	
X	\$$(1 - Cr(X))$
$\neg X$	\$$Cr(X)$

Bet B_2	
Y	\$$(1 - Cr(Y))$
$\neg Y$	\$$Cr(Y)$

Bet B_3	
$X \vee Y$	\$$(1 - Cr(X \vee Y))$
$\overline{X \vee Y}$	\$$Cr(X \vee Y)$

Then the agent's total payoffs are as follows:

$$
\begin{array}{c|c}
\neg X \wedge Y & -\$Cr(X) + \$(1 - Cr(Y)) - \$(1 - Cr(X \vee Y)) \\
X \wedge \neg Y & \$(1 - Cr(X)) - \$Cr(Y) - \$(1 - Cr(X \vee Y)) \\
\neg X \wedge \neg Y & -\$Cr(X) - \$Cr(Y) + \$Cr(X \vee Y)
\end{array}
$$

Since $Cr(X \vee Y) - Cr(X) - Cr(Y) < 0$, the agent loses money in all possible situations.