Chapter 3

Probability[1]

Department of Mathematics & Statistics
North Carolina A&T State University

[1] These notes use content from OpenIntro Statistics Slides by Mine Cetinkaya-Rundel.

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

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

$P(relapsed) = \frac{48}{72} \approx 0.67$

Joint probability

What is the probability that a patient received the antidepressant (desipramine) <u>and</u> 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) $= \frac{10}{72} \approx 0.14$

Conditonal probability

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

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

Conditonal probability

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

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

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

Conditonal probability

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

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

$$\begin{aligned} P(relapse|desipramine) &= \frac{P(relapse\ and\ desipramine)}{P(desipramine)} \\ &= \frac{10/72}{24/72} \end{aligned}$$

Conditonal probability

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

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

$$\begin{aligned} P(relapse|desipramine) &= \frac{P(relapse\ and\ desipramine)}{P(desipramine)} \\ &= \frac{10/72}{24/72} \\ &= \frac{10}{24} \end{aligned}$$

Conditonal probability

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

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

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

Conditional probability

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

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

$\mathrm{P}(\text{relapse} \mid \text{desipramine}) = \frac{10}{24} \approx 0.42$

Conditional probability

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) $= \frac{10}{24} \approx 0.42$

P(relapse | lithium) $= \frac{18}{24} \approx 0.75$

P(relapse | placebo) $= \frac{20}{24} \approx 0.83$

Conditional probability

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

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) $= \frac{10}{48} \approx 0.21$

Conditional probability

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

$\text{P(desipramine} \mid \text{relapse)} = \frac{10}{48} \approx 0.21$

$\text{P(lithium} \mid \text{relapse)} = \frac{18}{48} \approx 0.375$

$\text{P(placebo} \mid \text{relapse)} = \frac{20}{48} \approx 0.42$

General multiplication rule

- Earlier we saw that if two events are independent, their joint probability is simply the product of their probabilities. If the events are not believed to be independent, the joint probability is calculated slightly differently.

General multiplication rule

- Earlier we saw that if two events are independent, their joint probability is simply the product of their probabilities. If the events are not believed to be independent, the joint probability is calculated slightly differently.
- If A and B represent two outcomes or events, then

$$\text{P(A and B)} = \text{P(A} \mid \text{B)} \times \text{P(B)}$$

Note that this formula is simply the conditional probability formula, rearranged.

General multiplication rule

- Earlier we saw that if two events are independent, their joint probability is simply the product of their probabilities. If the events are not believed to be independent, the joint probability is calculated slightly differently.
- If A and B represent two outcomes or events, then

$$\text{P(A and B)} = \text{P(A} \mid \text{B)} \times \text{P(B)}$$

Note that this formula is simply the conditional probability formula, rearranged.

- It is useful to think of A as the outcome of interest and B as the condition.

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

	social science	non-social science	total
female	30	20	50
male	30	20	50
total	60	40	100

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

	social science	non-social science	total
female	30	20	50
male	30	20	50
total	60	40	100

- The probability that a randomly selected student is a social science major is

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

	social science	non-social science	total
female	30	20	50
male	30	20	50
total	60	40	100

- The probability that a randomly selected student is a social science major is $\frac{60}{100} = 0.6$.

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

	social science	non-social science	total
female	30	20	50
male	30	20	50
total	60	40	100

- The probability that a randomly selected student is a social science major is $\frac{60}{100} = 0.6$.
- The probability that a randomly selected student is a social science major given that they are female is

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

	social science	non-social science	total
female	30	20	50
male	30	20	50
total	60	40	100

- The probability that a randomly selected student is a social science major is $\frac{60}{100} = 0.6$.
- The probability that a randomly selected student is a social science major given that they are female is $\frac{30}{50} = 0.6$.

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

	social science	non-social science	total
female	30	20	50
male	30	20	50
total	60	40	100

- The probability that a randomly selected student is a social science major is $\frac{60}{100} = 0.6$.
- The probability that a randomly selected student is a social science major given that they are female is $\frac{30}{50} = 0.6$.
- Since P(SS | M) also equals 0.6, major of students in this class does not depend on their gender: $\text{P(SS} \mid \text{F)} = \text{P(SS)}$.

Independence and conditional probabilities

Generically, if $P(A \mid B) = P(A)$ then the events A and B are said to be independent.

Independence and conditional probabilities

Generically, if $P(A \mid B) = P(A)$ then the events A and B are said to be independent.

- Conceptually: Giving B doesn't tell us anything about A.

Independence and conditional probabilities

Generically, if P(A | B) = P(A) then the events A and B are said to be independent.

- Conceptually: Giving B doesn't tell us anything about A.
- Mathematically: We know that if events A and B are independent, P(A and B) = P(A) $\times$ P(B). Then,

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

Bayes' Theorem

- We can get the following by combining the conditional probability formula and the general multiplication rule.

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)}$$

Bayes' Theorem

- We can get the following by combining the conditional probability formula and the general multiplication rule.

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(B|A_1)P(A_1)}{P(B)}$$

Bayes' Theorem

- We can get the following by combining the conditional probability formula and the general multiplication rule.

$$P(A_1|B) \quad = \quad \frac{P(A_1 \cap B)}{P(B)} = \frac{P(B|A_1)P(A_1)}{P(B)}$$

- If events A_1 and A_2 partition a sample space S into mutually exclusive and exhaustive nonempty events, then the **total probability** of an event B can be written as:

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B)$$

Bayes' Theorem

$$
\begin{aligned}
P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B)} \\
&= \frac{P(B|A_1)P(A_1)}{P(A_1 \cap B) + P(A_2 \cap B)}
\end{aligned}
$$

Bayes' Theorem

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B)} \\ &= \frac{P(B|A_1)P(A_1)}{P(A_1 \cap B) + P(A_2 \cap B)} \\ &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} \end{aligned}$$

Bayes' Theorem

- The conditional probability formula we have seen so far is a special case of the Bayes' Theorem, which is applicable even when events have more than just two outcomes.

Bayes' Theorem

- The conditional probability formula we have seen so far is a special case of the Bayes' Theorem, which is applicable even when events have more than just two outcomes.
- **Bayes' Theorem**

$$P(outcome\ A_1\ of\ variable\ 1 \mid outcome\ B\ of\ variable\ 2)$$
$$= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1)+P(B|A_2)P(A_2)+\ldots+P(B|A_k)P(A_k)}$$

where $A_2, \ldots, A_k$ represent all other possible outcomes of variable 1.

Practice

Joe visits campus every Thursday evening. However, some days the parking garage is full, often due to college events. There are academic events on 35% of evenings, sporting events on 20% of evenings, and no events on 45% of evenings. When there is an academic event, the garage fills up about 25% of the time, and it fills up 70% of evenings with sporting events. On evenings when there are no events, it only fills up about 5% of the time. If Jose comes to campus and finds the garage full, what is the probability that there is a sporting event?

Practice

Joe visits campus every Thursday evening. However, some days the parking garage is full, often due to college events. There are academic events on 35% of evenings, sporting events on 20% of evenings, and no events on 45% of evenings. When there is an academic event, the garage fills up about 25% of the time, and it fills up 70% of evenings with sporting events. On evenings when there are no events, it only fills up about 5% of the time. If Jose comes to campus and finds the garage full, what is the probability that there is a sporting event?

The outcome of interest is whether there is a sporting event (call this A_1), and the condition is that the lot is full (B). Let A_2 represent an academic event and A_3 represent there being no event on campus. Then the given probabilities can be written as:

$$\begin{array}{rrr} P(A_1) = 0.2 & P(A_2) = 0.35 & P(A_3) = 0.45 \\ P(B|A_1) = 0.7 & P(B|A_2) = 0.25 & P(B|A_3) = 0.05 \end{array}$$

Practice

$$P(A_1) = 0.2 \qquad P(A_2) = 0.35 \qquad P(A_3) = 0.45$$
$$P(B|A_1) = 0.7 \qquad P(B|A_2) = 0.25 \qquad P(B|A_3) = 0.05$$

Bayes' Theorem can be used to compute the probability of a sporting event (A_1) under the condition that the parking lot is full (B):

Practice

$$\begin{aligned} P(A_1) &= 0.2 & P(A_2) &= 0.35 & P(A_3) &= 0.45 \\ P(B|A_1) &= 0.7 & P(B|A_2) &= 0.25 & P(B|A_3) &= 0.05 \end{aligned}$$

Bayes' Theorem can be used to compute the probability of a sporting event (A_1) under the condition that the parking lot is full (B):

$$P(A_1|B) \quad =$$

Practice

$$P(A_1) = 0.2 \qquad P(A_2) = 0.35 \qquad P(A_3) = 0.45$$
$$P(B|A_1) = 0.7 \qquad P(B|A_2) = 0.25 \qquad P(B|A_3) = 0.05$$

Bayes' Theorem can be used to compute the probability of a sporting event (A_1) under the condition that the parking lot is full (B):

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)}$$

Practice

$$P(A_1) = 0.2 \qquad P(A_2) = 0.35 \qquad P(A_3) = 0.45$$
$$P(B|A_1) = 0.7 \qquad P(B|A_2) = 0.25 \qquad P(B|A_3) = 0.05$$

Bayes' Theorem can be used to compute the probability of a sporting event (A_1) under the condition that the parking lot is full (B):

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\ &= \frac{(0.7)(0.2)}{(0.7)(0.2) + (0.25)(0.35) + (0.05)(0.45)} \end{aligned}$$

Practice

$$P(A_1) = 0.2 \qquad P(A_2) = 0.35 \qquad P(A_3) = 0.45$$
$$P(B|A_1) = 0.7 \qquad P(B|A_2) = 0.25 \qquad P(B|A_3) = 0.05$$

Bayes' Theorem can be used to compute the probability of a sporting event (A_1) under the condition that the parking lot is full (B):

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\ &= \frac{(0.7)(0.2)}{(0.7)(0.2) + (0.25)(0.35) + (0.05)(0.45)} \\ &= 0.56 \end{aligned}$$

Practice

$$
\begin{array}{rrr}
P(A_1) = 0.2 & P(A_2) = 0.35 & P(A_3) = 0.45 \\
P(B|A_1) = 0.7 & P(B|A_2) = 0.25 & P(B|A_3) = 0.05
\end{array}
$$

Bayes' Theorem can be used to compute the probability of a sporting event (A_1) under the condition that the parking lot is full (B):

$$
\begin{aligned}
P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\
&= \frac{(0.7)(0.2)}{(0.7)(0.2) + (0.25)(0.35) + (0.05)(0.45)} \\
&= 0.56
\end{aligned}
$$

Based on the information that the garage is full, there is a **56%** chance that a sporting event is being held on campus that evening.