BIOS 660/BIOS 672 (3 Credits):
Probability and Statistical Inference I

Jianwen Cai

https://sakai.unc.edu/portal/site/bios660-bios672-3-credits

Notes 5

Conditional Probability

If $P(B) > 0$ we define the *conditional probability* of the event A given B as

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

Intuitively, conditioning on B means reducing the original sample space S to B, which becomes the new, reduced sample space. All probabilities are computed with respect to B. Notice:

$$P(A|\Omega) = \frac{P(A \cap \Omega)}{P(\Omega)} = P(A)$$

Disjoint events: If $A \cap B = \emptyset$, then $P(A|B) = 0$ and $P(B|A) = 0$.

cont.

Conditional probability satisfies the axioms of probability:

1. $P(\Omega|B) = 1$
2. $P(A|B) \geq 0$
3. If $A_1, A_2, \ldots$ are mutually exclusive events, then $P\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) = \sum_{i=1}^{\infty} P(A_i|B)$

and all the other properties:

1. $P(\emptyset|B) = 0$
2. $P(A|B) \leq 1$
3. $P(A^c|B) = 1 - P(A|B)$

etc.

Independence

Two events A and B are said to be *independent* if

$$P(A \cap B) = P(A)P(B). \tag{1}$$

Why? Because then

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

This could have been taken as the definition, but (1) is easier to generalize.
If A and B are independent, then so are

- A^c and B
- A and B^c
- A^c and B^c

* Can two disjoint events be independent and vice versa? (HW)

Independence of many events

The events $A_1, A_2, \ldots, A_n$ are mutually independent if for *every* subcollection $A_{i_1}, \ldots, A_{i_k}$ of size $k = 2, \ldots, n$

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

Notice that is a very strong condition. But it is necessary to ensure that

$$P(A_j | A_{i_1}, \ldots, A_{i_k}) = P(A_j)$$

for every j and every subcollection $A_{i_1}, \ldots, A_{i_k}$ that does not include A_j. Pairwise independence is not enough.

Independence of many events (cont.)

Example: Toss a coin three times. Sample space Ω={HHH, HHT, HTH HTT, THH, THT, TTH, TTT}. Define the events:

- H_1 = { The outcome of the first toss is heads }
  ={HHH, HHT, HTH, HTT}
- H_2 = { The outcome of the second toss is heads }
  ={HHH, HHT, THH, THT}
- H_3 = { The outcome of the third toss is heads }
  ={HHH, HTH, THH, TTH}

Suppose every outcome is equally likely. Then these events are independent.
$H_1 \cap H_2 = \{HHH, HHT\}$; $H_1 \cap H_3 = \{HHH, HTH\}$; $H_2 \cap H_3 = \{HHH, THH\}$;
$P(H_1) = P(H_2) = P(H_3) = 4/8 = 1/2$; $P(H_1 \cap H_2) = 2/8 = 1/4 = P(H_1)P(H_2)$;
$H_1 \cap H_2 \cap H_3 = \{HHH\}$; $P(H_1 \cap H_2 \cap H_3) = 1/8 = P(H_1)P(H_2)P(H_3)$)

Independence of many events (cont.)

Now define the events:

- A_{12} = { The outcome of the first toss equals the second }
- A_{13} = { The outcome of the first toss equals the third }
- A_{23} = { The outcome of the second toss equals the third }

These events are pairwise independent but not mutually independent.

(A_{12}= { HHH, HHT, TTH, TTT }; A_{13}= {HHH, HTH, THT, TTT }; A_{23}={HHH, HTT, THH, TTT};
$P(A_{12}) = P(A_{13}) = P(A_{23}) = 4/8 = 1/2$;
$A_{12} \cap A_{13}$= {HHH, TTT };
$P(A_{12} \cap A_{13}) = 2/8 = 1/4$ =$P(A_{12})P(A_{13})$.

On the other hand,
$A_{12} \cap A_{13} \cap A_{23}$={HHH, TTT};
$P(A_{12} \cap A_{13} \cap A_{23}) = 2/8 = 1/4 \neq P(A_{12})P(A_{13})P(A_{23})$.)

Sequential conditioning

By the definition of conditional probability:

$$P(A \cap B) = P(A)P(B|A)$$
$$P(A \cap B) = P(B)P(A|B)$$

This is useful for computing probabilities of sequential events.
E.g: What is the probability of dealing two aces in a row?

More generally:

$$P(A_1 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\ldots P(A_n|A_1\ldots A_{n-1})$$

The Birthday Problem

In a group of n students in a class, what is the probability that at least two have the same birthday?

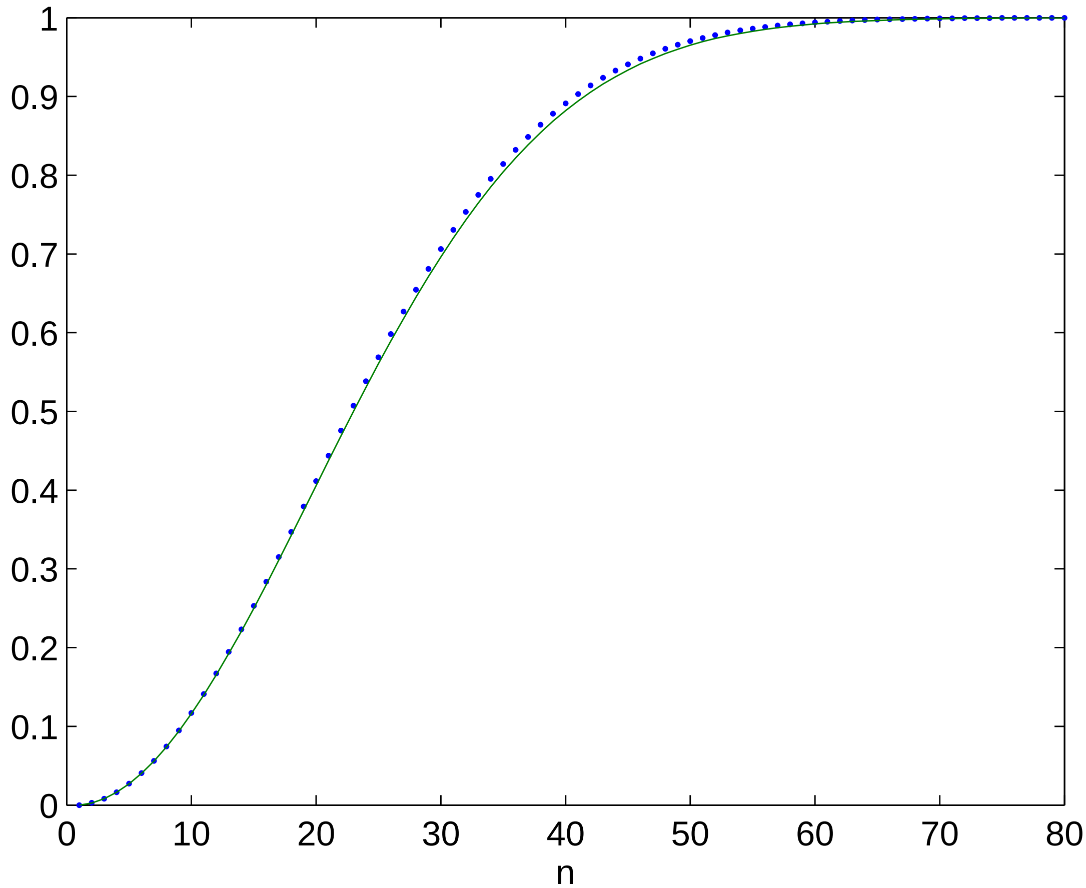
Solution:
Suppose we order the n students in an arbitrary order. Let D_j be the event that the first j have different birthdays. Based on page 22 in Notes 3, we have

$$P(D_j) = \frac{\text{No. of Samples with No Repetition}}{\text{No. of Samples}} = \frac{365!/(365-j)!}{365^j}$$

Let $D_j = \{A_1, A_2, A_3, \ldots, A_j\}$, where A_1 is the birth day of the first person, A_2 is the birth day of the second person, etc., and all the A_i ($i = 1, 2, \ldots, j$) are different. Based on sequential conditioning,

$$\begin{aligned} & P(D_j) = P(\{A_1, A_2, A_3, \ldots, A_j\}) \\ = \; & P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\ldots P(A_j|A_1\ldots A_{j-1}) \\ = \; & \frac{365}{365}\frac{365-1}{365}\cdots\frac{365-j+1}{365} = \frac{365!/(365-j)!}{365^j} \end{aligned}$$

The Birthday Problem



Turning around probabilities

Also by the definition of conditional probability:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2}$$

This is useful for computing conditional probabilities when the reverse conditioning is easier to compute.
E.g.: Prob. that it will rain given that it is thundering vs. prob. that it thundered given that it is raining.

(2) is called sometimes **Bayes' rule**. It is often used in a context where we want to know the probability that a particular hypothesis is true. We have an *a priori* belief in whether or not the hypothesis is true, then update that probability by collecting data.
E.g.: Suppose a priori boys are equally likely to be born as girls. Say 90% of boys play with trucks. Baby X plays with trucks. What is the probability that Baby X is a boy?

Decomposition Formula (Total Probability)

Let $\{A_1, A_2, \ldots\}$ be a partition of Ω. Let B be any subset in Ω. Then

$$P(B) = \sum_{i=1}^{\infty} P(A_i)P(B|A_i)$$

Proof:

(So called) Bayes' Theorem

Let $\{A_1, A_2, \ldots\}$ be a partition of Ω. Let B be any subset in Ω. Then

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{\infty} P(B|A_i)P(A_i)}$$

Proof:

Bayes and Screening

An important application of Bayes' theorem is screening.
Notation:

Let D be the disease:

D means diseased
$\overline{D}$ means "no disease"

and T be the diagnostic test:

T^+ means a positive test
T^- means a negative test.

Then we have that the **positive predictive value**

$$
\begin{aligned}
P(D|T^+) &= \frac{P(D)\,P(T^+|D)}{P(D)\,P(T^+|D) + P(\overline{D})\,P(T^+|\overline{D})} \\
&\equiv \frac{\text{prevalence} \times \text{sensitivity}}{\text{prev.} \times \text{sens.} + (1 - \text{prev.}) \times (1 - \text{specificity})}
\end{aligned}
$$

Papanicolaou Example

Let D be cervical cancer.
$P(D)$ we'll take to be 1 in 21,000, which is the approximate annual incidence rate in the US (SEER 2002 estimate).

$$P(D) = .00004762$$

Let us take the sensitivity ($P(T^+|D)$) to be 0.71 and the specificity ($1 - P(T^+|\overline{D})$) to be 0.75.
Thence the positive predictive value is:

$$
\begin{aligned}
P(D|T^+) &= \frac{0.00004762 \times 0.71}{0.00004762 \times 0.71 + (1 - 0.00004762) \times (1 - 0.75)} \\
&= 0.000135
\end{aligned}
$$

That means that for every 1,000,000 positive results, about 135 truly have cervical cancer. But for any particular patient, testing positive increases the probability of having the disease by a factor of 2.8!

Relative risks and relative odds

Given two conditions–smokers (S) and non-smokers ($\overline{S}$), say–then we say the *relative risk* of a disease (D)–lung cancer, say–due to smoking is:

$$RR = \frac{P(D|S)}{P(D|\overline{S})}$$

The *relative odds* of the disease (D) due to smoking is:

$$OR = \frac{\frac{P(D|S)}{1-P(D|S)}}{\frac{P(D|\overline{S})}{1-P(D|\overline{S})}}$$

Of course, if $P(D|S) \approx 0$ and $P(D|\overline{S}) \approx 0$, (*rare disease*) then

$$OR \approx \frac{\frac{P(D|S)}{1}}{\frac{P(D|\overline{S})}{1}} = RR$$

Bayes and Case Control Studies

Consider

$$OR(D|S) \quad \equiv \quad \frac{P(D|S)}{1-P(D|S)} \Big/ \frac{P(D|\overline{S})}{1-P(D|\overline{S})}$$

Now consider the numerator, which from Bayes theorem,

$$\frac{\frac{P(D)P(S|D)}{P(S)}}{\frac{P(\overline{D})P(S|\overline{D})}{P(S)}} \quad = \quad \frac{P(D)P(S|D)}{P(\overline{D})P(S|\overline{D})}$$

Do the same for the denominator, and get,

$$OR(D|S) \quad = \quad \frac{\frac{P(D)P(S|D)}{P(\overline{D})P(S|\overline{D})}}{\frac{P(D)P(\overline{S}|D)}{P(\overline{D})P(\overline{S}|\overline{D})}} = \frac{\frac{P(S|D)}{P(\overline{S}|D)}}{\frac{P(S|\overline{D})}{P(\overline{S}|\overline{D})}} = OR(S|D)$$

Additional Reading

See Chapter 1.2-1.3 in Casella and Berger.