Applied Bayesian Modeling module 2: **Bayesian inference for 1 discrete parameter**

Leontine Alkema, lalkema@umass.edu
Fall 2022

Applied Bayesian Modeling module 2:
Bayesian inference for 1 discrete parameter

Leontine Alkema, lalkema@umass.edu

Fall 2022

Recap from module 1: introduction to Bayesian inference

- In Bayesian inference, parameters are considered random variables
- We draw statistical conclusions about parameters of interest using probability statements.
- General approach for some outcome of interest θ, i.e. regression coefficient, is based on learning via Bayes' rule
  - start off with *prior* probability distribution to quantify information related to θ
  - collect data, and use Bayes' rule to update the prior into the posterior distribution
  - use posterior distribution to draw conclusions

Module 2: Bayesian inference for 1 discrete parameter

- Learning objectives
  - Quick review of basic probability theory (probability functions, conditional probabilities, Bayes' rule) for discrete outcome of interest
  - Worked example and HW exercise: apply Bayes' rule to calculate a conditional probability
- Note: basic prob theory is a pre-req for this course, slide set includes a quick review only, see readings on course website if you need more info.

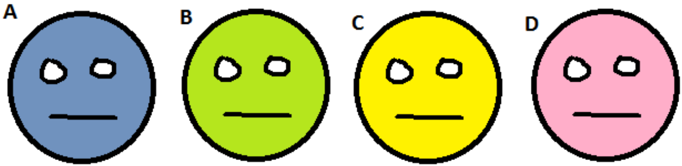
HW exercise: Breast cancer and mammogram screening

- For early detection of breast cancer, women are encouraged to have routine screening, even if they have no symptoms.
- The following information is available about asymptomatic women aged 40 to 50 who have mammography screening:
  - The probability an asymptomatic woman has breast cancer is 0.8%.
  - If she has breast cancer, the probability is 90% that she has a positive mammogram.
  - If she does not have breast cancer, the probability is 7% that she still has a positive mammogram.
- Suppose a woman has a positive mammogram: What is the probability she actually has breast cancer?
- Physicians' answers ranged from about 1% to about 90%.
  What do you think this probability is?

Crime investigation

Simple example to illustrate prob. rules, adapted from Kruschke

- You are investigating a crime
- You have identified 4 suspects, labeled A, B, C and D.
  You are 100% sure that the offender is A or B or C or D.
- You assume that each of them is equally likely to have committed the crime.
- The plan: carry out a Bayesian crime investigation;
  quantify "information" on who committed the crime using
  probability statements



How to assign probabilities?

Quick recap of basic probability theory for discrete random variables

- Suppose quantity of interest is a discrete random variable with a countable number of possible outcomes, i.e.
  - outcome of interest = who committed the crime?
  - possible outcomes (sample space): suspect A, B, C or D
- A probability function $Pr(\cdot)$ is a mathematical function with
  - input = possibility, chosen from a sample space;
  - output = a number (probability value) assigned to the possibility, which represents the probability that the possibility is correct/is true/has occured/will occur;
  - Example crime investigation: $Pr(A)$ = probability that suspect A commited the crime
- For a function to be a probability function, it has to satisfy three properties (simple summary Kruschke p.77, Kolmogorov, 1956):
  1. $Pr(\cdot) \geq 0$; a probability value must be nonnegative;
  2. For sample space S, $Pr(S) = 1$;
  3. For any two mutually exclusive events A and B, $Pr(A \cup B) = Pr(A) + Pr(B)$ (where $A \cup B$ means A OR B).

Bayesian crime investigation

- Outcome of interest: who committed the crime?
- Prior information given:
  Possibilities: A (committed the crime), B, C or D, with each possibility equally likely
- A probability function that reflects this state of prior knowledge:
  - Input: Sample space of possibilities $\{A, B, C, D\}$
  - Output: Probability function assigns a probability to each possibility, $Pr(A) = Pr(B) = Pr(C) = Pr(D) = 0.25$.
- Suppose that we find evidence that C is not guilty. What are the updated probabilities?
  - Even w/o formal probability rules, you probably agree that conditional on knowing that C is not guilty, the updated probabilities are: $Pr(A|notC) = Pr(B|notC) = Pr(D|notC) = 1/3$
  - Let's make this a bit more precise...

Conditional probabilities

- For events A and B, $Pr(A|B)$ is called a conditional probability, its definition is as follows (for an event B with $Pr(B) > 0$):

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

  where $Pr(A \cap B)$ refers to the probability of A AND B being true.
- Hence, in the crime example

$$Pr(A|notC) = \frac{Pr(A \cap notC)}{Pr(notC)} = \frac{1/4}{3/4} = 1/3.$$

- The probabilities $Pr(A \cap notC)$ and $Pr(notC)$ may be easy (or not?) to write down directly here. If they are not easy, additional probability rules come to the rescue.

Probability rules :)

- We defined a conditional probability $Pr(B|A)$ (with $Pr(B) > 0$) as follows:
$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}.$$
- From this definition, we get $Pr(A \cap B) = Pr(A|B)Pr(B)$
- Similarly, if $Pr(A) > 0$, we get $Pr(B|A) = \frac{Pr(B \cap A)}{Pr(A)}$.
- With the info given, we can derive Bayes' rule for $Pr(A), Pr(B) > 0$:

$$\begin{aligned} Pr(A|B) &= \frac{Pr(A \cap B)}{Pr(B)}, \\ &= \frac{Pr(B|A)Pr(A)}{Pr(B)}. \end{aligned}$$

Bayesian crime investigation (ctd)

- We can now write $Pr(A|notC)$ in terms of $Pr(notC|A)$:

$$
\begin{aligned}
Pr(A|notC) &= \frac{Pr(A \cap notC)}{Pr(notC)} \\
&= \frac{Pr(notC|A)Pr(A)}{Pr(notC)} \\
&= \frac{1 \cdot 1/4}{3/4} = 1/3.
\end{aligned}
$$

- What if $Pr(notC)$ is not that easy (i.e., in mammogram exercise example)?

5

Partitioning the sample space (Hoff Ch.2)

- Suppose that the set of possibilities $\{H_1, H_2, \ldots, H_K\}$ is a partion of the sample space, with
  - $\sum_{k=1}^{K} Pr(H_k) = 1$
  - $Pr(H_j \cap H_j) = 0$ for any $i \neq j$.
- Then for some specific possibility/event E (rule of marginal probability):

$$
Pr(E) = \sum_{k=1}^{K} Pr(E \cap H_k) = \sum_{k=1}^{K} Pr(E|H_k)Pr(H_k).
$$

- Applied to crime example:

$$
\begin{aligned}
Pr(notC) &= Pr(notC|A)Pr(A) + Pr(notC|B)Pr(B) + \\
&\quad Pr(notC|C)Pr(C) + Pr(notC|D)Pr(D) \\
&= 1 \cdot 1/4 + 1 \cdot 1/4 + 0 \cdot 1/4 + 1 \cdot 1/4 = 3/4.
\end{aligned}
$$

Recap module 2: Bayesian inference for 1 discrete parameter

- Bayesian inference $\rightarrow$ to draw statistical conclusions about parameters of interest using probability statements.
- General approach for some outcome of interest θ, i.e. who committed a crime?
  - start off with *prior* probability distributions to quantify information related to θ: $Pr(A) = Pr(B) = Pr(C) = Pr(D) = 1/4$
  - collect data, and use Bayes' rule to update the prior into the posterior distribution: $Pr(A|notC) = Pr(notC|a)P(A)/P(notC) = 1/3$
  - use posterior distribution to draw conclusions
- Outlook: Bayesian inference for a continuous parameter