Probability and Inference

Dr. Jarad Niemi

STAT 544 - Iowa State University

January 17, 2019

Outline

- Quick review of probability
    - Kolmogorov's axioms
    - Bayes' Rule
    - Application to Down's syndrome screening
- Bayesian statistics
    - Condition on what is known
    - Describe uncertainty using probability
    - Exponential example
- What is probability?
    - Frequency interpretation
    - Personal belief
- Why or why not Bayesian?

Events

Definition

The set, Ω, of all possible outcomes of a particular experiment is called the sample space for the experiment.

Events

Definition

The set, Ω, of all possible outcomes of a particular experiment is called the sample space for the experiment.

Definition

An event is any collection of possible outcomes of an experiment, that is, any subset of Ω (including Ω itself).

Craps

Craps:

Craps

Craps:

- $\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),(2,2),\ldots,(6,6)\}$

Craps

Craps:

- $\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),(2,2),\ldots,(6,6)\}$
- Come-out roll win: the sum of the dice is 7 or 11

Craps

Craps:

- $\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),(2,2),\ldots,(6,6)\}$
- Come-out roll win: the sum of the dice is 7 or 11
- Come-out roll loss: the sum of the dice is 2, 3, or 12

Craps

Craps:

- $\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),(2,2),\ldots,(6,6)\}$
- Come-out roll win: the sum of the dice is 7 or 11
- Come-out roll loss: the sum of the dice is 2, 3, or 12
- Come-out roll establishes a point: the sum of the dice is 4, 5, 6, 8, 9, or 10

Craps

Craps:

- $\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),(2,2),\ldots,(6,6)\}$
- Come-out roll win: the sum of the dice is 7 or 11
- Come-out roll loss: the sum of the dice is 2, 3, or 12
- Come-out roll establishes a point: the sum of the dice is 4, 5, 6, 8, 9, or 10
- Events:

Craps

Craps:

- $\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),(2,2),\ldots,(6,6)\}$
- Come-out roll win: the sum of the dice is 7 or 11
- Come-out roll loss: the sum of the dice is 2, 3, or 12
- Come-out roll establishes a point: the sum of the dice is 4, 5, 6, 8, 9, or 10
- Events:
  - the come-out roll wins

Craps

Craps:

- $\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),(2,2),\ldots,(6,6)\}$
- Come-out roll win: the sum of the dice is 7 or 11
- Come-out roll loss: the sum of the dice is 2, 3, or 12
- Come-out roll establishes a point: the sum of the dice is 4, 5, 6, 8, 9, or 10
- Events:
  - the come-out roll wins
  - the come-out roll loses

Craps

Craps:

- $\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),(2,2),\ldots,(6,6)\}$
- Come-out roll win: the sum of the dice is 7 or 11
- Come-out roll loss: the sum of the dice is 2, 3, or 12
- Come-out roll establishes a point: the sum of the dice is 4, 5, 6, 8, 9, or 10
- Events:
  - the come-out roll wins
  - the come-out roll loses
  - the come-out roll establishes a point

Pairwise disjoint

Definition

Two events A_1 and A_2 are disjoint (or mutually exclusive) if both A_1 and A_2 cannot occur simultaneously, i.e. $A_i \cap A_j = \emptyset$.

Pairwise disjoint

Definition

Two events A_1 and A_2 are disjoint (or mutually exclusive) if both A_1 and A_2 cannot occur simultaneously, i.e. $A_i \cap A_j = \emptyset$. The events $A_1, A_2, \ldots$ are pairwise disjoint (or mutually exclusive) if A_i and A_j cannot occur simultaneously for all $i \neq j$, i.e. $A_1 \cap A_2 = \emptyset$.

Pairwise disjoint

Definition

Two events A_1 and A_2 are disjoint (or mutually exclusive) if both A_1 and A_2 cannot occur simultaneously, i.e. $A_i \cap A_j = \emptyset$. The events $A_1, A_2, \ldots$ are pairwise disjoint (or mutually exclusive) if A_i and A_j cannot occur simultaneously for all $i \neq j$, i.e. $A_1 \cap A_2 = \emptyset$.

Craps pairwise disjoint examples:

- Win (A_1), Loss (A_2)

Pairwise disjoint

Definition

Two events A_1 and A_2 are disjoint (or mutually exclusive) if both A_1 and A_2 cannot occur simultaneously, i.e. $A_i \cap A_j = \emptyset$. The events $A_1, A_2, \ldots$ are pairwise disjoint (or mutually exclusive) if A_i and A_j cannot occur simultaneously for all $i \neq j$, i.e. $A_1 \cap A_2 = \emptyset$.

Craps pairwise disjoint examples:

- Win (A_1), Loss (A_2)
- Win (A_1), Loss (A_2), Point (A_3)

Pairwise disjoint

Definition

Two events A_1 and A_2 are disjoint (or mutually exclusive) if both A_1 and A_2 cannot occur simultaneously, i.e. $A_i \cap A_j = \emptyset$. The events $A_1, A_2, \ldots$ are pairwise disjoint (or mutually exclusive) if A_i and A_j cannot occur simultaneously for all $i \neq j$, i.e. $A_1 \cap A_2 = \emptyset$.

Craps pairwise disjoint examples:

- Win (A_1), Loss (A_2)
- Win (A_1), Loss (A_2), Point (A_3)
- $A_1 = (1,1), A_2 = (1,2), \ldots, A_6 = (1,6), A_7 = (2,1), \ldots, A_{12} = (2,6), \ldots, A_{36} = (6,6)$

Kolmogorov's axioms of probability

Definition

Given a sample space Ω and event space E, a probability is a function $P : E \to \mathbb{R}$ that satisfies

Kolmogorov's axioms of probability

Definition

Given a sample space Ω and event space E, a probability is a function $P : E \to \mathbb{R}$ that satisfies

1. $P(A) \geq 0$ for any $A \in E$

Kolmogorov's axioms of probability

Definition

Given a sample space Ω and event space E, a probability is a function $P : E \to \mathbb{R}$ that satisfies

1. $P(A) \geq 0$ for any $A \in E$
2. $P(\Omega) = 1$

Kolmogorov's axioms of probability

Definition

Given a sample space Ω and event space E, a probability is a function $P : E \to \mathbb{R}$ that satisfies

1. $P(A) \geq 0$ for any $A \in E$
2. $P(\Omega) = 1$
3. If $A_1, A_2, \ldots \in E$ are pairwise disjoint, then $P(A_1 \text{ or } A_2 \text{ or } \ldots) = \sum_{i=1}^{\infty} P(A_i)$.

Craps come-out roll probabilities

The following table provides the probability mass function for the sum of the two dice if we *believe* the probability of each elementary outcome is equal:

Craps come-out roll probabilities

The following table provides the probability mass function for the sum of the two dice if we *believe* the probability of each elementary outcome is equal:

Outcome	2	3	4	5	6	7	8	9	10	11	12	Sum
Combinations	1	2	3	4	5	6	5	4	3	2	1	36

Craps come-out roll probabilities

The following table provides the probability mass function for the sum of the two dice if we *believe* the probability of each elementary outcome is equal:

Outcome	2	3	4	5	6	7	8	9	10	11	12	Sum
Combinations	1	2	3	4	5	6	5	4	3	2	1	36
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

Craps come-out roll probabilities

The following table provides the probability mass function for the sum of the two dice if we *believe* the probability of each elementary outcome is equal:

Outcome	2	3	4	5	6	7	8	9	10	11	12	Sum
Combinations	1	2	3	4	5	6	5	4	3	2	1	36
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

Craps probability examples:

- P(Win) = P(7 or 11) $= 8/36 = 2/9$

Craps come-out roll probabilities

The following table provides the probability mass function for the sum of the two dice if we *believe* the probability of each elementary outcome is equal:

Outcome	2	3	4	5	6	7	8	9	10	11	12	Sum
Combinations	1	2	3	4	5	6	5	4	3	2	1	36
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

Craps probability examples:

- P(Win) $=$ P(7 or 11) $= 8/36 = 2/9$
- P(Loss) $=$ P(2, 3, or 12) $= 4/36 = 1/9$

Craps come-out roll probabilities

The following table provides the probability mass function for the sum of the two dice if we *believe* the probability of each elementary outcome is equal:

Outcome	2	3	4	5	6	7	8	9	10	11	12	Sum
Combinations	1	2	3	4	5	6	5	4	3	2	1	36
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

Craps probability examples:

- P(Win) $=$ P(7 or 11) $= 8/36 = 2/9$
- P(Loss) $=$ P(2, 3, or 12) $= 4/36 = 1/9$
- P(Point) $=$ P(4, 5, 6, 8, 9 or 10) $= 6/9$

Partition

Definition

A set of events, $\{A_1, A_2, \ldots\}$, is a partition of the sample space Ω

Partition

Definition

A set of events, $\{A_1, A_2, \ldots\}$, is a partition of the sample space Ω if and only if

- the events in $\{A_1, A_2, \ldots\}$ are pairwise disjoint and

Partition

Definition

A set of events, $\{A_1, A_2, \ldots\}$, is a partition of the sample space Ω if and only if

- the events in $\{A_1, A_2, \ldots\}$ are pairwise disjoint and
- $\cup_{i=1}^{\infty} A_i = \Omega$.

Partition

Definition

A set of events, $\{A_1, A_2, \ldots\}$, is a partition of the sample space Ω if and only if

- the events in $\{A_1, A_2, \ldots\}$ are pairwise disjoint and
- $\cup_{i=1}^{\infty} A_i = \Omega$.

Partition

Definition

A set of events, $\{A_1, A_2, \ldots\}$, is a partition of the sample space Ω if and only if

- the events in $\{A_1, A_2, \ldots\}$ are pairwise disjoint and
- $\cup_{i=1}^{\infty} A_i = \Omega$.

Craps partition examples:

- Win (A_1), Loss (A_2), Point (A_3)

Partition

Definition

A set of events, $\{A_1, A_2, \ldots\}$, is a partition of the sample space Ω if and only if

- the events in $\{A_1, A_2, \ldots\}$ are pairwise disjoint and
- $\cup_{i=1}^{\infty} A_i = \Omega$.

Craps partition examples:

- Win (A_1), Loss (A_2), Point (A_3)
- $A_1 = (1,1), A_2 = (1,2), \ldots, A_6 = (1,6), A_7 = (2,1), \ldots, A_{12} = (2,6), \ldots, A_{36} = (6,6)$

Conditional probability

Definition

If A and B are events in E, and $P(B) > 0$,

Conditional probability

Definition

If A and B are events in E, and $P(B) > 0$, then the conditional probability of A given B, written $P(A|B)$,

Conditional probability

Definition

If A and B are events in E, and $P(B) > 0$, then the conditional probability of A given B, written $P(A|B)$, is

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

Conditional probability

Definition

If A and B are events in E, and $P(B) > 0$, then the conditional probability of A given B, written $P(A|B)$, is

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

Example (Craps conditional probability)

$$P(7|\text{Win})$$

Conditional probability

Definition

If A and B are events in E, and $P(B) > 0$, then the conditional probability of A given B, written $P(A|B)$, is

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

Example (Craps conditional probability)

$$P(7|\text{Win}) = \frac{P(7 \text{ and Win})}{P(\text{Win})}$$

Conditional probability

Definition

If A and B are events in E, and $P(B) > 0$, then the conditional probability of A given B, written $P(A|B)$, is

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

Example (Craps conditional probability)

$$P(7|\text{Win}) = \frac{P(7 \text{ and Win})}{P(\text{Win})} = \frac{P(7)}{P(\text{Win})}$$

Conditional probability

Definition

If A and B are events in E, and $P(B) > 0$, then the conditional probability of A given B, written $P(A|B)$, is

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

Example (Craps conditional probability)

$$P(7|\text{Win}) = \frac{P(7 \text{ and Win})}{P(\text{Win})} = \frac{P(7)}{P(\text{Win})} = \frac{6/36}{8/36}$$

Conditional probability

Definition

If A and B are events in E, and $P(B) > 0$, then the conditional probability of A given B, written $P(A|B)$, is

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

Example (Craps conditional probability)

$$P(7|\text{Win}) = \frac{P(7 \text{ and Win})}{P(\text{Win})} = \frac{P(7)}{P(\text{Win})} = \frac{6/36}{8/36} = \frac{6}{8}$$

Law of Total Probability

Corollary (Law of Total Probability)

Let $A_1, A_2, \ldots$ be a partition of Ω and B is another event in Ω.

Law of Total Probability

Corollary (Law of Total Probability)

Let $A_1, A_2, \ldots$ be a partition of Ω and B is another event in Ω. The Law of Total Probability states that

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

Law of Total Probability

Corollary (Law of Total Probability)

Let $A_1, A_2, \ldots$ be a partition of Ω and B is another event in Ω. The Law of Total Probability states that

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

Law of Total Probability

Corollary (Law of Total Probability)

Let $A_1, A_2, \ldots$ be a partition of Ω and B is another event in Ω. The Law of Total Probability states that

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

Example (Craps Win Probability)

Let A_i be the event that the sum of two die rolls is i.

Law of Total Probability

Corollary (Law of Total Probability)

Let $A_1, A_2, \ldots$ be a partition of Ω and B is another event in Ω. The Law of Total Probability states that

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

Example (Craps Win Probability)

Let A_i be the event that the sum of two die rolls is i. Then

$$P(\text{Win})$$

Law of Total Probability

Corollary (Law of Total Probability)

Let $A_1, A_2, \ldots$ be a partition of Ω and B is another event in Ω. The Law of Total Probability states that

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

Example (Craps Win Probability)

Let A_i be the event that the sum of two die rolls is i. Then

$$P(\text{Win}) = \sum_{i=2}^{12} P(\text{Win and } A_i)$$

Law of Total Probability

Corollary (Law of Total Probability)

Let $A_1, A_2, \ldots$ be a partition of Ω and B is another event in Ω. The Law of Total Probability states that

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

Example (Craps Win Probability)

Let A_i be the event that the sum of two die rolls is i. Then

$$P(\text{Win}) = \sum_{i=2}^{12} P(\text{Win and } A_i) = P(7) + P(11)$$

Law of Total Probability

Corollary (Law of Total Probability)

Let $A_1, A_2, \ldots$ be a partition of Ω and B is another event in Ω. The Law of Total Probability states that

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

Example (Craps Win Probability)

Let A_i be the event that the sum of two die rolls is i. Then

$$P(\text{Win}) = \sum_{i=2}^{12} P(\text{Win and } A_i) = P(7) + P(11) = \frac{6}{36} + \frac{2}{36} = \frac{8}{36} = \frac{2}{9}.$$

Bayes' Rule

Theorem (Bayes' Rule)

If A and B are events in E with $P(B) > 0$,

Bayes' Rule

Theorem (Bayes' Rule)

If A and B are events in E with $P(B) > 0$, then Bayes' Rule states

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

Bayes' Rule

Theorem (Bayes' Rule)

If A and B are events in E with $P(B) > 0$, then Bayes' Rule states

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

Example (Craps Bayes' Rule)

$$P(7|\text{Win})$$

Bayes' Rule

Theorem (Bayes' Rule)

If A and B are events in E with $P(B) > 0$, then Bayes' Rule states

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

Example (Craps Bayes' Rule)

$$P(7|\text{Win}) = \frac{P(\text{Win}|7)P(7)}{P(\text{Win})}$$

Bayes' Rule

Theorem (Bayes' Rule)

If A and B are events in E with $P(B) > 0$, then Bayes' Rule states

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

Example (Craps Bayes' Rule)

$$P(7|\text{Win}) = \frac{P(\text{Win}|7)P(7)}{P(\text{Win})} = \frac{1 \cdot P(7)}{P(\text{Win})}$$

Bayes' Rule

Theorem (Bayes' Rule)

If A and B are events in E with $P(B) > 0$, then Bayes' Rule states

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

Example (Craps Bayes' Rule)

$$P(7|\text{Win}) = \frac{P(\text{Win}|7)P(7)}{P(\text{Win})} = \frac{1 \cdot P(7)}{P(\text{Win})} = \frac{6/36}{8/36} = \frac{6}{8}$$

Bayes' Rule

Theorem (Bayes' Rule)

If A and B are events in E with $P(B) > 0$, then Bayes' Rule states

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

Example (Craps Bayes' Rule)

$$P(7|\text{Win}) = \frac{P(\text{Win}|7)P(7)}{P(\text{Win})} = \frac{1 \cdot P(7)}{P(\text{Win})} = \frac{6/36}{8/36} = \frac{6}{8}$$

Down Syndrome screening

If a pregnant woman has a test for Down syndrome and it is positive, what is the probability that the child will have Down syndrome?

Down Syndrome screening

If a pregnant woman has a test for Down syndrome and it is positive, what is the probability that the child will have Down syndrome? Let D indicate a child with Down syndrome and D^c the opposite.

Down Syndrome screening

If a pregnant woman has a test for Down syndrome and it is positive, what is the probability that the child will have Down syndrome? Let D indicate a child with Down syndrome and D^c the opposite. Let '$+$' indicate a positive test result and $-$ a negative result.

Down Syndrome screening

If a pregnant woman has a test for Down syndrome and it is positive, what is the probability that the child will have Down syndrome? Let D indicate a child with Down syndrome and D^c the opposite. Let '$+$' indicate a positive test result and $-$ a negative result.

$$\text{sensitivity} \quad = P(+|D) \quad = 0.94$$

Down Syndrome screening

If a pregnant woman has a test for Down syndrome and it is positive, what is the probability that the child will have Down syndrome? Let D indicate a child with Down syndrome and D^c the opposite. Let '$+$' indicate a positive test result and $-$ a negative result.

$$\begin{array}{lll} \text{sensitivity} & = P(+|D) & = 0.94 \\ \text{specificity} & = P(-|D^c) & = 0.77 \end{array}$$

Down Syndrome screening

If a pregnant woman has a test for Down syndrome and it is positive, what is the probability that the child will have Down syndrome? Let D indicate a child with Down syndrome and D^c the opposite. Let '$+$' indicate a positive test result and $-$ a negative result.

$$\begin{array}{rll} \text{sensitivity} & = P(+|D) & = 0.94 \\ \text{specificity} & = P(-|D^c) & = 0.77 \\ \text{prevalence} & = P(D) & = 1/1000 \end{array}$$

Down Syndrome screening

If a pregnant woman has a test for Down syndrome and it is positive, what is the probability that the child will have Down syndrome? Let D indicate a child with Down syndrome and D^c the opposite. Let '$+$' indicate a positive test result and $-$ a negative result.

$$\begin{array}{rll} \text{sensitivity} & = P(+|D) & = 0.94 \\ \text{specificity} & = P(-|D^c) & = 0.77 \\ \text{prevalence} & = P(D) & = 1/1000 \end{array}$$

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

Down Syndrome screening

If a pregnant woman has a test for Down syndrome and it is positive, what is the probability that the child will have Down syndrome? Let D indicate a child with Down syndrome and D^c the opposite. Let '$+$' indicate a positive test result and $-$ a negative result.

$$\begin{array}{rll} \text{sensitivity} & = P(+|D) & = 0.94 \\ \text{specificity} & = P(-|D^c) & = 0.77 \\ \text{prevalence} & = P(D) & = 1/1000 \end{array}$$

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)} = \frac{P(+|D)P(D)}{P(+|D)P(D)+P(+|D^c)P(D^c)}$$

Down Syndrome screening

If a pregnant woman has a test for Down syndrome and it is positive, what is the probability that the child will have Down syndrome? Let D indicate a child with Down syndrome and D^c the opposite. Let '$+$' indicate a positive test result and $-$ a negative result.

$$\begin{aligned}
\text{sensitivity} &= P(+|D) &&= 0.94 \\
\text{specificity} &= P(-|D^c) &&= 0.77 \\
\text{prevalence} &= P(D) &&= 1/1000
\end{aligned}$$

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)} = \frac{P(+|D)P(D)}{P(+|D)P(D)+P(+|D^c)P(D^c)} = \frac{0.94 \cdot 0.001}{0.94 \cdot 0.001 + 0.23 \cdot 0.999}$$

Down Syndrome screening

If a pregnant woman has a test for Down syndrome and it is positive, what is the probability that the child will have Down syndrome? Let D indicate a child with Down syndrome and D^c the opposite. Let '$+$' indicate a positive test result and $-$ a negative result.

$$\begin{aligned}
\text{sensitivity} &= P(+|D) &&= 0.94 \\
\text{specificity} &= P(-|D^c) &&= 0.77 \\
\text{prevalence} &= P(D) &&= 1/1000
\end{aligned}$$

$$\begin{aligned}
P(D|+) &= \frac{P(+|D)P(D)}{P(+)} = \frac{P(+|D)P(D)}{P(+|D)P(D)+P(+|D^c)P(D^c)} = \frac{0.94\cdot 0.001}{0.94\cdot 0.001+0.23\cdot 0.999} \\
&\approx 1/250
\end{aligned}$$

Down Syndrome screening

If a pregnant woman has a test for Down syndrome and it is positive, what is the probability that the child will have Down syndrome? Let D indicate a child with Down syndrome and D^c the opposite. Let '$+$' indicate a positive test result and $-$ a negative result.

$$\begin{array}{rll} \text{sensitivity} & = P(+|D) & = 0.94 \\ \text{specificity} & = P(-|D^c) & = 0.77 \\ \text{prevalence} & = P(D) & = 1/1000 \end{array}$$

$$\begin{aligned} P(D|+) &= \frac{P(+|D)P(D)}{P(+)} = \frac{P(+|D)P(D)}{P(+|D)P(D)+P(+|D^c)P(D^c)} = \frac{0.94 \cdot 0.001}{0.94 \cdot 0.001 + 0.23 \cdot 0.999} \\ &\approx 1/250 \end{aligned}$$

$$P(D|-) \approx 1/10,000$$

A Bayesian statistician

Let

- y be the data we will collect from an experiment,

A Bayesian statistician

Let

- y be the data we will collect from an experiment,
- K be everything we know for certain about the world (aside from y), and

A Bayesian statistician

Let

- y be the data we will collect from an experiment,
- K be everything we know for certain about the world (aside from y), and
- θ be anything we don't know for certain.

A Bayesian statistician

Let

- y be the data we will collect from an experiment,
- K be everything we know for certain about the world (aside from y), and
- θ be anything we don't know for certain.

A Bayesian statistician

Let

- y be the data we will collect from an experiment,
- K be everything we know for certain about the world (aside from y), and
- θ be anything we don't know for certain.

My definition of a Bayesian statistician is an individual who makes decisions based on the probability distribution of those things we don't know conditional on what we know,

A Bayesian statistician

Let

- y be the data we will collect from an experiment,
- K be everything we know for certain about the world (aside from y), and
- θ be anything we don't know for certain.

My definition of a Bayesian statistician is an individual who makes decisions based on the probability distribution of those things we don't know conditional on what we know, i.e.

$$p(\theta|y, K).$$

Bayesian statistics (with explicit conditioning)

- Parameter estimation:

$$p(\theta|y, M)$$

where M is a model with parameter (vector) θ and y is data assumed to come from model M with true parameter θ_0.

Bayesian statistics (with explicit conditioning)

- Parameter estimation:

$$p(\theta|y, M)$$

  where M is a model with parameter (vector) θ and y is data assumed to come from model M with true parameter θ_0.

- Hypothesis testing/model comparison:

$$p(M_j|y, \mathcal{M})$$

  where $\mathcal{M}$ is a set of models with $M_j \in \mathcal{M}$ for $i = 1, 2, \ldots$ and y is data assumed to come from some model $M_0 \in \mathcal{M}$.

Bayesian statistics (with explicit conditioning)

- Parameter estimation:
$$p(\theta|y, M)$$
where M is a model with parameter (vector) θ and y is data assumed to come from model M with true parameter θ_0.
- Hypothesis testing/model comparison:
$$p(M_j|y, \mathcal{M})$$
where $\mathcal{M}$ is a set of models with $M_j \in \mathcal{M}$ for $i = 1, 2, \ldots$ and y is data assumed to come from some model $M_0 \in \mathcal{M}$.
- Prediction:
$$p(\tilde{y}|y, M)$$
where $\tilde{y}$ is unobserved data and y and $\tilde{y}$ are both assumed to come from M.

Bayesian statistics (with explicit conditioning)

- Parameter estimation:

$$p(\theta|y, M)$$

where M is a model with parameter (vector) θ and y is data assumed to come from model M with true parameter θ_0.

- Hypothesis testing/model comparison:

$$p(M_j|y, \mathcal{M})$$

where $\mathcal{M}$ is a set of models with $M_j \in \mathcal{M}$ for $i = 1, 2, \ldots$ and y is data assumed to come from some model $M_0 \in \mathcal{M}$.

- Prediction:

$$p(\tilde{y}|y, M)$$

where $\tilde{y}$ is unobserved data and y and $\tilde{y}$ are both assumed to come from M.

Bayesian statistics (with explicit conditioning)

- Parameter estimation:

$$p(\theta|y, M)$$

where M is a model with parameter (vector) θ and y is data assumed to come from model M with true parameter θ_0.

- Hypothesis testing/model comparison:

$$p(M_j|y, \mathcal{M})$$

where $\mathcal{M}$ is a set of models with $M_j \in \mathcal{M}$ for $i = 1, 2, \ldots$ and y is data assumed to come from some model $M_0 \in \mathcal{M}$.

- Prediction:

$$p(\tilde{y}|y, M)$$

where $\tilde{y}$ is unobserved data and y and $\tilde{y}$ are both assumed to come from M. Alternatively,

$$p(\tilde{y}|y, \mathcal{M})$$

where y and $\tilde{y}$ are both assumed to come from some $M_0 \in \mathcal{M}$.

Bayesian statistics (with implicit conditioning)

- Parameter estimation:

$$p(\theta|y)$$

  where θ is the unknown parameter (vector) and y is the data.

- Hypothesis testing/model comparison:

$$p(M_j|y)$$

  where M_j is one of a set of models under consideration and y is data assumed to come from one of those models.

- Prediction:

$$p(\tilde{y}|y)$$

  where $\tilde{y}$ is unobserved data and y and $\tilde{y}$ are both assumed to come from the same (set of) model(s).

Bayes' Rule

Bayes' Rule applied to a partition $P = \{A_1, A_2, \ldots\}$,

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

Bayes' Rule

Bayes' Rule applied to a partition $P = \{A_1, A_2, \ldots\}$,

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

Bayes' Rule also applies to probability density (or mass) functions, e.g.

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

where the integral plays the role of the sum in the previous statement.

Parameter estimation

Let y be data from some model with unknown parameter θ.

Parameter estimation

Let y be data from some model with unknown parameter θ. Then

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

Parameter estimation

Let y be data from some model with unknown parameter θ. Then

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

and we use the following terminology

Terminology	Notation
Posterior	$p(\theta\|y)$
Prior	$p(\theta)$
Model	$p(y\|\theta)$
Prior predictive distribution (marginal likelihood)	$p(y)$

Parameter estimation

Let y be data from some model with unknown parameter θ. Then

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

and we use the following terminology

Terminology	Notation
Posterior	$p(\theta\|y)$
Prior	$p(\theta)$
Model	$p(y\|\theta)$
Prior predictive distribution (marginal likelihood)	$p(y)$

If θ is discrete (continuous),
then $p(\theta)$ and $p(\theta|y)$ are probability mass (density) functions.

Parameter estimation

Let y be data from some model with unknown parameter θ. Then

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

and we use the following terminology

Terminology	Notation
Posterior	$p(\theta|y)$
Prior	$p(\theta)$
Model	$p(y|\theta)$
Prior predictive distribution (marginal likelihood)	$p(y)$

If θ is discrete (continuous),
then $p(\theta)$ and $p(\theta|y)$ are probability mass (density) functions.
If y is discrete (continuous),
then $p(y|\theta)$ and $p(y)$ are probability mass (density) functions.

Example: exponential model

Let $Y|\theta \sim Exp(\theta)$, then this defines the likelihood,

Example: exponential model

Let $Y|\theta \sim Exp(\theta)$, then this defines the likelihood, i.e.

$$p(y|\theta) = \theta e^{-\theta y}.$$

Example: exponential model

Let $Y|\theta \sim Exp(\theta)$, then this defines the likelihood, i.e.

$$p(y|\theta) = \theta e^{-\theta y}.$$

Let's assume a convenient prior $\theta \sim Ga(a, b)$,

Example: exponential model

Let $Y|\theta \sim Exp(\theta)$, then this defines the likelihood, i.e.

$$p(y|\theta) = \theta e^{-\theta y}.$$

Let's assume a convenient prior $\theta \sim Ga(a, b)$, then

$$p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}.$$

Example: exponential model

Let $Y|\theta \sim Exp(\theta)$, then this defines the likelihood, i.e.

$$p(y|\theta) = \theta e^{-\theta y}.$$

Let's assume a convenient prior $\theta \sim Ga(a, b)$, then

$$p(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}.$$

The prior predictive distribution is

$$p(y) = \int p(y|\theta)p(\theta)d\theta = \frac{b^a}{\Gamma(a)}\frac{\Gamma(a+1)}{(b+y)^{a+1}}.$$

Example: exponential model

Let $Y|\theta \sim Exp(\theta)$, then this defines the likelihood, i.e.

$$p(y|\theta) = \theta e^{-\theta y}.$$

Let's assume a convenient prior $\theta \sim Ga(a, b)$, then

$$p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}.$$

The prior predictive distribution is

$$p(y) = \int p(y|\theta)p(\theta)d\theta = \frac{b^a}{\Gamma(a)} \frac{\Gamma(a+1)}{(b+y)^{a+1}}.$$

The posterior is

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{(b+y)^{a+1}}{\Gamma(a+1)} \theta^{a+1-1} e^{-(b+y)\theta},$$

Example: exponential model

Let $Y|\theta \sim Exp(\theta)$, then this defines the likelihood, i.e.

$$p(y|\theta) = \theta e^{-\theta y}.$$

Let's assume a convenient prior $\theta \sim Ga(a, b)$, then

$$p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}.$$

The prior predictive distribution is

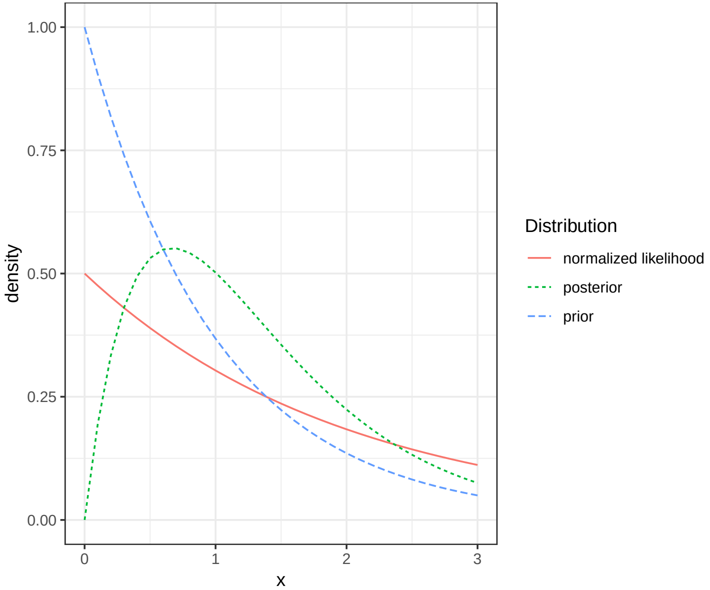
$$p(y) = \int p(y|\theta) p(\theta) d\theta = \frac{b^a}{\Gamma(a)} \frac{\Gamma(a+1)}{(b+y)^{a+1}}.$$

The posterior is

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{(b+y)^{a+1}}{\Gamma(a+1)} \theta^{a+1-1} e^{-(b+y)\theta},$$

thus $\theta|y \sim Ga(a+1, b+y)$.

```
a = 1; b = 1; y = 0.5
```


1.00
0.75
0.50
0.25
0.00
density
0
1
2
3
x
Distribution
normalized likelihood
posterior
prior

A shortcut

If

$$p(y) = \int p(y|\theta)p(\theta)d\theta < \infty,$$

A shortcut

If

$$p(y) = \int p(y|\theta)p(\theta)d\theta < \infty,$$

then we can actually use the following to find the posterior

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

A shortcut

If

$$p(y) = \int p(y|\theta)p(\theta)d\theta < \infty,$$

then we can actually use the following to find the posterior

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

where the $\propto$ signifies that terms not involving θ (or anything on the left of the conditioning bar) are irrelevant and can be dropped.

A shortcut

If

$$p(y) = \int p(y|\theta)p(\theta)d\theta < \infty,$$

then we can actually use the following to find the posterior

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

where the $\propto$ signifies that terms not involving θ (or anything on the left of the conditioning bar) are irrelevant and can be dropped.

In the exponential example

$$p(\theta|y) \propto p(y|\theta)p(\theta) \propto \theta e^{-\theta y}\theta^{a-1}e^{-b\theta} = \theta^{a+1-1}e^{-(b+y)\theta}$$

A shortcut

If

$$p(y) = \int p(y|\theta)p(\theta)d\theta < \infty,$$

then we can actually use the following to find the posterior

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

where the $\propto$ signifies that terms not involving θ (or anything on the left of the conditioning bar) are irrelevant and can be dropped.

In the exponential example

$$p(\theta|y) \propto p(y|\theta)p(\theta) \propto \theta e^{-\theta y}\theta^{a-1}e^{-b\theta} = \theta^{a+1-1}e^{-(b+y)\theta}$$

where we can recognize $p(\theta|y)$ as the kernel of a $Ga(a+1, b+y)$ distribution

A shortcut

If

$$p(y) = \int p(y|\theta)p(\theta)d\theta < \infty,$$

then we can actually use the following to find the posterior

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

where the $\propto$ signifies that terms not involving θ (or anything on the left of the conditioning bar) are irrelevant and can be dropped.

In the exponential example

$$p(\theta|y) \propto p(y|\theta)p(\theta) \propto \theta e^{-\theta y} \theta^{a-1} e^{-b\theta} = \theta^{a+1-1} e^{-(b+y)\theta}$$

where we can recognize $p(\theta|y)$ as the kernel of a $Ga(a+1, b+y)$ distribution and thus $\theta|y \sim Ga(a+1, b+y)$ and $p(y) < \infty$.

Independent data

Suppose $Y_i|\theta \stackrel{ind}{\sim} Exp(\theta)$ for $i = 1, \ldots, n$ and $y = (y_1, \ldots, y_n)$, then

$$p(y|\theta) = \prod_{i=1}^{n} p(y_i|\theta) = \theta^n e^{-\theta n \overline{y}}$$

Independent data

Suppose $Y_i|\theta \stackrel{ind}{\sim} Exp(\theta)$ for $i = 1, \ldots, n$ and $y = (y_1, \ldots, y_n)$, then

$$p(y|\theta) = \prod_{i=1}^{n} p(y_i|\theta) = \theta^n e^{-\theta n \overline{y}}$$

Then

$$p(\theta|y) \propto p(y|\theta)p(\theta) \propto \theta^{a+n-1} e^{-(b+n\overline{y})\theta}$$

Independent data

Suppose $Y_i|\theta \stackrel{ind}{\sim} Exp(\theta)$ for $i = 1, \ldots, n$ and $y = (y_1, \ldots, y_n)$, then

$$p(y|\theta) = \prod_{i=1}^{n} p(y_i|\theta) = \theta^n e^{-\theta n\overline{y}}$$

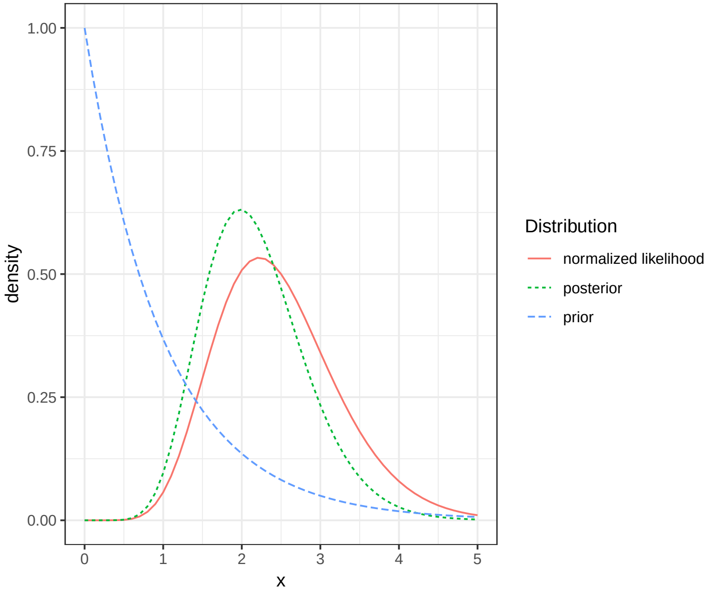
Then

$$p(\theta|y) \propto p(y|\theta)p(\theta) \propto \theta^{a+n-1} e^{-(b+n\overline{y})\theta}$$

where $n\overline{y} = \sum_{i=1}^{n} y_i$. We recognize this as the kernel of a gamma, i.e.

$\theta|y \sim Ga(a+n, b+n\overline{y})$.

```
a = 1; b = 1; set.seed(20141121); y = rexp(10, 2)
```



Bayesian learning (in parameter estimation)

So, Bayes' Rule provides a formula for updating from prior beliefs to our posterior beliefs based on the data we observe,

Bayesian learning (in parameter estimation)

So, Bayes' Rule provides a formula for updating from prior beliefs to our posterior beliefs based on the data we observe, i.e.

$$p(\theta|y) = \frac{p(y|\theta)}{p(y)}p(\theta) \propto p(y|\theta)p(\theta)$$

Bayesian learning (in parameter estimation)

So, Bayes' Rule provides a formula for updating from prior beliefs to our posterior beliefs based on the data we observe, i.e.

$$p(\theta|y) = \frac{p(y|\theta)}{p(y)}p(\theta) \propto p(y|\theta)p(\theta)$$

Suppose we gather $y_1, \ldots, y_n$ sequentially (and we assume y_i independent conditional on θ),

Bayesian learning (in parameter estimation)

So, Bayes' Rule provides a formula for updating from prior beliefs to our posterior beliefs based on the data we observe, i.e.

$$p(\theta|y) = \frac{p(y|\theta)}{p(y)}p(\theta) \propto p(y|\theta)p(\theta)$$

Suppose we gather $y_1, \ldots, y_n$ sequentially (and we assume y_i independent conditional on θ), then we have

$$p(\theta|y_1) \quad \propto p(y_1|\theta)p(\theta)$$

Bayesian learning (in parameter estimation)

So, Bayes' Rule provides a formula for updating from prior beliefs to our posterior beliefs based on the data we observe, i.e.

$$p(\theta|y) = \frac{p(y|\theta)}{p(y)}p(\theta) \propto p(y|\theta)p(\theta)$$

Suppose we gather $y_1, \ldots, y_n$ sequentially (and we assume y_i independent conditional on θ), then we have

$$\begin{aligned} p(\theta|y_1) &\propto p(y_1|\theta)p(\theta) \\ p(\theta|y_1, y_2) &\propto p(y_2|\theta)p(\theta|y_1) \end{aligned}$$

Bayesian learning (in parameter estimation)

So, Bayes' Rule provides a formula for updating from prior beliefs to our posterior beliefs based on the data we observe, i.e.

$$p(\theta|y) = \frac{p(y|\theta)}{p(y)}p(\theta) \propto p(y|\theta)p(\theta)$$

Suppose we gather $y_1, \ldots, y_n$ sequentially (and we assume y_i independent conditional on θ), then we have

$$\begin{aligned} p(\theta|y_1) &\propto p(y_1|\theta)p(\theta) \\ p(\theta|y_1, y_2) &\propto p(y_2|\theta)p(\theta|y_1) \end{aligned}$$

and

$$p(\theta|y_1, \ldots, y_i) \propto p(y_i|\theta)p(\theta|y_1, \ldots, y_{i-1})$$

Bayesian learning (in parameter estimation)

So, Bayes' Rule provides a formula for updating from prior beliefs to our posterior beliefs based on the data we observe, i.e.

$$p(\theta|y) = \frac{p(y|\theta)}{p(y)}p(\theta) \propto p(y|\theta)p(\theta)$$

Suppose we gather $y_1, \ldots, y_n$ sequentially (and we assume y_i independent conditional on θ), then we have

$$\begin{aligned} p(\theta|y_1) &\propto p(y_1|\theta)p(\theta) \\ p(\theta|y_1, y_2) &\propto p(y_2|\theta)p(\theta|y_1) \end{aligned}$$

and

$$p(\theta|y_1, \ldots, y_i) \propto p(y_i|\theta)p(\theta|y_1, \ldots, y_{i-1})$$

So Bayesian learning is

$$p(\theta) \to p(\theta|y_1) \to p(\theta|y_1, y_2) \to \cdots \to p(\theta|y_1, \ldots, y_n).$$

Model comparison

Formally, to compare models (or average over models), we use

$$p(M_j|y) \propto p(y|M_j)p(M_j)$$

Model comparison

Formally, to compare models (or average over models), we use

$$p(M_j|y) \propto p(y|M_j)p(M_j)$$

where

- $p(y|M_j)$ is the likelihood of the data when model M_j is true

Model comparison

Formally, to compare models (or average over models), we use

$$p(M_j|y) \propto p(y|M_j)p(M_j)$$

where

- $p(y|M_j)$ is the likelihood of the data when model M_j is true
- $p(M_j)$ is the prior probababability for model M_j

Model comparison

Formally, to compare models (or average over models), we use

$$p(M_j|y) \propto p(y|M_j)p(M_j)$$

where

- $p(y|M_j)$ is the likelihood of the data when model M_j is true
- $p(M_j)$ is the prior probabability for model M_j
- $p(M_j|y)$ is the posterior probability for model M_j

Model comparison

Formally, to compare models (or average over models), we use

$$p(M_j|y) \propto p(y|M_j)p(M_j)$$

where

- $p(y|M_j)$ is the likelihood of the data when model M_j is true
- $p(M_j)$ is the prior probabability for model M_j
- $p(M_j|y)$ is the posterior probability for model M_j

Model comparison

Formally, to compare models (or average over models), we use

$$p(M_j|y) \propto p(y|M_j)p(M_j)$$

where

- $p(y|M_j)$ is the likelihood of the data when model M_j is true
- $p(M_j)$ is the prior probabability for model M_j
- $p(M_j|y)$ is the posterior probability for model M_j

Thus, a Bayesian approach provides a natural way to learn about models, i.e. $p(M_j) \rightarrow p(M_j|y)$.

Prediction

Let y be observed data and $\tilde{y}$ be unobserved data from a model with parameter θ

Prediction

Let y be observed data and $\tilde{y}$ be unobserved data from a model with parameter θ where $\tilde{y}$ is conditionally independent of y given θ

Prediction

Let y be observed data and $\tilde{y}$ be unobserved data from a model with parameter θ where $\tilde{y}$ is conditionally independent of y given θ (true for many of the models we will discuss this semester)

Prediction

Let y be observed data and $\tilde{y}$ be unobserved data from a model with parameter θ where $\tilde{y}$ is conditionally independent of y given θ (true for many of the models we will discuss this semester) , then

$$p(\tilde{y}|y) \quad = \int p(\tilde{y}, \theta|y) d\theta$$

Prediction

Let y be observed data and $\tilde{y}$ be unobserved data from a model with parameter θ where $\tilde{y}$ is conditionally independent of y given θ (true for many of the models we will discuss this semester) , then

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y},\theta|y)d\theta \\ &= \int p(\tilde{y}|\theta,y)p(\theta|y)d\theta \end{aligned}$$

Prediction

Let y be observed data and $\tilde{y}$ be unobserved data from a model with parameter θ where $\tilde{y}$ is conditionally independent of y given θ (true for many of the models we will discuss this semester) , then

$$\begin{array}{rl} p(\tilde{y}|y) & = \int p(\tilde{y}, \theta|y) d\theta \\ & = \int p(\tilde{y}|\theta, y) p(\theta|y) d\theta \\ & = \int p(\tilde{y}|\theta) p(\theta|y) d\theta \end{array}$$

Prediction

Let y be observed data and $\tilde{y}$ be unobserved data from a model with parameter θ where $\tilde{y}$ is conditionally independent of y given θ (true for many of the models we will discuss this semester) , then

$$\begin{array}{rl} p(\tilde{y}|y) & = \int p(\tilde{y},\theta|y)d\theta \\ & = \int p(\tilde{y}|\theta,y)p(\theta|y)d\theta \\ & = \int p(\tilde{y}|\theta)p(\theta|y)d\theta \end{array}$$

where $p(\theta|y)$ is the posterior we obtained using Bayesian parameter estimation techniques.

Example: exponential distribution

From previous, let $y_i \stackrel{ind}{\sim} Exp(\theta)$ and $\theta \sim Ga(a, b)$, then $\theta|y \sim Ga(a + n, b + n\overline{y})$.

Example: exponential distribution

From previous, let $y_i \stackrel{ind}{\sim} Exp(\theta)$ and $\theta \sim Ga(a,b)$, then $\theta|y \sim Ga(a+n, b+n\overline{y})$. Suppose we are interested in predicting a new value $\tilde{y} \sim Exp(\theta)$

Example: exponential distribution

From previous, let $y_i \stackrel{ind}{\sim} Exp(\theta)$ and $\theta \sim Ga(a,b)$, then $\theta|y \sim Ga(a+n, b+n\overline{y})$. Suppose we are interested in predicting a new value $\tilde{y} \sim Exp(\theta)$ (conditionally independent of $y = (y_1, \ldots, y_n)$ given θ).

Example: exponential distribution

From previous, let $y_i \stackrel{ind}{\sim} Exp(\theta)$ and $\theta \sim Ga(a, b)$, then $\theta|y \sim Ga(a + n, b + n\overline{y})$. Suppose we are interested in predicting a new value $\tilde{y} \sim Exp(\theta)$ (conditionally independent of $y = (y_1, \ldots, y_n)$ given θ). Then we have

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta$$

Example: exponential distribution

From previous, let $y_i \stackrel{ind}{\sim} Exp(\theta)$ and $\theta \sim Ga(a,b)$, then $\theta|y \sim Ga(a+n, b+n\overline{y})$. Suppose we are interested in predicting a new value $\tilde{y} \sim Exp(\theta)$ (conditionally independent of $y = (y_1, \ldots, y_n)$ given θ). Then we have

$$\begin{array}{rl} p(\tilde{y}|y) & = \int p(\tilde{y}|\theta)p(\theta|y)d\theta \\ & = \int \theta e^{-\theta\tilde{y}} \frac{(b+n\overline{y})^{a+n}}{\Gamma(a+1)} \theta^{a+n-1} e^{-\theta(b+n\overline{y})} d\theta \end{array}$$

Example: exponential distribution

From previous, let $y_i \stackrel{ind}{\sim} Exp(\theta)$ and $\theta \sim Ga(a,b)$, then $\theta|y \sim Ga(a+n, b+n\overline{y})$. Suppose we are interested in predicting a new value $\tilde{y} \sim Exp(\theta)$ (conditionally independent of $y = (y_1, \ldots, y_n)$ given θ). Then we have

$$
\begin{array}{rl}
p(\tilde{y}|y) & = \int p(\tilde{y}|\theta)p(\theta|y)d\theta \\
& = \int \theta e^{-\theta\tilde{y}} \frac{(b+n\overline{y})^{a+n}}{\Gamma(a+1)} \theta^{a+n-1} e^{-\theta(b+n\overline{y})} d\theta \\
& = \frac{(b+n\overline{y})^{a+n}}{\Gamma(a+n)} \int \theta^{a+n+1-1} e^{-\theta(b+n\overline{y}+\tilde{y})} d\theta
\end{array}
$$

Example: exponential distribution

From previous, let $y_i \stackrel{ind}{\sim} Exp(\theta)$ and $\theta \sim Ga(a,b)$, then $\theta|y \sim Ga(a+n, b+n\overline{y})$. Suppose we are interested in predicting a new value $\tilde{y} \sim Exp(\theta)$ (conditionally independent of $y = (y_1, \ldots, y_n)$ given θ). Then we have

$$\begin{array}{rl} p(\tilde{y}|y) & = \int p(\tilde{y}|\theta)p(\theta|y)d\theta \\ & = \int \theta e^{-\theta\tilde{y}} \frac{(b+n\overline{y})^{a+n}}{\Gamma(a+1)} \theta^{a+n-1} e^{-\theta(b+n\overline{y})} d\theta \\ & = \frac{(b+n\overline{y})^{a+n}}{\Gamma(a+n)} \int \theta^{a+n+1-1} e^{-\theta(b+n\overline{y}+\tilde{y})} d\theta \\ & = \frac{(b+n\overline{y})^{a+n}}{\Gamma(a+n)} \frac{\Gamma(a+n+1)}{(b+n\overline{y}+\tilde{y})^{a+n+1}} \end{array}$$

Example: exponential distribution

From previous, let $y_i \stackrel{ind}{\sim} Exp(\theta)$ and $\theta \sim Ga(a,b)$, then $\theta|y \sim Ga(a+n, b+n\overline{y})$. Suppose we are interested in predicting a new value $\tilde{y} \sim Exp(\theta)$ (conditionally independent of $y = (y_1, \ldots, y_n)$ given θ). Then we have

$$
\begin{array}{rl}
p(\tilde{y}|y) & = \int p(\tilde{y}|\theta)p(\theta|y)d\theta \\
& = \int \theta e^{-\theta\tilde{y}} \frac{(b+n\overline{y})^{a+n}}{\Gamma(a+1)} \theta^{a+n-1} e^{-\theta(b+n\overline{y})} d\theta \\
& = \frac{(b+n\overline{y})^{a+n}}{\Gamma(a+n)} \int \theta^{a+n+1-1} e^{-\theta(b+n\overline{y}+\tilde{y})} d\theta \\
& = \frac{(b+n\overline{y})^{a+n}}{\Gamma(a+n)} \frac{\Gamma(a+n+1)}{(b+n\overline{y}+\tilde{y})^{a+n+1}} \\
& = \frac{(a+n)(b+n\overline{y})^{a+n}}{(\tilde{y}+b+n\overline{y})^{a+n+1}}
\end{array}
$$

This is the Lomax distribution for $\tilde{y}$ with parameters $a+n$ and $b+n\overline{y}$.

What is probability?

What is probability?

Consider the following three typical uses of the word "probability":

- What is the probability I will win on the come-out roll in craps?

What is probability?

Consider the following three typical uses of the word “probability”:

- What is the probability I will win on the come-out roll in craps?
- What is the probability my unborn child has Down’s syndrome given that they tested positive in an initial screening?

What is probability?

Consider the following three typical uses of the word "probability":

- What is the probability I will win on the come-out roll in craps?
- What is the probability my unborn child has Down's syndrome given that they tested positive in an initial screening?
- What is the probability the Green Bay Packers will win this year's superbowl?

Win on the come-out roll in craps

To win on the come-out roll in craps requires that the sum of two fair six-sided die is either a 7 or an 11.

Win on the come-out roll in craps

To win on the come-out roll in craps requires that the sum of two fair six-sided die is either a 7 or an 11. We calculated this probability earlier (based on equal probabilities of all simple outcomes) to be $2/9$.

Win on the come-out roll in craps

To win on the come-out roll in craps requires that the sum of two fair six-sided die is either a 7 or an 11. We calculated this probability earlier (based on equal probabilities of all simple outcomes) to be $2/9$. We likely meant that if we were to repeatedly roll the die, the long term proportion of wins (7s and 11s) would be $2/9$

Win on the come-out roll in craps

To win on the come-out roll in craps requires that the sum of two fair six-sided die is either a 7 or an 11. We calculated this probability earlier (based on equal probabilities of all simple outcomes) to be $2/9$. We likely meant that if we were to repeatedly roll the die, the long term proportion of wins (7s and 11s) would be $2/9$, i.e.

$$\text{if } X_i = \begin{cases} 1 & \text{if win on roll } i \\ 0 & \text{otherwise} \end{cases}$$

Win on the come-out roll in craps

To win on the come-out roll in craps requires that the sum of two fair six-sided die is either a 7 or an 11. We calculated this probability earlier (based on equal probabilities of all simple outcomes) to be $2/9$. We likely meant that if we were to repeatedly roll the die, the long term proportion of wins (7s and 11s) would be $2/9$, i.e.

$$\text{if } X_i = \left\{ \begin{array}{ll} 1 & \text{if win on roll } i \\ 0 & \text{otherwise} \end{array} \right. \qquad \text{then} \qquad \lim_{n\to\infty} \frac{\sum_{i=1}^n X_i}{n} \to \frac{2}{9}.$$

Definition

The frequency interpretation of probability is based on the relative frequency of an event (assumed to be performed in an identical manner).

Win on the come-out roll in craps

Definition

The frequency interpretation of probability is based on the relative frequency of an event (assumed to be performed in an identical manner).

Win on the come-out roll in craps

Definition

The frequency interpretation of probability is based on the relative frequency of an event (assumed to be performed in an identical manner).

Two problems with this frequency interpretation:

Win on the come-out roll in craps

Definition

The frequency interpretation of probability is based on the relative frequency of an event (assumed to be performed in an identical manner).

Two problems with this frequency interpretation:

- You cannot possibly throw the dice in an identical manner.

Win on the come-out roll in craps

Definition

The frequency interpretation of probability is based on the relative frequency of an event (assumed to be performed in an identical manner).

Two problems with this frequency interpretation:

- You cannot possibly throw the dice in an identical manner.
- If I knew enough physics, I could model each throw

Win on the come-out roll in craps

Definition

The frequency interpretation of probability is based on the relative frequency of an event (assumed to be performed in an identical manner).

Two problems with this frequency interpretation:

- You cannot possibly throw the dice in an identical manner.
- If I knew enough physics, I could model each throw and tell you exactly what the result would be

Win on the come-out roll in craps

Definition

The frequency interpretation of probability is based on the relative frequency of an event (assumed to be performed in an identical manner).

Two problems with this frequency interpretation:

- You cannot possibly throw the dice in an identical manner.
- If I knew enough physics, I could model each throw and tell you exactly what the result would be, i.e. the only randomness is because the throws are not identical.

Down's syndrome

What is the probability my unborn child has Down's syndrome given that they tested positive in an initial screening?

Down's syndrome

What is the probability my unborn child has Down's syndrome given that they tested positive in an initial screening?

Here the frequency interpretation makes no sense for two reasons:

Down's syndrome

What is the probability my unborn child has Down's syndrome given that they tested positive in an initial screening?

Here the frequency interpretation makes no sense for two reasons:

- There is only one child and thus no repeat of the experiment.
- There is no randomness: either the child has Down's syndrome or does not.

Down's syndrome

What is the probability my unborn child has Down's syndrome given that they tested positive in an initial screening?

Here the frequency interpretation makes no sense for two reasons:

- There is only one child and thus no repeat of the experiment.
- There is no randomness: either the child has Down's syndrome or does not.

Instead, we only have our own uncertainty about whether the child has Down's syndrome.

Down's syndrome

Also, why are we only conditioning on the positive test result, shouldn't we condition on everything else that is important, e.g. age.

Down's syndrome

Also, why are we only conditioning on the positive test result, shouldn't we condition on everything else that is important, e.g. age. Then the probability we care about is

$$P(D|+, \text{mother is 33}) = \frac{P(+|D, \text{mother is 33})P(D|\text{mother is 33})}{P(+|\text{mother is 33})}$$

Now the specificity, sensitivity, and prevalence are all the relative frequency of the event for this subpopulation.

Down's syndrome

Also, why are we only conditioning on the positive test result, shouldn't we condition on everything else that is important, e.g. age. Then the probability we care about is

$$P(D|+, \text{mother is 33}) = \frac{P(+|D, \text{mother is 33})P(D|\text{mother is 33})}{P(+|\text{mother is 33})}$$

Now the specificity, sensitivity, and prevalence are all the relative frequency of the event for this subpopulation.

But what about other measured variables, e.g. Caucasian, lives in MN, of Scandanavian descent, etc.

Down's syndrome

Also, why are we only conditioning on the positive test result, shouldn't we condition on everything else that is important, e.g. age. Then the probability we care about is

$$P(D|+, \text{mother is 33}) = \frac{P(+|D, \text{mother is 33})P(D|\text{mother is 33})}{P(+|\text{mother is 33})}$$

Now the specificity, sensitivity, and prevalence are all the relative frequency of the event for this subpopulation.

But what about other measured variables, e.g. Caucasian, lives in MN, of Scandanavian descent, etc. Taken to its logical extreme, each probability becomes a statement about one single event

Down's syndrome

Also, why are we only conditioning on the positive test result, shouldn't we condition on everything else that is important, e.g. age. Then the probability we care about is

$$P(D|+, \text{mother is 33}) = \frac{P(+|D, \text{mother is 33})P(D|\text{mother is 33})}{P(+|\text{mother is 33})}$$

Now the specificity, sensitivity, and prevalence are all the relative frequency of the event for this subpopulation.

But what about other measured variables, e.g. Caucasian, lives in MN, of Scandanavian descent, etc. Taken to its logical extreme, each probability becomes a statement about one single event, e.g. for this individual.

Superbowl Champions

What is the probability the Green Bay Packers win the Superbowl?

Superbowl Champions

What is the probability the Green Bay Packers win the Superbowl?

By similar arguments:

- There is only one Superbowl this year and only one Green Bay Packers.

Superbowl Champions

What is the probability the Green Bay Packers win the Superbowl?

By similar arguments:

- There is only one Superbowl this year and only one Green Bay Packers.
- Is the world random?

Superbowl Champions

What is the probability the Green Bay Packers win the Superbowl?

By similar arguments:

- There is only one Superbowl this year and only one Green Bay Packers.
- Is the world random? i.e. do we have free will?

Superbowl Champions

What is the probability the Green Bay Packers win the Superbowl?

By similar arguments:

- There is only one Superbowl this year and only one Green Bay Packers.
- Is the world random? i.e. do we have free will? If not, then (with enough time, computing power, money, etc) we could model the world and know what the result will be.

Superbowl Champions

What is the probability the Green Bay Packers win the Superbowl?

By similar arguments:

- There is only one Superbowl this year and only one Green Bay Packers.
- Is the world random? i.e. do we have free will? If not, then (with enough time, computing power, money, etc) we could model the world and know what the result will be. If yes, is there an objective probability that we could be estimating?

Personal belief

Definition

A subjective probability describes an individual's personal judgement about how likely a particular event is to occur.

http://www.stats.gla.ac.uk/glossary/?q=node/488

Personal belief

Definition

A subjective probability describes an individual's personal judgement about how likely a particular event is to occur.

http://www.stats.gla.ac.uk/glossary/?q=node/488

Remark Coherence of bets. The probability p you assign to an event E is the fraction at which you would exchange p for a return of 1 if E occurs.

Personal belief

Definition

A subjective probability describes an individual's personal judgement about how likely a particular event is to occur.

http://www.stats.gla.ac.uk/glossary/?q=node/488

Remark Coherence of bets. The probability p you assign to an event E is the fraction at which you would exchange p for a return of 1 if E occurs.

Rational individuals can differ about the probability of an event by having different knowledge, i.e. $P(E|K_1) \neq P(E|K_2)$.

Personal belief

Definition

A subjective probability describes an individual's personal judgement about how likely a particular event is to occur.

http://www.stats.gla.ac.uk/glossary/?q=node/488

Remark Coherence of bets. The probability p you assign to an event E is the fraction at which you would exchange p for a return of 1 if E occurs.

Rational individuals can differ about the probability of an event by having different knowledge, i.e. $P(E|K_1) \neq P(E|K_2)$. But given enough data, we might have $P(E|K_1, y) \approx P(E|K_2, y)$.

Personal belief

Using a personal belief definition of probability, it is easy to reconcile the use of probability in common language:

- What is the probability I will win on the come-out roll in craps?

Personal belief

Using a personal belief definition of probability, it is easy to reconcile the use of probability in common language:

- What is the probability I will win on the come-out roll in craps?
- What is the probability my unborn child has Down's syndrome given that they tested positive in an initial screening?

Personal belief

Using a personal belief definition of probability, it is easy to reconcile the use of probability in common language:

- What is the probability I will win on the come-out roll in craps?
- What is the probability my unborn child has Down's syndrome given that they tested positive in an initial screening?
- What is the probability the Green Bay Packers will win this year's superbowl?

Personal belief

Using a personal belief definition of probability, it is easy to reconcile the use of probability in common language:

- What is the probability I will win on the come-out roll in craps?
- What is the probability my unborn child has Down's syndrome given that they tested positive in an initial screening?
- What is the probability the Green Bay Packers will win this year's superbowl?
- What is the probability that global climate change is primarily driven by human activity?

Personal belief

Using a personal belief definition of probability, it is easy to reconcile the use of probability in common language:

- What is the probability I will win on the come-out roll in craps?
- What is the probability my unborn child has Down's syndrome given that they tested positive in an initial screening?
- What is the probability the Green Bay Packers will win this year's superbowl?
- What is the probability that global climate change is primarily driven by human activity?
- What is the probability the Higgs Boson exists?

Personal belief

Using a personal belief definition of probability, it is easy to reconcile the use of probability in common language:

- What is the probability I will win on the come-out roll in craps?
- What is the probability my unborn child has Down's syndrome given that they tested positive in an initial screening?
- What is the probability the Green Bay Packers will win this year's superbowl?
- What is the probability that global climate change is primarily driven by human activity?
- What is the probability the Higgs Boson exists?

Personal belief

Using a personal belief definition of probability, it is easy to reconcile the use of probability in common language:

- What is the probability I will win on the come-out roll in craps?
- What is the probability my unborn child has Down's syndrome given that they tested positive in an initial screening?
- What is the probability the Green Bay Packers will win this year's superbowl?
- What is the probability that global climate change is primarily driven by human activity?
- What is the probability the Higgs Boson exists?

and in the mathematical notation:

- $p(\theta) \to p(\theta|y)$

Personal belief

Using a personal belief definition of probability, it is easy to reconcile the use of probability in common language:

- What is the probability I will win on the come-out roll in craps?
- What is the probability my unborn child has Down's syndrome given that they tested positive in an initial screening?
- What is the probability the Green Bay Packers will win this year's superbowl?
- What is the probability that global climate change is primarily driven by human activity?
- What is the probability the Higgs Boson exists?

and in the mathematical notation:

- $p(\theta) \to p(\theta|y)$
- $p(H_1) \to p(H_1|y)$

Personal belief

Using a personal belief definition of probability, it is easy to reconcile the use of probability in common language:

- What is the probability I will win on the come-out roll in craps?
- What is the probability my unborn child has Down's syndrome given that they tested positive in an initial screening?
- What is the probability the Green Bay Packers will win this year's superbowl?
- What is the probability that global climate change is primarily driven by human activity?
- What is the probability the Higgs Boson exists?

and in the mathematical notation:

- $p(\theta) \to p(\theta|y)$
- $p(H_1) \to p(H_1|y)$
- $p(\tilde{y}|y)$

Why or why not Bayesian?

Why do a Bayesian analysis?

- Incorporate prior knowledge via $p(\theta)$

Why or why not Bayesian?

Why do a Bayesian analysis?

- Incorporate prior knowledge via $p(\theta)$
- Coherent, i.e. everything follows from specifying $p(\theta|y)$

Why or why not Bayesian?

Why do a Bayesian analysis?

- Incorporate prior knowledge via $p(\theta)$
- Coherent, i.e. everything follows from specifying $p(\theta|y)$
- Interpretability of results, e.g. the probability the parameter is in (L, U) is 95%

Why or why not Bayesian?

Why do a Bayesian analysis?

- Incorporate prior knowledge via $p(\theta)$
- Coherent, i.e. everything follows from specifying $p(\theta|y)$
- Interpretability of results, e.g. the probability the parameter is in (L, U) is 95%

Why or why not Bayesian?

Why do a Bayesian analysis?

- Incorporate prior knowledge via $p(\theta)$
- Coherent, i.e. everything follows from specifying $p(\theta|y)$
- Interpretability of results, e.g. the probability the parameter is in (L, U) is 95%

Why not do a Bayesian analysis?

- Need to specify $p(\theta)$

Why or why not Bayesian?

Why do a Bayesian analysis?

- Incorporate prior knowledge via $p(\theta)$
- Coherent, i.e. everything follows from specifying $p(\theta|y)$
- Interpretability of results, e.g. the probability the parameter is in (L, U) is 95%

Why not do a Bayesian analysis?

- Need to specify $p(\theta)$
- Computational cost

Why or why not Bayesian?

Why do a Bayesian analysis?

- Incorporate prior knowledge via $p(\theta)$
- Coherent, i.e. everything follows from specifying $p(\theta|y)$
- Interpretability of results, e.g. the probability the parameter is in (L, U) is 95%

Why not do a Bayesian analysis?

- Need to specify $p(\theta)$
- Computational cost
- Does not guarantee coverage, i.e. how well do the procedures work over all their uses (although *frequentist matching* priors are specifically designed to ensure frequentist properties, e.g. coverage)