Conditional probability

All probability judgements depend on the available information. We discuss the *conditional probability* $\mathbb{P}(E|F)$, i.e. the probability of an event E given that we know that an event F has occurred. After examine some examples, we arrive at

Definition

Let E and F two events in a sample space Ω with $\mathbb{P}(F) > 0$. $\mathbb{P}(E|F)$ is defined as:

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}. \tag{1}$$

Conditional probability

All probability judgements depend on the available information. We discuss the *conditional probability* $\mathbb{P}(E|F)$, i.e. the probability of an event E given that we know that an event F has occurred. After examine some examples, we arrive at

Definition

Let E and F two events in a sample space Ω with $\mathbb{P}(F) > 0$. $\mathbb{P}(E|F)$ is defined as:

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}. \tag{1}$$

Note: $\mathbb{P}(E|F)$ is a new probability of the same event E.

Product rule and independence

Multiplying both sides of (1) by $\mathbb{P}(F)$, we arrive at the *product rule*:

$$\mathbb{P}(E \cap F) = \mathbb{P}(F) \cdot \mathbb{P}(E|F) \tag{2}$$

Product rule and independence

Multiplying both sides of (1) by $\mathbb{P}(F)$, we arrive at the *product rule*:

$$\mathbb{P}(E \cap F) = \mathbb{P}(F) \cdot \mathbb{P}(E|F) \tag{2}$$

Note: We may use these relations either way. Sometimes we know $\mathbb{P}(E \cap F)$ and use (1) to compute $\mathbb{P}(E|F)$. In other cases, we know $\mathbb{P}(E|F)$ and use (2) to obtain $\mathbb{P}(E \cap F)$

Product rule and independence

Multiplying both sides of (1) by $\mathbb{P}(F)$, we arrive at the *product rule*:

$$\mathbb{P}(E \cap F) = \mathbb{P}(F) \cdot \mathbb{P}(E|F) \tag{2}$$

Note: We may use these relations either way. Sometimes we know $\mathbb{P}(E \cap F)$ and use (1) to compute $\mathbb{P}(E|F)$. In other cases, we know $\mathbb{P}(E|F)$ and use (2) to obtain $\mathbb{P}(E \cap F)$
We consider E and F *independent* if knowledge of F does not change the probability of E, i.e. if $\mathbb{P}(E|F) = \mathbb{P}(E)$.

Product rule and independence

Multiplying both sides of (1) by $\mathbb{P}(F)$, we arrive at the *product rule*:

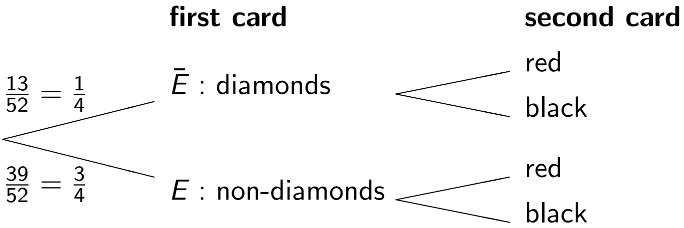
$$\mathbb{P}(E \cap F) = \mathbb{P}(F) \cdot \mathbb{P}(E|F) \tag{2}$$

Note: We may use these relations either way. Sometimes we know $\mathbb{P}(E \cap F)$ and use (1) to compute $\mathbb{P}(E|F)$. In other cases, we know $\mathbb{P}(E|F)$ and use (2) to obtain $\mathbb{P}(E \cap F)$
We consider E and F *independent* if knowledge of F does not change the probability of E, i.e. if $\mathbb{P}(E|F) = \mathbb{P}(E)$. Inserting this in (2), we obtain

$$E \text{ and } F \textit{ independent} \text{ if } \mathbb{P}(E \cap F) = \mathbb{P}(E) \cdot \mathbb{P}(F). \tag{3}$$

Tree diagram

A tree diagram is a graphical tool to represent chains of events. Suppose I draw 2 cards from a deck of 52. Then



Product and sum rules translate into visual rules:

- the probability of a chain (say 1st Diamonds-2nd Black) is obtained by multiplying the probabilities of each link;
- the probability of an event (say 2nd Black) is obtained by summing the probabilities of all chains leading to the event *(in this case one gets* $1/2$*)*.

Conditional probability. 2

Referring to the tree diagram, one can compute *conditional probabilities* either way:

- $\mathbb{P}(\text{2nd Black}|\text{1st Diamond}) = \frac{26}{51}$;
- $\mathbb{P}(\text{1st Diamond}|\text{2nd Black}) = ?$

To compute the latter, use the definition of conditional probability:

$$\mathbb{P}(\text{1st Diamond}|\text{2nd Black}) = \frac{\mathbb{P}(\text{1st Diamond } \textbf{and} \text{ 2nd Black})}{\mathbb{P}(\text{2nd Black})}$$
$$= \frac{\frac{13}{52} \cdot \frac{26}{51}}{1/2} = \frac{13}{51}$$

using the tree diagram to compute $\mathbb{P}(1stDiamond$ **and** 2nd Black$)$ and $\mathbb{P}(\text{2nd Black})$.

We can say we obtained the probability of the *cause* (the 1st card we draw), having observed a *consequence* (the 2nd card drawn)

Bayes' formula

The previous computation generalizes to Bayes' formula.

$$A_1, \ldots A_n \text{ alternative hypotheses}$$
$$[\ A_i \cap A_j = \emptyset \text{ for } i \neq j,\ A_1 \cap \ldots \cap A_n = \Omega\ ]$$

E = observed event. We know $P(E|A_j)$, $j = 1 \ldots n$. Then

$$P(A_j|E) = \frac{P(E|A_j) \cdot P(A_j)}{P(E)}$$
$$= \frac{P(E|A_j) \cdot P(A_j)}{P(A_1) \cdot P(E|A_1) + \cdots + P(A_n) \cdot P(E|A_n)}$$

$P(A_j|E)$ probability *a posteriori* of A_j.

Bayes' formula

The previous computation generalizes to Bayes' formula.

$$A_1, \ldots A_n \text{ alternative hypotheses}$$
$$[\ A_i \cap A_j = \emptyset \text{ for } i \neq j,\ A_1 \cap \ldots \cap A_n = \Omega\ ]$$

E = observed event. We know $P(E|A_j)$, $j = 1 \ldots n$. Then

$$\begin{aligned} P(A_j|E) &= \frac{P(E|A_j) \cdot P(A_j)}{P(E)} \\ &= \frac{P(E|A_j) \cdot P(A_j)}{P(A_1) \cdot P(E|A_1) + \cdots + P(A_n) \cdot P(E|A_n)} \end{aligned}$$

$P(A_j|E)$ probability *a posteriori* of A_j.
But to compute it we need to know also the probability *a priori* of A_j.

Approaches to inferential statistics

- Bayesian statistics: compute *a posteriori* estimates of parameters and scientific hypotheses (very little used until 10-20 years ago, mainly because of computational problems *(and also philosophical)*
- Frequentist statistics: observed data are only a sample from infinitely many other possibilities; we assess what could have happened (standard statistical methods: confidence intervals, hypothdsis testing. . .)

Bayesian approach to estimation.

We assume that data have been generated according to a model that includes the parameters ϑ.
In the Bayesian approach there is no *true value* of ϑ. There is a probability *a priori* for ϑ, and a probability *a posteriori* after the sample has been measured.
Correspondigly to Bayes formula:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

where B is the observation. $\mathbb{P}(A_i)$ is the probability *a priori*, while the conditional probability $\mathbb{P}(A_i|B)$ is the probability *a posteriori*.
in parameter estimation, one starts from a density *a priori* $\rho(\vartheta)$ to obtain a density *a posteriori* (after the sample $X_1, \ldots, X_n$) given by

$$f_{post}(\vartheta) = \frac{\mathbb{P}_\vartheta(X_1, \ldots, X_n)\rho(\vartheta)}{\int \mathbb{P}_\varphi(X_1, \ldots, X_n)\rho(\varphi)\, d\varphi}.$$

$\mathbb{P}_\vartheta(X_1, \ldots, X_n)$ is the probability of the data $(X_1, \ldots, X_n)$ if the parameter value is ϑ.

Bayesian example (elementary).

Assume that in a binomial phenomenon $n = 6$, $\#$ successes $= 4$, and p (which corresponds to the generic ϑ) is to be estimated. Then

$$\mathbb{P}_p(X_1, \ldots, X_n) = \binom{6}{4} p^4 (1-p)^2.$$

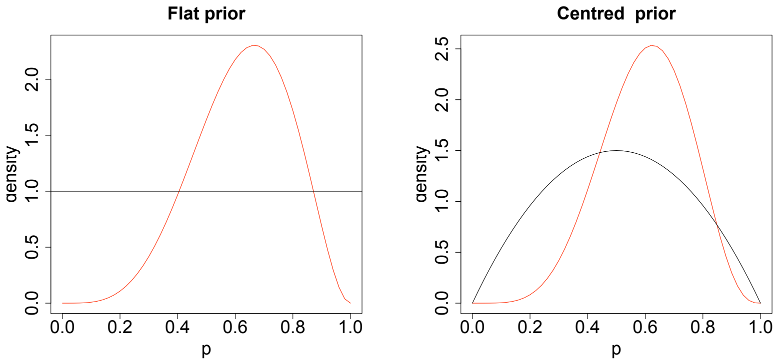
'Flat' a priori: $\rho(p) = 1$ for $p \in [0,1]$, Then

$$f_{post}(p) = \frac{\binom{6}{4} p^4 (1-p)^2}{\int_0^1 \binom{6}{4} q^4 (1-q)^2 \, dq} = Cp^4(1-p)^2 \text{ with } C = 105.$$

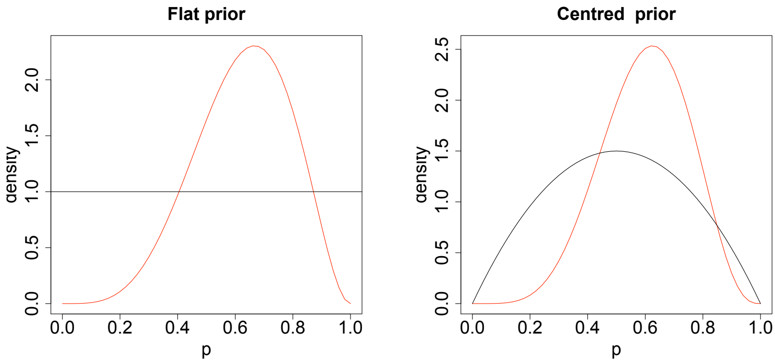
Assume instead $\rho(p) = 6p(1-p)$. Then

$$f_{post}(p) = Cp^5(1-p)^3 \text{ with } C = 504.$$

Bayesian example: figure

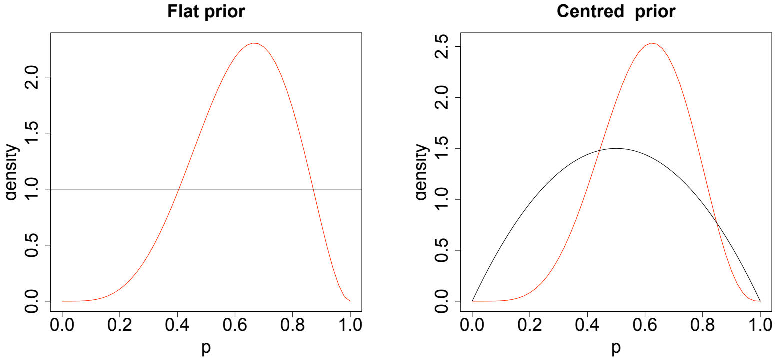


Bayesian example: figure



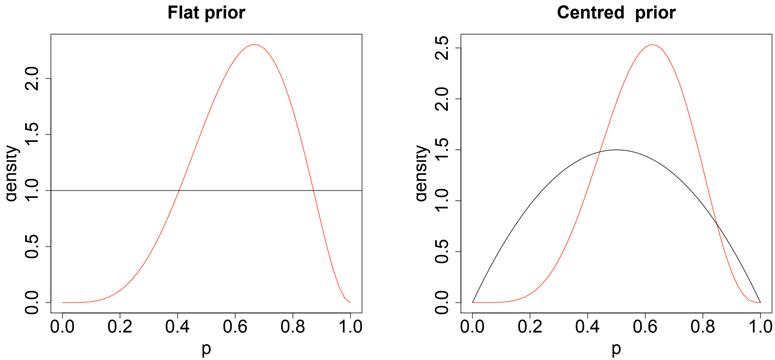
From the a posteriori plot, one can build 95% credible interval...

Bayesian example: figure



From the a posteriori plot, one can build 95% credible interval...
i.e. an interval I s.t. $P(p \in I) = 95\%$.

Bayesian example: figure



From the a posteriori plot, one can build 95% credible interval...

From the graphs one sees some effect of the prior. Increasing n would make the prior probability less relevant.

Bayesian analysis: conjugate distribution

The beta-binomial scheme:
If the prior distribution is Beta, and the likelihood is Binomial, the posterior distribution is again Beta (with different parameters).
See Script R for examples.