SECTION 1 - INTRODUCTION

1. An Interlab Problem: SRM 1946

2. Classical Solutions

3. BAYES Solution

THE PROBLEM

Multiple laboratories perform repeated measurements on the same quantity. The objective is to arrive at

- consensus value
- associated consensus measure of uncertainty.

EXAMPLE:

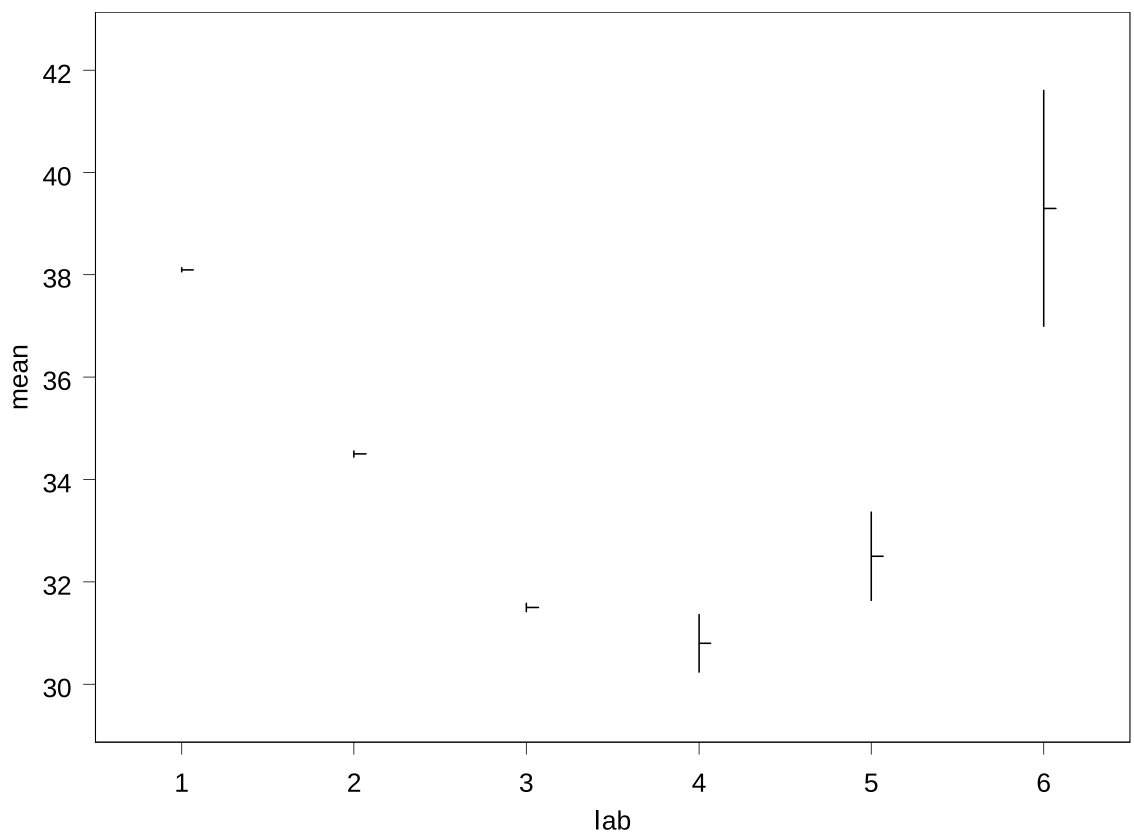
SRM 1946: Lake Superior Fish Tissue, analyzing for fatty acid and PCB content.

(Michelle Schantz, Curtis Phinney, Dianne Poster, Michael Welch, Steven Wise, CSTL):

PCB 101:

Lab ID	Mean Conc.	St. Dev.	# obs.
1	38.1	0.7	24
2	34.5	0.3	3
3	31.5	0.5	6
4	30.8	1.69	6
5	32.5	2.59	6
6	39.3	23.04	20

Graph of the Lab means ± 2 stdev.



CLASSICAL SOLUTIONS

Solution 1.
The Simplest.
GRAND MEAN:

consensus mean (μ)

estimated by the average of all data

$$\bar{\bar{Y}} = \frac{1}{N} \sum_j \sum_i Y_{ij} .$$

(36.50)

consensus uncertainty measure

estimated by the standard deviation of all data / $\sqrt{N}$.

(2.82)

Assumptions:

1. The labs all have the same mean.
2. The labs all have the same variability.
3. The data are random observations.

Advantages:

1. Conceptual simplicity.
2. Ease of calculation.

Disadvantages:

1. The assumptions are rarely met.

Solution 2.
THE MEAN OF MEANS:

consensus mean (μ)

estimated by the average of lab averages $\frac{1}{6}\sum_i \overline{Y}_i$.
(34.45)

consensus uncertainty measure
estimated by the standard deviation of the lab averages/ Sqrt(nlab).
(1.44)

Assumptions:

1. Within lab variability is negligible or the same across labs.
2. Data are random observations.

Advantages:
1. Simplicity.
2. Ease of calculation.
3. Assumptions are less restrictive than for Method 1.

Solution 3.
More sophisticated.
Maximum Likelihood(MLE) &
variants:

consensus mean (μ)

estimated by a weighted average of lab means.

(Weights are decreasing functions of lab standard deviation.)

(34.59)

consensus uncertainty measure

estimated using the within and between lab standard deviations, the lab means and the lab sample sizes.

(1.29)

Assumptions:

1.Large sample size for each lab.

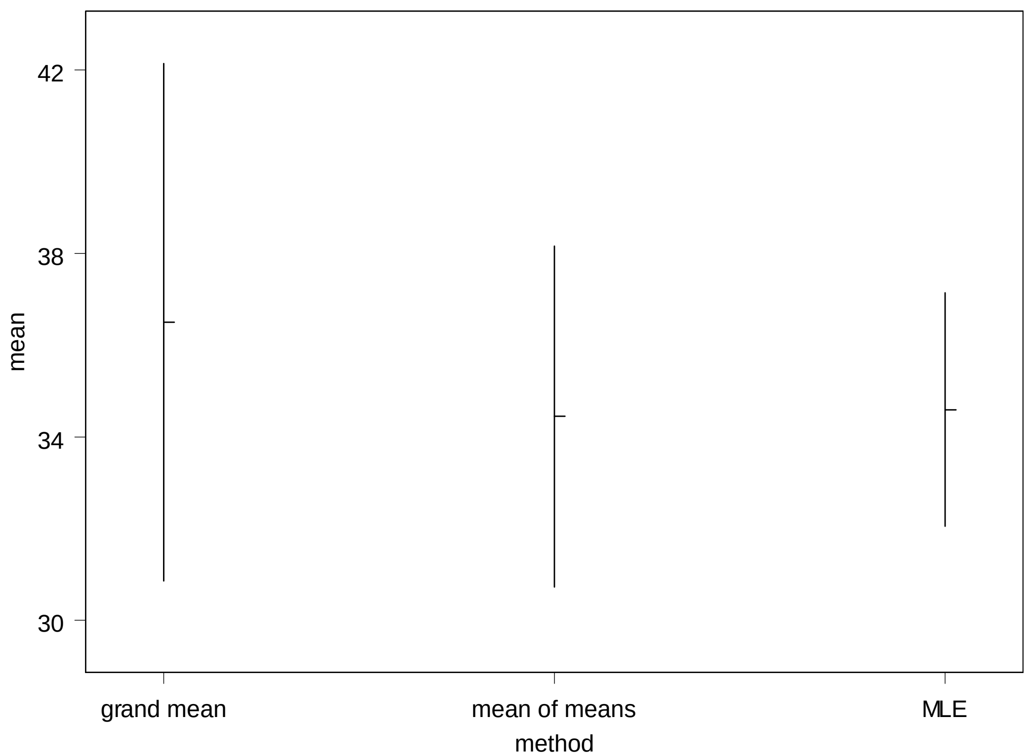
2.Number of labs >5.

Advantages:
The assumptions are the least restrictive so far and thus more likely to be met and produce accurate results.

Disadvantages:
More computationally demanding.

Summary of the Results for PCB 101:

Method	Consensus Mean	95% CI
Grand Mean	36.50	(30.86, 42.14)
Mean of Means	34.45	(30.73, 38.16)
MLE	34.59	(32.05, 37.14)



Notes:

1. The true lab means and standard deviations do not appear to be equal, the standard deviations are not all small, and the sample sizes vary a lot. These facts argue against the use of the Grand Mean and the Mean of Means methods.

The MLE is the preferred method, but the sample sizes and the number of laboratories are not large so that the asymptotic formula used to estimate consensus uncertainty may be misestimating the uncertainty size.

Solution 4.
A Bayesian Solution:

Classical:

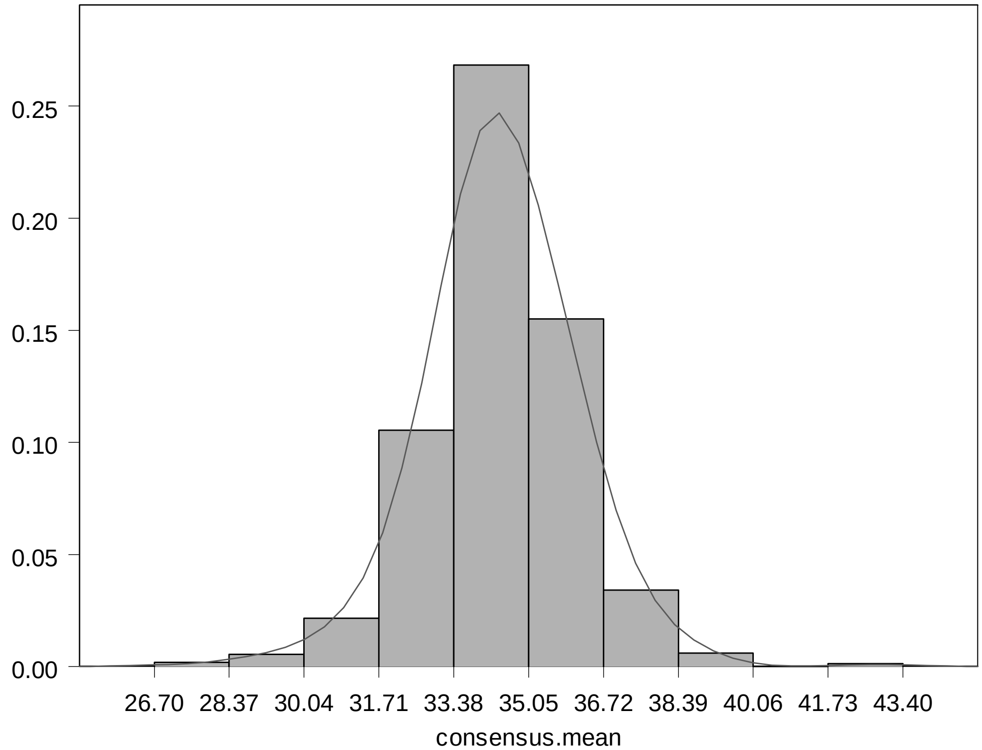
Parameter μ is fixed.
Data are random.

Bayes:

Data (once observed) are fixed.
Parameter μ is random.

Consensus mean μ has a probability distribution.

Before data - prior distribution.
After data - posterior distribution.
Plot of the posterior distribution for PCB 101:



Estimate μ by the posterior mean (34.33).

Estimate the consensus uncertainty by the posterior stdev. (0.8417) .

Advantages of the Bayesian formulation.

1. It enables us to make constructive use of expert opinion via the prior distribution.
2. It allows for rigorous incorporation of known physical constraints (e.g. $\mu > 0$) via the prior distribution.
3. It is better at handling complicated problems than the classical methods.
4. It can be employed to incorporate Type B error, even in complicated problems.
5. It allows naturally for successive updating of estimates upon the introduction of new data.

Disadvantages of the Bayesian formulation.

The prior distribution can be hard to specify because:

1. There may not be reliable expert opinion, or previous data experience.
2. There may be too much expert opinion which needs to be reconciled.
3. There may be many other parameters (nuisance) for which we need a prior distribution.

The posterior distribution can be hard to compute.

Specifying the prior distribution.

1. Using expert opinion. An expert may be able to give a range of possible values of μ with a probability distribution.

2. Using past data. That is, data from a related experiment can be used to give a mean and standard deviation for μ. Then a standard distribution such as the Gaussian can be used for the prior.

3. Using a so called “non-informative”, “vague” or “objective” prior. This models our ignorance about the parameter by assigning equal probability to values within some (usually large) interval. (Uniform distribution)

We will get back to the construction of priors later.

Now, more on the mechanics of Bayesian Statistics.

SECTION 2 – BAYESIAN STATISTICS 101

2.1 PROBABILITY

a. Definitions
b. Conditional Probability
c. Law of Total Probability
d. Bayes' Rule

2.2 MODELS FOR PROPORTIONS

a. Likelihood and Posterior Probabilities
b. Choice of a Prior Density
c. An Example
d. Comparing Two Proportions

2.3 MODELS FOR MEANS

a. Prior Densities and Normal Models
b. Comparing Two or More Means

2.1 PROBABILITY

Definition 1: Probability P(A) is a measure of the chance that an event A will happen.

Definition 2: Sample space S is the collection of all possible outcomes of an experiment.

Basic Properties:

1. $0 \leq P(A) \leq 1$.
2. $P(S) = 1$.
3. $P(\emptyset) = 0$.
4. $P(A)=1-P(\sim A)$
5. If A and B have no outcomes in common then $P(A \cup B) = P(A) + P(B)$.

Example 1: Throw a six-sided die.

Sample space $S = \{1, 2, 3, 4, 5, 6\}$

Event A: throw a 5,
Event B : throw a 6,
Event C: throw an even number.

Probability of A : $P(A) = P(B) = 1/6$.
Probability of C : $P(C) = 1/2$.
Probability of AUB : $P(A \cup B) = 1/3$
Probability of AUC : $P(A \cup C) = 2/3$.

Interpretations of Probability:

1. Long – run Frequency

Definition: The long-run frequency of an event is the proportion of time it occurs in a long sequence of trials.

Example 1, die tossing, is a good illustration of this.

2. Degree of Belief

Definition: A probability based on degree of belief is a subjective assessment of whether an event in question will occur.

Example 2: Weather forecasting.

Let T be the highest temperature that will occur outdoors tomorrow at 321 Penwood Drive.

We can assign probabilities to such events as:

Event A is that T   55˚ F

The associated probability is

$$P(A) = 0.3$$

Conditional Probability:

In Example 1, suppose that you know that the outcome must be an even number.

Conditionally on this fact the new sample space = {2,4,6}

Conditionally on this fact, the probability of getting a 6 is:

$$P(B|C) = 1/3.$$

This is called conditional probability .

Definition:
The conditional probability of B given A is

$$P(B|A) = P(A \quad B)/ P(A),$$

where $P(A \quad B)$ is the joint probability that both A and B occur.

Multiplication Rule:

$$P(A \quad B) = P(A)\, P(B|A).$$

Independence of events:

A and B are independent if $P(A|B) = P(A)$ or $P(B|A) = P(B)$.

For independent events,

$$P(A \quad B) = P(A)\, P(B).$$

In Example 1:
Event B : throw a 6,
Event C: throw an even number.

B   C is the event that 6 has occurred and an even number has occurred.

This is the set {6} and so P(B   C) = 1/6.

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{1/6}{1/2} = 1/3.$$

Example 3: Test a randomly selected subject's blood to determine whether infected by a disease.

Event A = subject is infected
Event B = test is positive

P(A) = probability of subject being infected.
(can be estimated based on the proportion of the population that is infected)

P(B|A) = probability of positive test result given that the subject is infected.
P(B|~A) = probability of positive test given subject not infected.

(both generally known by manufacturer of the test)

$P(A|B)$ = probability of subject being infected given a positive test result.

(Unknown - Main quantity of interest)

By definition:

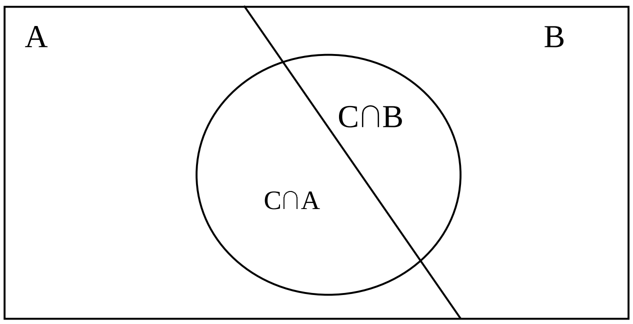
$$\boxed{\begin{aligned} P(A|B) &= P(A \quad B)/\ P(B) \\ &= P(B|A)\ P(A)\ /\ P(B)\ \ *** \end{aligned}}$$

We know: $P(A)$ and $P(B|A)$, $P(B|{\sim}A)$, and $P({\sim}A) = 1 - P(A)$.

We do not know: $P(B)$

*** This is the <u>Bayes Rule</u>.

LAW OF TOTAL PROBABILITY



$$P(C) = P(C \quad A) + P(C \quad B)$$
$$= P(C|A)\,P(A) + P(C|B)\,P(B)$$

Applying the Law of Total Probability to Example 3:

$$P(B) = P(B|A)\ P(A) + P(B|\sim A)\ P(\sim A)$$

So

$$\boxed{P(A|B) = P(B|A)\ P(A) \,/\, (P(B|A)\ P(A) + P(B|\sim A)\ P(\sim A))}$$

This result is referred to as the expanded form of the <u>Bayes' Rule</u>.

In Example 3:

Recall that

Event A = subject is infected

Event B = test is positive.

Let

$P(A) = 0.2$

$P(B|A) = 0.9$

($1 - P(B|A) = 0.1$ is the false negative rate)

$P(B|\sim A) = 0.05$

(the false positive rate)

$$
\begin{aligned}
P(B) &= P(B|A)P(A) + P(B|\sim A)P(\sim A) \\
&= 0.9\ (0.2) + 0.05\ (0.8) = 0.22
\end{aligned}
$$

$$
\begin{aligned}
P(A|B) &= P(B|A)\ P(A) / P(B) \\
&= 0.9(0.2)/0.22 = 0.82
\end{aligned}
$$

How do we apply BAYES THEOREM in interlab experiments ?

Event A - observe data Y
Event B - consensus mean
= some particular value

Wish to obtain P(B|A) = P(|Y)
Posterior distribution

Need : P(Y|) *Likelihood function*
P( ) *Prior distribution*

To apply Bayes Theorem:
P(|Y) = P(Y|) P( ) / P(Y)

Classical statistical models use only the function P(Y|) .

This can be used to produce statements such as:

Given that the true value of the consensus mean is 34 and that the standard deviation is 1.0, the probability of observing a measurement between 32 and 36 is 0.95.

Given data, this can be inverted into a confidence interval which enables us to say:

We are 95% confident that the true value of the consensus mean lies between 31 and 37.

Unfortunately, this is not a true probability statement.

Bayesian models use Bayes Theorem to obtain P(|Y) which enables us to say :

Given the observed measurements Y, the probability that the true value of the consensus mean is between 31 and 37 is 0.95.

This is a true probability statement.

We will now turn to the simpler situation of models for proportions to fully explain the concepts of prior distributions, likelihood functions and the use of Bayes Result to obtain posterior distributions.

2.2 MODELS FOR PROPORTIONS

Example 4: Cigarette Safety

Experiment to study how a cigarette causes ignition by transferring enough heat to fabric.

Two types of cigarettes:

low air permeability (# 529)

conventional air permeability (#531)

Data: proportion of ignitions

Cigarette						

Objective:

Compare the responses of the two types of cigarettes for the three types of substrate.

Let p_1 = probability that #529 ignites 10 layers (1993),

P_2 = probability that #531 ignites 10 layers (1993).

Classical Analysis:

Calculate a 95% confidence interval for p_1 and p_2 based on the sample proportions $\hat{p}_1 = {}^{0}\!/_{16}$, $\hat{p}_2 = {}^{16}\!/_{16}$, i.e.

$$\hat{p}_i - 1.96\sqrt{\frac{\hat{p}_i\left(1-\hat{p}_i\right)}{16}} \le p_i \le \hat{p}_i + 1.96\sqrt{\frac{\hat{p}_i\left(1-\hat{p}_i\right)}{16}}$$

We obtain:

$0.005728 \le p_1 \le 0.240736$

$0.759264 \le p_2 \le 0.994272$

To make a more direct comparison between the two proportions we can compute a 95% confidence interval for the difference $p_1 - p_2$:

$$\hat{p}_1 - \hat{p}_2 - 1.96\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{16} + \frac{\hat{p}_2(1-\hat{p}_2)}{16}}$$
$$\leq p_1 - p_2$$
$$\leq \hat{p}_1 - \hat{p}_2 + 1.96\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{16} + \frac{\hat{p}_2(1-\hat{p}_2)}{16}}$$

This interval is an approximation which in this case, due to the extreme values of p_1, p_2, does not work very well. In fact the classical 95% CI says that $p_1 - p_2 = -1$. (There are other forms of the classical CI that we could use see p. 229 of “An Introduction to Mathematical Statistics” by Larsen and Marx)

For Bayesian analysis:

Event A - observe data

Event B - p_1 and p_2 equal some particular values.

Wish to obtain: P(B|A).

Need:

1. prior probabilities for p_1 and p_2, i.e. P(B)

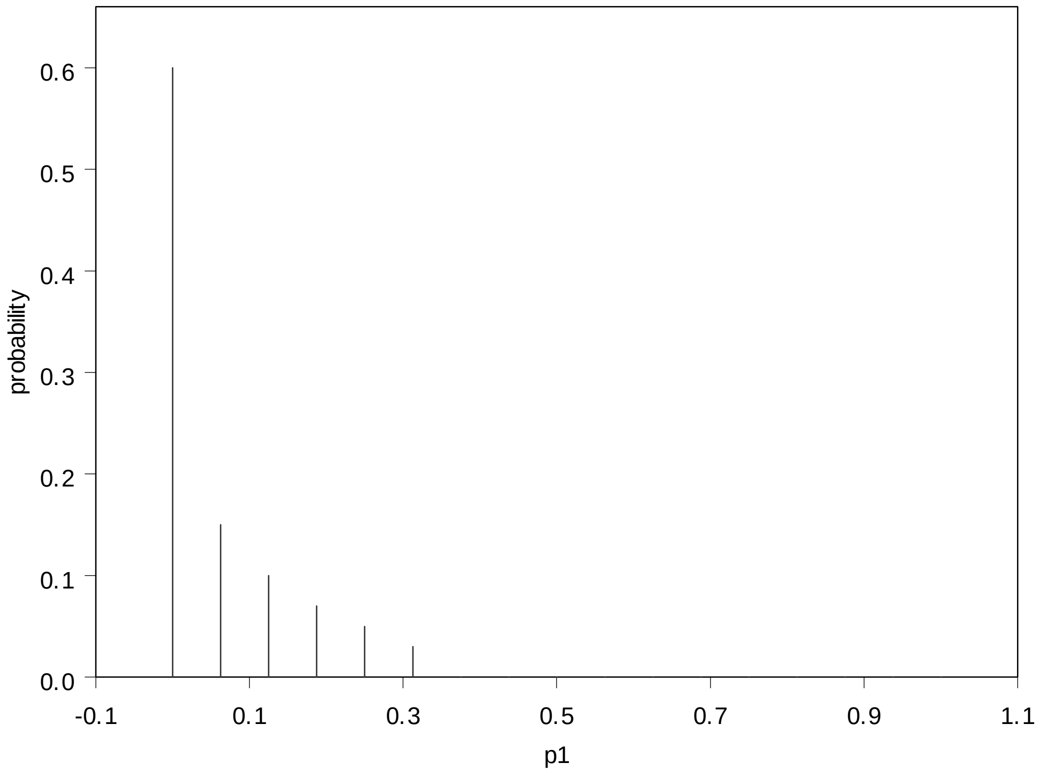
2. P(A|B) called the likelihood function.

Prior Distribution for p_1:

Even though p_1 can have one of infinitely many values between 0 and 1, we can make its range discrete. An example of a possible prior distribution is:

Value of p_1	Probability
0	0.6
0.0625	0.15
0.125	0.1
0.1875	0.07
0.25	0.05
0.3125	0.03
0.375	0
0.4375	0
0.5	0
0.5625	0
0.625	0
0.6875	0
0.75	0
0.8125	0
0.875	0
0.9375	0

1	0



The mean of a discrete distribution can be thought of as its center of gravity.

It is calculated as:

$$Mean = \sum_{all\ \pi} \pi \, P(p = \pi)$$

Mean of p_1 = 0.0625 (0.15) + 0.125 (0.1)

+ 0.1875 (0.07) + 0.25 (0.05)

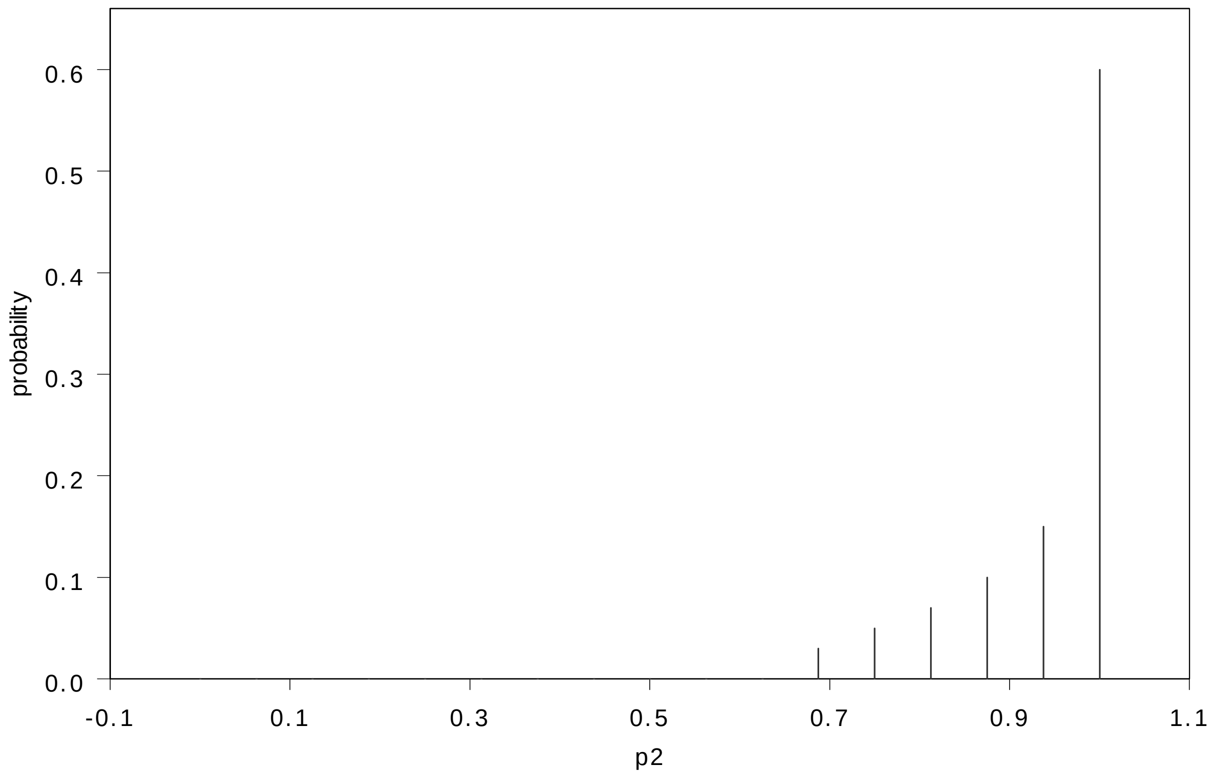
+ 0.3125 (0.03) = 0.057

Consider p_2 , possible prior distribution:

Value of p_2	Probability
0	0
0.0625	0
0.125	0
0.1875	0
0.25	0
0.3125	0
0.375	0
0.4375	0
0.5	0
0.5625	0
0.625	0
0.6875	0.03
0.75	0.05
0.8125	0.07
0.875	0.1
0.9375	0.15
1	0.6

Mean of p_2 = 0.9431

The prior distribution of p_2 :



Likelihood Function:

The Likelihood of a model is the probability of the data occurring, calculated assuming that model.

For cigarette # 529,

data = 0/16 (ignitions/ trials)

For cigarette # 531,

data = 16/16

We can obtain the likelihood values by assuming a binomial distribution.

If x = number of ignitions

P(data = $\frac{x}{16}$| p = π) =

$$\frac{16!}{x!\,(16-x)!}\;\pi^{x}\,(1-\pi)^{16-x}$$

The use of this distribution is justified by assuming that for each of the 16 cigarettes, the probability of ignition is some fixed number π, and that the fact that one cigarette ignites has no effect on the ignition of the next cigarette. (identical and independent trials)

We obtain for #529:

P(data= 0/16| p_1 = 0.0000) = 1.0

P(data= 0/16| p_1 = 0.0625) = 0.3561

P(data= 0/16| p_1 = 0.125) = 0.1181

P(data= 0/16| p_1 = 0.1875) = 0.0361

P(data= 0/16| p_1 = 0.2500) = 0.01

P(data= 0/16| p_1 = 0.3125) = 0.0025

P(data= 0/16| p_1 = 0.375) = 0.0005

P(data= 0/16| p_1 = 0.4375) = 0.0001

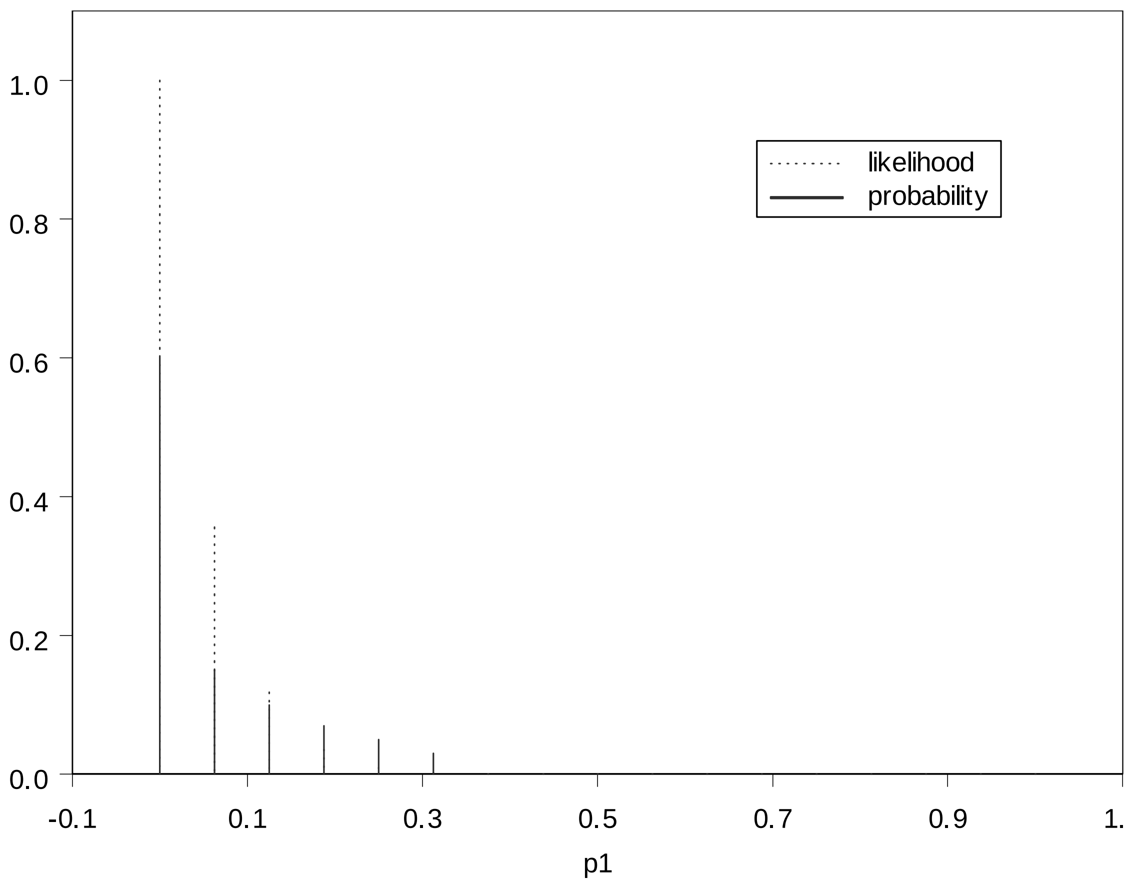
P(data= 0/16| p_1 = 0.5000) = 0.0000

P(data= 0/16| p_1 0.0625) = 0.0

In table form:

Value of p_1	Likelihood
0	1
0.0625	0.3561
0.125	0.1181
0.1875	0.0361
0.25	0.01
0.3125	0.0025
0.375	0.0005
0.4375	0.0001
0.5	0
0.5625	0
0.625	0
0.6875	0
0.75	0
0.8125	0
0.875	0
0.9375	0
1	0

The likelihood and the prior plotted on the same graph display the difference between our prior belief and the evidence given by the data:



The likelihood is much more concentrated on the values 0 to 0.1.

For cigarette #531:

In table form:

Value of p_2	Likelihood
0	0
0.0625	0
0.125	0
0.1875	0
0.25	0
0.3125	0
0.375	0
0.4375	0
0.5	0
0.5625	0.0001
0.625	0.0005
0.6875	0.0025
0.75	0.01
0.8125	0.0361
0.875	0.1181
0.9375	0.3561
1	1

To calculate the posterior:

Use Bayes' Rule:

$$\text{Posterior} = \frac{PRIOR \quad X \quad LIKELIHOOD}{P(Data)}$$

Where for example for p_1:

$$P(Data = {}^{0}\!/_{16}) = P\left(Data = {}^{0}\!/_{16} \mid p_1 = 0\right)P(p_1 = 0)$$
$$+ \ldots + P\left(Data = {}^{0}\!/_{16} \mid p_1 = 0.4375\right)P(p_1 = 0.4375) =$$
$$0.66833$$

.

We obtain for p_1:

Value of p_1	Prior	Likeli-hood	Prior X Likelihood	Poste-rior
0	0.6	1.0	0.6	0.898
0.0625	0.15	0.3561	0.0534	0.08
0.125	0.1	0.1181	0.01181	0.017
0.1875	0.07	0.0361	0.002527	0.004
0.25	0.05	0.01	0.00051	0.0007
0.3125	0.03	0.0025	0.000075	0.0001
0.375	0	0.0005	0	0
0.4375	0	0.0001	0	0
0.5	0	0	0	0

We obtain for p_2:

Value of p_2	Prior	Likeli-hood	Prior X Likelihood	Poste-rior
1.0	0.6	1.0	0.6	0.898
0.9375	0.15	0.3561	0.0534	0.08
0.875	0.1	0.1181	0.01181	0.017
0.8125	0.07	0.0361	0.002527	0.004
0.75	0.05	0.01	0.00051	0.0007
0.6875	0.03	0.0025	0.000075	0.0001
0.625	0	0.0005	0	0
0.5625	0	0.0001	0	0
0.5	0	0	0	0

Here P(data = 16/16) =0.66833.

To compare the three distributions for p_1: