

Introduction to Bayesian Statistics

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

by
Knut Helge Jensen

T. Bayes.



Thomas Bayes 1701 - 1761.

Most known for Bayes' Theorem which Bayesian statistical inference is based on. This work was published after his death by Richard Price (1763).

What is statistical inference?

Definition: Statistical inference is the process of drawing conclusions about population parameters based on a sample taken from the population.

Inferential statistics can be seen as a contrast to descriptive statistics which only is concerned with properties of the sampled data. In inferential statistics, where data is assumed to represent a larger population, the aim is to derive knowledge at the level of the population rather than the sampled data.

A very short history of statistical inference

Bayesian statistical inference



Thomas Bayes (1701 - 1761)

- Bayes' Theorem



Stanislaw Ulam (1901 - 1984)



Nicolas Metropolis (1915 - 1999)



Wilfred Keith Hastings (1930 - 2016)

- Markov Chain Monte Carlo methods (MCMC)

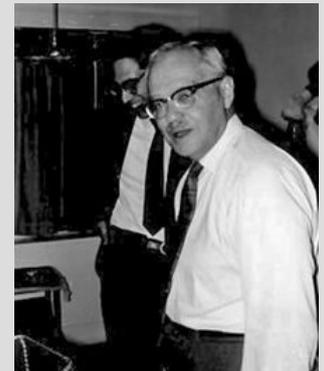
Frequentist statistical inference



Ronald A. Fisher (1890 - 1962)

- Maximum Likelihood
- ANOVA

...the theory of inverse probability is founded upon an error, and must be wholly rejected¹



Jerzy Neyman (1894 - 1981)

- Confidence interval

¹ Fisher (1925) on page 10 in *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh. Note! At that time Bayesian analysis was called inverse probability.

General probability rules and Bayes' theorem

Case study: A crab species appears in the colours green and red where the tendency to become red is related to age of the crab. The crab is sometimes infested with a parasite. The parasite has the two stages immature and adult where the adult stage is more common to find on old and red crabs.

We look at frequencies of crab colours and parasite stages to recall some probability rules and to explain Bayes' theorem.

Frequencies of colours of infected crabs and parasite stages.

Crab Colour	Parasite Stage		TOTAL
	Immature	Adult	
Green	151	29	180
Red	39	136	175
TOTAL	190	165	355

$P(A)$ and $P(B)$

Crab Colour	Parasite Stage		TOTAL
	Immature	Adult	
Green	151	29	180
Red	39	136	175
TOTAL	190	165	355

Let us use this table to define two events we want to work with:
Event A is that crab colour is green, and event B is that the parasite stage is Immature.

What is $P(A)$ and $P(B)$?

$$P(A) = 180/355 = 0.5070423$$

$$P(B) = 190/355 = 0.5352113$$

Joint probability: P(A and B)

Crab Colour	Parasite Stage		TOTAL
	Immature	Adult	
Green	151	29	180
Red	39	136	175
TOTAL	190	165	355

Note!

In statistical textbooks P(A and B) is often written as $P(A \cap B)$.

What is $P(A \text{ and } B)$?

Following the definition of event A and B, we can write it as:

$$P(\text{Crab}=\text{Green and Parasite}=\text{Immature}) = 151/355 = 0.4253521$$

Obviously, $P(\text{Crab}=\text{Green and Parasite}=\text{Immature})$, is the same as $P(\text{Parasite}=\text{Immature and Crab}=\text{Green})$. This gives the following rule:

$$P(A \text{ and } B) = P(B \text{ and } A) \quad (\text{Eq. 1})$$

Joint probability and independence

If the events Colour = Green and Stage = Immature is independent of each other, the joint probability is: $P(A) \times P(B)$.

Do we have independence in our case?

Crab Colour	Parasite Stage		TOTAL
	Immature	Adult	
Green	151	29	180
Red	39	136	175
TOTAL	190	165	355

$P(\text{Colour} = \text{Green}) \times P(\text{Stage} = \text{Immature})$

$$= (180/355) \times (190/355) = 0.2713747$$

The answer is different from the probability we calculated in the previous slide (0.4253521). Hence, A and B are not independent events.

Conditional probability

If the events Colour = Green and Stage = Immature is dependent of each other, we follow the rule for conditional probability:

$P(A|B)$, which means the probability of event A given event B.

If A and B are NOT independent events, we get:

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

This equation can be restructured to:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \quad (\text{Eq. 2})$$

The conditional probability $P(\text{Crab}=\text{Green} \mid \text{Parasite}=\text{Immature})$ is:

$$0.4253521/0.5352113 = 0.7947368$$

Bayes' Theorem

Recall from the previous slide that the following two equations hold:

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

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

From *Eq. 1* you learned that $P(A \text{ and } B) = P(B \text{ and } A)$. Thus, *Eq. 2* from the previous slide can replace the right side numerator from

$$P(A|B) \times P(B) \quad \text{to become} \quad P(B|A) \times P(A)$$

which gives:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (\text{Eq. 3})$$

This is Bayes' theorem!

Is Bayes' theorem correct?

	Parasite Stage		
Crab Colour	Immature	Adult	TOTAL
Green	151	29	180
Red	39	136	175
TOTAL	190	165	355

$$P(\text{Colour} = \text{Green} | \text{Stage} = \text{Immature}) = \frac{P(\text{Stage} = \text{Immature} | \text{Colour} = \text{Green}) \times P(\text{Colour} = \text{Green})}{P(\text{Stage} = \text{Immature})}$$

$$= \frac{\frac{151}{180} \times \frac{180}{355}}{\frac{190}{355}} = 0.7947368$$

Yes, it gives the same answer as calculating directly from the table:
 $151/190 = 0.7947368$

How Bayes' theorem are used in statistical inference

Obviously - the calculation $151/190$ is a shorter way to find the answer.

But the reason for using Bayes' theorem is the opportunities it provides for statistical inference.

We want to know something about parameters of a population and we may use Bayes' theorem as follows:

$$P(\beta|data) = \frac{P(data|\beta) \times P(\beta)}{P(data)}$$

Three reasons to use a Bayesian inference approach

1) Outputs are more in accordance with what you really are after.

In frequentist statistics:

$$P(\text{data} \mid \beta)$$

Hence, in null hypothesis testing or likelihood analysis you are interested in the probability of observing the data given the parameter(s).

It is more intuitive to ask for the probability of the parameter(s) given the data:

$$P(\beta \mid \text{data})$$

which is exactly what you do when using Bayesian statistics.

Frequentist
inference

Bayesian
inference

As a consequence of this difference, $P(\text{data} \mid \beta)$ versus $P(\beta \mid \text{data})$, the interpretation of a 95% confidence interval (e.g. 29.2, 32.4) in the frequentist world must NOT be interpreted as:

$$P[29.2 < \mu < 32.4] = 0.95.$$

This expression does not contain any random variable within the brackets, so we cannot speak about a probability.

However, you may say that you are 95% confident that the true mean lies within the given interval.

This means the following: If 95% confidence intervals are constructed *in an infinite number of independent experiments*, 95% of those intervals will contain the true value of the parameter.

Credible interval

In practice, we do not repeat an experiment many times. This means that the statement about p-values and confidence intervals in the frequentist world are based on some fictive data.

In the Bayesian world the interpretation is more in accordance to what we would like to say and do not depend on some fictive data: There is a 95% probability that the parameter β is within the given interval.

This analog to a confidence interval in Bayesian statistics is called the **credible interval**.

2) Possible to incorporate prior knowledge into the models.


$$P(\beta|data) = \frac{P(data|\beta) \times P(\beta)}{P(data)}$$

The posterior
(what you are after)

The likelihood

The Prior

The normalising constant

3) Sometimes only Bayesian based models give answers if you have a design with clustered data (random effect factors) and a non-normal response variable, i.e. when dealing with GLMM problems.

Examples are given in e.g. Zuur and Ieno (2016).

Many scientists automatically think about Bayesian based models when they see designs with random effect factors (clustered data) and non-normal response variables.

My guess is that we will see much more use of Bayesian analysis in the future - not only for GLMM problems.

Why NOT use a Bayesian inference approach?

- Bayesian inference statistics often requires much computing power.
- It is tradition to use frequentist approaches. Your collaborators and reviewers of scientific journals may not know Bayesian statistics good enough to evaluate the validity of your results.
- You want simple answers - reject H_0 or not, i.e. you and the rest of the scientific world want p-values.
- Frequentist analysis methods are more available in statistical packages than Bayesian methods.

A closer look on how to use Bayes' theorem for statistical inference

$$P(\beta|data) = \frac{\overset{\text{The likelihood}}{P(data|\beta)} \times \overset{\text{The Prior}}{P(\beta)}}{\underset{\text{The normalising constant}}{P(data)}}$$

The posterior
(what you are after)

$P(data)$ is a scaling factor so that the posterior distribution adds up to 1. It is not dependent on β , and therefore the equation is often written like:

$$P(\beta|data) \propto P(data|\beta) \times P(\beta)$$

This symbol means proportional to

Markov Chain Monte Carlo (MCMC)

$$P(\beta|data) = \frac{\overset{\text{The likelihood}}{P(data|\beta)} \times \overset{\text{The Prior}}{P(\beta)}}{\underset{\text{The normalising constant}}{P(data)}}$$

The posterior
(what you are after)

It is often difficult or impossible to calculate the normalising constant, which makes it difficult to calculate the posterior density directly.

However, there is a way around the problem, and it is called MCMC, which is a simulation method. The method is computer intensive and this is the reason why Bayesian inference has become popular in the age of computers.

How to do Bayesian statistics in R?

As always, there are many possibilities in R. To do it properly, R depends on some external software to do the MCMC.

Three commonly used external MCMC engines with good R integration are: JAGS (Plummer 2003), STAN (Gelman et al. 2015) and NIMBLE (de Valpine et al. 2017).

You may read more about them here:

JAGS: <http://mcmc-jags.sourceforge.net/>

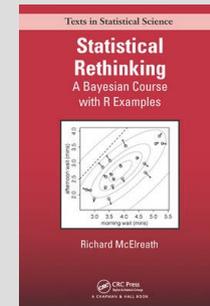
STAN: <http://mc-stan.org/>

NIMBLE: <https://r-nimble.org/>

I have chosen STAN of three main reasons:

1) It is more flexible and works faster than JAGS (Bürkner 2017, and references therein).

2) It is the software used in the book *Statistical Rethinking* by Richard McElreath. This book came in 2016 and it has already become a classic for those who want to learn Bayesian statistics.



3) The brms package for R use STAN and allows the user to benefit from the merits of STAN only by using simple, lme4-like formula syntax.

Case study: Recall the crab data from the table first shown on slide 5.

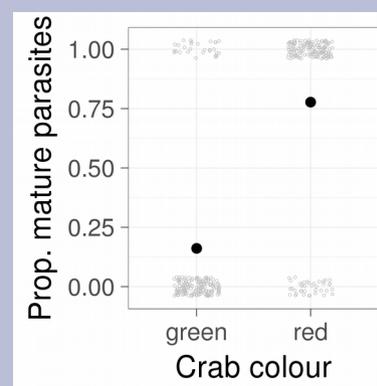
Crab Colour	Parasite Stage		TOTAL
	Immature	Adult	
Green	151	29	180
Red	39	136	175
TOTAL	190	165	355

You want to do a Bayesian analysis for the probability of finding a mature parasite depending on colour of the infected host crab.

However, let's assume that you have a clustering variable (random effect factor) called Site representing the location where each infected crab has been caught. Thus, you get a dataset for individual crabs that looks like the one to the right.

Site	Crab.colour	Parasite.stage
a	green	0
h	red	1
b	red	1
c	green	0
h	green	0
j	red	1
g	green	1
etc.	etc.	etc.

R syntax



#Importing dataset

```
crab.df <- read.table("https://folk.uib.no/nzlkj/data/crabdata.txt",  
header=T, sep=",")
```

#Plot

```
library(ggplot2)  
p1 <- ggplot(crab.df, aes(Crab.colour, Parasite.stage))  
p1 <- p1 + geom_jitter(width=0.2, height=0.04, shape=1, colour="gray")  
p1 <- p1 + theme_bw(base_size=20)  
p1 <- p1 + stat_summary(fun.y=mean, geom="point", shape=19, size=6)  
p1 <- p1 + labs(x="Crab colour", y="Prop. mature parasites")  
p1
```

#Performing the Bayesian analysis:

```
library(brms) # 1 Needs installation + RStan etc.  
fit1.brm <- brm(Parasite.stage ~ Crab.colour + (1|Site),  
family=bernoulli, data=crab.df)
```

¹ See slide 29

R output:

summary (fit1.brm)

```
Family: bernoulli
Links: mu = logit
Formula: Parasite.stage ~ Crab.colour + (1 | Site)
Data: crab.df (Number of observations: 355)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
        total post-warmup samples = 4000
ICs: LOO = NA; WAIC = NA; R2 = NA
```

Group-Level Effects:

~Site (Number of levels: 9)

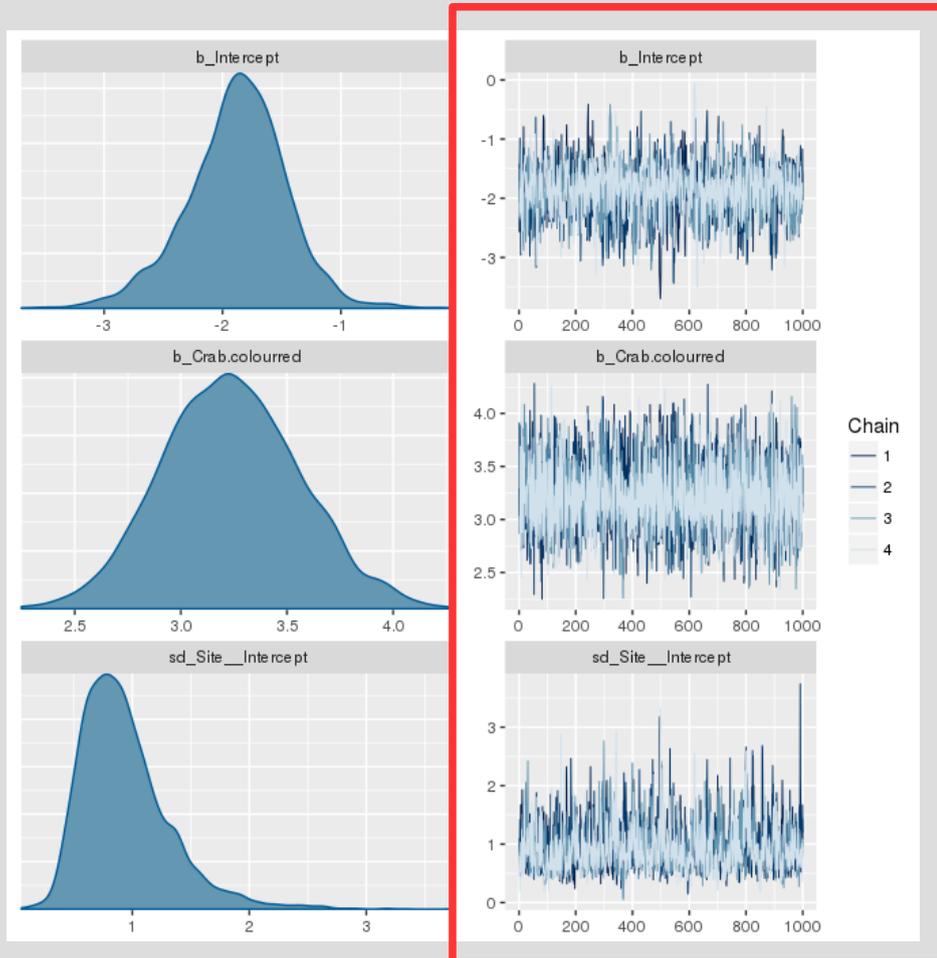
	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
sd(Intercept)	0.95	0.40	0.41	1.96	884	1.00

Population-Level Effects:

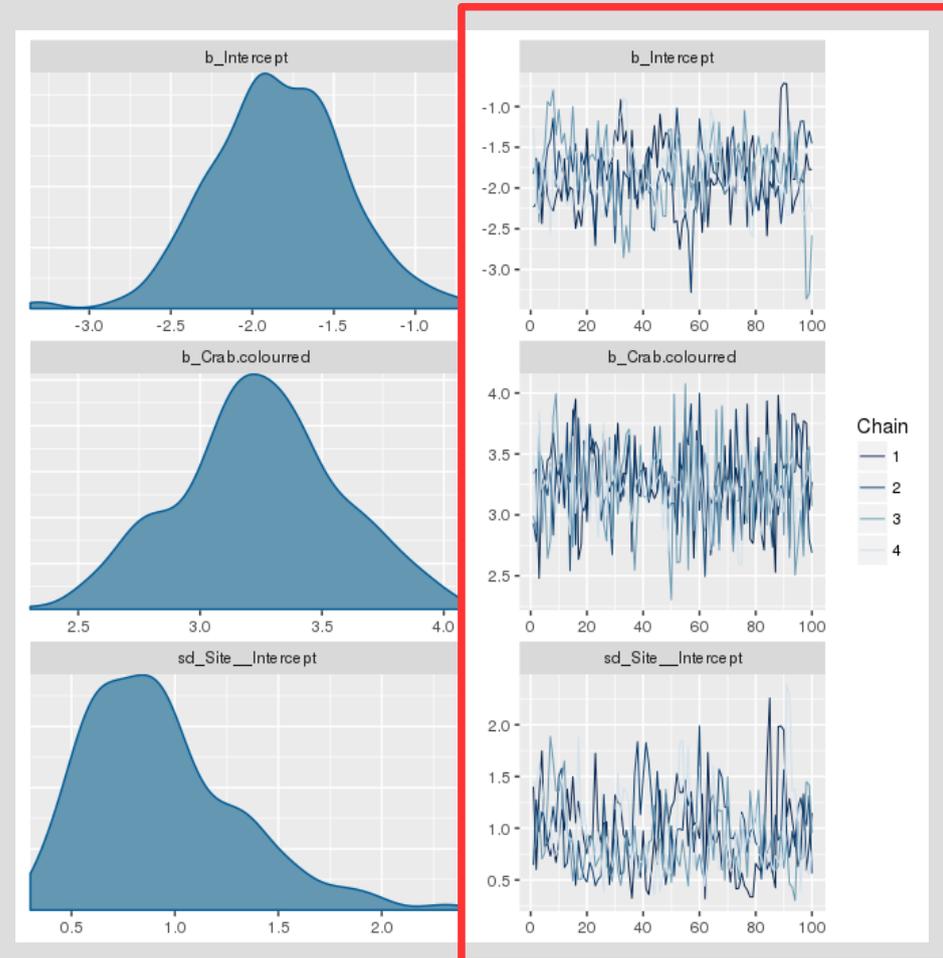
	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
Intercept	-1.86	0.41	-2.73	-1.11	1358	1.00
Crab.coloured	3.23	0.31	2.65	3.87	3512	1.00

Need to assess mixing:

`plot(fit1.brm)`



Example of bad mixing:



But what is mixing and why do we check it?

We do this to evaluate whether we can trust the MCMC iterations. If the iterations flip around an equilibrium then Bayesian theory dictates that the iterations converge to the real distribution (Zuur et al 2017).

Good mixing



www.alamy.com - BFAFH3

Bad mixing



© Thomas Jaehndel via CTBUH

The same model by glmm using lme4:

```
#Performing the Bayesian analysis:
```

```
library(lme4)
```

```
fit1.glmm <- glmer(Parasite.stage ~ Crab.colour + (1|Site),  
family=binomial, data=crab.df)
```

```
summary(fit1.glmm)
```

```
...
```

```
Fixed effects:
```

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.8273	0.3206	-5.70	1.2e-08	***
Crab.coloured	3.1644	0.3088	10.25	< 2e-16	***

For comparison - the corresponding output from the Bayesian analysis:

```
Population-Level Effects:
```

	Estimate	Est.Error	1-95% CI	u-95% CI	Eff.Sample	Rhat
Intercept	-1.86	0.41	-2.73	-1.11	1358	1.00
Crab.coloured	3.23	0.31	2.65	3.87	3512	1.00

How to install required software

You need to install two libraries in R called rstan and brms. RStan is the R interface to Stan. Stan is the module that performs MCMC sampling and is basically the engine you need for Bayesian analysis.

Before installing rstan and brms, there are some software that needs to be installed on your computer first. How to install these depends on your OS (Windows or Mac etc.). Just follow the link for your OS:

Windows:

<https://github.com/stan-dev/rstan/wiki/Installing-RStan-on-Windows>

Mac or Linux:

<https://github.com/stan-dev/rstan/wiki/Installing-RStan-on-Mac-or-Linux>

After following the link for your OS above, you install brms within R by:
`install.packages("brms")`, and you're done!

Don't want to or don't manage to install an external MCMC engine?

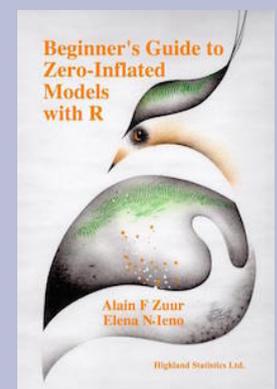
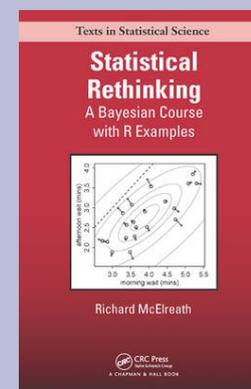
R has a solution for you as well:

You may install the MCMCglmm library by the command:
`install.packages("MCMCglmm")`

You'll find an intro about how to use this package at:

<https://github.com/tmalsburg/MCMCglmm-intro>

References and/or recommended books*



Bürkner P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80(1): 1-28.

Fisher R.A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

Gelman A., Lee D., Guo J. (2015). Stan: A Probabilistic Programming Language for Bayesian Inference and Optimization. *Journal of Educational and Behavioral Statistics* 40(5): 530-543.

*McElreath, Richard (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press. ISBN: 978-1-4822-5344-3.

Plummer M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*.

Price R. (1763). An Essay towards solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London* 53: 370-418.

de Valpine P., Turek D., Paciorek C.J., Anderson-Bergman C., Lang D.T. and Bodik R. (2017). Programming With Models: Writing Statistical Algorithms for General Model Structures With NIMBLE. *Journal of Computational and Graphical Statistics*: 26(2), 403-413.

*Zuur A.F. and Ieno E.N. (2016). *Beginner's Guide to Zero-Inflated Models with R*. Highland Statistics, Newburgh, UK. ISBN: 978-0-9571741-8-4.

Zuur A.F., Ieno E.N. and Saveliev A.A. (2017). *Beginner's Guide to Spatial, Temporal and Spatial-Temporal Ecological Data Analysis with R-INLA, Volume I*. Highland Statistics, Newburgh, UK. ISBN: 978-0-9571741-9-1.

A little exercise you can solve by using Bayes' theorem



You have been picked out from the audience of a game show called "Let's Make a Deal". You are placed in front of three doors and then the host of the game show says: "Behind one of the doors is a brand new Lamborghini Aventador. If you pick the right door, the Aventador is yours!"



You think that there is a $1/3$ probability of selecting the right door and say: "I want to open the first door!". Then the host of the game show, who knows where the car is, surprises you by opening door three to reveal that there is no car behind this door:



He then says: "Do you want to switch to door number two or do you want to keep door one?" – What should you do?