Bayesian approach to statistics

Gregor Gorjanc

University of Edinburgh / University of Ljubljana

2025-05-27

Sodobni statistični pristopi na doktorskem študiju Statistike

Ljubljana

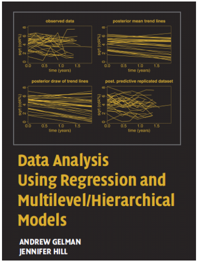
Provisional agenda

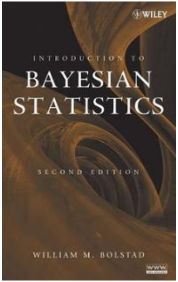
- Introduction
- Probability
- Likelihood
- Bayesian
- MCMC
- General Bayes programs/packages (demo)

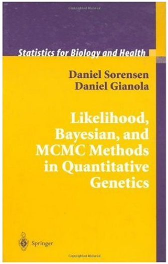
0. Introduction

- History (Laplace in Bayes)
- My experience (statistical/quantitative genetics)
- Terminology in Slovenian
  - Bayesova
  - Bayesovska
- Bayesian ="probabilistic modelling"

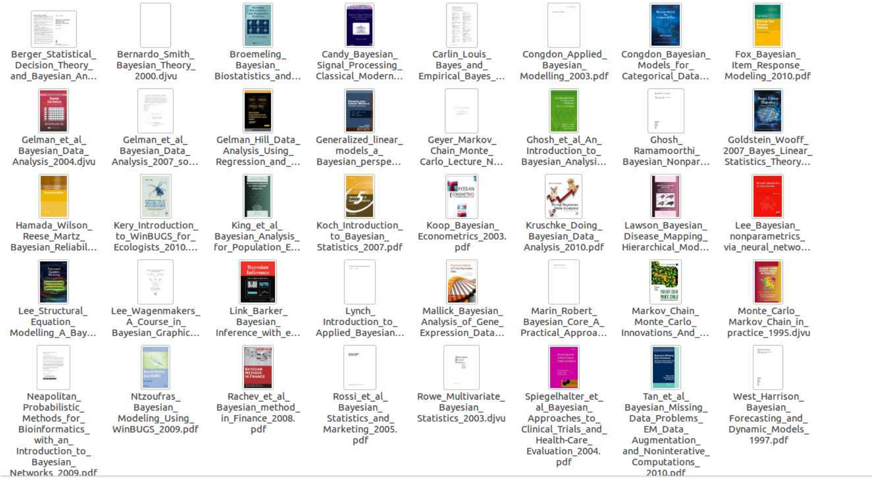
Literature (I studied from these sources)







Plenty of literature



Recommended

Richard McElreath

Anthropology, Evolutionary Ecology, Bayesian Data Analysis

HOME PUBLICATIONS SOFTWARE SR2 FOR INSTRUCTORS **STATISTICAL RETHINKING** TALKS TEACHING

Statistical Rethinking

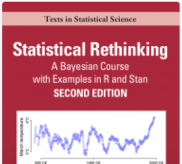
A Bayesian Course with Examples in R and Stan (& PyMC3 & brms & Julia too, see links below)

Most recent set of free lectures: Statistical Rethinking 2023

Second Edition

The second edition is now out in print. Publisher information on the CRC Press page. For more detail about what is new, look here.

Materials



2nd Edition

- Book: CRC Press
- Book sample: Chapters 1 and 2 (2MB PDF)
- Lectures and slides:

FOLLOW ME ON TWITTER

Tweets from @rlmcelreath

Richard McElr...
@. · Aug 16, 2023

I used to teach game thec
both undergrad & phd lev
One game I would do at s
is version of Keynesian
beauty contest: everyone
picks a number 0-100,
person closest to 2/3 of
average wins. Nash is 0. E
anyone choosing 0 loses,
the class aren't (yet) gam
theorists. >

100 6K

Richard McElr...
@ · Nov 30, 2022

Recommended (next level)

Michael Betancourt

Once and future physicist masquerading as an applied statistician.

- New York, NY
- Email
- Twitter
- GitHub

Part II: Modeling and Inference

Modeling and Inference
Generative Modeling
Prior Modeling
Introduction to Stan
Identifiability and Degeneracies
Principled Model Building Workflow

Part III: Modeling Techniques

Hierarchical Modeling
Factor Modeling
Modeling Sparsity
Modeling Ordinal Outcomes
Survival Modeling
Modeling Consensus HTML PDF
Variate-Covariate Modeling
Modeling Functional Behavior I: Linearized Models
Modeling Functional Behavior II: General Linearized Models
Underdetermined Linearized Regression
The QR Decompostion for Linearized Regression
Modeling Functional Behavior III: Gaussian Processes
Modeling Functional Behavior IV: Stochastic Differential Equations
Modeling Functional Behavior V: The Brownian Bridge

1. Probability

Probability

- Rolling a die
- Two views on probability:
  - probability to see six dots: $\Pr(6\,dots) = 1/6$
  - rolling a die and track frequency (= probability) of events
- Probability distribution, $\sum_i \Pr\left(i\ dots\right) = 1$:

No. dots	1	2	3	4	5	6
Probability	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

Exercise

- Start R
- Run this code

```
## Calculator
1 + 1
log(1)
## Rolling a die 10x
sample(x=1:6, size=10, replace=TRUE)
sample(x=1:6, size=10, replace=TRUE)
## ... 10.000x
x <- sample(x=1:6, size=10000, replace=TRUE)
t <- table(x)
p <- prop.table(x=t)
p
```

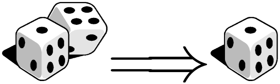
Joint probability

- Roll **two** dice
- Is no. of dots on the dies dependent? NO.
  –> Events are independent.
- Probability of joint event, to see 2x 6 dots:
  $\Pr(6\,dots, 6\,dots) = \Pr(6\,dots) \times \Pr(6\,dots) = {}^1/_6 \times {}^1/_6 = {}^1/_{36}$
- Probability distribution, $\sum_{i,j} \Pr(i\,dots, j\,dots) = 1$:

No. dots	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

Marginal probability

- Rolling **two** dice



- Probability distribution for 1, 2,. . . 6:

No. dots	1	2	3	4	5	6	Total
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
Total	1/6	1/6	1/6	1/6	1/6	1/6	1

$$\Pr(A) = \sum_B \Pr(A, B) \qquad p(a) = \int p(a, b)\, db$$

Exercise

- Run

```
## Joint probability
x1 <- sample(x=1:6, size=10000, replace=TRUE)
x2 <- sample(x=1:6, size=10000, replace=TRUE)
t <- table(x1, x2)
p <- prop.table(x=t)
p
## Marginal probability
rowSums(x=p)
colSums(x=p)
```

Joint probability II

- Rolling **one** die
- Is no. of dots on the die dependent on the no. being odd or even? YES.
  –> Events are dependent.
- Probability for an even number and 6 dots:

$$
\begin{aligned}
Pr(6\,dots, even\,no.) &= Pr(6\,dots|even\,no.) \times Pr(even\,no.)\\
Pr(6\,dots, even\,no.) &= {}^1/_6\\
Pr(even\,no.) &= {}^1/_2\\
Pr(6\,dots|even\,no.) &= \frac{Pr(6\,dots, even\,no.)}{Pr(even\,no.)} = \frac{{}^1/_6}{{}^1/_2} = {}^1/_3
\end{aligned}
$$

- $Pr(6\,dots|even\,no.)$ –> Conditional probability

Conditional probability

- Probability to observe 6 dots, conditionally on even no.: $\Pr(6\ \mathit{dots} | \mathit{even\ no.}) = 1/3$
- Probability distribution, $\sum_i \Pr\left(i\ \mathit{dots} | \mathit{odd\ no.}\right) = 1$:

Odd no.

No. dots	1	2	3	4	5	6
Probability	1/3	0	1/3	0	1/3	0

- Probability distribution, $\sum_i \Pr\left(i\ \mathit{dots} | \mathit{even\ no.}\right) = 1$:

Even no.

No. dots	1	2	3	4	5	6
Probability	0	1/3	0	1/3	0	1/3

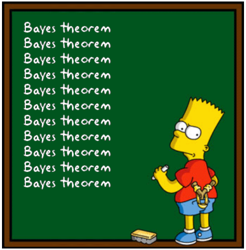
Bayes theorem

$$\begin{aligned}
\Pr(A, B) &= \Pr(A|B) \times \Pr(B) = \Pr(B|A) \times \Pr(A) \\
\Pr(A|B) &= \frac{\Pr(B|A) \times \Pr(A)}{\Pr(B)} \\
&= \frac{\Pr(B|A) \times \Pr(A)}{\sum_A \Pr(B|A) \times \Pr(A)} \\
\Pr(A|B) &\propto \Pr(B|A) \times \Pr(A)
\end{aligned}$$

Exercise

Frequency of diseased in population is0.008. There is a test for disease, with false positive rate of 10% and false negative rate of 5%.

- What is probability that an individual is diseased, given that they test positive?
- What is the impact of (apriori) information?
- What is probability that an individual is diseased, given that they test positive two times?



Exercise (solution)

Events:

- status: healthy, diseased
- test: negative, positive

Prior probability of being diseased without taking the test:

$$\Pr\left(\textit{diseased}\right) = 0.008$$

Test:

	Status	
Test	Healthy	Sick
Negative	$\Pr\left(\textit{negative} \mid \textit{healthy}\right) = 0.90$	$\Pr\left(\textit{negative} \mid \textit{diseased}\right) = 0.05$
Positive	$\Pr\left(\textit{positive} \mid \textit{healthy}\right) = 0.10$	$\Pr\left(\textit{positive} \mid \textit{diseased}\right) = 0.95$
Total	1	1

Exercise (solution) II

- Probability of being diseased given positive test?

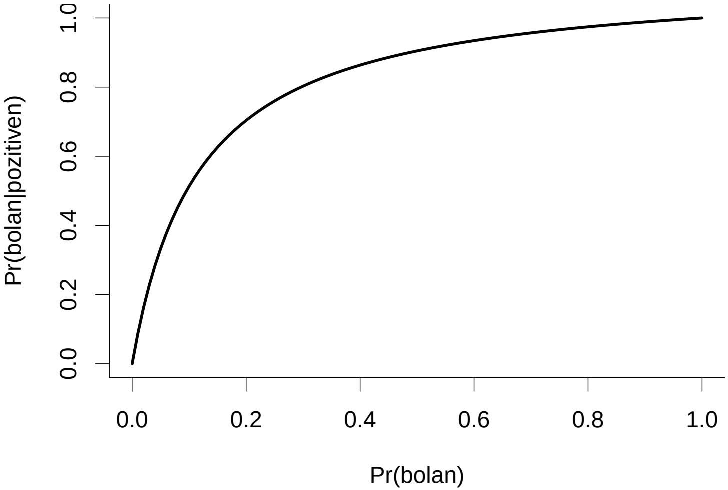
$$
\begin{aligned}
\Pr(\textit{diseased}|\textit{positive}) &= \frac{\Pr(\textit{positive}|\textit{diseased}) \times \Pr(\textit{diseased})}{\Pr(\textit{positive})} \\
&= \frac{\Pr(\textit{positive}|\textit{diseased}) \times \Pr(\textit{diseased})}{\Pr(\textit{positive}|\textit{diseased}) \times \Pr(\textit{diseased}) + \Pr(\textit{positive}|\textit{healthy}) \times \Pr(\textit{he}} \\
&= \frac{0.95 \times 0.008}{0.95 \times 0.008 + 0.1 \times 0.992} = \frac{0.0076}{0.1068} = 0.071
\end{aligned}
$$

$$
\begin{aligned}
\Pr(\textit{healthy}|\textit{positive}) &= \frac{\Pr(\textit{positive}|\textit{healthy}) \times \Pr(\textit{healthy})}{\Pr(\textit{positive})} \\
&= \frac{\Pr(\textit{positive}|\textit{healthy}) \times \Pr(\textit{healthy})}{\Pr(\textit{positive}|\textit{diseased}) \times \Pr(\textit{diseased}) + \Pr(\textit{positive}|\textit{healthy}) \times \Pr(\textit{hea}} \\
&= \frac{0.10 \times 0.992}{0.95 \times 0.008 + 0.1 \times 0.992} = \frac{0.0992}{0.1068} = 0.929
\end{aligned}
$$

$$
\begin{aligned}
\Pr(\textit{diseased}|\textit{positive}) + \Pr(\textit{healthy}|\textit{positive}) &= 0.071 + 0.929 = 1 \\
\Pr(\textit{positive}) &= 0.0076 + 0.0992 = 0.1068
\end{aligned}
$$

Exercise (solution) III

- What is the impact of (apriori) information?



Exercise (solution) IV

- What is probability that an individual is diseased, given that they test positive two times?

$$
\begin{aligned}
\Pr(\textit{diseased}|\textit{positive}) &= 0.071 \Longrightarrow \Pr(\textit{diseased})\\
\Pr(\textit{diseased}|2x\ \textit{positive}) &\Longrightarrow \Pr(\textit{diseased}|\textit{positive})\\
\Pr(\textit{diseased}|\textit{positive}) &= \frac{\Pr(\textit{positive}|\textit{diseased}) \times \Pr(\textit{diseased})}{\Pr(\textit{positive})}\\
&\frac{\Pr(\textit{positive}|\textit{diseased}) \times \Pr(\textit{diseased})}{\Pr(\textit{positive}|\textit{diseased}) \times \Pr(\textit{diseased}) + \Pr(\textit{positive}|\textit{healthy}) \times \mathrm{F}}\\
&= \frac{0.95 \times 0.071}{0.95 \times 0.071 + 0.1 \times 0.929} = \frac{0.06745}{0.16035} = 0.421
\end{aligned}
$$

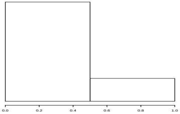
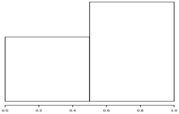
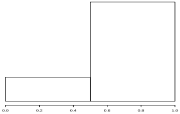
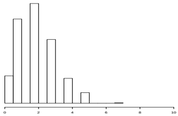
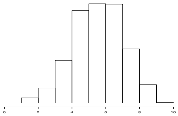
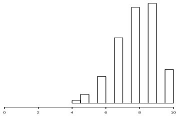
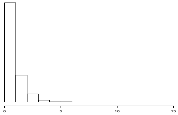
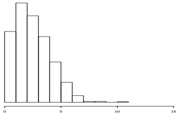
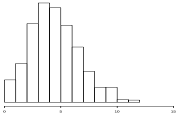
Questions?

Distributions

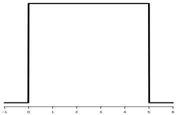
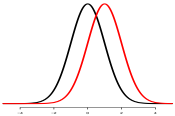
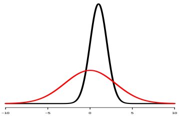
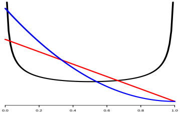
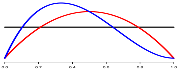
Common distributions used in statistics

Distributions - discrete

Distribution	Notation	$\Pr(x)$	$E(x)$	$Var(x)$
Bernoulli	$B(1,\theta)$	$\theta^x (1-\theta)^{1-x}$	θ	$\theta(1-\theta)$
Binomial	$B(n,\theta)$	$\binom{n}{x}\theta^x (1-\theta)^{n-x}$	$n\theta$	$n\theta(1-\theta)$
Poisson	$P(\lambda)$	$\frac{\lambda^x \exp(-\lambda)}{x!}$	λ	λ

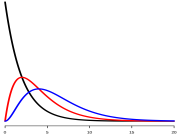
Distributions - continuous

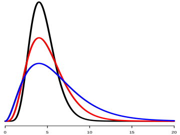
Distribution	Notation	$p(x)$	$E(x)$	$Var(x)$
Uniform	$U(\alpha,\beta)$	$\frac{1}{\beta-\alpha}$	$\frac{\alpha+\beta}{2}$	$\frac{(\beta-\alpha)^2}{12}$
				
Gauss	$N(\mu,\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ^2
				
Beta[1]	$Be(\alpha,\beta)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
				

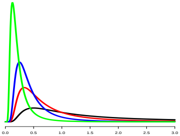
[1] $\Gamma(x)$ = gamma function

Distributions - continuous II

Distribution	Notation	$p(x)$	$E(x)$	$Var(x)$
Gamma[2]	$G(\alpha, \beta)$	$\beta^{\alpha} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$	α/β	α/β^2
	$G(\alpha, \theta = 1/\beta)$	$\frac{1}{\theta^{\alpha}} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\theta}\right)$	$\alpha\theta$	$\alpha\theta^2$
Inverse gamma	$IG(\alpha, \beta)$	$\beta^{\alpha} \frac{1}{\Gamma(\alpha)} (x)^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right)$	$\frac{\beta}{\alpha-1}$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$







[2] $\Gamma(x)$ = gamma function

Exercise

Explain probability mass function for binomial distribution:

$$\Pr(x) = \binom{n}{x} \theta^{x} (1 - \theta)^{n-x}$$

Exercise (solution)

Probability mass function for binomial distribution $B(1, \theta)$:

$$\Pr(x) = \theta^{x} (1 - \theta)^{1-x}$$

- range of value $[0, 1]$
  (0=head or 1=tail for a coin, 0=healthy or 1=diseased,...)
- probability of event 1: $\Pr(x = 1) = \theta$
  ($\theta = 1/2$ for a fair coin)
- genrally (see above):
  - $\Pr(x = 0) = \theta^{0} (1 - \theta)^{1-0} = 1 (1 - \theta)^{1} = 1 - \theta$
  - $\Pr(x = 1) = \theta^{1} (1 - \theta)^{1-1} = \theta^{1} (1 - \theta)^{0} = \theta$

For multiple events from $B(1, \theta)$ the number of combination is $\binom{n}{x}$.
Hence probability mass function for binomial distribution $B(n, \theta)$:

$$\Pr(x) = \binom{n}{x} \theta^{x} (1 - \theta)^{n-x}$$

Exercise

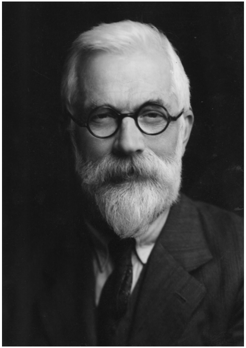
- Run the code

```
n <- 10; theta <- 0.5
## Probability to see x=1, x=5, or x=9?
x <- c(1, 5, 9)
choose(n, x)*theta^x*(1-theta)^(n-x)
## Sample from Binomial distribution
x <- rbinom(n=100, size=n, prob=theta)
## Inspect the sample
table(x); hist(x)
## Calculate mean and variance
n*theta; n*theta*(1-theta)   ## expected
mean(x); var(x)              ## observed
```

- Increase sample size from 100 to 100.000 and repeat. What do you observe?
- Homework: Study documentation:?rpois,?runif,?rnorm,. . .

Questions?
(break)

2. Likelihood



Example

We visit7 independent locations and conduct5 independent and accurate tests for disease. Results are:

Location	1	2	3	4	5	6	7	Total
No. positive	2	0	1	1	1	1	2	8

- What is the proportion (rate) of diseased in the population?
  - No. of positive tests per location:i: y_i (known)
  - Proportion of diseased: θ (unknown)
  - Assumed model: $\Pr(y_i|n=5,\theta) \sim B(5,\theta)$
  - Likelihood principle (Fisher):"**Find parameter values** (θ),**which will maximise the likelihood of observed data** ($\mathbf{y}$)."
  - Probability density/mass $\Pr(\mathbf{y}|n=5,\theta)$ and likelihood $L(\mathbf{y}|n=5,\theta)$:

$$\Pr(\mathbf{y}|n=5,\theta) = \prod_{i=1}^{7} \binom{n}{y_i} \theta^{y_i} (1-\theta)^{n-y_i} \qquad L(\mathbf{y}|n=5,\theta) = \prod_{i=1}^{7} \binom{n}{y_i} \theta^{y_i} (1-\theta)^{n-y_i}$$

Example II

Find parameter values (θ), which will maximise the likelihood of observed data ($\mathbf{y}$). –> Find maximum of $L(y|n,\theta)$.

Standard procedure:

1. "Computers don't like small numbers and multiplications are more work than additions"
   $\log(L(y|n,\theta)) = l(y|n,\theta)$
2. "Massage" $l(y|n,\theta)$
3. Take the first derivative and find where it's equal to zero (check for second too)
4. Solve expression –> calculate estimate $\left(\hat{\theta}_{MLE}\right)$ (direct or numerically)
5. Calculate variance of estimate from information matrix $\left(Var\left(\hat{\theta}_{MLE}\right) = \left(-E\left(\frac{\partial^2 \log(L(y|n,\theta))}{\partial\theta\partial\theta'}\right)\right)^{-1}\right)$

MLE = eng. Maximum Likelihood Estimate, slo. cenilka po metodi največjega verjetja

Example II

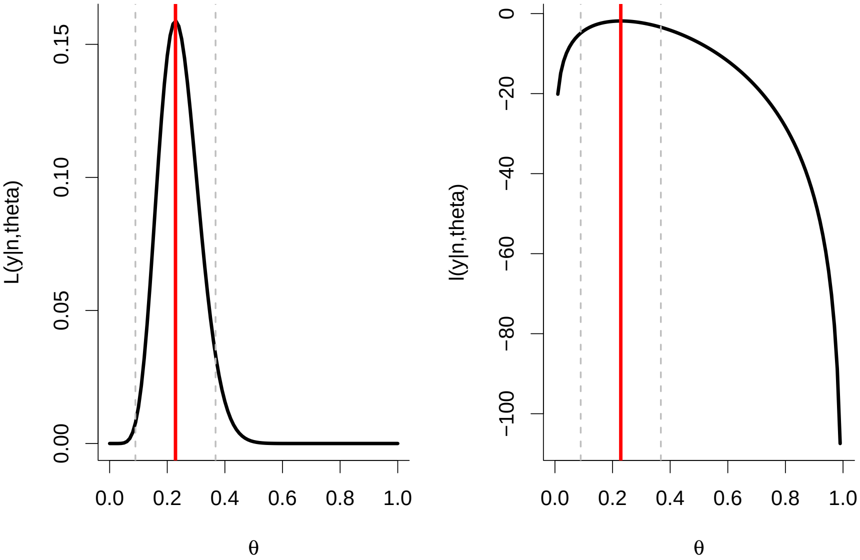
Estimate of the proportion of diseased given the observed data:

$$\begin{aligned}\hat{\theta}_{MLE} &= \frac{y}{n} \\ &= \frac{8}{35} = 0.23\end{aligned}$$

Standard error of the estimate:

$$\begin{aligned}SE\left(\hat{\theta}_{MLE}\right) &= \sqrt{\frac{-1}{-\frac{y}{p^2} - \frac{n-y}{(1-p)^2}}} \\ &= \sqrt{\frac{-1}{-\frac{8}{0.23^2} - \frac{35-8}{(1-0.23)^2}}} = 0.07\end{aligned}$$

Example III



L(y|n,theta)
0.00
0.05
0.10
0.15
0.0
0.2
0.4
0.6
0.8
1.0
θ
l(y|n,theta)
0
−20
−40
−60
−80
−100
0.0
0.2
0.4
0.6
0.8
1.0
θ

Exercise

- Use this code

```
n <- 35
y <- 8
p <- seq(from=0, to=1, by=0.01)
MLE <- y/n
MLE_SE <- sqrt(-1 / (-y/MLE^2 - (n-y)/(1-MLE)^2))
MLE_CI <- MLE + c(-1.96, 0, 1.96) * MLE_SE
L <- choose(n, y)*p^y*(1-p)^(n-y)
l <- log(L)
par(mfrow=c(1, 2))
plot(L ~ p, type="l"); abline(v=MLE_CI)
plot(l ~ p, type="l"); abline(v=MLE_CI)
```

- What changes if $n = 2$ and we have two positive outcomes or $n = 1000$ and 230 positive outcomes?

3. Bayesian statistics

Bayes theorem

- What is the proportion (rate) of diseased in the population?
  - No. of all positive tests: $n = 35, y = 8$ (known)
  - Proportion of diseased: θ (unknown)

$$
\begin{aligned}
p(\theta|y,n) &= \frac{p(y|n,\theta) \times p(\theta)}{p(y)} \\
&= \frac{p(y|n,\theta) \times p(\theta)}{\int_\theta p(y|n,\theta) \times p(\theta)} \\
p(\theta|y,n) &\propto p(y|n,\theta) \times p(\theta)
\end{aligned}
$$

"posterior is proportional to likelihood times prior"

- Glossary:
  - $p(\theta|y,n)$ = posterior distribution (result)
  - $p(y|n,\theta)$ = likelihood (assumed model + data)
  - $p(\theta)$ = prior distribution (assumption = prior knowledge)

Type of distribution?

- What is the proportion (rate) of diseased in the population?
  - No. of all positive tests: $n = 35, y = 8$ (known)
  - Proportion of diseased: θ (unknown)

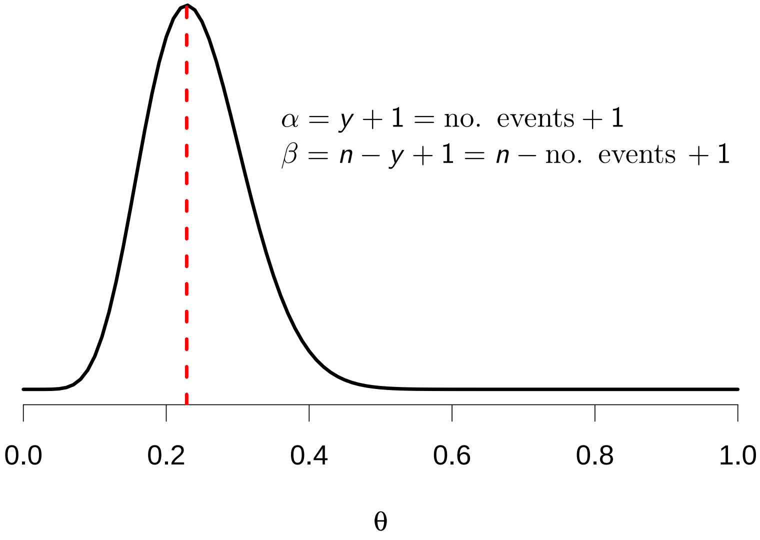
$$
\begin{aligned}
p(\theta|y,n) &\propto p(y|n,\theta) \times p(\theta) \\
&\propto \binom{n}{y} \theta^y (1-\theta)^{n-y} \times p(\theta) \\
&\propto \theta^y (1-\theta)^{n-y} \times p(\theta)
\end{aligned}
$$

"after the inspection beta distribution can be recognized"

$$
x \sim Be(\alpha, \beta) \qquad p(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}
$$

$$
p(y|n,\theta) = Be(y+1, n-y+1)
$$

Type of distribution $Be\,(y+1, n-y+1)$



Conjugate prior distribution

- Conjugate prior distribution has the same "form" as likelihood:

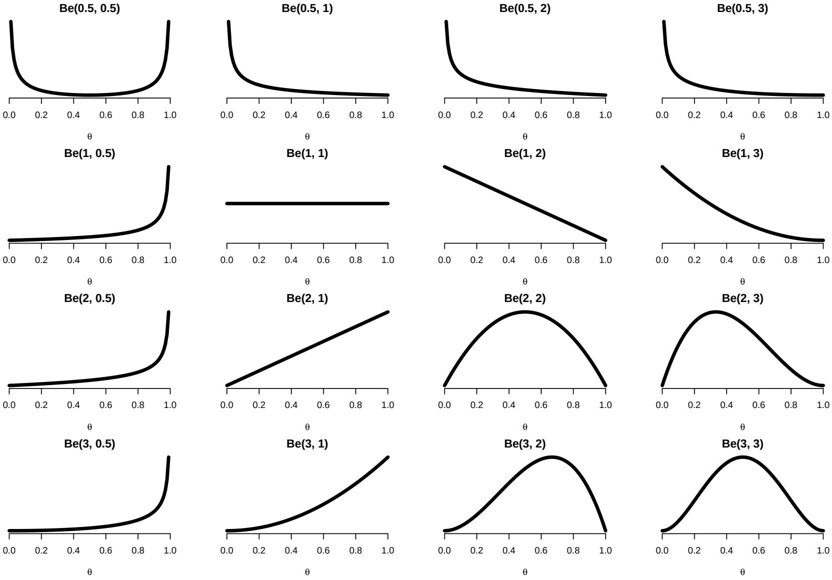
$$\begin{aligned} p(y|n,\theta) &= Be(y+1, n-y+1) \\ p(\theta) &= Be(\alpha,\beta) \\ p(\theta|y,n) &\propto \theta^{y}(1-\theta)^{n-y} \times \theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &\propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1} \end{aligned}$$

"after the inspection beta distribution can be recognized"

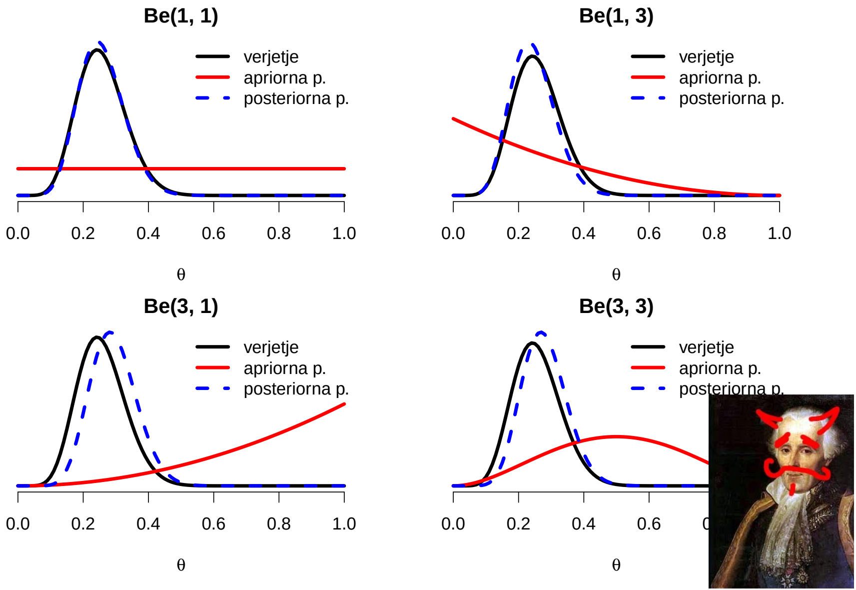
$$p(\theta|y,n) = Be(y+1+\alpha, n-y+1+\beta)$$

Provide parameters of the prior distribution: α, β

Prior distribution $Be(\alpha, \beta)$



Posterior distribution $Be\left(y + 1 + \alpha, n - y + 1 + \beta\right)$



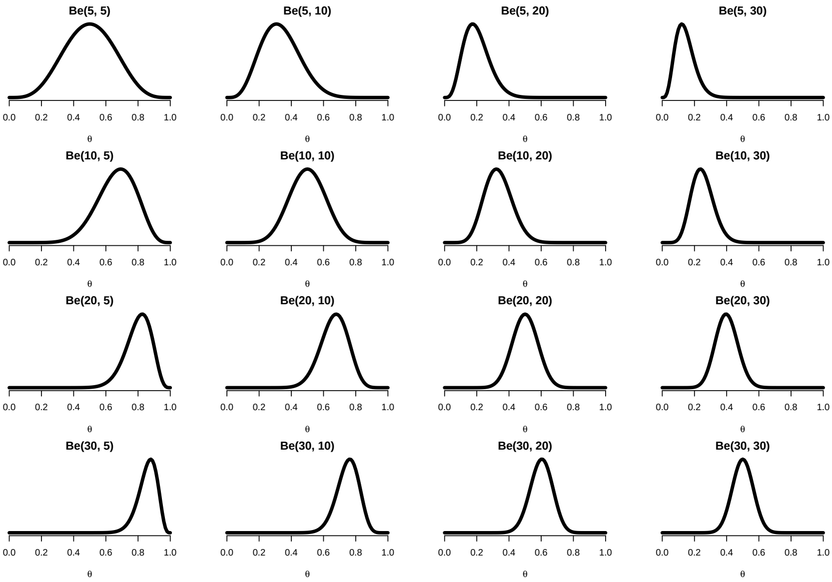
Posterior distribution $Be\left(y+1+\alpha, n-y+1+\beta\right)$

$$E\left(x|\alpha,\beta\right)=\frac{\alpha}{\alpha+\beta} \qquad Var\left(x|\alpha,\beta\right)=\frac{\alpha\beta}{\left(\alpha+\beta\right)^2\left(\alpha+\beta+1\right)}$$

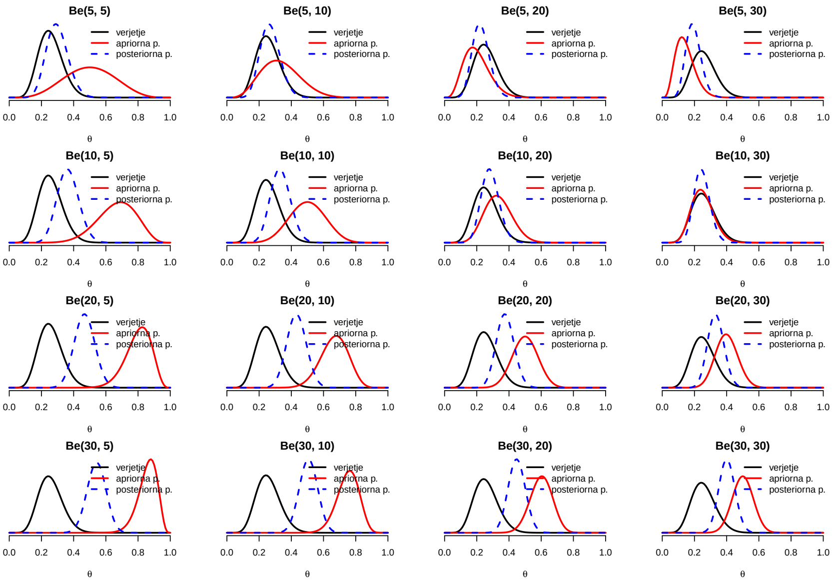
Prior d.	$E\left(\theta\|y,n,\alpha,\beta\right)$	$SE\left(\theta\|y,n,\alpha,\beta\right)$
$Be\left(1,1\right)$	0.256	0.069
$Be\left(1,3\right)$	0.243	0.066
$Be\left(3,1\right)$	0.293	0.070
$Be\left(3,3\right)$	0.279	0.068

$$\hat{\theta}_{MLE}=0.23, \quad SE\left(\hat{\theta}_{MLE}\right)=0.07$$

Prior distribution $Be(\alpha, \beta)$



Posterior distribution $Be\left(y + 1 + \alpha, n - y + 1 + \beta\right)$



Exercise

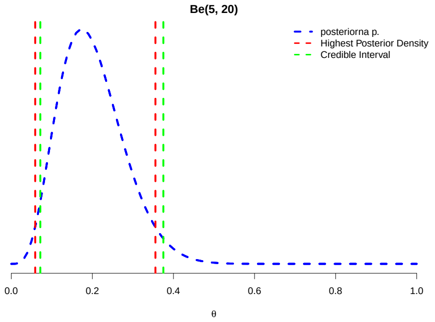
- Using the code, calculate and plot posterior distribution for:
  - sample size 10 and 2 positive events
  - sample size 1000 and 230 positive events

```
n <- 35; y <- 8
alpha <- 1; beta <- 10
a <- y + alpha; b <- n - y + beta
MLE <- y/n
MLE_SE <- sqrt(-1 / (-y/MLE^2 - (n-y)/(1-MLE)^2))
POS <- a / (a + b)
POS_SE <- sqrt(a * b / (((a + b)^2) * (a + b + 1)))
p <- seq(from=0, to=1, by=0.01)
dL   <- dbeta(x=p, shape1=y+1,   shape2=n-y+1)
dPRI <- dbeta(x=p, shape1=alpha, shape2=beta)
dPOS <- dbeta(x=p, shape1=a,     shape2=b)
matplot(y=cbind(dL, dPRI, dPOS), x=p, type="l",
        col=c("black", "red", "blue"))
abline(v=MLE)
```

Interval

- Credible interval (analogous to confidence interval; when we know the posterior distribution we can obtain quantiles)
- Highest Posterior Density (HPD) (shortest interval covering % of distribution!)

$$CI\,(\theta|y,n,\alpha,\beta) = 0.07 - 0.38, \quad HPD\,(\theta|y,n,\alpha,\beta) = 0.06 - 0.36$$



Exercise II

- Using the code, sample from posterior distribution and calculate:
  - mean
  - standard deviation
  - credible interval
  - highest posterior density

```
n <- 35; y <- 8
alpha <- 1; beta <- 1
a <- y + 1 + alpha; b <- n - y + 1 + beta
x <- rbeta(n=100000, shape1=a, shape2=b)
hist(x)
mean(x)
sd(x)
mean(x) + c(-1, 1)*1.96*sd(x)
library(package="LaplacesDemon")
p.interval(x, plot=TRUE)
p.interval(x, plot=TRUE, HPD=FALSE)
```

Questions?

Prior distribution

"Bayesians start with a vague expectation of seeing a horse, but at a glimpse of a donkey conclude they have seen a mule." Senn (1997)

- Non-informative prior does not exist. Even uniform prior distribution says that all values are equally likely.
- The are approaches for deriving "non-informative prior distribution
  - Jeffrey's prior distribution
  - reference prior distribution (Bernardo)
  - ...
- Sensitivity analysis
- Use of prior information can bias estimates, which have smaller variance (combination of two sources of information −> smaller variance)
- Preference for conjugate priors to simplify calculations
- Priors less important with more records
- Can improve model fitting!

"General" prior distribution

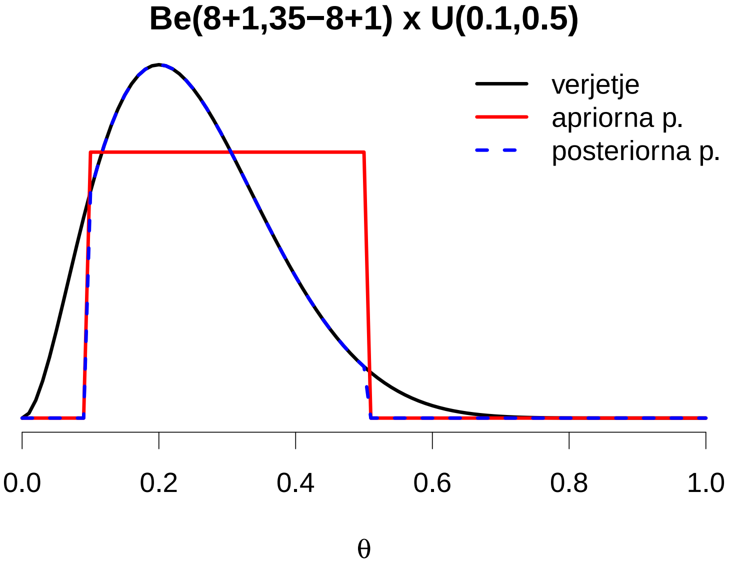
- For prior distribution we can pick any valid (proper) distribution:

$$
\begin{aligned}
p(y|n,\theta) &= Be(y+1, n-y+1) \\
p(\theta) &= U(\alpha,\beta) \\
p(\theta|y,n) &\propto \theta^{y}(1-\theta)^{n-y} \times \frac{1}{\beta-\alpha} \\
&\propto \int_{\alpha}^{\beta} Be(y+1, n-y+1)\frac{1}{\beta-\alpha}d\theta \\
&\propto ????
\end{aligned}
$$

"after the inspection no standard distribution can be recognized"

Cannot directly solve this. Need an alternative approach!

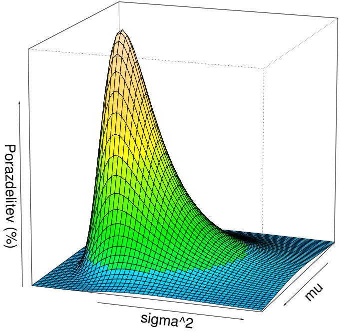
“General” prior distribution - uniform



More parameters?

- Statistical models have usually more than one parameter!
- For example in Gauss distribution

$$\begin{aligned}
p\left(y|\mu,\sigma^2\right) &= N\left(\mu,\sigma^2\right)\\
p\left(\mu\right) &= N\left(\ldots\right)\\
p\left(\sigma^2\right) &= IG\left(\ldots\right)\\
p\left(\mu,\sigma^2|y\right) &\propto p\left(y|\mu,\sigma^2\right)p\left(\mu\right)p\left(\sigma^2\right)
\end{aligned}$$

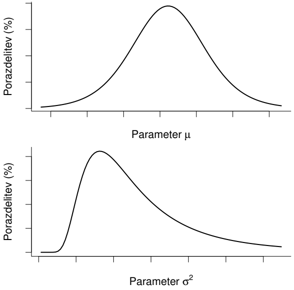


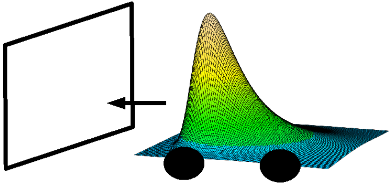
But, we want as result marginal (one-dimensional) posterior distributions: $p\left(\mu|y\right)$in$p\left(\sigma^2|y\right)$

More parameters?

But, we want as result marginal (one-dimensional) posterior distributions!

$$p(\mu|y) = \int_{0}^{\infty} p(\mu, \sigma^2|y)\, d\sigma^2 \qquad p(\sigma^2|y) = \int_{-\infty}^{+\infty} p(\mu, \sigma^2|y)\, d\mu$$





Even more parameters?

$$
\begin{aligned}
p(\theta_1, \theta_2, \ldots, \theta_p | y) &\propto p(y | \theta_1, \theta_2, \ldots, \theta_p)\, p(\theta_1)\, p(\theta_2) \ldots p(\theta_p) \\
p(\theta_1 | y) &= \int_{\theta_2} \int_{\ldots} \int_{\theta_p} p(\theta_1, \theta_2, \ldots, \theta_p | y)\, d\theta_2 d \ldots d\theta_p \\
p(\theta_2 | y) &= \int_{\theta_1} \int_{\ldots} \int_{\theta_p} p(\theta_1, \theta_2, \ldots, \theta_p | y)\, d\theta_1 d \ldots d\theta_p \\
\ldots & \\
\boldsymbol{\theta} &= \theta_1, \theta_2, \ldots, \theta_p \\
p(\theta_i | y) &= \int_{\boldsymbol{\theta}_{-i}} p(\boldsymbol{\theta} | y)\, d\boldsymbol{\theta}_{-i}
\end{aligned}
$$

Usually analytically impossible to solve!!!

Modern computational methods for Bayesian statistics

- MCMC
- Variational Bayes (Expectation Propagation, Mean-Field, ...)
- INLA (Integrated Nested Laplace Approximation) - for some models
- ...

Questions?
(break)

4. MCMC

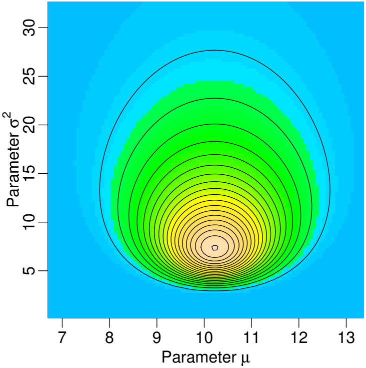
MCMC

- MCMC
  - eng. Monte Carlo Markov Chain
  - slo. Monte Carlo z Markovskimi verigami
- Monte Carlo - stochastic method for solution of compute intense problems
- Markov chain - current value of a variable depends only on the previous value
- In Bayesian statistics we use MCMC for sampling from (conditional) posterior distributions
- Algorithms
  - Metropolis in Metropolis-Hastings sampling
  - Gibbs sampling
  - NUTS
  - HMC
  - ...

Larger number of parameters?

- Statistical models have usually more than one parameter!
- For example in Gauss distribution

$$
\begin{aligned}
p\left(y|\mu,\sigma^2\right) &= N\left(\mu,\sigma^2\right)\\
p\left(\mu\right) &= N\left(\ldots\right)\\
p\left(\sigma^2\right) &= IG\left(\ldots\right)\\
p\left(\mu,\sigma^2|y\right) &\propto p\left(y|\mu,\sigma^2\right)p\left(\mu\right)p\left(\sigma^2\right)
\end{aligned}
$$



But, we want as result marginal (one-dimensional) posterior distributions:

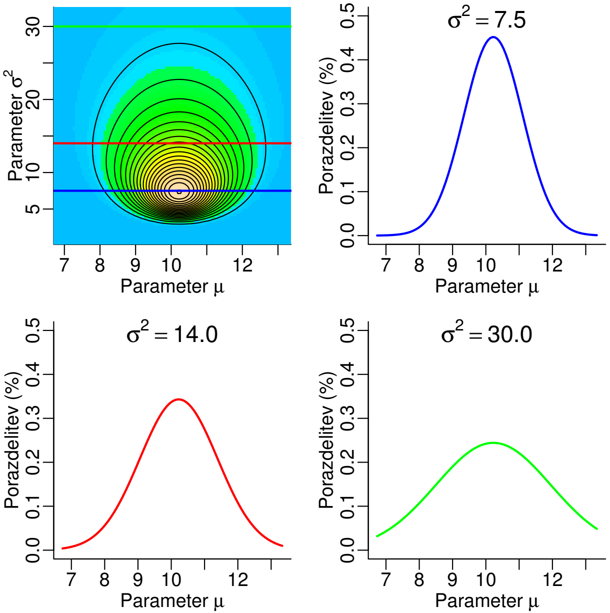
$p\left(\mu|y\right)$in$p\left(\sigma^2|y\right)$

MCMC - conditional distributions

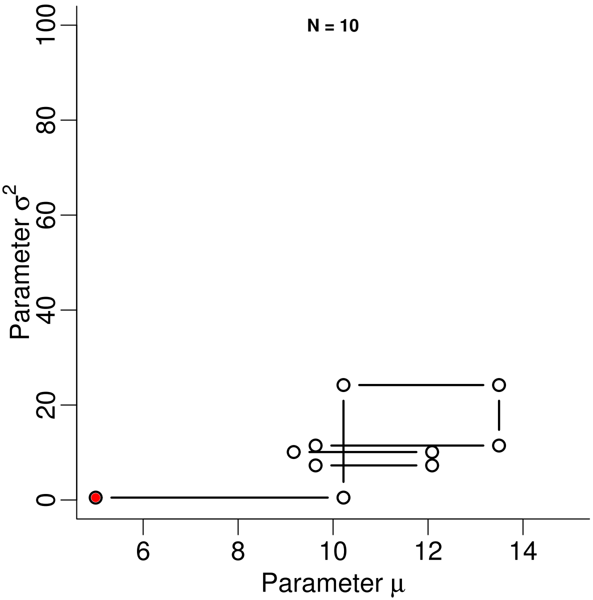
- Derive conditional distributions for each parameter (possibly for a group/block of parameters)
- Simple to sample from one-dimensional conditional distributions
- Conditional distributions for Gauss' model:

$$
\begin{aligned}
p\left(\mu, \sigma^2 | y\right) &\propto p\left(y | \mu, \sigma^2\right) p\left(\mu\right) p\left(\sigma^2\right) \\
p\left(\mu | y, \sigma^2\right) &\propto p\left(y | \mu, \sigma^2\right) p\left(\mu\right) \cancel{p\left(\sigma^2\right)} \\
p\left(\sigma^2 | y, \mu\right) &\propto p\left(y | \mu, \sigma^2\right) \cancel{p\left(\mu\right)} p\left(\sigma^2\right)
\end{aligned}
$$

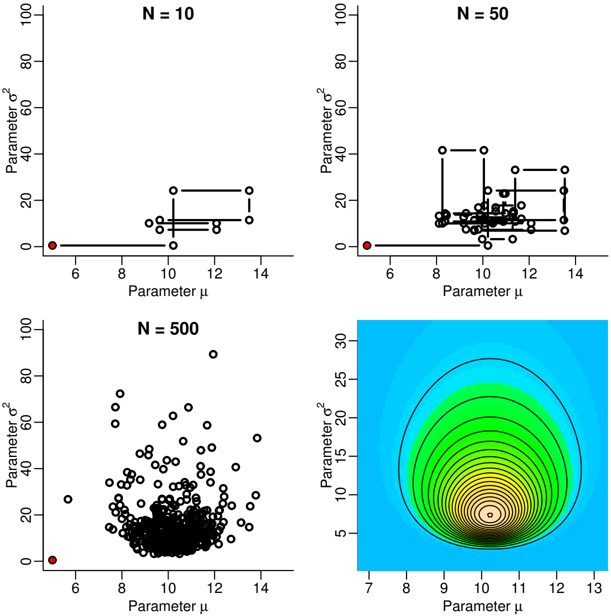
MCMC - conditional distributions (demo)



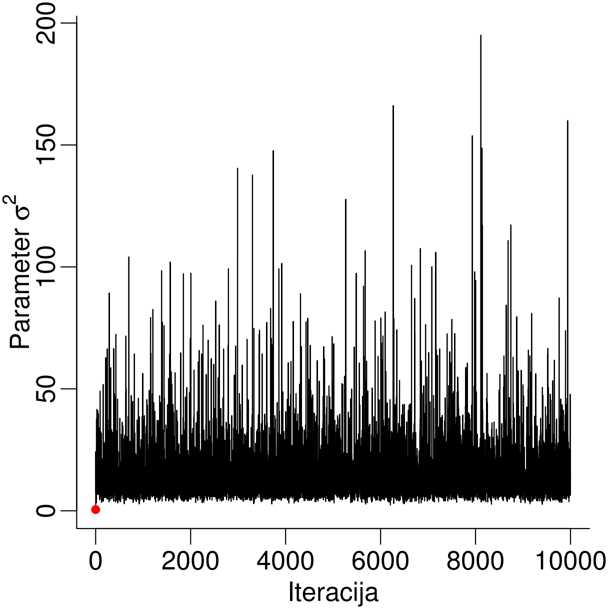
MCMC - sampling



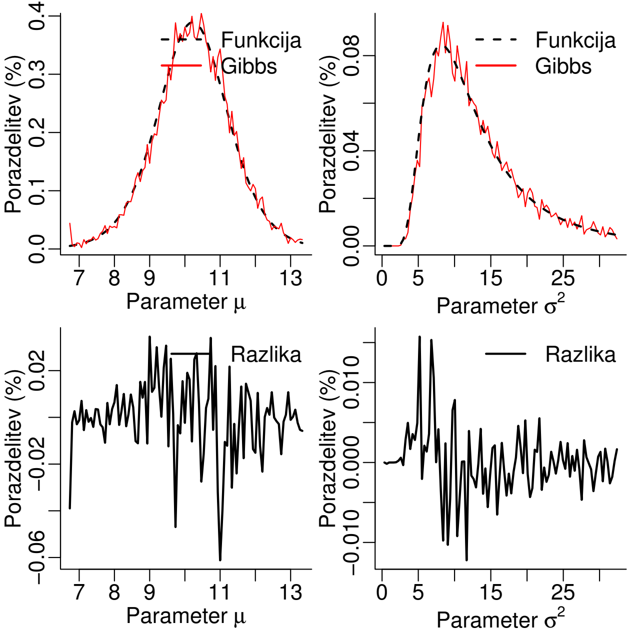
MCMC - sampling II



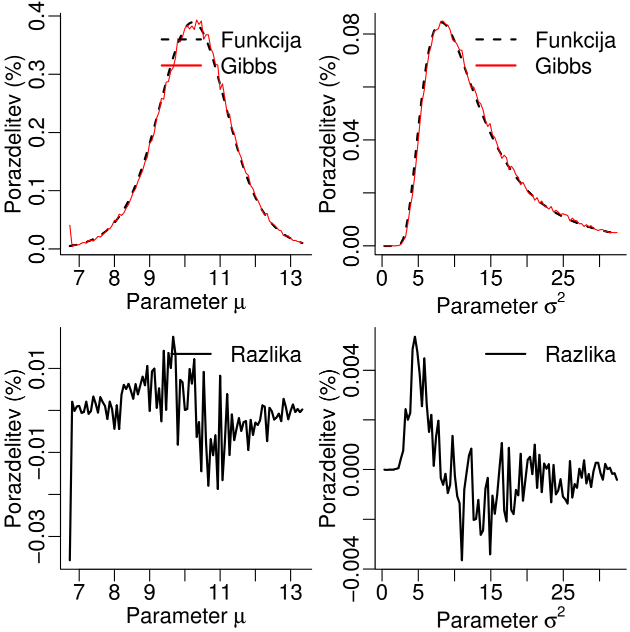
MCMC - Markov chain



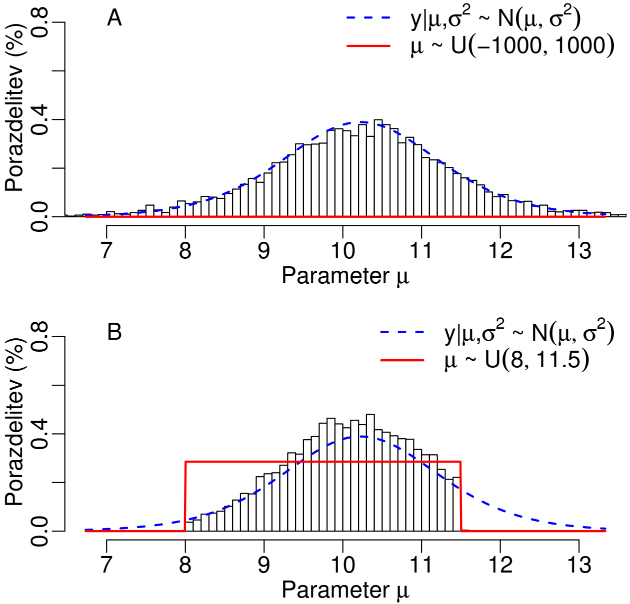
MCMC - marginal distributions



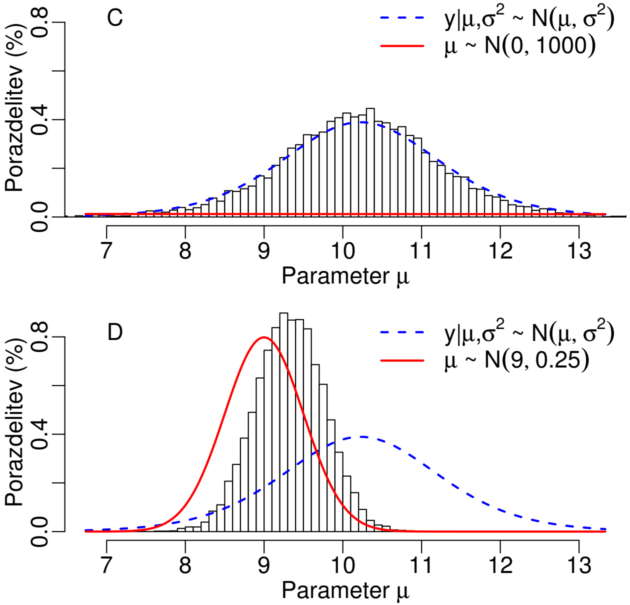
MCMC - marginal distributions II



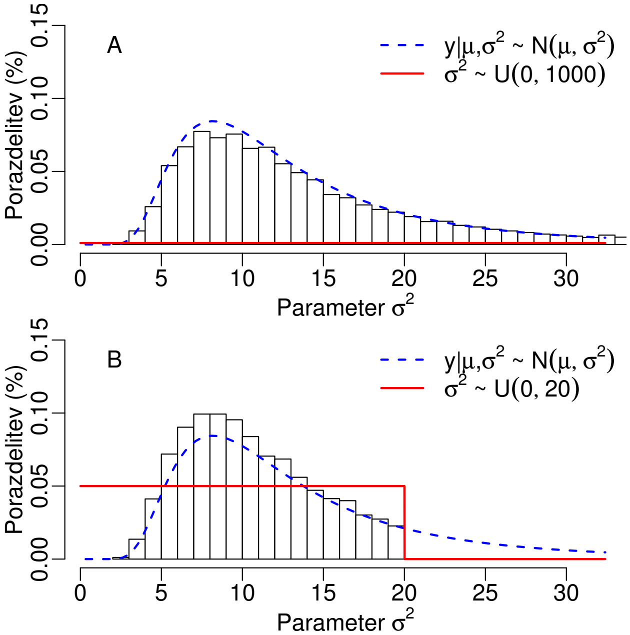
MCMC - different priors for μ



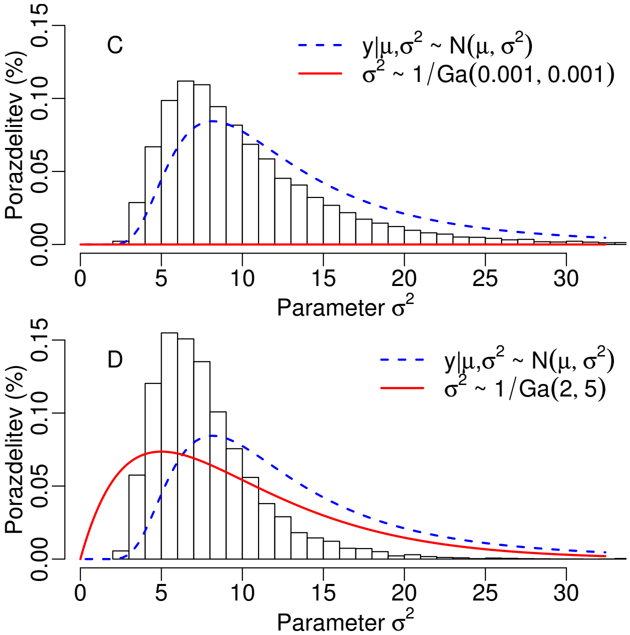
MCMC - different priors for μ II



MCMC - different priors for σ^2



MCMC - different priors for σ^2 II



Questions?

General Bayes programs/packages

- BUGS, JAGS, and Stan family
- R paketi
  http://cran.r-project.org/web/views/Bayesian.html
  - brms
    https://cran.r-project.org/web/packages/brms/index.html
  - rstan https:
    //cran.r-project.org/web/packages/rstan/index.html
  - R-INLA https://https://www.r-inla.org
  - ...