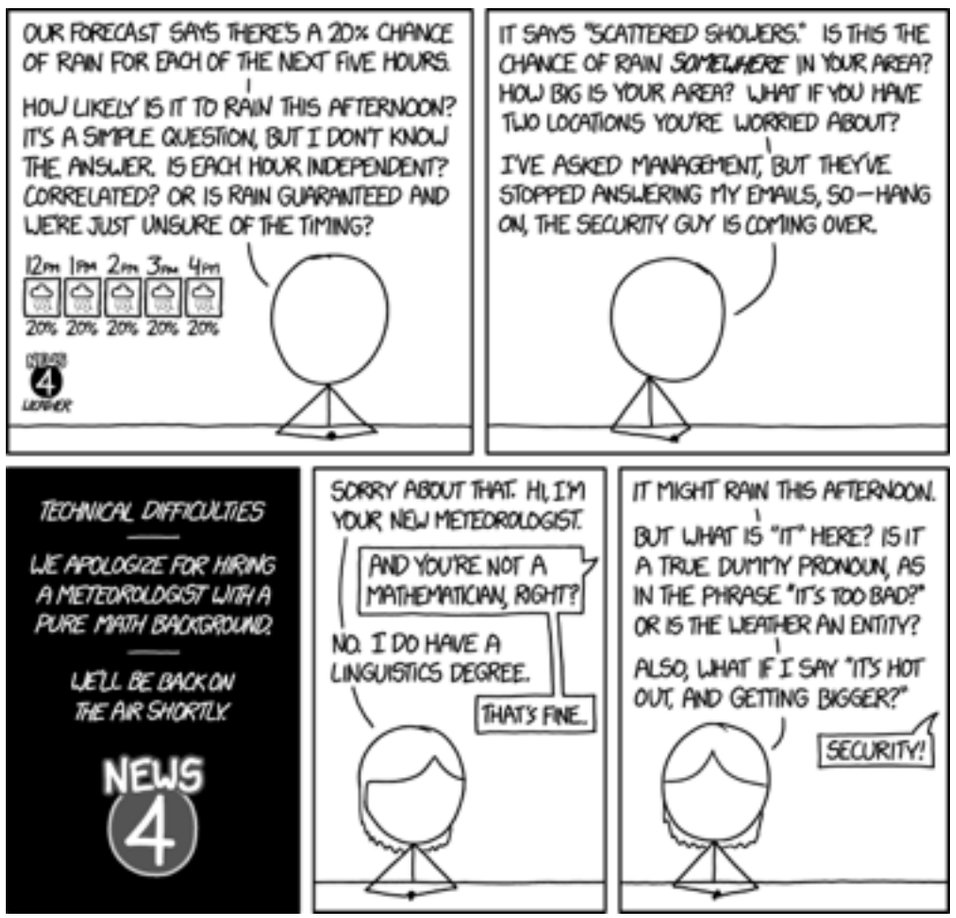
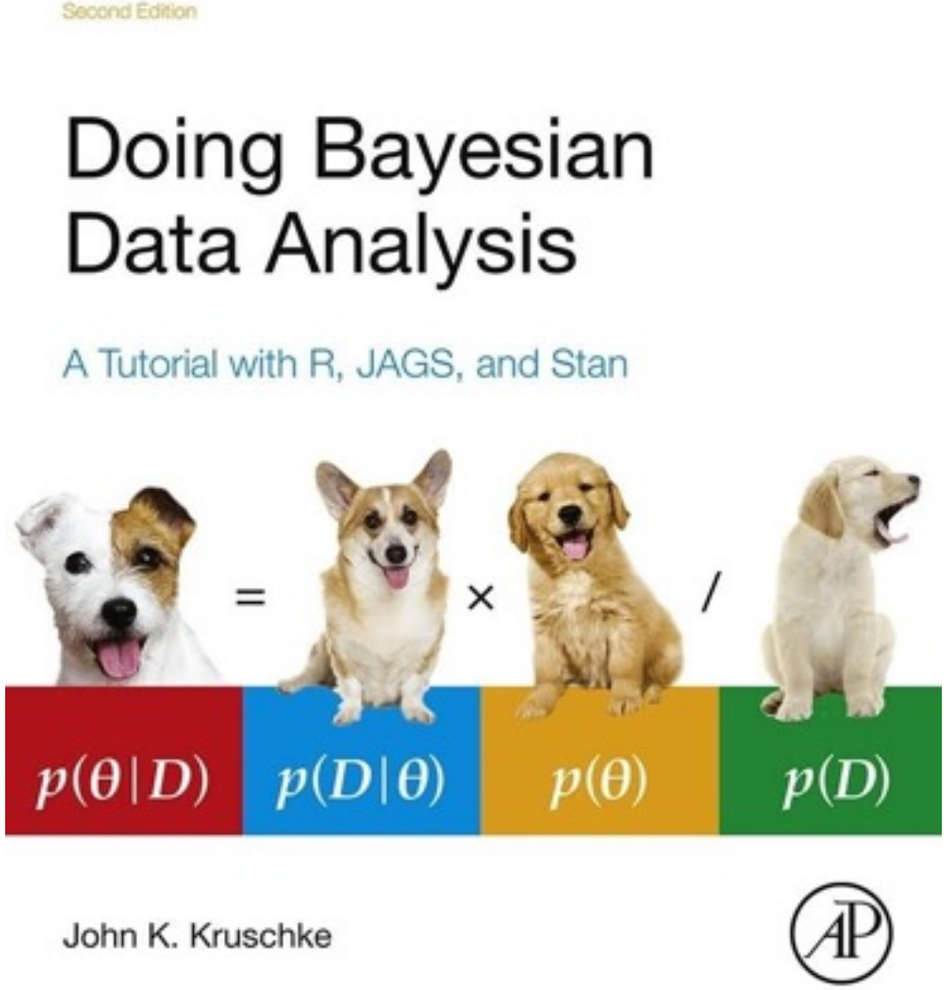
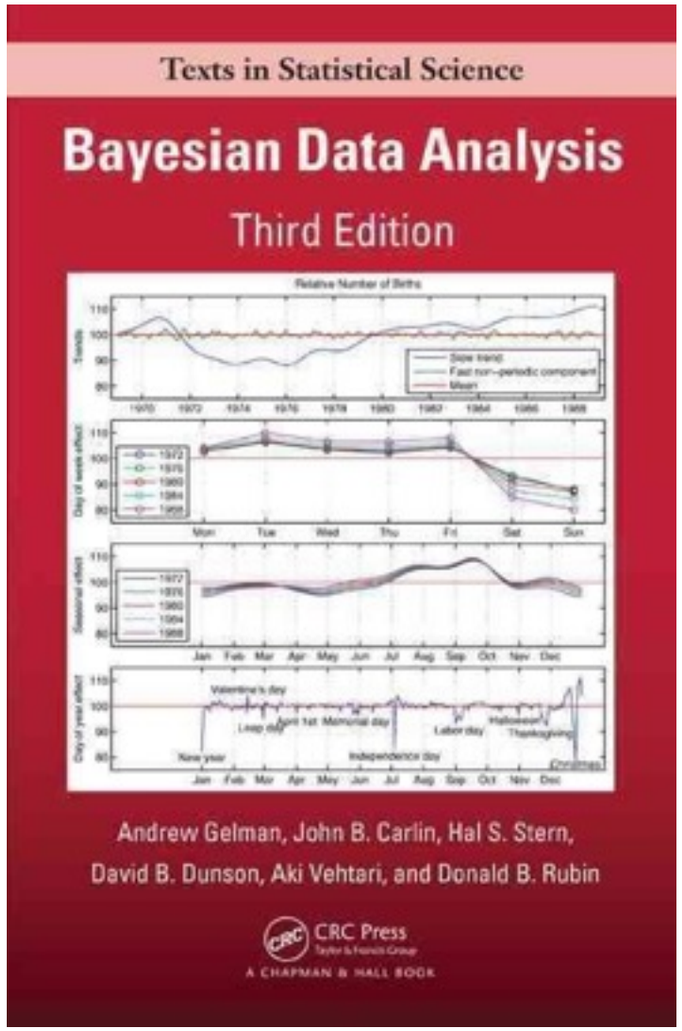
Introduction to Bayesian data analysis



xkcd.com

José P. Ossandón
jose.ossandon@uni-hamburg.de





In press, *Journal of Experimental Psychology: General*
Version of May 31, 2012

Bayesian estimation supersedes the *t* test

John K. Kruschke
Indiana University, Bloomington

Outline:

1. **The Big problem**

2. **‘Technical’ problems: NHST**

3. **Overview Bayes**

4. **Overview Bayes modelling**

5. **Estimation continuous variable** (effects of prior and amount of data)

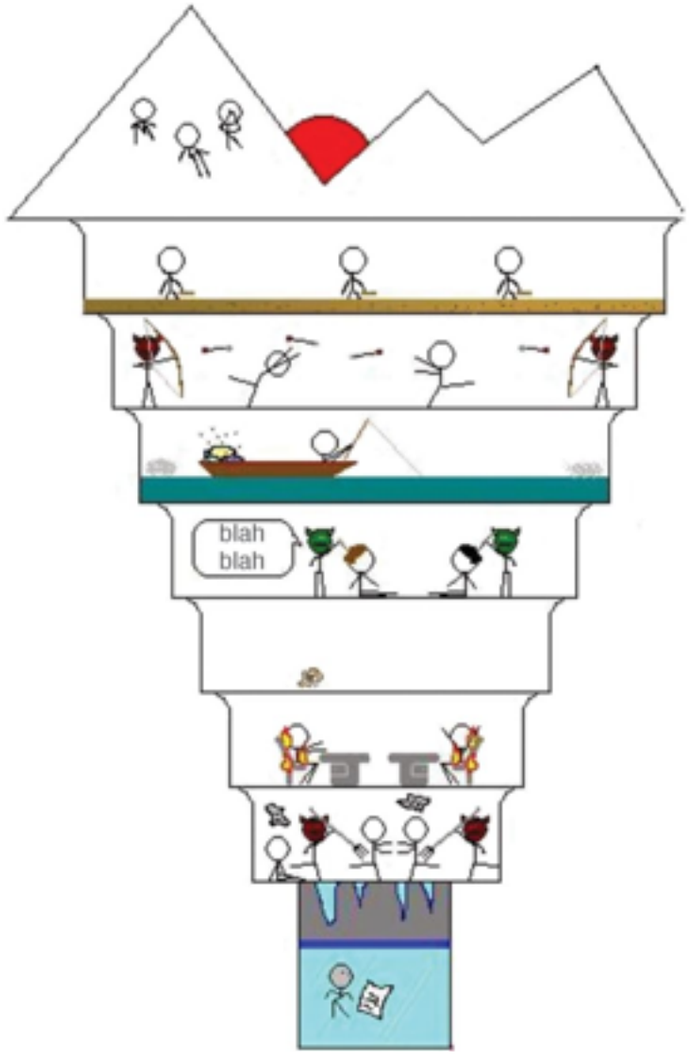
6. **Difference continuous variables** (convergence checks, comparison between normal and t likelihood distribution, effect size)

7. **Multiple groups continuous variable - hierarchical linear model** (ANOVA like, establishing differences: ROPE, shrinkage)

8. **Hierarchical linear model continuous variable, within subjects** (model comparison)

The big problem

Neuroskeptic's 9 circles of scientific hell:



I	Limbo
II	Overselling
III	Post-Hoc Storytelling
IV	P-Value Fishing
V	Creative Outliers
VI	Plagiarism
VII	Non-Publication
VIII	Partial Publication
IX	Inventing Data

The big problem

ACADEMIA AND CLINIC

Toward Evidence-Based Medical Statistics. 1: The *P* Value Fallacy

Steven N. Goodman, MD, PhD

The Earth Is Round ($p < .05$)

Jacob Cohen

Published in: D. Kaplan (Ed.). (2004). *The Sage handbook of quantitative methodology for the social sciences* (pp. 391–408). Thousand Oaks, CA: Sage.

The Null Ritual
What You Always Wanted to Know About Significance Testing but Were Afraid to Ask

Gerd Gigerenzer, Stefan Krauss, and **Oliver Vitouch**[1]

The big problem

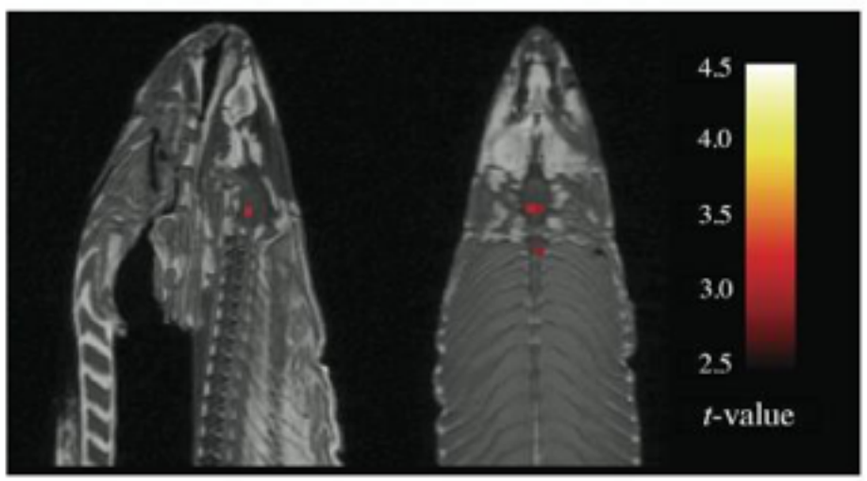
Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition[1]

Edward Vul,[1] Christine Harris,[2] Piotr Winkielman,[2] & Harold Pashler[2]

[1]*Massachussetts Institute of Technology and* [2]*University of California, San Diego*



Open access, freely available online

Essay

Why Most Published Research Findings Are False

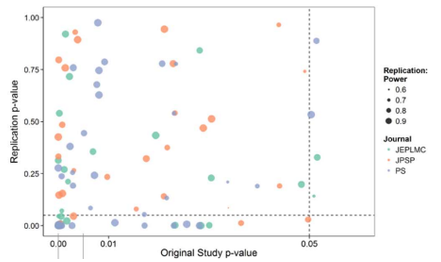
John P. A. Ioannidis

RESEARCH ARTICLE SUMMARY

PSYCHOLOGY

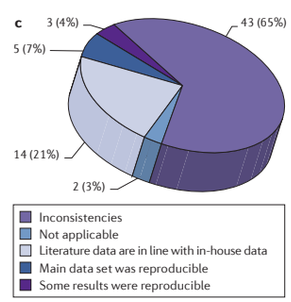
Estimating the reproducibility of psychological science

Open Science Collaboration*



Believe it or not: how much can we rely on published data on potential drug targets?

Florian Prinz, Thomas Schlange and Khusru Asadullah



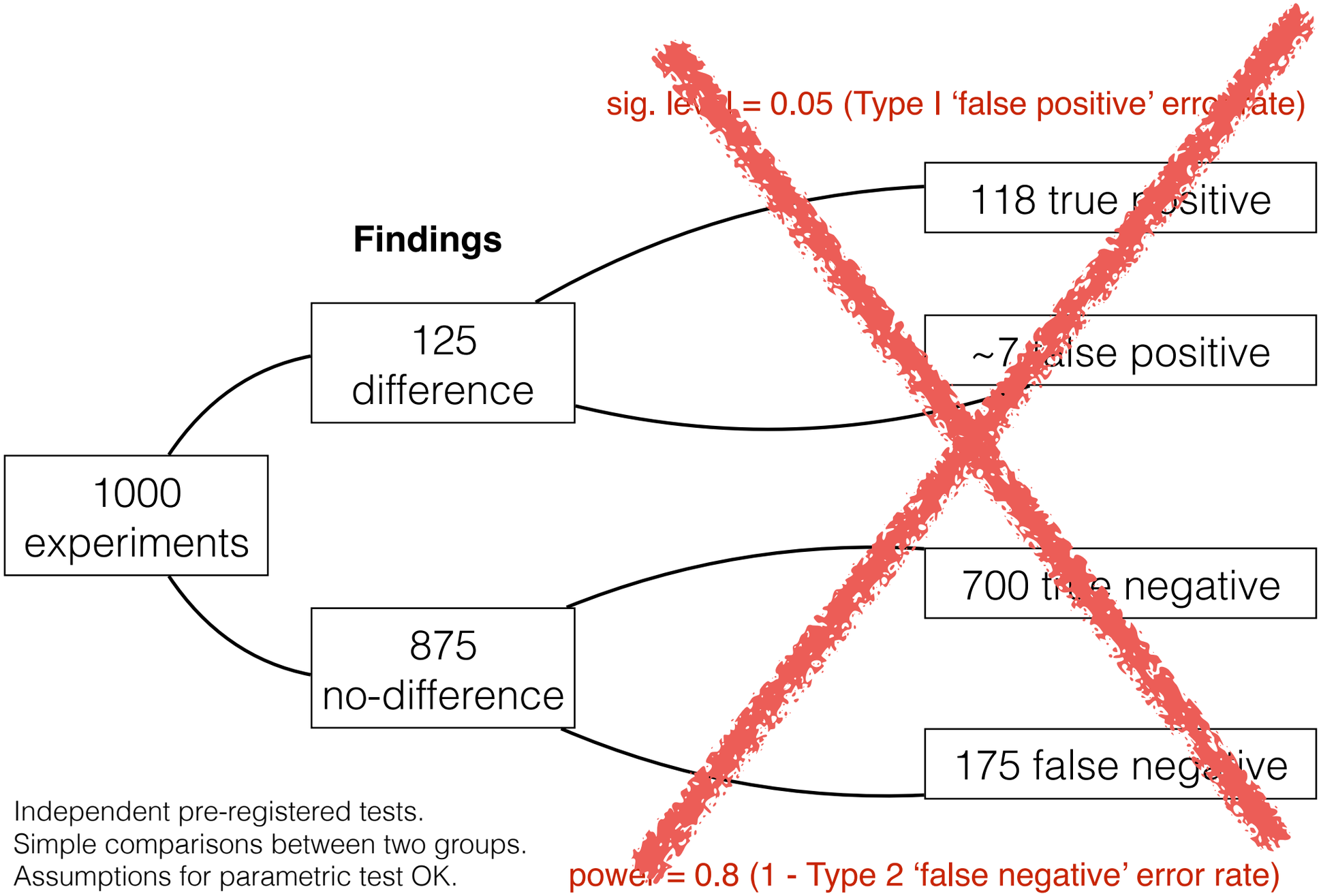
The problems with NHST

Null Hypothesis Significant Testing:

p(data|hypothesis): probability of the data given a null hypothesis.

- 'Interpretation' problem: alpha threshold and p values do not determine error rates.
- 'Practical' problems: depends on testing intention and stopping criterions, possibilities ('forked paths'). Focus on 'significant' difference results. Low threshold for significance. Underpowered studies.

NHST problems: what is my error rate?



5 % errors when saying there is something and 20 % when there is nothing?

(this would be still being wrong ~18% of the time, *if you have a power of 80 %,* btw)

NHST problems: what is my error rate?

p(data|hypothesis): probability of the data given a null hypothesis.

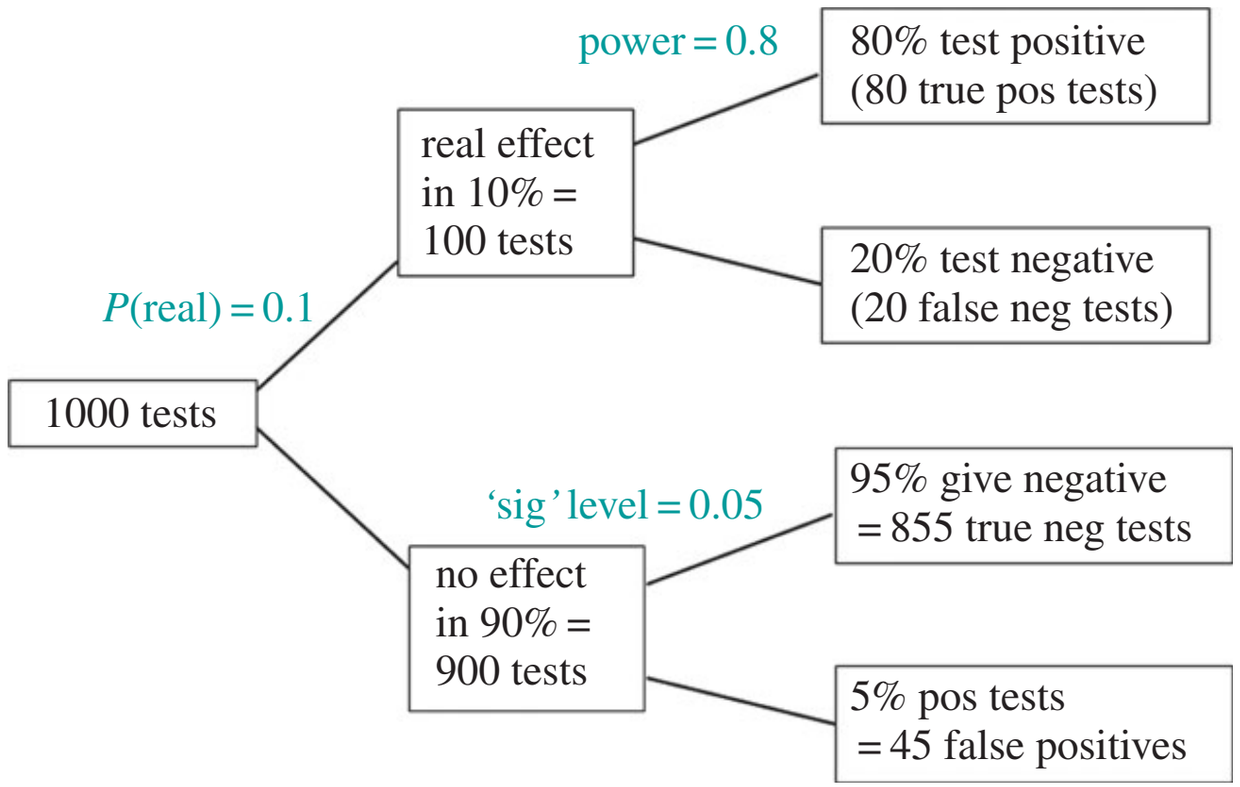


Figure 2. Tree diagram to illustrate the false discovery rate in significance tests. This example considers 1000 tests, in which the prevalence of real effects is 10%. The lower limb shows that with the conventional significance level, $p = 0.05$, there will be 45 false positives. The upper limb shows that there will be 80 true positive tests. The false discovery rate is therefore $45/(45 + 80) = 36\%$, far bigger than 5%.

Colquhoun D. An investigation of the false discovery rate and the misinterpretation of *p*-values.
R. Soc. open sci. **1**: 140216. (2014)

NHST problems: what is my error rate?

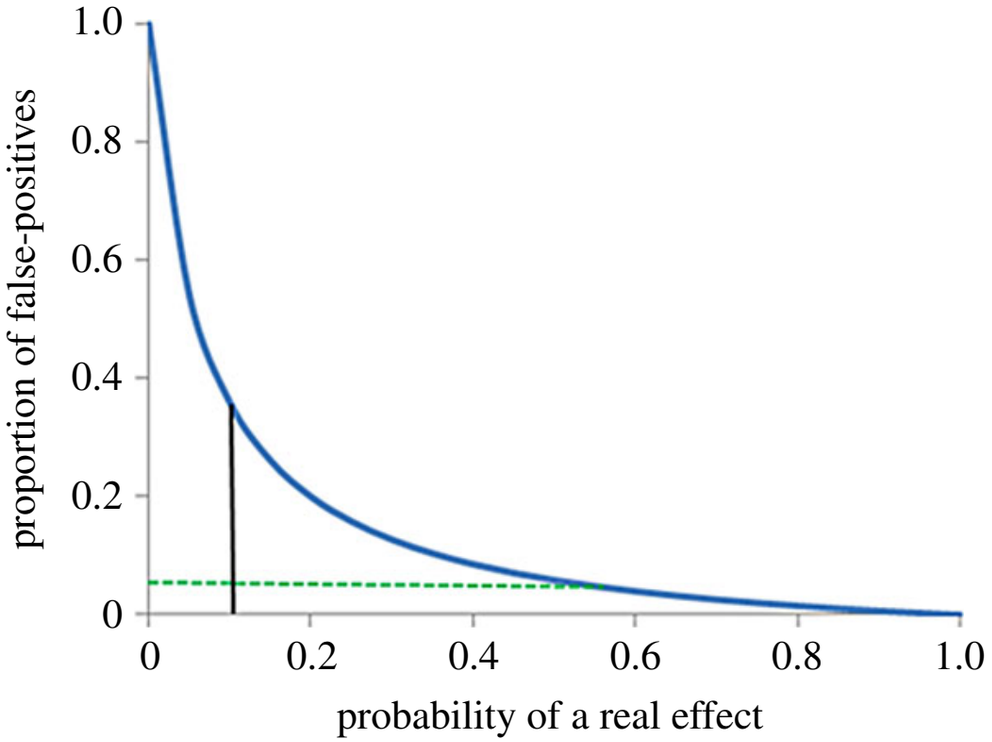


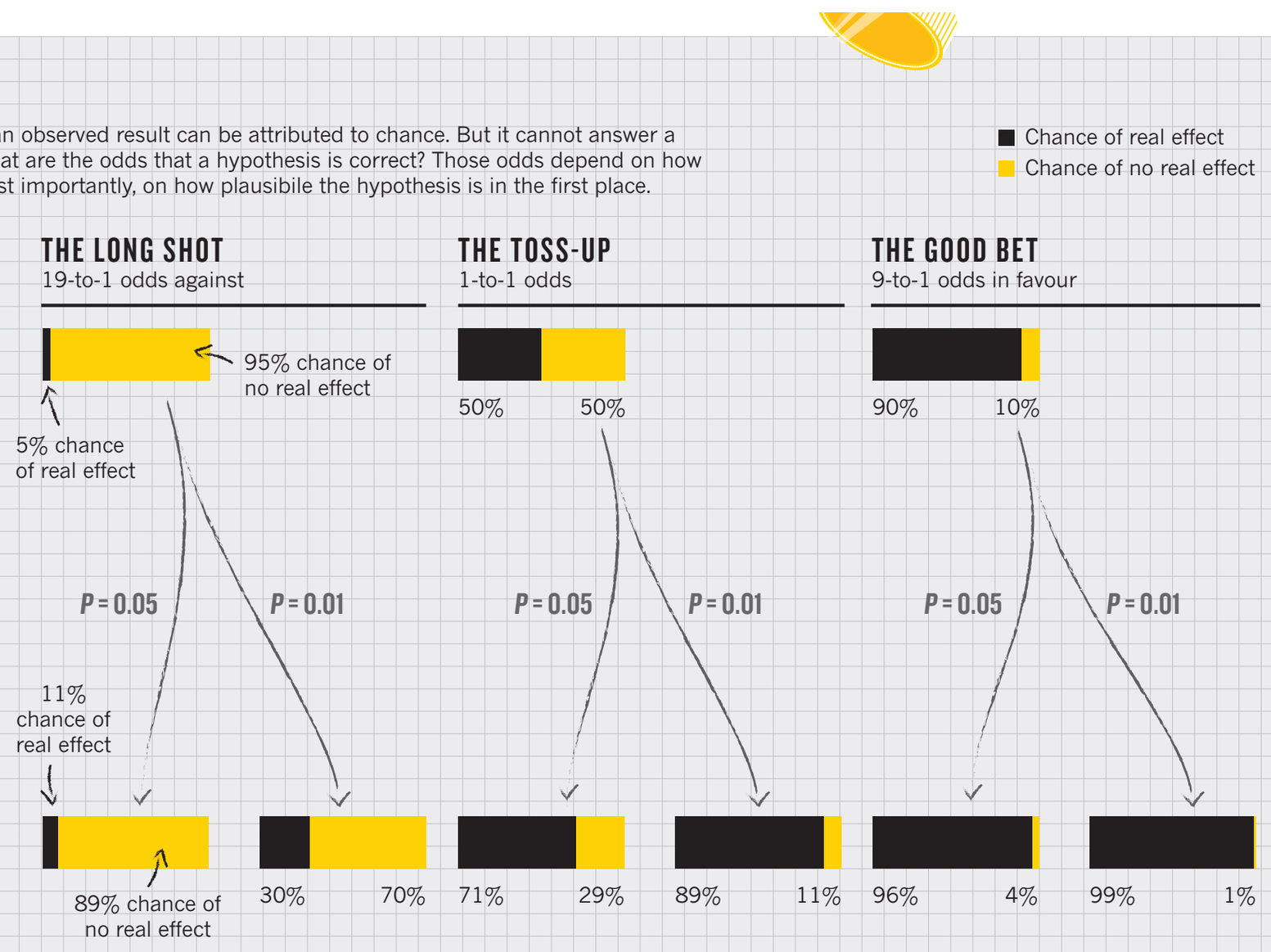
Figure 1. Proportion of 'false-positive' rejections of the null hypothesis as a function of the probability that the hypothesized effect exists (i.e. is 'real'). The curve is drawn for the case when α (the 'significance level' or the putative risk of a type I error) is 0.05 and the power of the test (i.e. the probability of correctly rejecting the null hypothesis when it is false) has the value 0.8 (mimicking the value adopted by Colquhoun [32]).

Loiselle D, Ramchandra R. 2015 A counterview of 'An investigation of the false discovery rate and the misinterpretation of *p*-values' by Colquhoun (2014). *R. Soc. open sci.* **2**: 150217.

NHST problems: what is my error rate?

PROBABLE CAUSE

A *P* value measures whether an observed result can be attributed to chance. But it cannot answer a researcher's real question: what are the odds that a hypothesis is correct? Those odds depend on how strong the result was and, most importantly, on how plausibile the hypothesis is in the first place.



Before the experiment
The plausibility of the hypothesis — the odds of it being true — can be estimated from previous experiments, conjectured mechanisms and other expert knowledge. Three examples are shown here.

The measured *P* value
A value of 0.05 is conventionally deemed 'statistically significant'; a value of 0.01 is considered 'very significant'.

After the experiment
A small *P* value can make a hypothesis more plausible, but the difference may not be dramatic.

The problems with NHST

Null Hypothesis Significant Testing:

p(data|hypothesis): probability of the data given a null hypothesis.

- Interpretation problem: alpha threshold and p values do not determine error rates.

- Practice problems: depends on testing intention, possibilities ('forked paths'), and stopping criterions. Focus on 'significant' difference results. Low threshold for significance. Underpowered studies.

NHST problems: researcher degrees of freedom

Table 1. Likelihood of Obtaining a False-Positive Result

Researcher degrees of freedom	Significance level		
	$p < .1$	$p < .05$	$p < .01$
Situation A: two dependent variables ($r = .50$)	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

Note: The table reports the percentage of 15,000 simulated samples in which at least one of a set of analyses was significant. Observations were drawn independently from a normal distribution. Baseline is a two-condition design with 20 observations per cell. Results for Situation A were obtained by conducting three *t* tests, one on each of two dependent variables and a third on the average of these two variables. Results for Situation B were obtained by conducting one *t* test after collecting 20 observations per cell and another after collecting an additional 10 observations per cell. Results for Situation C were obtained by conducting a *t* test, an analysis of covariance with a gender main effect, and an analysis of covariance with a gender interaction (each observation was assigned a 50% probability of being female). We report a significant effect if the effect of condition was significant in any of these analyses or if the Gender × Condition interaction was significant. Results for Situation D were obtained by conducting *t* tests for each of the three possible pairings of conditions and an ordinary least squares regression for the linear trend of all three conditions (coding: low = −1, medium = 0, high = 1).

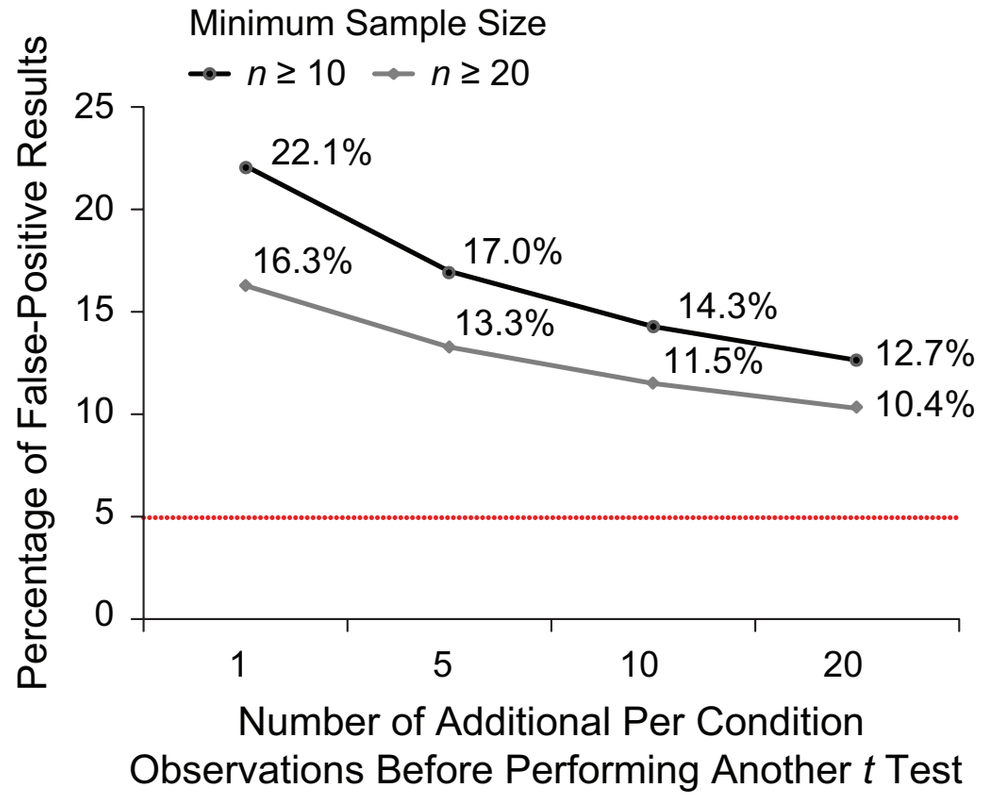


Fig. 1. Likelihood of obtaining a false-positive result when data collection ends upon obtaining significance ($p \leq .05$, highlighted by the dotted line). The figure depicts likelihoods for two minimum sample sizes, as a function of the frequency with which significance tests are performed.

Simons, Nelson & Simonsohn, False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant Psychological Science 22(11) 1359–1366

NHST problems: forking paths

You have a problem with researchers degrees of freedom even when you do just one or two things

"Consider the following testing procedures:

1. Simple classical test based on a unique test statistic, T , which when applied to the observed data yields T(y).

2. Classical test pre-chosen from a set of possible tests: thus, T(y;φ), with preregistered φ. …

3. Researcher degrees of freedom without fishing: computing a single test based on the data, but in an environment where a different test would have been performed given different data; thus T(y;φ(y)), where the function φ(·) is observed in the observed case.

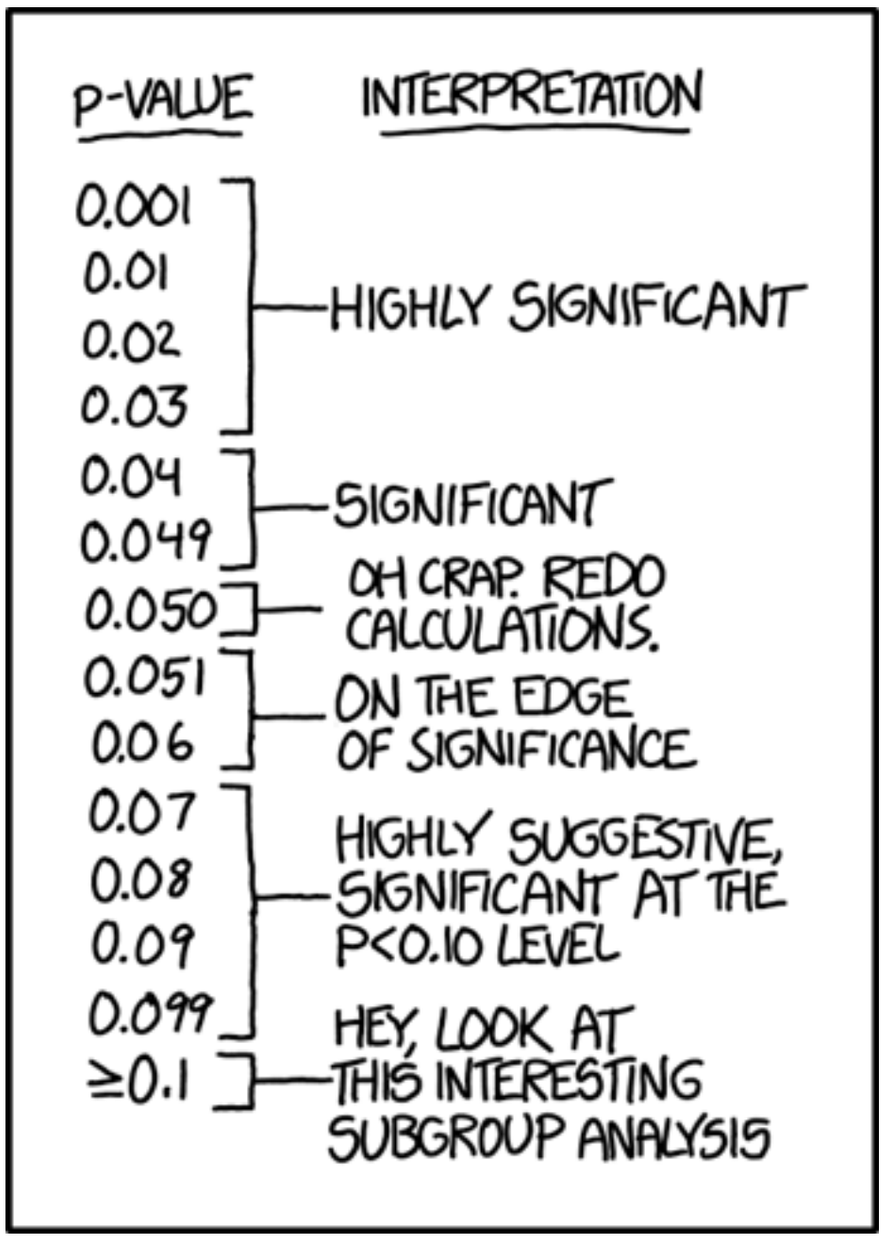
4. "Fishing": computing T(y;φj) for j = 1,...,J: that is, performing J tests and then reporting the best result given the data, thus T (y; φ^{best}(y)).

There is a one-to-many mapping from scientific to statistical hypotheses. "

German & Loken, The garden of forking paths (2014)

Basically, if we look at our data before deciding our complete analysis we are doomed.

NHST problems: sig. threshold, yes-no decisions



https://xkcd.com/1478/

Revised standards for statistical evidence

Valen E. Johnson[1]

Department of Statistics, Texas A&M University, College Station, TX 77843-3143

Edited by Adrian E. Raftery, University of Washington, Seattle, WA, and approved October 9, 2013 (received for review

Recent advances in Bayesian hypothesis testing have led to the development of uniformly most powerful Bayesian tests, which represent an objective, default class of Bayesian hypothesis tests that have the same rejection regions as classical significance tests. Based on the correspondence between these two classes of tests, it is possible to equate the size of classical hypothesis tests with evidence thresholds in Bayesian tests, and to equate *P* values with Bayes factors. An examination of these connections suggest that recent concerns over the lack of reproducibility of scientific studies can be attributed largely to the conduct of significance tests at unjustifiably high levels of significance. To correct this problem, evidence thresholds required for the declaration of a significant finding should be increased to 25–50:1, and to 100–200:1 for the declaration of a highly significant finding. In terms of classical hypothesis tests, these evidence standards mandate the conduct of tests at the 0.005 or 0.001 level of significance.

the average value of the sampling
under each of the two hypothes
the prior density specified on the
each hypothesis.

Paradoxically, the two approac
often produce results that are see
For instance, many statisticians ha
may correspond to Bayes factors
hypothesis by odds of 3 or 4–1 (1
ancy stems from the fact that the
testing are based on the calculat
P values and significance tests are
ability of observing test statistics
extreme than the test statistic actu
factors represent the relative pr
served data under each of the con

Redefine statistical significance

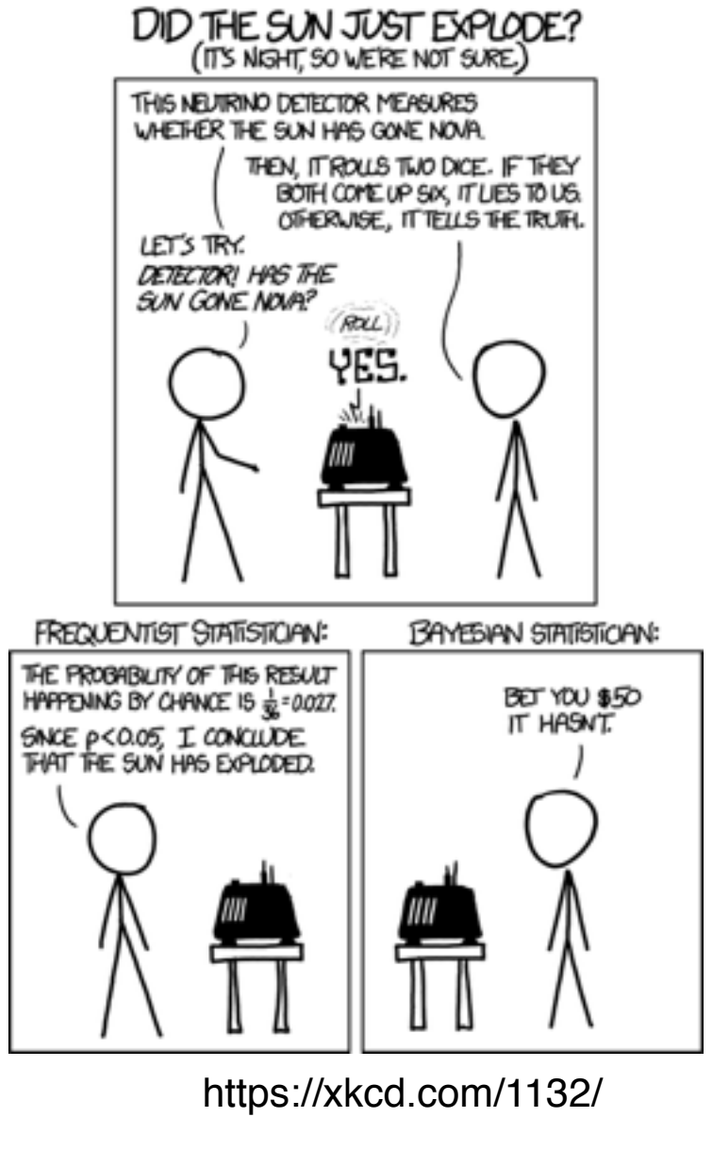
We propose to change the default *P*-value threshold for statistical significance from 0.05 to 0.005 for claims of new discoveries.

Daniel J. Benjamin, James O. Berger, Magnus Johannesson, Brian A. Nosek, E.-J. Wagenmakers, Richard Berk, Kenneth A. Bollen, Björn Brembs, Lawrence Brown, Colin Camerer, David Cesarini, Christopher D. Chambers, Merlise Clyde, Thomas D. Cook, Paul De Boeck, Zoltan Dienes, Anna Dreber, Kenny Easwaran, Charles Efferson, Ernst Fehr, Fiona Fidler, Andy P. Field, Malcolm Forster, Edward I. George, Richard Gonzalez, Steven Goodman, Edwin Green, Donald P. Green, Anthony Greenwald, Jarrod D. Hadfield, Larry V. Hedges, Leonhard Held, Teck Hua Ho, Herbert Hoijtink, Daniel J. Hruschka, Kosuke Imai, Guido Imbens, John P. A. Ioannidis, Minjeong Jeon, James Holland Jones, Michael Kirchler, David Laibson, John List, Roderick Little, Arthur Lupia, Edouard Machery, Scott E. Maxwell, Michael McCarthy, Don Moore, Stephen L. Morgan, Marcus Munafó, Shinichi Nakagawa, Brendan Nyhan, Timothy H. Parker, Luis Pericchi, Marco Perugini, Jeff Rouder, Judith Rousseau, Victoria Savalei, Felix D. Schönbrodt, Thomas Sellke, Betsy Sinclair, Dustin Tingley, Trisha Van Zandt, Simine Vazire, Duncan J. Watts, Christopher Winship, Robert L. Wolpert, Yu Xie, Cristobal Young, Jonathan Zinman and Valen E. Johnson

NHST vs Bayes

NHST:
p(data|hypothesis): probability of the data given a null hypothesis.

Bayes:
p(parameter|data): probability of parameter given the data.



https://xkcd.com/1132/

Bayes' rule derivation:

$$p(A|B) = \frac{p(A,B)}{p(B)}$$

$$p(B|A) = \frac{p(A,B)}{p(A)}$$

$$p(A,B) = p(B|A)p(A)$$

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

Bayes' rule derivation:

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

$$p(B) = p(A_1, B) + p(A_2, B) + ... + p(A_N, B) = \sum_{i=1}^{N} p(A_i, B)$$

$$\sum_{i=1}^{N} p(A_i, B) = \sum_{i=1}^{N} p(B|A_i)p(A_i)$$

$$p(A|B) = \frac{p(B|A)p(A)}{\sum_{i=1}^{N} p(B|A_i)p(A_i)}$$

$$p(\theta|data) = \frac{p(data|\theta)p(\theta)}{\sum_{i=1}^{N} p(data|\theta_i)p(\theta_i)}$$

Bayes' rule: the typical example

$$p(disease|+) = \frac{p(+|disease)p(disease)}{p(+|disease)p(disease) + p(+|healthy)p(healthy)}$$

True positive rate ('Sensitivity'): 99.9%
False positive rate (1-'Specificity):0.001%
Disease prevalence: 0.1%

$$p(disease|+) = \frac{0.999 * .001}{0.999 * .001 + 0.001 * 0.999}$$

$$= \frac{0.000999}{0.000999 + 0.000999} = 0.5$$

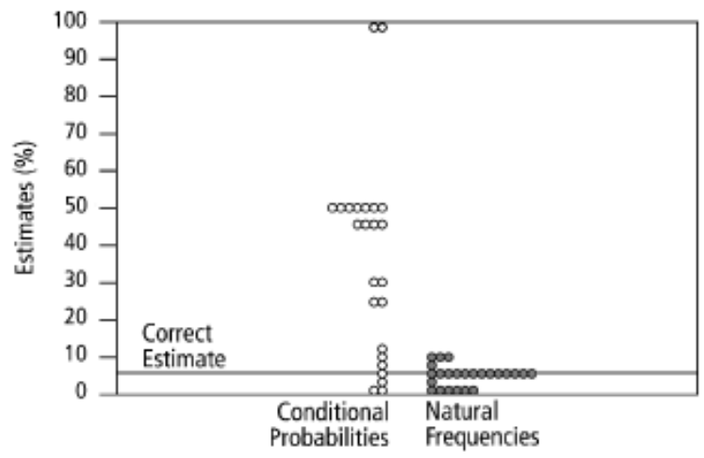


Fig. 9. How to reduce the variability in physicians' judgments. Shown are individual estimates by physicians that a person has colorectal cancer given a positive fecal occult blood test when information was given in conditional probabilities (left) versus natural frequencies (right). Variability decreased dramatically and the correct answer was given more often when numerical information was in natural frequencies (Hoffrage & Gigerenzer, 1998).

Bayes' rule: the typical example

Dichotomous data (-,+)

$$p(disease|+) = \frac{p(+|disease)p(disease)}{p(+|disease)p(disease) + p(+|healthy)p(healthy)}$$

Dichotomous parameter space (disease, healthy)

True positive rate ('Sensitivity'): 99.9%
False positive rate (1-'Specificity):0.001%
Disease prevalence: 0.1%

$$p(disease|+) = \frac{0.999 * .001}{0.999 * .001 + 0.001 * 0.999}$$
$$= \frac{0.000999}{0.000999 + 0.000999} = 0.5$$

$$p(healthy|+) = 1 - p(disease|+) = 0.5$$

Probability mass and density functions

Probability distribution indicate the probability of all possible values of a variable.

Discrete variable either by probability mass or density functions. Probability masses across all possible values (or intervals) are < 1 and sum to 1.

Continuous variable are described by probability density functions. Probability density values are >0 and do not sum to 1. Density value tells about the amount of probability mass relative to the scale (infinitesimal interval)

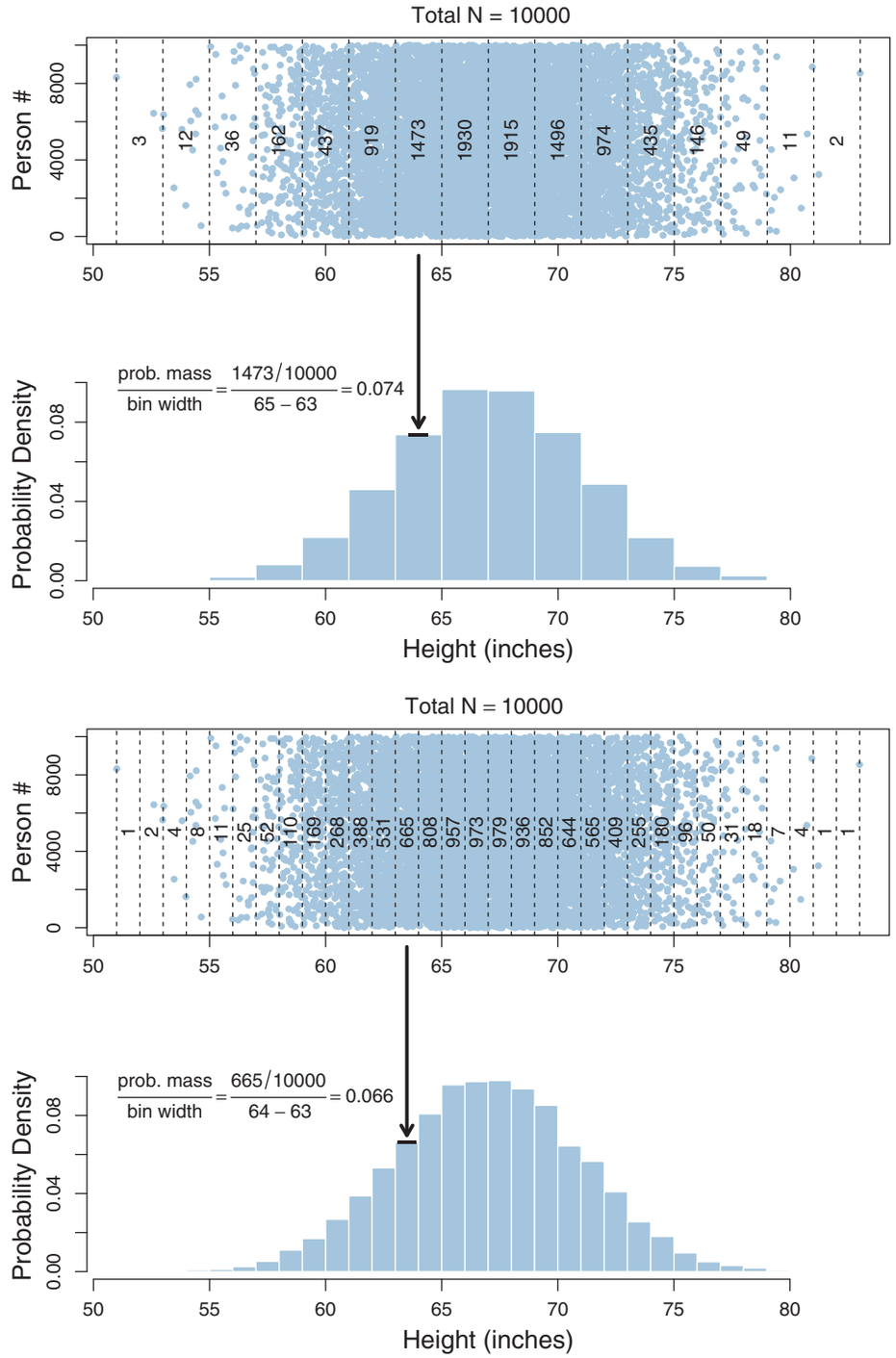


Figure 4.2 Examples of computing probability density. Within each main panel, the upper plot shows a scatter of 10,000 heights of randomly selected people, and the lower plot converts into probability density for the particular selection of bins depicted.

HDI: high density intervals

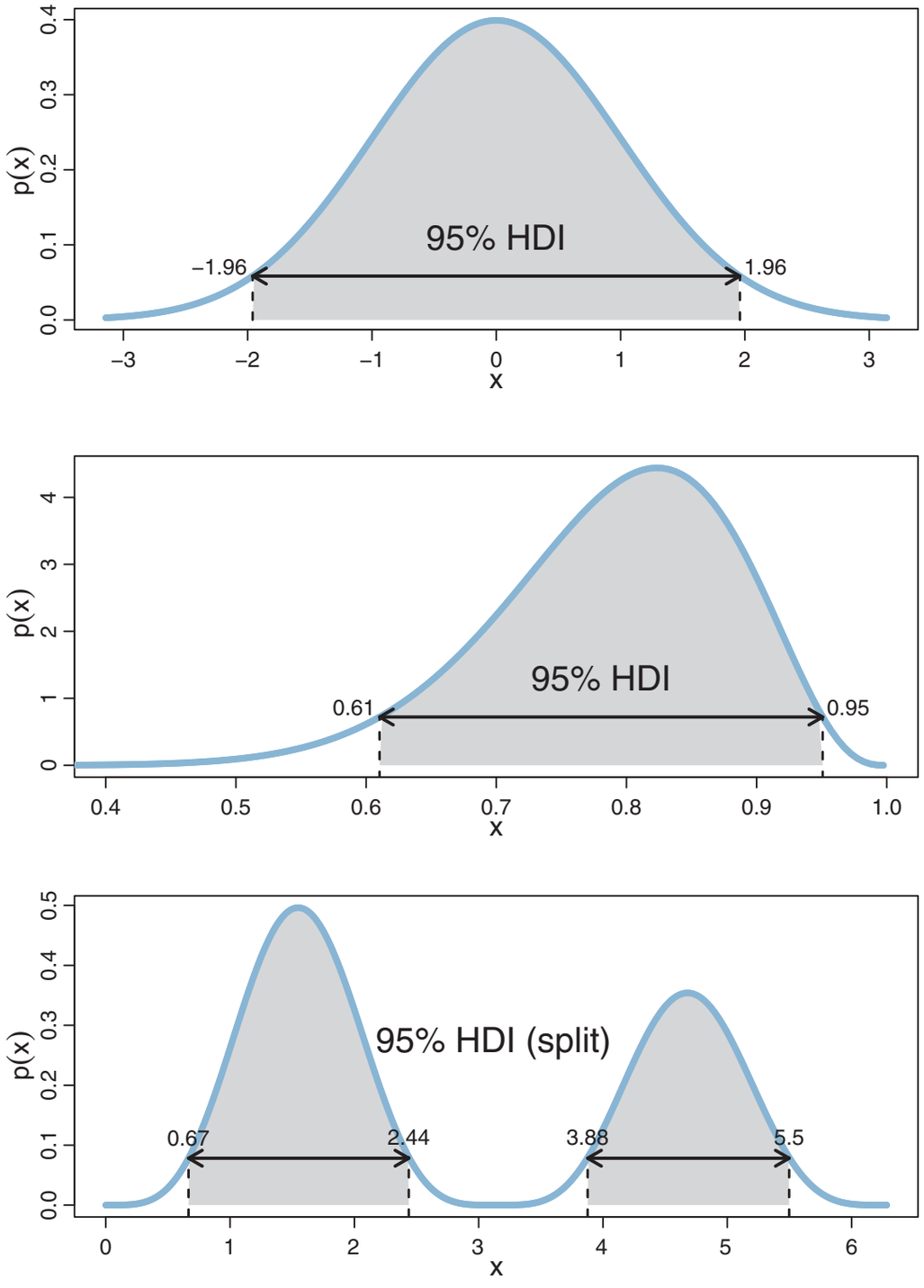


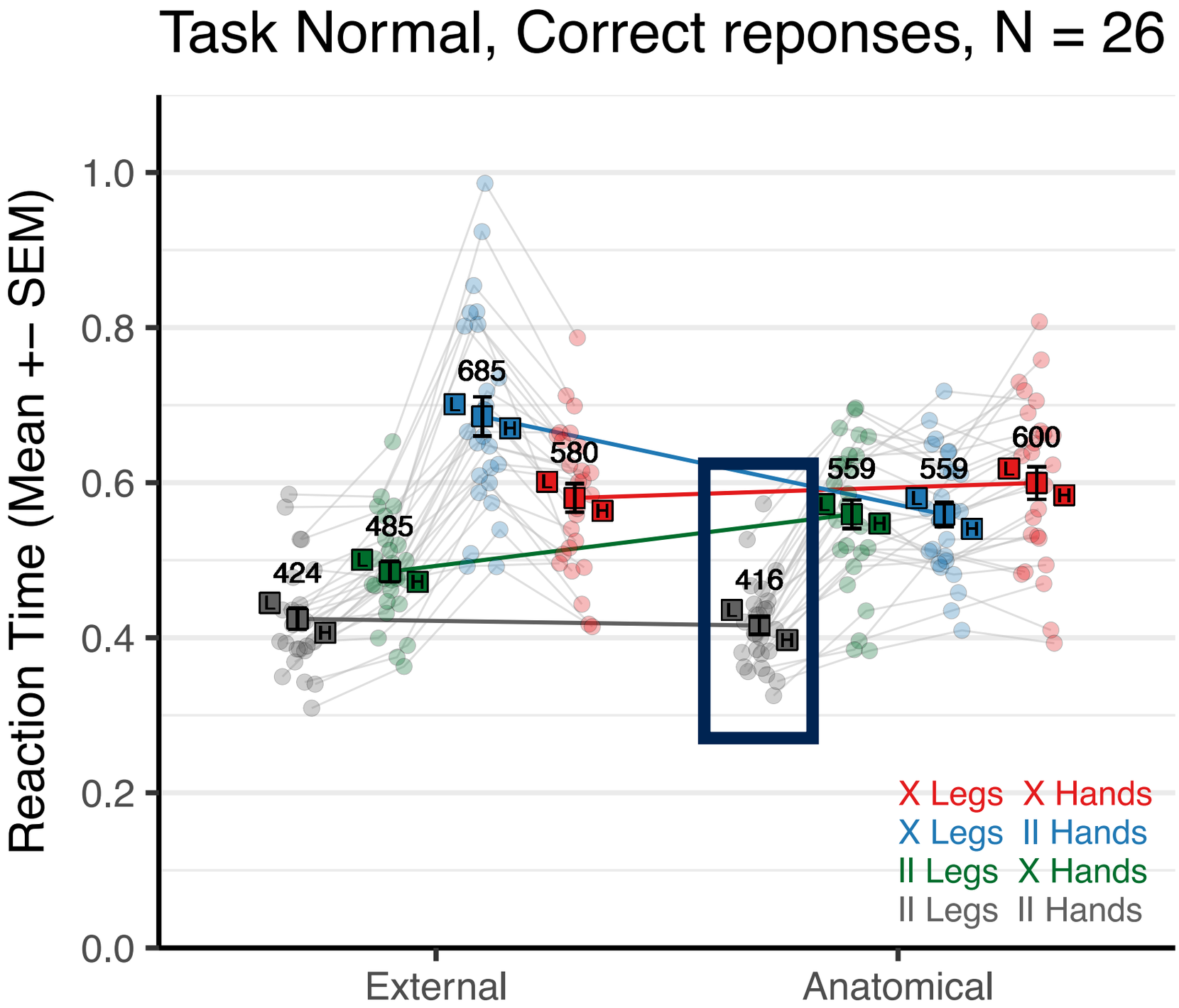
Figure 4.5 Examples of 95% highest density intervals (HDIs). For each example, all the *x* values inside the interval have higher density than any *x* value outside the interval, and the total mass of the points inside the interval is 95%. The 95% area is shaded, and it includes the zone below the horizontal arrow. The horizontal arrow indicates the width of the 95% HDI, with its ends annotated by (rounded) *x* values. The height of the horizontal arrow marks the minimal density exceeded by all *x* values inside the 95% HDI.

Exercise 1

Estimation continuous variable

(effects of prior and amount of data)

data: subjects means || Legs || hands



Estimation of a continuous variable μ and σ

Continuous normal variable posterior for μ and σ is given by:

'Continuous' data (e.g. 0-2000 ms)

Likelihood **bivariate** prior

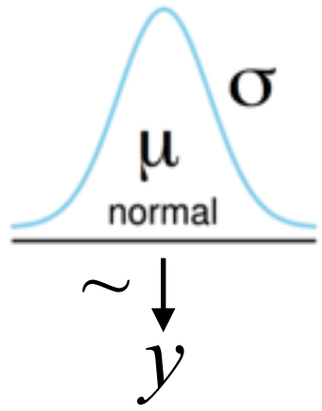
$$p(\mu, \sigma|y) = \frac{p(y|\mu, \sigma)p(\mu, \sigma)}{\int\int p(y|\mu, \sigma)p(\mu, \sigma)\ d\mu\ d\sigma}$$

Evidence

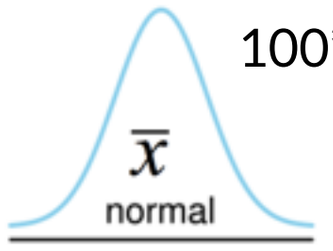
Continuous 2D parameter space

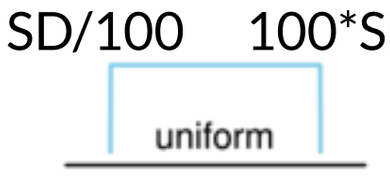
We can use the normal distribution probability density as the likelihood function:

$$p(y|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$



Priors are needed for both the μ and σ parameter. An uninformative μ prior is a normal variable centered at the sample mean with large SD (e.g. 100 x sample SD). An uninformative SD prior is an uniform distribution between 0 and a large number (e.g. sample SD*1000)





Estimation of a continuous variable μ and σ

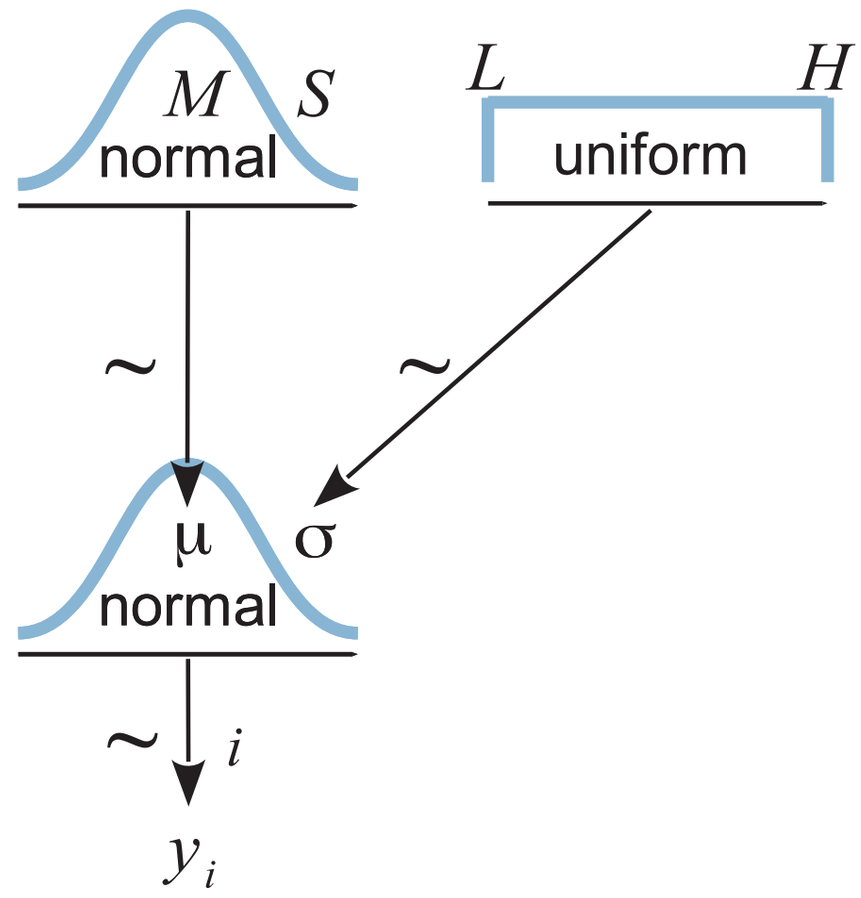
Continuous normal variable posterior for μ and σ is given by:

Likelihood **bivariate** prior

$$p(\mu, \sigma|y) = \frac{p(y|\mu, \sigma)p(\mu, \sigma)}{\int\int p(y|\mu, \sigma)p(\mu, \sigma)\ d\mu\ d\sigma}$$

how do we compute this??

$$p(y|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$



How do we get the posterior distribution?

The posterior distribution can be computed:

- by direct calculation with discrete and small parameters spaces.
- by analytic solution, needs conjugate priors (this is possible for simple models).
- by grid estimation, dividing a continuous parameter space in small segments, this is only possible when estimating a few parameters. When the parameter space has n1 parameters dimension with n2 possible values each, then the number of possible parameters combinations is n2^n1
- Sampling a sufficient number of values from the posterior distribution by Markov chain Monte Carlo (metropolis, gibbs, hamiltonian, etc). This is what are we going to do with R and JAGS (Just Another Gibbs Sampler).

Metropolis (easier to explain than Gibbs)

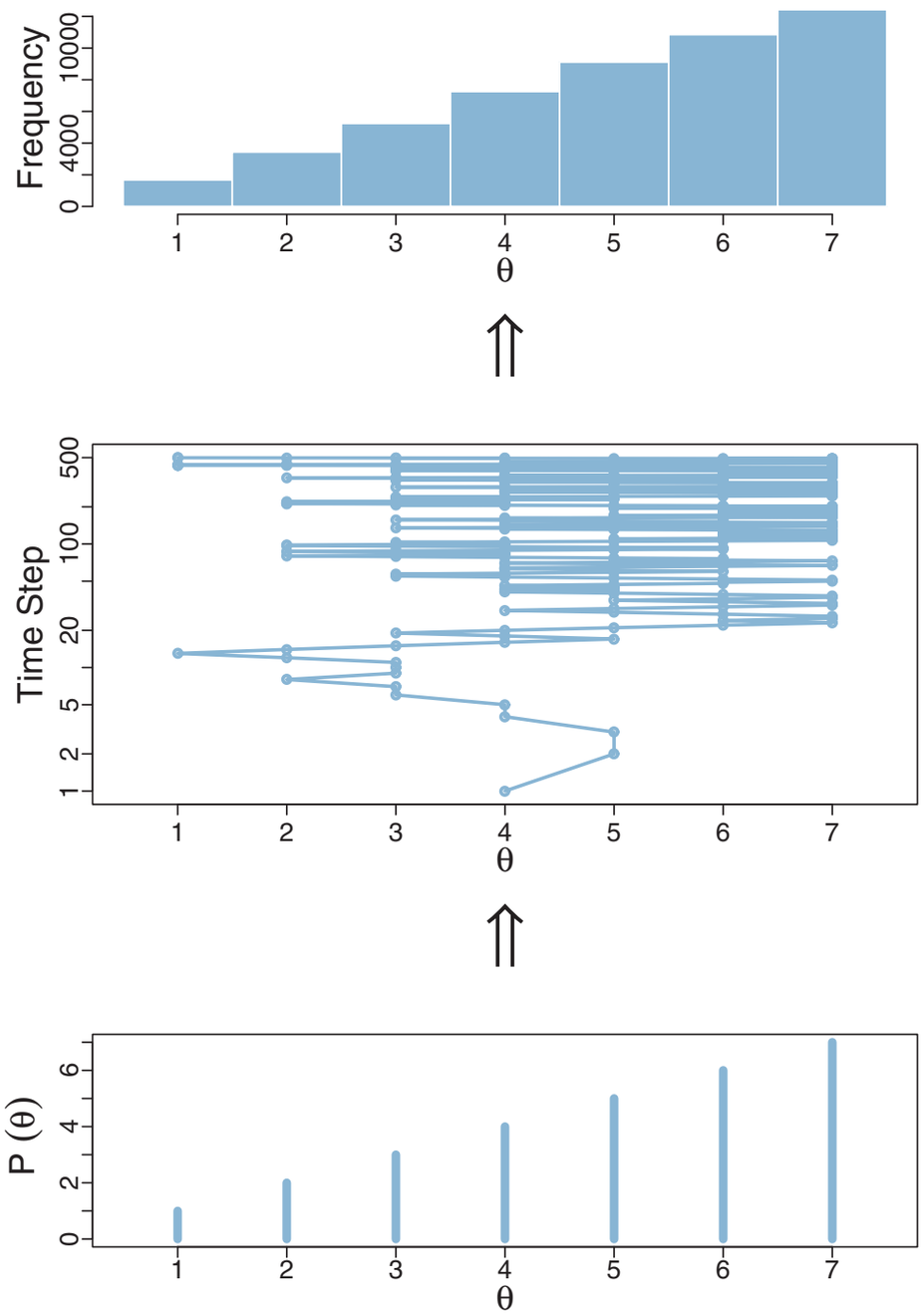


Figure 7.2 Illustration of a simple Metropolis algorithm. The bottom panel shows the values of the target distribution. The middle panel shows one random walk, at each time step proposing to move either one unit right or one unit left, and accepting the proposed move according the heuristic described in the main text. The top panel shows the frequency distribution of the positions in the walk.

Likelihood bivariate

$$p(\mu,\sigma|data) = \frac{p(data|\mu,\sigma)p(\mu,\sigma)}{\int\int p(data|\mu,\sigma)p(\mu,\sigma)d\mu\, d\sigma}$$

Evidence

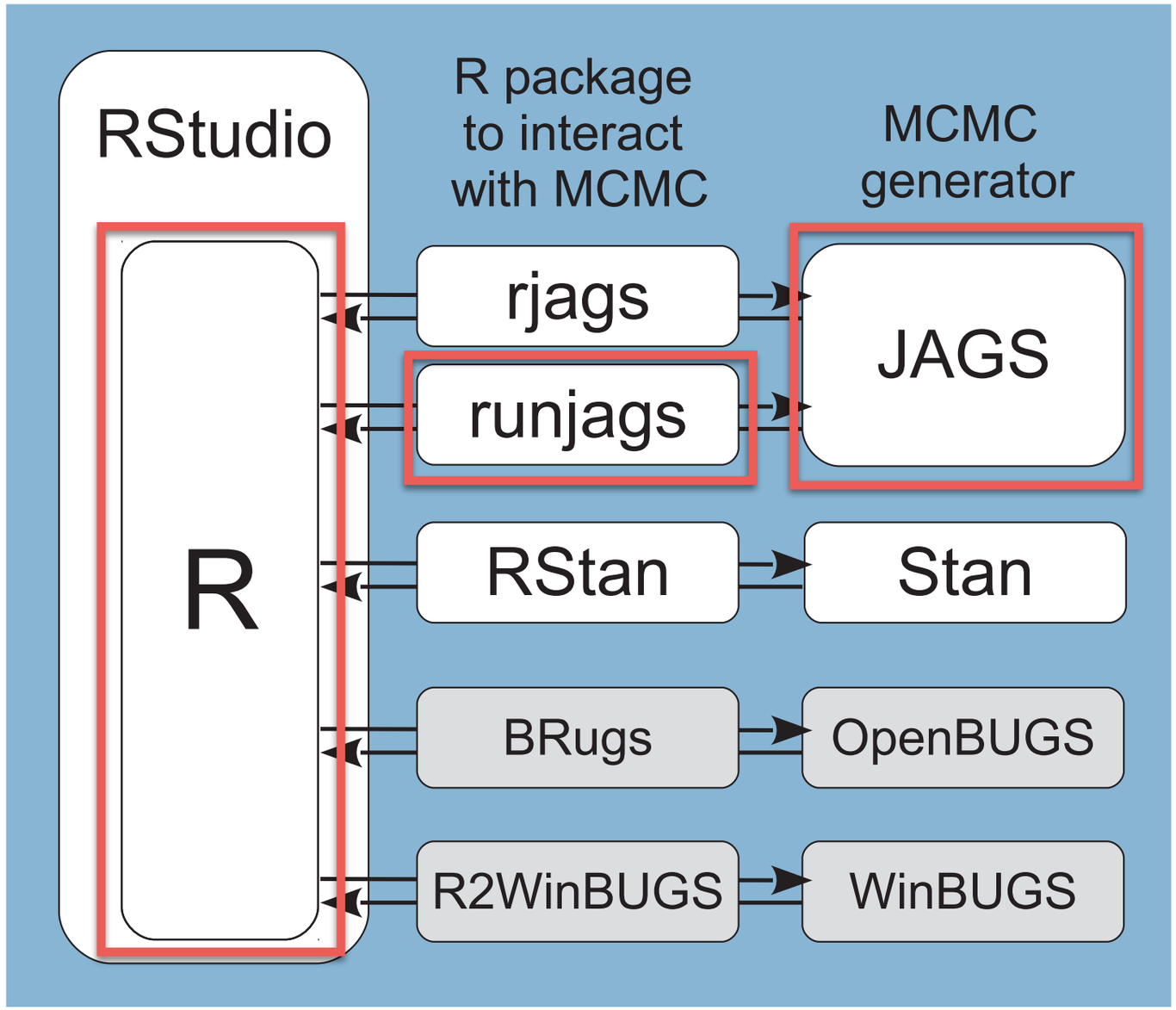
$$p(\theta|data) = \frac{p(data|\theta)p(\theta)}{\sum_{i=1}^{N} p(data|\theta_i)p(\theta_i)}$$

1. Start at some position
2. Choose randomly a proposed new location
3. Calculate likelihood*prior at current and proposed

$$\frac{p(\theta_p|data)}{p(\theta_c|data)} = \frac{\frac{p(data|\theta_p)*p(\theta)}{p(data)}}{\frac{p(data|\theta_c)*p(\theta)}{p(data)}} = \frac{p(data|\theta_p)*p(\theta)}{p(data|\theta_c)*p(\theta)}$$

4. Go to *θp if p(θp|data) > p(θc|data)*
5. *If <, go to θp if a random number between 0 and 1 is <* $\frac{p(\theta_p|data)}{p(\theta_c|data)}$
6. This guarantees that *θ* values will be visited proportionally to their posterior distribution, thus giving us the posterior.

Using R and JAGS



Using R and JAGS

Get the data:

```
load(…)
read.csv(…)
…
```

Data as a list:

```
dataList =
list(y = data,
     N = length(data),
     …)
```

Model as a text:

```
mStr = "model{
         for(i in 1:nTotal){
           y[i] ~ <likelihood>
         }
         <priors>
        }"
writeLines(mStr, con="model.txt" )
```

Initialize the chains with a list:

initsList = list(<prior values>)

run the model:

```
model = run.jags(file   = model.txt,
                 data   = dataList
                 inits = initsList,
                 …)
```

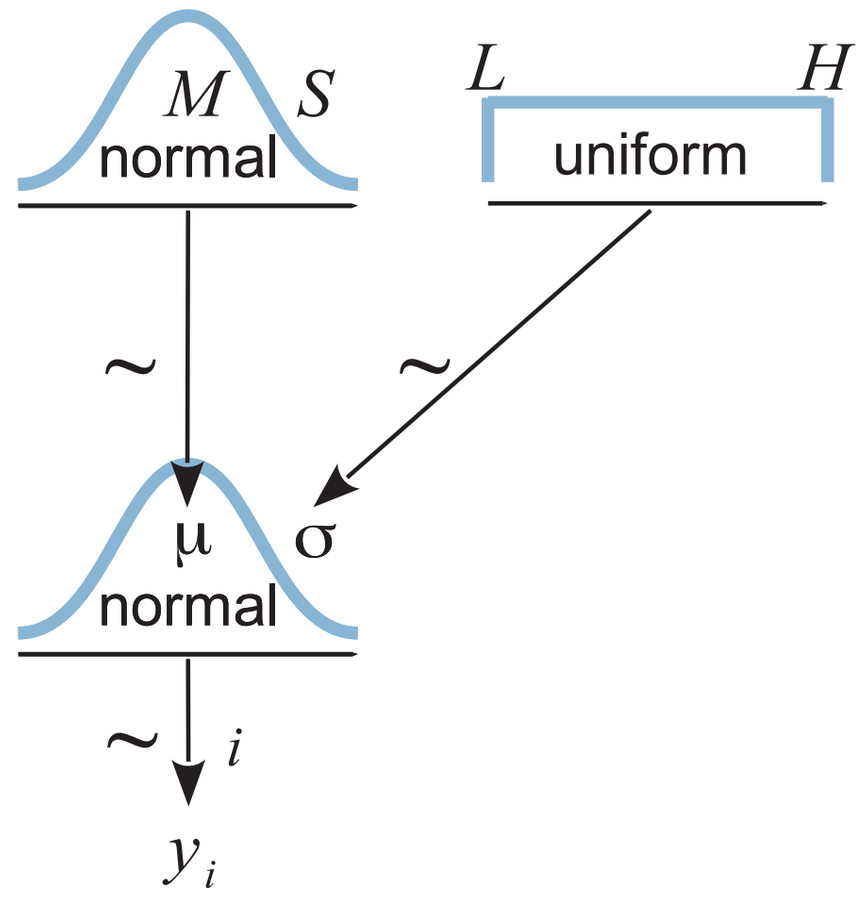
Estimation of a continuous variable μ and σ

Continuous normal variable posterior for μ and σ is given by:

Likelihood **bivariate** prior

$$p(\mu, \sigma|y) = \frac{p(y|\mu, \sigma)p(\mu, \sigma)}{\int\int p(y|\mu, \sigma)p(\mu, \sigma)\ d\mu\ d\sigma}$$

$$p(y|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

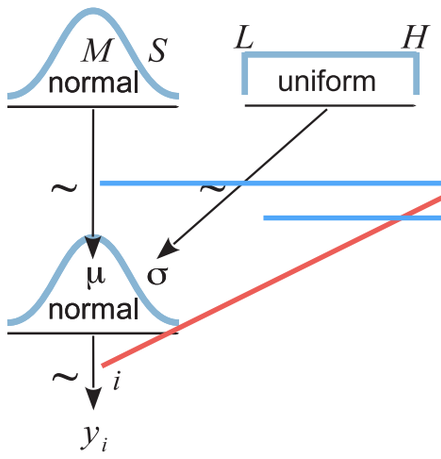


Using R and JAGS

Data as a list:

```
dataList = list(y          = dFc$mRT,
                nTotal     = length(dFc$mRT),
                MPriorstd  = sd(data)*100,
                MPriormean = mean(data),
                SPriorL    = sd(dFc$mRT)/1000,
                SPriorH    = sd(dFc$mRT)*1000)
```

Model as a text:




```
modelStr = 'model{
        for ( i in 1:nTotal ) {
            y[i] ~ dnorm(mu, 1/sigma^2)
        }
        mu ~ dnorm(MPriormean, 1/(MPriorstd)^2)
        sigma ~ dunif(SPriorL, SPriorH )
            }'
writeLines(modelStr, con="model.txt")
```

'precision' instead of SD

Likelihood

priors

Initialize the chains with a list:

```
initsList = list(mu    = mean(data),
                 sigma = sd(data))
```

run the model:

```
parameters    = c("mu","sigma")
runjagsModel  = run.jags("model.txt",
                         data     = dataList,
                         sample   = 1000,
                         monitor  = parameters,
                         inits    = initsList,
                         n.chains = 4,…)
```

R exercise: effects of prior and data size

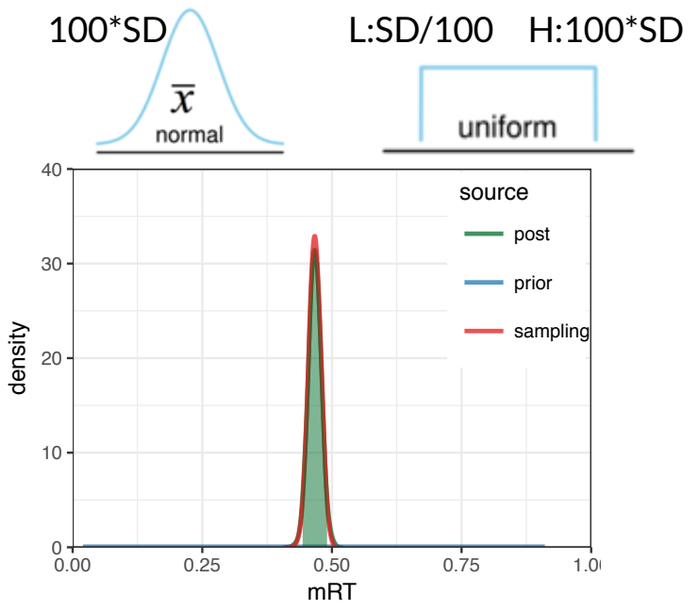
1. Read and run continuous1.R

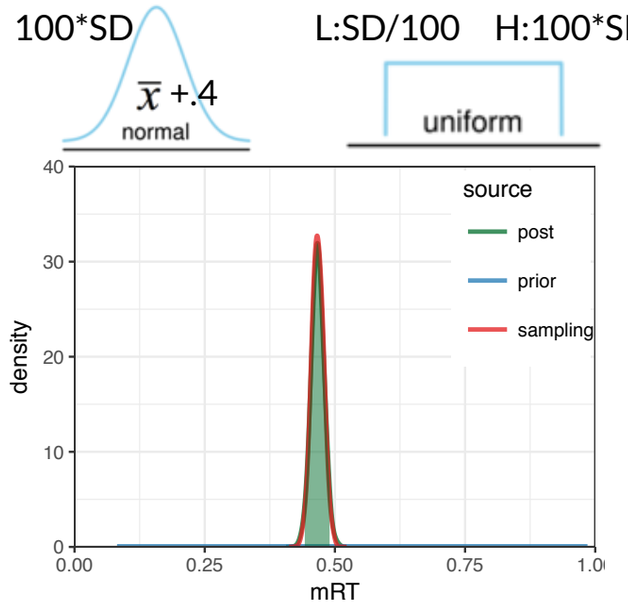
2. Compare posterior to prior and sampling distribution when the prior are uninformative.

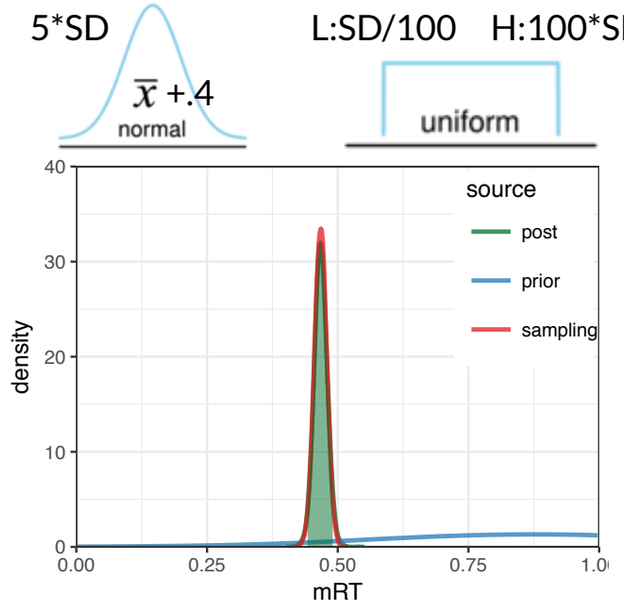
3. Modify the posterior estimate by changing the prior of the mean:
   - By changing the prior mean.
   - By changing the prior standard deviation.
   - By changing both.

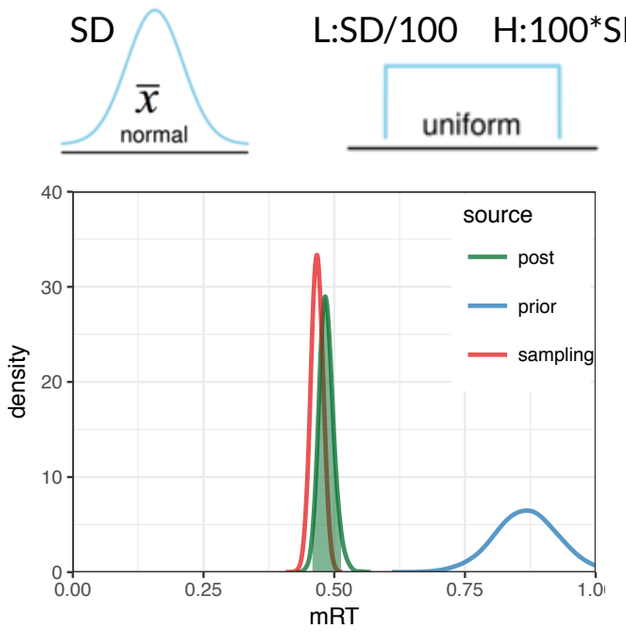
4. Keep an informative prior that does not change much the posterior from an uninformative prior, see what happened when the sample is changed:
   - Lower the sample to 50% of participants.

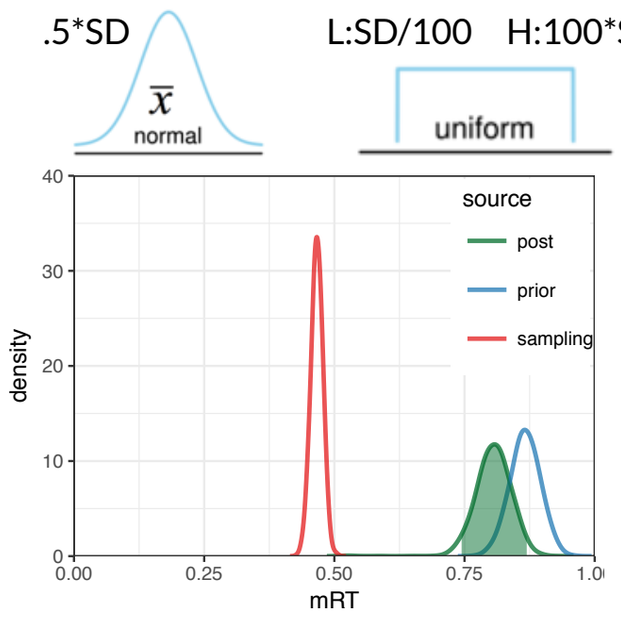
R exercise: effects of prior and data size

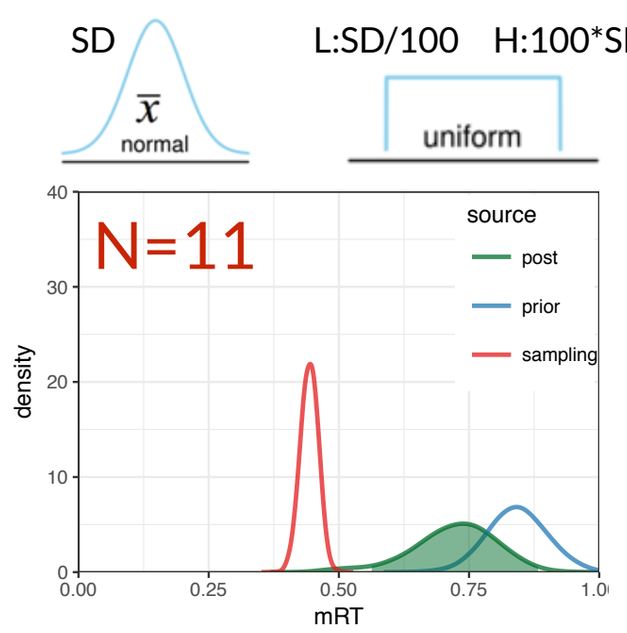










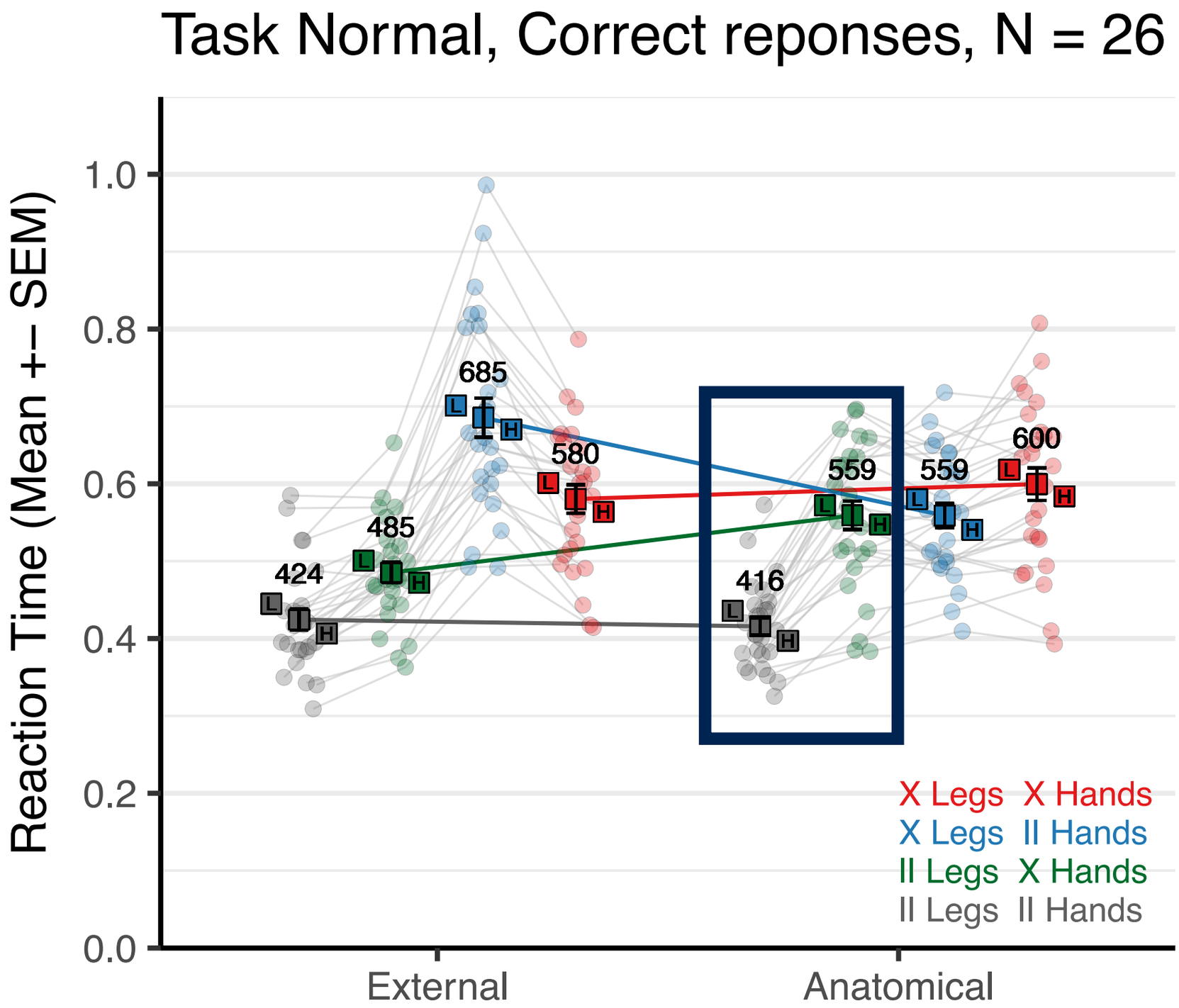


Exercise 2

Difference continuous variables

(convergence checks, comparison between normal and t likelihood distribution, effect size)

data: subjects means || Legs || hands & || Legs x Hands



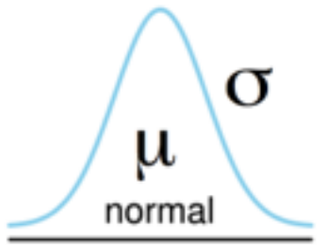
Comparison of a continuous variable μ and σ

Continuous normal variable posterior for μ and σ is given by:

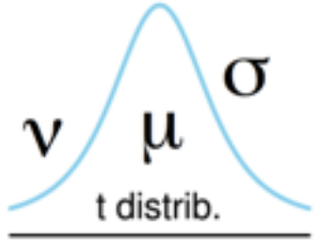
Likelihood **bivariate** prior

$$p(\mu, \sigma|y) = \frac{p(y|\mu, \sigma)p(\mu, \sigma)}{\int\int p(y|\mu, \sigma)p(\mu, \sigma)\ d\mu\ d\sigma}$$

We can use the normal distribution probability density function or the t distribution pdf as the likelihood function:

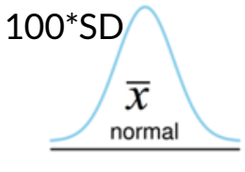


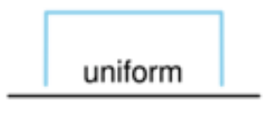
$$p(y|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

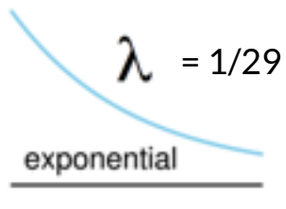


$$p(y|\mu, \sigma, \nu) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})\sigma\sqrt{\pi\nu}} \left(1 + \frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}$$

Priors are needed for μ , σ, and ν (student t). An uninformative μ prior is a normal variable centered at the sample mean with large SD (e.g. 100 x sample SD). An uninformative SD prior is an uniform distribution between 0 and a large number (e.g. sample SD*1000). An uninformative ν prior is taken from an exponential distribution with a small λ rate (e.g. 1/29)



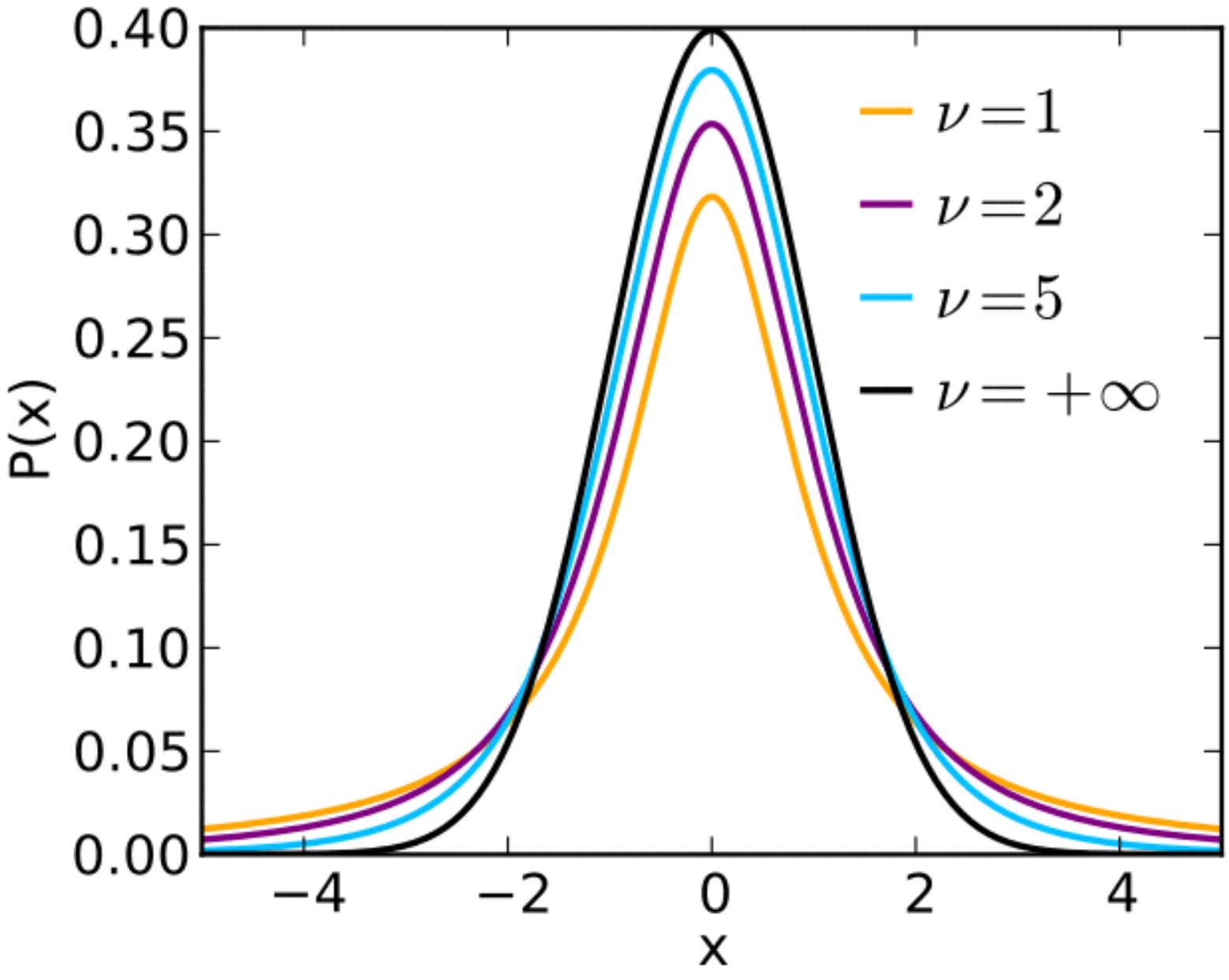




$\nu = \lambda+1$

Student's t PDF

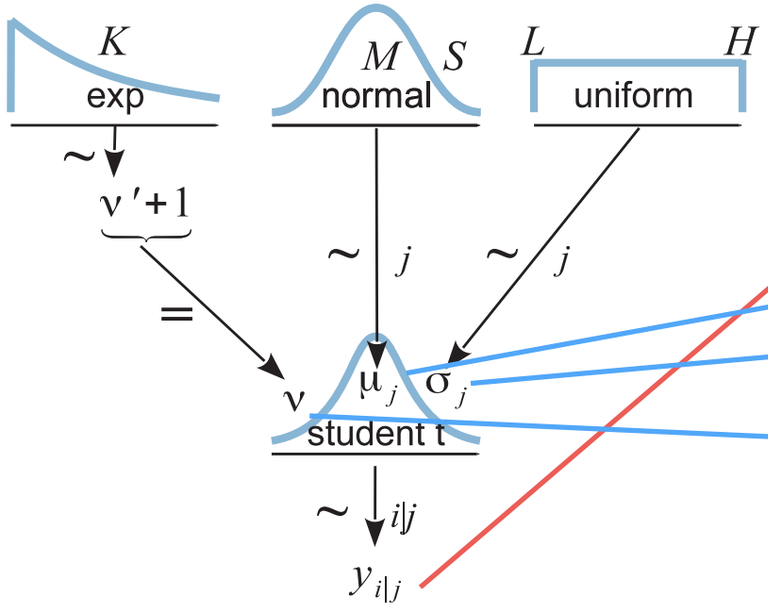
$$p(y|\mu,\sigma,\nu) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})\sigma\sqrt{\pi\nu}}\left(1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}$$



NORMAL PDF

Comparison of a continuous variable μ and σ

```
cond        = as.numeric(dFc$cond)
dataList    = list(y         = dFc$mRT,
                   cond      = cond,
                   nCond     = length(unique(cond)),
                   nTotal    = length(dFc$mRT),
                   MPriormean = mean(dFc$mRT),
                   MPriorstd  = sd(dFc$mRT)*100,
                   SPriorL    = sd(dFc$mRT)/1000,
                   SPriorH    = sd(dFc$mRT)*1000)
```




```
modelString = 'model{
  for ( i in 1:nTotal ) {
    y[i] ~ dt( mu[cond[i]] , 1/sigma[cond[i]]^2 , nu )
  }
  for ( j in 1:nCond ) {
    mu[j] ~ dnorm(MPriormean, 1/(MPriorstd)^2)
    sigma[j] ~ dunif(SPriorL, SPriorH )
  }
  nu <- nuMinusOne+1
  nuMinusOne ~ dexp(1/29)
}'
writeLines(modelString, con="model2.txt" )
```

Now we fit *j* ('nCond') parameters for μ , σ

R exercise: convergence, effects size, outliers

1. Read and run continuous2.R.

2. First we check the convergence tests. Run the model again with adapt=1, burn=1, and sample = 100, then 1000, then 10000.

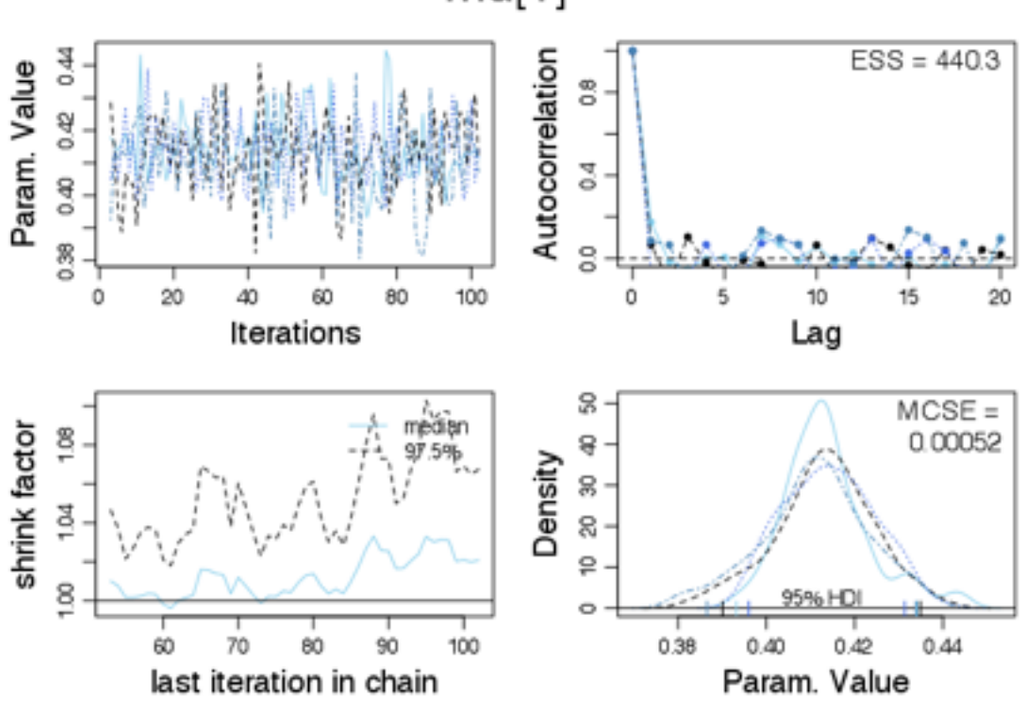
3. Plot all the parameters and their difference

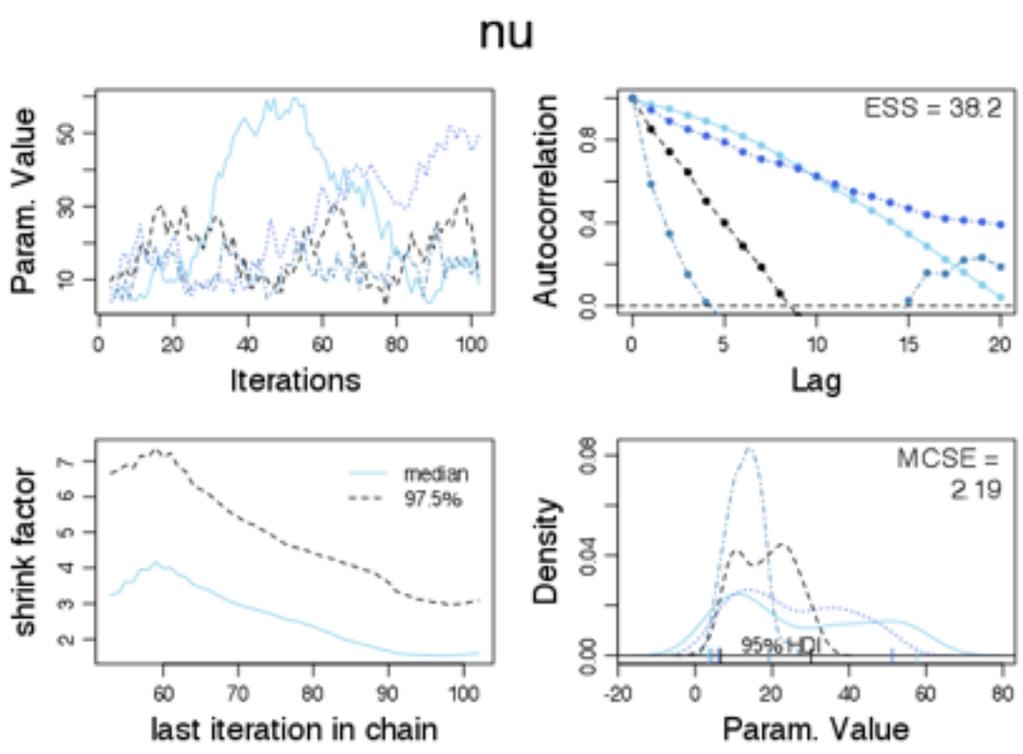
4.. Add a plot of the effect size $Effect\ Size = \frac{\mu_2 - \mu_1}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}$

6. Add one subject with an extreme outlier value in one group and a mean value in the other one, run again and look at the mean, sd, and effect size estimates. Change the model to use a normal likelihood like in continuous1.R and compare.

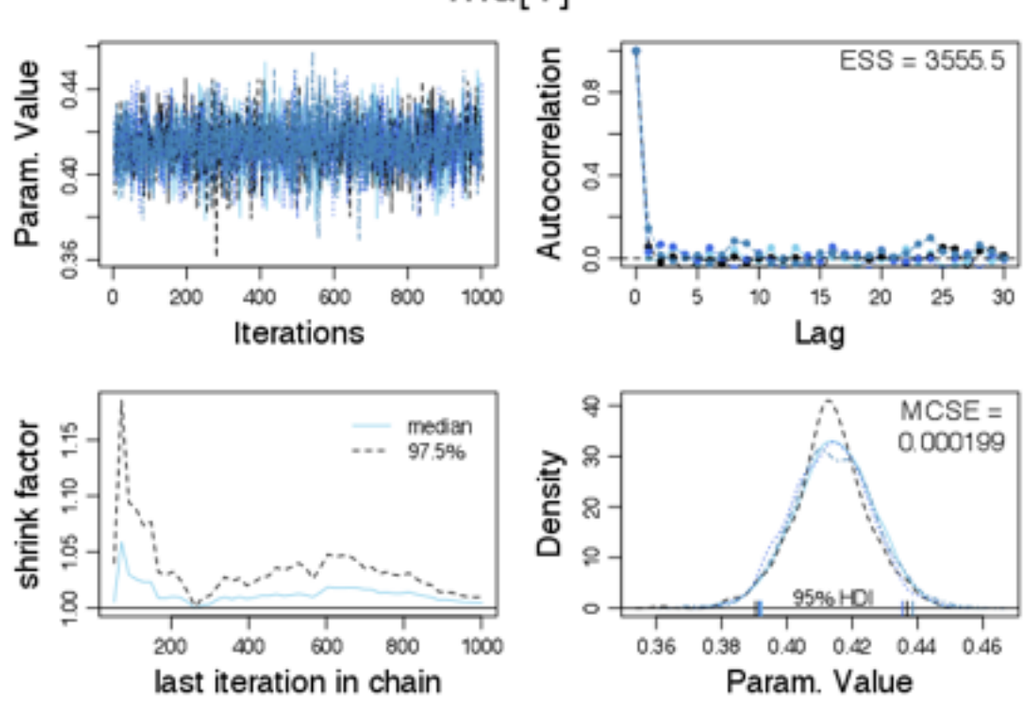
Convergence 100 sampleschecks

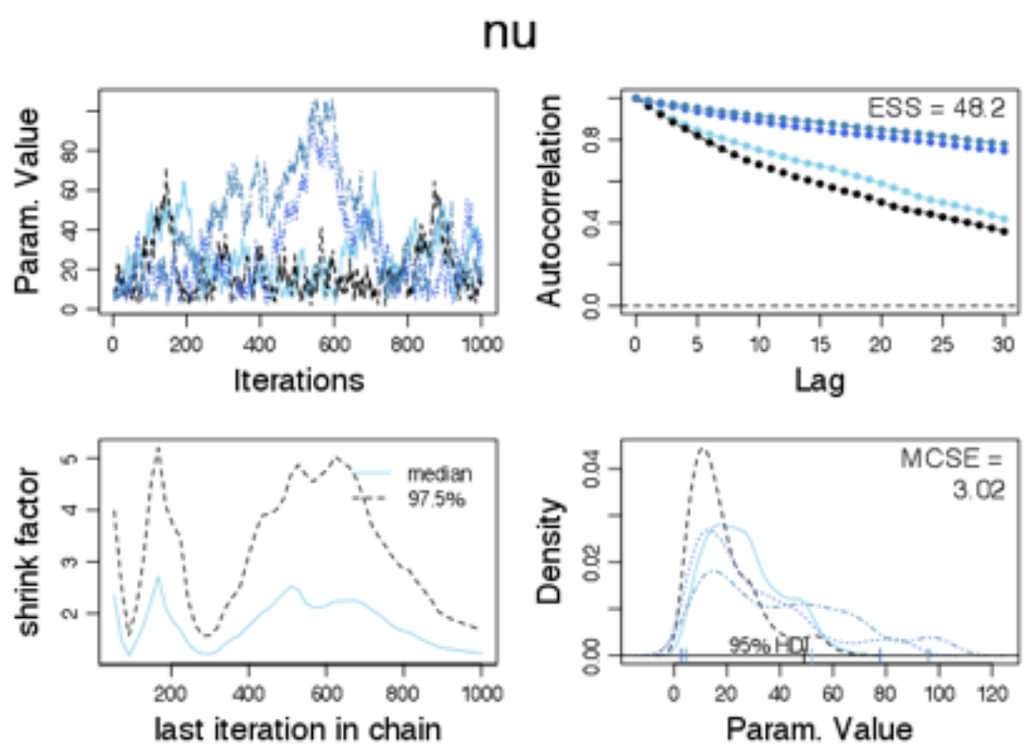
100 samples





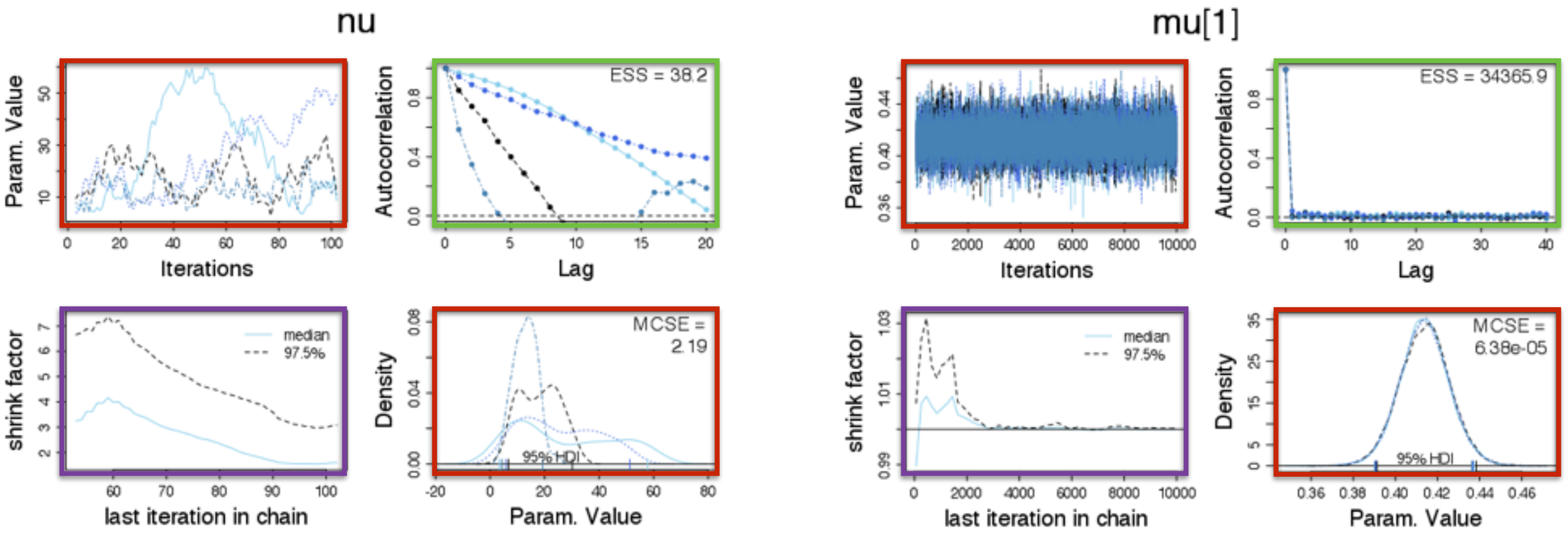
1000 samples





Convergence checks

1. Chain convergence
2. Variability between/within chain < 1.1
3. Low autocorrelation between samples.
Effective sample size (ESS ~ sample size/ACF) ~10000



Adapt samples is for JAGS/BUGS to do some magic
Burn-in samples is to discard initial sample when chains have not converged yet

Convergence checks: autocorrelation

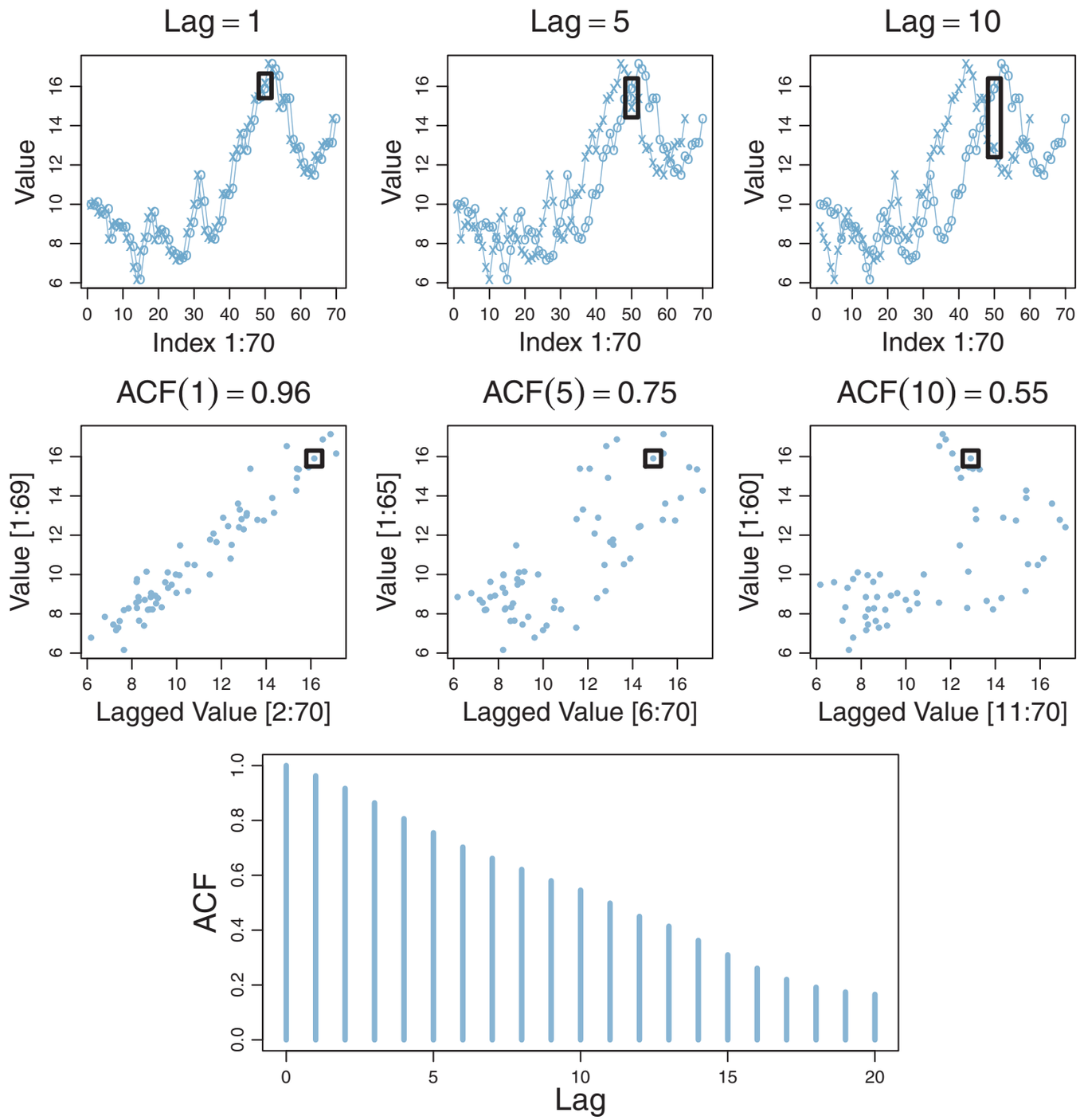
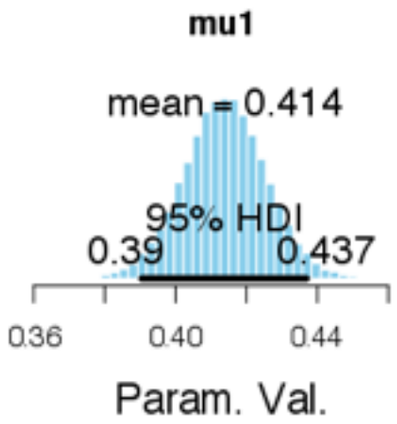
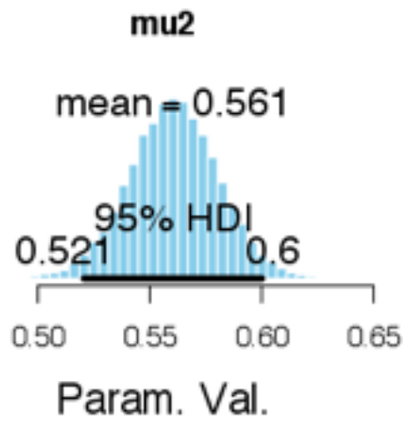
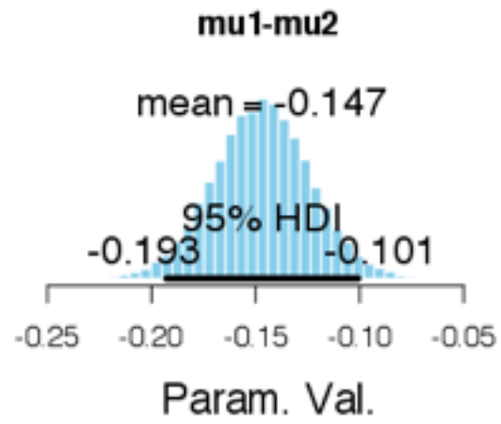


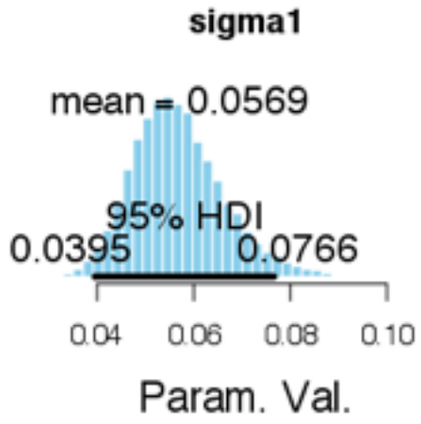
Figure 7.12 Autocorrelation of a chain. Upper panels show examples of lagged chains. Middle panels show scatter plots of chain values against lagged chain values, with their correlation annotated. Lowest panel shows the autocorrelation function (ACF).

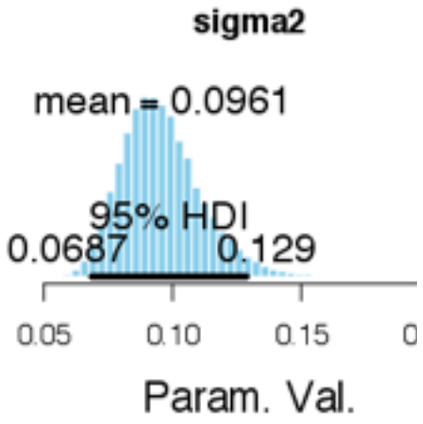
Comparison of a continuous variable μ and σ

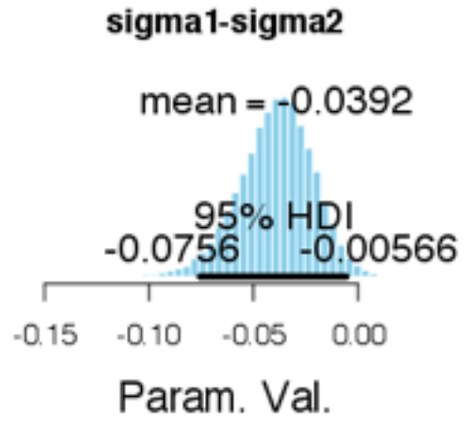


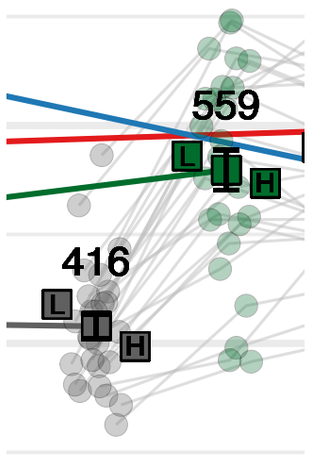


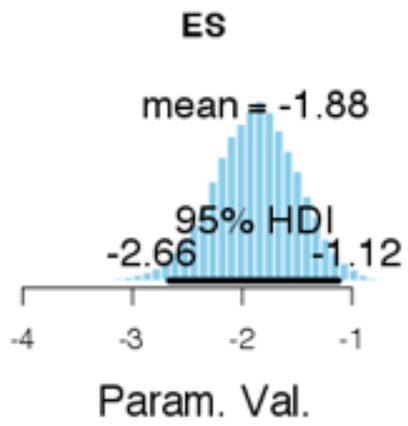


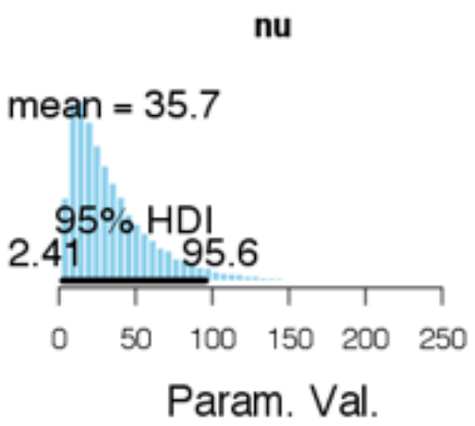




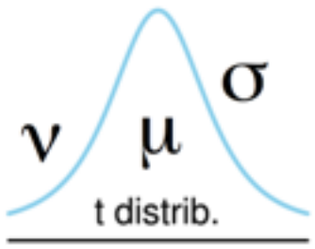


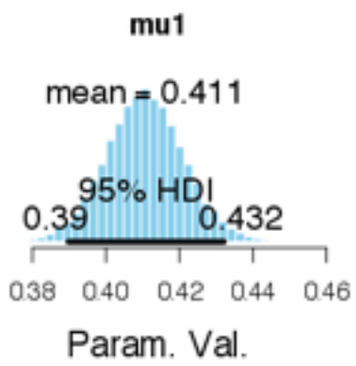


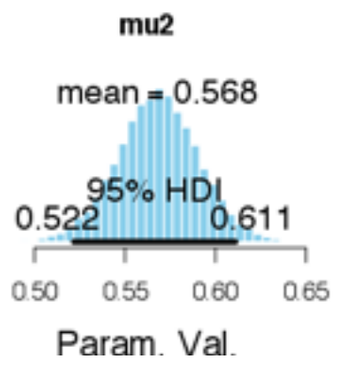


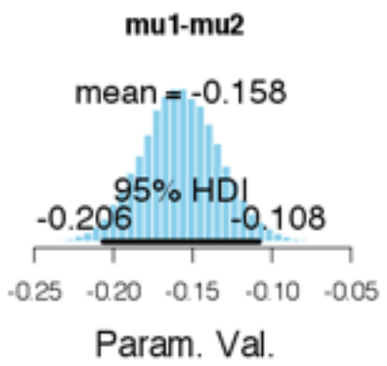


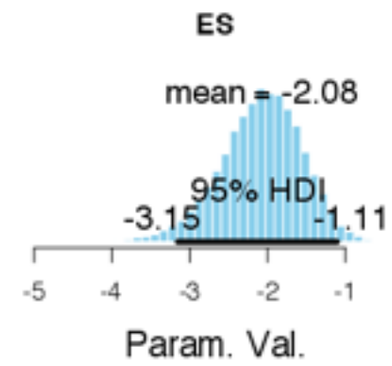
Comparison of models with outlier 2.5*mean

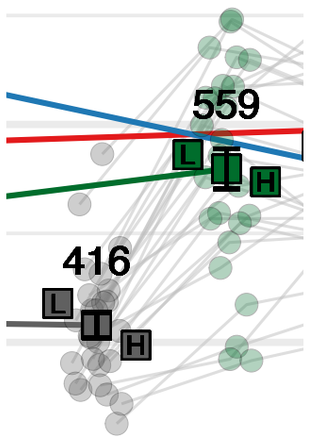


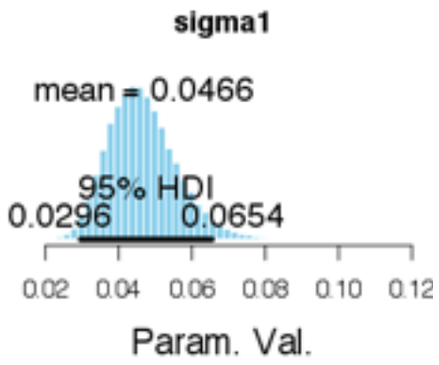


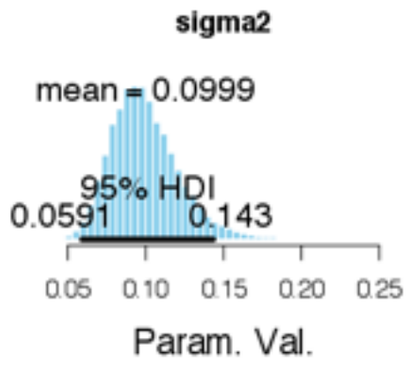


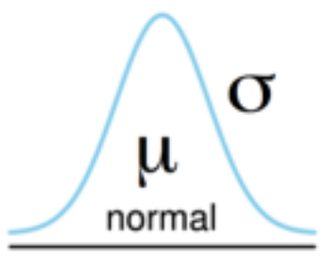


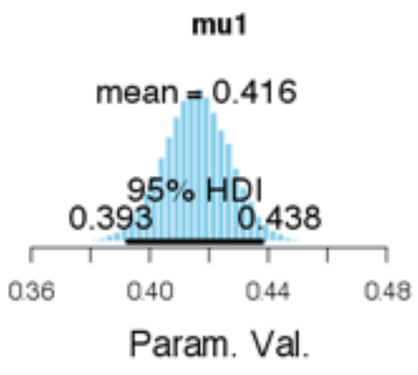


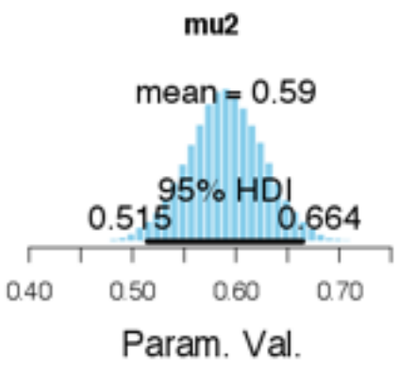


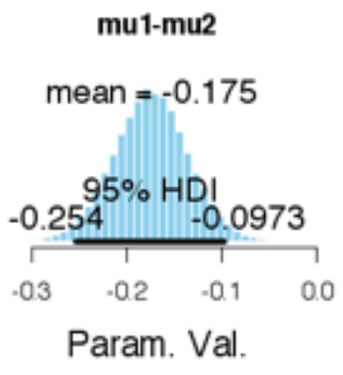


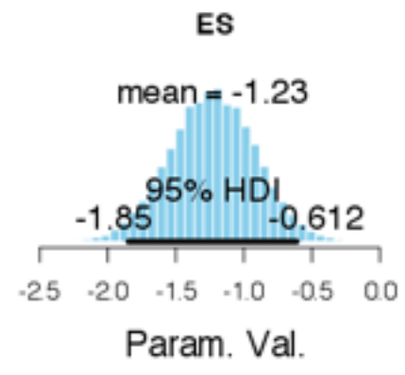


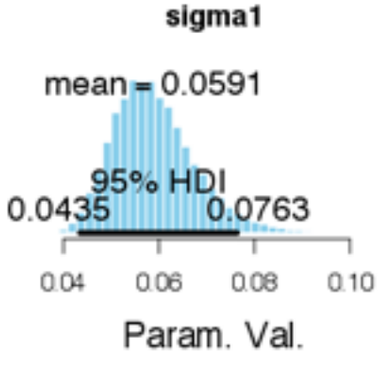


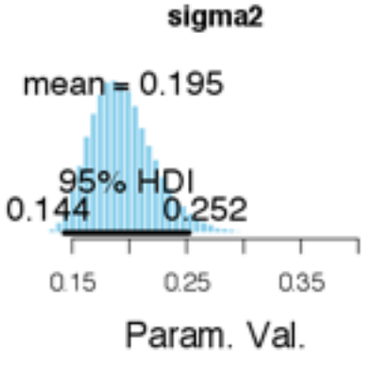










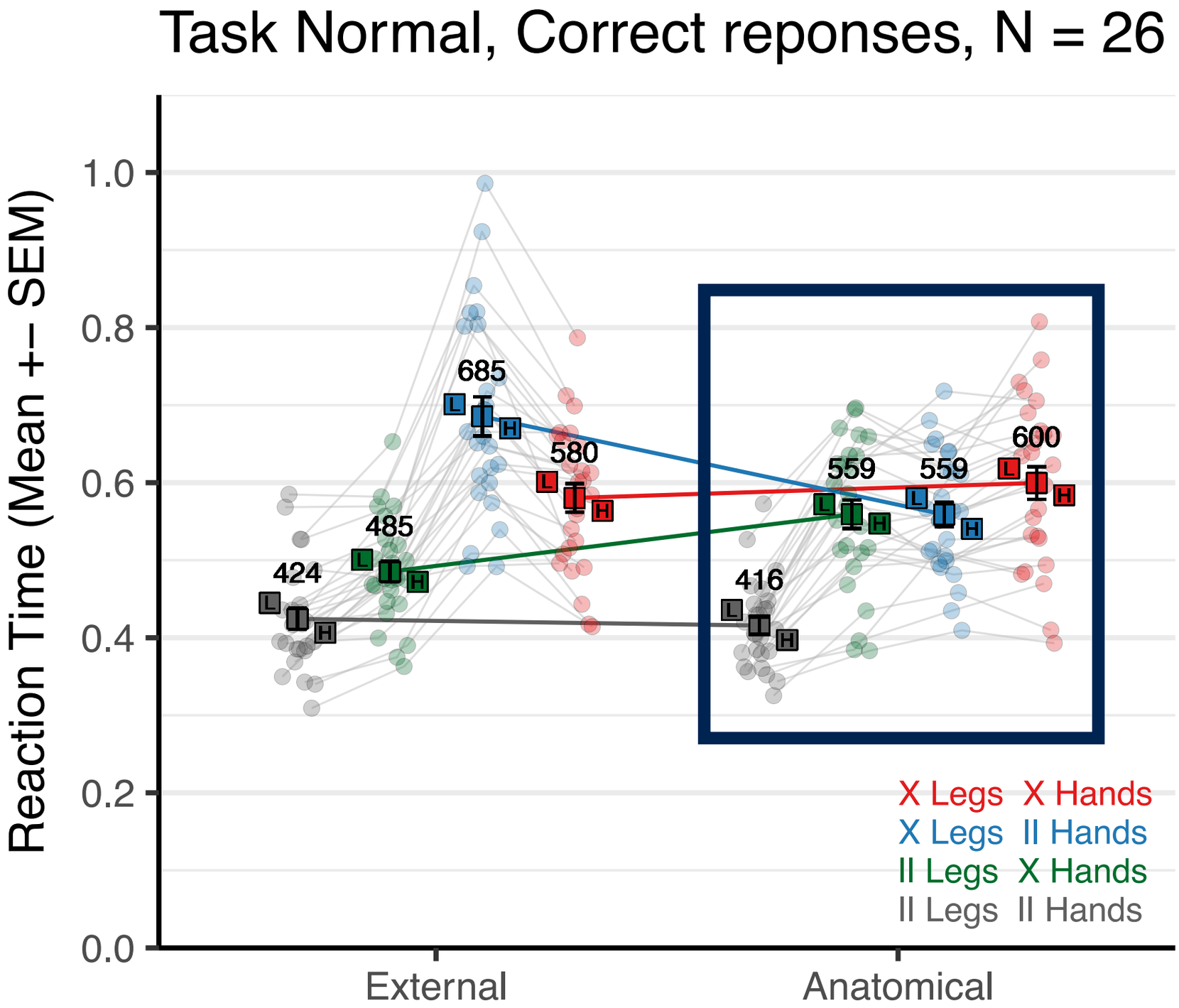


Exercise 3

Multiple groups continuous variable
hierarchical linear model

(ANOVA like, establishing differences: ROPE, shrinkage)

data: subjects means anatomical



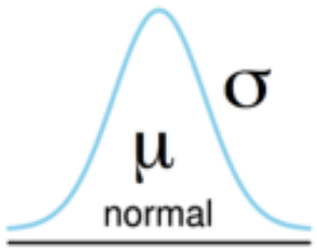
Comparison of a continuous variable μ and σ

Continuous normal variable posterior for μ and σ is given by:

Likelihood **bivariate** prior

$$p(\mu, \sigma|y) = \frac{p(y|\mu, \sigma)p(\mu, \sigma)}{\int\int p(y|\mu, \sigma)p(\mu, \sigma)\ d\mu\ d\sigma}$$

We can use the the t distribution pdf as the likelihood function:

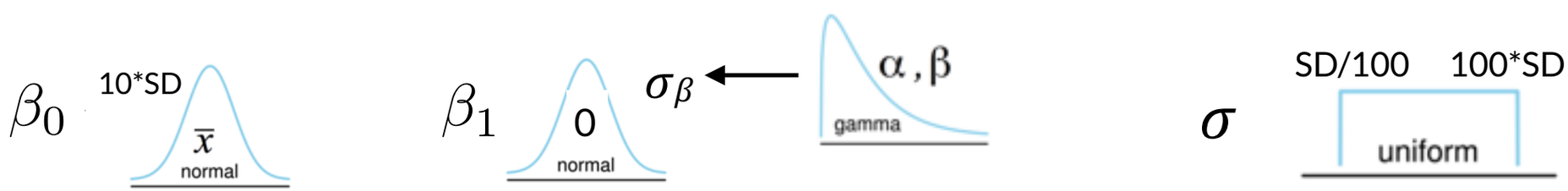


$$p(y|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

And we need a model

$$\mu_i = \beta_0 + \beta_{1(cond[i])} \qquad \sum_{j=1}^{nCond} \beta_j = 0$$

Priors are needed for β_0, β_{1j} and σ



ANOVA model

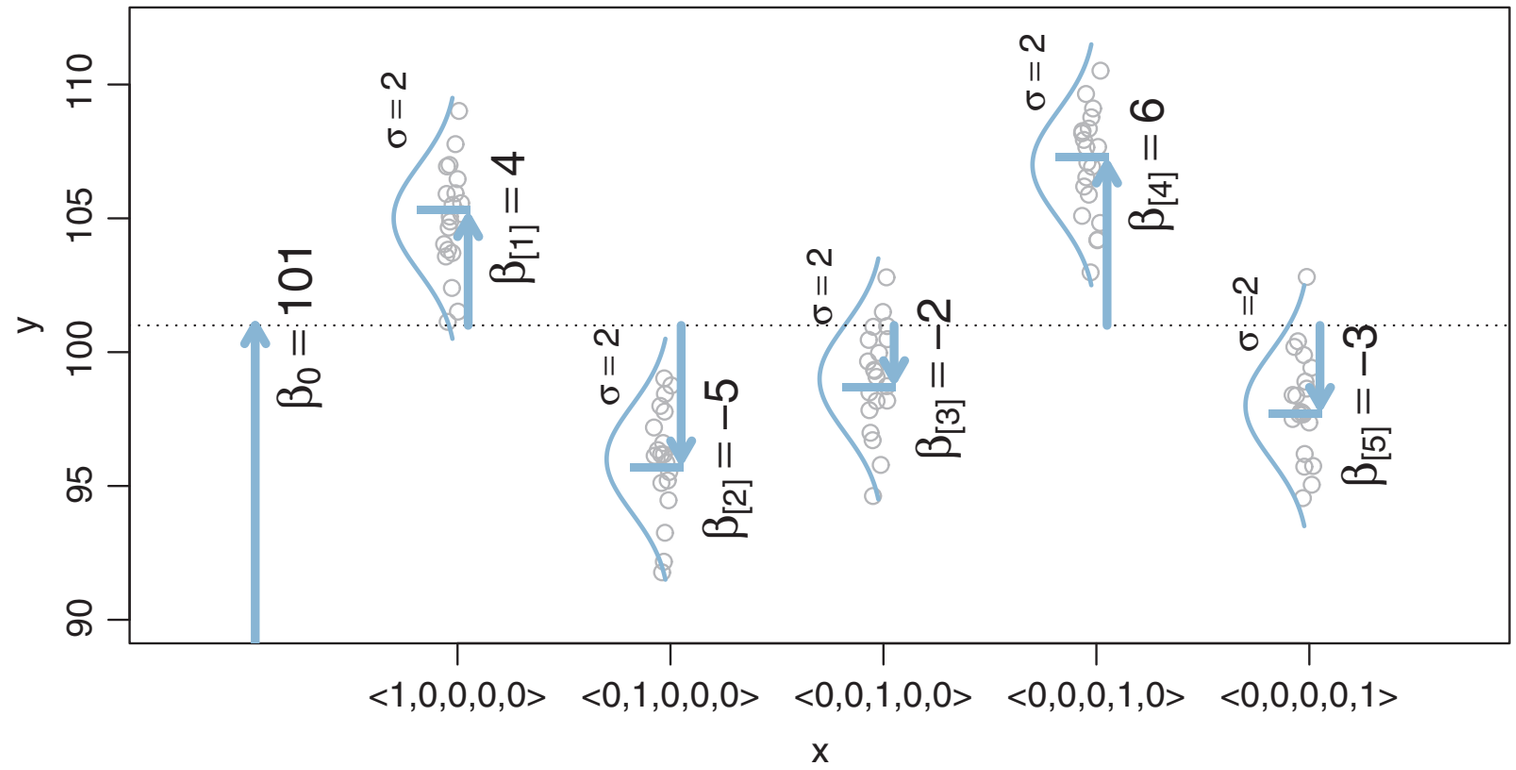
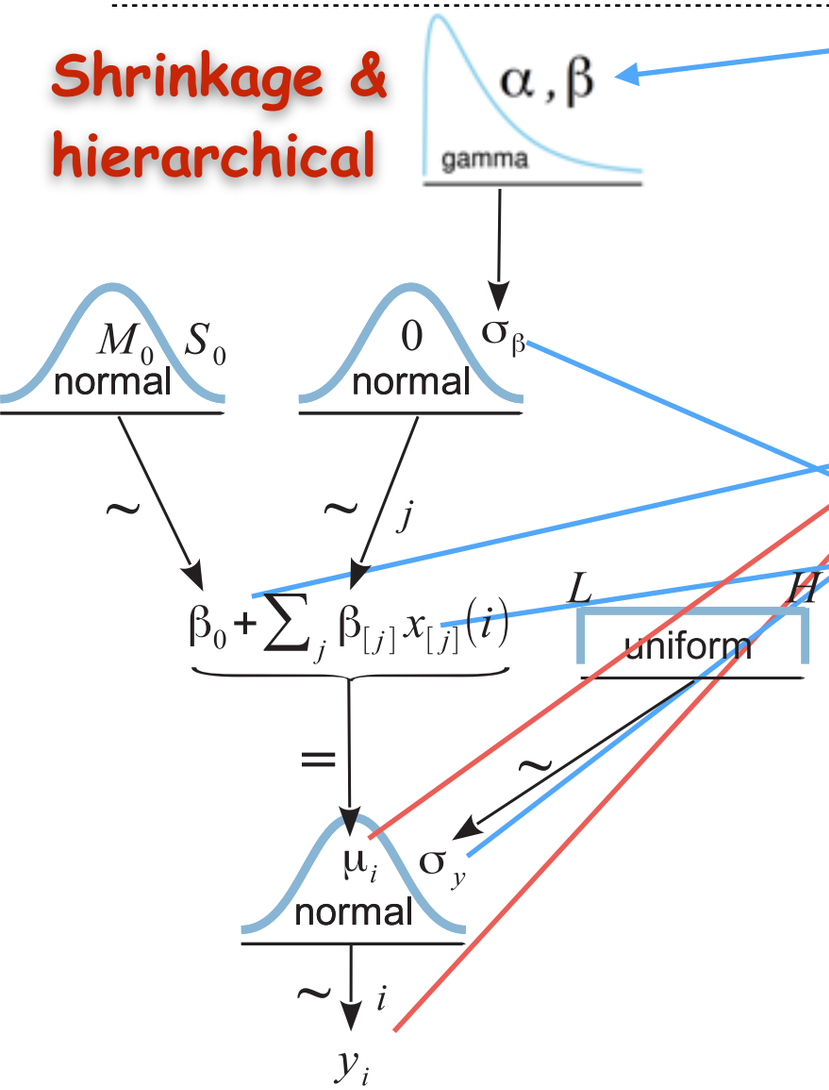


Figure 19.1 Description of data as normally distributed around group means that are conceptualized as deflections from an overall baseline. Data are indicated by circular dots (jittered left-right for visibility). The standard deviation of the data within groups is assumed to be the same for all groups and is indicated as σ. Baseline and deflections are indicated by arrows and β values. Notice that the deflections from baseline sum to zero.

Comparison of a continuous variable μ and σ

```
dataList = list(y          = dFc$mRT,
                cond       = cond,
                nCond      = length(unique(cond)),
                nTotal     = length(dFc$mRT),
                MPriormean = mean(dFc$mRT),
                MPriorstd  = sd(dFc$mRT)*10,
                SPriorL    = sd(dFc$mRT)/100,
                SPriorH    = sd(dFc$mRT)*100,
                aGammaShRa = gammaShRaFromModeSD(mode = sd(dFc$mRT)/2,
                                                 sd = sd(dFc$mRT)*2))
```

Shrinkage & hierarchical




```
modelString = " model {
  for ( i in 1:nTotal ) {
    y[i]  ~ dnorm( mu[i] , 1/sigma^2)
    mu[i] <- a0 + a1[cond[i]]}
  sigma ~ dunif(SPriorL, SPriorH )
  a0 ~ dnorm(MPriormean,1/MPriorstd^2)
  for ( j in 1:nCond ) {
     a1[j] ~ dnorm(0, 1/(aSigma)^2)}
  aSigma ~ dgamma( aGammaShRa[1] , aGammaShRa[2] )

 # Convert a0,a[] to sum-to-zero b0,b[] :
  for ( j in 1:nCond ) { m[j] <- a0 + a1[j] }
  b0 <- mean( m[1:nCond] )
  for ( j in 1:nCond ) { b1[j] <- m[j] - b0 }
}"
```

What is a ‘credible’ parameter difference?

ROPE: region of practical equivalence, range of equivalence

“A *region of practical equivalence* (ROPE) indicates a small range of parameter values that are considered to be practically equivalent to the null value for purposes of the particular application.”

“A parameter value is declared to be not credible, or rejected, if its entire ROPE lies outside the 95% highest density interval (HDI) of the posterior distribution of that parameter.”

“A parameter value is declared to be accepted for practical purposes if that value’s ROPE completely contains the 95% HDI of the posterior of that parameter. “

HDI: high density intervals

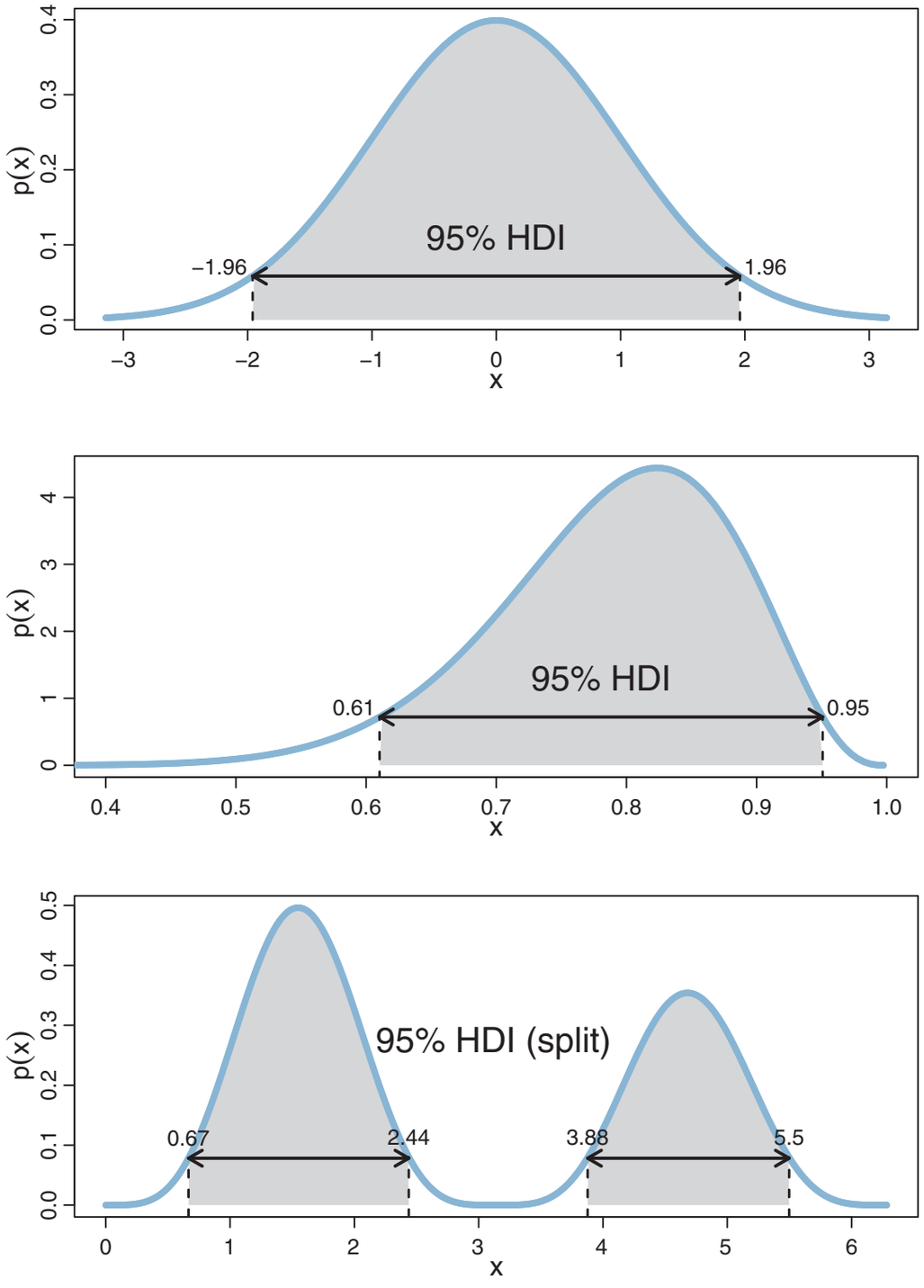


Figure 4.5 Examples of 95% highest density intervals (HDIs). For each example, all the *x* values inside the interval have higher density than any *x* value outside the interval, and the total mass of the points inside the interval is 95%. The 95% area is shaded, and it includes the zone below the horizontal arrow. The horizontal arrow indicates the width of the 95% HDI, with its ends annotated by (rounded) *x* values. The height of the horizontal arrow marks the minimal density exceeded by all *x* values inside the 95% HDI.

R exercise: credible differences, shrinkage

1. Read and run continuous3.R.

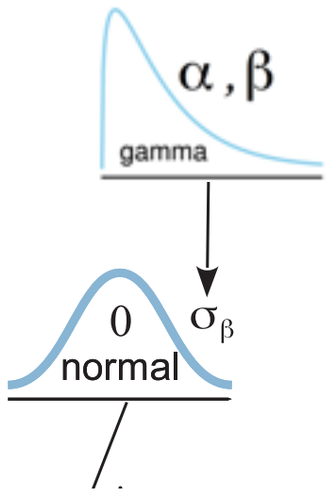
2. Check model convergence

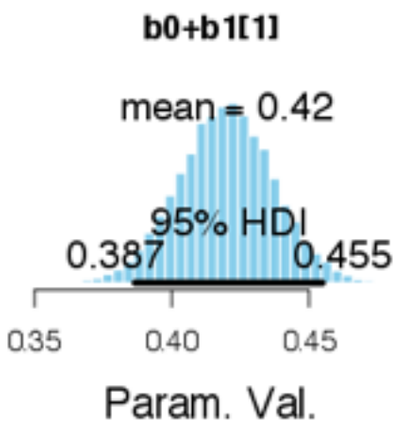
3. Plot β_1 parameters, their differences, and effect size of the diffrence. Define a ROPE of +-10ms to define what group differences are credible different and in which one we define that there is no difference.

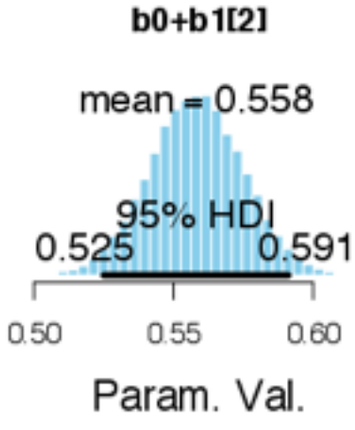
4. Change the aSigma prior to a constant to eliminate shrinkage, compare the results with the model with shrinkage.

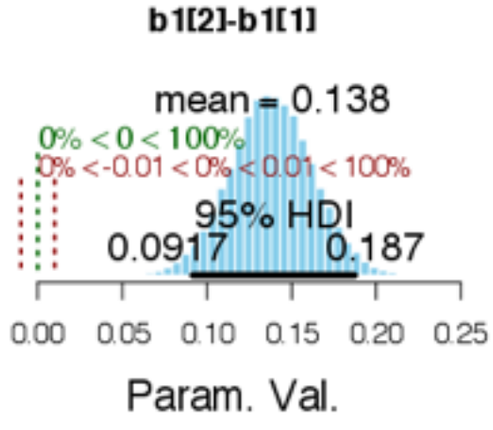
R exercise: credible differences

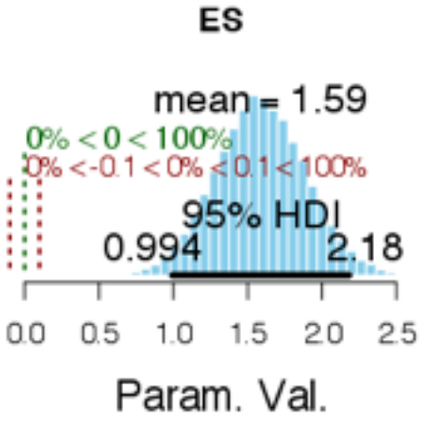
With shrinkage

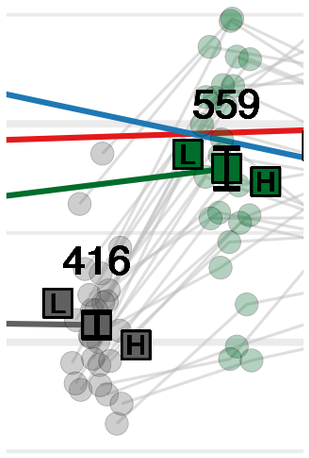




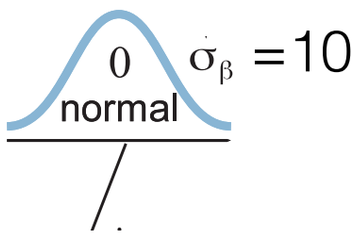


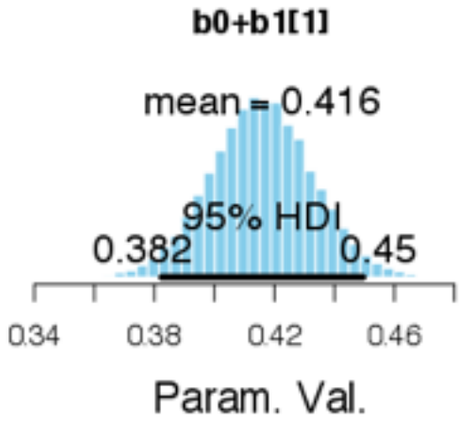


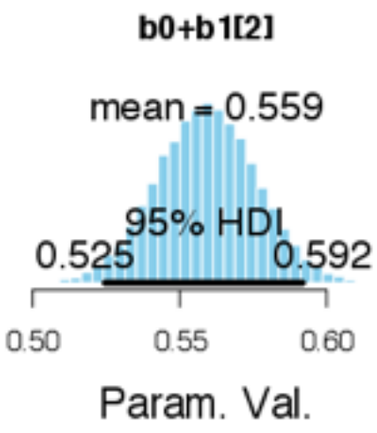


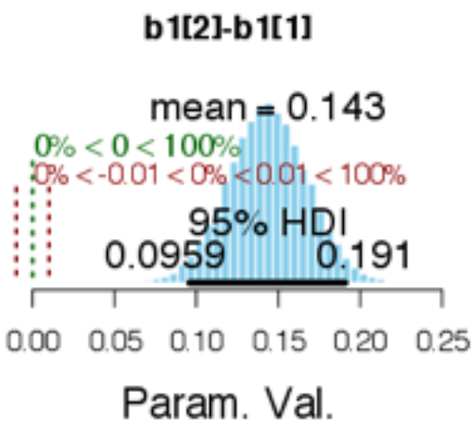


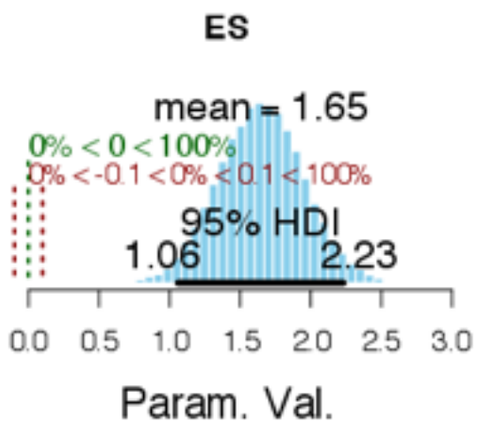
Without shrinkage:









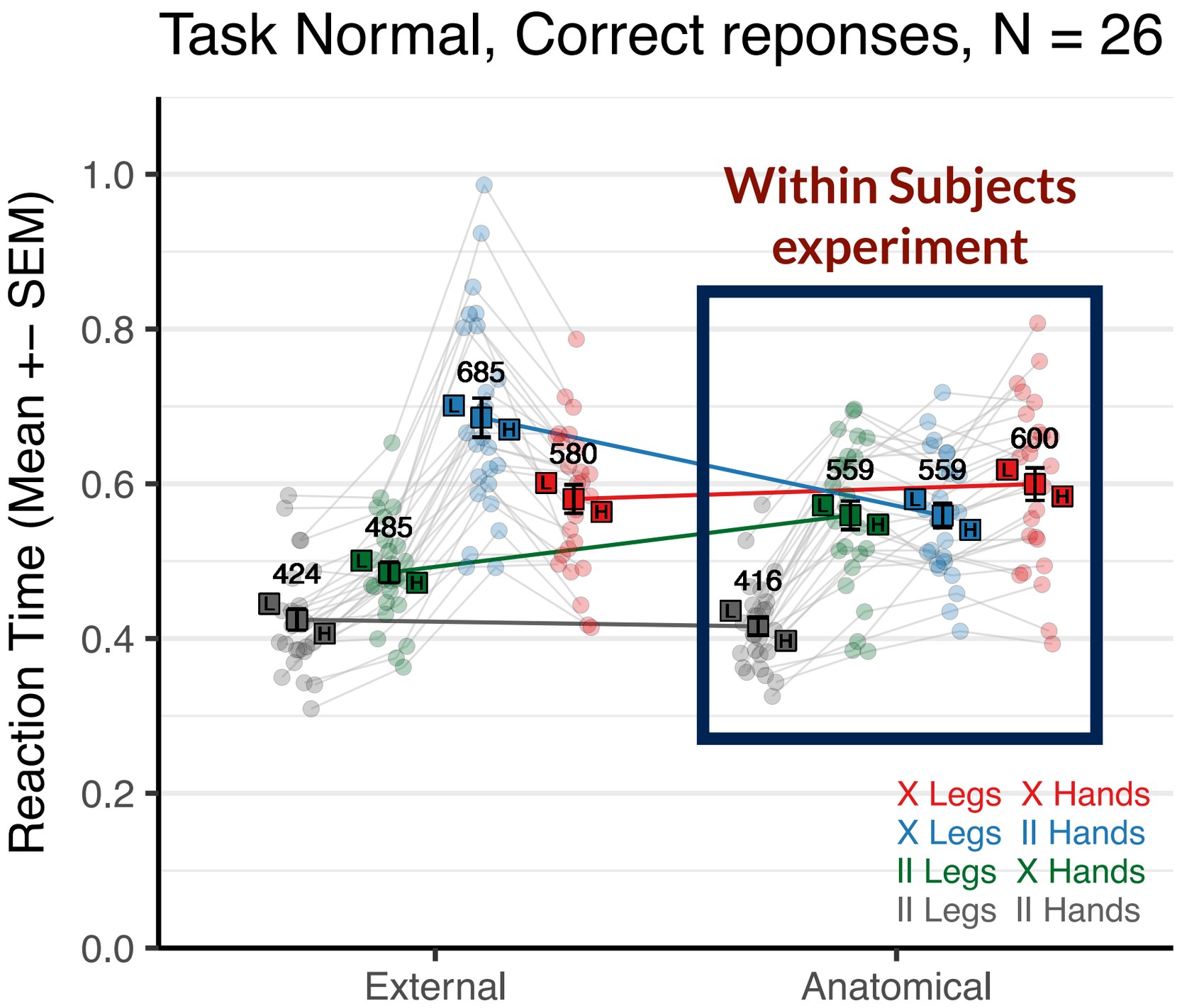


Exercise 4

Hierarchical linear model continuous variable within subjects

(subject factor, model comparison)

data: subjects means anatomical



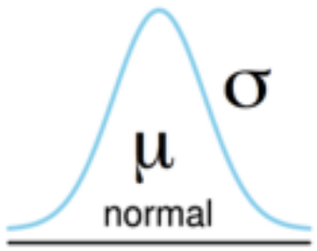
Comparison of a continuous variable μ and σ

Continuous normal variable posterior for μ and σ is given by:

Likelihood **bivariate** prior

$$p(\mu, \sigma|y) = \frac{p(y|\mu, \sigma)p(\mu, \sigma)}{\int\int p(y|\mu, \sigma)p(\mu, \sigma)\ d\mu\ d\sigma}$$

We can use the the t distribution pdf as the likelihood function:

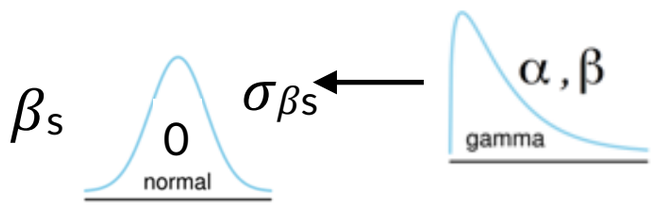


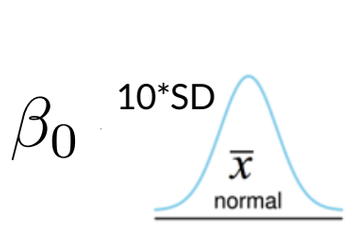
$$p(y|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

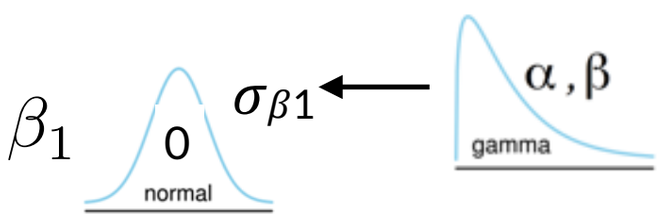
And we need a model

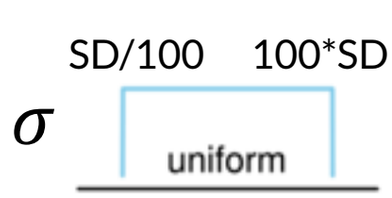
$$\mu_i = \beta_0 + \beta_{1(cond[i])} + \beta_{s(subj[i])} \qquad \sum_{j=1}^{nCond} \beta_j = 0$$

Priors are needed for β_0, β_{1j}, β_{sk} and σ









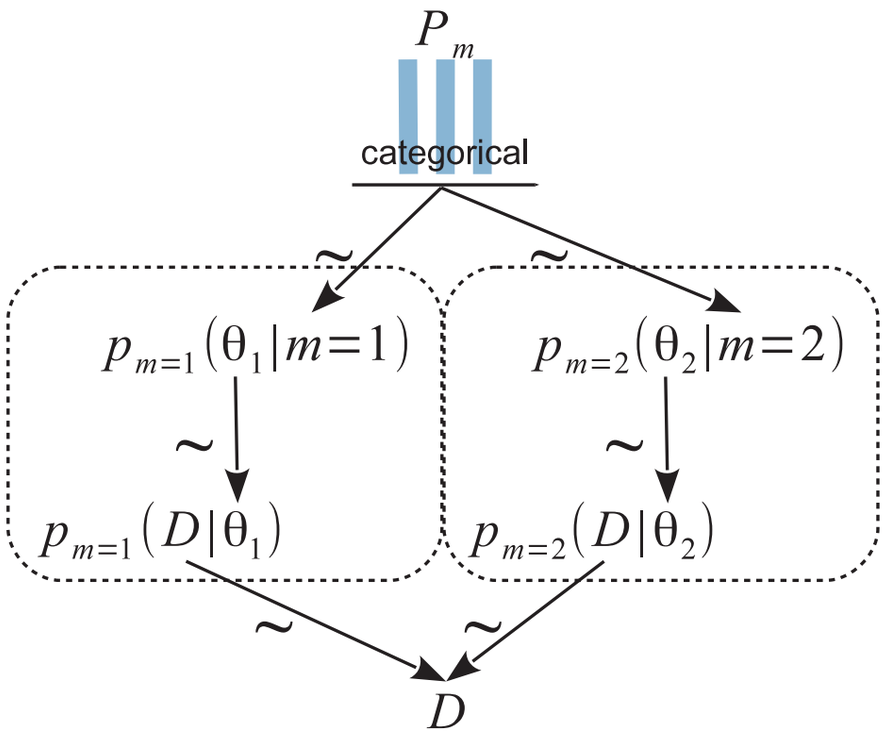
Comparison of a continuous variable μ and σ

```
dataList = list(y          = dFc$mRT,
                cond       = cond,
                subj       = subj,
                nCond      = length(unique(cond)),
                nTotal     = length(dFc$mRT),
                nSubj      = length(unique(subj)),
                MPriormean = mean(dFc$mRT),
                MPriorstd  = sd(dFc$mRT)*10,
                SPriorL    = sd(dFc$mRT)/100,
                SPriorH    = sd(dFc$mRT)*100,
                aGammaShRa = gammaShRaFromModeSD(mode = sd(dFc$mRT)/2,
                             sd = sd(dFc$mRT)*2))
```

```
modelString = "model {for ( i in 1:nTotal ) {
         y[i]  ~ dnorm( mu[i] , 1/sigma^2)
         mu[i] <- a0 + a1[cond[i]] +aS[subj[i]]}
         sigma ~ dunif(SPriorL, SPriorH )
         a0 ~ dnorm(MPriormean,1/MPriorstd^2)
        for ( j in 1:nCond ) {a1[j] ~ dnorm(0, 1/(aSigma)^2)}
         aSigma ~ dgamma( aGammaShRa[1], aGammaShRa[2])
        for (j in 1:nSubj) {aS[j] ~ dnorm(0.0, 1/(sSigma)^2 )}
         sSigma ~ dgamma(aGammaShRa[1], aGammaShRa[2])

 # Convert a0,a[] to sum-to-zero b0,b[] :
  for ( j in 1:nCond ) { for (s in 1:nSubj){
            m[j,s] <- a0 + a1[j] + aS[s] }} # cell means
  b0 <- mean( m[1:nCond,1:nSubj] )
  for ( jS in 1:nSubj ) { bS[jS] <- mean( m[1:nCond, jS] ) - b0}
  for ( jC in 1:nCond ) { b1[jC] <- mean( m[jC,1:nSubj] ) - b0}}"
```

model comparison



prior odds

$$\underbrace{\frac{p(m=1|D)}{p(m=2|D)}}_{\text{posterior odds}} = \underbrace{\frac{p(D|m=1)}{p(D|m=2)}}_{\text{BF}} \frac{p(m=1)}{p(m=2)} \underbrace{\frac{/\sum_m p(D|m)\,p(m)}{/\sum_m p(D|m)\,p(m)}}_{=1}$$

When p(m=1)=p(m=2)=0.5 posterior odds are equal to Bayes Factor. Otherwise multiply the posterior odd with p(m=2)/p(m=1)

```
modelString = "
          model {
              for ( i in 1:nTotal ) {
                y[i]  ~ dnorm( mu[i] , 1/sigma^2)
               mucond[i] <- a0 + a1[cond[i]] + aS[subj[i]]
               mucond2[i] <- a0 + a1[cond2[i]]  + aS[subj[i]]
               mu[i]  <- equals(mC,1)*mucond[i] + equals(mC,2)*mucond2[i]
                  }
               mC ~ dcat( mPriorProb[] )
               mPriorProb[1] <- .5
               mPriorProb[2] <- .5
               …}
```

2 likelihoods formulas

Selection line

2 categories same p

the rest of the variables will be a mixture of variables under model 1 and 2

R exercise: within subjects model, model comparison

1. Read and run continuous4.R.

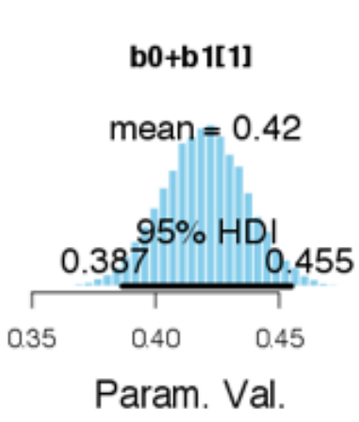
2. Check model convergence

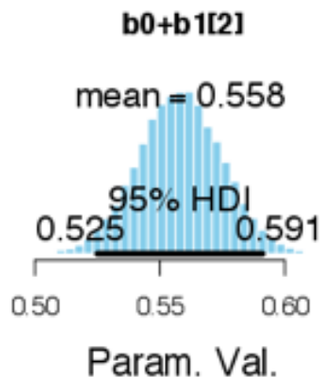
3. Compare β_1 parameters and their differences in the within subject models to the ones from exercise 3.

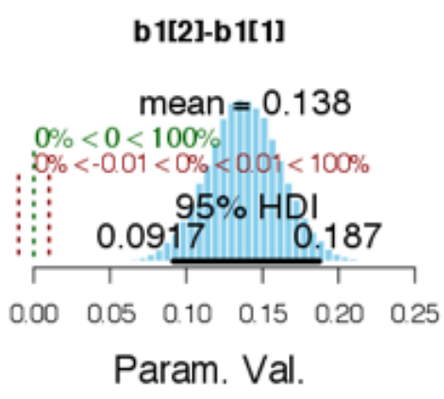
4. Uncomment the model comparison lines. Calculate posterior odds/ bases factor for the comparison between a model in which the conditions are divided in four levels to one in which there is only two levels (1st one against the other three)

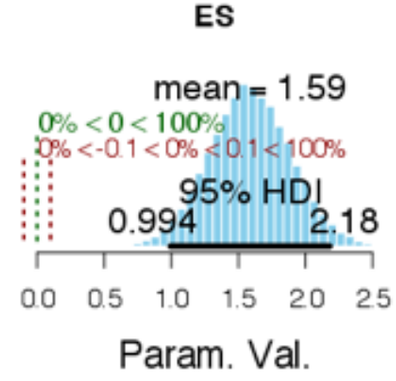
R exercise: within subjects vs between subjects

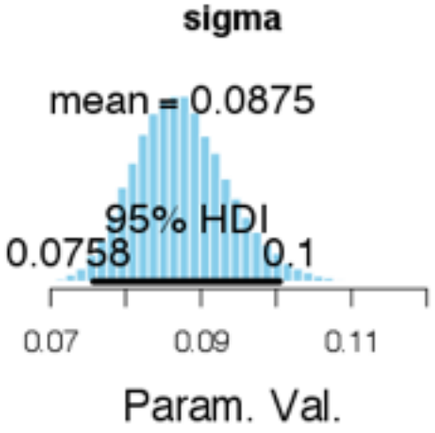
Model exercise 3
(with shrinkage)



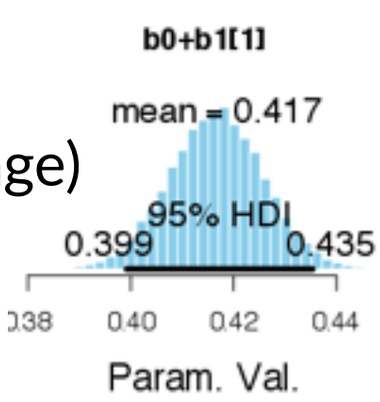


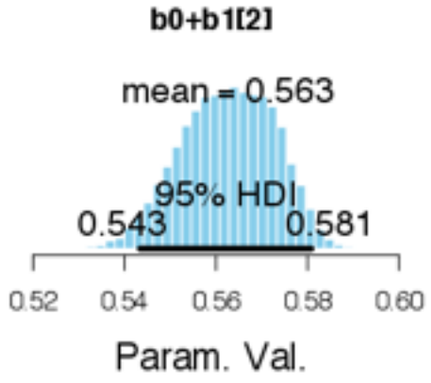


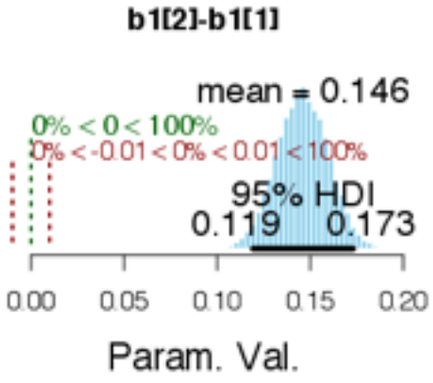


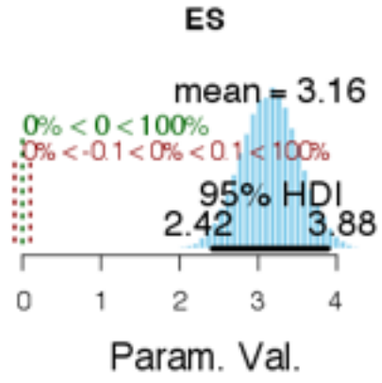


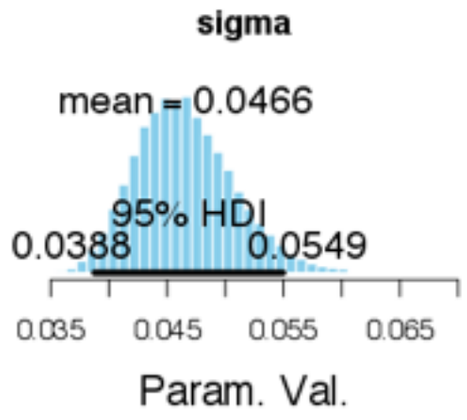
Model exercise 4
within (with shrinkage)



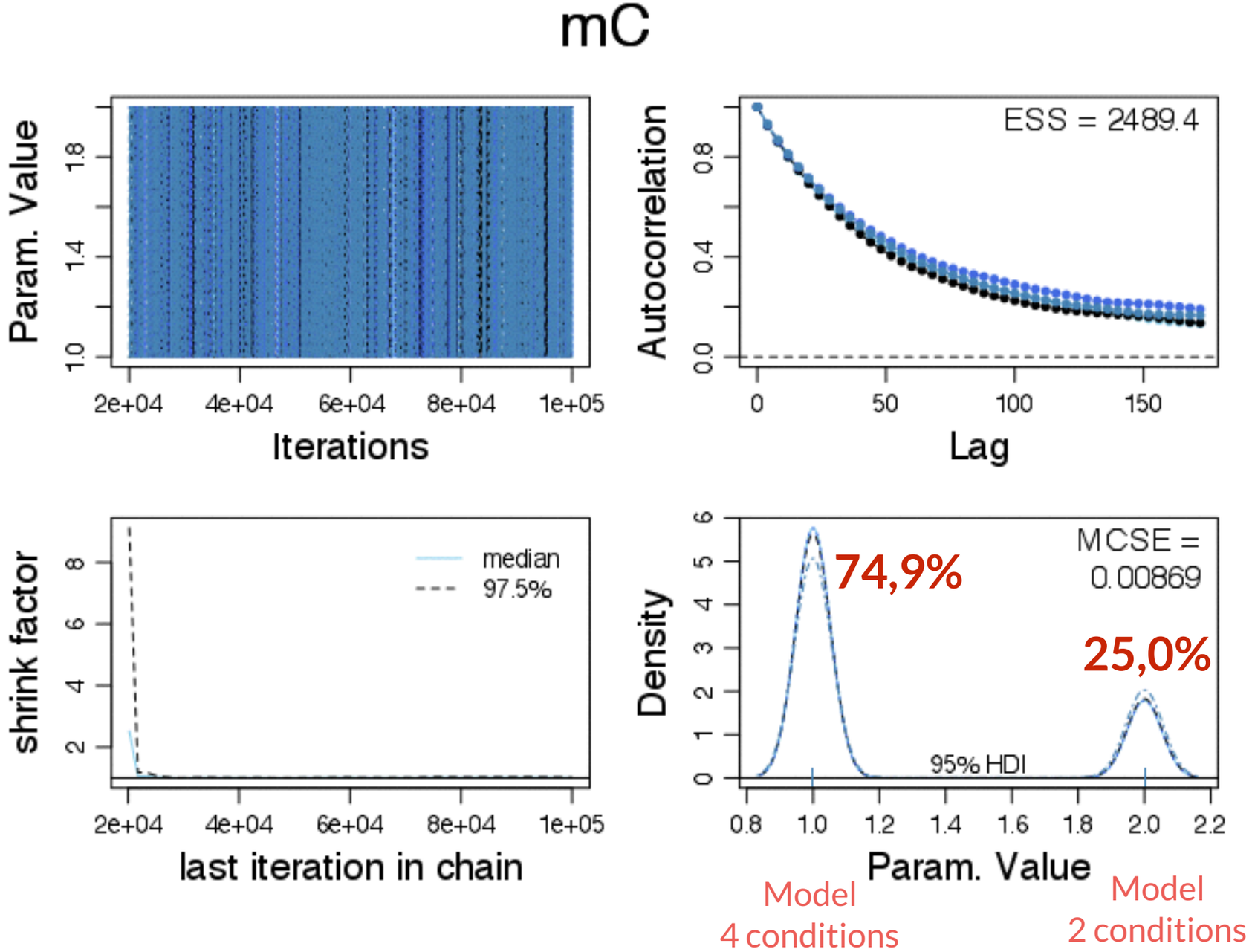








R exercise: model comparison



$$\frac{p(m=1|D)}{p(m=2|D)} = \frac{0.749}{0.2509} = 2.984 = BF$$