Use of historical data

Methods, applications and implementation with the R package RBesT

Heinz Schmidli, Sebastian Weber

Novartis, Statistical Methodology and Consulting

Workshop of IBS-GR WG Bayes Methods

Göttingen, Germany, 6th Dec 2018

Introduction

Prior derivation with gMAP

Using MAP priors for clinical trials

Probability of Success

Useful Resources

Excersises

Backup

Introduction

Disclaimer

The views and opinions expressed in this presentation and on the slides are solely those of the presenters and are not necessarily those of Novartis. Novartis does not guarantee the accuracy or reliability of the information provided herein.

Executive Summary

RBesT: R Bayesian evidence synthesis tools

- Facilitates the application of the Meta-Analytic-Predictive (MAP) approach in clinical trials
- RBesT is designed as a modern R library
  - Fully **documented with examples**
  - Standard R formula syntax supported
  - Fully **unit-tested** software
  - **Fast and accurate analytical** computations
- Supports **binary, normal (known σ) and Poisson** endpoints
- High abstraction level makes (complex) computations straightforward and user-friendly (trial statistician-friendly)

`RBesT` facilitates applying the MAP Approach

RBesT supports in *using* historical data for clinical trials

1. Assess historical data for relevance
   - **exchangeability assumption justifiable?**
   - between-trial heterogeniety τ?
2. Run MAP analysis to obtain
   informative prior in *parametric form*
   - Analyse historical data using MCMC (Stan)
     `gMAP`
   - Approximate MCMC MAP prior with parametric density
     `mixfit` or `automixfit`
   - Consider robustification
     `robustify`
3. Evaluate frequentist design properties
   `oc1S` or `oc2S`
4. Run final analysis
   `postmix`

Prior derivation with gMAP

Ankylosing Spondylitis, *The Lancet*, 2013, (382)
Double-blinded POC to test *secukinumab* against *placebo*

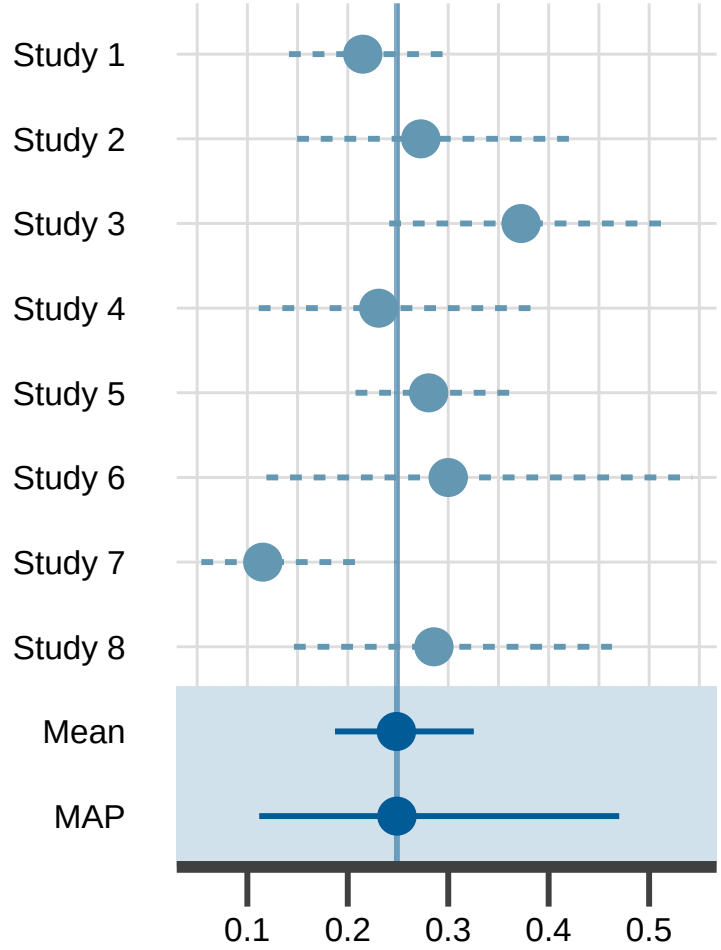
- Endpoint is binary ASAS20 at week 6 (higher response rate is better)
- Bayesian design
  - $P(\pi_p - \pi_t \leq 0 | y) > 0.95$
  - Placebo: MAP derived,

$$\pi_p \sim \text{Beta}(11, 32)$$

  - Active:

$$\pi_t \sim \text{Beta}(0.5, 1)$$

- 4:1 randomization (24 vs 6)



NOVARTIS

Generalized Meta-Analytic-Predictive model

Hierarchical model to obtain predictive of mean parameter

Y is the (control) group summary data for H historical trials

$$Y_h|\theta_h \sim f(\theta_h) \qquad \forall\ h \in [1, H]$$
$$Y_*|\theta_* \sim f(\theta_*) \qquad \text{for new trial}$$

Exchangeability assumption:

$$g(\theta_h)|\beta, \tau \sim \text{Normal}(\beta, \tau^2) \qquad \forall\ h \in [1, H]$$
$$g(\theta_*)|\beta, \tau \sim \text{Normal}(\beta, \tau^2) \qquad \text{for new trial}$$

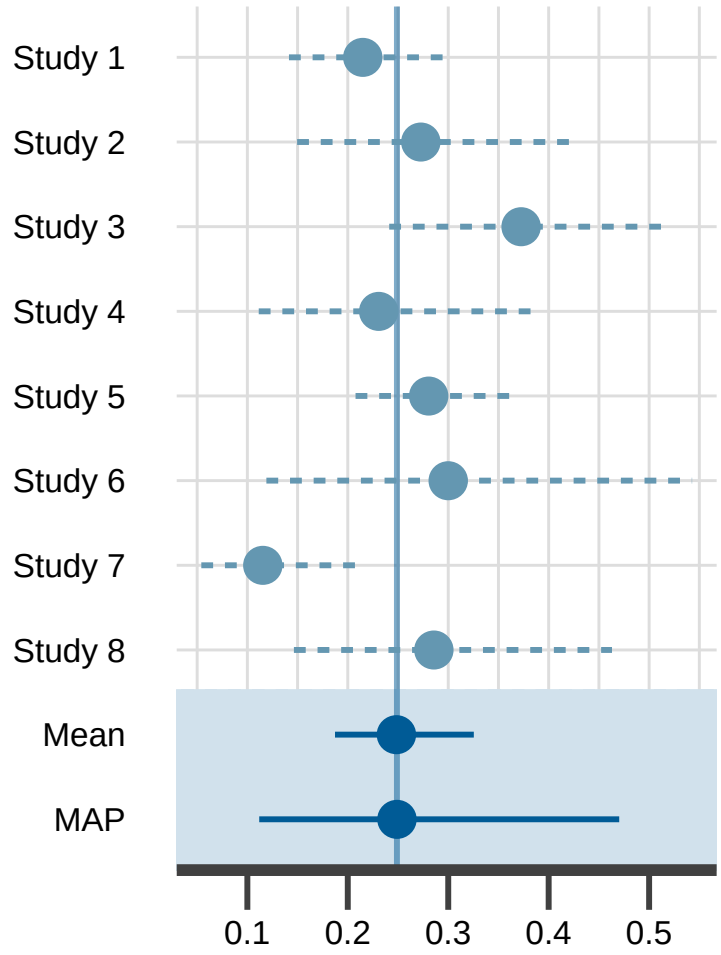
- f likelihood and g link function
  Binomial/logit, Normal (fixed σ)/identity or Poisson/log
- β population mean with prior Normal(m_β, s_β^2)
- τ between-trial heterogeniety with prior P_τ
  $\tau \to 0 \Rightarrow$ pooling / unbound use of historical data
  $\tau \to \infty \Rightarrow$ stratification / no use of historical data

Meta-Analytic-Predictive approach in words

A MAP prior is the predictive for the mean of a future trial

- A MAP analysis is a standard meta-analysis **plus** a prediction
- $p(\beta|y)$ is the population mean or the *typical trial result*
- $p(\theta_*|y)$ is the MAP or the **predictive distribution** for the mean of a *future trial*
- between-trial heterogeniety τ critically governs borrowing:
  $\tau \to 0 \Rightarrow$ pooling
  $\tau \to \infty \Rightarrow$ stratification

Ankylosing Spondylitis Example



NOVARTIS

Conservative prior choices for τ and β

Binary and normal endpoints

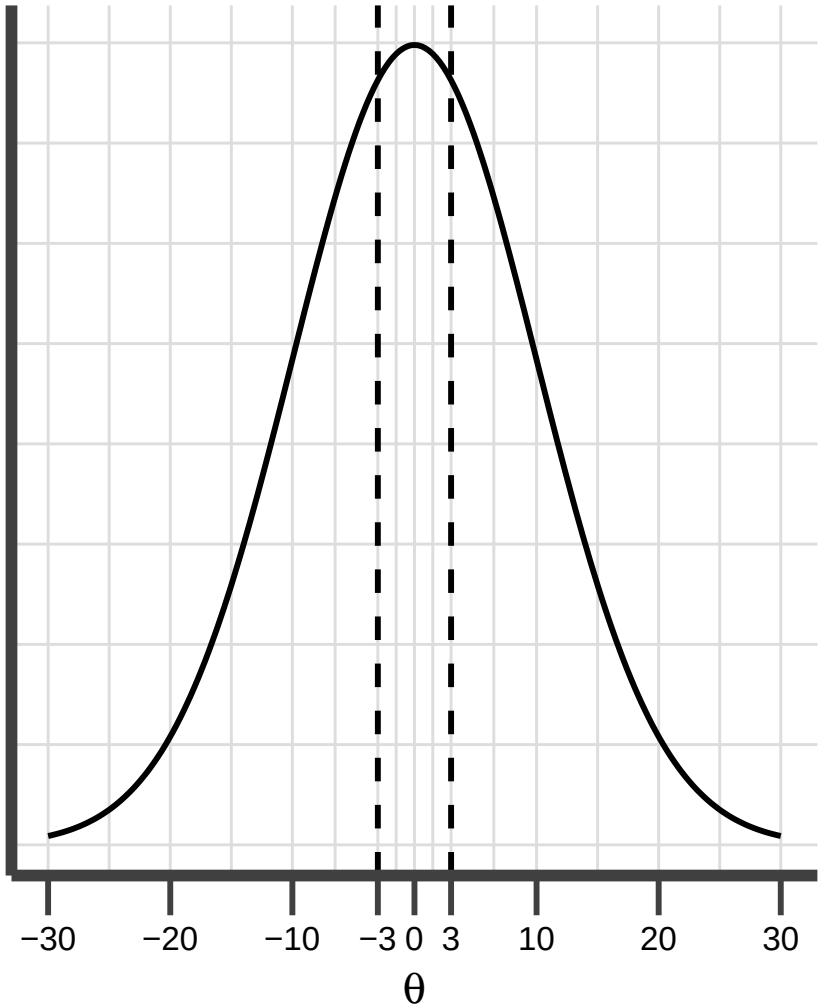
Endpoint		very conservative[1] τ prior	conservative[1,2] τ prior	β prior[3]
Binary	$0.2 < \pi < 0.8$	$N^+(0, 1)$	$N^+(0, (1/2)^2)$	$N(0, 2^2)$
Normal	known σ	$N^+(0, (\sigma/2)^2)$	$N^+(0, (\sigma/4)^2)$	$N(\mu_0, \sigma^2)$

1. very conservative, see *Neuenschwander et al., 2010*
2. less heterogeneous data as often seen empirically in meta-analysis, see *Friede et al., 2016*
3. unit-information prior for β (single observation of no effect), see *Kass & Wasserman, 1995*
   μ_0 set problem dependent (often 0)

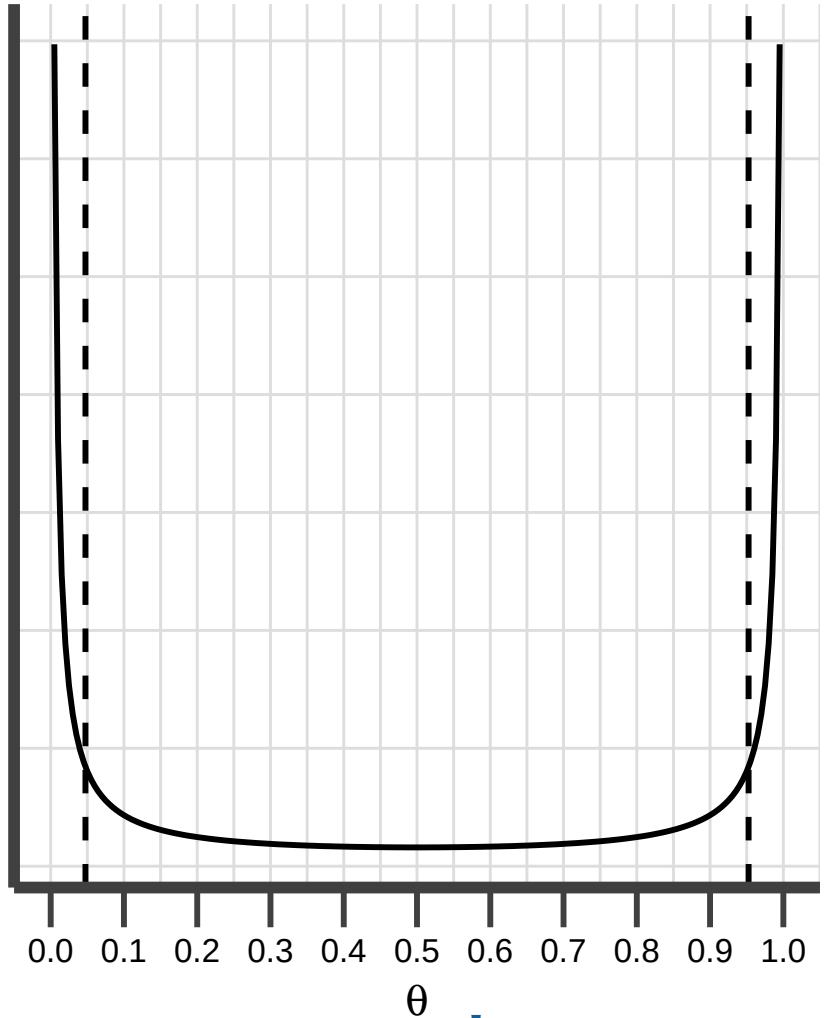
Detour: Non-informative priors on logit scale?

Consider $\beta \sim N(0, 10^2)$ on the logit scale:

N(0,10^2) on a logit scale...



... transformed to 0−1 scale

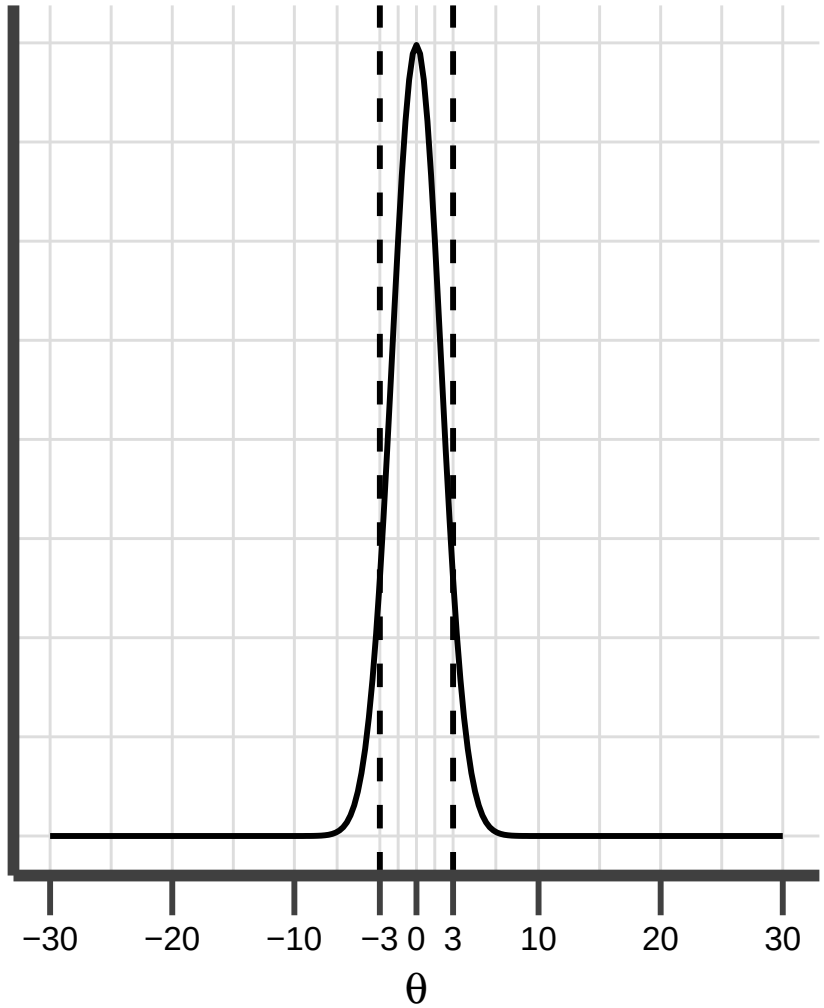


NOVARTIS

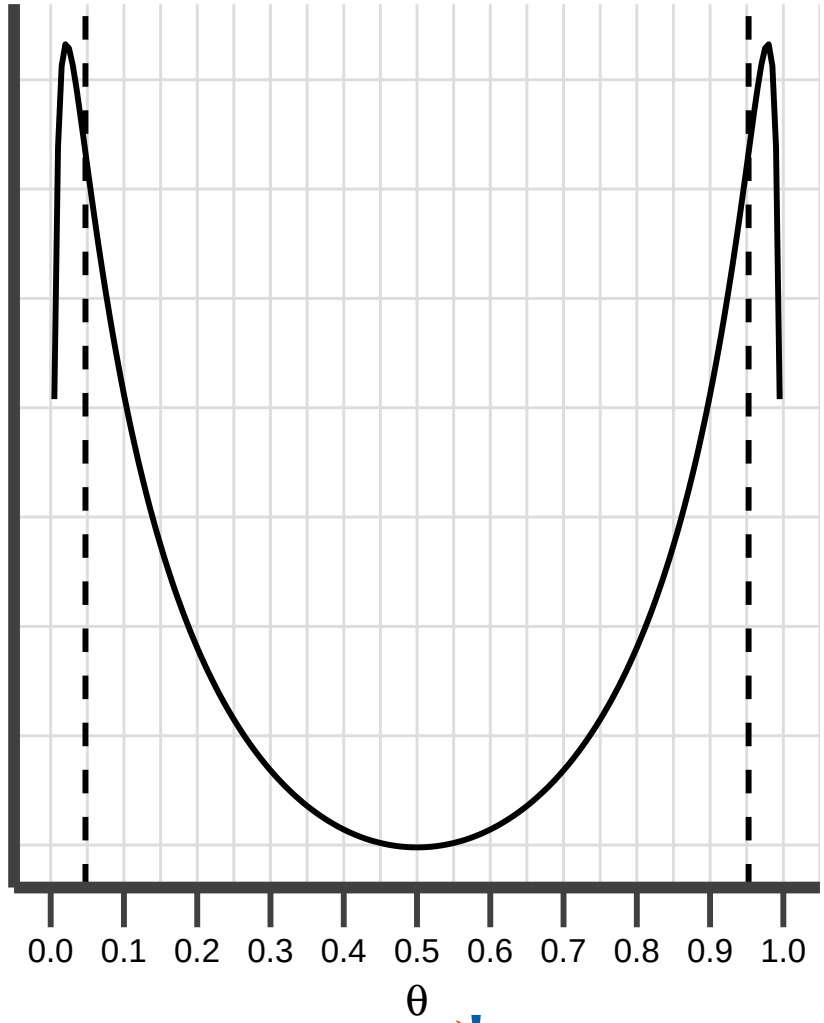
Detour: Non-informative priors on logit scale?

Consider now a $\beta \sim N(0, 2^2)$ on the logit scale:

N(0,2^2) on a logit scale...



... transformed to 0−1 scale



NOVARTIS

Running the MAP analysis with gMAP
Let's apply it to the AS data-set:

```
set.seed(123234)
map_mc <- gMAP(cbind(r, n-r) ~ 1 | study, data=AS, family=binomial,
               tau.dist="HalfNormal", tau.prior=1, beta.prior=2)
```

- `set.seed` ensures exact reproducibility
- model formula follows standard R conventions
  `cbind(# responders, # non-responders) ~ 1 | study`
- data-set `AS` (part of `RBesT`) passed in as `data.frame`
- `family` selects likelihood (and link function)
- τ prior must be set (very conservative here)
- β prior *should* be given (very conservative here)

gMAP results

```
print(map_mc)
```

```
## Generalized Meta Analytic Predictive Prior Analysis
##
## Call:  gMAP(formula = cbind(r, n - r) ~ 1 | study, family = binomial,
##     data = AS, tau.dist = "HalfNormal", tau.prior = 1, beta.prior = 2)
##
## Exchangeability tau strata: 1
## Prediction tau stratum    : 1
## Maximal Rhat              : 1
##
## Between-trial heterogeneity of tau prediction stratum
##   mean     sd   2.5%    50%  97.5%
## 0.3750 0.2130 0.0375 0.3480 0.8870
##
## MAP Prior MCMC sample
##   mean     sd   2.5%    50%  97.5%
## 0.2590 0.0877 0.1100 0.2490 0.4740
```

`RBesT` supports standard generic functions
Analyses result object have standard query functions

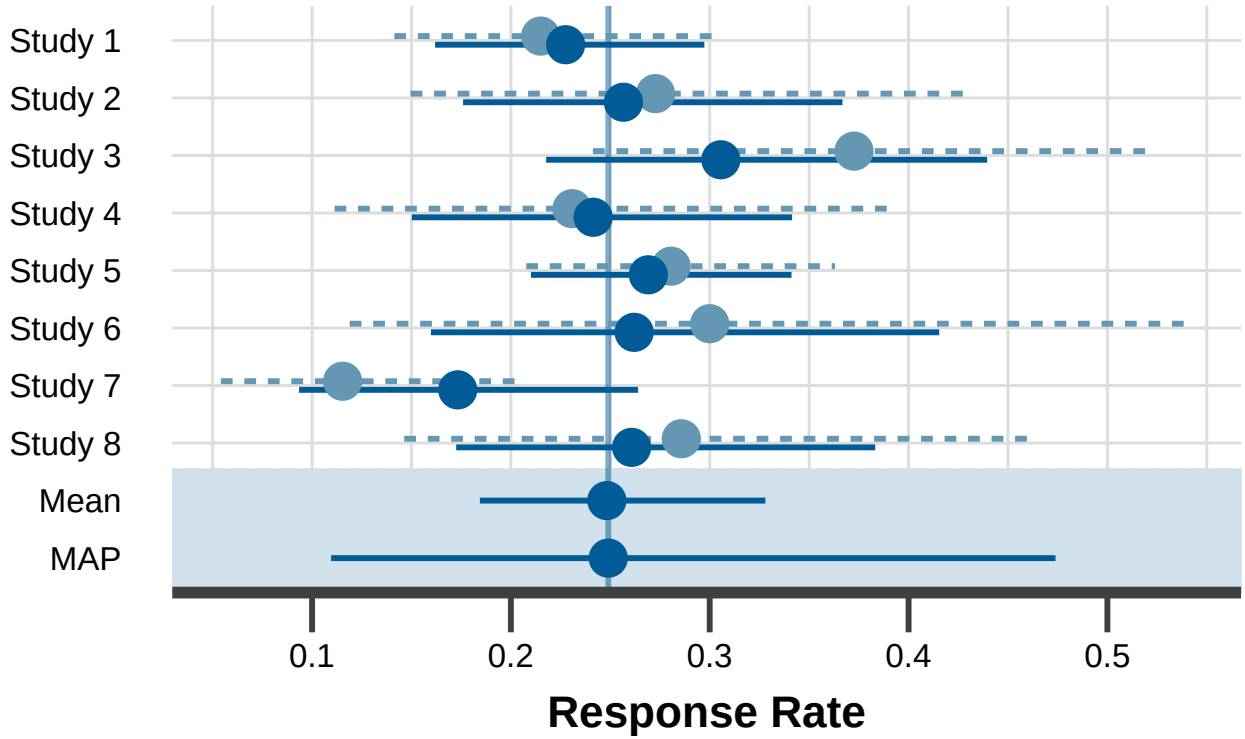
Generics defined for a `gMAP` analysis object:

Function	`gMAP` context
`print`	Key analysis printout
`summary`	**Model summary**
`fitted`	Fitted responses
`coef`	Fitted model parameters
`predict`	Obtain predictions (MAP prior with covariates)
`plot`	MCMC diagnostics, densities, **MAP model forest plot**
`as.matrix`	obtain MCMC sample (advanced)

Graphical model diagnostics

Standard forest plot with meta-analytic model estimates

```
plot(map_mc, size=0.5)$forest_model
```



- `forest_plot` function produces customizable forest plots
- please refer to “Customizing RBesT Plots” vignette for basic customization of `ggplot2` plots

Impact of prior choices

Example: Use less conservative prior for τ

```
map_mc_alt <- update(map_mc, tau.dist="HalfNormal", tau.prior=0.5)
rbind(summary(map_mc)$tau,
      summary(map_mc_alt)$tau)
```

```
##              mean        sd       2.5%       50%     97.5%
## tau[1] 0.3753758 0.2129691 0.03751009 0.3482959 0.8867479
## tau[1] 0.3354001 0.1778245 0.03811913 0.3151081 0.7574960
```

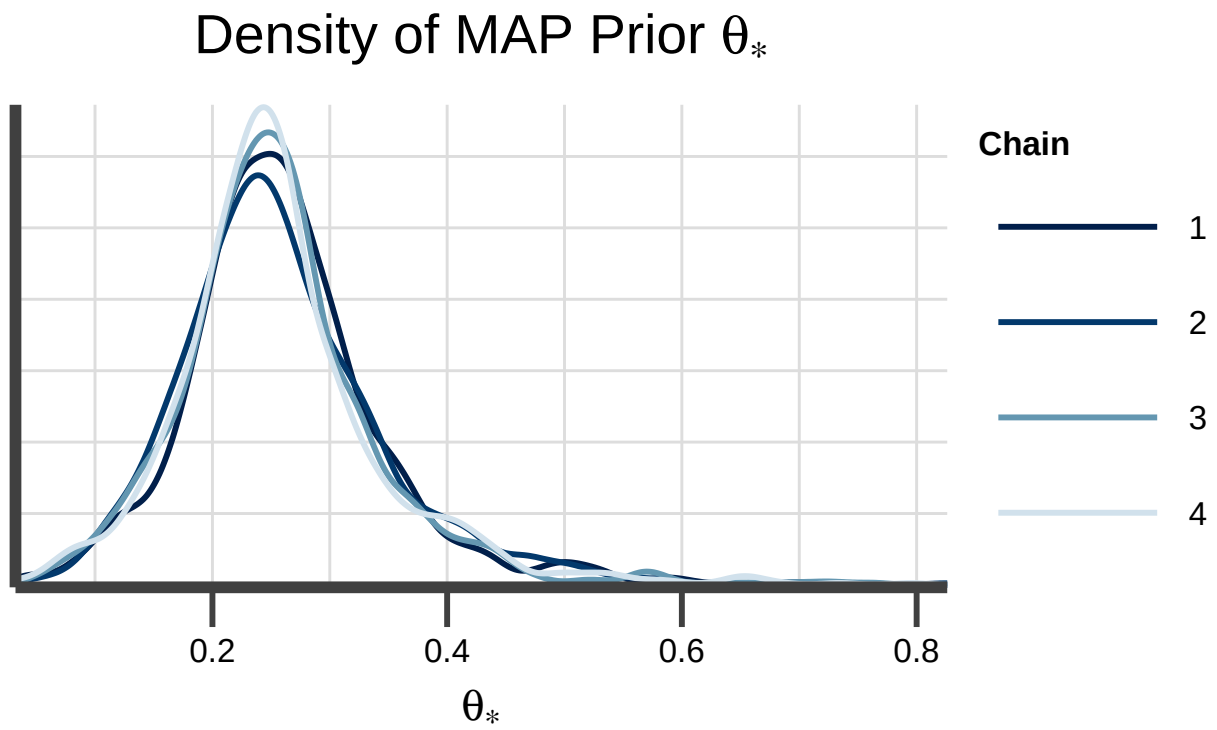
```
rbind(summary(map_mc)$theta.pred,
      summary(map_mc_alt)$theta.pred)
```

```
##                       mean         sd      2.5%       50%     97.5%
## theta_resp_pred 0.2585327 0.08768081 0.1095699 0.2489713 0.4739009
## theta_resp_pred 0.2571384 0.07744917 0.1248146 0.2492036 0.4468847
```

Using MAP priors for clinical trials

MAP analysis result is a MCMC sample

```
plot(map_mc)$densityThetaStar
```



A MCMC sample of 4×10^3 *draws is unconvenient to communicate. . .*

NOVARTIS

Turning MAP into a parametric density
Parametric densities have many practical advantages

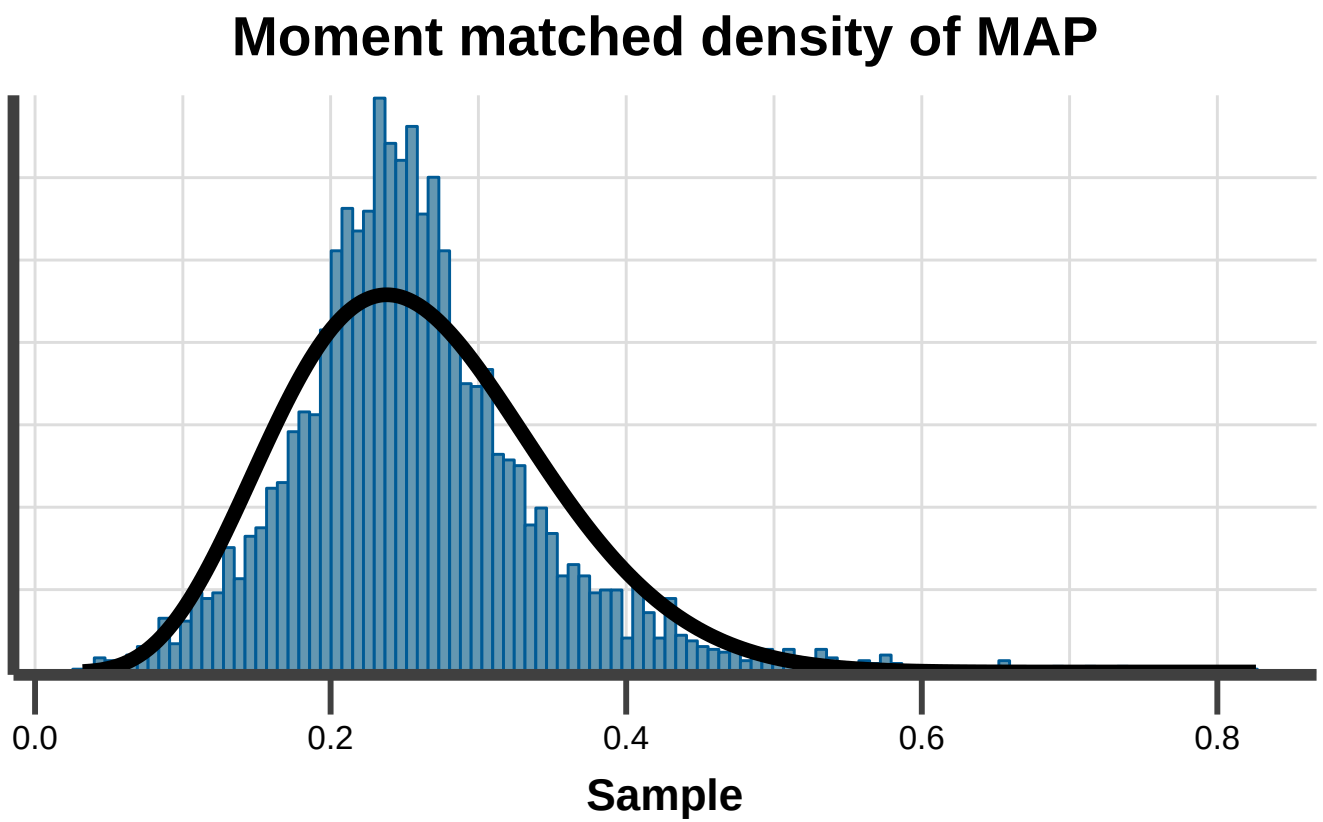
- **Conjugate** priors allow for **fast analytic** manipulations
  the posterior is then given by the same distributional class as the prior

Likelihood	Prior	Posterior
Binomial	Beta	Beta
Normal (known σ)	Normal	Normal
Poisson	Gamma	Gamma

- Simple moment matching often not accurate (heavy tails)
  $\Rightarrow$ mixtures are arbitrarily accurate and maintain conjugacy

Limitations of Moment Matching
Heavy tails of MAP priors lead to misfit

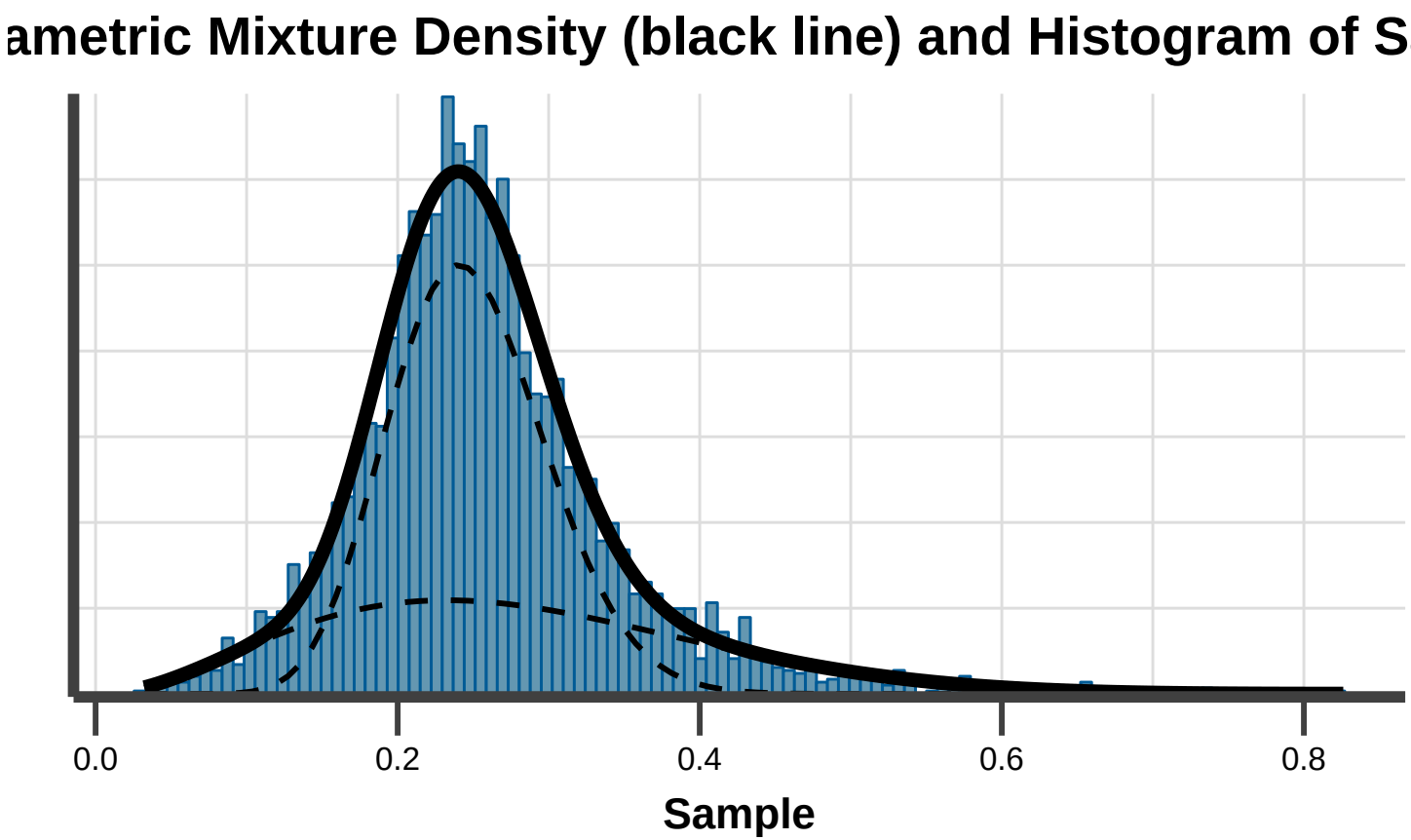
```
map_moment_match <- mixfit(map_mc, Nc = 1)
plot(map_moment_match)$mix + ggtitle("Moment matched density of MAP")
```



Mixtures improve accuracy of parametric MAP

Inference with EM of a 2-component mixture...

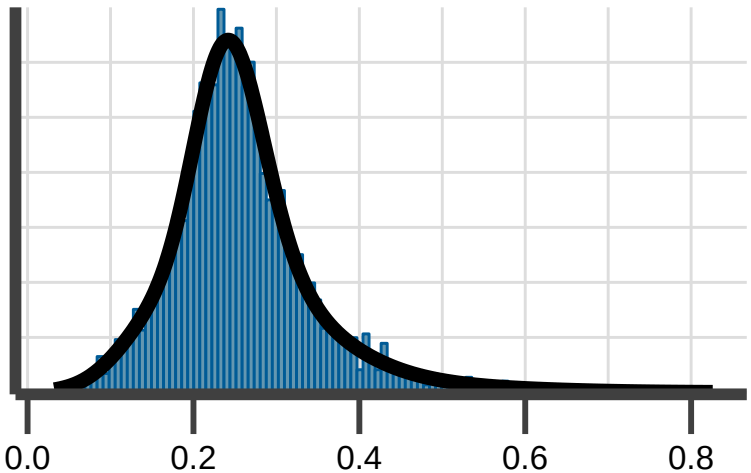
```
map_mix <- mixfit(map_mc, Nc = 2)
plot(map_mix)$mix + overlay_function(fun = dmix, args = list(mix = map_mix[[1]]),
    linetype = 2) + overlay_function(fun = dmix, args = list(mix = map_mix[[2]]),
    linetype = 2)
```



NOVARTIS

Mixtures improve accuracy of parametric MAP
... or automatic AIC based selection for number components

```
map_automix <- automixfit(map_mc)  # fits EM with 1-5 components and selects model with lowest AIC
plot(map_automix)$mix + ggtitle(NULL) + xlab(NULL)
```




```
print(map_automix)
```

```
## EM for Beta Mixture Model
## Log-Likelihood = 4443.006
##
## Univariate beta mixture
## Mixture Components:
##    comp1       comp2       comp3
## w  0.4802886  0.3950534  0.1246580
## a  6.0794485 28.4176518  2.5062925
## b 17.8864109 86.3824536  5.5103510
```

Generic Interface for Mixtures
Support for mixture Beta, Normal, Gamma & its predictive

- Mixtures are created in `RBesT` with `mixbeta`, `mixnorm` or `mixgamma` (or via `auto/mixfit`)
- Mixture components are defined by a **triplet** (w_k, a_k, b_k)

$$p(x, \mathbf{w}, \mathbf{a}, \mathbf{b}) = \sum_{k=1}^{K} w_k \, p_k(x, a_k, b_k)$$

- All standard `R` functions are supported `(d/p/q/r)mix`

```
# create a beta mixture with named components
bm <- mixbeta(inf1=c(0.5, 12, 7), inf2=c(0.5, 12, 4))
dmix(bm, x=c(0.1, 0.5)) # density
pmix(bm, q=c(0.1, 0.5)) # cumulative density
qmix(bm, p=c(0.1, 0.5)) # quantile function
rmix(bm, n=100)         # random number generation

# commands work the same for any mixture (replace bm with nm):
nm <- mixnorm(rob=c(0.2, 0, 2), inf=c(0.8, 2, 2), sigma=5)
```

Evaluating trial designs classically
Binary responder analysis

- Type I error α for no effect hypothesis
  $\theta_p = \theta_t$
- Sample size per group N_p & N_t chosen under true effect assumption (alternative) and desired type II error β
  $\theta_p, \theta_t = \theta_p + \delta$
- 1:1 randomization has highest efficiency
  $N_p = N_t$
- Type I error is controlled for any θ at a fixed α under the null

Evaluating trial designs with RBesT
Binary responder analysis

```
alpha <- 0.05
## 1. Define decision criterium for success
##    here: 2-sample decision criterium, P(p_placebo - p_treat <= 0) > 0.95
decision <- decision2S(1 - alpha, 0, lower.tail=TRUE)

## 2. Define design (priors, sample size, decision)
uniform_prior <- mixbeta(c(1, 1, 1))
design_uniform_classic <- oc2S(uniform_prior, uniform_prior, 24, 24, decision)

## 3. Evaluate power (type I error is controlled by design)
design_uniform_classic(0.25, 0.25)
```

```
## [1] 0.04927474
```

```
design_uniform_classic(0.25, 0.25 + 0.35)
```

```
## [1] 0.8187252
```

Evaluating trial designs with `RBesT` using MAP
Using MAP priors allows to reduce (control) sample size

- Informative MAP priors enable unequal randomization by *substituting* sample size of the control by prior information
- An informative prior can be considerd to have an *effective sample size* (ess)

```
## 0. derive MAP prior
## 1. perform classic operating characeristics
## 2. use ess as initial guess for sample size reduction

ess(map_automix, "moment") ## default (conservative)
```

```
## [1] 24
```

```
ess(map_automix, "morita") ## Morita et al. (2008)
```

```
## [1] 77
```

So we may substantially reduce the control group here!

Operating Characeristics for $N_p = 6$ and $N_t = 24$
First definition of design, then exact calculations (binary case)

```
# Define decision criterium, P(p_placebo - p_treat <= 0) > 0.95
decision <- decision2S(0.95, 0, lower.tail=TRUE)

treat_prior   <- mixbeta(c(1, 0.5, 1)) # Prior for treatment arm
placebo_prior <- mixbeta(c(1,11  ,32)) # Prior for placebo arm as used
uniform_prior <- mixbeta(c(1, 1  , 1)) # Uniform prior for comparison
map_robust    <- robustify(map_automix, weight=0.2, mean=0.5) # robust MAP

# Calculate design properties (depends on priors, sample size & decision)
design_uniform <- oc2S(uniform_prior, uniform_prior, 6, 24, decision)
design_trial   <- oc2S(placebo_prior, treat_prior  , 6, 24, decision)
design_robust  <- oc2S(map_robust   , treat_prior  , 6, 24, decision)

# Note: decision functions take mixtures as arguments and return
# 0="NO GO", 1="GO"
decision(postmix(map_robust, r=1, n=6), postmix(treat_prior, r=15, n=24))
## [1] 1
```

Analytic Operating Characteristics in `RBesT`
RBesT calculates OCs for one-sided designs fast

The decision function $D(y_1, y_2)$, priors and sample sizes uniquely define the **decision boundary** $D_1(y_2)$ (conditional critical values):

$$D_1(y_2) = \sup_{y_1}\{D(y_1, y_2) = 1\},$$

$$\iint f_1(y_1|\theta_1)\, D(y_1, y_2)\, f_2(y_2|\theta_2)\, dy_1\, dy_2 = \int F_1(D_1(y_2)|\theta_1)\, f_2(y_2|\theta_2)\, dy_2.$$

$D_1(y_2)$ is calculated when calling `oc2S`. Then all calls to the returned function evaluate the frequency for 1 assuming that y_1 (y_2) is distributed according to the assumed true value of θ_1 (θ_2).

Binary case calculation is exact, other endpoints use adaptive quadrature integration.

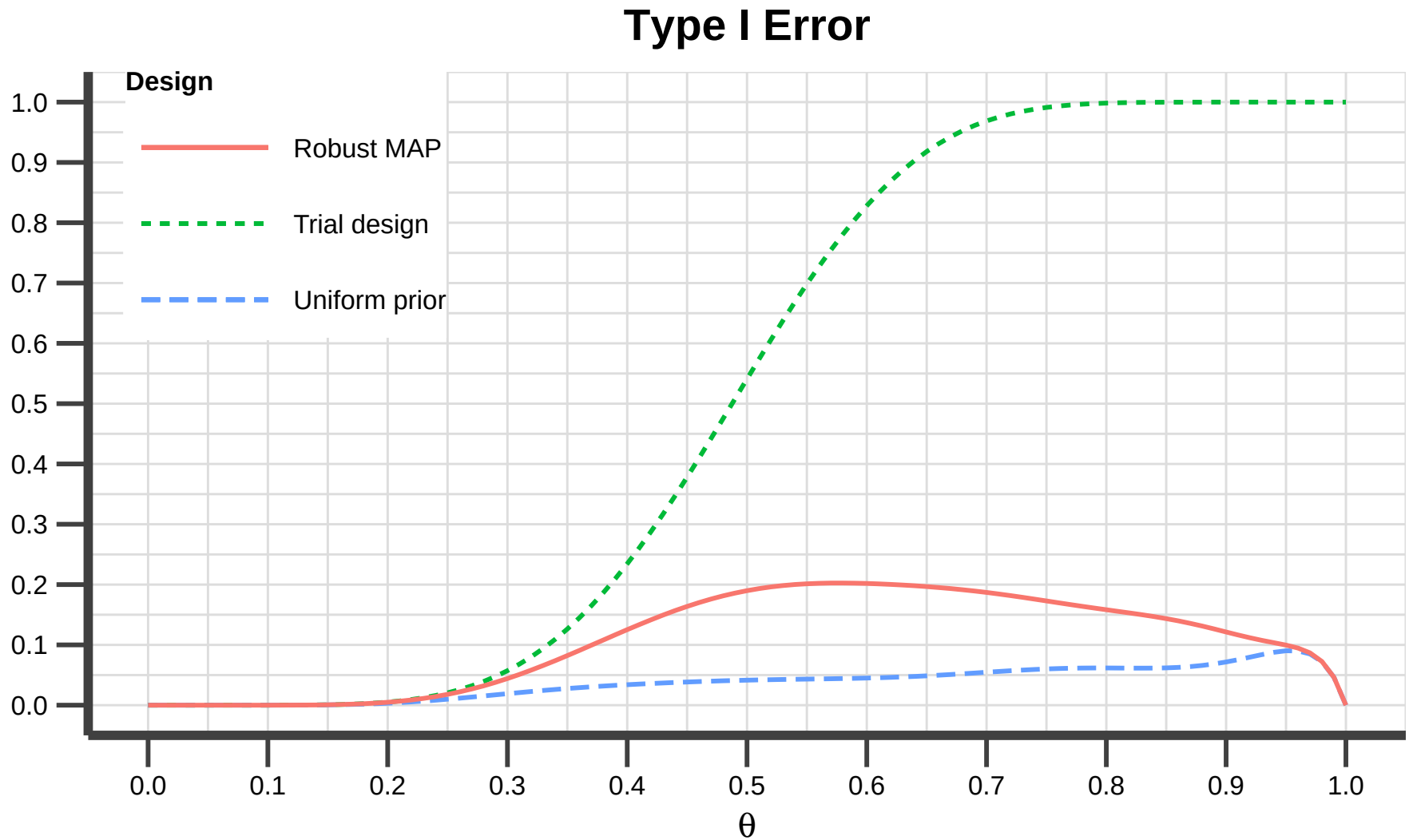
Operating Characteristics

```
delta <- c(0, 0.3, 0.6)
oc <- data.frame(theta_p=c(rep(0.25, 3), 0.25+delta),
                 theta_t=0.25+delta,
                 delta=c(delta, rep(0, 3))) %>%
    mutate(oc_uniform=design_uniform(theta_p, theta_t),
           oc_trial  =design_trial(  theta_p, theta_t),
           oc_robust =design_robust( theta_p, theta_t))
kable(oc, digits=2)
```

theta_p	theta_t	delta	oc_uniform	oc_trial	oc_robust
0.25	0.25	0.0	0.01	0.02	0.02
0.25	0.55	0.3	0.34	0.82	0.67
0.25	0.85	0.6	0.91	1.00	0.98
0.25	0.25	0.0	0.01	0.02	0.02
0.55	0.55	0.0	0.04	0.70	0.20
0.85	0.85	0.0	0.06	1.00	0.14

Type I Error, Frequency of GO for $\theta = \theta_t = \theta_p$

Comparing designs: robust MAP, trial, uniform

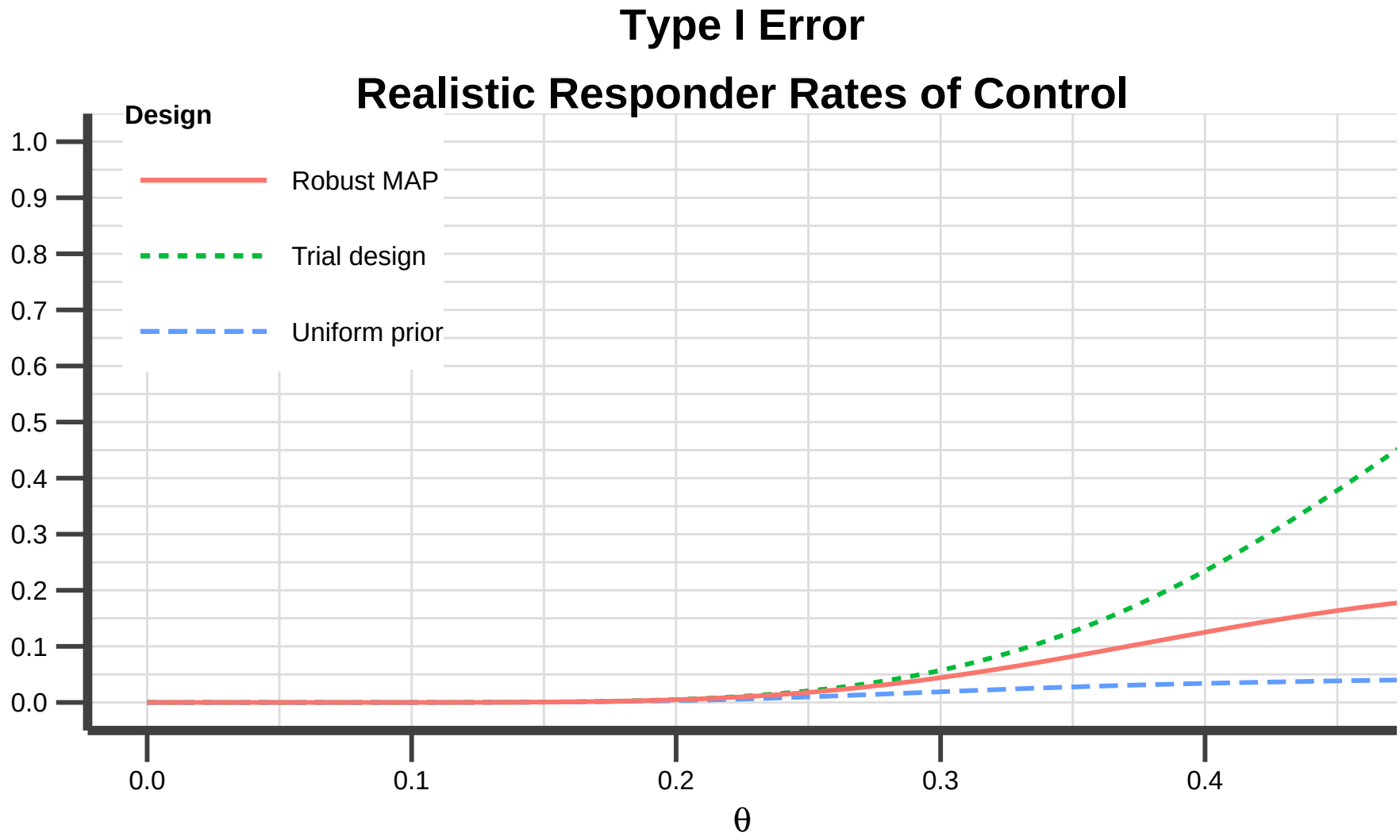


(code for the above plot is in the vignette)

NOVARTIS

Type I Error, Frequency of GO for $\theta = \theta_t = \theta_p$

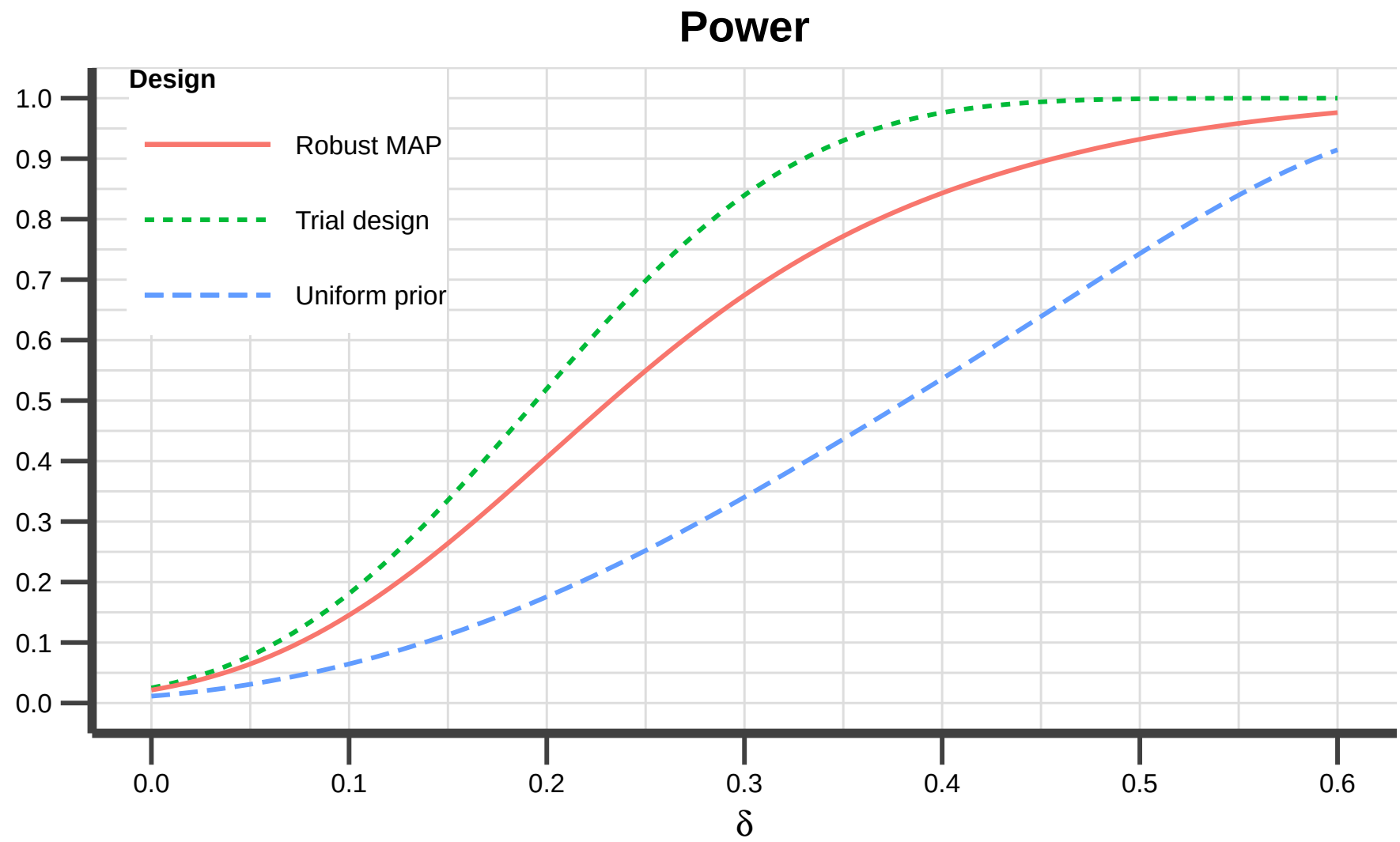
Comparing designs: robust MAP, trial, uniform



NOVARTIS

Power, Frequency of GO for $\theta_t = \theta_p + \delta$ $(\theta_p = \bar{\theta}_p)$

Comparing designs: robust MAP, trial, uniform

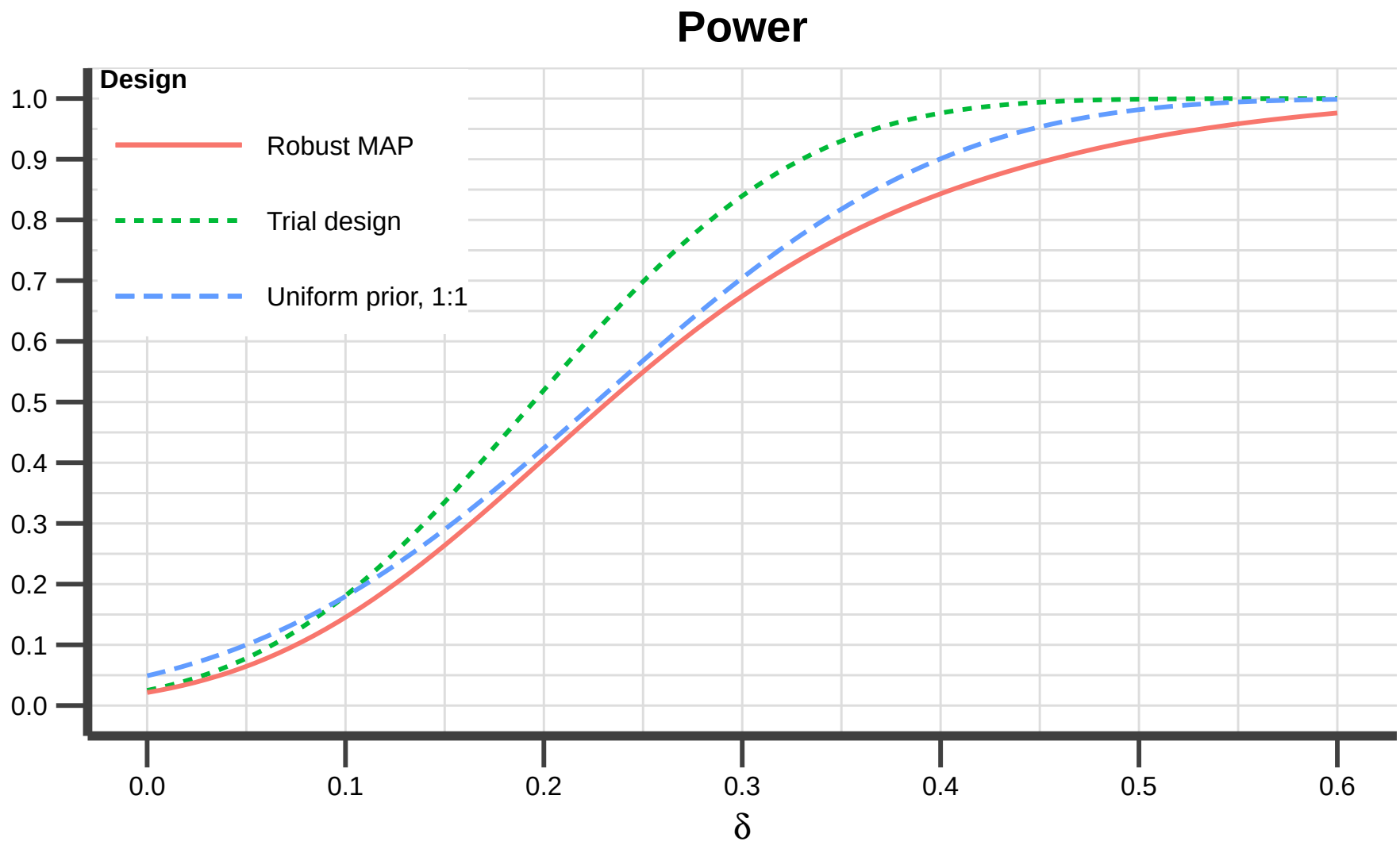


(code for the above plot is in the vignette)

NOVARTIS

Power, Frequency of GO for $\theta_t = \theta_p + \delta$ $(\theta_p = \bar{\theta}_p)$

Comparing designs: robust MAP, trial, uniform 1:1 (24 vs 24)

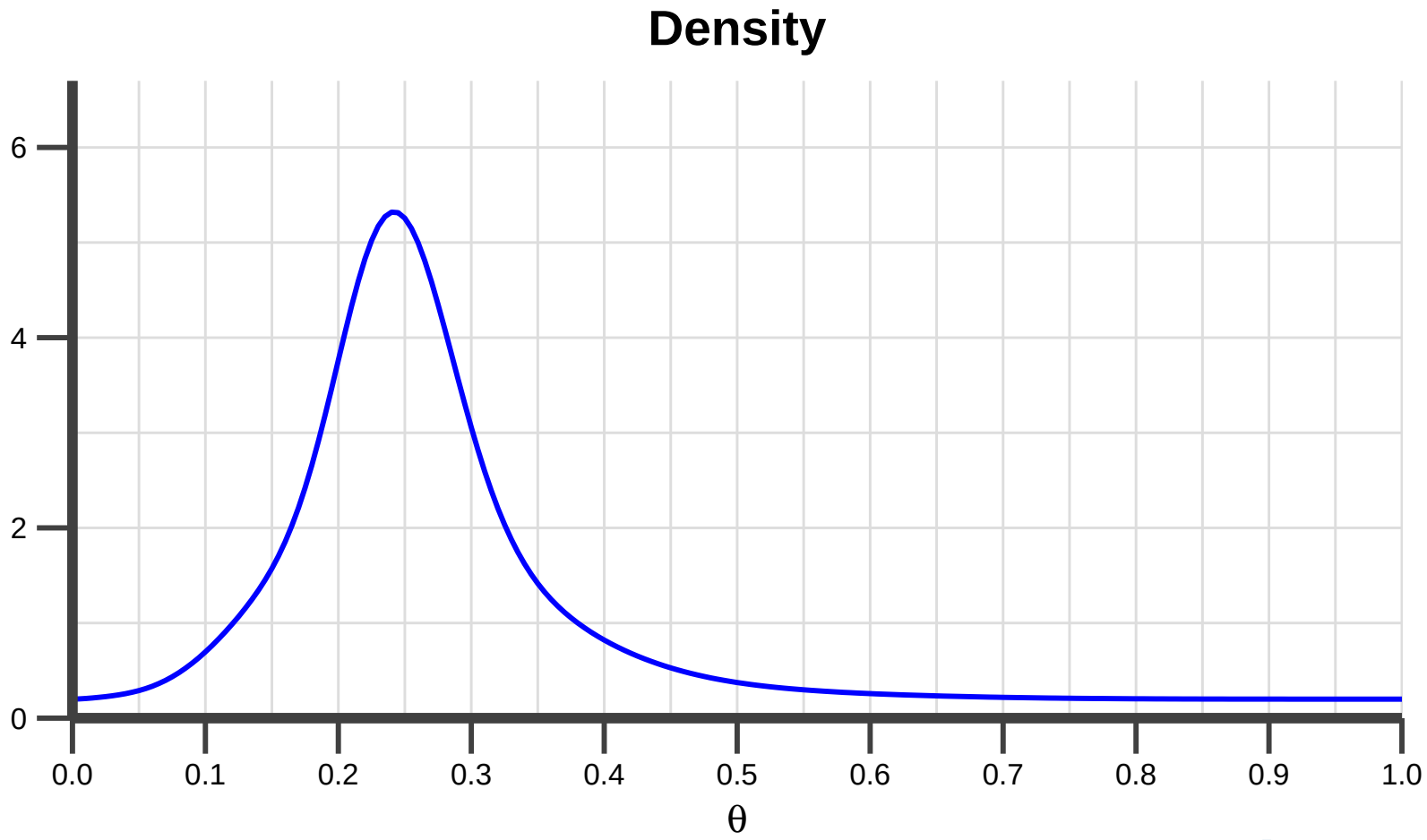


(code for the above plot is in the vignette)

NOVARTIS

Graphical Analysis of Control Densities
robust prior (blue)

$$p(\theta|y) = p(\theta)\, p(y|\theta)/p(y)$$

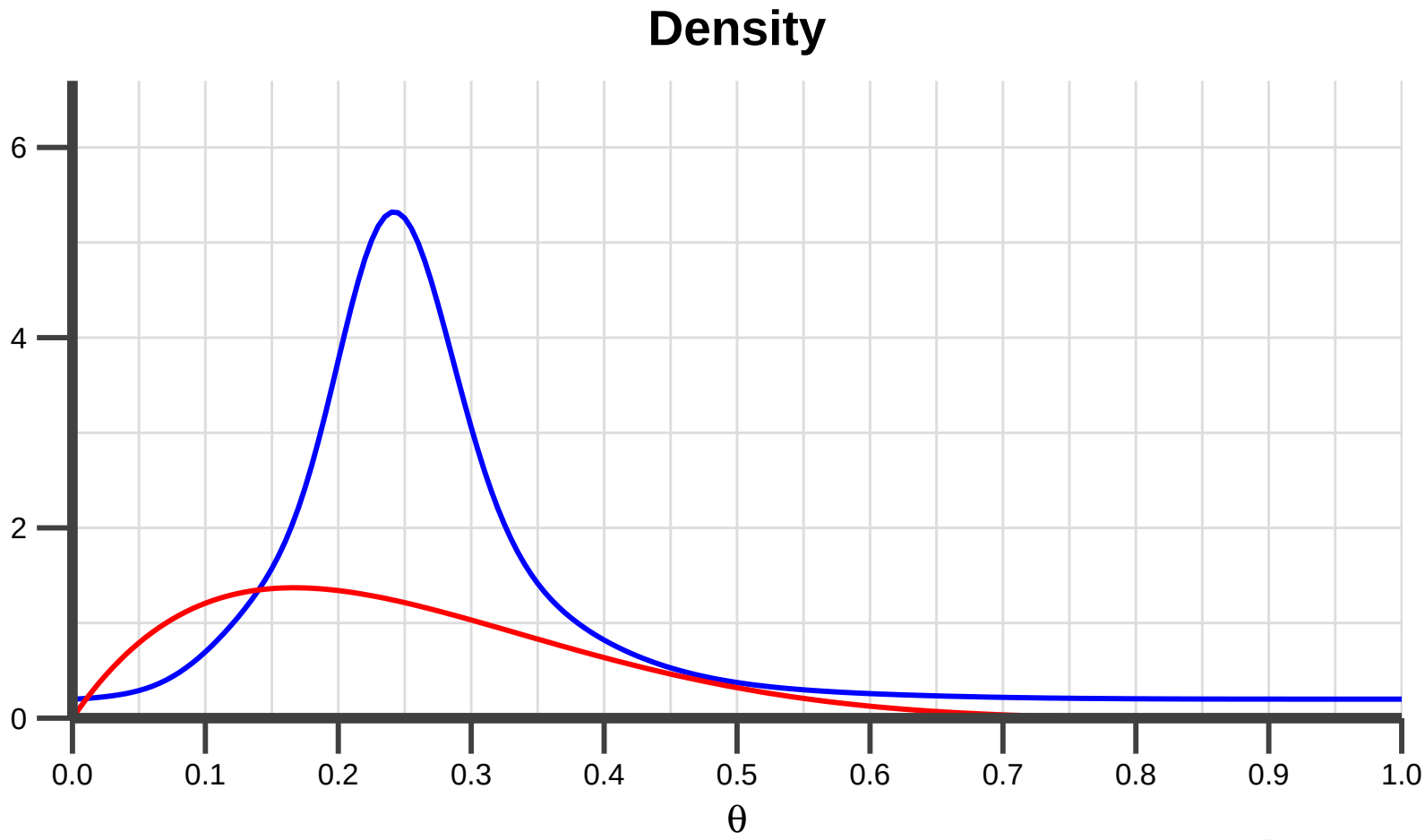


NOVARTIS

Graphical Analysis of Control Densities

robust prior (blue), likelihood / p(y) (red)

$$p(\theta|y) = p(\theta)\, p(y|\theta)/p(y)$$

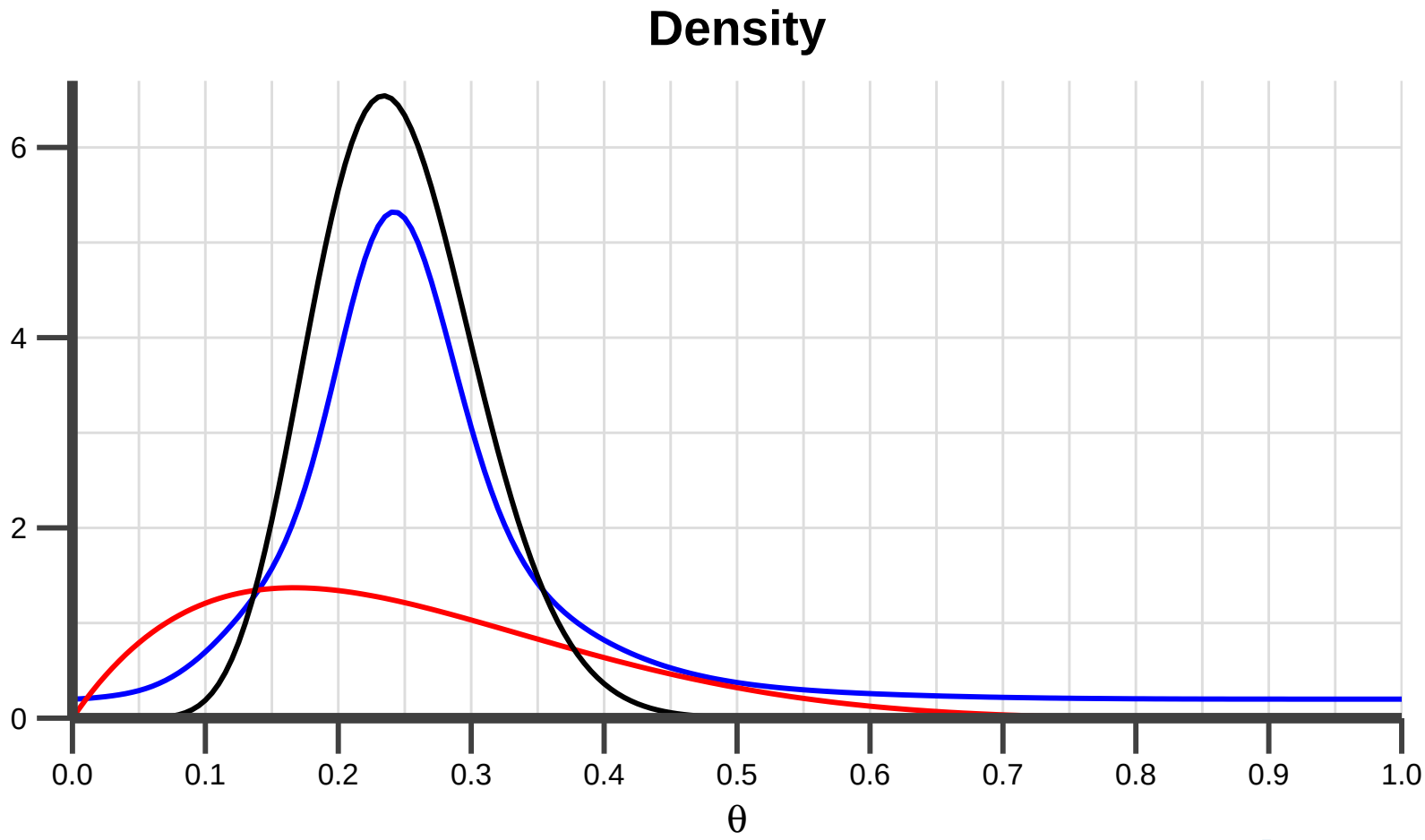


NOVARTIS

Graphical Analysis of Control Densities

robust prior (blue), likelihood / p(y) (red), posterior (black)

$$p(\theta|y) = p(\theta)\, p(y|\theta)/p(y)$$



NOVARTIS

Probability of Success

Getting RBesT and Help
RBesT is integrated into the R system

```
install.packages("RBesT") ## download and install RBesT from CRAN
library(RBesT)            ## load the library
?gMAP                     ## get help for gMAP
example(gMAP)             ## run the example for gMAP
help.search("postmix")    ## find help page for postmix
```

- Inter-linked HTML pages with `help.start()`
  opens a web-browser or RStudio help then follow
  `Packages -> RBesT`
- PDF reference distributed with RBesT (LaTeX formulas)
- **Vignettes**
  - introduction: Getting started (binary endpoint)
  - introduction normal: Getting started (normal endpoint)
  - customizing plots: Plotting help
  - robustMAP: Reproduces Schmidli et al. (2014)
  - ...

Useful Resources

Useful Resources

- RBesT R help
  `R> ?gMAP` for help on gMAP
- Vignettes
- CRAN: `https://cran.r-project.org/package=RBesT`
  - **Vignettes** binary & normal endpoint, plotting
  - Reference PDF manual `RBesT.pdf`
- Install RBesT (on CONNECT):
  `install.packages("RBesT", dependencies=TRUE)`
- Using RBesT: `library(RBesT)`

Excersises

Starting with `RBesT`

- Start with installing `RBesT` from CRAN

  ```
  R> install.packages("RBesT", dependencies=TRUE)
  ```

- Load the package and run your first analysis
  - `R> library(RBesT)`
  - `R> example(gMAP)`
- Open the help for the `gMAP` command using one of
  - `R> ?gMAP`
  - `R> help.search("gMAP")`
- Navigate to the additional help which include the vignettes
  - `R> help.start()`
    in Rstudio you can also click on the Packages tab at the right
  - Follow the links
    Packages -> RBesT -> User guides, package vignettes, ...
  - Explore the different documents

Vignette: Getting started with `RBesT` (binary)

1. Work through the vignette
   "Getting started with `RBesT` (binary)"
   Hint: You can download the R code from the help
   $\Rightarrow$ "R code" link at the right of the overview page **Note:**
   Please uncomment the line `## is_CRAN <- FALSE`
2. How much is the type I error inflated for the robust MAP prior? Consider why this happens.
3. Compare operating characteristics for a robust MAP prior with 80% and 50% weight on the MAP prior.
4. Evaluate further the difference the prior on τ makes.
   Compare the $\tau \sim \text{HalfNormal}(0, 1)$ with the
   $\tau \sim \text{HalfNormal}(0, (1/2)^2)$ prior.
   - 4.1 How do the posteriors for each quantity differ (β, τ, θ_*)?
   - 4.2 Differences in power and type I error?
   - 4.3 Repeat the comparison, but only use the first 3 studies

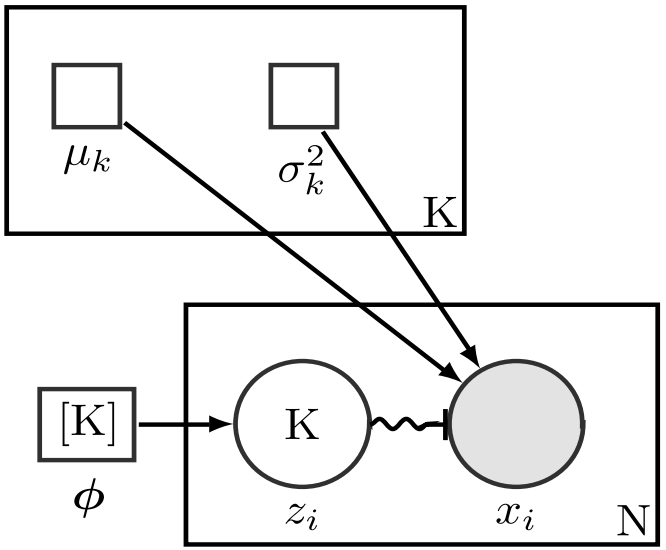
Probability of Success

1. Work through the probability of success material.
2. What is the probability of success for a phase III trial to be successful **before** initiating these based on the PoC and phase II data only?
3. How large is the probability of success for **both** phase III studies to be successfull based on the historical data only?

Backup

Mixture Models
Estimation with Expectation-Maximization (EM)

$$\log p(\boldsymbol{x}|\boldsymbol{w},\boldsymbol{a},\boldsymbol{b}) = \sum_{n=1}^{N} \log \left[\sum_{k=1}^{K} w_k \, p(x_n|a_k, b_k) \right]$$



Example: Univariate normal

Source: Wikipedia

EM "trick" is to extend the likelihood

$$p(\boldsymbol{x}|\boldsymbol{w},\boldsymbol{a},\boldsymbol{b}) = \int p(\boldsymbol{x},\boldsymbol{z}|\boldsymbol{w},\boldsymbol{a},\boldsymbol{b})\, d\boldsymbol{z}$$

- x **observed** data as recorded
- z **latent** data, i.e. component indicator
- (x, z) **complete** data

Posterior Analysis for Mixture Priors
Fixed prior weights change in the posterior

Assume a mixture prior for some parameter θ

$$p(\theta, \mathbf{w}, \mathbf{a}, \mathbf{b}) = \sum_{k=1}^{K} w_k \, p_k(\theta, a_k, b_k)$$

for data y and likelihood $f(y|\theta)$, then the posterior is again a mixture equal to the posterior of each component and updated weights

$$p(\theta, \mathbf{w}, \mathbf{a}, \mathbf{b}|y) = \sum_{k=1}^{K} w'_k \, p_k(\theta, a_k, b_k|y)$$

Note: The prior weights w_k, are not random (fixed) but are still updated to $w'_k = w^*_k / \sum_{k=1}^{K} w^*_k$ with

(marginal likelihood) $w^*_k \equiv w_k \int f(y|\theta) \, p_k(\theta, a_k, b_k) \, d\theta = w_k \, p_k(y)$

References

Baeten, D., Baraliakos, X., Braun, J., Sieper, J., Emery, P., Heijde, D. van der, et al. (2013) Anti-interleukin-17A monoclonal antibody secukinumab in treatment of ankylosing spondylitis: A randomised, double-blind, placebo-controlled trial. *The Lancet*, **382**, 1705–1713.

Dalal, S.R. and Hall, W.J. (1983) Approximating Priors by Mixtures of Natural Conjugate Priors. *Journal of the Royal Statistical Society. Series B (Methodological)*, **45**, 278–286.

Neuenschwander, B., Capkun-Niggli, G., Branson, M. and Spiegelhalter, D.J. (2010) Summarizing historical information on controls in. *Clin Trials*, **7**, 5–18.

Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D. and Neuenschwander, B. (2014) Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biom*.

Spiegelhalter, D.J., Abrams, K.R. and Myles, J.P. (2004) *Bayesian Approaches to Clinical Trials and Health-Care Evaluation: Spiegelhalter/Clinical Trials and Health-Care Evaluation*. John Wiley & Sons, Ltd, Chichester, UK.

West, M. (1993) Approximating Posterior Distributions by Mixture. *Journal of the Royal Statistical Society. Series B (Methodological)*, **55**, 409–422.