Bayesian Predictive power: the bathtub phenomenon

Kaspar Rufibach[1,2]
Joint work with Markus Abt[3] *and Uli Burger*[1,3]
[1] *Methods, Collaboration, and Outreach (MCO),* [2] *Oncology Biostatistics,*
[3] *CNS & Metabolism Biostatistics*
BBS Seminar, Basel, 27th October 2015

Generated 2015-10-27 at 13:37:12.

Agenda

1. Bayesian Predictive Power
2. Excursion: summary measure of a distribution
3. Choice of prior for Bayesian Predictive Power
4. "Quantify uncertainty" for Bayesian Predictive Power?

Agenda

Bayesian Predictive Power

Continuous endpoint, true effect Δ, estimator assumed to follow Normal distribution.

Estimate $\widehat{\Delta}_{\text{final}}$ at final analysis of pivotal trial, based on n_{final} observations:

$$\widehat{\Delta}_{\text{final}} \quad \sim \quad N(\Delta, \sigma^2_{\text{final}} = \sigma^2 / n_{\text{final}}).$$

Pivotal trial is called a success if $\widehat{\Delta}_{\text{final}} \leq \Delta_{\text{suc}}$ (think of log hazard ratio).

Δ_{suc}: can be

- **Minimal detectable difference** (MDD), i.e. effect size such that trial is just significant: Equate standardized test statistic to critical value of z-test: $\Delta_{\text{suc}}/\text{SE}(\widehat{\Delta}_{\text{final}}) = z_{\alpha_{\text{final}}/2} \Rightarrow \Delta_{\text{suc}} = z_{\alpha_{\text{final}}/2}\text{SE}(\widehat{\Delta}_{\text{final}})$.
- Any **other quantity of interest**, e.g. assumed alternative in sample size planning = target product profile (TPP).

Bayesian Predictive Power

Quantity of interest = **power function**:

$$P(\widehat{\Delta}_{\text{final}} \leq \Delta_{\text{suc}}) \quad = \quad \Phi\Big(\frac{\Delta_{\text{suc}} - \Delta}{\sigma_{\text{final}}}\Big).$$

Depends on assumed (or true) effect Δ! What can we do?

1. Provide $P(\widehat{\Delta}_{\text{final}} \leq \Delta_{\text{suc}})$ for different assumed values of Δ.
2. Assume distribution over Δ with density q and average:

$$\begin{aligned} \text{PoS} &= \mathbb{E}_{\Delta}\Big(P_{\Delta}(\widehat{\Delta}_{\text{final}} \leq \Delta_{\text{suc}})\Big) \\ &= \int_{-\infty}^{\infty} \Phi\Big(\frac{\Delta_{\text{suc}} - \Delta}{\sigma_{\text{final}}}\Big) q(\Delta) d\Delta. \end{aligned}$$

Bayesian predictive power initially introduced in Spiegelhalter et al. (1986).

Various names in the literature, we use "Probability of Success" (PoS).

Bayesian Predictive Power

$$\begin{aligned} \mathrm{PoS} &= \mathbb{E}_\Delta\Big(P_\Delta(\widehat{\Delta}_{\text{final}} \leq \Delta_{\text{suc}})\Big) \\ &= \int_{-\infty}^{\infty} \Phi\Big(\frac{\Delta_{\text{suc}} - \Delta}{\sigma_{\text{final}}}\Big) q(\Delta) d\Delta. \end{aligned}$$

Why "**Bayesian**"?

- At the time, maybe simply because a distribution on a parameter was assumed.
- Start with $q \Rightarrow$ after interim where you **learn** interim estimate (collaborative group framework) update prior $q = q_{\text{prior}}$ with data likelihood to get posterior $q_{\text{posterior}} \Rightarrow$ use this to update PoS.
- Can also update q_{prior} with external data, e.g. other studies, competitor data, etc.

After interim: Power $P_\Delta(\widehat{\Delta}_{\text{final}} \leq \Delta_{\text{suc}})$ becomes **conditional power** $P_\Delta(\widehat{\Delta}_{\text{final}} \leq \Delta_{\text{suc}} | \widehat{\Delta}_{\text{interim}} = \Delta_{\text{interim}})$!

Update after **blinded** interim: Rufibach et al. (2015).

Quantities

1. **Power** $P(\widehat{\Delta}_{\text{final}} \leq \Delta_{\text{suc}})$. At trial start, function of assumed effect Δ.
2. **Conditional power**: "Updated" power after trial has started, function of Δ.
3. **Bayesian predictive power** (BPP): average over (conditional) power with respect to distribution over Δ.
4. **Predictive probability**: At interim, what is probability to beat Δ_{suc} at final, given current information (expectation of posterior conditional on every possible future outcome)? Depends on prior parameters. See e.g. Berry et al. (2011) for details.

Different quantities that

- depend on different assumptions,
- have different properties,
- have different interpretations.

Keep them apart!

Time-to-event framework

Approximate distribution of **estimated log(hazard ratio)** $\widehat{\theta} := \log \widehat{\lambda}$:

$$\widehat{\theta} \approx N(\theta, 4/d).$$

- $\theta = \log \lambda$: **true underlying effect**, true log-hazard ratio.
- 1:1 randomized trial: $\mathrm{Var}(\widehat{\theta}) = 4/d$.
- d: total number of events in both arms.

In context of pivotal trial:

- Random variable $\widehat{\theta}_{\text{final}} \sim N(\theta, \sigma^2_{\text{final}} = 4/d_{\text{final}})$.
- d_{final}: number of events at final analysis.
- α_{final}: significance level at final analysis. May be adjusted for group-sequential design.

Closed form of PoS if prior is Normal

Lemma (Explicit computation of PoS)

Assuming the prior is Normal with density q_{prior}, mean θ_0, variance σ_0^2, and is independent of the random variable $\widehat{\theta}_{final}$. Then

$$PoS := \int P_\theta(\widehat{\theta}_{final} \leq \theta_{suc}) q_{prior}(\theta) d\theta \quad = \quad \Phi\Big(\frac{\theta_{suc} - \theta_0}{\sqrt{\sigma_{final}^2 + \sigma_0^2}}\Big).$$

Proof: Use law of total probability and properties of Normal distribution. See Rufibach et al. (2015).

References containing alternative proofs: Spiegelhalter et al. (1986), O'Hagan et al. (2005), Proschan et al. (2006), or Dmitrienko and Wang (2006).

Example

Assumptions:

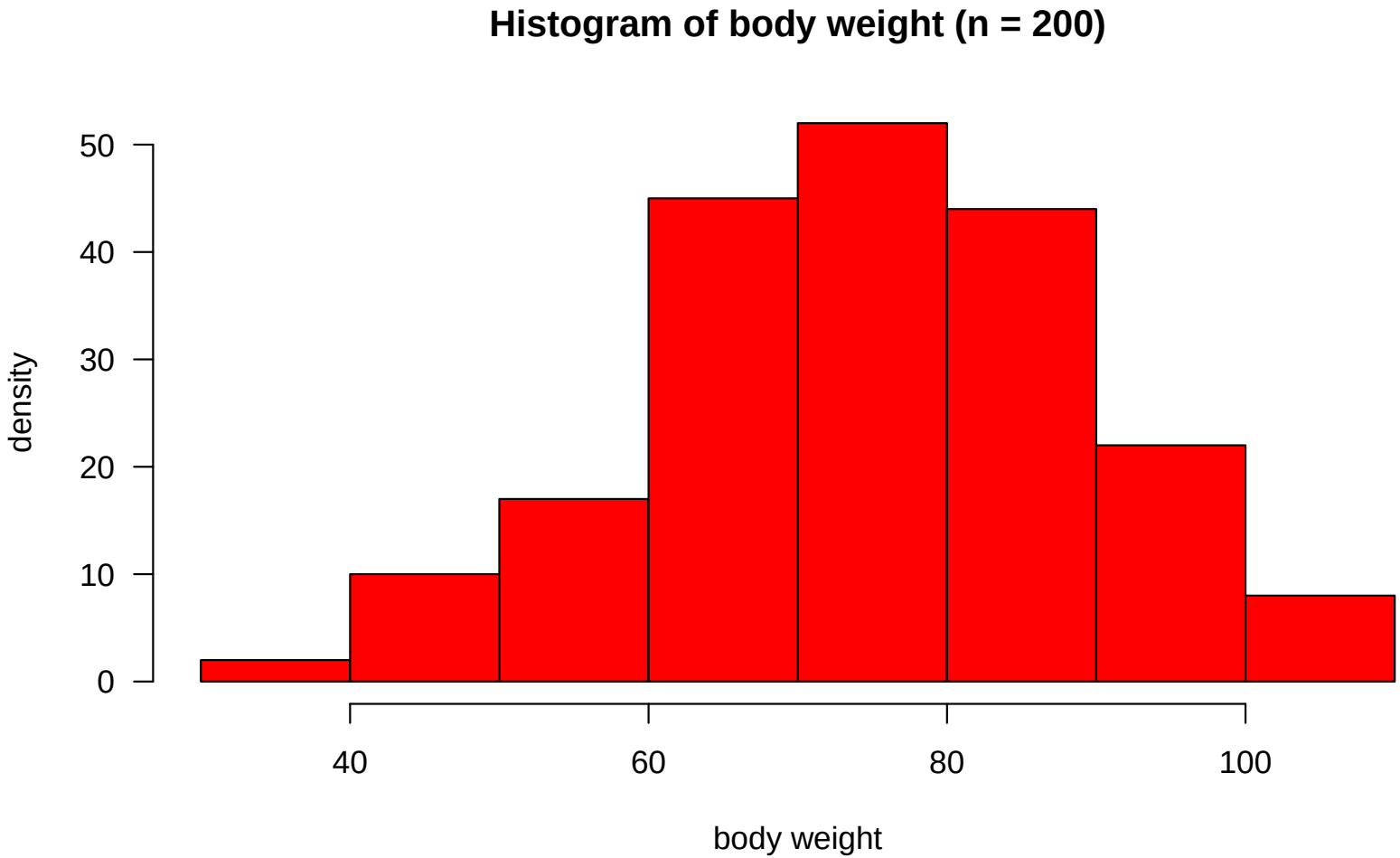
- Phase 2 result: $\widehat{\theta}_{\text{Phase 2}} = \log(\mathbf{0.700})$, based on $d_{\text{prior}} = 50$ **events**.
- Hazard ratio used as alternative in sample size computation: 0.75.
- Final analysis after $d_{\text{final}} = 380$ events.
- $\alpha_{\text{final}} = 0.050$.
- Minimal detectable difference: $\theta_{\text{suc}} = \log(0.818)$.

PoS at start of Phase 3 trial, assuming we know the Phase 2 result:

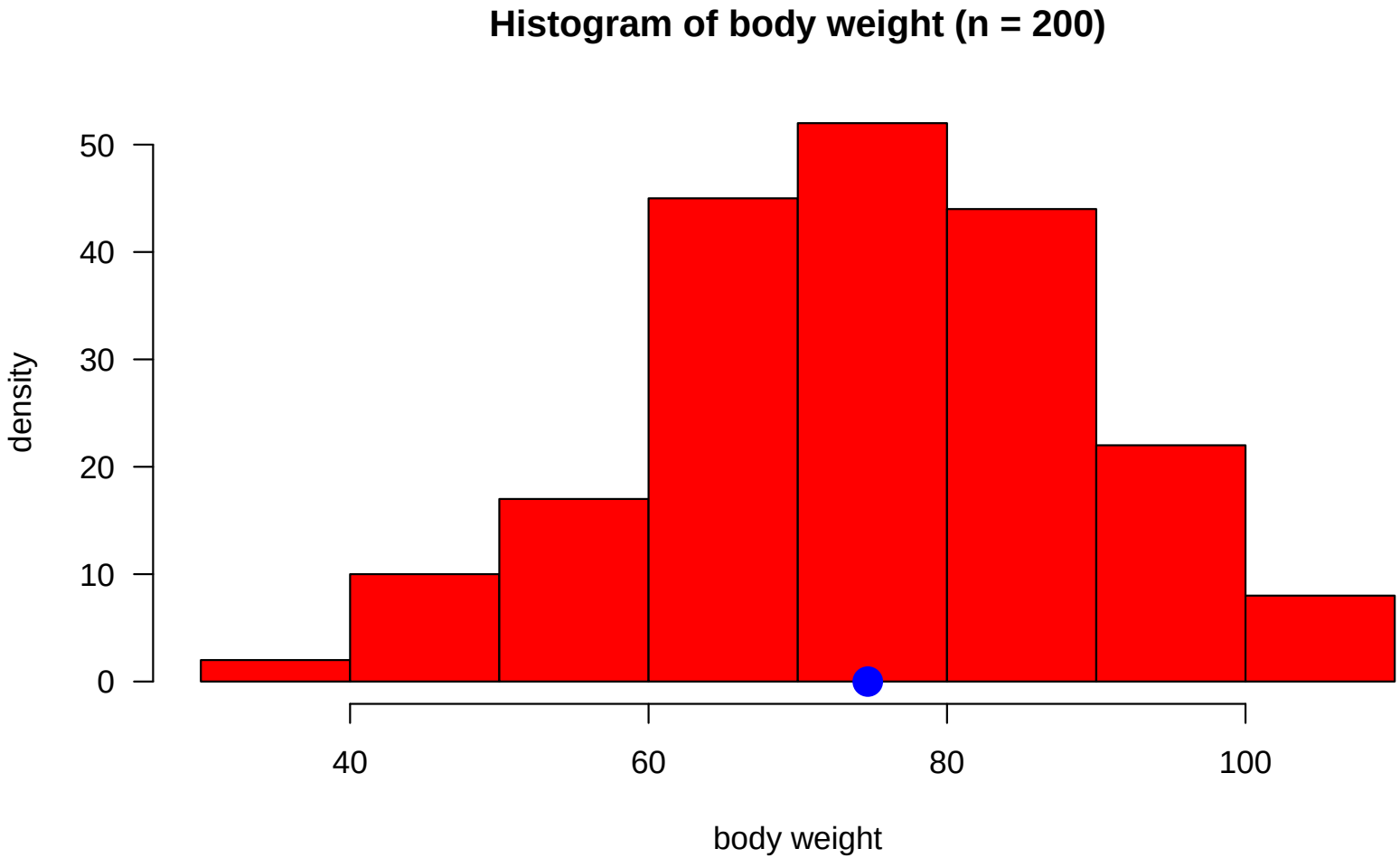
$$\text{PoS} = \int_{-\infty}^{\infty} P_\theta(\widehat{\theta}_{\text{final}} \leq \theta_{\text{suc}}) \phi_{\mu=\log(0.700),\sigma^2=4/50}(\theta) d\theta = \mathbf{0.697}.$$

Agenda

How do you summarize this density in one number?



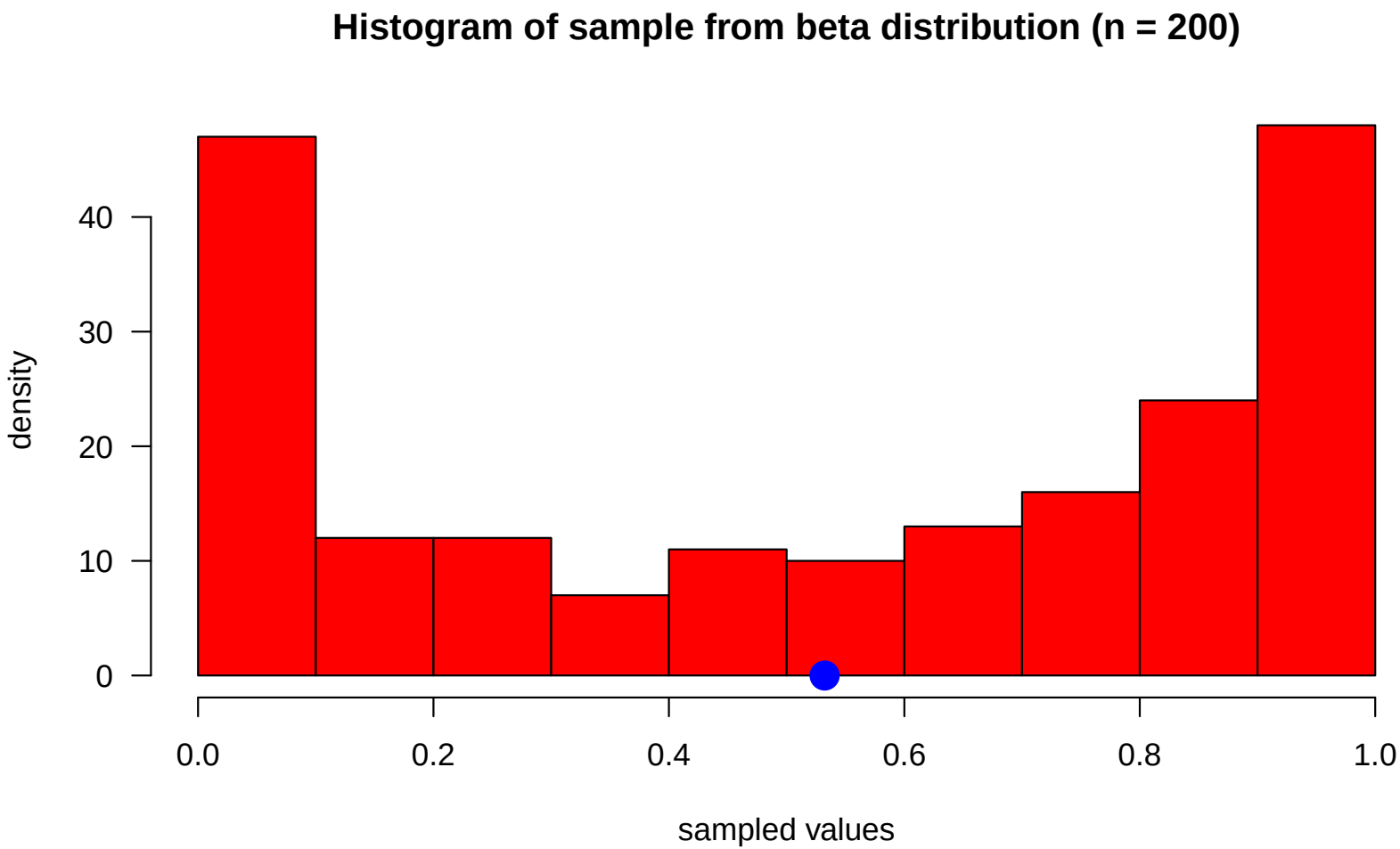
How do you summarize this density in one number?



Using the **mean**!

- Value that is most common, i.e. mean $\approx$ mode.
- Represents center of data.

How do you summarize this density in one number?



Really using the **mean**? Or rather provide histogram? Or table with frequencies?

Agenda

Recall definitions and example

PoS definition:

$$\begin{aligned} \text{PoS} &= \mathbb{E}_\theta\Big(P_\theta(\widehat{\theta}_{\text{final}} \leq \theta_{\text{suc}})\Big) \\ &= \int_{-\infty}^{\infty} P_\theta(\widehat{\theta}_{\text{final}} \leq \theta_{\text{suc}}) q_{\text{prior}}(\theta) d\theta. \end{aligned}$$

Power function:

$$T(\theta) := P_\theta(\widehat{\theta}_{\text{final}} \leq \theta_{\text{suc}}) = \Phi\Big(\frac{\theta_{\text{suc}} - \theta}{\sigma_{\text{final}}}\Big).$$

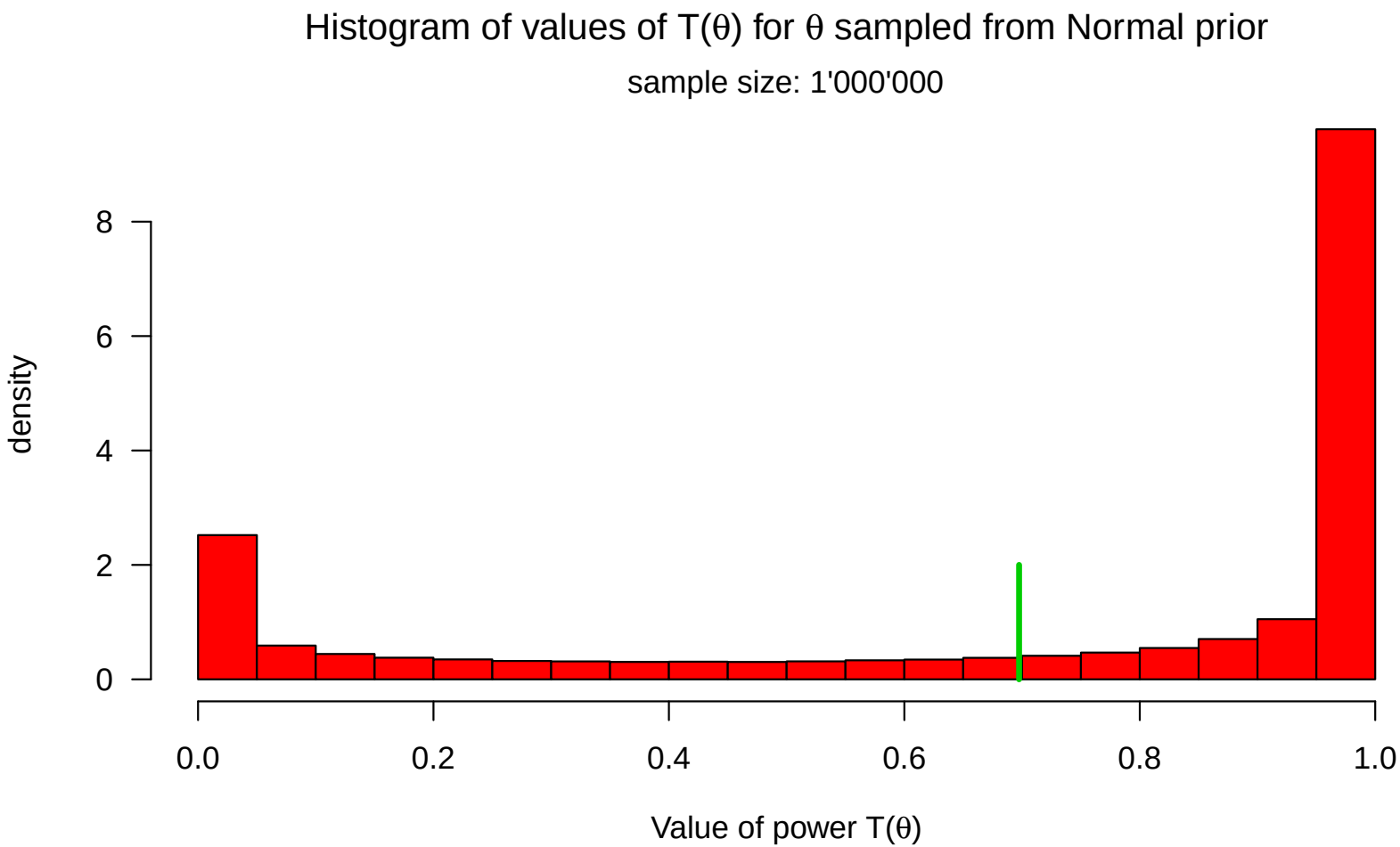
Compute PoS via simulation (law of large numbers!):

- Draw a sample $(\widehat{\theta}_1, \ldots, \widehat{\theta}_M)$ from prior.
- Compute $(T(\widehat{\theta}_1), \ldots, T(\widehat{\theta}_M))$.
- PoS = average over these values.

Recall example:

- Phase 2 result: $\widehat{\theta}_{\text{Phase 2}} = \log(\mathbf{0.700})$, based on 50 events.
- Phase 3 final analysis: Minimal detectable difference $\theta_{\text{suc}} = \log(\mathbf{0.818})$ based on $d_{\text{final}} = 380$ events.

Simulate PoS in example



1. Is mean really appropriate number to **summarize** this histogram?
2. Can we compute this density?

Density of power $T(\Theta)$

Assume prior r.v. Θ with PDF q, CDF Q, and define $Y := T(\Theta)$ with PDF g, CDF G.

Use **transformation theorem** and **rule about derivative of an inverse** to get:

$$\begin{aligned} G(y) &= 1 - Q(\theta_{\text{suc}} - \sigma_{\text{final}} z), \\ g(y) &= q(\theta_{\text{suc}} - \sigma_{\text{final}} z) \frac{\sigma_{\text{final}}}{\phi(z)} \end{aligned}$$

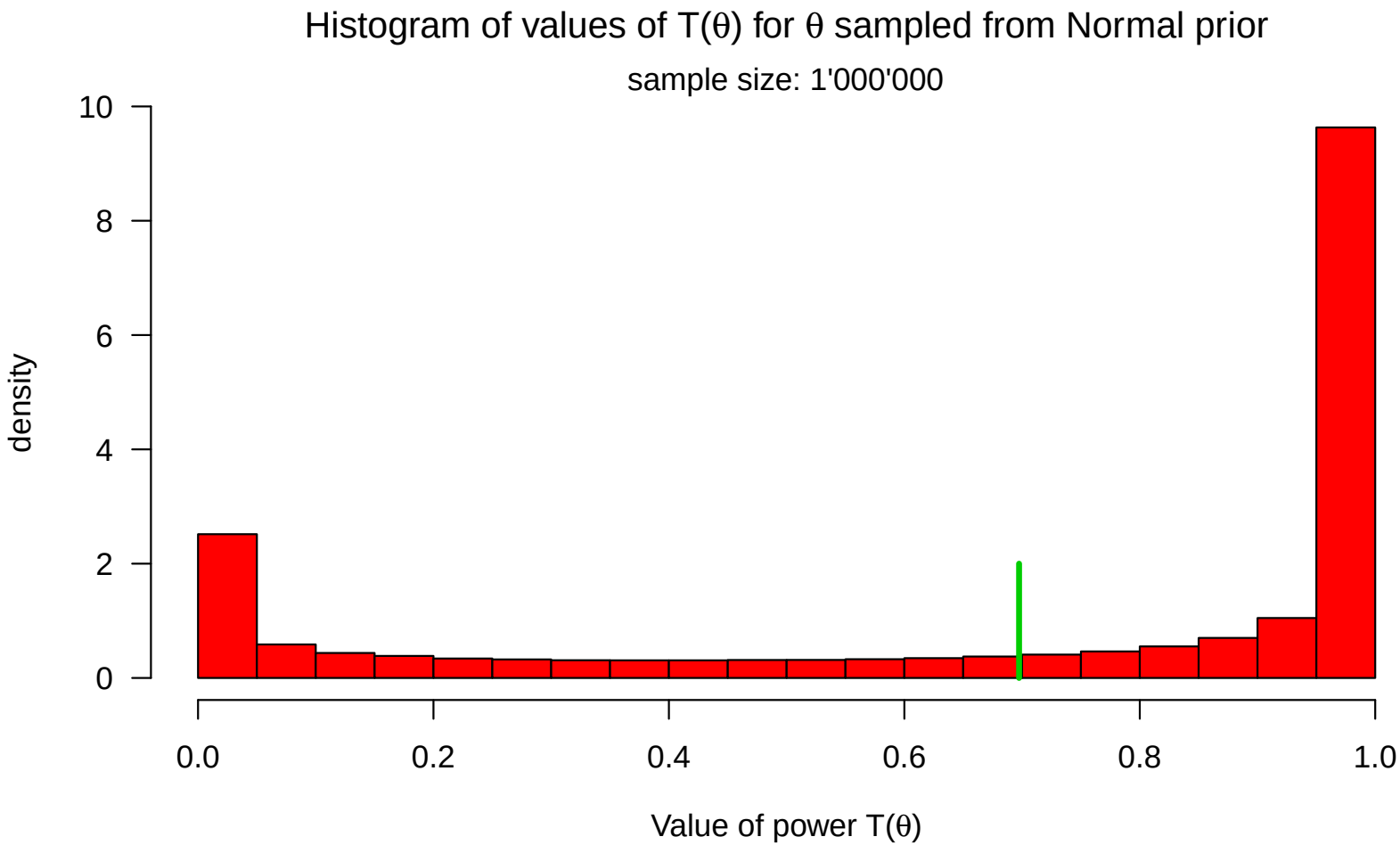
where $z := \Phi^{-1}(y)$ and ϕ the standard Normal density function.

If we assume $\Theta \sim N(\theta_0, \sigma_0^2)$:

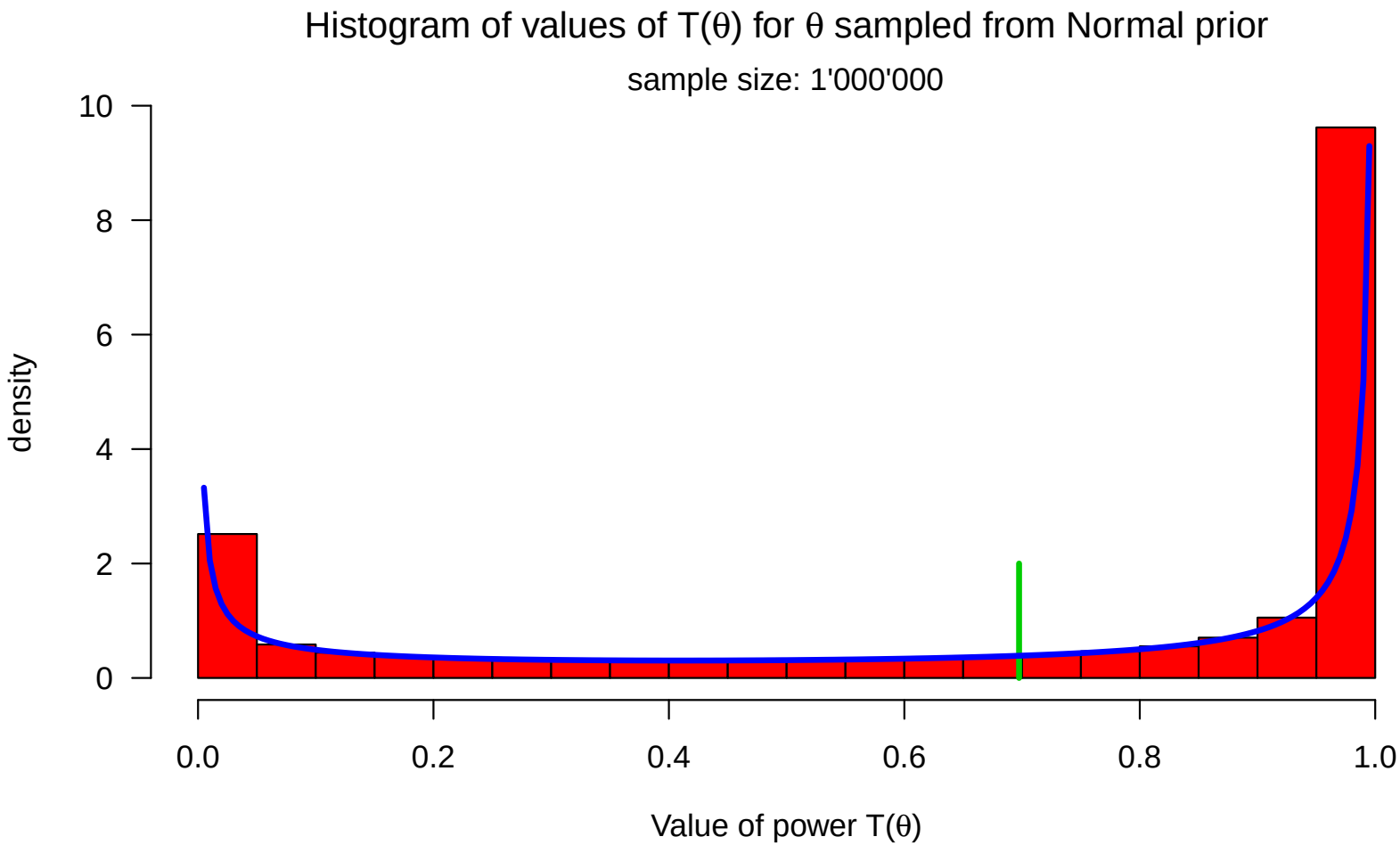
$$\begin{aligned} G(y) &= 1 - \Phi(\beta - \alpha z), \\ g(y) &= \alpha \phi(\beta - \alpha z) \Big[\phi(z)\Big]^{-1} \end{aligned}$$

where $\alpha = \sigma_{\textbf{final}}/\sigma_0 > 0$ and $\beta = (\theta_{\textbf{suc}} - \theta_0)/\sigma_0$.

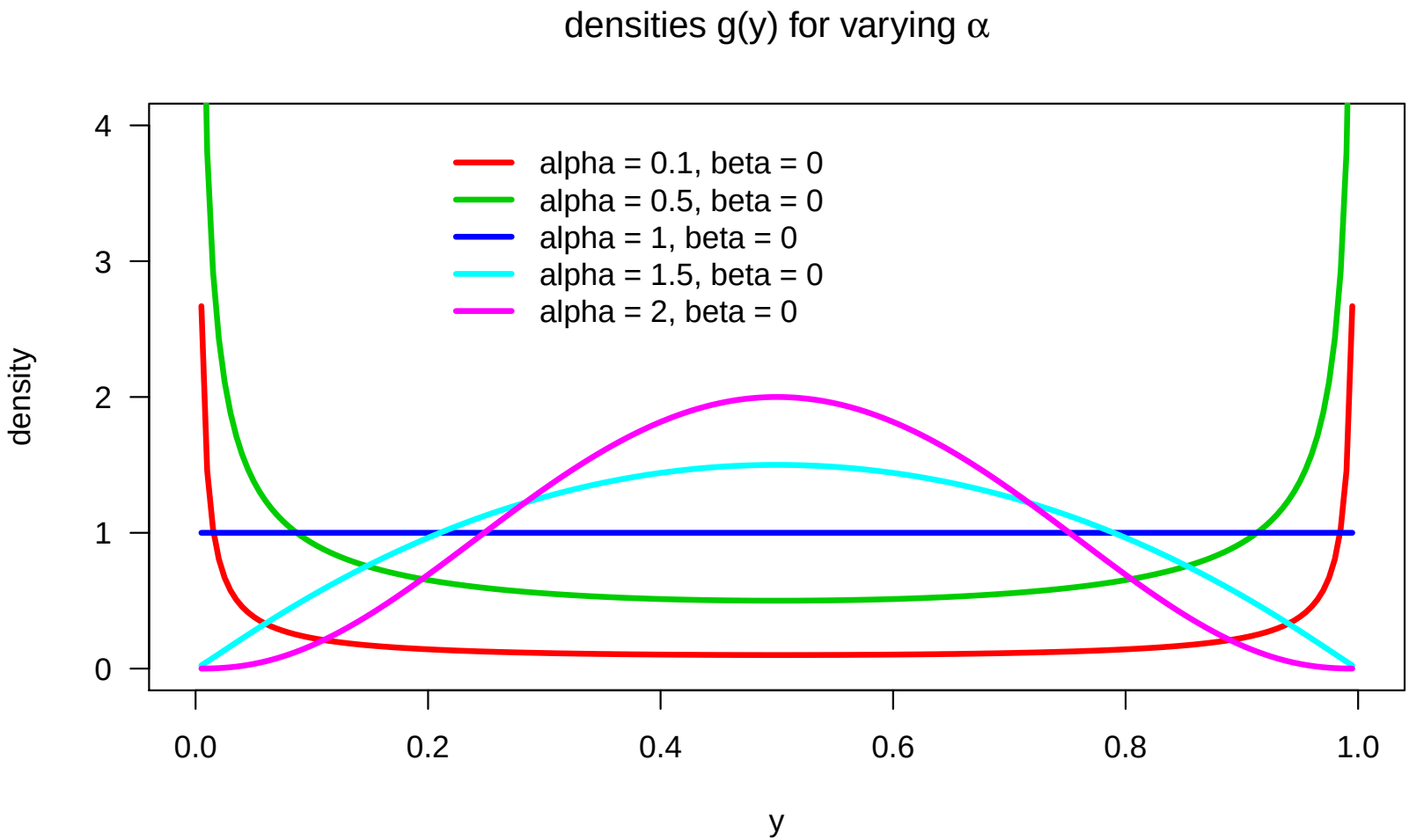
Simulate PoS in example



Simulate PoS in example

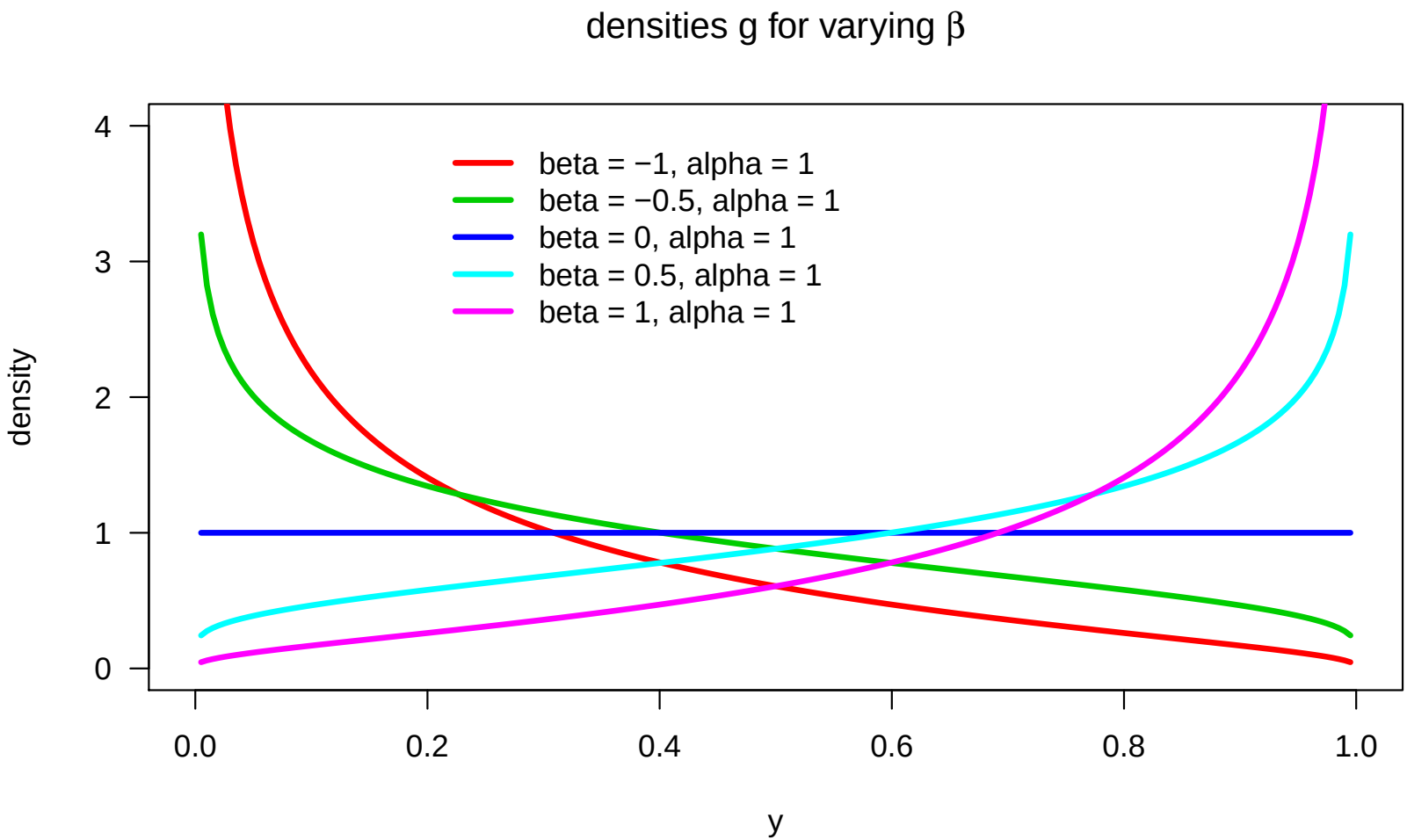


Density g as a function of α, for $\beta = 0$



- When summarizing g with PoS $\Rightarrow$ unimodal density most sensible?
- $\alpha = 1$: transition between "bathtub-shaped" (even convex?) and unimodal (obviously not concave).
- Make qualitative features precise!

Density g as a function of β, for $\alpha = 1$



- Make qualitative features precise!

Qualitative features of g

Theorem (Qualitative features of g)

We have the following statements:

1. *If* $\alpha = 1$, *then* g *is*

$$\begin{cases} \textit{strictly decreasing for} & \beta < 0, \\ \textit{constant for} & \beta = 0, \\ \textit{strictly increasing for} & \beta > 0. \end{cases}$$

   on $[0, 1]$. *Minima and maxima of* g *are accordingly either at 0 or 1.*

2. *If* $\alpha \neq 1$ *then* g

$$\begin{cases} \textit{has a minimum at } y_m \textit{ if} & \alpha < 1, \\ \textit{has a maximum at } y_m \textit{ if} & \alpha > 1, \end{cases}$$

   for $y_m = \Phi(\alpha\beta/(\alpha^2 - 1))$. *Furthermore,* g

$$\begin{cases} \textit{is decreasing for } y < y_m \textit{ and increasing for } y > y_m \textit{ if} & \alpha < 1, \\ \textit{is increasing for } y < y_m \textit{ and decreasing for } y > y_m \textit{ if} & \alpha > 1. \end{cases}$$

Proof: Compute g', g'', discuss these.

Why? And what does it mean?

Simplest case: $\alpha = \beta = 0 \Rightarrow d_{\text{prior}} = d_{\text{final}}, \theta_0 = \theta_{\text{suc}} \Rightarrow g$ uniform.

Prior and distribution of pivotal effect size have same variance $\Rightarrow$ power becomes uniform, either you beat θ_{suc} or not, with equal probability.

Why P(extreme PoS values) so high if $\alpha < 1$? $d_0 < d_{\text{final}} \Rightarrow$ high variance of prior $\Rightarrow$ high probability to have extreme HRs $\Rightarrow$ power for these is either almost 0 or 1.

g unimodal if $\alpha > 1 \Rightarrow \sigma_{\text{final}} > \sigma_0 \Rightarrow d_{\text{final}} < d_0$. **Unrealistic** in clinical development!

How should we choose prior to get unimodal PoS distribution?

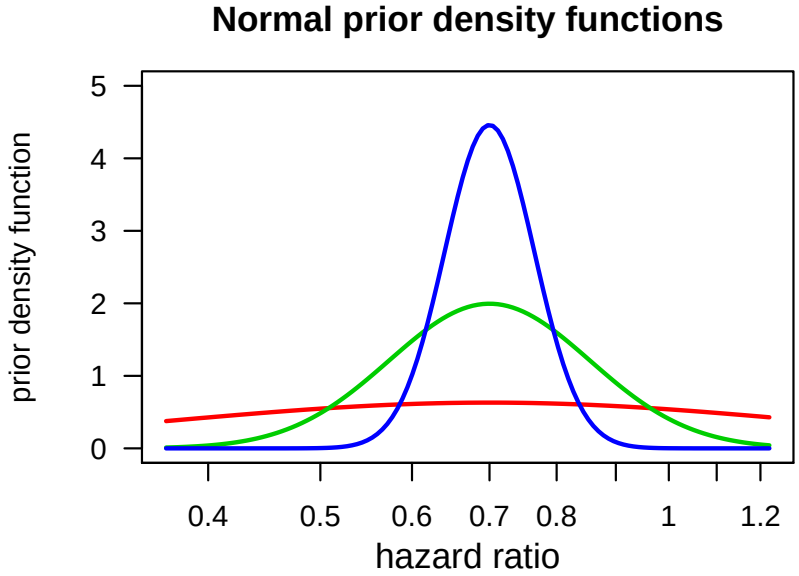
Alternative prior 1: simply choose a large variance

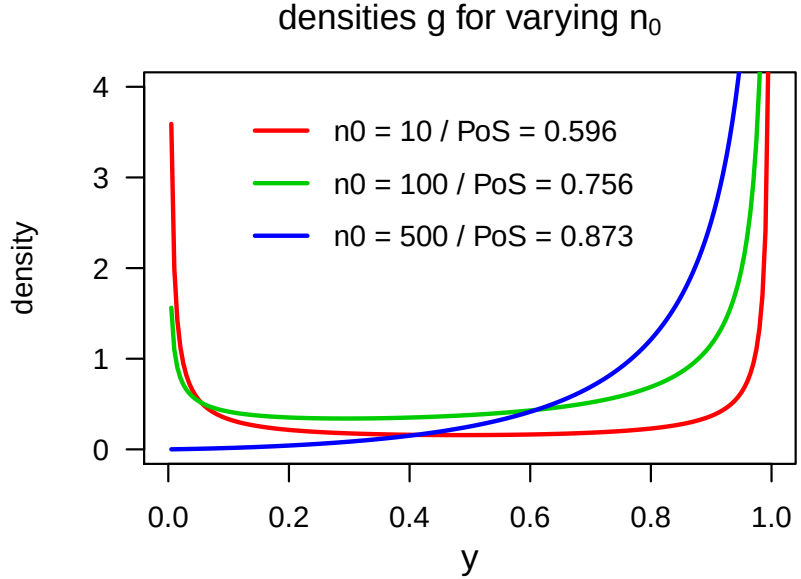
Is Normal prior with large variance "uninformative" - **no**!

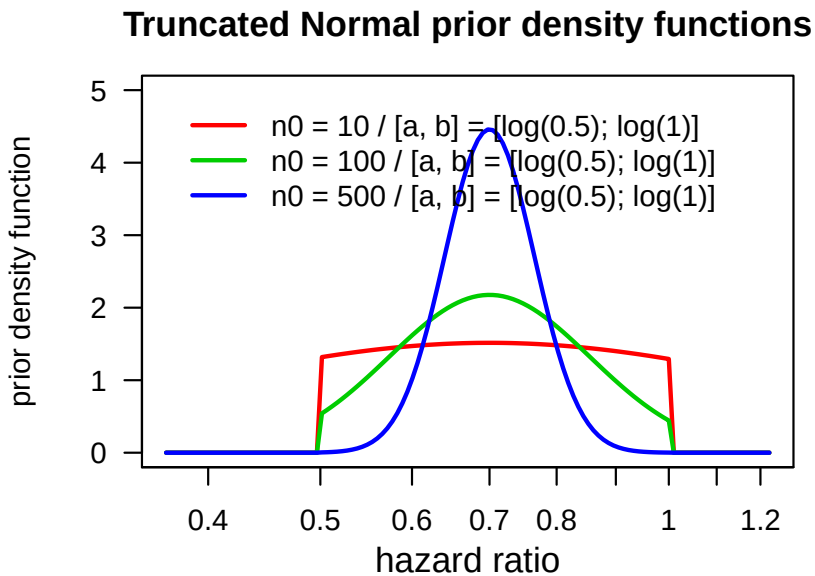
If $\sigma_0 > \sigma_{\text{final}} \Rightarrow \alpha < 1 \Rightarrow g$ is bathtub-shaped.

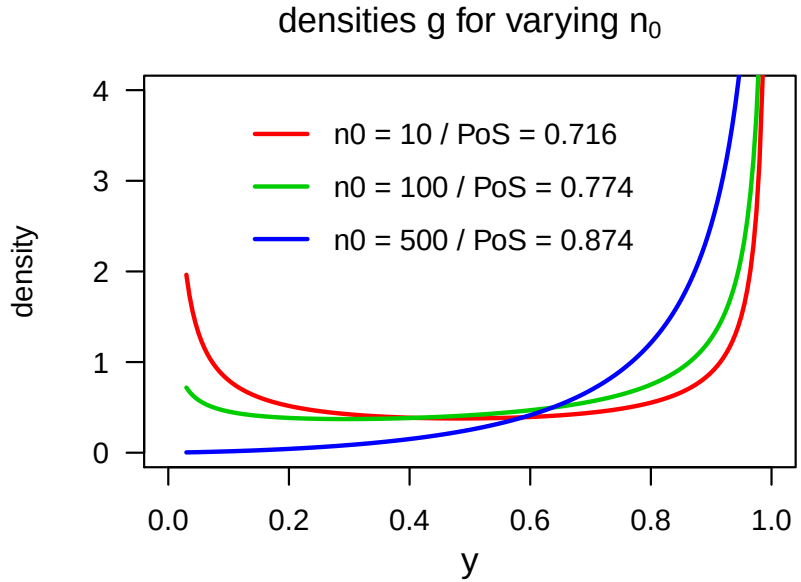
If $\sigma_0 \to \infty \Rightarrow \text{PoS} \to 0.5$: improper uniform prior $\Rightarrow$ basically symmetric around θ_{suc} $\Rightarrow$ get same distribution of power values left and right of $\theta_{\text{suc}} \Rightarrow$ average of them is 0.5.

Alternative prior 2: truncated Normal prior density







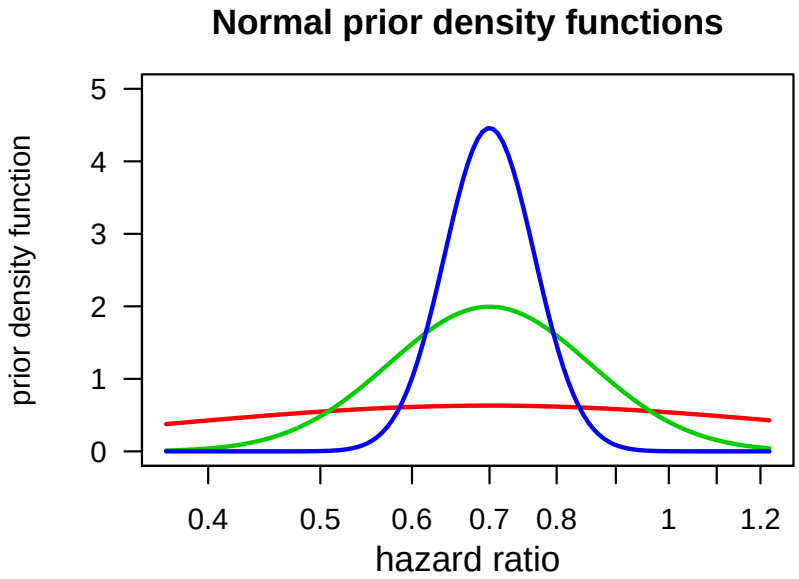


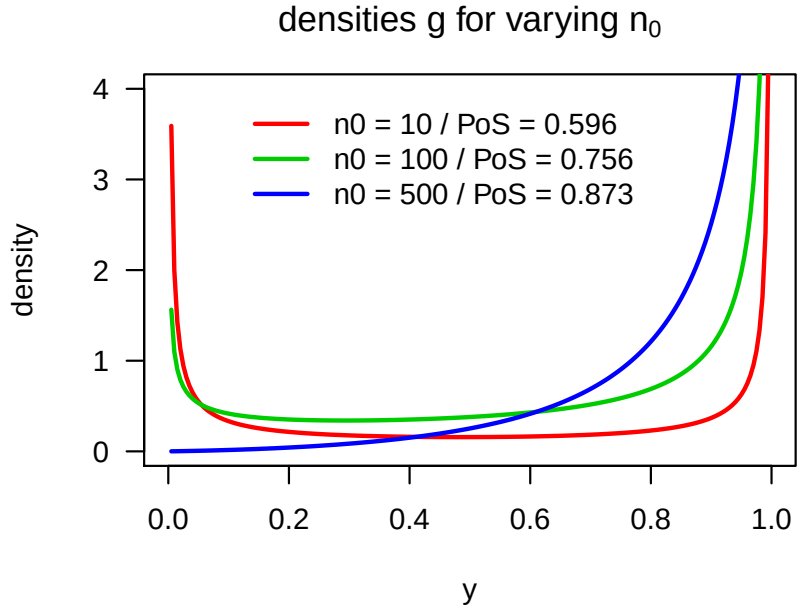
Alternative prior 2: truncated Normal prior density

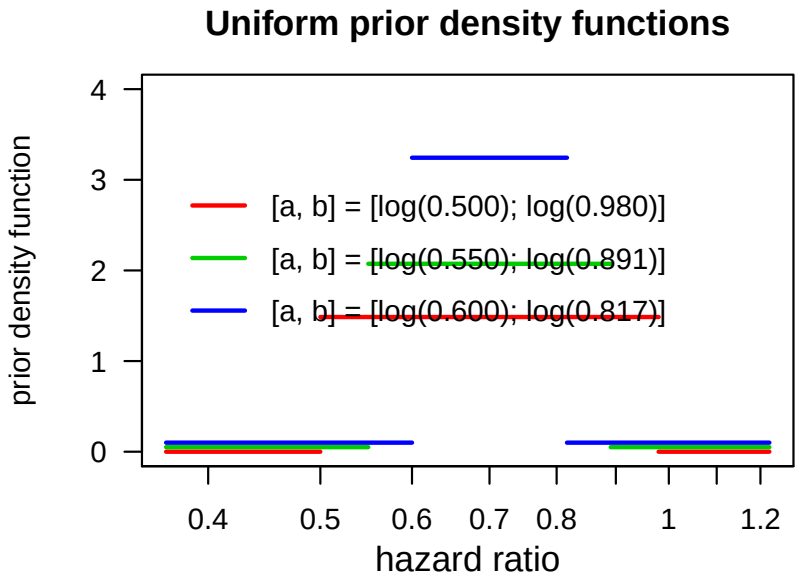
Observations:

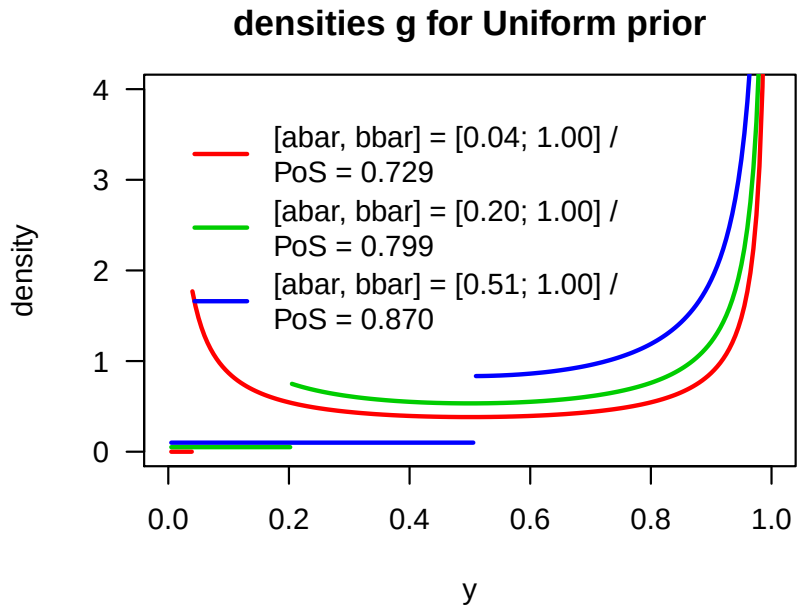
- Prior effect $< \theta_{\text{suc}} \Rightarrow$ the more weight to prior, the higher PoS.
- Truncation only useful if σ_0 not too small $\Rightarrow$ corresponds to low $n_0 \Rightarrow$ still get bathtub-shaped g.
- Prior symmetric but g not (unless $\theta_0 = \theta_{\text{suc}}$). Prior lives on log-scale!

Alternative prior 3: Uniform prior density









Alternative prior 3: Uniform prior density

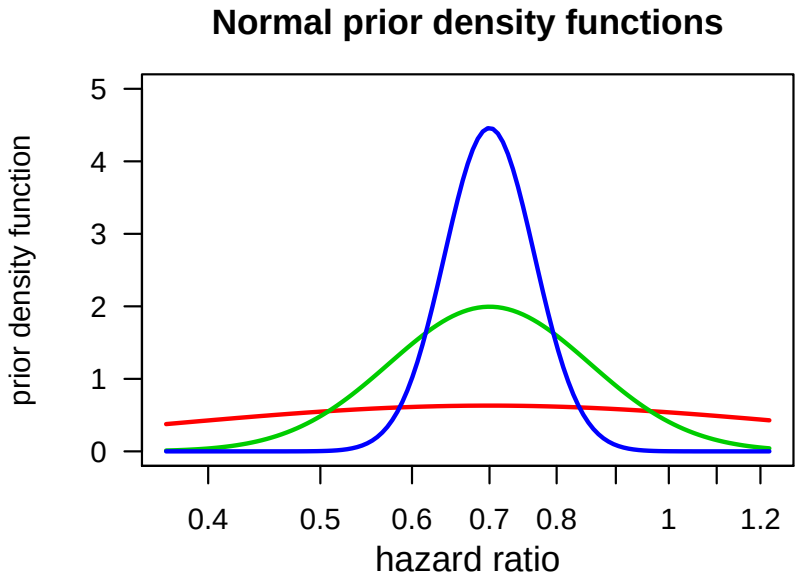
Uniform prior:

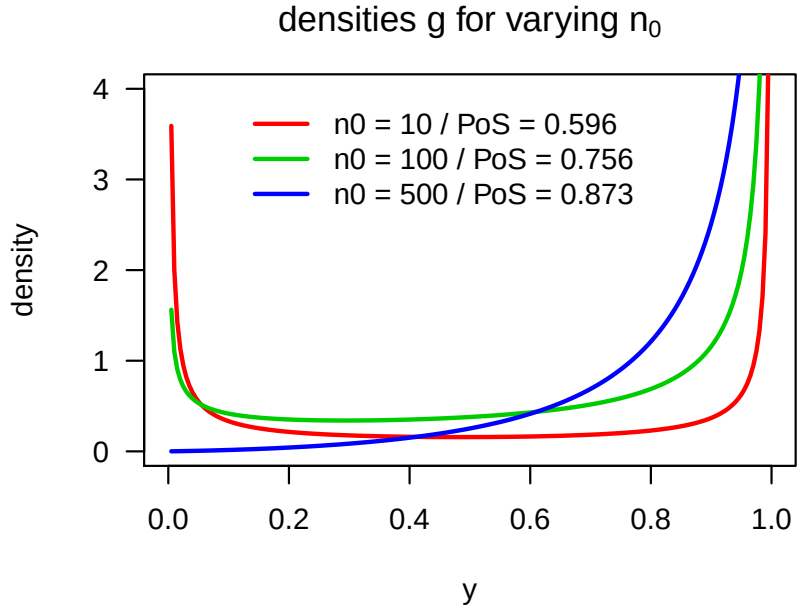
- Uniform priors centered at $\widehat{\theta}_{\text{Phase 2}} = \log(0.700)$.
- When updating with external knowledge or at interim: No matter what we observe $\Rightarrow$ posterior will only have mass in $[a, b]$.

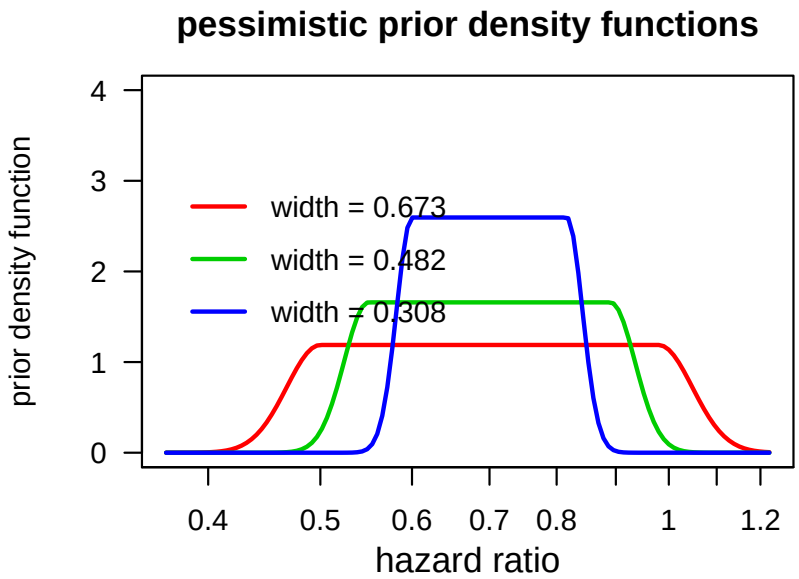
Observations:

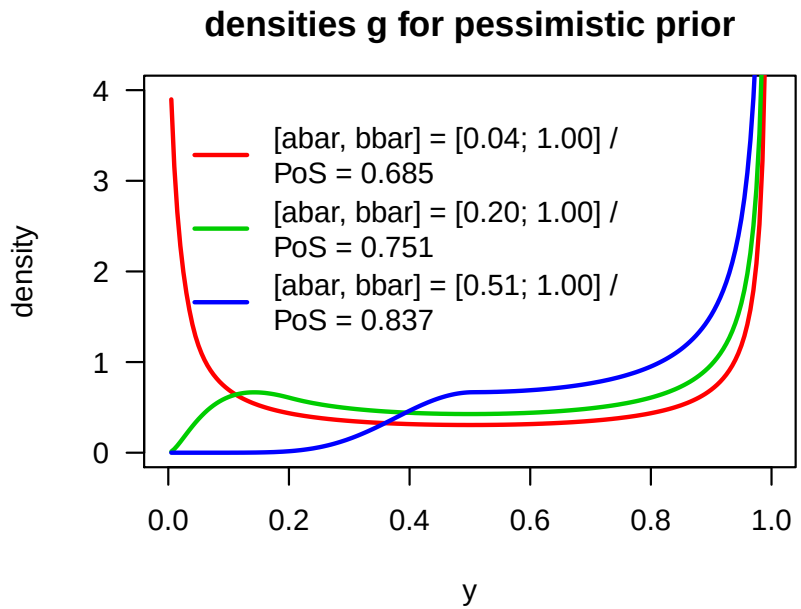
- g has **restricted support**.
- Restriction larger the more weight on Phase 2 (i.e. the larger σ_0^2).

Alternative prior 4: Uniform with Normal tails









Alternative prior 4: Uniform with Normal tails

Proposed in Rufibach et al. (2015) as "pessimistic prior".

Still have many low (only if prior is uncertain) and high power values.

Agenda

What is "uncertainty" for Bayesian Predictive Power?

What is BPP **not**?

- Not a population parameter to be estimated in a frequentist sense.
- Not a parameter with a prior distribution in a Bayesian framework

What is BPP?

- Average of a transformed effect size (the power function) with respect to a prior on that effect size.

What should "uncertainty" mean in this context? Rather **sensitivity** to assumptions!

If prior knowledge is based on 500 instead of 50 events $\Rightarrow$ should be reflected in BBP.

Sensitivity interval for BPP

Use quantiles of g to compute **sensitivity interval** corresponding to a level of γ (typically chosen to be 0.95):

$$[G^{-1}((1-\gamma)/2), G^{-1}((1+\gamma)/2)].$$

Generic example: increase number of events from 50 to 500 $\Rightarrow$ BBP increases from 0.697 to 0.873.

Why does BBP increase? We assumed that $\theta_0 < \theta_{\text{suc}}$.

Sensitivity intervals: $[0.000, 1.000]$ and $[0.424, 0.999]$.

Interpretation: 95% of power values lie in interval when θ drawn from prior p.

Intervals **not of much practical use**, do not restrict interval of plausible BBP values that are compatible with the prior.

Reason: if prior does not carry a lot of information on the underlying true effect $\Rightarrow$ bathtub-shaped g $\Rightarrow$ wide confidence interval.

Discussion

Observations:

- Density of power values often **bathtub-shaped**.
- **Does it make sense** to summarize this distribution in one number which we call BBP?
- Prior with large variance is **not** uninformative!

General questions/comments on bathtub phenomenon:

- How to quantify our prior belief about hazard ratio when we observe $\widehat{\theta}_{\text{Phase 2}} = \log(0.700)$ based on 50 events in Phase 2?
- Are we clear what properties PoS should have, as a function of all its input parameters? Dmitrienko and Wang (2006): choice of prior depends on
  - trial's objective,
  - development phase,
  - indication / patient population.

  Sensible/possible to come up with a **"one-size-fits-all"** concept?
- Include plot of g in discussion with teams!

Thank you for your attention.

References

- Berry, S. M., Carlin, B. P., Lee, J. J. and Müller, P. (2011). *Bayesian adaptive methods for clinical trials*, vol. 38 of *Chapman & Hall/CRC Biostatistics Series*. CRC Press, Boca Raton, FL. With a foreword by David J. Spiegelhalter.
- Dmitrienko, A. and Wang, M.-D. (2006). Bayesian predictive approach to interim monitoring in clinical trials. *Statistics in Medicine* **25** 2178–2195.
- O'Hagan, A., Stevens, J. W. and Campbell, M. J. (2005). Assurance in clinical trial design. *Pharmaceutical Statistics* **4** 187–201.
- Proschan, M., Lan, K. and Wittes, J. (2006). *Statistical Monitoring of Clinical Trials: A Unified Approach*. Springer, New York.
- Rufibach, K., Jordan, P. and Abt, M. (2015). Sequentially Updating the Likelihood of Success of a Phase 3 Pivotal Time-To-Event Trial based on Interim Analyses or External Information. *Journal of Biopharmaceutical Statistics*, to appear.
- Spiegelhalter, D., Reedman, L. and Blackburn, P. (1986). Monitoring clinical trials - conditional power or predictive power. *Control Clin Trials* **7** 8–17.

Backup slides.

Why P(extreme PoS values) so high?

Recall g:

$$g(y) \quad = \quad q(\theta_{\text{suc}} - \sigma_{\text{final}} z) \frac{\sigma_{\text{final}}}{\phi(z)}$$

for $z = \Phi^{-1}(y)$.

Why P(extreme PoS values) so high?

Recall g:

$$g(y) \quad = \quad q(\theta_{\text{suc}} - \sigma_{\text{final}} z) \frac{\sigma_{\text{final}}}{\phi(\mathbf{z})}$$

for $z = \Phi^{-1}(y)$.

Increase at both ends:

- $\phi(z)$ in denominator becomes small for $y \to 0, 1$.
- $\phi(z)$ in denominator derivative of Normal power function $\Rightarrow$ can only be **"nullified"** by choice of prior, but not removed.
- If σ_0 large $\Rightarrow$ high probability for very small / large hazard ratios $\Rightarrow$ power at these hazard ratios virtually 0 or 1.

Doing now what patients need next

R version and packages used to generate these slides:

R version: R version 3.1.1 (2014-07-10)

Base packages: stats / graphics / grDevices / utils / datasets / methods / base

Other packages: mvtnorm / reporttools / xtable / DDCP

This document was generated on 2015-10-27 at 13:37:12.