Including historical data in the analysis of clinical trials using the modified power priors: theoretical overview and sampling algorithms

Joost van Rosmalen[1], David Dejardin[2,3], and Emmanuel Lesaffre[1,2]

[1]Department of Biostatistics, Erasmus MC, Rotterdam, the Netherlands

[2]I-BioStat, KU Leuven, Leuven, Belgium

[3]Biometrics, Roche Pharmaceutical Research and Early Development, Basel, Switzerland

Bayes Pharma meeting 2014

Contents

- Introduction
- Theoretical overview
- Sampling algorithms
- Behaviour of modified power prior
- Discussion

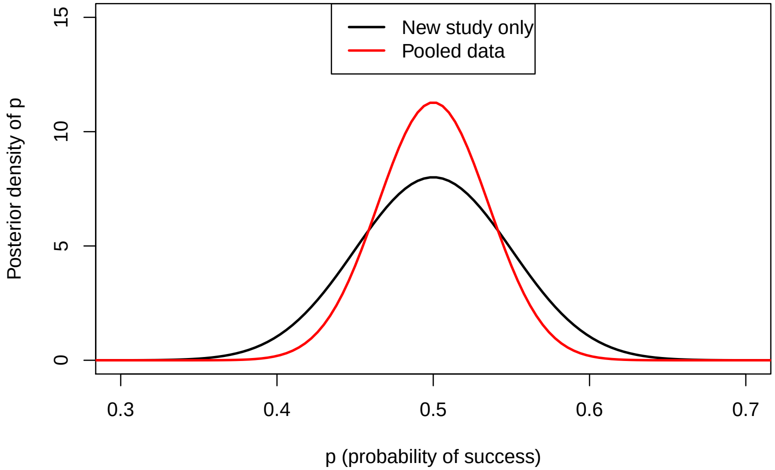
Introduction

- Including historical data in analysis of clinical studies is a tantalizing prospect.
- Advantages: improved precision and statistical power, fewer patients needed in new study.
- Common example: RCTs in which control group receives same treatment as in previous studies.
- Including historical data in analysis is controversial.

Introduction

- Hypothetical example: probability of success of new surgical technique
- Current study: 50 successes out of 100 cases
- Previous study: 50 successes out of 100 cases
- Binomial model, uniform prior for probability

Introduction



Introduction

- Alternative to pooling: include historical data in the analysis, but assign lower weight to historical data.
- Ibrahim & Chen (Stat. Science, 2000) proposed power prior:

$$\pi^{CPP}(\beta|H,\theta) \propto L_H(\beta)^{\theta} p(\beta),$$

  with $p(\beta)$ the prior density of model parameters, H the historical data, and θ (with $0 \leq \theta \leq 1$) the relative weight of the historical data.
- Referred to as conditional power prior (CPP)
- Combining with likelihood of new data set D yields posterior:

$$p^{CPP}(\beta|D,H,\theta) \propto L_D(\beta) L_H(\beta)^{\theta} p(\beta).$$

Theoretical overview

- Chen & Ibrahim (Bayesian Analysis, 2006): specific choices of θ in CPP equivalent to meta-analysis.
- Gaussian model for current $(y_1^D, \dots y_{n_D}^D)$ and hist. data $(y_1^H, \dots y_{n_H}^H)$

$$\begin{aligned} y_1^D, \dots y_n^D \sim N(\mu_D, \sigma_D^2) \quad &\text{and} \quad y_1^H, \dots y_{n_H}^H \sim N(\mu_H, \sigma_H^2) \\ \mu_D \sim N(\mu, \tau^2) \quad &\text{and} \quad \mu_H \sim N(\mu, \tau^2) \\ p(\mu) &\propto 1 \end{aligned}$$

- Result: distribution of μ_H given $y_1^D, \dots y_n^D$ and $y_1^H, \dots y_{n_H}^H$ equal to CPP with:

$$\theta = \frac{1}{2\tau^2 n_H / \sigma_H^2 + 1}$$

- Results can be extended to linear regression, GLMs.

Erasmus MC
University Medical Center Rotterdam

Theoretical overview

- Alternative approach: estimate θ using observed data.
- Result: adaptive pooling of historical and current data.
- Original power prior (Ibrahim & Chen, 2000):

$$\pi^{OPP}(\beta, \theta | D, H) \propto L_H(\beta)^{\theta} p(\beta) p(\theta),$$

with $p(\theta)$ prior distribution of weight parameter.

Erasmus MC
University Medical Center Rotterdam

Theoretical overview

- Duan et al. (Environmetrics, 2006): $\pi^{OPP}(\beta, \theta | D, H)$ does not satisfy likelihood principle.
- Multiplying $L_H(\beta)$ by arbitrary constant k yields

$$\pi^{OPP}(\beta, \theta | D, H) \propto k^{\theta} L_H(\beta)^{\theta} p(\beta) p(\theta),$$

  thus changing posterior distribution of θ.
- Original power prior gives low values of θ in large data sets, even if $L_D(\beta) = L_H(\beta)$.

Theoretical overview

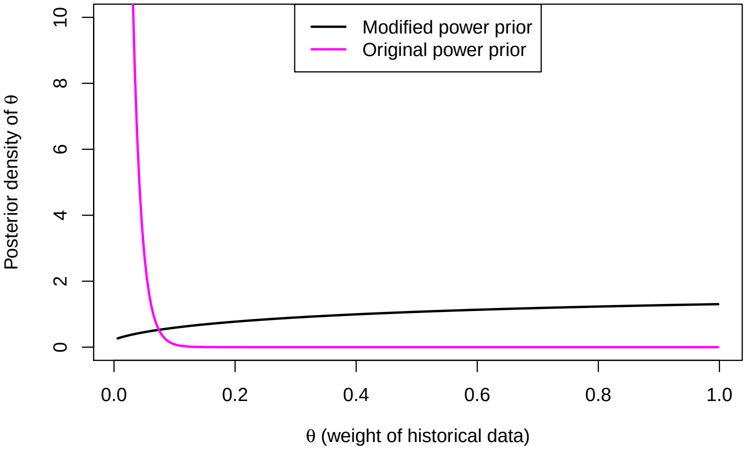
- Duan et al. (2006) proposed modified power prior (MPP):

$$\pi^{MPP}(\beta, \theta | H) \propto C(\theta) L_H(\beta)^{\theta} p(\beta) p(\theta),$$

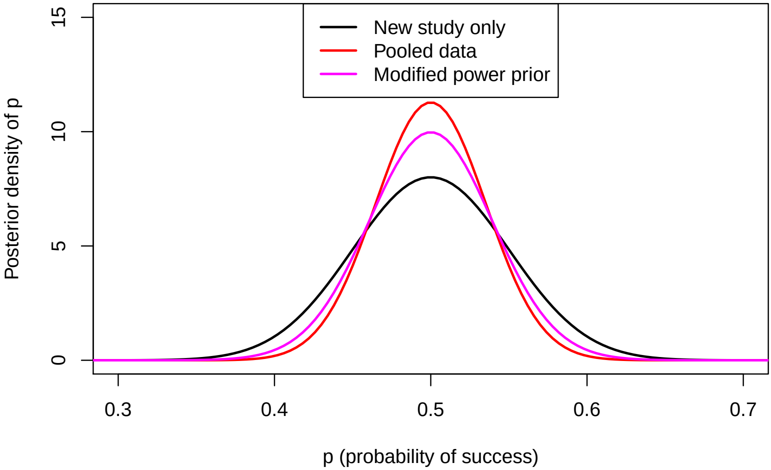
with

$$C(\theta) = \frac{1}{\int L_H(\beta)^{\theta} p(\beta)\, \mathrm{d}\beta}.$$

Theoretical overview



Theoretical overview



Erasmus MC
University Medical Center Rotterdam

Sampling algorithms

- Using MPP, prior involves integral $\int L_H(\beta)^\theta p(\beta)\,\mathrm{d}\beta$ in $C(\theta)$.
- In MCMC sampler, $C(\theta)$ must be available in every iteration: large computation time not acceptable.
- Possible approaches for simple models:
  - Analytical integration
  - Numerical integration
- MPP has only been applied in models with closed-form posterior.
- New algorithms for complex models:
  - Laplace approximation (talk of David)
  - Path sampling

Sampling algorithms: path sampling

- Friel and Pettitt (JRSS-B, 2008): calculation of marginal likelihood using power posteriors.
- Weight parameter used as auxiliary variable for calculating marginal likelihood (i.e. $\int L_H(\beta)^\theta p(\beta)\,\mathrm{d}\beta$ with $\theta = 1$).
- Idea of algorithm: logarithm of $C(\theta)$ equals integrated log-likelihood:

$$\log\left(\int L_H(\beta)^\theta p(\beta)\,\mathrm{d}\theta\right) = \int_{\theta^*=0}^{\theta} \mathbf{E}_{L_H(\beta)^{\theta^*} p(\beta)}\left[\log(L_H(\beta))\right]\,\mathrm{d}\theta^*,$$

with β distributed as $L_H(\beta)^{\theta^*} p(\beta)$.

Sampling algorithms: path sampling

Path sampling algorithm:

- Choose Δ_θ (step size for θ) and n_{iter} (iterations per step). Set $\theta = 0$.
- Repeat until $\theta = 1$:
  1. Increase θ by Δ_θ.
  2. Sample n_{iter} MCMC iterations from $L_H(\beta)^\theta p(\beta)$.
  3. Calculate $\overline{\log(L_H(\beta))}$ as the average of $\log(L_H(\beta))$ using βs sampled for the current value of θ
- Calculate cumulative sum of $\overline{\log(L_H(\beta))}$, as a function of θ.
- $\int L_H(\beta)^\theta p(\beta)\,\mathrm{d}\beta$ is proportional to exponential of the cumulative sum.

- $C(\theta)$ calculated for finite number of points in $[0, 1]$
- For intermediate values: use linear interpolation of $\log(C(\theta))$
- Few or no burn-in iterations required per value of θ.
- Numerical comparisons show that path sampling algorithm is accurate.
- For posterior results of MPP: use MCMC sampler and look up value of $C(\theta)$ in every iteration.

Behaviour of modified power prior

- If $L_H(\beta) = L_D(\beta)$ and $p(\theta) = 1$, posterior mode of θ is 1.
- But posterior CI of θ often very wide.
- Illustration with Gaussian model using results of Duan et al. (2006).

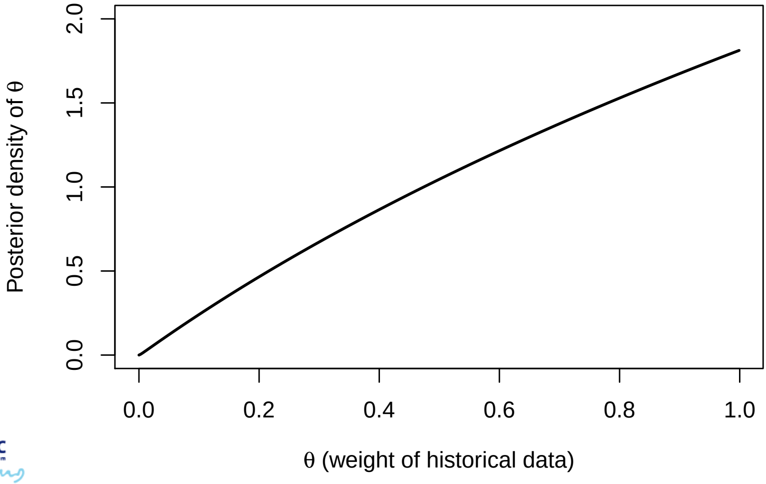
Behaviour of modified power prior

- Gaussian model: historical and current data $N(\mu, \sigma^2)$
- Beta(1,1) prior (i.e. uniform) for θ
- Jeffreys prior for μ and σ^2, $\pi(\mu, \sigma^2) = (1/\sigma^2)^{3/2}$
- $n = 100$, $\bar{x} = 0$, $s^2 = 1$ for current data
- Three settings for historical data:
  1. $n_0 = 100$, $\bar{x_H} = 0$, $s_0^2 = 1$
  2. $n_0 = 100$, $\bar{x_H} = 0.5$, $s_0^2 = 1$
  3. $n_0 = 100$, $\bar{x_H} \in [-4, 4]$, $s_0^2 = 1$

Erasmus MC
University Medical Center Rotterdam

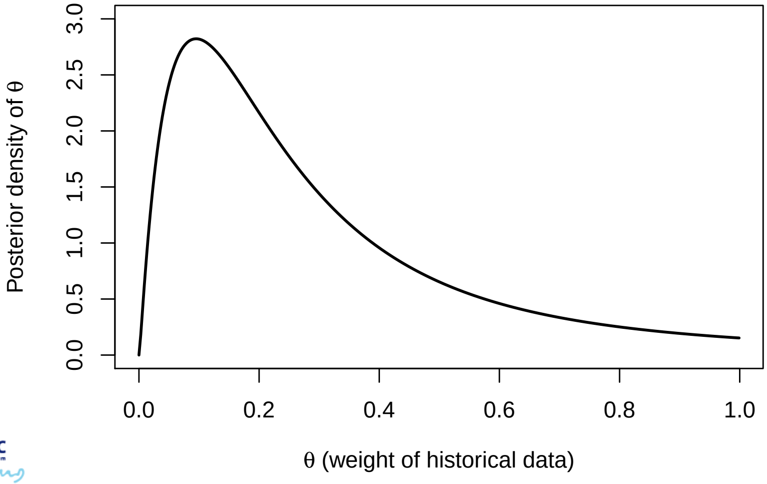
Behaviour of modified power prior

Posterior distribution of θ with equal historical and current data



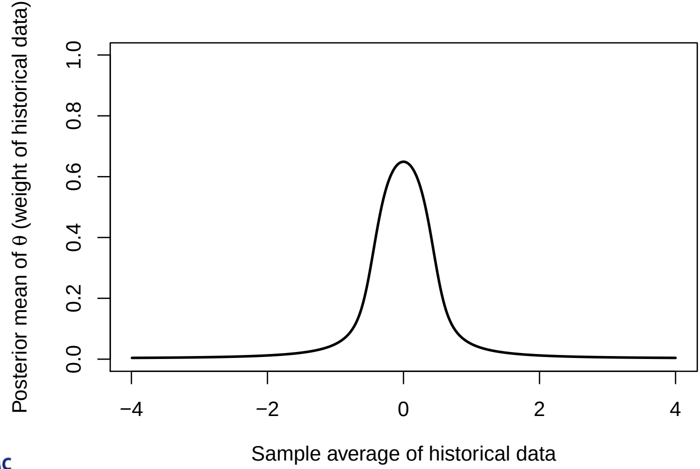
Behaviour of modified power prior

Posterior distribution of θ with $\bar{x} = 0$ and $\bar{x_H} = 0.5$



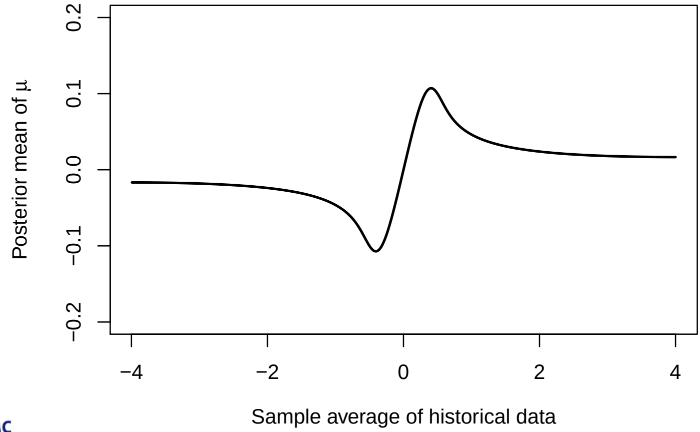
Behaviour of modified power prior

Posterior mean of θ as a function of $\bar{x}_H$



Behaviour of modified power prior

Posterior mean of μ as a function of $\bar{x_H}$



Behaviour of modified power prior

- Largest effect of historical data when $\bar{x} = 0$ and $\bar{x_H} \approx 0.41$, corresponding with significant difference in frequentist analysis (p-value two-sample t-test: 0.005).
- In Gaussian model, posterior mean of θ never becomes 1 using flat prior for θ.
- Scaling constant $\int_\beta L_H(\beta)^\theta p(\beta)\,\mathrm{d}\beta$ is not finite with reference prior $f(\mu, \sigma^2) = \frac{1}{\sigma^2}$ (Duan et al. 2006)

Discussion

- Modified power prior is an interesting and versatile method.
- Advantages:
  - Applicable to any parametric Bayesian model
  - Normality assumption for parameters not needed.
  - Adaptive pooling happens automatically.
- Limitations:
  - Any discrepancy between historical and current data leads to downweighting.
  - Amount of discounting often considered too severe (Neelon & O'Malley, J Biomet Biostat 2010).
  - Unclear how prior of θ must be chosen.
  - Implementation requires more work than standard Bayesian analysis.

Discussion

- Topics for future research:
  - Definition of modified power prior with multiple historical data sets.
  - Comparison of MPP with other methods, e.g. meta-analysis.
  - What to do if $\int_\beta L_H(\beta)^\theta p(\beta)\,\mathrm{d}\beta$ is not finite for all $\theta \in (0,1)$
  - Frequentist characteristics
- Theoretical understanding of MPP needs to be improved.