Wasserstein Consensus for Bayesian Sample Size Determination

Tullia Padellini

tulliapadellini.github.io
tullia.padellini@uniroma1.it

SAPIENZA
UNIVERSITÀ DI ROMA

Michele Cianfriglia
Pierpaolo Brutti

Dipartimento di Scienze Statistiche

Napoli – 9 July 2019

Sample Size Determination

the big picture

Pre-sperimental problem consisting of **choosing the size of the sample,** n, typically trying to minimize uncertainty under some cost constraint.

Sample Size Determination

the big picture

Pre-sperimental problem consisting of **choosing the size of the sample,** n, typically trying to minimize uncertainty under some cost constraint.

In **Clinical Trials** this translates into:

- **Cost** - every patient is "precious" for both ethics and finances
- **Uncertainty** - we cannot risk introducing a dangerous treatment

Sample Size Determination

the big picture

Pre-sperimental problem consisting of **choosing the size of the sample,** n, typically trying to minimize uncertainty under some cost constraint.

In **Clinical Trials** this translates into:

- **Cost** - every patient is "precious" for both ethics and finances
- **Uncertainty** - we cannot risk introducing a dangerous treatment

GOAL: find a sample size that induces agreement between different parties

The (bayesian) state of the art

the main ingredients

- Analysis Prior $\pi_A(\theta)$:

  models pre-experimental information to be used to obtain the **posterior distribution**

- Design Prior $\pi_D(\theta)$:

  models uncertainty on the experiment to be used to obtain the **predictive distribution**

The (bayesian) state of the art

the main ingredients

- Analysis Prior $\pi_A(\theta)$:

  models pre-experimental information to be used to obtain the **posterior distribution**

- Design Prior $\pi_D(\theta)$:

  models uncertainty on the experiment to be used to obtain the **predictive distribution**

Select n in order to satisfy some inferential goal, to be formalized in terms of a summary of the posteriors

$$\rho_{\pi_A}(\theta|y_n) = \int g(\theta)\pi_A(\theta|y_n)\mathrm{d}\theta$$

The (bayesian) state of the art

in the two prior approach

The design predictive distribution $m_D(y)$ removes the dependency of $\rho_{\pi_A}(\theta|y_n)$ from the observed sample y_n

The (bayesian) state of the art

in the two prior approach

The design predictive distribution $m_D(y)$ removes the dependency of $\rho_{\pi_A}(\theta|y_n)$ from the observed sample y_n

- **PEC** - Predictive Expectation Criterion

$$e(n) = \mathbb{E}_{m_D}[\rho_{\pi_A}(\theta|Y_n)] \qquad n^* = \min\{n \in \mathbb{N} : e(n) > \eta\}$$

The (bayesian) state of the art

in the two prior approach

The design predictive distribution $m_D(y)$ removes the dependency of $\rho_{\pi_A}(\theta|y_n)$ from the observed sample y_n

- **PEC** - Predictive Expectation Criterion

$$e(n) = \mathbb{E}_{m_D}[\rho_{\pi_A}(\theta|Y_n)] \qquad n^* = \min\{n \in \mathbb{N} : e(n) > \eta\}$$

- **PPC** - Predictive Probability Criterion

$$p(n) = \mathbb{P}_{m_D}[\rho_{\pi_A}(\theta|\mathbf{X}_n) > \gamma] \qquad n^* = \min\{n \in \mathbb{N} : p(n) > \eta\}$$

The (bayesian) state of the art

in the two prior approach

The design predictive distribution $m_D(y)$ removes the dependency of $\rho_{\pi_A}(\theta|y_n)$ from the observed sample y_n

- **PEC** - Predictive Expectation Criterion

$$e(n) = \mathbb{E}_{m_D}[\rho_{\pi_A}(\theta|Y_n)] \qquad n^* = \min\{n \in \mathbb{N} : e(n) > \eta\}$$

- **PPC** - Predictive Probability Criterion

$$p(n) = \mathbb{P}_{m_D}[\rho_{\pi_A}(\theta|\mathbf{X}_n) > \gamma] \qquad n^* = \min\{n \in \mathbb{N} : p(n) > \eta\}$$

η and γ are clinically relevant thresholds and depend on the problem.

Multiple priors

when should we look for "consensus"?

- diverging expert opinions
- multiple scenarios to take into account
- data from previous studies

Multiple priors

when should we look for "consensus"?

- diverging expert opinions
- multiple scenarios to take into account
- data from previous studies

Community of priors problem: how to combine multiple sources of pre-sperimental information into the analysis?

The standard Solution

mixtures of priors

Aggregate multiple priors into one and then use the approach of your likings.

The standard Solution

mixtures of priors

Aggregate multiple priors into one and then use the approach of your likings.

$$\pi_1(\theta), \ldots, \pi_K(\theta)$$

The standard Solution
mixtures of priors

Aggregate multiple priors into one and then use the approach of your likings.

$$\pi_1(\theta), \ldots, \pi_K(\theta) \longrightarrow \pi_A(\theta) = \sum_{i=1}^{K} \omega_{0,i} \pi_i(\theta)$$

The standard Solution

mixtures of priors

Aggregate multiple priors into one and then use the approach of your likings.

$$\pi_1(\theta), \ldots, \pi_K(\theta) \quad \longrightarrow \quad \pi_A(\theta) = \sum_{i=1}^{K} \omega_{0,i}\pi_i(\theta)$$

$$\pi_A(\theta|y_n) = \sum_{i=1}^{K} \frac{\omega_{0,i} m_i(y_n)}{\sum_{r=1}^{K} \omega_{0,r} m_r(y_n)} \times \pi_i(\theta|y_n)$$

The standard Solution
mixtures of priors

Aggregate multiple priors into one and then use the approach of your likings.

$$\pi_1(\theta), \ldots, \pi_K(\theta) \longrightarrow \pi_A(\theta) = \sum_{i=1}^{K} \omega_{0,i}\pi_i(\theta)$$

$$\pi_A(\theta|y_n) = \sum_{i=1}^{K} \frac{\omega_{0,i} m_i(y_n)}{\sum_{r=1}^{K} \omega_{0,r} m_r(y_n)} \times \pi_i(\theta|y_n)$$

Will the *i*-th clinician believe us?

Brutti, P., De Santis, F., & Gubbiotti, S. (2009). *Mixtures of prior distributions for predictive Bayesian sample size calculations in clinical trials.* Statistics in medicine

Our Solution

enforcing “consensus” between sources

- (possibly) conflicting priors π_1, π_2
- resulting posteriors $\pi_{1,y}, \pi_{2,y}$

Two experts “agree” if their inferential conclusions are the same

Our Solution

enforcing "consensus" between sources

- (possibly) conflicting priors π_1, π_2
- resulting posteriors $\pi_{1,y}, \pi_{2,y}$

Two experts "agree" if their inferential conclusions are the same, hence if their **posterior distributions are close enough**.

Our Solution

enforcing “consensus” between sources

- (possibly) conflicting priors π_1, π_2
- resulting posteriors $\pi_{1,y}, \pi_{2,y}$

Two experts “agree” if their inferential conclusions are the same, hence if their **posterior distributions are close enough**.

We formalize **agreement** or **consensus** in terms of distance between $\pi_{1,y}$ and $\pi_{2,y}$

Formally

how does this relate to the standard framework?

We can still adopt the Predictive approach, as this it's just another way of defining the summary statistic:

$$\rho_{\pi_A}(\theta|y_n)$$

Formally

how does this relate to the standard framework?

We can still adopt the Predictive approach, as this it's just another way of defining the summary statistic:

$$\rho_{\pi_A}(\theta|y_n) \quad \longrightarrow \quad d(\pi_{1,y}, \pi_{2,y})$$

Formally

how does this relate to the standard framework?

We can still adopt the Predictive approach, as this it's just another way of defining the summary statistic:

$$\rho_{\pi_A}(\theta|y_n) \quad \longrightarrow \quad d(\pi_{1,y}, \pi_{2,y})$$

- **PEC** - Predictive Expectation Criterion

$$e_{1,2}(n) = \mathbb{E}_{m_D}[d(\pi_{1,y}, \pi_{2,y})] \qquad\qquad n^* = \min\{n \in \mathbb{N} : e_{1,2}(n) < \eta\}$$

Formally

how does this relate to the standard framework?

We can still adopt the Predictive approach, as this it's just another way of defining the summary statistic:

$$\rho_{\pi_A}(\theta|y_n) \quad \longrightarrow \quad d(\pi_{1,y}, \pi_{2,y})$$

- **PEC** - Predictive Expectation Criterion

$$e_{1,2}(n) = \mathbb{E}_{m_D}[d(\pi_{1,y}, \pi_{2,y})] \qquad n^* = \min\{n \in \mathbb{N} : e_{1,2}(n) < \eta\}$$

- **PPC** - Predictive Probability Criterion

$$p_{1,2}(n) = \mathbb{P}_{m_D}[d(\pi_{1,y}, \pi_{2,y}) > \gamma] \qquad n^* = \min\{n \in \mathbb{N} : p_{1,2}(n) < \eta\}$$

Formally

how does this relate to the standard framework?

We can still adopt the Predictive approach, as this it's just another way of defining the summary statistic:

$$\rho_{\pi_A}(\theta|y_n) \longrightarrow d(\pi_{1,y}, \pi_{2,y})$$

- **PEC** - Predictive Expectation Criterion

$$e_{1,2}(n) = \mathbb{E}_{m_D}[d(\pi_{1,y}, \pi_{2,y})] \qquad n^* = \min\{n \in \mathbb{N} : e_{1,2}(n) < \eta\}$$

- **PPC** - Predictive Probability Criterion

$$p_{1,2}(n) = \mathbb{P}_{m_D}[d(\pi_{1,y}, \pi_{2,y}) > \gamma] \qquad n^* = \min\{n \in \mathbb{N} : p_{1,2}(n) < \eta\}$$

we just need to pick a distance

Wasserstein distance

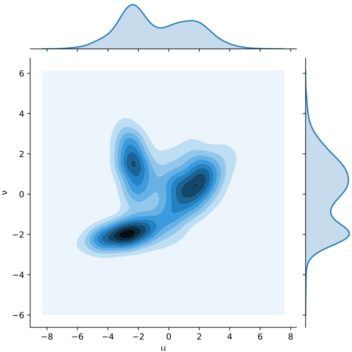
a.k.a. Kantorovic, Earth Mover

$(p, d)-$ **Wasserstein distance**

$X \sim P$ and $Y \sim Q$, $p \geqslant 1$ and d ground distance

$$W_{d,p}(P, Q) = \left(\inf_J \int_{\mathcal{X} \times \mathcal{Y}} d(x, y)^p \; \mathrm{d}J(x, y) \right)^{1/p}$$

where the infimum is over **all joint** distributions J having P and Q as marginals.



Wasserstein distance

it's this popular for a good reason

- it "metricises" convergence in distribution

    if two distributions are close w.r.t. the Wasserstein distance they are probabilistically similar.

Verdinelli, I., Wasserman, L. (2018) Hybrid Wasserstein Distance and Fast Distribution Clustering. *arXiv preprint arXiv:1812.11026*

Wasserstein distance

it's this popular for a good reason

- it "metricises" convergence in distribution

  if two distributions are close w.r.t. the Wasserstein distance they are probabilistically similar.

- it tells us why the distributions differ

  with the Wasserstein distance is associated to a map (transport plan) that shows us how we have to move the mass of P to morph it into Q.

Verdinelli, I., Wasserman, L. (2018) Hybrid Wasserstein Distance and Fast Distribution Clustering. *arXiv preprint arXiv:1812.11026*

Wasserstein distance

it's this popular for a good reason

- it "metricises" convergence in distribution

  if two distributions are close w.r.t. the Wasserstein distance they are probabilistically similar.

- it tells us why the distributions differ

  with the Wasserstein distance is associated to a map (transport plan) that shows us how we have to move the mass of P to morph it into Q.

- it is sensible to the geometry of the space

  it's not just about the location!

Verdinelli, I., Wasserman, L. (2018) Hybrid Wasserstein Distance and Fast Distribution Clustering. *arXiv preprint arXiv:1812.11026*

Multivariate Gaussian distributions

computing the Wasserstein distance

Let $X \sim \mathrm{N}(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ and $Y \sim \mathrm{N}(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$, when **the ground distance is taken to be the L_2 distance**, we have a closed form expression for Wasserstein:

$$\mathrm{W}_{L^2,2}(X, Y) = \|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|_2^2 + \mathtt{B}^2(\boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Y)$$

Multivariate Gaussian distributions

computing the Wasserstein distance

Let $X \sim \mathrm{N}(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ and $Y \sim \mathrm{N}(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$, when **the ground distance is taken to be the L_2 distance**, we have a closed form expression for Wasserstein:

$$\mathrm{W}_{L^2,2}(X, Y) = \|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|_2^2 + \mathrm{B}^2(\boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Y)$$

$\mathrm{B}^2(\boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Y) = \mathrm{tr}\left[\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_Y - 2\sqrt{\boldsymbol{\Sigma}_X^{1/2}\boldsymbol{\Sigma}_Y\boldsymbol{\Sigma}_X^{1/2}}\right]$ is the **Bures distance**.

Multivariate Gaussian distributions

computing the Wasserstein distance

Let $X \sim \mathrm{N}(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ and $Y \sim \mathrm{N}(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$, when **the ground distance is taken to be the L_2 distance**, we have a closed form expression for Wasserstein:

$$\mathrm{W}_{L^2,2}(X, Y) = \|\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y\|_2^2 + \mathrm{B}^2(\boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Y)$$

$\mathrm{B}^2(\boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Y) = \mathrm{tr}\left[\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_Y - 2\sqrt{\boldsymbol{\Sigma}_X^{1/2}\boldsymbol{\Sigma}_Y\boldsymbol{\Sigma}_X^{1/2}}\right]$ is the **Bures distance**.

distance between the means **+** distance between the variances

Conjugate Univariate Gaussian Model

computing the Wasserstein distance

Likelihood: $\mathsf{N}(\theta, \sigma^2)$, with σ^2 known.

$$\pi(\theta) = \mathsf{N}\left(\theta; \mu_0, \frac{\sigma^2}{n_0}\right)$$

Conjugate Univariate Gaussian Model

computing the Wasserstein distance

Likelihood: $\mathrm{N}(\theta, \sigma^2)$, with σ^2 known.

$$\pi(\theta) = \mathrm{N}\left(\theta; \mu_0, \frac{\sigma^2}{n_0}\right) \qquad \pi(\theta | y_n) = \mathrm{N}\left(\theta; \frac{n_0\mu_0 + n\bar{y}_n}{n + n_0}, \frac{\sigma^2}{n + n_0}\right)$$

Conjugate Univariate Gaussian Model

computing the Wasserstein distance

Likelihood: $\mathrm{N}(\theta, \sigma^2)$, with σ^2 known.

$$\pi(\theta) = \mathrm{N}\left(\theta; \mu_0, \frac{\sigma^2}{n_0}\right) \qquad \pi(\theta|y_n) = \mathrm{N}\left(\theta; \frac{n_0\mu_0 + n\bar{y}_n}{n + n_0}, \frac{\sigma^2}{n + n_0}\right)$$

If we have two priors, Wasserstein between the corresponding posteriors is:

$$\mathrm{W}_{L^2,2}(\pi_{1,y}, \pi_{2,y}) = (\mu_{1,\mathrm{P}} - \mu_{2,\mathrm{P}})^2 + (\sigma_{1,\mathrm{P}} - \sigma_{2,\mathrm{P}})^2 .$$

Conjugate Gaussian Model

in the Bayesian predictive approach to SSD

Under the usual $\pi_D(\theta) \sim \mathrm{N}(\mu_D, \sigma/n_D)$ assumption:

Conjugate Gaussian Model

in the Bayesian predictive approach to SSD

Under the usual $\pi_D(\theta) \sim \mathrm{N}(\mu_D, \sigma/n_D)$ assumption:

PEC: $$e_{1,2}(n) = \widetilde{\mu}^2 + \sigma^2 \left(w_n^2 \left[\frac{1}{n} + \frac{1}{n_D} \right] + \left[\frac{1}{\sqrt{n+n_1}} - \frac{1}{\sqrt{n+n_2}} \right]^2 \right)$$

Conjugate Gaussian Model

in the Bayesian predictive approach to SSD

Under the usual $\pi_D(\theta) \sim \mathrm{N}(\mu_D, \sigma/n_D)$ assumption:

PEC: $$e_{1,2}(n) = \widetilde{\mu}^2 + \sigma^2 \left(w_n^2 \left[\frac{1}{n} + \frac{1}{n_D} \right] + \left[\frac{1}{\sqrt{n+n_1}} - \frac{1}{\sqrt{n+n_2}} \right]^2 \right)$$

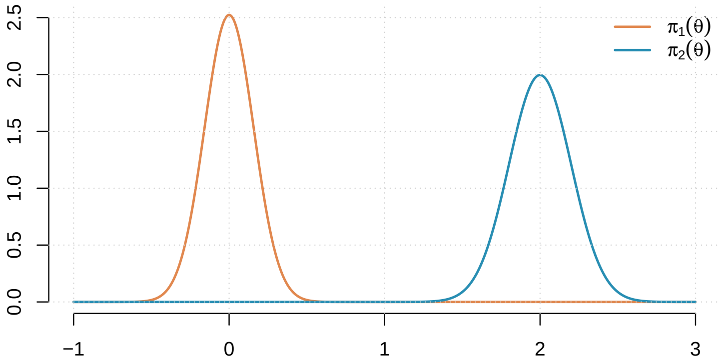
PPC: $$p_{1,2}(n) = 1 - F_{\chi^2} \left(\frac{\gamma - \mathrm{B}^2_{\sigma^2}}{\widetilde{\sigma}^2}; df = 1, \widetilde{\mu}^2 \right)$$

- $w_1 = n_1/(n+n_1)$
- $w_2 = n_2/(n+n_2)$
- $w_n = (1-w_1) - (1-w_2)$
- $\widetilde{\mu} = w_1\mu_1 - w_2\mu_2 + w_n\mu_D$
- $\widetilde{\sigma}^2 = w_n^2\sigma^2(1/n + 1/n_{\mathrm{D}})$
- $\mathrm{B}^2_{\sigma^2} = (\sigma_{1,\mathrm{P}} - \sigma_{1,\mathrm{P}})^2$

A Toy Example

mildly informative priors

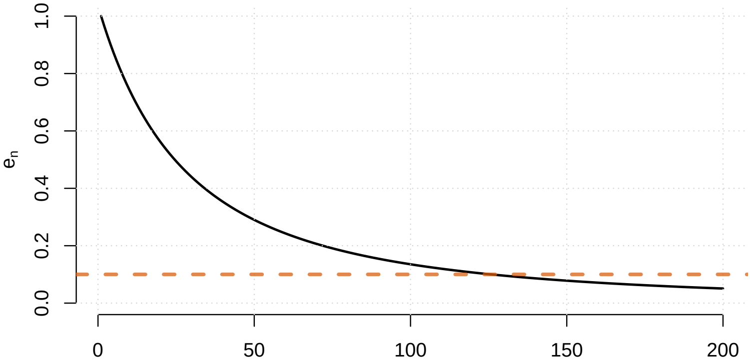
$$\pi_1(\theta) = \mathrm{N}(0, 2/80) \qquad \pi_2(\theta) = \mathrm{N}(2, 2/50)$$



A Toy Example

mildly informative priors

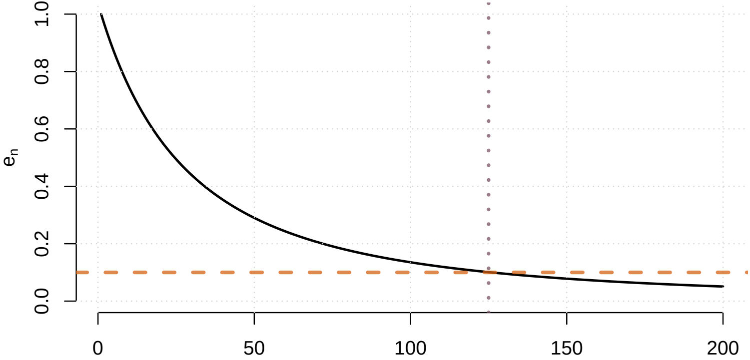
$\eta = 0.1$ $\quad$ $n^* = 125$



A Toy Example

mildly informative priors

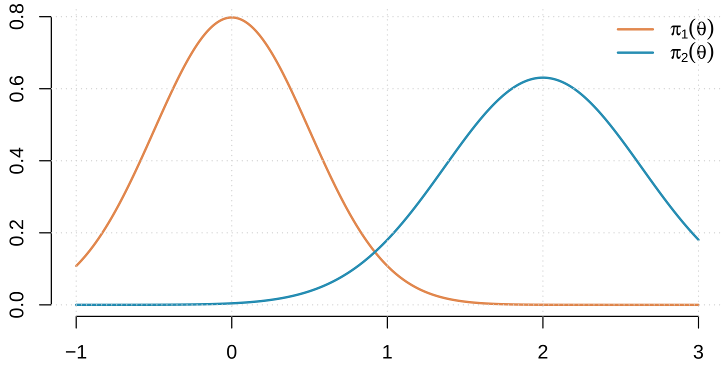
$\eta = 0.1$ $n^* = 125$



Another Toy Example

weakly informative priors

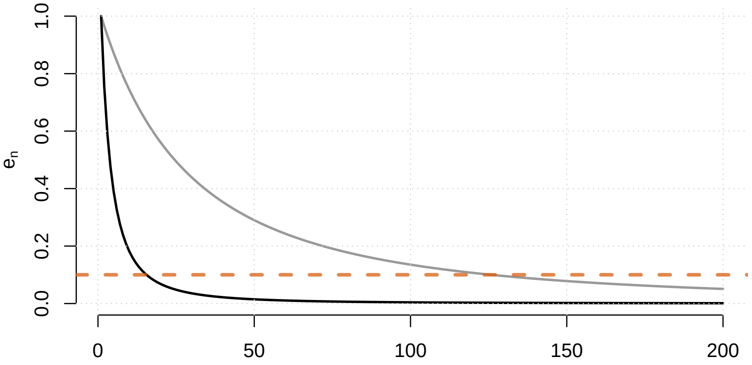
$$\pi_1(\theta) = \mathrm{N}(0, 2/8) \qquad \pi_2(\theta) = \mathrm{N}(2, 2/5)$$



Another Toy Example

weakly informative priors

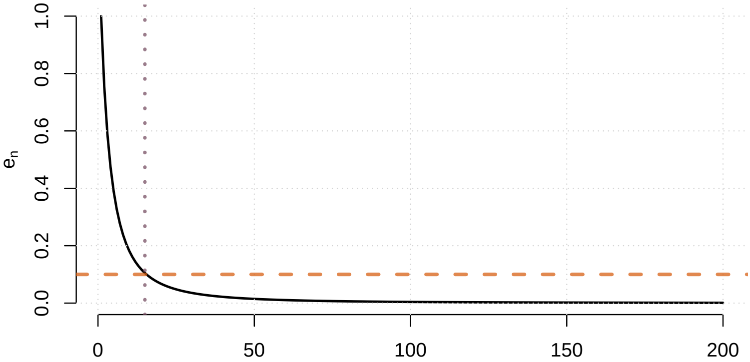
$\eta = 0.1$ $n^* = 15$



Another Toy Example

weakly informative priors

$\eta = 0.1$ $\quad\quad n^* = 15$



How to select η?

a small bump in the road

Given a $\beta \in (0, 1)$, choose η as

$$\beta \times \arg\max_n e_{1,2}(n)$$

How to select η?

a small bump in the road

Given a $\beta \in (0, 1)$, choose η as

$$\beta \times \arg\max_n e_{1,2}(n)$$

It turns out that under some regularity assumptions, $e_{1,2}(n)$ **can be monotone in** n.

How to select η?

a small bump in the road

Given a $\beta \in (0, 1)$, choose η as

$$\beta \times \arg\max_n e_{1,2}(n)$$

It turns out that under some regularity assumptions, $e_{1,2}(n)$ **can be monotone in** n.

When this happens

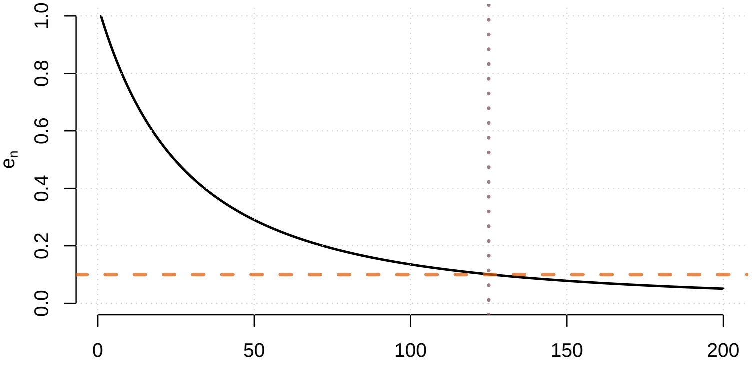
$$\arg\max_n e_{1,2}(n) = e_{1,2}(1)$$

β represent how much difference we can tolerate with respect to the minimum sample size possible.

A Toy Example

Reprise

$\eta = 0.1$ $n^* = 125$



A Real Data Example

from Spiegelhlter et al. (2004)

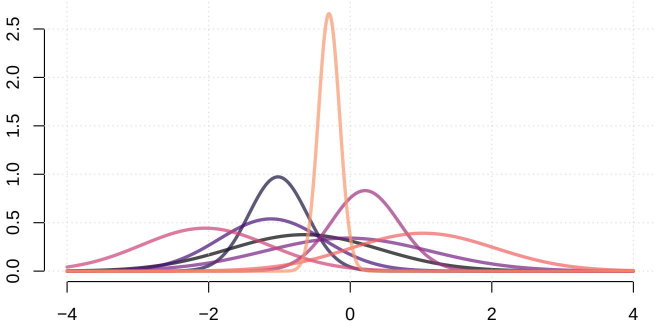
$\theta = \log$ OR of intravenous magnesium sulphate after acute myocardial infarction with respect to placebo.

A Real Data Example

from Spiegelhlter et al. (2004)

$\theta = \log \mathrm{OR}$ of intravenous magnesium sulphate after acute myocardial infarction with respect to placebo.

A bunch of priors encoding evidence from previous experiments:



An Unfair Comparison

was this really necessary?

- **Likelihood** Gaussian with unknown mean θ and $\sigma^2 = 4$
- **Design Prior** Gaussian with mean $\mu_D = 0.058$ and variance σ^2/n_D
- **Threshold** $\eta = 0.05$

An Unfair Comparison

was this really necessary?

- **Likelihood** Gaussian with unknown mean θ and $\sigma^2 = 4$
- **Design Prior** Gaussian with mean $\mu_D = 0.058$ and variance σ^2/n_D
- **Threshold** $\eta = 0.05$

$\boldsymbol{n_D}$	$\boldsymbol{n^*_{\texttt{WASS}}}$
4319	361
432	371
43	468

An Unfair Comparison

was this really necessary?

- **Likelihood** Gaussian with unknown mean θ and $\sigma^2 = 4$
- **Design Prior** Gaussian with mean $\mu_D = 0.058$ and variance σ^2 / n_D
- **Threshold** $\eta = 0.05$

n_D	$n^*_{\texttt{WASS}}$	$n^*_{\texttt{MIXT}}$
4319	361	498
432	371	509
43	468	190

Consensus does not typically come "for free"

Brutti, P., De Santis, F., & Gubbiotti, S. (2009). *Mixtures of prior distributions for predictive Bayesian sample size calculations in clinical trials.* Statistics in medicine

Conjugate Beta-Binomial

moving beyond gaussianity

When the posterior distributions are not Gaussian, the Wasserstein distance does not necessarily have an analyitic expression.

This is the case for the Beta-Binomial conjugate model.

Conjugate Beta-Binomial

moving beyond gaussianity

When the posterior distributions are not Gaussian, the Wasserstein distance does not necessarily have an analyitic expression.

This is the case for the Beta-Binomial conjugate model.

Possible solutions are:

- Numerical evaluation of the Wasserstein distance
- Approximation of the Wasserstein distance via Stein's method

Stein’s method

quickest introduction ever

X, Y random variables (typically X is "what you have", Y is "what you want")

1. rewrite the distance between X and Y as the **expectation** of a functional $h(X)$

Stein's method

quickest introduction ever

X, Y random variables (typically X is "what you have", Y is "what you want")

1. rewrite the distance between X and Y as the **expectation** of a functional $h(X)$

Stein's method

quickest introduction ever

X, Y random variables (typically X is "what you have", Y is "what you want")

1. rewrite the distance between X and Y as the **expectation** of a functional $h(X)$
2. **bound** such expectation

Stein's method
quickest introduction ever

X, Y random variables (typically X is "what you have", Y is "what you want")

1. rewrite the distance between X and Y as the **expectation** of a functional $h(X)$
2. **bound** such expectation

Stein's method

quickest introduction ever

X, Y random variables (typically X is "what you have", Y is "what you want")

1. rewrite the distance between X and Y as the **expectation** of a functional $h(X)$
2. **bound** such expectation

If we compare X and Y via the L_1 Wasserstein distance, we can derive tight bounds for it.

Ley, C., Reinert, G., & Swan, Y. (2017). Stein's method for comparison of univariate distributions. *Probability Surveys*.

Stein's bound for the B-B case

the second most famous framework in clinical trials

Likelihood: Binomial(θ, N), T events in the sample.

$$\pi(\theta) = \text{Beta}\,(\theta; \alpha, \beta)$$

Stein's bound for the B-B case

the second most famous framework in clinical trials

Likelihood: Binomial(θ, N), T events in the sample.

$$\pi(\theta) = \text{Beta}\,(\theta; \alpha, \beta) \qquad\qquad \pi(\theta|y_n) = \text{Beta}\,(\theta; \alpha + t, \beta + n - t)$$

Stein's bound for the B-B case

the second most famous framework in clinical trials

Likelihood: Binomial(θ, N), T events in the sample.

$$\pi(\theta) = \text{Beta}\left(\theta; \alpha, \beta\right) \qquad \pi(\theta|y_n) = \text{Beta}\left(\theta; \alpha + t, \beta + n - t\right)$$

$$d_W(\pi_{1,y}, \pi_{2,y}) \leq \frac{|\alpha_1 - \alpha_2|}{\alpha_1 + \beta_1 + n}(1 - \mu_{2,P}) + \frac{|\beta_2 - \beta_1|}{\alpha_1 + \beta_1 + n}\mu_{2,P}$$

Stein's bound for the B-B case

the second most famous framework in clinical trials

Likelihood: Binomial(θ, N), T events in the sample.

$$\pi(\theta) = \text{Beta}\left(\theta; \alpha, \beta\right) \qquad \pi(\theta|y_n) = \text{Beta}\left(\theta; \alpha + t, \beta + n - t\right)$$

$$d_W(\pi_{1,y}, \pi_{2,y}) \leq \frac{|\alpha_1 - \alpha_2|}{\alpha_1 + \beta_1 + n}(1 - \mu_{2,P}) + \frac{|\beta_2 - \beta_1|}{\alpha_1 + \beta_1 + n}\mu_{2,P}$$

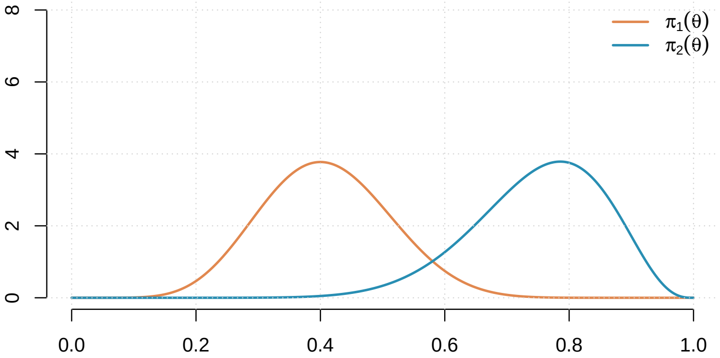
If we assume $\pi_D(\theta) = \text{Beta}(\theta; \alpha_D, \beta_D)$ it is possible to bound the **PEC** and **PPC** just by remembering:

$$\mathbb{E}_{m_D}[T] = \frac{n\alpha_D}{\alpha_D + \beta_D}$$

Yet Another Toy Example

the more the merrier

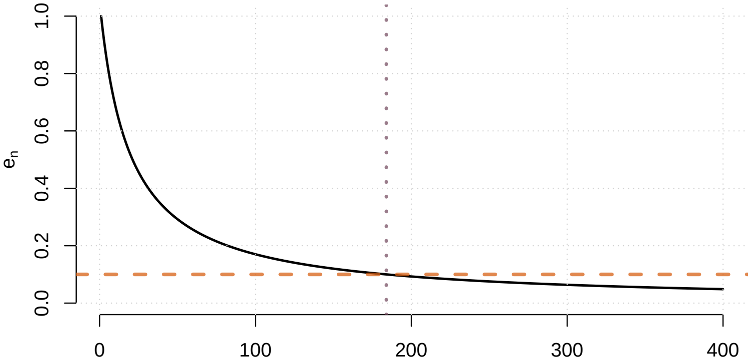
$$\pi_1(\theta) = \text{Beta}(9, 13) \qquad \pi_2(\theta) = \text{Beta}(12, 4)$$



Yet Another Toy Example

the more the merrier

$\eta = 0.1$ $\quad\quad$ $n^* = 184$



R-package coming soon!

SAPIENZA
UNIVERSITÀ DI ROMA

Thanks!

tulliapadellini.github.io
tullia.padellini@uniroma1.it