SAPIENZA
UNIVERSITÀ DI ROMA

Bayesian Sample size for calibration of approximate intervals in clinical trials

Fulvio De Santis & Stefania Gubbiotti

Dipartimento di Scienze Statistiche
Sapienza Università di Roma

Milano, 11 ottobre 2019

Outline

- Interval estimation of θ
- C : exact Highest Posterior Density (**HPD**) interval
- $\widetilde{C}$: Likelihood Normal Approximate (**LNA**) interval

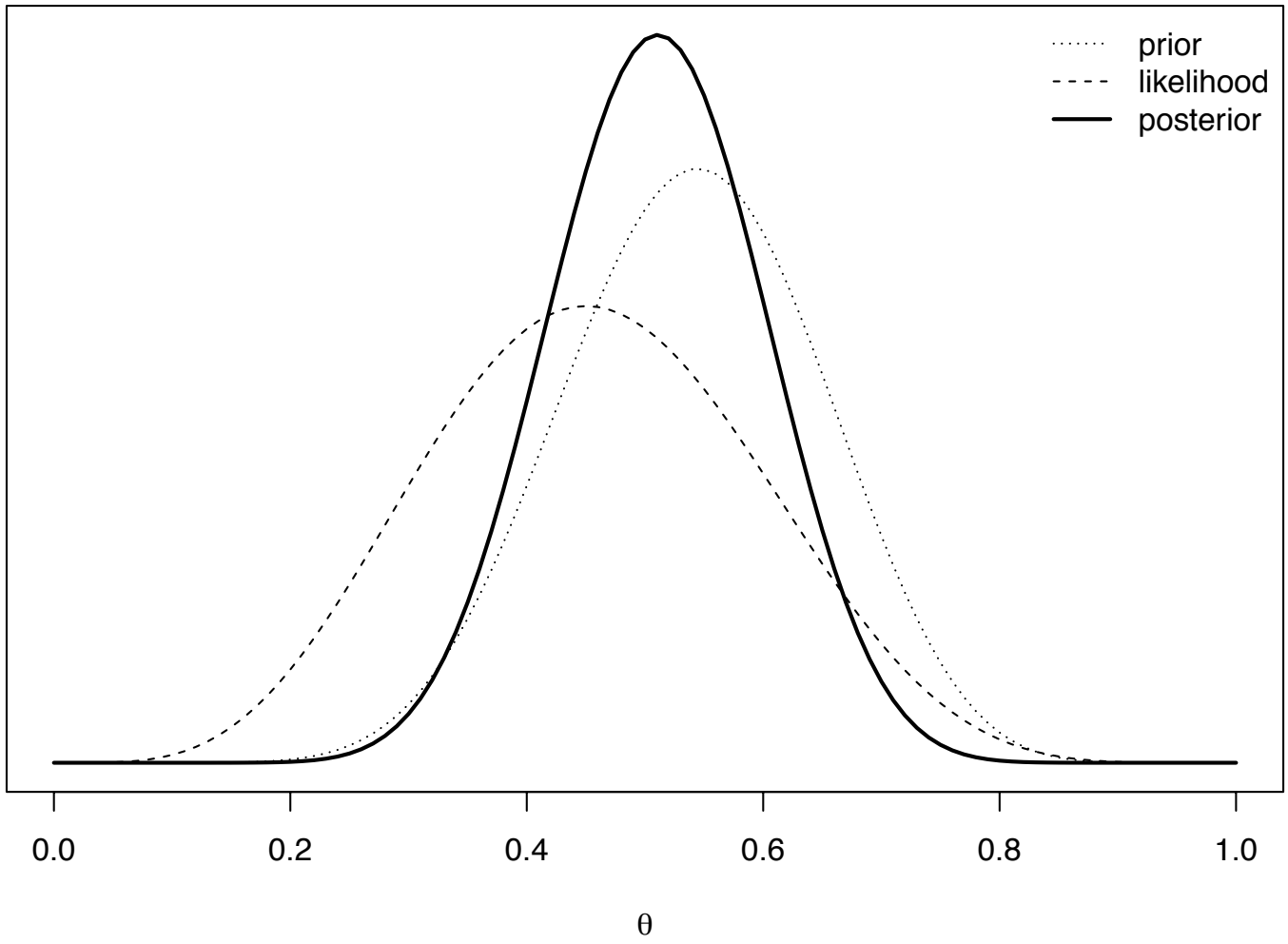
Problem: what n such that $\widetilde{C} \simeq C$?

Model

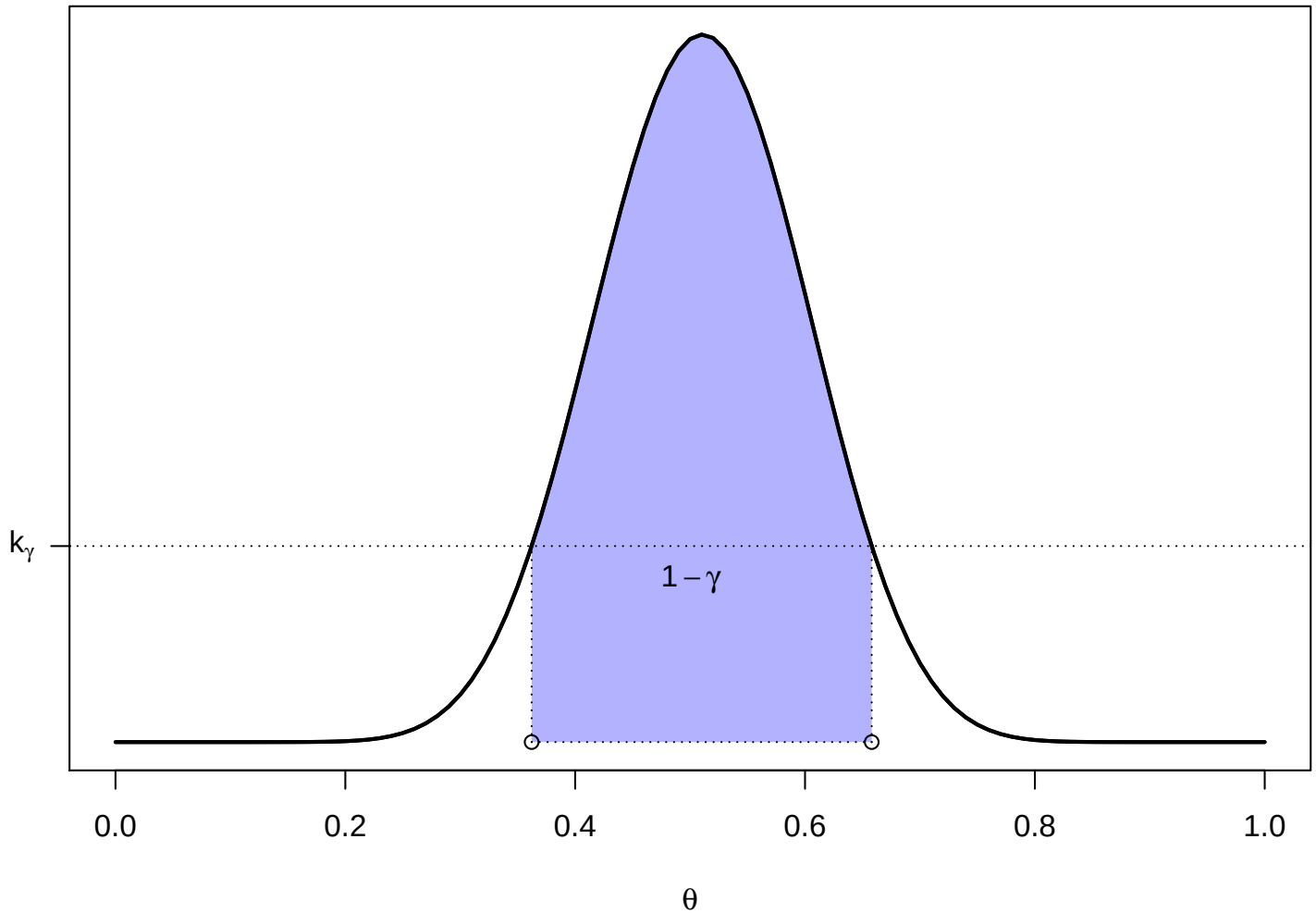
Beta-Binomial model

$$\left.\begin{array}{l} \boldsymbol{X_n} = (X_1, \ldots, X_n) \\ X_i|\theta \sim \text{Ber}(\theta) \\ \theta \sim \text{Be}(\alpha, \beta) \end{array}\right\} \quad \Longrightarrow \quad \theta|\boldsymbol{x_n} \sim \text{Be}(\bar{\alpha}, \bar{\beta})$$

Posterior distribution



Exact credible intervals for θ: HPD



Exact credible intervals for θ

HPD: $C(\boldsymbol{x_n}) = [\ell(\boldsymbol{x_n}), u(\boldsymbol{x_n})]$

- $C(\boldsymbol{x_n}) = \{\theta \in \Theta : \pi(\theta|\boldsymbol{x_n}) \geq k_\gamma\}$,
- $\mathbb{P}[\theta \in C(\boldsymbol{x_n})|\boldsymbol{x_n}] = 1 - \gamma$.

Exact credible intervals for θ

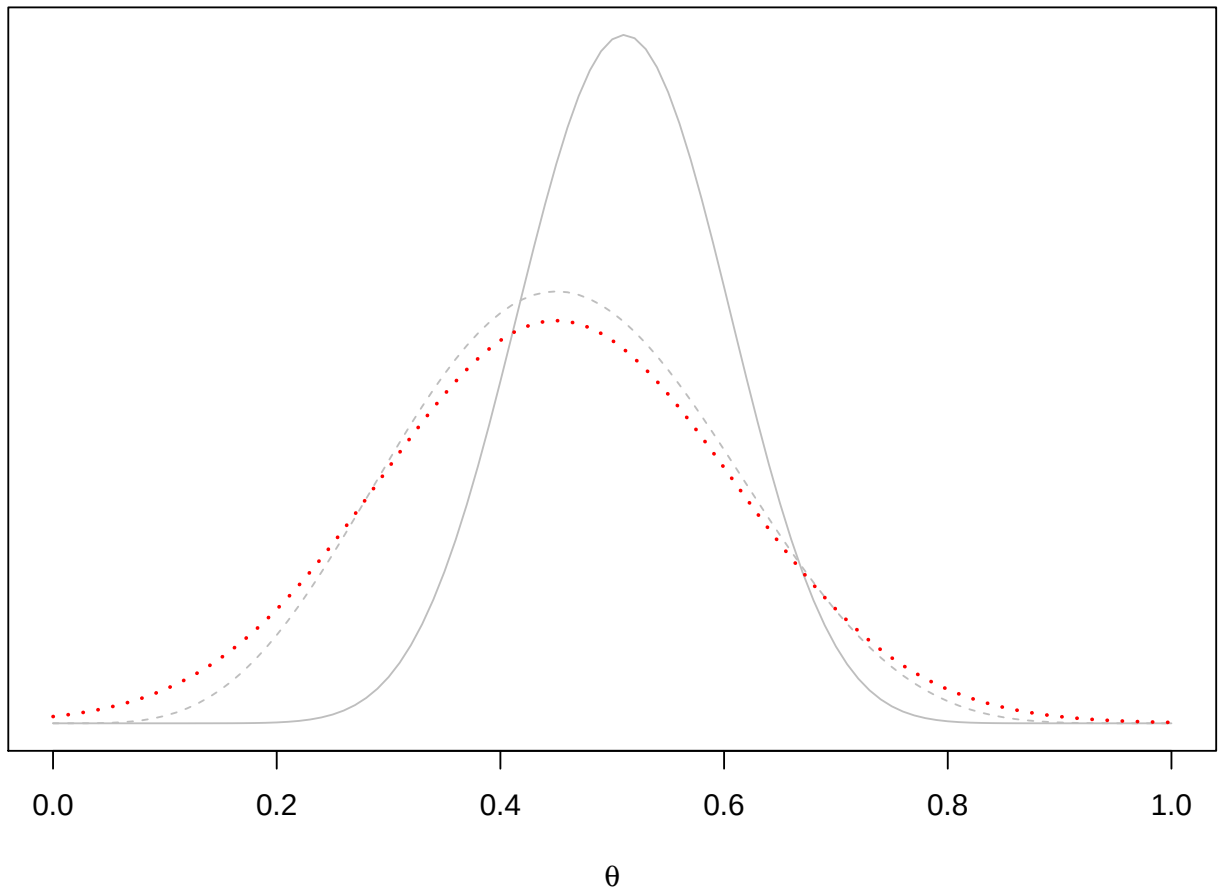
HPD: $C(\boldsymbol{x_n}) = [\ell(\boldsymbol{x_n}), u(\boldsymbol{x_n})]$

- $(1 - \gamma)$ shortest intervals
- ***but*** no closed-form expression

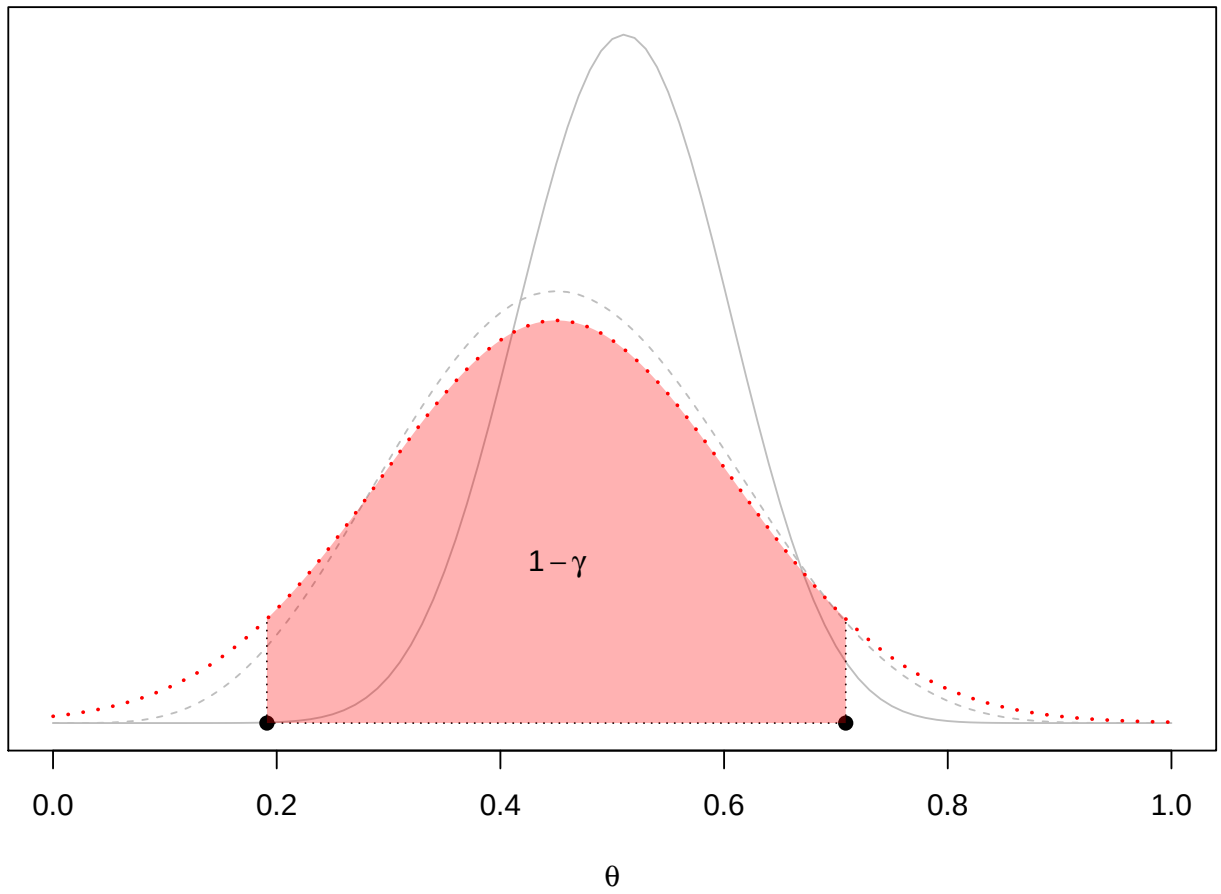
Likelihood Normal Approximation of the Posterior

$$\theta|\boldsymbol{x_n} \dot{\sim} \mathrm{N}[\hat{\theta}, I_n(\hat{\theta})^{-1}]$$

with $\hat{\theta}$ MLE and $I_n(\theta)$ expected Fisher Information



Likelihood Normal Approximation of the Posterior



Approximate credible intervals for θ

LNA: $\widetilde{C}(\boldsymbol{x_n}) = [\tilde{\ell}(\boldsymbol{x_n}), \tilde{u}(\boldsymbol{x_n})]$

- $\tilde{\ell} = \hat{\theta} - z_{1-\frac{\gamma}{2}}\ I_n(\hat{\theta})^{-1/2} = \bar{x}_n - z_{1-\frac{\gamma}{2}}\sqrt{\bar{x}_n(1-\bar{x}_n)/n}$
- $\tilde{u} = \hat{\theta} + z_{1-\frac{\gamma}{2}}\ I_n(\hat{\theta})^{-1/2} = \bar{x}_n + z_{1-\frac{\gamma}{2}}\sqrt{\bar{x}_n(1-\bar{x}_n)/n}$

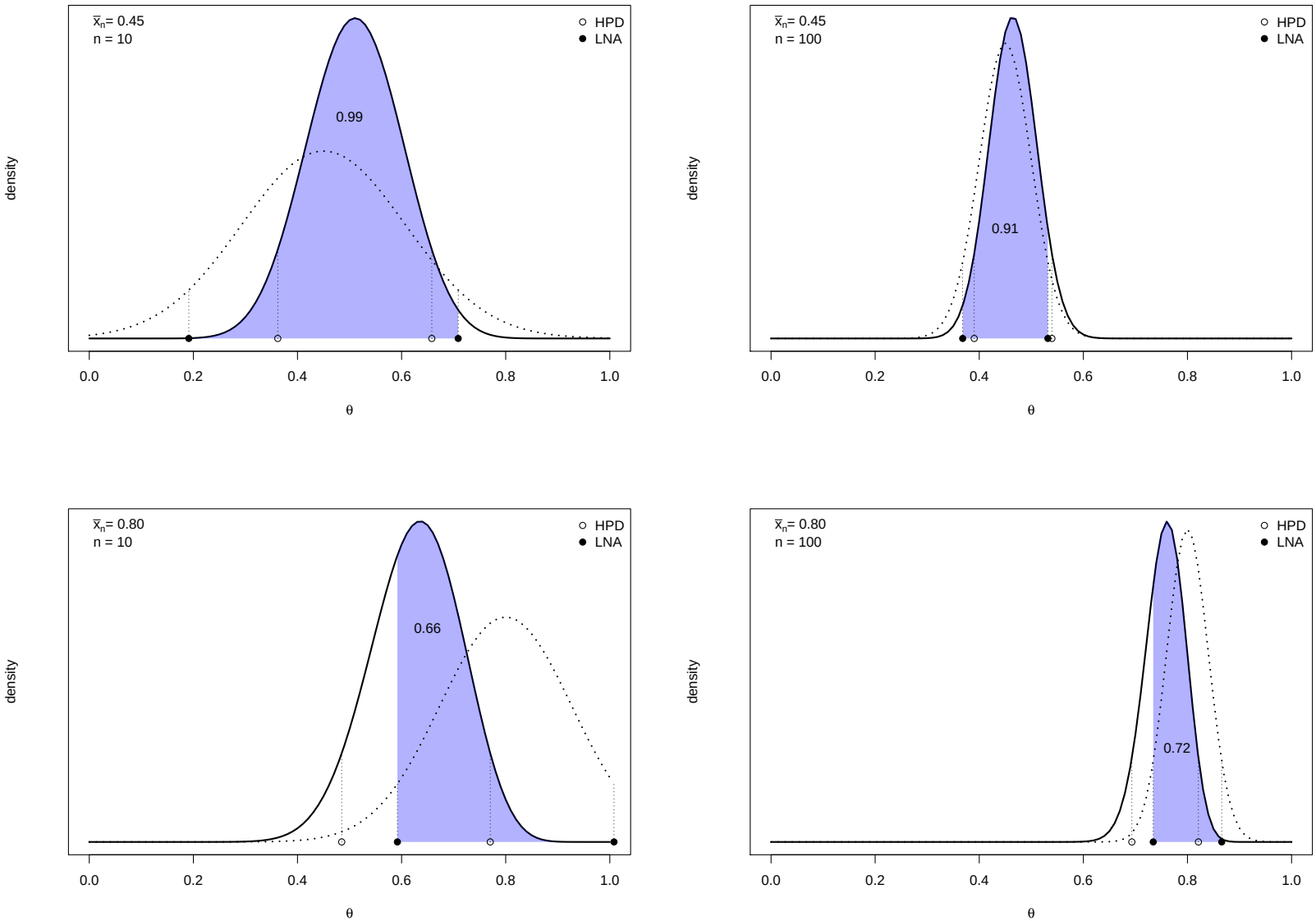
Calibrated approximate intervals

- In general, the *actual* posterior probability of $\widetilde{C}$ may be different from its *nominal* value $1 - \gamma$.
- We say $\widetilde{C}$ is *calibrated* if its exact posterior probability is equal to $1 - \gamma$.
- As n increases C and $\widetilde{C}$ converge, i.e.

$$\mathbb{P}[\theta \in \widetilde{C} | \boldsymbol{x_n}] \to (1 - \gamma).$$

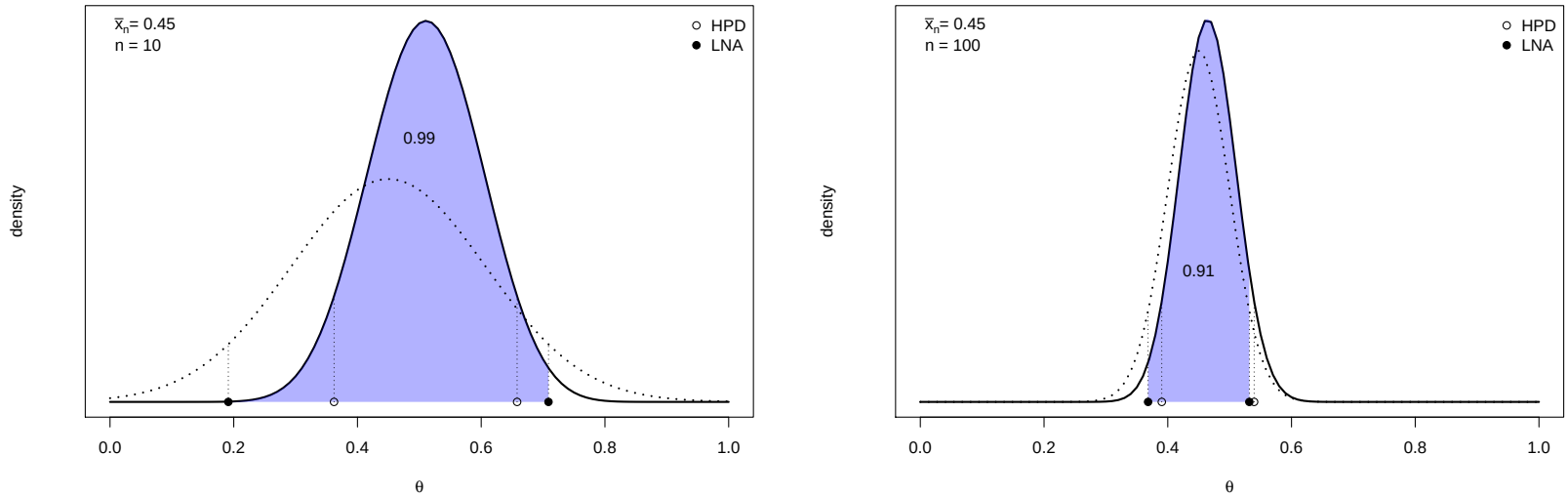
Posterior density VS its likelihood approximation

$1 - \gamma = 0.9$



Posterior density VS its likelihood approximation

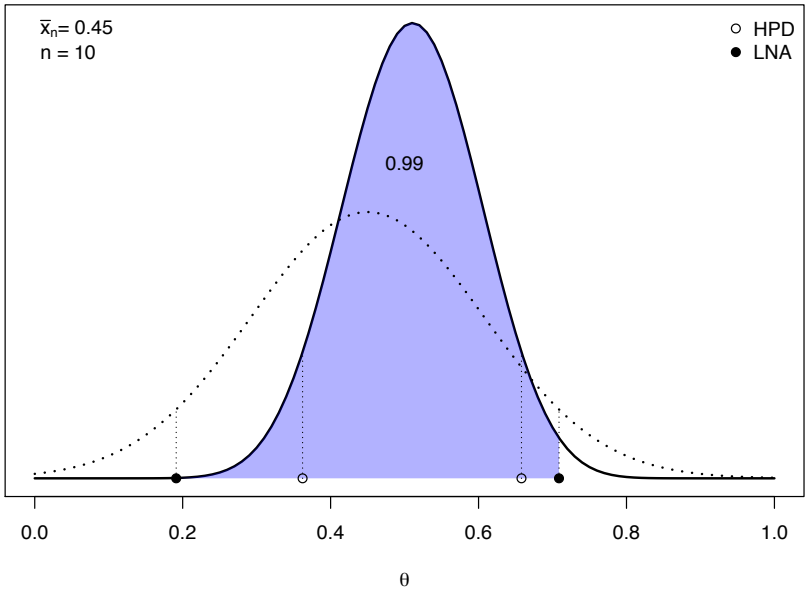
$1 - \gamma = 0.9$

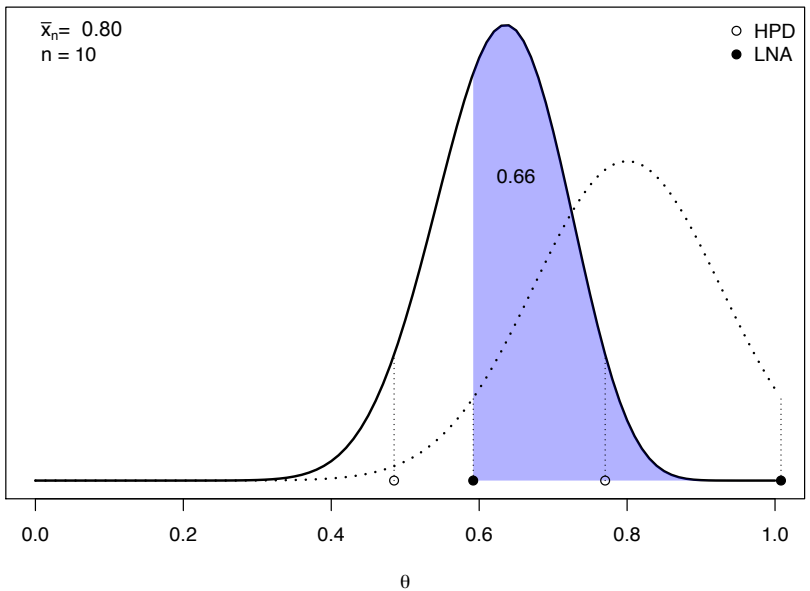


larger sample sizes ▶ better approximations

Posterior density VS its likelihood approximation

$1 - \gamma = 0.9$





larger distance between likelihood mode (MLE) and posterior mode

▼

greater discrepancy between HPD and LNA

Idea

A measure of lack of calibration of $\widetilde{C}$, is:

$$\left|\mathbb{P}[\theta \in \widetilde{C}|\boldsymbol{x_n}] - (1-\gamma)\right|$$

- quantifies the discrepancy between the actual posterior probability of $\widetilde{C}$ and $(1-\gamma)$
- ranges in $(0, 1-\gamma)$

Hence, a **relative measure of discrepancy** is

$$P(\boldsymbol{x_n}) = \frac{\left|\mathbb{P}[\theta \in \widetilde{C}|\boldsymbol{x_n}] - (1-\gamma)\right|}{1-\gamma}$$

Sample size determination

- Before observing the data, $P(\boldsymbol{X_n})$ is random.
- The calibration of $\widetilde{C}(\boldsymbol{X_n})$ can be studied by looking at

$$e_n^P = \mathbb{E}_d[P(\boldsymbol{X_n})]$$

computed w.r.t. the distribution of $\boldsymbol{X_n}$ for a design value θ_d.

- To obtain a calibrated interval, select

$$n_P^\star = \min\{n \in \mathbb{N} : e_n^P < \epsilon\}.$$

Computation of e_n^P

In simple cases with exact calculations...more often (*always*) via

Monte Carlo simulation

For each n, θ_d and $N = 10000$,

1. draw N samples $\boldsymbol{x_n}^{(1)}, \ldots, \boldsymbol{x_n}^{(N)}$ from $f_n(\cdot; \theta_d)$;
2. compute $\tilde{\ell}(\boldsymbol{x_n}^{(j)})$ and $\tilde{u}(\boldsymbol{x_n}^{(j)})$, for $j = 1, \ldots, N$;
3. compute $P(\boldsymbol{x_n}^{(j)})$, for $j = 1, \ldots, N$;
4. set $e_n^P \simeq \frac{\sum_{j=1}^{N} P(\boldsymbol{x_n}^{(j)})}{N}$.

Example: single-arm phase II trial

- **Setup**: based on a trial conducted between 2009 and 2011 to test the combination of *lenalidomideandrituximab* in patients with *recurrent indolent non-follicular lymphoma* (see Sacchi et al, 2016, and Sambucini, 2019)

- **Goal**: interval estimation of the treatment response rate θ

- **Endpoint**: sample treatment response rate $\hat{\theta}$, i.e. proportion of patients achieving complete/partial response

Example: single-arm phase II trial

- **Hystorical data**: 21 responses out of 39 patients
- **Prior elicitation**: $\theta \sim Be(\alpha, \beta)$ based on hystorical data
  $\Rightarrow$ prior mean $\alpha/(\alpha + \beta) = 0.54$
- **Sensitivity analysis**: different amount of information
  $\Rightarrow$ prior sample sizes $\alpha + \beta = (5, 10, 20)$
- **Non-informative case**: $\alpha = \beta = 1 \Rightarrow \theta \sim Unif(0, 1)$

Example: single-arm phase II trial

- **Design values**:

  ⋆ lowest acceptable value: $\theta_d = 0.45$

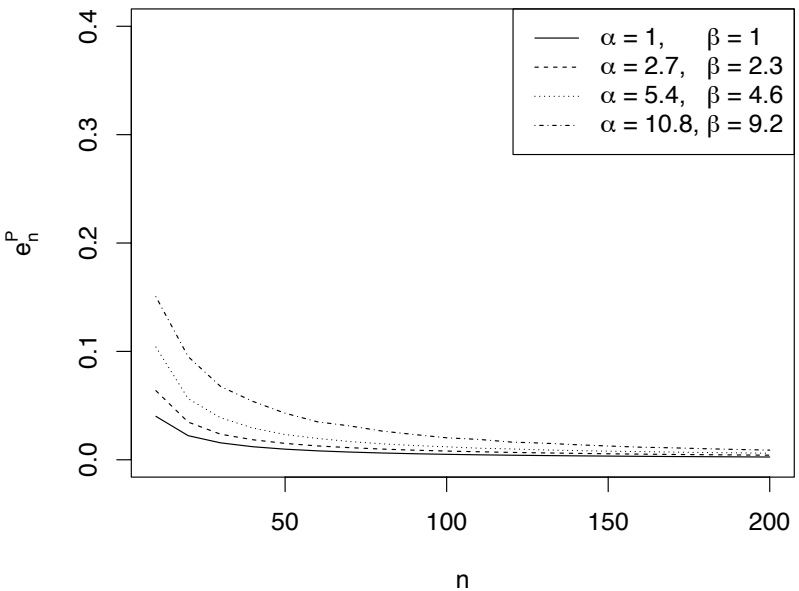
  ⋆ optimistic design value: $\theta_d = 0.80$

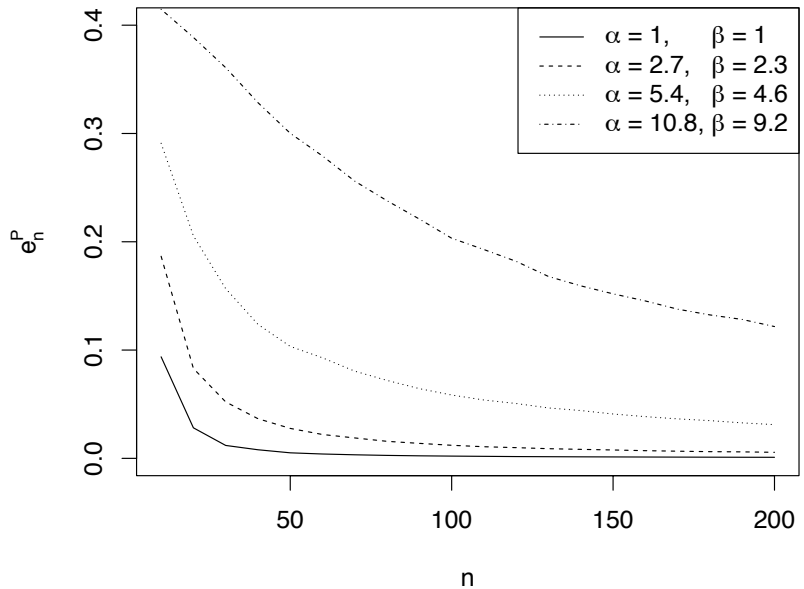
- **SSD criterion threshold**:

  $\epsilon = 0.01$

Example: single-arm phase II trial

Plots of e_n^P w.r.t. n for several values of (α, β) with $\theta_d = 0.45$ and $\theta_d = 0.8$



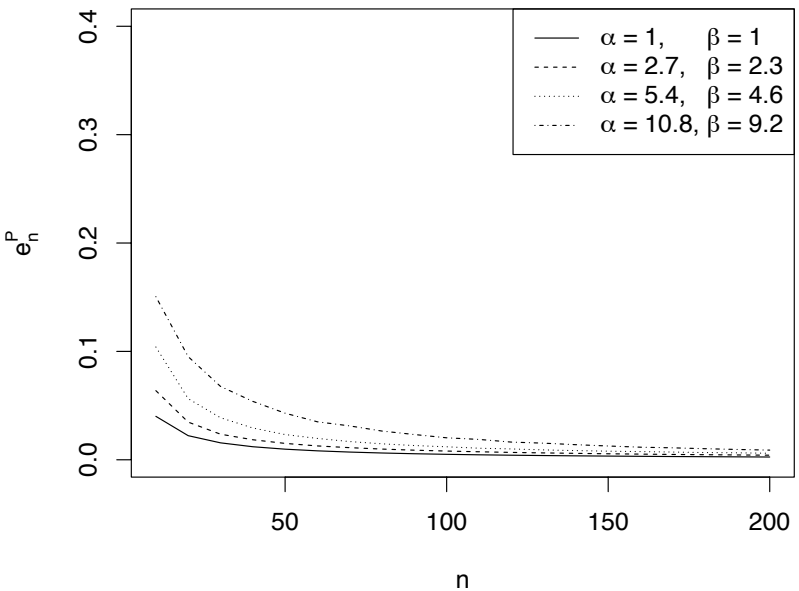


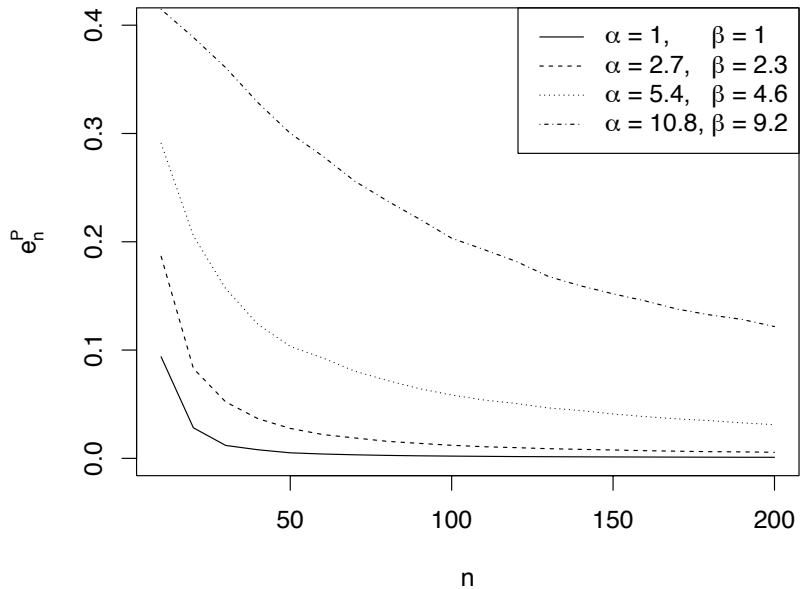
Remarks

1. *Effect of n.* Values of e_n^P decrease as n increases

Example: single-arm phase II trial

Plots of e_n^P w.r.t. n for several values of (α, β) with $\theta_d = 0.45$ and $\theta_d = 0.8$



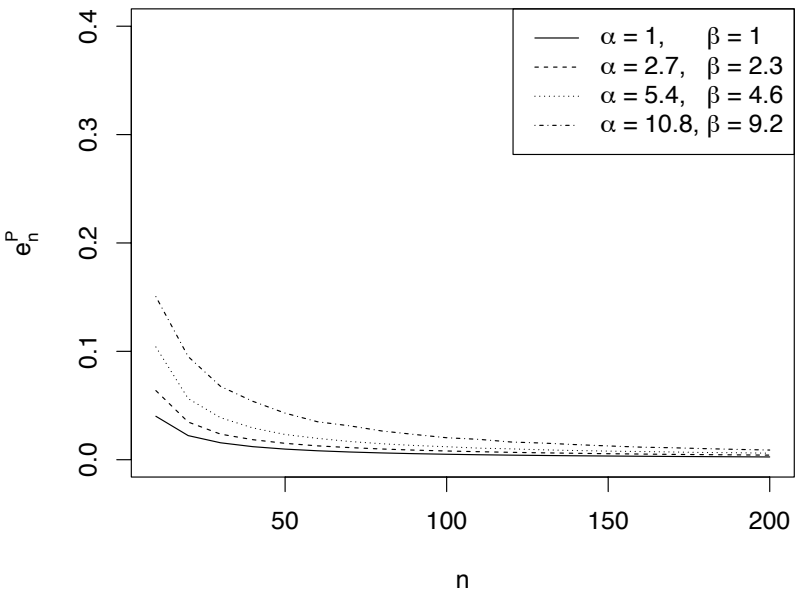


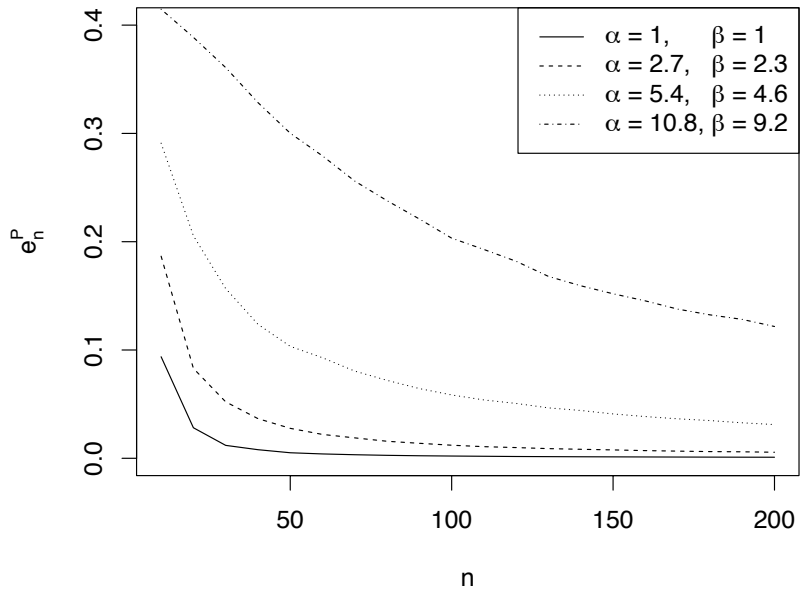
Remarks

2. *Effect of prior sample size.* Large $\alpha + \beta \Rightarrow$ large e_n^P

Example: single-arm phase II trial

Plots of e_n^P w.r.t. n for several values of (α, β) with $\theta_d = 0.45$ and $\theta_d = 0.8$





Remarks

3. *Effect of* $|\theta_d - \mathbb{E}(\theta)|$. If $\alpha + \beta \gg n$, large $|\theta_d - \mathbb{E}(\theta)| \Rightarrow$ large e_n^P

Example: single-arm phase II trial

Optimal sample sizes for different prior parameters, design values, $\epsilon_P = 0.01$, $\epsilon_L = 0.1$.

θ_d	(α, β)	$(1, 1)$	$(2.7, 2.3)$	$(5.4, 4.6)$	$(10.8, 9.2)$
0.45	$n_P^\star$	49	80	119	182
	$n_L^\star$	265	262	257	247
	$n_{\tilde{L}}^\star$	267	267	267	267
0.80	$n_P^\star$	35	118	646	2911
	$n_L^\star$	170	169	169	167
	$n_{\tilde{L}}^\star$	172	172	172	172

Example: single-arm phase II trial

Optimal sample sizes for different prior parameters, design values, $\epsilon_P = 0.01$, $\epsilon_L = 0.1$.

θ_d	(α, β)	$(1, 1)$	$(2.7, 2.3)$	$(5.4, 4.6)$	$(10.8, 9.2)$
0.45	$n_P^\star$	49	80	119	182
	$n_L^\star$	265	262	257	247
	$n_{\widetilde{L}}^\star$	267	267	267	267
0.80	$n_P^\star$	35	118	646	2911
	$n_L^\star$	170	169	169	167
	$n_{\widetilde{L}}^\star$	172	172	172	172

Remarks

4. *Comparison with ALC.* Opposite behaviour of $n_P^\star$ and $n_L^\star$.
Large $\alpha + \beta \Rightarrow$ short intervals $\Rightarrow$ small $n_L^\star$

Comments · Objections · Extensions

The traditional way of using Bayesian intervals...

- pre-posterior control of length and position of credible interval

A different look at the problem...

- number of units needed to *safely* use LNA instead of HPD
- price of using LNA instead of HPD *in terms of expected discrepancy*

Our criterion can be put beside additional criteria related to the main goal of the trial

Comments · Objections · Extensions

One drawback of $\widetilde{C}$ is that it may include values outside $[0, 1]$

- Common solution: $\psi = g(\theta)$ (e.g. the logodds)
- Normal approximation improves
- Implementation is substantially similar...
  - HPD bounds are obtained via MC
  - Delta method provides closed form expressions of LNA
- Similar comments apply

References

Bayes for CT & SSD methods based on intervals

- Spiegelhalter et al (2004) Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Statistics in Practice. Wiley.
- Joseph L, Wolfson D B, Berger R Du (1995) Sample Size Calculations for Binomial Proportions via Highest Posterior Density Intervals. J Roy Stat Soc: Series D (Statistician), 44: 143-154.
- M'Lan C E, Joseph L, Wolfson D B (2008) Bayesian sample size determination for binomial proportions. Bayesian Anal, 3: 269-296.
- Cao J, Lee J Jack, Alber S (2009) Comparison of Bayesian sample size criteria: ACC, ALC, and WOC. J Stat Plan Inference, 139: 4111-4122.
- Joseph L, Belisle P (2019) Bayesian consensus-based sample size criteria for binomial proportions. Stat Med, doi: 10.1002/sim.8316.

References

Our contributions to interval-based Bayes SSD

- Brutti P, De Santis F (2008) Robust Bayesian sample size determination for avoiding the range of equivalence in clinical trials. J Stat Plan Inference, 138: 1577-1591.
- Gubbiotti S, De Santis F (2011) A Bayesian method for the choice of the sample size in equivalence trials. Aust N Z J Stat, 53: 443-460.
- Brutti P, De Santis F, Gubbiotti S (2014) Predictive measures of the conflict between frequentist and Bayesian estimators. J Stat Plan Inference, 148: 111-122.

References

Application

- Sacchi et al (2016) Safety and efficacy of lenalidomide in combination with rituximab in recurrent indolent non-follicular lymphoma: final results of a phase II study conducted by the Fondazione Italiana Linfomi *Haematologica*, 101 (5), e196-e199.
- Sambucini, V (2019) Bayesian predictive monitoring with bivariate binary outcomes in phase II clinical trials. *Computational Statistics & Data Analysis*, Elsevier, vol. 132(C), 18-30.