Bayesian and Mixed Bayesian/Likelihood Criteria for Sample Size Determination

Statistics in Medicine, Vol. 16, 769-781 (1997)

Lawrence Joseph, Roxane Du Berger, Patrick Bélisle

October 21, 2021

Sample Sizes

Usually sample sizes are calculated to ensure a certain level of power for significance tests.

But Bayesians often avoid significance tests, so what should they do to determine an appropriate sample size when designing experiments?

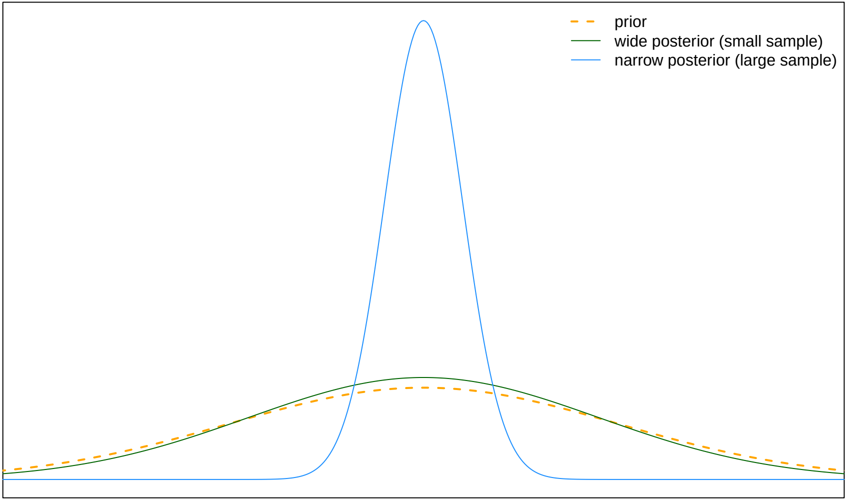
Spiegelhalter Says...

> "[In regards to proper Bayesian designed experiments] there is in principle no need for pre-planned sample sizes... **Alternatively, it is natural to focus on the eventual precision of the posterior distribution..."**

- Spiegelhalter et. al., Section 6.5[1] (emphasis mine).

[1] Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). Bayesian approaches to clinical trials and health-care evaluation (Vol. 13). John Wiley & Sons.

Posterior Precision



Credible Intervals

The posterior

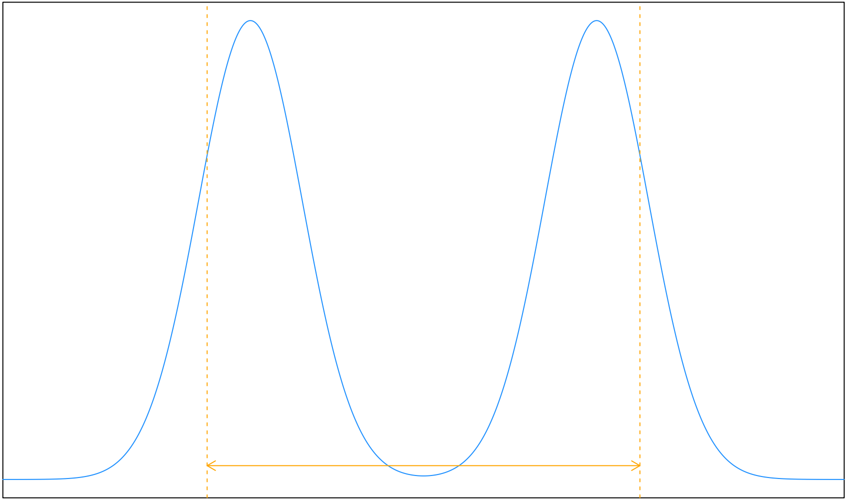
$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int_{\Theta} f(x|\theta)f(\theta)d\theta}$$

can be summarised by a credible interval, which is any interval $\mathcal{I}$ that satisfies

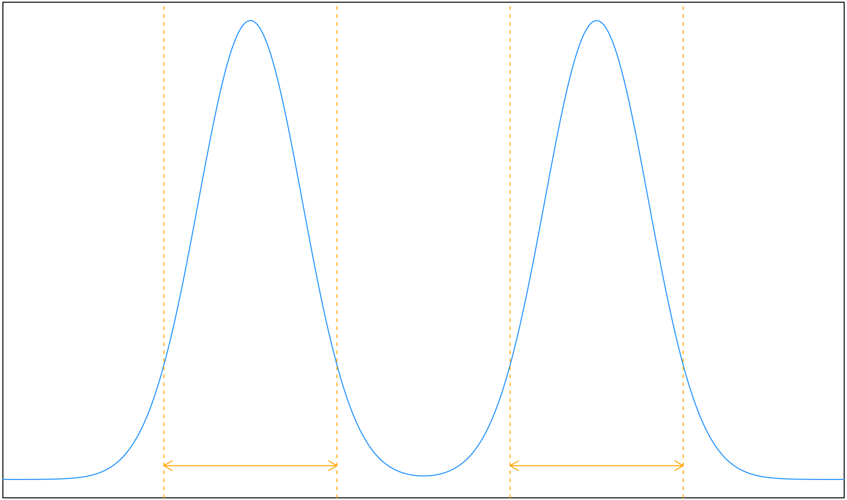
$$\int_{\theta \in \mathcal{I}} f(\theta|x)d\theta = 1 - \alpha$$

where 1 - α is the pre-specified desired coverage probability. The paper focuses on the *Highest Posterior Density* (HPD) interval.

Equal Tailed Intervals



HPD Intervals



Average Coverage Criterion (ACC)

Let $a(x, n)$ be the lower limit of the HPD. If we are set on a fixed length l, but less concerned about the coverage probability, we can calculate the minimum value of n that satisfies:

$$\int_{x \in \mathcal{X}} \left\{ \int_{a(x,n)}^{a(x,n)+l} f(\theta|x) d\theta \right\} f(x) dx \geq 1 - \alpha$$

which tells us that the average HPD computed from an n subset collection of the data space $\mathcal{X}$ will *on average* at least be as large as 1 - α.

Average Length Criterion (ALC)

We could do it the other way around: fix the coverage probability $1 - \alpha$. We can compute the minimum sample size n that satisfies

$$\int_{x \in \mathcal{X}} l'(x, n) f(x) dx \leq l$$

for HPD length $l'(x, n)$, i.e., satisfying:

$$\int_{a(x,n)}^{a(x,n)+l'(x,n)} f(\theta|x) d\theta = 1 - \alpha$$

Modified Worst Outcome Criterion (MWOC)

If averages are not sufficient, and we can ensure that we cover at least $1 - \alpha$ of the distribution with HPD length l if we use the MWOC:

$$\inf_{x \in \mathcal{S}} \left\{ \int_{a(x,n)}^{a(x,n)+l} f(\theta|x) d\theta \right\} \geq 1 - \alpha$$

where $\mathcal{S} \subseteq \mathcal{X}$.

Differences of Binomial Proportions Example

Table I. Sample sizes for example 1, using fully Bayesian, mixed Bayesian/likelihood, and standard frequentist criteria

	ACC	ALC	MWOC(95)	MWOC(99)	WOC
Full Bayes	1799	1763	2582	2687	3033
Mixed Bayes/likelihood	1840	1794	2625	2731	3070
Frequentist	1899		2825	2903	3074

Bayes returns smaller results thanks to the prior!

Mixed Bayesian Approach?

The paper suggests that a non-Bayesian may wish to utilise the above Bayesian methods to determine sample size, even if they intend to analyse the data in a non-Bayesian way!

The simple way they suggest to do this is to use the true prior $f(\theta)$ for the sample size calculations, but then revert to using a uniform when analysing the data.

Question for the group: would the above methods appeal to non-Bayesians for sample size determination?

Further Reading

- Wang, F., & Gelfand, A. E. (2002). A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. Statistical Science, 193-208.[2]
- Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). Bayesian approaches to clinical trials and health-care evaluation (Vol. 13). John Wiley & Sons.

[2] The publisher lists the authors in reverse order for some reason.