ORIGINAL PAPER

The variation of the posterior variance and Bayesian sample size determination

Jörg Martin[1] · Clemens Elster[1]

© The Author(s) 2020

Abstract

We consider Bayesian sample size determination using a criterion that utilizes the first two moments of the posterior variance. We study the resulting sample size in dependence on the chosen prior and explore the success rate for bounding the posterior variance below a prescribed limit under the true sampling distribution. Compared with sample size determination based on the average of the posterior variance the proposed criterion leads to an increase in sample size and significantly improved success rates. Generic asymptotic properties are proven, such as an asymptotic expression for the sample size and a sort of phase transition. Our study is illustrated using two real world datasets with Poisson and normally distributed data. Based on our results some recommendations are given.

1 Introduction

Sample size determination (SSD) is the attempt to estimate the data size that is needed in order to meet a certain criterion (Desu 2012). This task is usually performed at a planning stage before any data is actually measured or recorded so that especially in the context of high financial or temporal expenses a careful SSD becomes indispensable. In the design, say, of animal experiments or clinical trials SSD can even have an ethical dimension (Charan and Kantharia 2013; Dell et al. 2002). In this article we study a Bayesian method for SSD that limits the expected fluctuations of the uncertainty of the result. By "uncertainty" we will here mean (the square root of) the posterior variance. For n data points $\mathbf{x}_n = (x_1, \ldots, x_n)$ drawn from a sampling distribution $p(\mathbf{x}_n|\theta)$ with parameter θ the posterior distribution is defined by

✉ Jörg Martin
joerg.martin@ptb.de

Clemens Elster
clemens.elster@ptb.de

[1] Physikalisch-Technische Bundesanstalt, Abbestraße 2, 10587 Berlin, Germany

🖄 Springer

$$\pi(\theta|\mathbf{x}_n) \propto \pi(\theta) \cdot p(\mathbf{x}_n|\theta), \tag{1}$$

where $\pi(\theta)$ denotes the prior for the parameter θ. The posterior variance is then given as

$$u_n^2 := \mathrm{Var}_{\theta \sim \pi(\theta|\mathbf{x}_n)}(\theta). \tag{2}$$

In practice, a scientist performing an experiment might desire to specify her/his result with an according uncertainty, say

$$\hat{\theta} \pm u_n,$$

with u_n being the square root of u_n^2 as defined in (2) and with $\hat{\theta}$ being the posterior mean. In order for this result to be precise enough the scientist might desire to fulfill a condition such as $u_n < \varepsilon$ or, equivalently,

$$u_n^2 < \varepsilon^2 \tag{3}$$

for some small, positive ε that is chosen a priori. As the posterior distribution is dependent on the data $\mathbf{x}_n$, so is u_n^2. Choosing an appropriate sample size n so that (3) is guaranteed *before* $\mathbf{x}_n$ is known is only possible for a few restricted scenarios, for instance Bernoulli distributed samples (Pham-Gia and Turkkan 1992; Joseph and Bélisle 2019). A more generally applicable criterion is to require instead of (3)

$$\overline{u_n^2} = \mathbb{E}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}[u_n^2] < \varepsilon^2, \tag{4}$$

where $m(\mathbf{x}_n) = \int p(\mathbf{x}_n|\theta)\pi(\theta)\mathrm{d}\theta$ denotes the *prior predictive*. This is known as the *average posterior variance criterion* (APVC) in the literature (Wang and Gelfand 2002; Pham-Gia and Turkkan 1992; De Santis 2007). For many standard cases explicit expressions for $\overline{u_n^2}$ can be derived. The usage of the prior predictive $m(\mathbf{x}_n)$ is quite natural as it describes what is known about the data $\mathbf{x}_n$ given our prior knowledge. We will denote the smallest n such that the APVC (4) is satisfied throughout this article by $\widetilde{n}_\varepsilon$. In the literature many alternative criteria can be found that replace u_n^2 by some other, data dependent random variable $T(\mathbf{x}_n)$, compare for instance (Adcock 1997; M'lan et al. 2008; De Santis 2006; Wang and Gelfand 2002; Rubin and Stern 1998). While we will stick in this article to the choice $T(\mathbf{x}_n) = u_n^2$ many of the ideas presented here can, in principle, be translated to such approaches.[1]

The APVC has a rather obvious drawback: it only guarantees (3) to hold on average. Consequently, one expects that $u_{\widetilde{n}_\varepsilon}$ is larger than ε for certain data samples $\mathbf{x}_{\widetilde{n}_\varepsilon}$, cf. also Fig. 3a below. To get a better grasp on the variability of the uncertainty one can make more extensive usage of the prior predictive $m(\mathbf{x}_n)$ (De Santis 2006, 2007; Brutti et al. 2008; Gubbiotti and De Santis 2011; Sambucini 2008; Sahu and Smith 2006). This article aims at studying the behavior of such criteria and

---

[1] Section 3 of this article, for instance, relies on Assumption 3.1. Provided similar assumptions hold for $T(\mathbf{x}_n)$ then all the proofs given there carry through.

thereby to give some guidance or, at least, a deeper understanding for a SSD based on $m(\mathbf{x}_n)$. In order to do so we will use an extension of the APVC (4) which we will call the *variation of the posterior variance criterion* (VPVC) and which takes the form

$$\overline{u_n^2} + k \cdot \Delta u_n^2 < \varepsilon^2. \tag{5}$$

where k is a parameter to be chosen and where

$$\Delta u_n^2 = (\mathrm{Var}_{x \sim m(\mathbf{x}_n)}(u_n^2))^{1/2}. \tag{6}$$

Throughout this article we will denote the smallest n such that (5) is satisfied by n_ε. We will see in Sect. 2 below that taking (5) into account leads to substantially different sample sizes than the sole consideration of the APVC (4) and to a better compliance with (3). We will often take $k = 2$ in this article (loosely motivated from the normal distribution) but will show that for $\varepsilon \to 0$ there is an optimal k_*. We will give some ideas on how to guess k_* in practice. Moreover, we will provide an asymptotic formula for the sample size n_ε in the small ε regime.

As the VPVC (5) only uses the first two moments it is of course not as exhaustive in its description as a criterion that considers the full law $m(\mathbf{x}_n)$—compare for instance (De Santis 2006, 2007; Brutti et al. 2008; Gubbiotti and De Santis 2011; Sambucini 2008), but for different criteria than the one considered here. In Sect. 2.1 below we will compare ourselves to a more extensive usage of $m(\mathbf{x}_n)$ and will find that in many cases (5) is a good approximation. Taking this approximation on the other hand is quite convenient for our purposes: it provides us with explicit expressions, spares us numerical issues and allows for a rather concise discussion of asymptotic properties in Sect. 3. We will discuss our method for two common cases, namely Poisson and normally distributed data, and illustrate our discussion with real world datasets.

The authors are, to the best of their knowledge, not aware of work in the literature that considers a criterion in the exact same shape of (5). (In Pham-Gia and Turkkan 1992) Pham-Gia and Turkkan consider an object such as Δu_n^2 from (6) for a Binomial distribution but apply it in a different manner. Our approach is inspired from the quite common idea of using $m(\mathbf{x}_n)$ for studying u_n^2. We believe that the discussion in this paper deepens the understanding of SSD methods built on $m(\mathbf{x}_n)$ in general. The paper is organized as follows: Sect. 2 discusses the application of the VPVC to Poisson and normally distributed data. We will compare our results to the simpler AVPC method (4), visualize the effect of using a prior that is inconsistent with the underlying parameter and debate how a conservative SSD could be performed. For this purpose we will use actual datasets, namely the daily number of accidents in Leeds in 2018 (Leeds City Council 2019) and the mercury concentration in the blood of alligators in South Carolina and Florida (Lawson and Jodice 2019; Nilsen et al. 2019). In Sect. 3 we will look at the behavior of the VPVC for $\varepsilon \to 0$. The results from that section, especially Theorem 3.5, indicate that SSD methods based on $m(\mathbf{x}_n)$ exhibit some sort of phase transition in this limit. We will show that a k_* exists such that for $k > k_*$ a SSD based on the VPVC (5) will

ensure that (3) will *asymptotically be true with probability 1* under the true sampling distribution. As k_* will depend on the (usually unknown) true value of θ, we will discuss a method to get an upper bound based on the prior knowledge. Moreover, in Lemma 3.4 we will provide a generic, asymptotic formula for the sample n_ε predicted by the VPVC.

2 SSD based on the variation of the posterior variance

In this section we study the effects of using the VPVC as proposed in (5) and discuss its dependency on the prior knowledge and the true parameter. The discussion is first carried out for the case of a single parameter and Poisson distributed data. In Sect. 2.2 we then consider an example that involves nuisance parameters and normally distributed data.

2.1 Single parameter: poisson distributed data

The histogram in Fig. 1 shows the daily number of traffic accidents that happened in Leeds (UK), taken from Leeds City Council (2019). The number of accidents is an important metric when judging the effect of traffic planning (Hadayeghi et al. 2010). While evaluating a whole year will give a good precision on the expected daily number of accidents, this might be an intolerable time scale for judging the effect of traffic measures such as a revised speed limit. Planning an evaluation at a too early stage on the other hand might turn out to be too vague and therefore useless.

What is a suitable number of days before we can give a decent estimate? To formalize this question let us suppose that the data for n days follows the product of n Poisson distributions with the (unknown) parameter θ:[2]

$$p(\mathbf{x}_n|\theta) = \prod_{i=1}^{n} \text{Poi}(x_i|\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} e^{-\theta}, \tag{7}$$

where $\mathbf{x}_n = (x_1, \ldots, x_n)$ and each x_i for $i = 1, \ldots, n$ should be read as the number of accidents on a day. We want to find n such that the uncertainty u_n about θ is less than some ε, for which we take $\varepsilon = 0.3$ in this section. To specify our prior knowledge about θ we will use a gamma distribution $\pi(\theta) = \text{Gamma}(\theta|\alpha, \beta)$ with shape α and rate β. As uncertainty u_n for our result we take the square root of the posterior variance $u_n^2 = \frac{\alpha + \sum_{i=1}^{n} x_i}{(n+\beta)^2}$. The inequality $u_n^2 < \varepsilon^2$ has no solution for n that holds for any $\mathbf{x}_n$. However, we can average over prior predictive $m(\mathbf{x}_n)$ and choose the smallest $n \geq 1$ such that

$$\overline{u_n^2} = \mathbb{E}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}[u_n^2] = \frac{\alpha}{\beta} \cdot \frac{1}{n+\beta} < \varepsilon^2. \tag{8}$$

We referred to this as the APVC in the introduction and denoted the corresponding n

---

[2] This is an approximation. The ratio of variance and mean is around 2 for the full dataset, however to keep the setup simple we will stick to the Poisson assumption and treat the mean, i.e. the MLE, as the true parameter.

Fig. 1 Daily number of accidents in Leeds in 2018, taken from (2018)
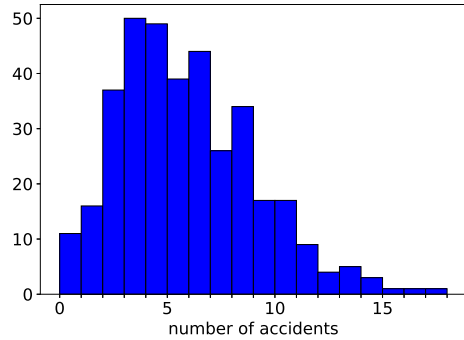
by $\widetilde{n}_\varepsilon$. The result of such an SSD for $\varepsilon = 0.3$ is shown in Fig. 2a: for various values of the prior mean $\mathbb{E}_{\theta \sim \pi(\theta)}[\theta] = \frac{\alpha}{\beta}$ and its standard deviation $((\mathrm{Var}_{\theta \sim \pi(\theta)}(\theta)))^{1/2} = \frac{\alpha^{1/2}}{\beta}$ we plotted the sample size $\widetilde{n}_\varepsilon$ predicted by (8). Naturally, as the mean of $\pi(\theta)$ increases the sample size $\widetilde{n}_\varepsilon$ increases as well since θ is linked to the variance of the Poisson distributed data. The standard deviation of $\pi(\theta)$ seems to have only a minor influence on the sample size with one exception: below a certain threshold the prior variance pushes the posterior variance into the right ballpark so that even a minimal sample size of $\widetilde{n}_\varepsilon = 1$ is enough to fulfill (8).

Figure 3a illustrates for which prior choices the SSD based on the APVC is successful based on the full year accident dataset we depicted in Fig. 1: for each prior mean and standard deviation and the corresponding sample sizes $\widetilde{n}_\varepsilon$ from Fig. 2a 100 random samples of size $\widetilde{n}_\varepsilon$ were drawn from the Leeds accident data and the corresponding $u_{\widetilde{n}_\varepsilon}$ were computed. Figure 3a shows the fraction of $u_{\widetilde{n}_\varepsilon}$ that is actually below ε.

For comparison the "true" value (the average of the full data shown in Fig. 1) is depicted by the dashed line. The result is rather disappointing. Only beyond the true θ of 5.46 the quota rises above 60%. The reason is quite apparent: according to the APVC $u_{\widetilde{n}_\varepsilon}^2$ is expected to be small enough only *on average*.

As we pointed out in the introduction we therefore here use a refined criterion, which we called the VPVC and which takes the form

$$\overline{u_n^2} + k \cdot \Delta u_n^2 < \varepsilon^2, \tag{9}$$

where $\Delta u_n^2 = (\mathrm{Var}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}(u_n^2))^{1/2} = \frac{\alpha^{1/2}}{\beta} \cdot \frac{n^{1/2}}{(n+\beta)^{3/2}}$ and where we choose $k = 2$. The result of the VPVC is depicted in Fig. 2b. There are two clear differences compared to Fig. 2a: first, the sample sizes n_ε are higher than the numbers $\widetilde{n}_\varepsilon$ we obtained from the APVC, which was expected as we added an additional positive term $k \Delta u_n^2$ to the left hand side of the criterion. Second, and perhaps more important, the result of the SSD substantially increases once we increase the standard deviation of the prior. This allows us to make our sample choice more conservative for a given prior mean by increasing its standard deviation, i.e. our prior uncertainty about θ, which is quite natural. The most conservative SSD is thus located in the right upper corner of the
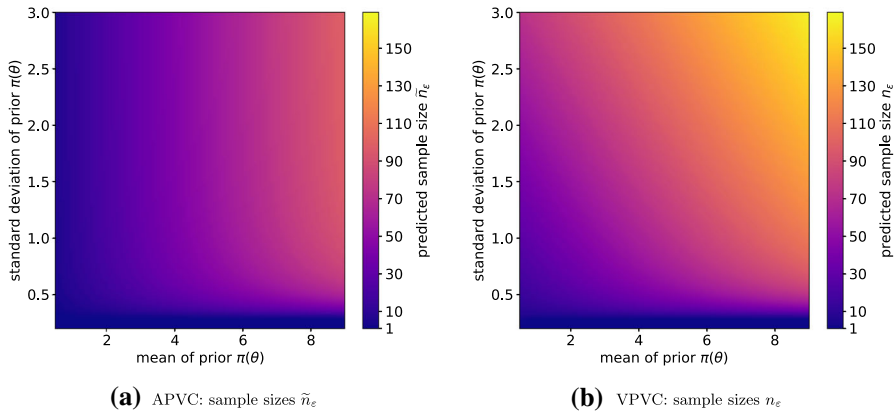
(a) APVC: sample sizes $\widetilde{n}_\varepsilon$

(b) VPVC: sample sizes n_ε

Fig. 2 Left: sample sizes predicted for $\mathrm{Poi}(\mathbf{x}_n|\theta)$ distributed data and various priors $\pi(\theta) = \mathrm{Gamma}(\theta|\alpha,\beta)$ for $\varepsilon = 0.3$ by the APVC (8). Right: sample sizes n_ε predicted by the VPVC (9) for the same setup
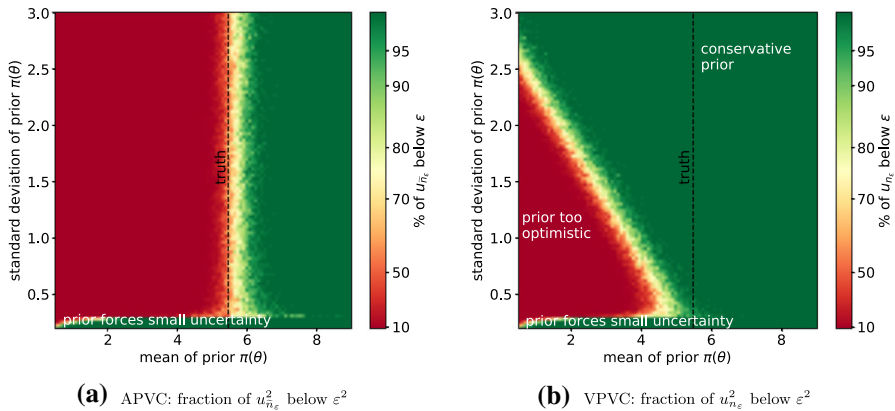

(a) APVC: fraction of $u_{\widetilde{n}_\varepsilon}^2$ below ε^2

(b) VPVC: fraction of $u_{n_\varepsilon}^2$ below ε^2

Fig. 3 Left: fraction of $u_{\widetilde{n}_\varepsilon}^2$ below ε^2 for the data from Fig. 1 using the sample sizes $\widetilde{n}_\varepsilon$ predicted by the APVC (Fig. 2a). Right: Fraction of $u_{n_\varepsilon}^2$ below ε^2 for the sample sizes n_ε predicted by the VPVC (Fig. 2b)

plot. This phenomenon for parameters that describe the variance of the data will appear again in the next subsection.

The success of the refined criterion is depicted in Fig. 3b. The percentage of u_{n_ε} below ε reaches now much higher values, at many positions beyond 95%. This success is also more robust against deviations from the true value. Even for a mean of the prior $\pi(\theta)$ which is well below the true θ a high enough standard deviation of $\pi(\theta)$ will allow to reach a quota of 95%.

Figure 3b therefore splits in three areas. As in Fig. 3a there is a bottom part where the posterior variance is dominated by the small prior uncertainty so that $u_{n_\varepsilon} < \varepsilon$ is easily satisfied even for the minimal sample size of $n_\varepsilon = 1$. As the standard deviation of $\pi(\theta)$ gets larger a too small value of the prior mean will result in a low quota: the prior was chosen too optimistic. Increasing either the mean of $\pi(\theta)$ or its

standard deviation will lead however to a more conservative SSD and to a better compliance of $u_{n_\varepsilon} < \varepsilon$. The highest percentage can therefore be found in the upper right corner of Fig. 3b and in the bottom area.

We already pointed out in the introduction of this article that enhancing the mean by the variance as in the VPVC (5) is, in a certain sense, only an approximation and an exhaustive treatment would use the full distribution $m(\mathbf{x}_n)$. In fact many methods in the literature (De Santis 2006, 2007; Brutti et al. 2008; Gubbiotti and De Santis 2011; Sambucini 2008; Wang and Gelfand 2002; Chen et al. 2011; Psioda and Ibrahim 2019) use a criterion that is formulated as a probability when $\mathbf{x}_n$ is drawn from $m(\mathbf{x}_n)$. The cited works use different criteria that are not directly comparable to the VPVC presented here. To allow for some comparison let us forge (5) into a probabilistic shape

$$\mathbb{P}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}\left(u_n^2 < \varepsilon^2\right) \geq p \qquad (10)$$

where p can be chosen between 0 and 1. We will denote the smallest n for which (10) is satisfied as $n_\varepsilon^{\mathbb{P}}$. In a way (10) resembles the length probability criterion (LPC) from the literature, compare for example (Santis et al. 2004; De Santis and Pacifico 2004; De Santis 2007; Joseph and Belisle 1997). For most cases $n_\varepsilon^{\mathbb{P}}$ can only be estimated numerically via sampling and even for rather easy cases, such as the one from the next subsection, we observed that finding a stable $n_\varepsilon^{\mathbb{P}}$ is rather elaborate. For the Poisson setup a computation shows that $\sum_{i=1}^n x_i = (n+\beta)^2 u_n^2 - \alpha$ follows a negative binomial distribution with failures α and success probability $\frac{n}{n+\beta}$ so that we can in fact precisely compute $n_\varepsilon^{\mathbb{P}}$ in this case. Figure 9 in the appendix shows the relative deviation between n_ε and $n_\varepsilon^{\mathbb{P}}$ for various choices of p in (10) together with the corresponding success rates $u_{n_\varepsilon^{\mathbb{P}}} < \varepsilon$. For values of p between 0.95 and 0.97 we observe a good accordance between the VPVC and (10): the relative deviation between n_ε and $n_\varepsilon^{\mathbb{P}}$ is below 5% for almost all priors, with exception of extremely skew choices (in the upper left corner) where a strong dependency of $n_\varepsilon^{\mathbb{P}}$ on p can be observed. A value of $p = 0.98$, which would arise for a normal distribution for $k = 2$ and thus ignore any skewness, produces for a range of prior choices substantially larger sample sizes.

2.2 Nuisance parameters: normally distributed data

Figure 4 shows 356 measurements of mercury concentration in the whole blood of alligators, collected at the Yawkey Wildlife Center in South Carolina (Lawson and Jodice 2019) and at the Merritt Island National Wildlife Refuge in Florida (Nilsen et al. 2019) between 2010–2017 and 2007–2014. The bioaccumulation of mercury in alligators can serve, due their role as top predators, as an indicator of pollution (Rumbold et al. 2002; Grillitsch and Schiesari 2010).

We want to study the minimal number of alligators that is needed in order to specify the Hg concentration up to a pre-specified precision ε and will, in a similar manner as in Sect. 2.1, use the data collection from (Lawson and Jodice 2019; Nilsen et al. 2019) to validate the success of this sample size planning.

It is common to model the concentration of pollutants via a log-Normal distribution (Ott 1990): for n measurements of mercury concentration $c_{\mathrm{Hg},i}$ we set

Fig. 4 Logarithmic concentration of mercury measured in the whole blood of alligators in South Carolina and Florida. This dataset is a combination of (Lawson and Jodice 2019, Nilsen et al. 2019)
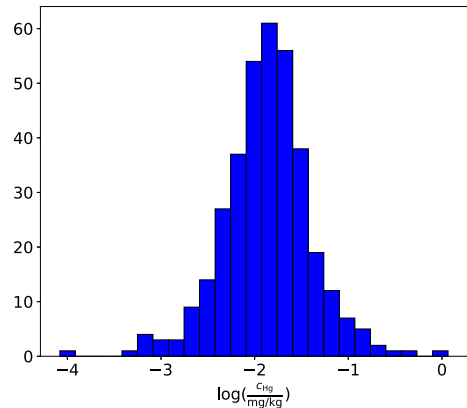
$x_i = \log(\frac{c_{\text{Hg},i}}{\text{mg/kg}})$ and assume that $p(\mathbf{x}_n|\mu, \sigma^2) = \prod_{i=1}^{n} \mathcal{N}(x_i|\mu, \sigma^2)$ with mean μ and standard deviation σ^2. Both parameters μ and σ^2 are unknown, with μ as the parameter of interest and σ^2 as nuisance parameter. We take the normal-inverse-gamma prior (Fink 1997) $\pi(\mu, \sigma^2) = \mathcal{N}(\mu|\mu_0, \lambda\sigma^2) \cdot \text{IG}(\sigma^2|\alpha, \beta)$, where IG denotes the inverse-gamma distribution and where $\alpha > 2, \lambda, \beta > 0$ and μ_0 are hyperparameters. The squared uncertainty u_n^2 for μ is given by the variance of the marginal posterior $\pi(\mu|\mathbf{x}_n)$, that is $u_n^2 = \text{Var}_{\mu \sim \pi(\mu|\mathbf{x}_n)}(\mu) = \frac{2\beta + \sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{n}{2n_\lambda}(\bar{x} - \mu_0)^2}{n_\lambda(n + 2\alpha - 2)}$ with $n_\lambda = n + \lambda^{-1}$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Let n_ε be again the smallest n such that $\overline{u_n^2} + k \cdot \Delta u_n^2 < \varepsilon^2$, where $\overline{u_n^2} = \mathbb{E}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}[u_n^2] = \frac{\beta}{n_\lambda(\alpha - 1)}$,

$$\Delta u_n^2 = (\text{Var}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}(u_n^2))^{1/2} = \frac{\beta}{n_\lambda(\alpha - 1)(\alpha - 2)^{1/2}}\left(\frac{n}{n + 2\alpha - 2}\right)^{1/2}$$

and where we choose once more $k = 2$. The sample sizes n_ε for various choices of the prior $\pi(\mu, \sigma^2)$ and $\varepsilon = 0.1$ are depicted in Fig. 5.

To visualize the impact of the prior knowledge we varied one of the marginals $\pi(\mu)$ and $\pi(\sigma^2)$ while keeping the other one fixed. In Fig. 5a we varied the marginal prior $\pi(\mu)$, while fixing the marginal $\pi(\sigma^2)$ to have a mean of 0.15 and a standard deviation of 0.10, for comparison this $\pi(\sigma^2)$ was marked by the black cross in Fig. 5b. For Fig. 5b we fixed $\pi(\mu)$ to have a mean -1.4 and a standard deviation of 0.6 (marked by the black cross in Fig. 5a) and varied $\pi(\sigma^2)$. In particular, the two crosses in Fig. 5a, b both mark positions with an equal sample size of $n = 33$. The APVC criterion from (4) and the average coverage criterion from (Adcock 1988, Joseph et al. 1995) both yield a sample size of 15 for this prior.

Note, that the effect of the marginal $\pi(\mu)$ is rather minimal, once $\pi(\sigma^2)$ is kept fixed. In fact, the VPVC turns out to be independent of the hyperparameter μ_0—compare the vertical symmetry in Fig. 5a. Varying the standard deviation of $\pi(\mu)$ while keeping $\pi(\sigma^2)$ fixed will only affect the hyperparameter λ which has only a minor influence on the SSD result. The only exception to this is the bottom area of Fig. 5 where the small standard deviation of $\pi(\mu)$ forces the posterior variance to be
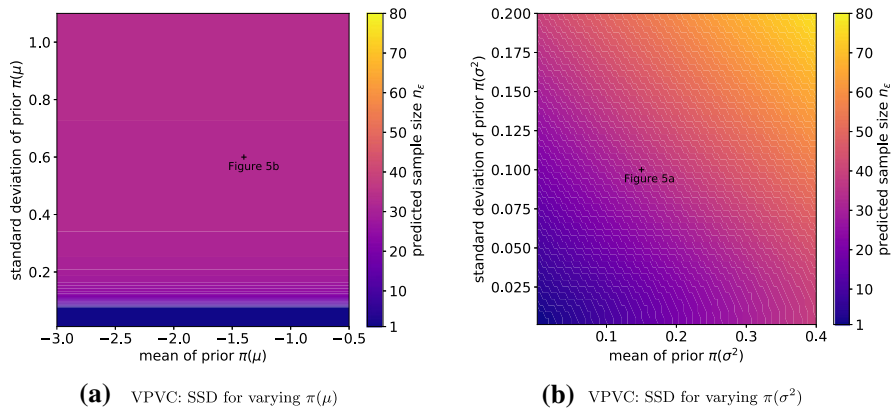
(a) VPVC: SSD for varying $\pi(\mu)$

(b) VPVC: SSD for varying $\pi(\sigma^2)$

Fig. 5 Sample sizes predicted by the VPVC for different $\pi(\mu)$ and $\pi(\sigma^2)$. Left: $\pi(\sigma^2)$ is fixed as indicated by the black cross in Fig. 5b and the sample sizes n is depicted for various $\pi(\mu)$. Right: Sample sizes for various $\pi(\sigma^2)$ and a fixed $\pi(\mu)$ marked by the black cross in Fig. 5a

small and thus predicts a small sample size of 1, which is similar to a phenomenon we observed in Sect. 2.1. Figure 5b shows that the variance parameter σ^2 has a similar influence on the sample size as the Poisson parameter in Sect. 2.1: increasing either the mean or the standard deviation of $\pi(\sigma^2)$ will increase the sample size so that the most conservative experimental design is located in the upper right corner. This behavior can again be expected since increasing the mean of $\pi(\sigma^2)$ one will expect data that have larger variability and contain less information. In contrast to Figs. 2b and 5a there is no distinct bottom area of minimal sample sizes.

As in Sect. 2.1 we can use the full dataset, the one depicted in Fig. 4, to judge the success of our sample size planning. Figure 6 displays, for the same priors as in Fig. 5 the proportion of u_{n_ε} below ε for u_{n_ε} computed from random samples of size n_ε, drawn from full dataset. The left plot, Fig. 6a, reveals that the choice of the marginal $\pi(\mu)$ has mostly an impact on the success of the SSD if its mean is far-off from the true parameter. Moreover, a very precise prior knowledge about μ, will result in a high proportion of $u_{n_\varepsilon} < \varepsilon$, which is similar to the observations we made in Sect. 2.1. The behavior of the percentage for which $u_{n_\varepsilon} < \varepsilon$ when varying $\pi(\sigma^2)$ is similar to what we have observed for the Poisson parameter in Sect. 2.1. For small values of the mean of $\pi(\sigma^2)$ the quota drops, which can be cured by increasing the standard deviation of $\pi(\sigma^2)$. The highest percentage can be found in the upper right corner. For the prior marked by the black crosses in Fig. 6, the same as the one marked in Fig. 5, the quota is around 95% while the sample size predicted by the APVC and average coverage criterion only yields a percentage of around 35%. The bottom area of Fig. 6b, where the posterior variance is heavily influenced by the prior, is not as distinct as in Fig. 3b, this role seems to be played in this context by μ - compare Fig. 6a.

Let us shortly summarize the most important observations we have made concerning SSD based on the prior predictive $m(\mathbf{x}_n)$ via using the VPVC (5). There are two strategies to have a good chance of achieving $u_n < \varepsilon$:
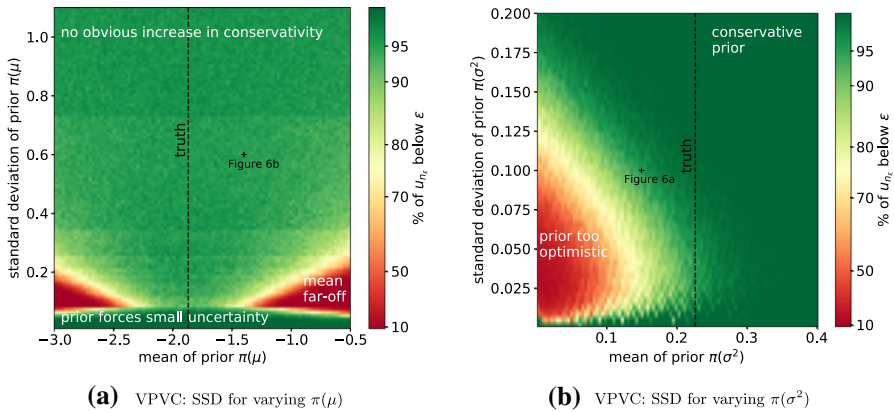
(a) VPVC: SSD for varying $\pi(\mu)$

(b) VPVC: SSD for varying $\pi(\sigma^2)$

Fig. 6 Percentage of $u^2_{n_\varepsilon} < \varepsilon^2$ for the sample sizes n_ε from Fig. 5 within the dataset from Fig. 4. Axes of both plots and the positions of the black crosses are the same as in Fig. 5. The "true" value within the full dataset is depicted by the dashed lines

1. A precise prior knowledge about the parameter of interest. This will result in small sample sizes
2. A conservative estimate of the parameter that determines the variation of the data. This will result in larger sample sizes. We have observed two different ways to achieve this:

   - Make a conservative (that is rather large) guess for the mean of the prior marginal w.r.t this parameter.
   - Choose a large enough standard deviation for this marginal prior. In order for this to work it is not sufficient to base the SSD on the average $\overline{u^2_n}$ only but higher moments as in the VPVC have to be taken into account.

In the Poisson example from Sect. 2.1 both strategies 1 and 2 concern the same parameter θ. For the case of normally distributed data we have observed that strategy 1 concerns the parameter of interest μ while to follow strategy 2 the parameter σ^2, that describes the variance of the data, was central.

3 Behavior for small ε

We have seen in Sect. 2 that the determination of the sample size via the VPVC can be successful, provided the prior knowledge is either precise enough or sufficiently conservative. Being "conservative" is of course a rather vague quality, but in this section we want to answer what happens if ε from (5) becomes smaller. Will it become easier or harder to be conservative? In Theorem 3.5 we show that as $\varepsilon \to 0$ the VPVC tends to satisfy (3) perfectly provided k is bigger then some threshold k_*. In other words for such a k the VPVC has a tendency to become conservative. If k is

smaller than k_* on the other hand we will show that, asymptotically, the uncertainty will almost surely be above the expression used for the VPVC. This phase transition with respect to k, compare Fig. 8 in the appendix, gives some insight on SSD based on the prior predictive in the small ε regime. We will give some ideas how to get an upper bound on k_*. Another result of this section is Lemma 3.4, where we show an asymptotic formula for the sample size n_ε determined by the VPVC.

First let us fix our setup. Similar as in Sect. 2 we will assume that there is a univariate parameter of interest θ_0 and a, possibly empty or multivariate, nuisance parameter θ', so that the total parameter that determines our sampling distribution $p(\mathbf{x}_n|\theta) = p(x_1, \ldots, x_n|\theta)$ is given by

$$\theta = (\theta_0, \theta'). \tag{11}$$

We will denote the Fisher information matrix of the single sample distribution $p(x_1|\theta)$ by I_θ and its first component (the one linked to θ_0) by I_{θ_0}. Note, that since θ_0 is univariate I_{θ_0} is a non-negative number. The knowledge about θ is described by the prior $\pi(\theta)$ with marginals $\pi(\theta_0)$ and $\pi(\theta')$. We further fix a value $\theta_{\text{true}} = (\theta_{\text{true},0}, \theta'_{\text{true}})$ in the support of $\pi(\theta)$ that we treat as the true parameter. In practice θ_{true} will of course be unknown. Given some data $\mathbf{x}_n = (x_1, \ldots, x_n)$ we introduce as above the squared uncertainty by $u_n^2 = \text{Var}_{\theta_0 \sim \pi(\theta_0|\mathbf{x}_n)}(\theta_0)$ as well as $\overline{u_n^2} = \mathbb{E}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}[u_n^2]$, $\Delta u_n^2 = (\text{Var}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}(u_n^2))^{1/2}$ with $m(\mathbf{x}_n) = \int p(\mathbf{x}_n|\theta)\pi(\theta)\mathrm{d}\theta$ and formulate the VPVC for $\varepsilon > 0$ as

$$\overline{u_n^2} + k\,\Delta u_n^2 < \varepsilon^2. \tag{12}$$

The smallest n satisfying (12) will be called once more n_ε.

Assumption 3.1 We assume that for $\pi(\theta)$ and θ_{true} it holds:

1. *Fisher information sufficiently regular* The Fisher information I_{θ_0} is strictly positive for $\theta = \theta_{\text{true}}$ and almost every θ in the domain of $\pi(\theta)$. Moreover we assume that the second moment $\mathbb{E}_{\theta \sim \pi(\theta)}[I_{\theta_0}^{-2}]$ exists and that $\text{Var}_{\theta \sim \pi(\theta)}(I_{\theta_0}^{-1}) > 0$.
2. *B.-v.-Mises limit holds in L^2* The quantity $n \cdot u_n^2$ converges in L^2 against $I_{\theta_0}^{-1}$ conditional under $\theta = \theta_{\text{true}}$ and $\pi(\theta)$ in the sense that:

$$\lim_{n\to\infty} \mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}[|n \cdot u_n^2 - I_{\theta_{\text{true}},0}^{-1}|^2] = 0 \text{ and}$$
$$\lim_{n\to\infty} \mathbb{E}_{\theta \sim \pi(\theta)}[\mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta)}[|n \cdot u_n^2 - I_{\theta_0}^{-1}|^2]] = 0. \tag{13}$$

3. *No trivial SSD* For all n we have $\overline{u_n^2} > 0$.

From these assumptions point 2 is probably the one that needs the most explanation. The Bernstein-von-Mises theorem (Van der Vaart 2000) indicates that under relatively mild conditions, and conditional on θ, the posterior distribution is asymptotically normal with variance $\frac{1}{n}I_{\theta_0}^{-1}$, which motivates (13). When considering variances, as in this paper, it is then quite natural to require convergence in L^2, while

the B.-v.-Mises theorem only ensures convergence in probability. While standard methods, such as Vitali's convergence theorem, could be applied to provide the first convergence (13) these seems more involved for the second limit as it is in a rather non-standard form. Examining this point in more detail would stray far away from the scope of this paper and could be subject to future research. However, we found that both identities (13) were rather easy to check for many standard cases such as the ones from Sect. 2 or a Bernoulli distribution with a Beta prior.

Under Assumptions 3.1 the posterior variance is, conditional on θ, asymptotically proportional to $I_{\theta_0}^{-1}$. The following quantity therefore describes the variation of the asymptotic posterior variance given the prior knowledge $\pi(\theta)$:

$$\gamma = \frac{(\mathrm{Var}_{\theta \sim \pi(\theta)}(I_{\theta_0}^{-1}))^{1/2}}{\mathbb{E}_{\theta \sim \pi(\theta)}[I_{\theta_0}^{-1}]}. \tag{14}$$

Let us make these considerations more precise.

Lemma 3.2 *Define the coefficient of variation under the prior predictive $m(\mathbf{x}_n)$ and under θ_{true} as*

$$c_n = \frac{\Delta u_n^2}{\overline{u_n^2}} \quad \text{and} \quad c_{\mathrm{true},n} = \frac{(\mathrm{Var}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\mathrm{true}})}(u_n^2))^{1/2}}{\mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\mathrm{true}})}[u_n^2]}.$$

Under Assumptions 3.1 we then have

$$\lim_{n \to \infty} c_n = \gamma \quad \text{and} \quad \lim_{n \to \infty} c_{\mathrm{true},n} = 0.$$

Remark 3.3 The limit of $c_{\mathrm{true},n}$ is a special case of the one for c_n by taking the delta distribution $\delta_{\theta_{\mathrm{true}}}$, centered at θ_{true}, as a prior.

Proof Let us first consider the limit of c_n. Note that the second condition of point 2 in Assumption 3.1 states that $X_n := n u_n^2$ converges to $I_{\theta_0}^{-1}$ in L^2 with respect to the law $p(\theta, \mathbf{x}_n) = \pi(\theta) \cdot p(\mathbf{x}_n|\theta)$. Since X_n is not dependent on θ, I_{θ_0} is not dependent on $\mathbf{x}_n$ and the marginals of $p(\theta, \mathbf{x}_n)$ are $\pi(\theta)$ and $m(\mathbf{x}_n)$ we conclude that

$$\mathrm{Var}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}(X_n) = \mathrm{Var}_{(\theta,\mathbf{x}_n) \sim p(\theta,\mathbf{x}_n)}(X_n)$$
$$\to \mathrm{Var}_{(\theta,\mathbf{x}_n) \sim p(\theta,\mathbf{x}_n)}(I_{\theta_0}^{-1}) = \mathrm{Var}_{\theta \sim \pi(\theta)}(I_{\theta_0}^{-1})$$

and similar $\mathbb{E}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}[X_n] \to \mathbb{E}_{\theta \sim \pi(\theta)}[I_{\theta_0}^{-1}]$. From this we obtain the claimed limit of c_n:

$$c_n \overset{X_n = n u_n^2}{=} \frac{(\mathrm{Var}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}(X_n))^{1/2}}{\mathbb{E}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}[X_n]} \to \frac{(\mathrm{Var}_{\theta \sim \pi(\theta)}(I_{\theta_0}^{-1}))^{1/2}}{\mathbb{E}_{\theta \sim \pi(\theta)}[I_{\theta_0}^{-1}]} = \gamma.$$

For the second part of the claim we use that by the first condition in point 2 of Assumption 3.1 we have L^2 convergence of X_n conditional on θ_{true} from which we obtain indeed

$$c_{\text{true},n} = \frac{(\text{Var}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}(X_n))^{1/2}}{\mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}[X_n]} \to \frac{(\text{Var}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}(I_{\theta_{\text{true},0}}^{-1}))^{1/2}}{\mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}[I_{\theta_{\text{true},0}}^{-1}]} = 0,$$

where we used on the right hand side that $I_{\theta_{\text{true},0}}$ does not depend on $\mathbf{x}_n$. $\qquad\square$

We see from Lemma 3.2 that the variation of u_n^2 under the marginal $m(\mathbf{x}_n)$ is behaving in a different manner as under the true parameter θ_{true}. The object $\Delta u_n^2 = (\text{Var}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}(u_n^2))^{1/2}$ decays at the same order as $\overline{u_n^2} = \mathbb{E}_{\mathbf{x}_n \sim m(\mathbf{x}_n)}[u_n^2]$ whereas, being conditional on a parameter, the standard deviation $(\text{Var}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}(u_n^2))^{1/2}$ decays faster than $\mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}[u_n^2]$. The distributions $m(\mathbf{x}_n)$ and $p(\mathbf{x}_n|\theta_{\text{true}})$ result in quite different asymptotic behavior for the moments of the posterior variance. We will see in Theorem 3.5 below that these different asymptotic properties will provoke some phase transition with respect to k in the limit $\varepsilon \to 0$.

As a first consequence of the observations from Lemma 3.2 we will now see that the variance term $k\,\Delta u_n^2$ in the VPVC will strongly affect the sample size even as $\varepsilon \to 0$.

Lemma 3.4 (VPVC sample size for small ε) *Provided Assumption 3.1 is true, we have*

$$\lim_{\varepsilon \to 0} \frac{n_\varepsilon}{\varepsilon^{-2} \cdot (1 + k\,\gamma)\mathbb{E}_{\theta \sim \pi(\theta)}[I_{\theta_0}^{-1}]} = 1 \tag{15}$$

Proof We have used already in the first part of the proof of Lemma 3.2 that $n\overline{u_n^2} \to \mathbb{E}_{\theta \sim \pi(\theta)}[I_{\theta_0}^{-1}]$ so that applying in addition the convergence $c_n \to \gamma$ from Lemma 3.2 we arrive at

$$s_n := n\left(\overline{u_n^2} + k\,\Delta u_n^2\right) = n\overline{u_n^2}(1 + k\,c_n) \overset{n\to\infty}{\longrightarrow} (1 + k\,\gamma)\mathbb{E}_{\theta \sim \pi(\theta)}[I_{\theta_0}^{-1}] =: s_\infty. \tag{16}$$

Next, observe that since $\overline{u_n^2} > 0$ by point 3 of Assumption 3.1 this implies that we can find a constant $c > 0$ such that for any n we have $\frac{c}{n} < \overline{u_n^2} + k\Delta u_n^2$, from which we get $n_\varepsilon > c \cdot \varepsilon^{-2}$ and in particular

$$\lim_{\varepsilon \to 0} n_\varepsilon = \infty. \tag{17}$$

Let us rewrite the claim of the lemma as

$$\lim_{\varepsilon \to 0} \frac{n_\varepsilon}{\varepsilon^{-2} \cdot s_\infty} = 1. \tag{18}$$

From (16), and the choice of n_ε, we already have one inequality of (18):

$$\liminf_{\varepsilon \to 0} \frac{n_\varepsilon}{\varepsilon^{-2} s_\infty} \overset{\text{def. of } n_\varepsilon}{\geq} \liminf_{\varepsilon \to 0} \frac{n_\varepsilon \cdot (\overline{u_{n_\varepsilon}^2} + k\Delta u_{n_\varepsilon}^2)}{s_\infty} = \liminf_{\varepsilon \to 0} \frac{s_{n_\varepsilon}}{s_\infty} \overset{(16),(17)}{=} 1.$$

To show the opposite inequality we argue by contradiction. Suppose that there is a $\delta > 0$ such that

$$\limsup_{\varepsilon \to 0} \frac{n_\varepsilon}{\varepsilon^{-2} \cdot s_\infty} \geq 1 + 2\,\delta. \tag{19}$$

Define $m_\varepsilon = \lfloor s_\infty \varepsilon^{-2} \rfloor$, the largest integer less equal $s_\infty \varepsilon^{-2}$, and note that for ε small enough there is an integer n'_ε such that the following nested inequality is true

$$1 \leq m_\varepsilon < (1 + \delta/2)(1 + m_\varepsilon) \leq n'_\varepsilon \leq m_\varepsilon (1 + \delta) \overset{(19)}{<} n_\varepsilon. \tag{20}$$

Choose further ε small enough such that for any $n > m_\varepsilon$ we have $s_n/s_\infty < (1 + \delta/2)$. But then the following inequality holds:

$$\overline{u_{n'_\varepsilon}^2} + k \cdot \Delta u_{n'_\varepsilon}^2 = \frac{s_{n'_\varepsilon}}{n'_\varepsilon} \overset{n'_\varepsilon > m_\varepsilon}{<} \frac{(1 + \delta/2) \cdot s_\infty}{n'_\varepsilon} \overset{(20)}{\leq} \frac{s_\infty}{1 + \lfloor s_\infty \varepsilon^{-2} \rfloor} \leq \varepsilon^2.$$

By (20) we have thus found a smaller integer $1 \leq n'_\varepsilon < n_\varepsilon$ that satisfies the VPVC (12), which contradicts the choice of n_ε. $\square$

For the setup considered in Sect. 2.2 we sketched the convergence of the object from Lemma 3.4 for various priors in Fig. 7a.

The formula from Lemma 3.4 allows for a few interesting observations. First, we see that asymptotically the prior $\pi(\theta)$ has an impact on the sample size through $\mathbb{E}_{\theta \sim \pi(\theta)}[I_{\theta_0}^{-1}]$ (the expectation of the rescaled B.-v.-Mises limit of the posterior variance) and through the coefficient of variation γ. The higher the latter the more impact will the choice of k have on the sample size. Moreover, comparing the sample sizes n_ε from the VPVC and $\tilde{n}_\varepsilon$ from the APVC ($k = 0$) we see that taking the variance into account will increase the sample size by a factor of $\frac{n_\varepsilon}{\tilde{n}_\varepsilon} \simeq 1 + k \cdot \gamma$. Finally, note that we do not need any explicit expression for u_n^2 when using the asymptotic expression of n_ε. Given the Fisher information I_{θ_0} of the sampling distribution and any prior $\pi(\theta)$ we can directly compute the sample size for small ε without any need of computing the posterior distribution.

We will now turn to the main result of this section. In Sect. 2 we have evaluated the "success" of the SSD by evaluating how often we will have $u_{n_\varepsilon} < \varepsilon$ on the actual dataset, recall for instance Fig. 3b. In the following we want to do this for $\varepsilon \to 0$ in a purely generic setup by assuming that the data $\mathbf{x}_n$ follows the distribution $p(\mathbf{x}_n|\theta_{\text{true}})$ with the true parameter θ_{true}. We have already observed in Lemma 3.2 that conditioning on a parameter like θ_{true} leads to a different asymptotic behavior than considering the marginal $m(\mathbf{x}_n)$. The following theorem shows the consequence of this disparity concerning the success of the SSD.

Theorem 3.5 (VPVC becomes conservative for $k > k_*$ and small ε) *Assume that Assumptions* 3.1 *hold and define*

$$k_* = k_*(\theta_{\text{true}}) := \frac{\max(I_{\theta_{\text{true},0}}^{-1} - \mathbb{E}_{\theta \sim \pi(\theta)}[I_{\theta_0}^{-1}], 0)}{(\text{Var}_{\theta \sim \pi(\theta)}(I_{\theta_0}^{-1}))^{1/2}}. \tag{21}$$

We then have for any $k > k_*$

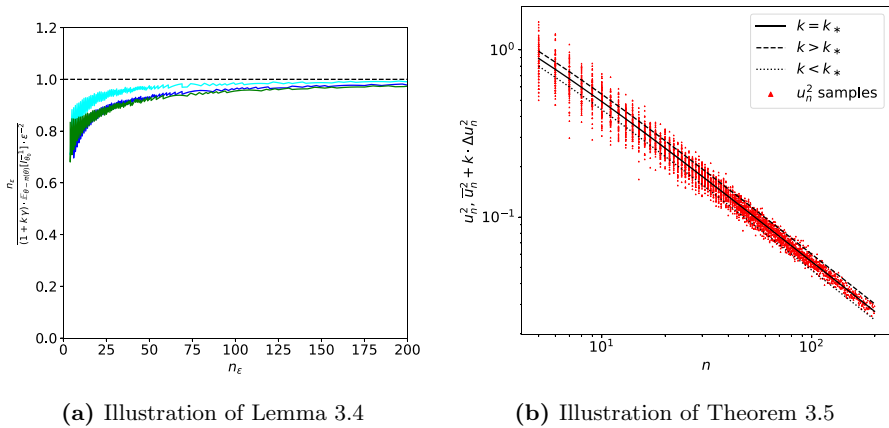**(a)** Illustration of Lemma 3.4   **(b)** Illustration of Theorem 3.5

Fig. 7 Convergence of the expression from Lemma 3.4 forthe setup from Sect. 2.2 $k = 2$ various priors$\pi(\sigma^2)$ and the fixed prior $\pi(\mu)$marked in Fig 5a with mean and standard deviation -1.4 ± 0.6. Choices for$\pi(\sigma^2)$: *blue* 0.30 ± 0.10, *cyan*: 0.15 ± 0.10, *green*: 0.20 ± 0.05. Illustration of Theorem 3.5 for an actual dataset. Samples of u_n^2 (in red) for different sample sizes using the data and setup from Sect 2.1 and a prior $\pi(\theta)$with mean and standard deviation 4.0 ± 2.0. The black lines mark $\overline{u_n^2} + k\Delta u_n^2$ for $k = k_* = 0.73$ (computed from the full dataset) and $k = k_* \pm 0.30$

$$\lim_{n \to \infty} \mathbb{P}_{\mathbf{x}_n \sim p(\mathbf{x}_n | \theta_{\text{true}})} \left(u_n^2 > \overline{u_n^2} + k \, \Delta u_n^2 \right) = 0. \tag{22}$$

In particular, we have for any SSD built on the VPVC (12) *with such a k*

$$\lim_{\varepsilon \to 0} \mathbb{P}_{\mathbf{x}_{n_\varepsilon} \sim p(\mathbf{x}_{n_\varepsilon} | \theta_{\text{true}})} \left(u_{n_\varepsilon}^2 \geq \varepsilon^2 \right) = 0. \tag{23}$$

Moreover, k_* is optimal in the sense that for any $0 < k < k_*$ we have

$$\lim_{n \to \infty} \mathbb{P}_{\mathbf{x}_n \sim p(\mathbf{x}_n | \theta_{\text{true}})} \left(u_n^2 < \overline{u_n^2} + k\Delta u_n^2 \right) = 0. \tag{24}$$

Proof We introduce $\rho = \dfrac{I_{\theta_{\text{true}},0}^{-1}}{\mathbb{E}_{\theta \sim \pi(\theta)}[I_{\theta_0}^{-1}]} - 1$ so that k_* can be written as $k_* = \max(\rho, 0)/\gamma$ with γ as in Lemma 3.2. Similar as in the proof of Lemma 3.2 we see that

$$\rho = \lim_{n \to 0} \frac{\mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n | \theta_{\text{true}})}[u_n^2]}{\overline{u_n^2}} - 1. \tag{25}$$

Let us first show (22). We can rewrite the inequality inside the probability as

$$u_n^2 - \mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n | \theta_{\text{true}})}[u_n^2] > \overline{u_n^2} - \mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n | \theta_{\text{true}})}[u_n^2] + k\Delta u_n^2. \tag{26}$$

The expression on the right hand side becomes finally positive when n is high enough. Indeed, we can we reshape it, using point 3 of Assumption 3.1 and the object c_n from Lemma 3.2, as

$$\overline{u_n^2} \cdot \left(1 - \frac{\mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}[u_n^2]}{\overline{u_n^2}} + kc_n\right) =: \overline{u_n^2} \cdot h_n$$

where h_n satisfies $\lim_{n\to\infty} h_n = -\rho + k \cdot \gamma > 0$ due to Lemma 3.2, (25) and the choice $k > k_* = \max(\rho, 0)/\gamma$. For n large enough such that the right hand side of (26) is positive, we can apply Chebyshev's inequality, which yields

$$\limsup_{n\to\infty} \mathbb{P}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}\left(u_n^2 > \overline{u_n^2} + k\Delta u_n^2\right)$$

$$\leq \limsup_{n\to\infty} \frac{\text{Var}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}(u_n^2)}{(\overline{u_n^2} - \mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}[u_n^2] + k\Delta u_n^2)^2}$$

$$= \limsup_{n\to\infty} c_{\text{true},n}^2 \cdot \frac{\mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}[u_n^2]^2}{\overline{u_n^2} \cdot h_n^2} \leq \lim_{n\to\infty} c_{\text{true},n}^2 \cdot \left(\frac{1+\rho}{-\rho + k\gamma}\right)^2 = 0,$$

where we used once more (25) and that we know from Lemma 3.2 that $c_{\text{true},n}$ converges to 0. From (22) we immediately obtain (23) by the definition of n_ε and the fact that $n_\varepsilon \to \infty$, due to Lemma 3.4. To show (24) we reshape the expression inside the parentheses of (24) into

$$\mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}[u_n^2] - u_n^2 > \mathbb{E}_{\mathbf{x}_n \sim p(\mathbf{x}_n|\theta_{\text{true}})}[u_n^2] - (\overline{u_n^2} + k\Delta u_n^2). \tag{27}$$

The rest of the argument then follows along the lines of the first part of the proof. □

Theorem 3.5 reveals a sort of phase transition. For a k smaller than some threshold k_* the squared uncertainty u_n^2 will asymptotically be above $\overline{u_n^2} + k\Delta u_n^2$, while for $k > k_*$ we will have $u_n^2 < \overline{u_n^2} + k\Delta u_n^2$ with probability almost 1 for large n (or small ε). This effect is shown in Fig. 7b for the Leeds accident data from Sect. 2.1 and a fixed prior.

The behavior predicted by Theorem 3.5 might be surprising at first glance: recall that $\overline{u_n^2}$ is defined as an average and that the object Δu_n^2 is the according standard deviation. Naively one might expect that the range of k times the standard deviation around the mean might always cover some percentage strictly between 0 and 1. The reason behind the observed phenomenon is that we use one distribution, $m(\mathbf{x}_n)$, to perform the SSD and another one, $p(\mathbf{x}_n|\theta_{\text{true}})$, to evaluate its success. The different asymptotics between these two distributions then gives rise to the erratic behavior around $k = k_*$. The effect for a fixed k and various priors is depicted by Fig. 8 in the appendix: As ε gets small we get two sharply separated regions for the hyperparameters where one region has a quota of $u_{n_\varepsilon}^2 < \varepsilon^2$ around 0% while the other one approaches 100%. We expect that this phenomenon is not solely adherent to the VPVC. For instance, it seems reasonable that for many cases a credible interval for u_n^2 under $m(\mathbf{x}_n)$ is, at least loosely, related to a certain k so that similar effects are likely to arise for a framework based on other approaches using $m(\mathbf{x}_n)$. Investigating such a point further might be an interesting subject for future research.

We have seen in Theorem 3.5 that for small ε the success of the SSD crucially depends on having $k > k_*$. Now, k_* is dependent on the true parameter θ_{true} and thus

unknown in practice. For the prior used for Fig. 7b we have a k_* around 0.21 and therefore much smaller than the $k = 2$ used in the analysis of Sect. 2.1. In general one might choose for an area Θ in the domain of $\pi(\theta)$ as an upper bound for k_*

$$k_*^{up.b.} = \sup_{\theta \in \Theta} k_*(\theta). \qquad (28)$$

Taking a single standard deviation around the prior mean for $\pi(\theta)$ and $\pi(\sigma^2)$ respectively we get for the prior from Fig. 7b $k_*^{up.b.} = 1.7$ and for the one marked by the black cross in Fig. 6b $k_*^{up.b.} = 1.0$, which are both below the choice $k = 2$ we used in Sect. 2. The actual values of $k_\star$, with "true" parameter estimated from the full dataset, are 0.73 and 0.75 respectively.

4 Conclusions

We discussed a Bayesian criterion for sample size determination, based on the prior predictive, which we called the variation of posterior variance criterion (VPVC). Compared with the classical average posterior variance criterion (APVC) this criterion leads to a better compliance with the objective to restrain the uncertainty by some ε, while still providing explicit expressions in contrast to a full treatment of the law of the uncertainty under the prior predictive. In particular this allows to treat the asymptotic behavior for small ε in a generic manner and thus to enhance the understanding of sample size methods based on the prior predictive.

Using two different datasets we discussed the dependency of the sample size determination and its success on the chosen prior and the true parameter and deduced concepts on how to choose the prior to lead the sample size determination more likely to a success. In particular, we observed that a part of these strategies cannot be applied for the APVC but only for methods such as the VPVC that take higher moments with respect to the prior predictive into account.

Finally, we gave some results concerning the behavior of the VPVC for $\varepsilon \to 0$. We proved an explicit formula for the predicted sample size in this regime and showed that there is an exact limit for the portion of the variance w.r.t. prior predictive that can be used in order to guarantee the success of the sample size determination in this limit.

Appendix

See Figs. 8 and 9.

(a) $\varepsilon = 0.20$

(b) $\varepsilon = 0.10$

(c) $\varepsilon = 0.05$
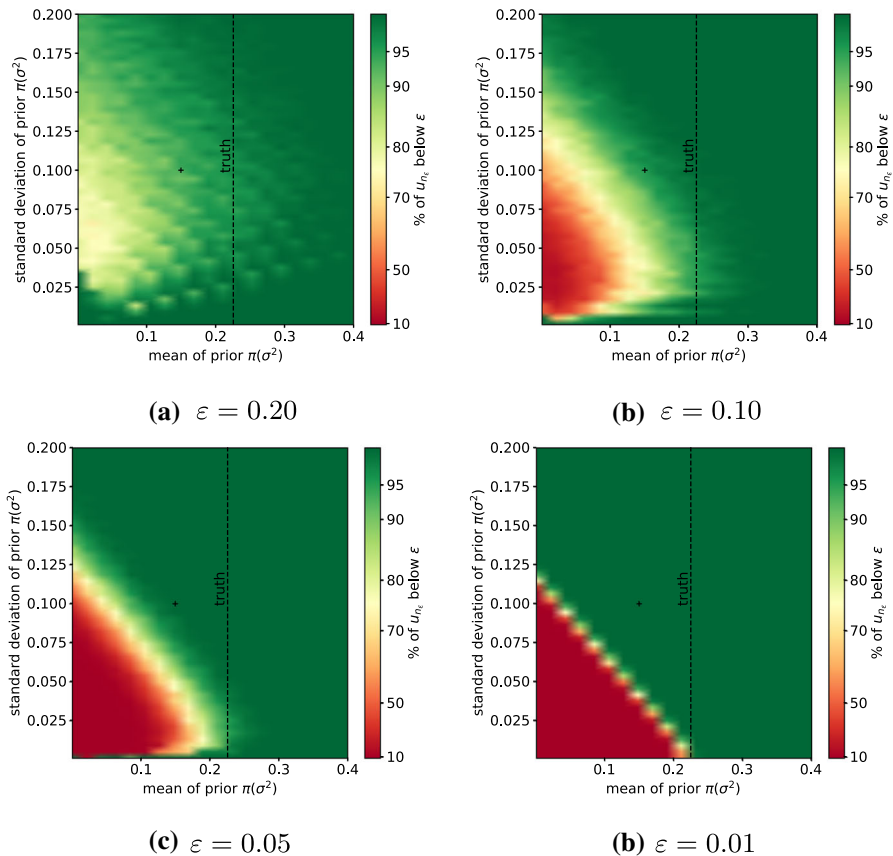
(b) $\varepsilon = 0.01$

Fig. 8 Percentage of $u_{n_\varepsilon} < \varepsilon$ for the setup and data from Sect. 2.2 for various ε (in decreasing order), $k = 2$, various marginals $\pi(\sigma^2)$ and the same fixed marginal $\pi(\mu)$ as in Fig. 6b. The "true" value of σ^2 (computed from the dataset) is depicted by the dashed lines. We used a smaller resolution compared to Fig. 6b for numerical reasons

Fig. 9 Left column: Relative deviation $(n_\varepsilon - n_\varepsilon^{\mathbb{P}})/n_\varepsilon^{\mathbb{P}}$ between the sample sizes n_ε from the VPVC and $n_\varepsilon^{\mathbb{P}}$ ▶ predicted by (10) in steps of 5% for the setup from Sect. 2.1 and various choices of p in (10). Right column: Percentage of $u_{n_\varepsilon^{\mathbb{P}}} < \varepsilon$ for the data from Fig. 1
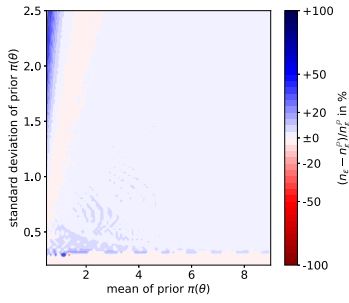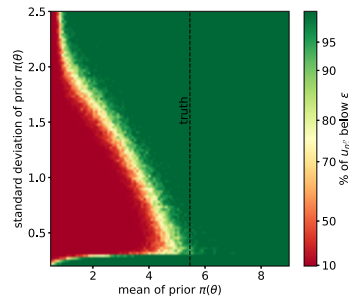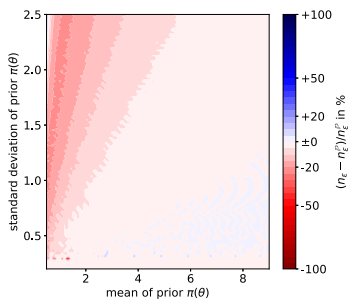
(a) $p = 0.95$
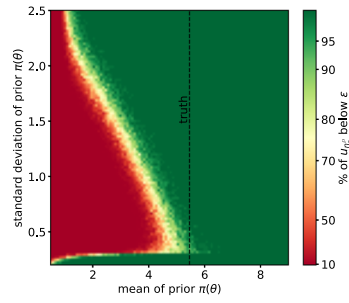
(b) $p = 0.95$
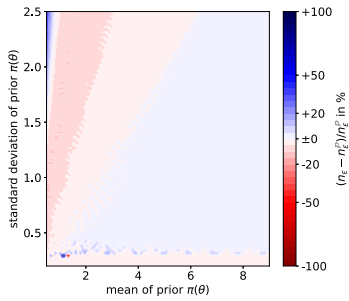
(c) $p = 0.96$

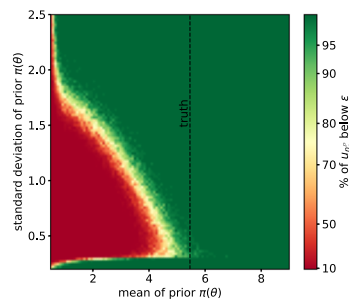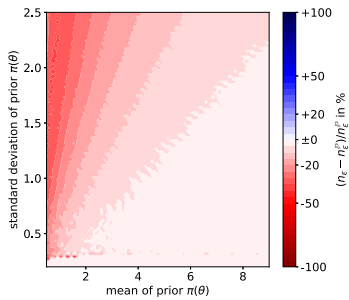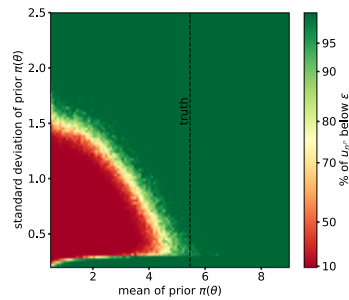**(d)** $p = 0.96$

(e) $p = 0.97$

(f) $p = 0.97$

(g) $p = 0.98$

(h) $p = 0.98$

References

Adcock C (1988) A bayesian approach to calculating sample sizes. J R Stat Soc Ser D (The Statistician) 37(4–5):433–439

Adcock C (1997) Sample size determination: a review. J R Stat Soc Ser D (The Statistician) 46(2):261–283

Brutti P, De Santis F, Gubbiotti S (2008) Robust Bayesian sample size determination in clinical trials. Stat Med 27(13):2290–2306

Charan J, Kantharia N (2013) How to calculate sample size in animal studies? J Pharmacol Pharmacother 4(4):303

Chen M-H, Ibrahim JG, Lam P, Yu A, Zhang Y (2011) Bayesian design of noninferiority trials for medical devices using historical data. Biometrics 67(3):1163–1170

De Santis F (2006) Sample size determination for robust bayesian analysis. J Am Stat Assoc 101(473):278–291

De Santis F (2007) Using historical data for Bayesian sample size determination. J R Stat Soc Ser A (Statistics in Society) 170(1):95–113

De Santis F, Pacifico MP (2004) Two experimental settings in clinical trials: predictive criteria for choosing the sample size in interval estimation. Applied Bayesian statistical studies in biology and medicine. Springer, Boston, pp 109–130

Dell RB, Holleran S, Ramakrishnan R (2002) Sample size determination. ILAR J 43(4):207–213

Desu M (2012) Sample size methodology. Elsevier, Amsterdam

Fink D (1997) A compendium of conjugate priors. Tech. Rep

Grillitsch B, Schiesari L (2010) The ecotoxicology of metals in reptiles. In: Sparling D, Linder G, Bishop C, Krest S (eds) Ecotoxicol Amphib Reptiles. https://doi.org/10.1201/EBK1420064162

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(4):443–460

Hadayeghi A, Shalaby AS, Persaud BN (2010) Development of planning level transportation safety tools using geographically weighted poisson regression. Accid Anal Prev 42(2):676–688

Joseph L, Belisle P (1997) Bayesian sample size determination for normal means and differences between normal means. J R Stat Soc Ser D (The Statistician) 46(2):209–226

Joseph L, Bélisle P (2019) Bayesian consensus-based sample size criteria for binomial proportions. Stat Med 38(23):4566–4573

Joseph L, Wolfson DB, Berger RD (1995) Sample size calculations for binomial proportions via highest posterior density intervals. J R Stat Soc Ser D (The Statistician) 44(2):143–154

Lawson A, Jodice P (2019) Mercury concentrations in American alligators in South Carolina, 2010–2017. U.S, Geological Survey data release

M'lan CE, Joseph L, Wolfson DB et al (2008) Bayesian sample size determination for binomial proportions. Bayesian Anal 3(2):269–296

Nilsen F, Lawson A, Lowers R, Jodice P (2019) Total mercury concentrations in whole blood of American alligators in South Carolina and Florida. Mendeley Data 1. https://doi.org/10.17632/vycpgt6ycf.1

Ott WR (1990) A physical explanation of the lognormality of pollutant concentrations. J Air Waste Manag Assoc 40(10):1378–1383

Pham-Gia T, Turkkan N (1992) Sample size determination in Bayesian analysis. J R Stat Soc Ser D (The Statistician) 41(4):389–397

Psioda MA, Ibrahim JG (2019) Bayesian clinical trial design using historical data that inform the treatment effect. Biostatistics 20(3):400–415

Road traffic accidents in 2018 (2019) Published by Leeds City Council, Available on https://data.gov.uk. Accessed 01 June 2020

Rubin DB, Stern HS (1998) Sample size determination using posterior predictive distributions. Sankhyā Indian J Stat Ser B 60:161–175

Rumbold D, Fink L, Laine K, Niemczyk S, Chandrasekhar T, Wankel SD, Kendall C (2002) Levels of mercury in alligators (alligator mississippiensis) collected along a transect through the florida everglades. Sci Total Environ 297(1–3):239–252

Sahu S, Smith T (2006) A Bayesian method of sample size determination with practical applications. J R Stat Soc Ser A (Statistics in Society) 169(2):235–253

Sambucini V (2008) A bayesian predictive two-stage design for phase ii clinical trials. Stat Med 27(8):1199–1224

Santis FD, Pacifico MP, Sambucini V (2004) Optimal predictive sample size for case-control studies. J R Stat Soc Ser C (Applied Statistics) 53(3):427–441

Van der Vaart AW (2000) Asymptotic statistics, vol 3. Cambridge University Press, Cambridge

Wang F, Gelfand AE et al (2002) A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. Stat Sci 17(2):193–208