On the discrepancy between Bayes credibility and frequentist probability of coverage

Asma Bahamyirou & Éric Marchand[a]

a Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, J1K 2R1

(e-mails: eric.marchand@usherbrooke.ca; asma.bahamyirou@usherbrooke.ca)

Abstract. We consider interval estimation of a bounded normal mean $\theta \in [a, b]$ with known variance σ^2. We establish properties of both Bayes credible intervals with fixed credibility and interval estimators with confidence level $1-\alpha$. For a large class of Bayes credible intervals, we show that the minimal frequentist coverage probability must be bounded above by $\Phi(\frac{b-a}{\sigma})$, which is indicative of a significant discrepancy between the credibility and the coverage probability when the former is large and $\frac{b-a}{\sigma}$ is small. Analogously, for a large class of interval estimators with confidence level $1-\alpha$, included previously studied estimators in the literature, and whenever $\frac{b-a}{2\sigma} < \Phi^{-1}(1-\alpha)$, we show that the Bayesian credibility must be equal to one for some values of the sample space indicative of the inability of such estimators to report a strict subset of the parameter space $[a, b]$ with probability one.

Keywords: Bayesian methods, Bounded normal mean, Credibility, Frequentist coverage probability, Interval estimation.

1. Introduction

This paper is concerned with both: **(i)** the frequentist probability of coverage of Bayes credible sets, and **(ii)** the Bayes credibility of exact frequentist coverage methods, for estimating a normal mean θ bounded to an interval $[a, b]$, and based on $X \sim N(\theta, \sigma^2)$ with known σ^2 (a sample of size one without loss of generality). With respect to **(i)**, we prove (i.e., Theorem 1) the existence of a discrepancy, which can significant, between a given credibility $1-\alpha$ and the minimal frequentist probability coverage. Indeed, we show that the latter is quite a bit lower, for a very large class (essentially all priors with a symmetric density about 0) of priors supported on $[a, b]$ and choice of Bayes credible set, whenever the relative width $m = \frac{b-a}{2\sigma}$ is small and $1-\alpha$ is not small. With respect to **(ii)**, we show that confidence intervals $I(x) = [l(x), u(x)]$ which are equivariant with respect to a sign change (i.e., $l(x) = -u(-x)$ for all x), which have monotone increasing in x endpoints $l(x)$

and $u(x)$, and which have exact coverage probability $1-\alpha$ must be equal to the full parameter space $[a,b]$ for a range of x values, (i.e., have credibility equal to one for any prior), whenever $m=\frac{b-a}{2\sigma}<\Phi^{-1}(1-\alpha)$. So exact coverage comes at the price of this unattractive feature. The particular cases of the truncated Pratt interval (Evans, Stark and Hansen, 2005) and the so-called "unified method" (Feldman and Cousins, 1998) which is the inversion of a likelihood ratio test, serve as illustrations for **(ii)**, while the Bayes HPD credible set with respect to the uniform prior on $[a,b]$ serves as an illustration for **(i)**.

The negative results for **(i)** stand in contrast to a series of findings for a lower bounded space of the form $\theta \geq 0$ which limit the discrepancy and which also apply to a wide array of situations (e.g., Mandelkern, 2002; Roe and Woodroofe, 2000; Zhang and Woodroofe, 2002, 2003; Marchand and Strawderman, 2006, 2013; Marchand et al., 2008). The context of our findings relates to the intrinsic interest in findings objective priors with near probability matching properties without relying on asymptotics and in the presence of a bounded parameter space. The context and motivation was described by Marchand and Strawderman (2013) as follows:

> Bayesian credible sets are not designed (e.g., Robert, 2011) and are far from guaranteed (Fraser, 2011) to have satisfactory, exact or precise frequentist coverage but it is nevertheless of interest to investigate (Wasserman, 2011) to what extent there is convergence or divergence in various situations.

With respect to **(ii)**, it is useful to have available procedures, adapted to the parameter constraint and that may even be optimal in a certain sense (e.g., the truncated Pratt interval has a minimax interpretation), that guarantee exact coverage, but we believe that the drawback of having to report interval estimates equal to the full parameter space for some observed values is not well understood or not known. [1]

Subsections 2.1 and 2.2 contains the results corresponding to **(i)** and **(ii)** respectively, with remarks and illustrations in complement. Final remarks conclude the presentation.

[1] The opposite situation where the credibility is equal to 0 for some values of x is more familiar, and arises for the "standard" confidence interval $X \pm \sigma\Phi^{-1}(1-\alpha)$ truncated to $[a,b]$, for $x > b+\sigma\Phi^{-1}(1-\alpha)$ or $x < a-\sigma\Phi^{-1}(1-\alpha)$.

2. Main Results and Illustrations

2.1. On Bayesian confidence intervals

Without loss of generality, we assume hereafter $b = -a = m$ and $\sigma = 1$. Here is the first main result concerning the frequentist probability of coverage of Bayes credible sets.

Theorem 1. *Let $X|\theta \sim N(\theta, 1)$ with $|\theta| \leq m$. Let π be a continuous prior proper density for θ supported on $[-m, m]$ which is an even function. Let $I_\pi(X) = [l_\pi(X), u_\pi(X)]$ be a $1-\alpha$ Bayes credible set associated with π such that the endpoints are non-decreasing as a function of x and satisfy an equivariance property $l_\pi(-x) = -u_\pi(x)$ for all x. Then, we must have*

$$\inf_{\theta \in [-m,m]} C_{\mathbb{I}_\pi}(\theta) \leq \Phi(m)\,, \tag{1}$$

where Φ is the $N(0,1)$ cdf, and $C_{\mathbb{I}_\pi}(\theta) = \mathbb{P}(I_\pi(X) \ni \theta|\theta)$ is the frequentist probability of coverage.

Proof. We make use of the fact that a Bayes credible set $I_\pi(x)$ must be a strict subset of $[-m, m]$ for all x. [2] Observe that we must have $l_\pi(x) > -m$ for $x > 0$ since, otherwise we would have for $u_\pi(-x) = -l_\pi(x) = m$ by symmetry, $l_\pi(-x) = -m$ given the non-decreasing property, and corresponding credibility for such an x equal to 1 which is not allowed. Since, $l_\pi(x) > -m$ for $x > 0$, the probability of non-coverage at $\theta = -m$ is bounded below by $P_{-m}(X > 0) = 1 - \Phi(m)$, which leads to the result. □

Remark 1. *Depending on the values of m and α, but certainly for m not too large and $1-\alpha$ not too small, the above result is indicative of a possibly substantial discrepancy between Bayes credibility and frequentist probability of coverage. For instance, if the credibility is equal to $1-\alpha = 0.95$ and (i) $m = 1$, (ii) $m = 0.5$ (i.e., the mean θ is known to within* **(i)** *one,* **(ii)** *one half standard deviation), we have a minimal probabilities of coverage bounded above by $\Phi(1) \approx 0.84$ and $\Phi(0.5) \approx 0.69$*

[2] In terms of the plausibility of the assumptions, with the problem being invariant with respect to sign changes, symmetric priors lead naturally to equivariant interval estimators $I(X) = [l(X), u(X)]$ such that $l(-x) = -u(x)$ for all x. Also, with the family of $N(\theta, 1)$ distributions possessing an increasing in X monotone likelihood ratio with parameter θ and with the model densities satisfying the property $f(x-\theta) = f(\theta - x)$ for all x, θ, it follows that the family of posterior distributions $\theta|x$ possesses also an increasing in θ monotone likelihood ratio with parameter x. This tells us that the class of Bayes credible sets $I_\pi(X)$ with non-decreasing endpoints is of primordial interest.

respectively (in fact, these upper bounds hold for all $1-\alpha$), illustrating the possible degrees of discrepancy.

Remark 2. *When the parameter space is larger, the upper bound of course does not signal a discrepancy and there is good reason to believe that several choices of π and $I_\pi(X)$ may yield satisfactory level of matching between credibility and coverage probability. Indeed, for the analogue to the large m case of a lower bound constraint $\theta \geq a$ for some a, and the choices of the π_U uniform prior on $[a,\infty)$ and the HPD credible set, Roe and Woodroofe (2000) established the lower bound $\frac{1-\alpha}{1+\alpha}$ for frequentist coverage $C_{\pi_U}(\theta)$ limiting the amount of discrepancy. In fact, for $\alpha < 1/3$, the more precise statements: $1-\frac{3\alpha}{2} \leq C_{\pi_U}(\theta) \leq 1-\frac{\alpha}{2}$ for all $\theta \geq a$, and $\inf_{\theta\geq a} \in [1-\frac{3\alpha}{2}, 1-\frac{3\alpha}{2}+\frac{\alpha^2}{1+\alpha}]$ were given by Marchand et al. (2008). And, as well, the lower bound $\frac{1-\alpha}{1+\alpha}$ arises in a vast number of settings and for a class of Bayes credible sets associated with the truncation of a non-informative prior (i.e., right Haar invariant prior), such as π_U in the above illustration (see Marchand and Strawderman, 2006, 2013).*

Remark 3. *It is important to realize that the average frequentist coverage of probability of a $(1-\alpha)\times 100\%$ Bayesian credible set $I_\pi(X)$ with respect to π is equal to the credibility $1-\alpha$. Noting $f(\cdot|\theta)$ the model density and f_X the marginal density equal to $f_X(x) = \int_{(-m,m)} f(x|\theta)\,\pi(\theta)\,d\theta$, this follows since*

$$\begin{aligned}
\int_{(-m,m)} C_{\mathbb{I}_\pi}(\theta)\,\pi(\theta)\,d\theta &= \int_{(-m,m)}\int_{\mathbb{R}} \mathbb{I}\left(\mathbb{I}_\pi(x)\ni\theta\right) f(x|\theta)\,\pi(\theta)\,dx\,d\theta \\
&= \int_{\mathbb{R}}\left(\int_{(-m,m)} \mathbb{I}\left(\mathbb{I}_\pi(x)\ni\theta\right)\pi(\theta|x)\,d\theta\right) f_X(x)dx \\
&= \int_{\mathbb{R}} (1-\alpha)\,f_X(x)\,dx = 1-\alpha\,.
\end{aligned}$$

Paired with Theorem 1's maximal upper bound for minimal coverage, it must not only be the case that the frequentist coverage $C_{\mathbb{I}_\pi}(\theta)$ fluctuates to some extent below and above the credibility, but also to a rather large extent when $\Phi(m)$ is quite a bit smaller than the credibility $1-\alpha$. Figure 1 is illustrative of this for varying m, for the uniform prior density π_U on $[-m,m]$, and the $(1-\alpha)\times 100\%$ HPD credible set I_{π_U} (say). The associated posterior density is a (unimodal) truncated to $[-m,m]$ $N(x,1)$ density and it is verified in a straightforward manner that $I_{\pi_U}(x) = [l_{\pi_U}(x), u_{\pi_U}(x)]$, with

$l_{\pi_U}(x) = -u_{\pi_U}(-x)$, *and*

$$u_{\pi_U}(x) = \min\left(m, x + \max\{(1-\alpha)\,\Phi(m-x) - \alpha\Phi(-m-x)\,,\, \frac{1}{2} + \frac{1-\alpha}{2}\,(\Phi(m-x) - \Phi(-m-x))\}\right).$$

The research question that led to Theorem 1 came about following partial analysis and numerical evaluations of this coverage probability (e.g., Lmoudden, 2008). Realizing that coverage could be quite poor, attempts to find a Bayesian credible set with high infimum coverage probability were directed to either changing the prior, or departing from the HPD criteria and focussing on a different selection procedure (still for π_U) (such as in Marchand and Strawderman, 2013). But Theorem 1, which applies to a large class of choices π and of the Bayes confidence interval $I_\pi(X)$ tells us indeed that such a search is illusory unless m and α are large enough. For moderate or large m though, the minimal frequentist coverage appears to be less unsatisfactory as illustrated by Figure 1 for $m = 1.5$.

Example 1. *Figure 1 is illustrative of the points made above, namely in Remarks 1 and 3, and presents the frequentist coverage probability $C(\theta)$ of the 90% and 95% HPD credible sets associated with the uniform prior on $[-m, m]$, and for $m = 0.5, 1.0, 1.5$. For instance, looking at the case of credibility 95% and $m = 1$, we see that the coverage probability it at least 0.95 for a large part of the parameter space (approx. for $|\theta| \leq 0.69$), but drops down sharply when θ approaches the boundary to a minimum value of about 0.816 in comparison to Theorem 1's lower bound of $\Phi(1.0) \approx 0.84$. The other cases for $m = 0.5, 1.0$ are similar but the discrepancy is less pronounced for credibility 0.90, and more pronounced for $m = 0.5$. For larger m such as $m = 1.50$, Theorem 1 still applies (i.e., minimal coverage bounded above by $\Phi(1.50) \approx 0.933$ but does not imply a significant discrepancy for the chosen credibilities of 0.90 and 0.95. The graphs suggest here that the coverage fluctuates to a much lesser degree around the credibility.*

2.2. On interval estimators with exact frequentist coverage

As a corollary to the above Theorem 1, confidence intervals $I(X)$ with exact frequentist coverage $1-\alpha$ which satisfy the symmetry assumptions of Theorem 1 cannot yield exact credibility $1-\alpha$ whenever the latter is larger than $\Phi(m)$. In fact, one can say more and we prove in Theorem

1 that, whenever $m < \Phi^{-1}(1-\alpha)$, such confidence intervals must be such that $I(x) = [-m, m]$ for a positive Lebesgue measure set of x values. Such a possibility is illustrated with the simple choice $X \pm \Phi^{-1}(1-\alpha/2)$ truncated to $[-m, m]$, which has exact frequentist coverage $1-\alpha$ for all $\theta \in [-m, m]$, and which is equal to $[-m, m]$ whenever $|x| \leq \Phi^{-1}(1-\frac{\alpha}{2})-m$. The following Theorem establishes that such a phenomenon is inevitable for a large class of interval estimators.

Theorem 2. *Let $X|\theta \sim N(\theta, 1)$ with $|\theta| \leq m$. Let $I(X) = [l(X), u(X)]$ be a confidence interval with frequentist coverage $\geq 1-\alpha$ for all $\theta \in [-m, m]$, such that l and u are non-decreasing, and such that $I(X)$ satisfies the invariance property $l(x) = -u(-x)$ for all $x \in \mathbb{R}$. Then, whenever $m < \Phi^{-1}(1-\alpha)$, it must be the case that $P_\theta(I(X) = [-m, m]) > 0$ for all $\theta \in [-m, m]$.*

Proof. We show that the contrary condition that $I(x)$ be a strict subset (a.e. x) of $[-m, m]$ leads to a contradiction, namely that the frequentist coverage cannot be $1-\alpha$. In such a case, we must have $l(x_0) > -m$ for any $x_0 > 0$ since, otherwise $l(x)$ would be equal to $-m$ for all $x < 0$ in view of the non-decreasing property, $u(-x)$ would then be equal to m for the same $x < 0$ which is not permitted since $I(x)$ is a strict subset of $[-m, m]$ by assumption. But if $l(x_0) > 0$ for all $x_0 > 0$, we would have for the coverage probability at $\theta = -m$:

$$C(-m) = P_{-m}(I(X) \ni -m) \leq P_{-m}(X \leq x_0) = \Phi(x_0 + m)$$

for all $x_0 > 0$. Taking $x_0 \to 0$, we would infer that $C(-m) \leq \Phi(m) < 1-\alpha$ which is indeed a contradiction. □

Example 2. *(inversion of likelihood ratio test)*
A standard method to derive a confidence interval with exact frequentist coverage $1-\alpha$; also referred to as the "unified method" (e.g. Feldman and Cousins, 1998); is to first consider the acceptance regions $A_{\theta_0} \subset \mathbb{R}, \theta_0 \in [-m, m]$, associated with the likelihood ratio test of significance level α for $H_0 : \theta = \theta_0$ versus $H_a : \theta \neq \theta_0$, $\theta \in [-m, m]$, and then invert the test to obtain

$$I_{LRT1}(x) = \{\theta_0 \in [-m, m] : x \in A_{\theta_0}\}.$$

By construction, we obtain indeed coverage probability $P_{\theta_0}(I(X) \ni \theta_0) = P_{\theta_0}(X \in A_{\theta_0}) = 1-\alpha$ for all $\theta_0 \in [-m, m]$. Now, in our case with $X|\theta \sim N(\theta, 1)$, we obtain $A_m = [m - \Phi^{-1}(1-\alpha), \infty)$, $A_{-m} = (-\infty, \Phi^{-1}(1-\alpha) - m]$. For $\theta_0 \in (-m, m)$, the acceptance regions A_{θ_0} are of the form $[\,c_1(\theta_0, c_2(\theta_0)\,]$

with $c_1(\theta_0) < m - \Phi^{-1}(1-\alpha)$ *and* $c_2(\theta_0) > \Phi^{-1}(1-\alpha) - m$.[3]

Now, Theorem 2 applies here (one can also show that the endpoints are increasing and that the equivariance property is satisfied) so that $I_{LRT}(x)$ *must be equal to* $[-m, m]$ *for some values of* x *whenever* $m < \Phi^{-1}(1-\alpha)$. *Indeed, with the above representations of* A_{θ_0}*, we observe that* $I_{LRT}(x) = [-m, m] \Longleftrightarrow x \in A_{\theta_0}$ *for all* $\theta_0 \in [-m, m] \Longleftrightarrow x \in A_m \cap A_{-m} \Longleftrightarrow |x| \leq \Phi^{-1}(1-\alpha) - m$. *Graphs of Bayesian credibility with respect to a uniform prior on* $[-m, m]$ *(or equivalently the posterior probability of coverage where the posterior is a truncated to* $[-m, m]$ $N(x, 1)$ *distribution) are presented in Figure 2 for* $m = 1.0$ *and confidence levels* 0.90 *and* 0.95*. The credibility equals* 1 *for* $|x| \leq \Phi^{-1}(1-\alpha) - 1.0$*, as shown above, and decreases as a function of* $|x|$ *to levels well below the confidence level.*

Example 3. *The Pratt interval (Pratt, 1961), given* $I_P(x) = [\min(0, X - c), \max(0, X + c)]$ *where* $c = \Phi^{-1}(1-\alpha)$*, has frequentist probability coverage equal to* 1 *for* $\theta = 0$*, and* $1 - \alpha$ *for* $\theta \neq 0$*. Its expected length is less than the expected length of the usual choice* $X \pm \Phi^{-1}(1 - \alpha/2)$ *when* $|\theta|$ *is close to* 0*. For the truncated case* $\theta \in [-m, m]$*, Evans, Hansen and Stark (2005) establish an optimality property of its truncated version* $I_p(X) \cap [-m, m]$ *for* $m \leq 2\Phi^{-1}(1-\alpha)$*, namely minimaxity in terms of minimizing maximum expected length among all confidence intervals with minimal frequentist coverage equal to* $1-\alpha$ *for all* $\theta \in [-m, m]$*. As inferred above, with the endpoints non-decreasing and the interval equivariant, it must be the case that the interval* $I_p(x) \cap [-m, m]$ *equals* $[-m, m]$ *for relatively small* m*. Indeed it is easy to verify in this case that* $I_p(x) \cap [-m, m]$ *equals* $[-m, m]$ *whenever* $\Phi(m) < 1 - \alpha$ *and* $|x| \leq \Phi^{-1}(1-\alpha) - m$*.*

Concluding Remarks

We have established, discussed, and illustrated a discrepancy between Bayesian credibility and frequentist probability of coverage $C(\theta)$ that arises for interval estimators of the mean θ of a $N(\theta, 1)$ distribution under the constraint $|\theta| \leq m$. On one hand, the minimal value of $C(\theta)$ cannot exceed $\Phi(m)$ for a vast class of Bayesian estimators. On the other hand, interval estimators $I(x)$ with exact

[3] It is easy to verify also that $A_{\theta_0} = \theta_0 \pm \Phi^{-1}(1 - \frac{\alpha}{2})$ whenever $\Phi^{-1}(1 - \frac{\alpha}{2}) \leq m - |\theta_0|$. Otherwise, the bounds $c_1(\theta_0)$ and $c_2(\theta_0)$ require a numerical evaluation.

frequentist coverage $1-\alpha$ must be equal to $[-m, m]$ with positive probability whenever $m < \Phi^{-1}(m)$. As suggested by the proofs, these features appear to intimately related to the "smallness" of the parameter space and either are attenuated or vanish for larger parameter spaces. For this reason, it seems quite plausible that such phenomena recur for other models and further investigation would be useful. Although we have focussed on what can be judged as negative traits, our findings may lead to useful prescriptions such as adjusting the credibility of Bayes estimators or adjusting the confidence level of frequentist procedures to the size of the parameter space.

Acknowledgements

Eric Marchand's research is supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada. We are thankful to Bill Strawderman and Aziz Lmoudden for helpful discussions.

References

Evans, S.N., Hansen, B.B. & Stark, P.B. (2005). Minimax expected measure confidence sets for restricted location parameters. *Bernoulli*, **11**, 571-590.

Feldman, G.J. and Cousins, R. (1998). Unified approach to the classical statistical analysis of small signals. *Physical Review D*, **57**, 3873-3889.

Fraser, D.A.S. (2011). Is Bayes posterior just quick and dirty confidence? *Statistical Science*, **26**, 299-316.

Lmoudden, A. (2008). *Sur les intervalles de confiance bayésiens pour des espaces de paramètres contraints.* M.Sc. thesis, Département de mathématiques, Université de Sherbrooke (http://savoirs.usherbrooke.ca/handle/11143/4790)

Mandelkern, M. (2002). Setting Confidence Intervals for Bounded Parameters with discussion. *Statistical Science*, **17**, 149-172.

Marchand, É. and Strawderman, W. E. (2013). On Bayesian credible sets, restricted parameter spaces and frequentist coverage, *Electronic Journal of Statistics*, **7**, 1419-1431.

Marchand, É., Strawderman, W. E., Bosa, K., and Lmoudden, A. (2008). On the frequentist coverage of Bayesian credible intervals for lower bounded means. *Electronic Journal of Statistics*, **2**, 1028-1042.

Marchand, É. and Strawderman, W. E. (2006). On the behaviour of Bayesian credible intervals for some restricted parameter space problems. *Recent Developments in Nonparametric Inference and Probability : A Festschrift for Michael Woodroofe,*, IMS Lecture Notes-Monograph Series, **50**, pp. 112-126.

Robert, C.P. (2011). Discussion of "Is Bayes posterior just quick and dirty confidence?" by D.A.S. Fraser. *Statistical Science*, **26**, 317-318.

Roe, B. and Woodroofe, M. (2000). Setting confidence belts. *Physical Review D*, **63**, 013009/01-09.

Wasserman, L. (2011). Frasian inference. *Statistical Science*, **26**, 322-325.

Zhang, T. and Woodroofe, M. (2003). Credible and confidence sets for restricted parameter spaces. *Journal of Statistical Planning and Inference*, **115**, 479-490.

Zhang, T. and Woodroofe, M. (2002). Credible and confidence sets for the ratio of variance components in the balanced one-way model. *Sankhyā: Special issue in memory of D. Basu*, **64**, 545-560.

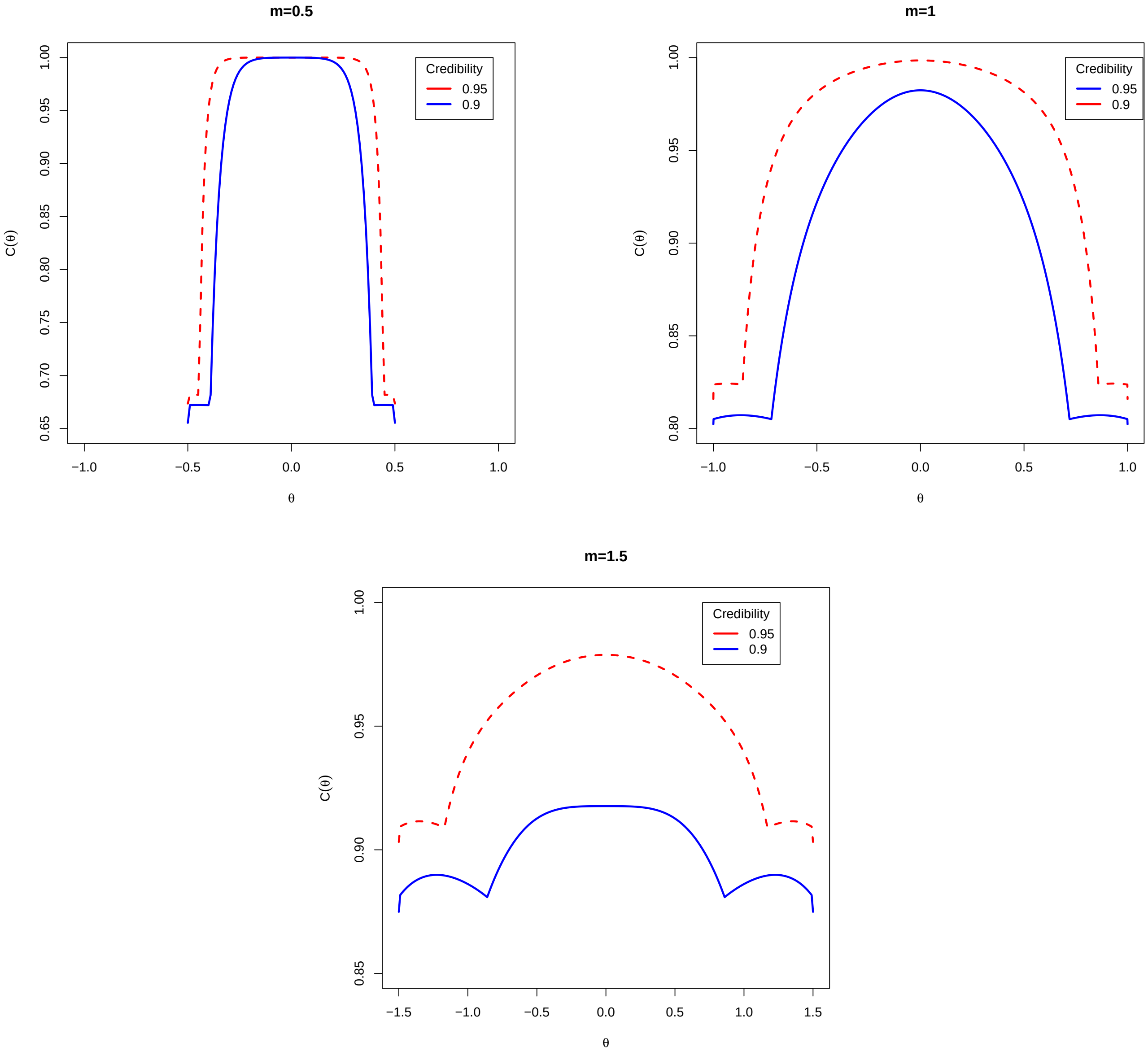

Figure 1: Coverage probability $C(\theta)$ of the HPD credible set as a function of θ for varying m and credibility.

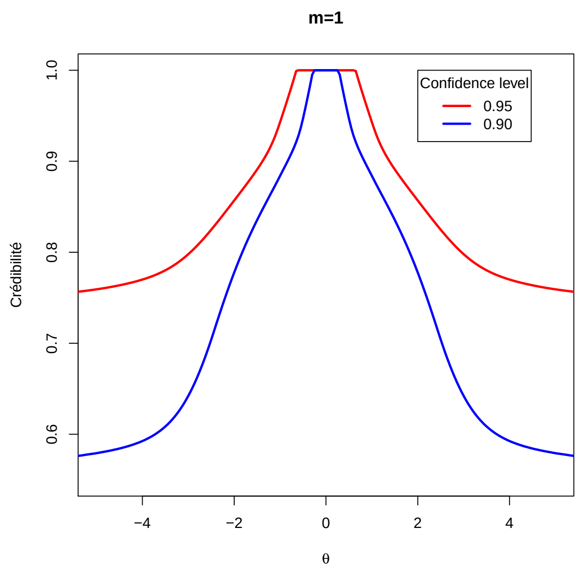

Figure 2: Bayesian credibility of the 90% and 95% confidence interval $I_{LRT}(X)$ with respect to the uniform prior for $m = 1.0$.