Using prior knowledge in frequentist tests

Use of Bayesian posteriors with informative priors as optimal frequentist tests

Christian Bartels
Email: h.christian.bartels@gmail.com
Allschwil, 15.3.2019

Reference as:
Bartels, C., 2019, Using prior knowledge in frequentist tests. Figshare.
https://dx.doi.org/10.6084/m9.figshare.4819597

Versions:
13.4.2017 - corrected error in references
23.7.2017 - corrected typo in definition of total risk
15.3.2019 - clarified chosen relation between decision space and parameter space

Introduction

Making decisions under uncertainty based on limited data is important and challenging. Decision theory provides a framework to reduce risks of decisions under uncertainty with typical frequentist test statistics being examples for controlling errors, e.g., Dudley (2003) or Rüschendorf (2014). This strong theoretical framework is mainly applicable to comparatively simple problems. For more complex problems and/or if there is only limited data, it is often not clear how to apply the strong framework to the actual problem at hand (e.g., Altonji, 1996). In practice, careful iterative model building and checking seems to be the best what can be done - be it using Bayesian methods or applying frequentist approaches that were established for simpler problems or for the limit of large samples.

This manuscript aims at expanding the armory for decision making under uncertainty with complex models, focusing on trying to expand the reach of decision theoretic, frequentist methods. In prior work (Bartels, 2015), an efficient integration method was re-evaluated for repeated calculation of statistical integrals for a set of of hypotheses (e.g., p-values, confidence intervals). Key to the method was the use of importance sampling. Subsequently, pointwise mutual information was proposed as an efficient test statistics and shown to be optimal under certain conditions. Here, proposals are made for optimal frequentist test statistics that can take into account prior knowledge.

Proposal

Essentially, it is re-proposed to use the Bayesian posterior distributions or alternatively and with the same result the pointwise mutual information as optimal frequentist test statistics. The resulting statistics give optimal equalizer (and thus minimax) decision rules with the same risk of erroneous decisions for all values of the parameter. Prior knowledge differentiating potential losses for different decisions is taken into account. Key to the derivation of these results is a change of perspective. Usually, one would postulate some loss function, and then try to determine the prior that gives a minimax rule (e.g., Dudley, 2003 or Rüschendorf, 2014). Here, the prior is assumed to be informative, e.g., to result from an evaluation process that is part of careful iterative model building (e.g., Gelman, 2014). The informative prior is then used to derive a loss function, such that the corresponding decision rule is an optimal equalizer rule. With both perspectives external knowledge has to be used to inform the decision process. Either the external knowledge is used to select a loss function, or it is used to select a prior. Having selected either the prior or the loss function, the other is determined via an optimization of the resulting decision function. The change of perspective, pre-specifying the prior rather than the loss function, has the advantage of simplifying and rendering possible the required calculations, which otherwise would remain elusive.

The proposed approach is positioned within decision theory (e.g., Dudley, 2003 or Rüschendorf, 2014; see also appendix). In decision theory, there is a measurable space of decisions called the decision space. A decision set $\delta(x, d) \in [0, 1]$ that associates observation x with decisions d. The rule expressed by the decision set says that if x is observed, then the action d should should be considered as a possible option with probability $\delta(x, d)$; if $\delta(x, d) = 0$, the decision d should not be considered anymore; if $\delta(x, d) = 1$, the decision d should be considered as an option. In what follows, the decision space is chosen to be the set of all possible parameters $\theta = (\mu, \sigma)$, and the decision rule, $\delta(x, \mu, \sigma) \in [0, 1]$, says whether or not parameters should still be considered as possible after having observed x. Other possible choices for the decision space are not considered here.

Optimal decision sets

Decision sets are considered that maintain the risk at or below a pre-specified level and have the smallest possible size (similar to Schafer, 2009). Typical examples are 95% confidence intervals or sets that maintain the risk of not including the true value of the parameter at or below 5%.

Here, the risk for given parameters, μ and σ, is given by

$$R_{\mu,\sigma} = \int l(x, \mu, \sigma)\, \{1 - \delta(x, \mu, \sigma)\}\, f(x, \mu, \sigma)\, d\nu(x)$$

Using as reference the total risk that does not depend on the decision set

$$TR_{\mu,\sigma} = \int l(x, \mu, \sigma) f(x, \mu, \sigma)\, d\nu(x)$$

this gives

$$R_{\mu,\sigma} = TR_{\mu,\sigma} - \int l(x, \mu, \sigma)\, \delta(x, \mu, \sigma) f(x, \mu, \sigma)\, d\nu(x)\,.$$

The size of the decision set is

$$S_{\mu,\sigma} = \int \delta(x, \mu, \sigma)\, d\nu(x)$$

To reduce the size of a decision for a given risk, data, x, must preferentially be included in the decision set, $\delta(x, \mu, \sigma)$ for which $l(x, \mu, \sigma) f(x, \mu, \sigma)$ is large.

If no nuisance parameters need to be eliminated, decision sets can be defined for each value of the parameters and may be adjusted to maintain the risk at a pre-specified limit.

Such decision sets are referred to as equalizer rules since they have equal risk independent of the values of the parameters.

Correspondence between prior distribution and loss function

Two cases are considered to establish a correspondence between prior information encoded in a loss function and prior information encoded in an informative prior distribution. The first case considers an informative loss function together with a noninformative reference prior. The second case considers a non-informative 0-1 loss function together with an informative prior. Both cases consider the same conditional likelihood of the data given the parameters. For case 1, let $l_1(x, \mu, \sigma)$ denote the informative loss function, and $d\pi_1(\mu, \sigma)$ the noninformative reference prior with $d\nu_1(x)$ and $f_1(x, \mu, \sigma)$ being the corresponding marginal measures of the data and the pointwise mutual information, respectively. For case 2, the loss function is noninformative and equal to 1 for all combinations of parameters and data, and $d\pi_2(\mu, \sigma)$ is the informative prior with corresponding $d\nu_2(x)$ and $f_2(x, \mu, \sigma)$. This gives for the two cases the risks

$$R_{1,\mu,\sigma} = \int l_1(x, \mu, \sigma)\,\{1 - \delta(x, \mu, \sigma)\}\, f_1(x, \mu, \sigma)\, d\nu_1(x) \text{ and}$$

$$R_{2,\mu,\sigma} = \int \{1 - \delta(x, \mu, \sigma)\}\, f_2(x, \mu, \sigma)\, d\nu_2(x).$$

The same likelihood is used for both cases, implying

$$f_1(x, \mu, \sigma)\, d\nu_1(x) = f_2(x, \mu, \sigma)\, d\nu_2(x) = dP_{\mu,\sigma}(x),$$

and both expressions for the risks use the same probability measure for the integration. From optimality considerations above, optimal decision functions are constructed for any value of the parameters, μ, σ by preferentially including data, x, in the decision set, $\delta(x, \mu, \sigma)$ for which $l(x, \mu, \sigma) f(x, \mu, \sigma)$ is large. For case 1, data is included for which $l_1(x, \mu, \sigma) f_1(x, \mu, \sigma)$ is large. For case 2, data is included for which $f_2(x, \mu, \sigma)$ is large. The two cases include the same data preferentially in the decision set, if

$$l_1(x, \mu, \sigma) f_1(x, \mu, \sigma) = f_2(x, \mu, \sigma).$$

Substituting $f_1(x, \mu, \sigma)$ by $f_2(x, \mu, \sigma)$ using the equivalence of the likelihood, it follows that

$$l_1(x, \mu, \sigma) = \frac{d\nu_1(x)}{d\nu_2(x)}$$

Thus the informative prior defines a non-uniform loss function that depends on the data, x.

In general, the risks $R_{1,\mu,\sigma}$ and $R_{2,\mu,\sigma}$ will be different for any given decision set $\delta(x, \mu, \sigma)$. However, for any one given optimal decision set, $\delta_{opt}(x, \mu, \sigma)$, the risks can be made equal by scaling the loss function with an appropriate constant $l_1(\mu, \sigma)$. The resulting expression of the loss function is $l_1(x, \mu, \sigma) = \frac{d\nu_1(x)}{d\nu_2(x)} l_1(\mu, \sigma)$.

Illustration

The following examples illustrate the effect of using informative priors on the resulting decision rules. To facilitate illustration, a one-dimensional problem with known closed form solutions for the likelihood and the posteriors has been chosen: the data is assumed to be generated from a normal distribution with unknown mean, μ, and known standard deviation, σ_{true}. I.e., $n = 10$ numbers, x_i, are drawn from a normal distribution $N(\mu, \sigma_{true})$. First confidence intervals (CI) are derived using a non-informative prior resulting in CI as they are usually defined[1].

The usual definition for this example of confidence intervals with a coverage of C (e.g. C=95%) is the central part $\bar{x} \pm z^* \frac{\sigma_{true}}{\sqrt{n}}$ of the normal distribution with mean, $\bar{x}$, standard deviation, $\frac{\sigma_{true}}{\sqrt{n}}$, and the coverage C defining the critical value z^* via $\int_{y=-z*}^{+z*} N(y,0)dy = C$. Or, written as a set

$$\forall \bar{x} \;\; CI_{\bar{x}} = \left\{ \mu \mid |\bar{x} - \mu| < z^* \tfrac{\sigma_{true}}{\sqrt{n}} \right\}$$

Non-informative loss function

The same confidence interval is obtained with the proposed optimal decision rule using a non-informative, improper prior $d\pi(\mu) = d\mu$ ($d\mu$ being the Lebesgue measure of the mean parameter). With this prior, the marginal distribution of the observations $d\nu(x)$ is equal to the Lebesgue measure $d\bar{x}$ of the sufficiency statistics $\bar{x}$, and the conditional posterior probability of the parameters given some data, $f(x, \mu, \sigma)\, d\pi(\mu, \sigma)$, is equal to $N\left(\mu, \bar{x}, \frac{\sigma_{true}}{\sqrt{n}}\right) d\mu$. Independent of the prior, the conditional likelihood $f(x, \mu, \sigma)\, d\nu(x)$ is equal to the normal distribution $N\left(\bar{x}, \mu, \frac{\sigma_{true}}{\sqrt{n}}\right) d\bar{x}$. The conditional posterior probability is used as test statistics to select the data to be included by preference into the decision set. Data should be included by preference for which the test statistics is large, i.e., decision rules

$$\delta_{z'}(\bar{x}, \mu) = 1 \text{, if } |\mu - \bar{x}| < z'$$

$$\delta_{z'}(\bar{x}, \mu) = 0 \text{, if } |\mu - \bar{x}| \geq z'$$

are considered with z' being the parameter that specifies the size of the decision set. The size, $z' = z^* \frac{\sigma_{true}}{\sqrt{n}}$, is fixed to give the desired coverage $\int_{\bar{x}=\mu-z'}^{\mu+z'} N(\mu - \bar{x}, 0, \frac{\sigma_{true}}{\sqrt{n}})\, d\bar{x} = C$.

Thus, the decision sets, DS_μ are

$$\forall \mu \;\; DS_\mu = \left\{ \bar{x} \mid |\bar{x} - \mu| < z^* \tfrac{\sigma_{true}}{\sqrt{n}} \right\}.$$

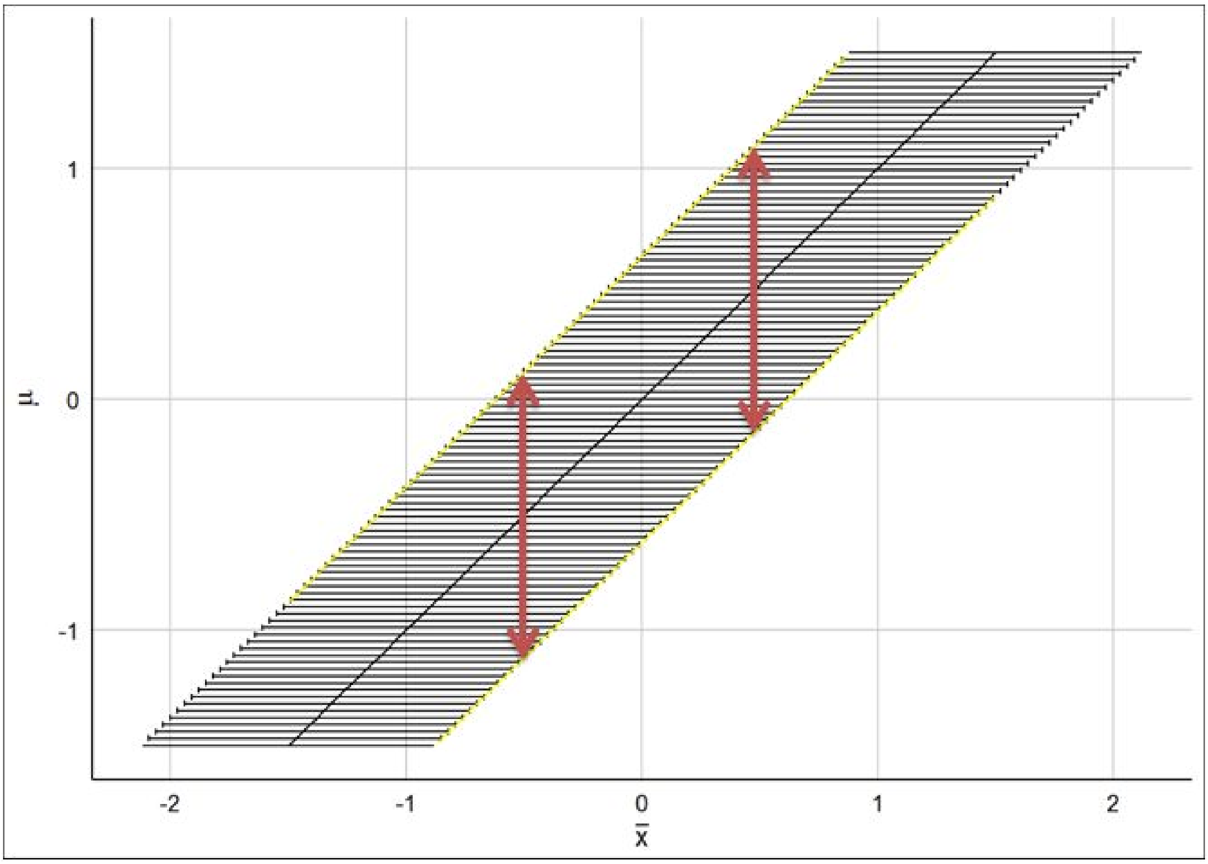
This holds for any given μ, and may be written as a set $\left\{ (\bar{x}, \mu) \mid |\bar{x} - \mu| < z^* \frac{\sigma_{true}}{\sqrt{n}} \right\}$ of parameters, μ, and data, $\bar{x}$, included by the decision rule. Changing perspective, and looking at parameters included in the set for each value of the data, $\bar{x}$, gives the confidence intervals

$$\forall \bar{x} \;\; CI_{\bar{x}} = \left\{ \mu \mid |\bar{x} - \mu| < z^* \tfrac{\sigma_{true}}{\sqrt{n}} \right\}.$$

The decision sets and the confidence intervals are illustrated in Fig. 1. The decision sets, DS_μ, defined to guarantee the desired coverage are shown as black lines. The yellow lines indicate the usual CI, $\bar{x} \pm z^* \frac{\sigma_{true}}{\sqrt{n}}$. The two red arrows illustrate two CI's defined as the parameters included in the decision sets for any given value, $\bar{x}$, of the data. The figure is entitled as being based on a non-informative loss function to emphasize that no prior information was used to differentiate different parameters.

[1] http://www.stat.yale.edu/Courses/1997-98/101/confint.htm

Figure 1. Decision sets and confidence interval based on non-informative loss function



Informative loss function

Confidence intervals with the same coverage can be derived for the same example using informative priors. As example, the prior $d\pi(\mu) = N\left(\mu, 0, s_{prior}\right) d\mu$ is used, with $s_{prior} = 0.5$. Independent of the prior, the conditional likelihood $f(x, \mu, \sigma)\, d\nu(x)$ is equal to the normal distribution $N\left(\bar{x}, \mu, \frac{\sigma_{true}}{\sqrt{n}}\right) d\bar{x}$. The conditional posterior probability of the parameters given some data, $f(x, \mu, \sigma)\, d\pi(\mu, \sigma)$, is equal to $N\left(\mu, m_{post}, s_{post}\right) d\mu$. The central value and the standard deviation of the posterior distribution are given by

$$s_{post} = \sqrt{\left(1/s_{prior}^2 + n/\sigma_{true}^2\right)^{-1}}$$

$$m_{post} = \alpha\bar{x} \text{ with } \alpha = \frac{n \times s_{post}^2}{\sigma_{true}^2}$$

The conditional posterior probability is used as test statistics to select the data to be included by preference into the decision set. Data should be included by preference for which the test statistics is large, i.e., decision rules

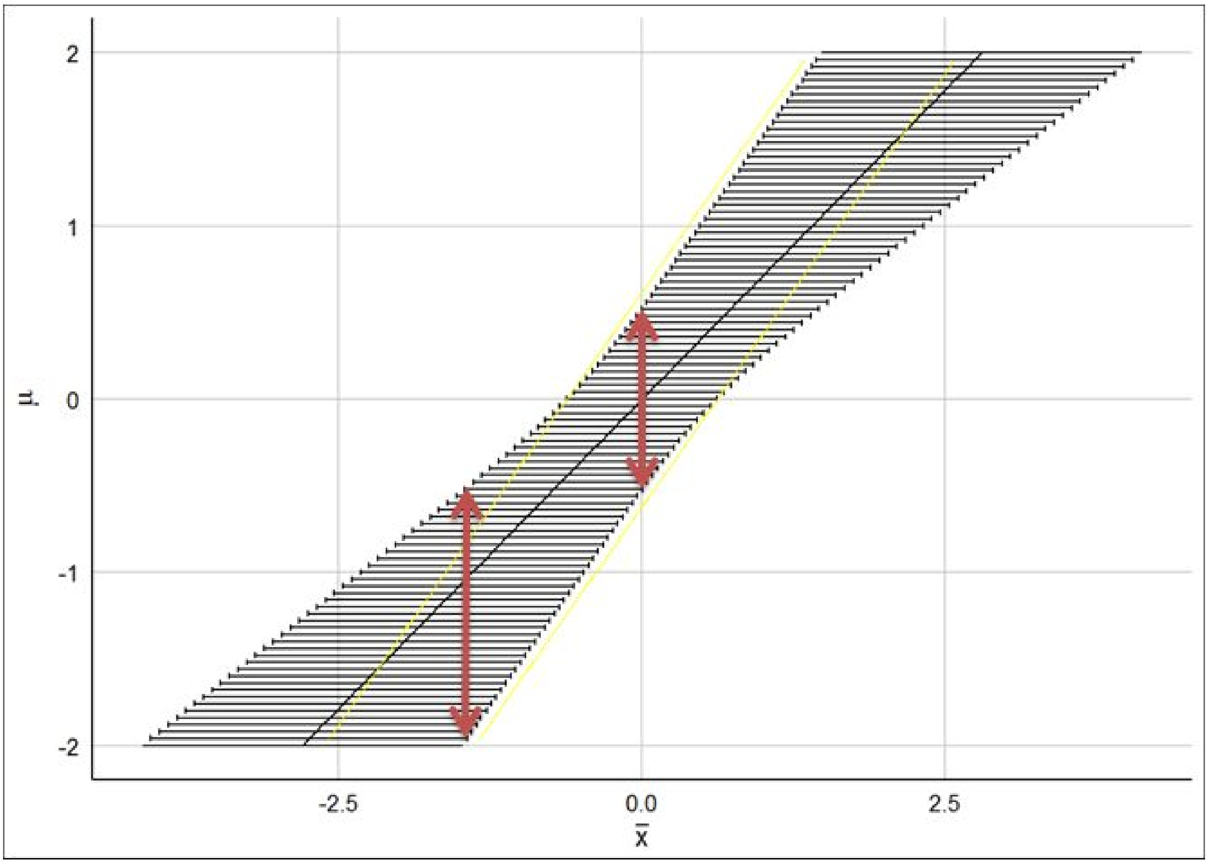
$$\delta_{z'}(\bar{x}, \mu) = 1 \text{, if } |\mu - \alpha\bar{x}| < z'$$

$$\delta_{z'}(\bar{x}, \mu) = 0 \text{, if } |\mu - \alpha\bar{x}| \geq z'$$

are considered with z' being the parameter that specifies the size of the decision set. The size, z', is fixed to give the desired coverage $\int_{\bar{x}=\mu/\alpha - z'}^{\mu/\alpha + z'} N(\mu - \bar{x}, 0, \frac{\sigma_{true}}{\sqrt{n}})\, d\bar{x} = C$.

The decision sets for fixed values of the parameter μ are illustrated in Figure 2 as black lines. For μ equal to 0, the decision set is the same as the one derived for the case with an uninformative prior that is indicated by the yellow lines. For parameters μ smaller than 0, the decision set starts and ends at smaller (more negative) values of $\bar{x}$ than for the case with an uninformative prior. For parameters μ larger than 0, the decision set starts and ends at larger values of $\bar{x}$ than for the case with an uninformative prior. As a result, the CI at $\bar{x}$ equal to 0 (red arrow) is smaller than the CI obtained with with an uninformative prior and does not extend out to the yellow lines. At $\bar{x}$ different from 0, the CI get larger and start to extend beyond one of the limits of the CI obtained with an uninformative prior.

Figure 2. Decision sets and confidence interval based on an informative loss function



To evaluate the impact of the informative prior, one can evaluate the decisions that are taken, if the CI intervals would be used to test against the null hypothesis of the mean μ being equal to 1. If observations are made that have a high probability based on the prior, i.e., close to zero, the confidence intervals based on the informative prior have a higher power to reject the null hypothesis than the CI intervals obtained with the non-informative loss function. If observations are made that have a lower probability based on the prior, e.g., $\bar{x}$ being equal to 2, then the confidence intervals based on the informative prior have a lower power to reject the null hypothesis than the CI intervals obtained with the non-informative loss function. This illustrates how prior information may be used to set up decision functions with improved power.

Bayesian excursion

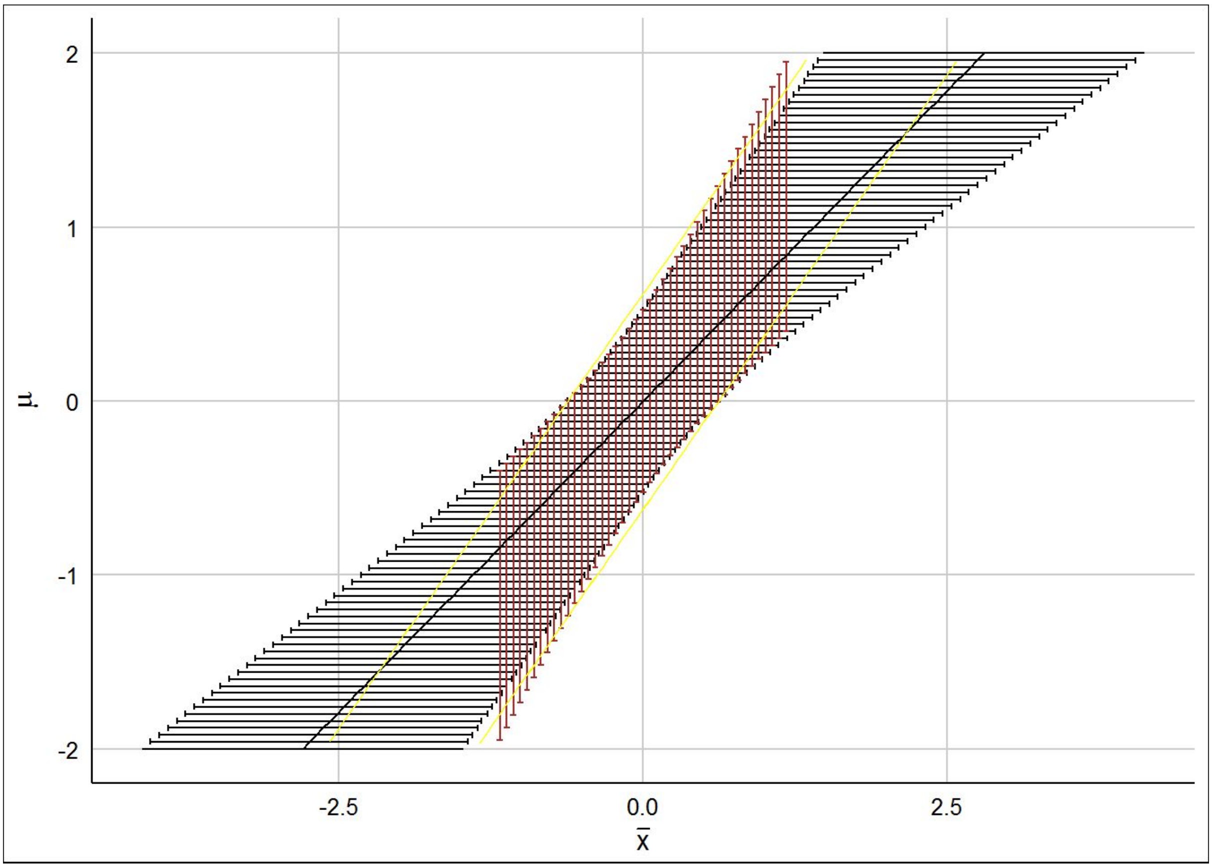
For completeness, Bayesian decision sets derived with the same optimality criteria and the same informative prior are presented in this section. The resulting credible intervals are consistent with the confidence intervals in that the same criterion is used to include decisions into the decision set. The question being asked is different, and the credible intervals are neither identical to the confidence intervals nor do they give guarantees on type I errors of decisions. The credible intervals are helpful in that they make statements about hypothetical objects that cannot be observed - the true value of the parameters. In the limit of a large number of observations, confidence and credible intervals as defined here and as evaluated in detail by Evans (2016) are identical.

Bayesian decision sets fix the coverage for integrals over parameters, $N\left(\mu, \alpha\bar{x}, s_{post}\right) d\mu$, given some observation, $\bar{x}$, and use the likelihood function, $N\left(\bar{x}, \mu, \frac{\sigma_{true}}{\sqrt{n}}\right) d\bar{x}$, as test function. The resulting Bayesian credible intervals are defined by the integrals

$$\int_{\mu=\bar{x}-z'}^{\bar{x}+z'} N(\mu - \alpha\bar{x}, 0, s_{post})\, d\mu = C \quad \text{or} \quad \int_{\mu=\bar{x}-z'}^{\bar{x}+z'} N(\mu, \alpha\bar{x}, s_{post})\, d\mu = C$$

The Bayesian decision sets are illustrated in Fig. 3 using red lines and compared to the corresponding frequentist decision sets illustrated by black lines. At $\bar{x}$ or μ equal to 0, the intervals coincide. At other values, the intervals and the regions covered by the two sets are similar but differ.

Figure 3. Comparison of frequentist and Bayesian confidence and credible intervals based on the same informative loss function



Discussion

Making decisions under uncertainty based on limited data is important and challenging. Combining the sampling scheme proposed in (Bartles, 2015) with the use of prior information as proposed here gives a method with a series of desired properties:

- Generic: The approach is generic in that confidence intervals, credible intervals or p-values can be determined for any problem at hand as long as the likelihood function is defined
- Practicable: even though the method relies on numerical evaluation of integrals via sampling, which inherently requires some computational resources, use of importance resampling lessens the computational burden sufficiently for it to be implemented and used on commonly available hardware.
- Prior knowledge: an approach is proposed to use prior information for decision making within a frequentist framework. Use of prior information encoded as informative loss functions is an inherent part of decision theory. Translating this theoretical possibility into an implementation of an optimal frequentist decision method given informative loss functions is however at least challenging and tends not be done in practice.
- Optimal: the confidence intervals and credible intervals are optimal in a sense similar to Stark and colleagues (Schafer, 2009) in that they have the smallest average size for any given coverage. Also, using the posterior distribution is equivalent to using the pointwise mutual information as test statistics. The latter can also be used as an optimal test statistics for Bayesian set selection and has been illustrated here for the simple example at hand. The pointwise mutual information is referred to as relative belief by Evans (2016) and argued to be an optimal measure of statistical evidence.
- Exact: the approach is exact and does not rely on large sample approximations. Even if there is only very limited data, intervals to support the decisions will have the correct coverage.
- Consistent: the approach gives confidence intervals and credible intervals that are consistent with each other in that the same criterion is used to prioritize inclusion of parameters or data into the decision sets. For both questions, confidence intervals and credible intervals, the pointwise mutual information is used as criterion.

An important limitation of the proposed approach is that the decision space is set equal to the space of the possible parameters. Other choices are not discussed. Also, it has not been described how to handle nuisance parameters. In principle, this can be handled via an integrated likelihood approach. Integrating the joint distribution over the nuisance parameters gives a joint distribution without nuisance parameters from which the corresponding conditional or marginal distributions may be derived.

Conclusions

An approach for frequentist decisions making has been presented that takes into account prior knowledge on the decision to be taken. The approach has been positioned within decision theory (e.g., Dudley, 2003; Rüschendorf, 2014). The approach limits type I errors and confidence intervals have the correct coverage. The approach makes optimal decisions given an informative loss function with the loss function being encoded via the specification

of a prior distribution on the parameters. Taking into account appropriate prior information may increase the power of frequentist tests.
Given the prior, the posterior distribution is used as test statistics to define observations to be included by preference into the decision sets, and the size of the decision sets is adjusted to give the desired frequentist coverage. Nuisance parameters could be handled by integrating them out.
The approach has been illustrated with a simple example taking advantage of simple closed form solutions that exist for this example. The illustration focused on exemplifying the effect of using prior information - confidence intervals got smaller for observations that were consistent with the prior information. The proposed approach can also be implemented using numerical integration methods (Bartels 2015). As such the approach is generic and can be applied whenever a likelihood function is defined.

References

Altonji, J.G. and Segal, L.M., 1996. Small-sample bias in GMM estimation of covariance structures. *Journal of Business & Economic Statistics*, *14*(3), pp.353-366.

Bartels, C., 2015, Generic and consistent confidence and credible regions. Figshare. https://dx.doi.org/10.6084/m9.figshare.1528163.v5

Dudley, R.M., 2003. Mathematical Statistics, 18.466 lecture notes, Spring 2003. On MIT OCW (OpenCourseWare) website. Archived at http://hdl.handle.net/1721.1/103814

Evans, M., 2016. Measuring statistical evidence using relative belief. *Computational and structural biotechnology journal*, *14*, pp.91-96.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B., 2014. *Bayesian data analysis* (Vol. 2). Boca Raton, FL, USA: Chapman & Hall/CRC.

Rüschendorf, L., 2014. *Mathematische statistik*. Springer Spektrum.

Schafer, C.M. and Stark, P.B., 2009. Constructing confidence regions of optimal expected size. *Journal of the American Statistical Association*, *104*(487), pp.1080-1089.

Appendix

Nomenclature

Parameters and observations

Set selection and interval selection are considered. The parameter of interest is μ, e.g., the mean of a normal distribution or a discrete variable differentiating between two or more alternative hypotheses. Nuisance parameters are denoted by σ. Observations are denoted by x, and the actually observed data by x_0.

Probability distributions

The joint probability of the parameters and the observations, $dP(x, \mu, \sigma)$, is assumed to exist, and to be factorizable into the product of the marginal distribution of the observations, $d\nu(x)$, the marginal distribution of the parameters, $d\pi(\mu,\sigma)$, and the pointwise mutual information $f(x,\mu,\sigma)$. This notation is special in that the marginal distribution of the observations is used as reference measure. This results in simple, symmetric expressions for the different distributions:

- $f(x,\mu,\sigma)\,dv(x)\,d\pi(\mu,\sigma)$ - joint distribution $dP(x,\mu,\sigma)$
- $f(x,\mu,\sigma)\,dv(x)$ - conditional likelihood of the observations given the parameters
- $f(x,\mu,\sigma)\,d\pi(\mu,\sigma)$ - conditional posterior probability of the parameters given some data

Decision Theory

The proposed approach is positioned within decision theory (e.g., Dudley, 2003 or Rüschendorf, 2014). In decision theory, there is a measurable space of decisions called the decision space. A decision set $\delta(x,d) \in [0,1]$ that associates observation x with decisions d. The rule expressed by the decision set says that if x is observed, then the action d should should be considered as a possible option with probability $\delta(x,d)$; if $\delta(x,d)=0$, the decision d should not be considered anymore; if $\delta(x,d)=1$, the decision d should be considered as an option. In what follows, the decision space is chosen to be the set of all possible parameters $\theta=(\mu,\sigma)$, and the decision rule, $\delta(x,\mu,\sigma) \in [0,1]$, says whether or not parameters should still be considered as possible after having observed x.

Decision functions into the space of possible parameters

Decisions are made on the associations of outcomes with model parameters: Should a parameter or hypothesis be considered still as possible after having observed the data x_0? Nuisance parameters are considered as parameters that are required to fully specify the model but that are not of interest for the decision. A special case of this setup are hypothesis tests where the parameter of interest is a binary indicator that distinguishes between two disjunct sets of the values of nuisance parameters with the disjunct sets referred to as hypotheses. Decision functions, $\delta(x,\mu,\sigma) \in [0,1]$ are use to denote the association with 0 indicating that the particular association should not anymore be considered, 1 indicating that the association should still be considered. Values between 0 and 1 may be used and are often more of theoretical and the practical relevance; they indicate that a random experiment should be carried out that accepts the association with a probability equal to the value $\delta(x,\mu,\sigma)$.

If σ is handled as nuisance parameter, the decision function may not depend on it, $\delta(x,\mu,\sigma) \equiv \delta(x,\mu)$. Associations are made between the parameter of interest, μ, and the observations irrespective of the value of the nuisance parameter which is neither known nor of interest.

Losses, risks and sizes

Identification of optimal decision functions requires, as input, information on the value of different correct or wrong decisions. Commonly, losses of wrong decisions, $l(x,\mu,\sigma)$, are considered.

With these definitions risk can be defined as integrals of the probability of making certain decisions and their associated loss. E.g., assuming that the parameter values μ and σ correspond to the truth, the risk of making a wrong decision is

$$R_{\mu,\sigma} = \int l\,(x, \mu, \sigma)\ \{1 - \delta\,(x, \mu, \sigma)\}\, f\,(x, \mu, \sigma)\, d\nu\,(x)\,.$$

With nuisance parameters, one usually considers the maximal risk over all possible values of the nuisance parameter

$$R_{\mu} = sup_{\sigma}\ R_{\mu,\sigma}$$

Similarly, if evaluating decision rules that are to be applied for different possible values of the parameter of interest (rather than evaluating separate decision functions for each of the values of the parameter of interest), one may use the maximal risk over these values to assess the risk of the rule.

Risks of decision functions may not be sufficient to identify optimal decision functions and sizes may have to be used as additional optimality criteria or as constraints. This is the case, in particular, also for set and interval estimation, as considered here, and for confidence and credible intervals, respectively.