A default prior for regression coefficients

Erik van Zwet ®

Abstract

When the sample size is not too small, M-estimators of regression coefficients are approximately normal and unbiased. This leads to the familiar frequentist inference in terms of normality-based confidence intervals and p-values. From a Bayesian perspective, use of the (improper) uniform prior yields matching results in the sense that posterior quantiles agree with one-sided confidence bounds. For this, and various other reasons, the uniform prior is often considered objective or non-informative. In spite of this, we argue that the uniform prior is not suitable as a default prior for inference about a regression coefficient in the context of the bio-medical and social sciences. We propose that a more suitable default choice is the normal distribution with mean zero and standard deviation equal to the standard error of the M-estimator. We base this recommendation on two arguments. First, we show that this prior is non-informative for inference about the sign of the regression coefficient. Second, we show that this prior agrees well with a meta-analysis of 50 articles from the MEDLINE database.

1 Introduction

Regression modeling plays a central role in the bio-medical and social sciences. Linear and generalized linear models (with and without random effects), generalized estimating equations (GEE) and quantile regression offer great flexibility and are easy to use. Also, simple two group comparisons can usually be viewed in terms of a regression model. When the sample size is not too small, the statistical analysis can be based on the fact that M-estimators of regression coefficients are approximately normal and unbiased.[1] This leads to the familiar (frequentist) inference in terms of normality-based confidence intervals and p-values.

We will avoid small sample issues by assuming that we have a normally distributed, unbiased estimator B of a regression coefficient β with known standard error se. Then we have for any fixed β

$$P_\beta(B - 1.96\,\text{se} < \beta < B + 1.96\,\text{se}) = 0.95. \tag{1}$$

While our set-up may appear to be overly simplistic, we emphasize that inference about regression parameters based on Wald type confidence intervals (and associated p-values) is very common throughout the life sciences. Exact intervals based on the t-distribution are available in linear models, but the difference is already very small when the degrees of freedom exceed, say, 40.

Statement (1) describes the long-run coverage performance of the random interval $[B - 1.96\,\text{se}, B + 1.96\,\text{se}]$. However, this confidence statement is often mistakenly interpreted as the conditional probability statement

$$P(B - 1.96\,\text{se} < \beta < B + 1.96\,\text{se}|B = b) = 0.95 \tag{2}$$

Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

Corresponding author:
Erik van Zwet, Department of Biomedical Data Sciences, Leiden University Medical Center, The Netherlands.
Email: E.W.van_Zwet@lumc.nl

where b is the observed value of the estimator and β is viewed as a random variable. We refer to Greenland et al.[2] for a discussion of this misinterpretation. Statement (2) is arguably more relevant than (1) because it refers to the data at hand, rather than the procedure being used. This may explain, at least in part, the pervasiveness of the misinterpretation; it is what researchers want to know.

Statement (2) is actually only valid if we assume that β has the (improper) uniform or "flat" prior distribution. This has led some authors to consider the uniform prior to be an objective or non-informative prior.[3] Many other criteria have been proposed by which a prior may be considered to be objective,[4] but in the normal location model with known standard deviation they all yield the (improper) uniform distribution as the unique objective prior.

We find that the flat prior is used for Bayesian inference about regression coefficients in two distinct situations. It is used explicitly with the goal of objective Bayesian inference[5] and implicitly whenever the confidence interval for a regression coefficient is interpreted as a credibility interval.

This paper consists of two parts. In the first part (section 2), we discuss the objective Bayesian approach. Loosely speaking, the goal of this approach is to minimize the influence of the prior on the posterior. However, this depends on which aspect of the posterior we are considering. While the flat prior may well be considered non-informative for β, it is very informative *both* for the magnitude and the sign of β. This is just a consequence of the fact that a very diffuse prior favors large values of $|\beta|$. Hence, use of the flat prior implies that the magnitude of β will be inflated and the evidence about its sign will be exaggerated.

This is problematic because for regression coefficients (other than the intercept) we are typically most interested in the sign and the magnitude. In section 2.3, we go one step further and argue that the sign is often of primary interest. We start from the premise that associations studied in the life sciences are almost never exactly zero and that it is therefore of special importance to quantify the evidence for the *direction* of a certain association. From a Bayesian perspective, this means that we are primarily interested in $P(\beta > 0|B)$.

Now, if we want to avoid undue influence of the prior on $P(\beta > 0|B)$, it is natural to use a prior for β such that $P(\beta > 0|B)$ has the standard uniform distribution. Theorem 1 asserts that the normal distribution with mean zero and standard deviation equal to the standard error of the unbiased estimator is such a prior. In other words, this prior may be considered to be non-informative for inference about the sign of β.

In the second part of the paper (section 3), we turn to the Empirical Bayesian approach.[6,7] MEDLINE is an extensive bibliographic database of life sciences and biomedical information. Compiled by the United States National Library of Medicine, it is freely available on the Internet and searchable via PubMed. In the absence of additional prior information, we may consider papers from MEDLINE to be *exchangeable*. This implies that we can use a sample of MEDLINE papers to estimate a suitable prior distribution for regression coefficients. Such an estimated prior is objective in the sense that it is essentially free of any personal opinions or biases. Based on a sample of 50 MEDLINE articles, we estimate the distribution of regression coefficients (other than the intercept) to be normal with mean zero and standard deviation 1.28 se.

Objective Bayes and Empirical Bayes both aim for objectivity, but from very different points of view. It is therefore quite remarkable that both approaches lead to such similar priors. This supports the main conclusion of this paper that the flat prior is not suitable for objective inference about regression coefficients.

In section 3.2, we discuss several reasons why the factor of 1.28 is likely to be an overestimate. Therefore, we recommend that the normal distribution with mean zero and standard deviation se is a suitably conservative default prior. This prior combines the theoretical support of section 2.3 with the empirical evidence from section 3.1. We do stress that this default prior is not always appropriate. In section 4, we discuss several situations in which it is *not* to be used.

Upon observing $B = b$, the resulting posterior is the normal distribution with mean $b/2$ and standard deviation $se/\sqrt{2}$. Since statement (1) holds for all β, it remains valid under the proposed prior. However, the conditional coverage of the usual confidence interval becomes

$$P(B - 1.96\,se < \beta < B + 1.96\,se|B = b) = \Phi\left(\frac{b}{\sqrt{2}\,se} + 1.96\sqrt{2}\right) - \Phi\left(\frac{b}{\sqrt{2}\,se} - 1.96\sqrt{2}\right) \tag{3}$$

where Φ is the standard normal cumulative distribution function. We show the conditional coverage in Figure 1. Instead of b, we put the two-sided p-value $2\Phi(-|b|/\,se)$ on the x-axis to emphasize that with increasing significance, the coverage probability decreases. Also, we have

$$P\left(\frac{B}{2} - 1.96\frac{se}{\sqrt{2}} < \beta < \frac{B}{2} + 1.96\frac{se}{\sqrt{2}}\Big|B = b\right) = 0.95 \tag{4}$$
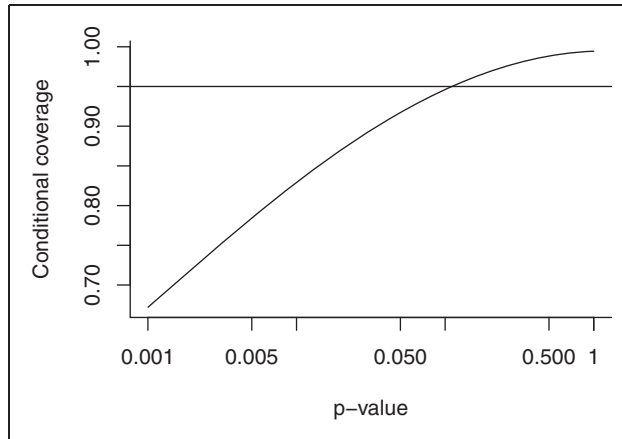
Figure 1. Conditional coverage when β has the normal prior with mean zero and standard deviation se, as a function of the two-sided p-value.

and

$$P(\beta > 0 | B = b) = \Phi(b/\sqrt{2}\, \text{se}) \tag{5}$$

Finally, $P(\beta < 0 \mid B = b) = 1 - P(\beta > 0 | B = b)$

We note that several other authors[8–12] have warned against using the flat prior as a default choice and some of those authors have proposed priors that are qualitatively similar to ours, in the sense of being symmetric and concentrated around zero.

2 Objective Bayes

The Bayesian method offers a principled and coherent approach to statistical inference, but it does require the specification of a prior distribution. If one is unable or unwilling to formalize external information into a prior, it is tempting to use a so-called objective (or non-informative) prior to try to avoid undue influence on the posterior.[5]

Many criteria have been proposed by which a prior may be considered to be objective.[4] In the normal location model with known standard deviation, they all point to the flat prior. Here we review two important criteria; one based on invariance of the statistical decision problem and one based on the principle of parameterization invariance. We will argue that neither criterium is as compelling as it may seem. An information theoretic argument in section 2.3 leads us to propose the normal distribution with mean zero and standard deviation equal to the standard error of the unbiased estimator as a suitable objective prior.

2.1 Invariant decision problems

The statistical decision problem of estimating the mean β of a normal distribution under squared error loss is invariant under translation.[13] In the absence of prior information, it is sensible that the prior distribution on β should be translation-invariant because that will imply that the Bayes rule is translation-invariant as well.[13] The uniform distribution is the only such distribution over the real numbers (up to a multiplicative constant).

Applied researchers are often particularly interested in the sign or the magnitude of a regression coefficient. This focus may be reflected formally by a loss function that depends on β only through its sign or magnitude. Under such loss functions, the decision problem is no longer invariant and the above argument no longer applies.

Suppose we want to estimate the magnitude of β. Since B is unbiased for β, it follows by Jensen's inequality that $|B|$ is positively biased for $|\beta|$. For fixed β, $|B|$ has the folded normal distribution with mean

$$E_\beta|B| = |\beta| + \sqrt{\frac{2}{\pi}}\, \text{se}\; e^{-\beta^2/2\text{se}^2} - 2|\beta|\Phi\left(-\frac{|\beta|}{\text{se}}\right) \tag{6}$$

The bias $E_\beta |B| - |\beta|$ is maximal at $\beta = 0$ when it is equal to $\sqrt{2/\pi}$ se ≈ 0.8 se. Now, if we use the flat prior, then the posterior distribution of β given $B = b$ is normal with mean x and standard deviation se. Hence, the posterior distribution of $|\beta|$ is the folded normal with mean

$$E(|\beta| | B = b) = |b| + \sqrt{\frac{2}{\pi}} \text{se} \, e^{-b^2/2\text{se}^2} - 2|b|\Phi\left(-\frac{|b|}{\text{se}}\right) \tag{7}$$

This posterior mean is the Bayes estimator of $|\beta|$ under squared error loss. We see that it is even larger than $|B|$, which is already positively biased for $|\beta|$. Evidently, the flat prior is very informative for $|\beta|$.

Similar problems arise with inference about the sign of β. If we use the flat prior then

$$P(\text{sgn}(\beta) = \text{sgn}(B)|B = b) = \Phi\left(\frac{|b|}{\text{se}}\right) \tag{8}$$

Now, the unimodal, symmetric priors are a natural class to consider for objective inference about the sign of β. The following proposition is essentially due to Casella and Berger.[14]

Proposition 1. *Suppose β has a prior distribution π which has a unimodal density and is symmetric about zero. Also, suppose that conditionally on β, B has the normal distribution with mean β and standard deviation se > 0. For any b*

$$P_\pi(\text{sgn}(\beta) = \text{sgn}(B)|B = b) \le \Phi\left(\frac{|b|}{\text{se}}\right) \tag{9}$$

The proposition asserts that the *perceived* evidence that β has the same sign as B is maximal under the uniform prior among all unimodal, symmetric priors. We conclude that the flat prior is very informative for inference about the sign. This issue is has long been recognized and was discussed, for instance, by Berger and Mortera.[8]

2.2 Parameterization invariance

Parameterization invariance refers to the principle that the construction of a non-informative prior should not depend on the parameterization that happens to have been chosen. This principle leads to Jeffreys rule[15] which is to define the prior as (proportional to) the square root of the Fisher information. This construction is unaffected by smooth reparameterization. In the normal location problem with known standard deviation, Jeffreys prior is the uniform distribution. We will now present a non-smooth reparameterization that does affect Jeffreys rule.

We can reparameterize the location problem quite naturally in terms of the sign and absolute value of β. This reparameterization is of course not smooth and the Fisher information is not defined because the sign is discrete. However, it is natural to put the Bernoulli(1/2) prior on the sign. Writing $\theta = |\beta|$, the distribution of B becomes a two-component mixture with density

$$f(b|\theta) = \frac{1}{2\,\text{se}} \varphi\left(\frac{b+\theta}{\text{se}}\right) + \frac{1}{2\,\text{se}} \varphi\left(\frac{b-\theta}{\text{se}}\right) \tag{10}$$

where φ is the standard normal density function. Jeffreys prior on θ is defined as

$$f(\theta) \propto \left(E_\theta \left(\frac{\partial}{\partial\theta} \log f(b|\theta) \right)^2 \right)^{1/2} \tag{11}$$

where

$$\frac{\partial}{\partial\theta} \log f(b|\theta) \propto \frac{-(b+\theta)\varphi((b+\theta)/\text{se}) + (b-\theta)\varphi((b-\theta)/\text{se})}{\varphi((b+\theta)/\text{se}) + \varphi((b-\theta)/\text{se})} \tag{12}$$

A closed form expression for (11) is not available, but we can easily compute it. We show Jeffreys prior for various values of se in Figure 2. We see that this prior is no longer uniform. In fact, it is even more widely dispersed.
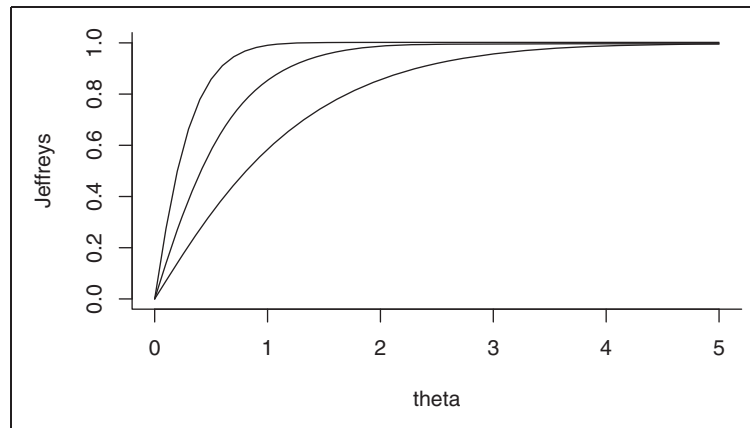
Figure 2. Jeffreys prior for $\theta = |\beta|$ when se $= 0.5, 1, 2$ (from left to right).

2.3 The parameter of interest

Almost all criteria that have been put forward for constructing a non-informative prior agree on the uniform distribution for a location parameter. However, this prior leads to overestimation of the magnitude and overconfidence about the sign. Here, we take a different approach.

Much research in the bio-medical and social sciences is aimed at assessing the evidence about the association between two variables, often after correcting for additional variables. The inferential approach that is most often used is to perform a hypothesis test. Hypothesis testing is ubiquitous, but it is not without criticism. We refer to the American Statistical Association's statement[16] for a discussion of the issues involved in the so-called "p-value debate". Much of the criticism of hypothesis testing centers on the fact that the p-value is often misinterpreted as the posterior probability that the null hypothesis of no association is true.

In many situations, it is unlikely that β is *exactly* zero. If one assumes from the outset that β is probably not exactly zero, then attention naturally shifts to inference about the sign of β.[9] John Tukey has argued this point particularly forcefully: "All we know about the world teaches us that the effects of A and B are always different – in some decimal place – for any A and B. Thus asking 'Are the effects different?' is foolish. What we should be answering first is 'Can we tell the direction in which the effects of A differ from the effects of B? In other words, can we be confident about the direction from A to B?".[17]

Following this reasoning, we propose that the (data-dependent) parameter $P(\beta > 0|B)$ is of primary importance in much empirical research. If indeed $P(\beta > 0|B)$ is of primary interest and if one wants to avoid undue influence of the prior, then it is quite natural to put the uniform prior on this parameter.[18–20] It is not immediately obvious which prior on β implies the uniform prior on $P(\beta > 0|B)$, but we have the following connection between the uniform and the normal distribution, which is new as far as we know.

Theorem 1. *Suppose the prior distribution of β is normal with mean 0 and standard deviation* se *and conditionally on β, B has the normal distribution with mean β and standard deviation* se. *Then $P(\beta > 0|B)$ has the standard uniform distribution.*

We believe that $N(0, \text{se}^2)$ is actually the only prior for β such that $P(\beta > 0|B)$ has the uniform distribution, but we have not been able to prove it. It is of course unique among all normal distributions. Now, on the basis of Theorem 1, we propose that the $N(0, \text{se}^2)$ prior on β is appropriate for objective Bayesian inference about the sign of β.

It is interesting to mention that our proposal would follow from an application of the reference method of Bernardo[21] to the parameter $P(\beta > 0|B)$. However, it should be stressed that the formal definition of the reference prior does not cover data-dependent parameters.

Recall that se represents the standard error of the estimator B and therefore depends on the sample size. Consequently, the proposed prior depends on the sample size as well. This may seem awkward from a subjective Bayesian point of view and requires justification. Now, objective Bayesian inference aims to avoid undue influence of the prior on the posterior. Since the prior exerts its influence on the posterior through the

likelihood, the choice of objective prior will often depend on the likelihood. While this may be at odds with the subjective Bayesian point of view, it is inherent in the goal of objectivity.[22]

A researcher's prior beliefs about β are surely not informed by the sample size, but it does work the other way around. It is typically the case that the researcher's prior beliefs influence the sample size through (formal or informal) sample size calculations. Therefore, it is not unreasonable for the reader of a bio-medical paper to take the sample size into account in his or her prior beliefs about β.

3 Empirical Bayes

The fact that we are considering a regression coefficient in the context of the bio-medical or social sciences may in itself be taken as prior information. Here, we will use external information from the extensive MEDLINE database to inform our prior beliefs, but do so in an objective manner. In the absence of additional prior information, we consider papers from MEDLINE to be *exchangeable*. This implies that we can use a sample of MEDLINE papers to estimate a suitable prior. Estimating the prior is often referred to as Empirical Bayes[6,7]

Let β_{ij} denote the i-th regression coefficient in paper j and assume the following hierarchical model.

- g_j has the normal distribution with mean g and variance σ^2
- Conditionally on g_j, β_{ij} has the normal distribution with mean zero and variance $g_j se_{ij}^2$
- Conditionally on β_{ij}, B_{ij} has the normal distribution with mean β_{ij} and variance se_{ij}^2

It follows that conditionally on g_j, the "z-value" $Z_{ij} = B_{ij}/se_{ij}$ has the normal distribution with mean zero and variance $g_j + 1$. The idea behind the model is that some studies have a larger sample size or a more predictable outcome than others, and that will lead to z-values of larger magnitude. To capture this in our model, we have the variance of the z-values depend on the study.

Now, conditionally on g_j, Z_{ij}^2 has the Gamma distribution with mean $g_j + 1$ and shape $1/2$. Therefore, based on a sample of z-values, we can estimate the parameters of this model by fitting a generalized linear mixed model with Gamma distribution, shape $1/2$, identity link and Gaussian random effect per study, to the squared z-values. If we use an offset of one, then the intercept of this model estimates g. Note that the prior we proposed in section 2.3 for objective inference about the sign has $g = 1$.

We will now describe how we collected our data.

3.1 MEDLINE

It is well known that for various reasons (p-hacking, fishing, file drawer effect, etc.) reported effects tend to be inflated.[23–27] We have tried to collect "honest" effects as follows. It is a fairly common practice in the life sciences to build multivariate regression models in two steps. First, the researchers run a number of univariate regressions for all predictors that they believe could have an important effect. Next, those predictors with a p-value below some threshold are selected for the multivariate model. While this approach is statistically unsound, we believe that the univariate regressions should be largely unaffected by selection on significance, simply because that selection is still to be done!

We entered the search term "univariate multivariate regression" into the PubMed system which yielded over 20,000 results. We selected the 80 most recent (in August 2018), consecutive results. Of these, 50 reported univariate and multivariate p-values. The 30 other papers were either unavailable under the license of the University of Leiden, or did not report p-values precisely (summaries such as $p < 0.01$ or $p < 0.05$ or "NS" are common). In case results were reported for more than one outcome, we used only the first. We collected in total 576 univariate p-values (all two-sided). We did not collect p-values for the intercept, or p-values based on F tests. Also, we did not collect p-values below 0.001. There are two reasons for that. First, because p-values below 0.001 are usually only reported as such. Second, because we would consider such a small p-value as evidence against our default prior and would not recommend its use in those cases, c.f. section 4. All data are available as a supplement to this article.

We converted the two-sided p-values to absolute z-values through $|z| = |\Phi^{-1}(p/2)|$ and display the result in Figure 3. The fact that we do not have the sign of the z-values is not relevant for our purpose, because we consider only symmetric distributions as plausible default priors.
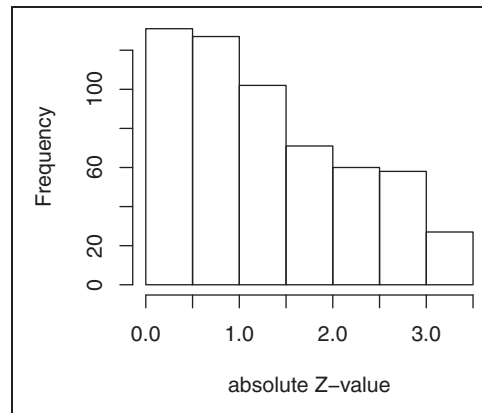
Figure 3. Histogram of 576 absolute z-values from 50 MEDLINE articles. Restricted to $[0, 3.29]$.

3.2 Results

We have fitted the Gamma regression model as described above to 576 squared z-values obtained from 50 MEDLINE articles. We are primarily interested in $\sqrt{g}$ which is the relation between the standard deviation of the prior and the standard error of the unbiased estimator. We estimate $\sqrt{g}$ to be 1.28 (95% CI: 1.12 to 1.42). The details of our analysis including the R code are in the Supplemental material. In addition to the (conditional) mixed model, we also fitted a marginal model. As the results were very similar, we do not report them here. They are included in the Supplemental Material.

Our proposal in section 2.3 for an objective prior for inference about the sign has $\sqrt{g} = 1$, while we estimate $\sqrt{g} = 1.28$ for a typical study. We consider this to be quite close, especially compared to the flat prior. Moreover, the value 1.28 is likely to be an overestimate, because it is safe to assume that even the univariate z-values we have collected are inflated. For example, researchers may have dichotomized certain predictor variables in a favorable way, or decided not to report predictors that showed no association with the outcome at all.

4 Discussion

The flat prior is used for Bayesian inference about regression coefficients in two distinct situations. It is used explicitly with the goal of objective Bayesian inference[5] and implicitly whenever the confidence interval for a regression coefficient is interpreted as a credibility interval. This is problematic, because the uniform distribution is not realistic at all in the context of bio-medical research. Consequently, its use leads to overestimation of the magnitude of the regression coefficient and overconfidence about its sign.

In this paper, we have proposed a different prior to be used as a default. Suppose we have an unbiased, normally distributed estimator B of β with standard error se. Then we have argued that the normal distribution with mean zero and standard deviation se is more suitable as a default prior than the uniform distribution. Equivalently, upon observing $B = b$, the normal distribution with mean $b/2$ and standard deviation $se/\sqrt{2}$ is a more suitable default posterior than the normal distribution with mean b and standard deviation se.

We based our proposal on an information theoretic argument (Theorem 1), but also demonstrated that our prior agrees quite well with data about regression coefficients we gathered from 50 papers from the bio-medical and social sciences. We do want to stress that our default prior is not meant to be a universal prior. There are, at least, four circumstances when our proposed default prior should *not* be used.

(1) Our prior is concentrated around zero which is usually not appropriate for the intercept as there is no reason a priori why the intercept should be close to zero.
(2) A different prior should be used when additional external prior information is available. We do feel, however, that it is the responsibility of the researcher to convince the reader (i.e. the scientific community) that his or her study is different from a typical MEDLINE study.
(3) Our proposed prior should not be used if a two-sided p-value less than 0.001 is observed. This is a form of prior-data conflict,[28] because such a small p-value is quite unlikely under our prior. To be precise, a two-sided

p-value less than 0.001 corresponds to a *z*-value exceeding 3.29 in absolute value, and the probability of that event is about 2% under our prior.

(4) Our proposed prior should not be used in situations in high dimensional situations where we can use empirical Bayes methods to reliably estimate a prior that is specific to the study under consideration. Many examples of such situations are discussed in a book by Efron about large scale inference.[7]

A limitation of our work is that have only considered the case where we have an unbiased, normally distributed estimator with known standard deviation. This is indeed restrictive, but note that the usual frequentist inference about regression coefficients almost always relies on the (asymptotic) normality and unbiasedness of their estimators. This is the case for linear and generalized linear models with and without random effects, as well as for generalized estimating equations (GEE) and quantile regression.[1]

The fact that the standard deviation of the estimator is typically unknown and must be estimated, can be taken into account – to some extent – as follows. Suppose we have observed $B = b$ and its associated two-sided *p*-value p. Then we can compute the absolute *z*-value as $|z| = |\Phi^{-1}(p/2)|$ and an "implied standard error" as $se = |b|/|z|$. This implied value may be larger than the estimated standard error, depending on how the *p*-value was calculated.

Declaration of conflicting interests

Funding

ORCID iD

Erik van Zwet http://orcid.org/0000-0001-5537-3179

Supplemental material

Supplemental material is available for this article online.

References

1. Stefanski LA and Boos DD. The calculus of m-estimation. *Am Stat* 2002; **56**: 29–38.
2. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016; **31**: 337–350.
3. Ghosh M, et al. Objective priors: An introduction for frequentists. *Stat Sci* 2011; **26**: 187–202.
4. Kass RE and Wasserman L. The selection of prior distributions by formal rules. *J Am Stat Assoc* 1996; **91**: 1343–1370.
5. Berger JO. The case for objective Bayesian analysis. *Bayes Analys* 2006; **1**: 385–402.
6. Carlin BP and Louis TA. *Bayes and empirical Bayes methods for data analysis*. New York: Chapman and Hall/CRC, 2010.
7. Efron B. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*. Vol 1. New York: Cambridge University Press, 2012.
8. Berger JO and Mortera J. Default Bayes factors for nonnested hypothesis testing. *J Am Stat Assoc* 1999; **94**: 542–554.
9. Gelman A and Tuerlinckx F. Type s error rates for classical and Bayesian single and multiple comparison procedures. *Computat Stat* 2000; **15**: 373–390.
10. Senn S. Trying to be precise about vagueness. *Stat Med* 2007; **26**: 1417–1430.
11. Gelman A, Jakulin A, Pittau MG, et al. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat* 2008; **2**: 1360–1383.
12. Seaman III JW, Seaman Jr JW and Stamey JD. Hidden dangers of specifying noninformative priors. *Am Stat* 2012; **66**: 77–84.
13. Berger JO. *Statistical decision theory and Bayesian analysis*. New York: Springer Science & Business Media, 2013.
14. Casella G and Berger RL. Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *J Am Stat Assoc* 1987; **82**: 106–111.
15. Jeffreys H. An invariant form for the prior probability in estimation problems. *Proc R Soc Lond A* 1946; **186**: 453–461.

16. Wasserstein RL and Lazar NA. The ASA's statement on *p*-values: context, process, and purpose. *The American Statistician* 2016; **70**: 129–133.

17. Tukey JW. The philosophy of multiple comparisons. *Statistical science* 1991; **6**: 100–116.

18. Bayes T, Price R and Canton J. *An essay towards solving a problem in the doctrine of chances*. London, UK: C. Davis, Printer to the Royal Society of London, 1763.

19. Laplace PS. *Philosophical essay on probabilities: translated from the fifth French edition of 1825 with notes by the translator*, volume 13. New York: Springer Science & Business Media, 2012.

20. Jaynes ET. Prior probabilities. *IEEE Transactions on systems science and cybernetics* 1968; **4**: 227–241.

21. Berger JO, Bernardo JM and Sun D. The formal definition of reference priors. *Ann Stat* 2009; **37**: 905–938.

22. Gelman A, Simpson D and Betancourt M. The prior can generally only be understood in the context of the likelihood. *Entropy* 2017; **19**: 1–13.

23. Rothstein HR, Sutton AJ and Borenstein M. *Publication bias in meta-analysis: prevention, assessment and adjustments*. Chichester: John Wiley & Sons, 2006.

24. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005; **294**: 218–228.

25. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* 2008; **19**: 640–648.

26. Button KS, Ioannidis JP, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013; **14**: 365.

27. Gelman A and Loken E. *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no fishing expedition or p-hacking and the research hypothesis was posited ahead of time*. Columbia University, 2013. Available at: http://www.stat.columbia.edu/˜gelman/research/unpublished/p_hacking.pdf.

28. Evans M, Moshonov H, et al. Checking for prior-data conflict. *Bayesian Analysis* 2006; **1**: 893–914.