Stockholms universitet

Research Report

Department of Statistics

Research Report

Department of Statistics

No. 2018:4

A Classroom Approach to the Construction of Bayesian Credible Intervals of a Poisson Mean

Per Gösta Andersson

+++++++++++++++

No. 2018:4

A Classroom Approach to the Construction of Bayesian Credible Intervals of a Poisson Mean

Per Gösta Andersson

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

A Classroom Approach to the Construction of Bayesian Credible Intervals of a Poisson Mean

Per Gösta Andersson, Department of Statistics, Stockholm University

Abstract

The Poisson distribution is here used to illustrate Bayesian inference concepts with the ultimate goal to construct credible intervals for a mean. The evaluation of the resulting intervals is in terms of potential negative effects of mismatched priors and posteriors. The discussion is in the form of an imaginary dialogue between a teacher and a student, who have met earlier, discussing and evaluating the Wald and score confidence intervals, as well as confidence intervals based on transformation and bootstrap techniques. From the perspective of the student the learning process is akin to a real research situation. By this time the student is supposed to have studied mathematical statistics for at least two semesters.

KEY WORDS: Gamma distribution; Posterior; Prior.

1. INTRODUCTION

For illustration of statistical theory and practice the Poisson distribution has proved to be of great value, due to its properties and simplicity, see e.g. Casella and Berger (2002) and Hogg, McKean and Craig (2005). This paper addresses Bayesian issues by way of a discussion between a teacher and a student. They meet on three occasions, when the student is gradually introduced to basic concepts. This ultimately leads to an understanding of the construction of credible intervals and their properties for, in this instance, a Poisson mean.
It is to be understood that during the meetings the teacher and student have access to a whiteboard to facilitate the interaction between them.

2. THE FIRST MEETING

Teacher: Once again we meet to discuss interval estimation and as before we are going to make use of the Poisson distribution for illustrative purposes. This time however we will not focus on confidence interval constructions, involving various types of approximations and transformations, but instead primarily deal with an altogether different approach, involving Bayesian credible intervals.
Student: I must admit that after our previous sessions I felt a certain weariness of confidence intervals in combination with the Poisson distribution. Though when you mention a Bayesian method, I become curious! Some of my recent courses included Bayesian ideas, but not in a very deep and systematic way. Primarily we studied procedures under normal distribution assumptions. As I understand it, the teachers were not fully committed Bayesians.
Reply: I believe that some scholars have fully adapted the Bayesian methodology and others consider themselves pragmatic and use Bayesian methods where they seem useful. There are also the "anti-Bayesians" who believe that the concept is fundamentally wrong, mostly due to its elements of subjectivity. A major complication for us now is that comparisons with our previous results using the classical so called frequentist approach are difficult to make, due to the fundamentally different interpretation of the unknown Poisson parameter θ from a Bayesian point of view.
Student: Well, that much I have understood. A Bayesian treats θ as the outcome of a random variable, say Θ, with a prior distribution, which is updated when we obtain data, thus arriving at a posterior distribution. The resulting inference is then conditional on observed data, but is that not always the case?
Reply: This is true enough, but a "frequentist" is of course not able to make a probabilistic statement about the **resulting** confidence interval in terms of θ. Now, to get things going, let us start with the fundamental idea of

Bayesian inference, which is the following application of Bayes' rule:

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{g(\mathbf{x})},$$

where $g(\mathbf{x})$ represents the joint marginal distribution of $\mathbf{X} = (X_1, \ldots, X_n)$, $\pi(\theta)$ the prior distribution of θ, $f(\mathbf{x}|\theta)$ the conditional distribution of $\mathbf{X}$ given θ and $\pi(\theta|\mathbf{x})$ the conditional posterior distribution of θ given $\mathbf{x}$.

Student: So here we assume a random sample $X_1, \ldots, X_n$, where $X_i \sim Poi(\theta)$, $i = 1, \ldots, n$. Thus

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} \exp(-\theta) = \frac{1}{\prod_{i=1}^{n} x_i!} \, \theta^{\sum_{i=1}^{n} x_i} \exp(-n\theta)$$

Reply: Yes, so far this is all we can say. But now we have the flexibility to choose a prior $\pi(\theta)$. Actually, we can start with a situation where we want to be neutral.

Student: I guess that a natural candidate as a prior of θ would then be a uniform continuous distribution, but in that case I must decide on an upper bound for the support.

Reply: Already we have another decision to make! But, as it turns out, $\pi(\theta)$ does not have to be a proper probability density function (pdf).

Student: In that case, I simply let $\pi(\theta) = c, \;\; 0 < \theta < \infty \;\; (c > 0)$.

Reply: This will work! We can furthermore note that $f(\mathbf{x}|\theta)\pi(\theta)$ represents the so called mixed discrete continuous joint pdf $f(\mathbf{x}, \theta)$. Also, $\pi(\theta)$ is an example of a flat and improper prior.

Student: But will $\pi(\theta|\mathbf{x})$ be a proper pdf if $\pi(\theta)$ is not a proper pdf? Do not say anything, I will check it!

$$f(\mathbf{x}|\theta)\pi(\theta) = \frac{1}{\prod_{i=1}^{n} x_i!} \, \theta^{\sum_{i=1}^{n} x_i} \exp(-n\theta) \cdot c$$

and

$$g(\mathbf{x}) = \int_0^\infty f(\mathbf{x}|\theta)\pi(\theta)\, d\theta = (\sum_{i=1}^{n} x_i = k) = \frac{c}{\prod_{i=1}^{n} x_i!} \int_0^\infty \theta^k \exp(-n\theta)\, d\theta$$

I can integrate $\theta^k \exp(-n\theta)$ by parts repeatedly. I recall from a math course on Fourier series something called the Kronecker lemma, which leads to

$$\begin{aligned} \int_0^\infty \theta^k \exp(-n\theta)\, d\theta &= [-\theta^k \frac{\exp(-n\theta)}{n} - k\theta^{k-1} \frac{\exp(-n\theta)}{n^2} - \cdots - k! \frac{\exp(-n\theta)}{n^{k+1}}]_0^\infty \\ &= \frac{k!}{n^{k+1}} \end{aligned}$$

So

$$g(\mathbf{x}) = \frac{c}{\prod_{i=1}^{n} x_i!} \, \frac{(\sum_{i=1}^{n} x_i)!}{n^{\sum_{i=1}^{n} x_i + 1}}$$

Thus

$$\pi(\theta|\mathbf{x}) = \frac{n^{\sum_{i=1}^n x_i+1}}{(\sum_{i=1}^n x_i)!}\, \theta^{\sum_{i=1}^n x_i} \exp(-n\theta)$$

From the integration just performed I can see that this is indeed a proper posterior pdf!
Reply: Good! Do you recognize the posterior distribution?
Student: Not immediately, but let me see (checking a list of distributions). It is a gamma $\Gamma(\alpha, \beta)$ with $\alpha = k + 1 = \sum_{i=1}^n x_i + 1$ and $\beta = 1/n$.
Reply: Correct. By the way, you used Kronecker's lemma (*Berlin sitzungsberichte* 1885 and 1889!) for the integration, but you might instead have considered manipulating the pdf of a gamma distribution.
Student: Of course, that old trick!

$$\int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp(-x/\beta)dx = 1$$

If I convert the notation to our situation: $x = \theta$, $\alpha = k + 1$, $\beta = 1/n$, so

$$\int_0^\infty \theta^k \exp(-n\theta)d\theta = \Gamma(k + 1) \left(\frac{1}{n}\right)^{k+1} \cdot 1 = \frac{k!}{n^{k+1}}$$

Reply: You can well imagine that integration to obtain a density soon gets really complicated for more complex situations, such as when we construct so called hierarchical models, where priors for hyperparameters, like α and β for a gamma distribution, are taken into account. Computer intensive methods like MCMC (Markov Chain Monte Carlo) to simulate samples from distributions are then of great help
If we continue to consider the gamma distribution, what happens if we let $\pi(\theta) \sim \Gamma(\alpha, \beta)$? ($\alpha$ and β are assumed to be known.)
Student: When $\pi(\theta)$ was flat, the result was a gamma distribution, so a not very wild guess would be that a gamma prior leads to a gamma posterior!
Reply: Correctly "guessed"! What are the resulting gamma parameters?
Student: If I start with the prior $\pi(\theta) \sim \Gamma(\alpha, \beta)$,

$$\begin{aligned} f(\mathbf{x}|\theta)\pi(\theta) &= \frac{1}{\prod_{i=1}^n x_i!} \theta^k \exp(-n\theta)\, \frac{1}{\Gamma(\alpha)\beta^\alpha}\, \theta^{\alpha-1} \exp(-\theta/\beta) \\ &= \frac{1}{\prod_{i=1}^n x_i!}\, \frac{1}{\Gamma(\alpha)\beta^\alpha}\, \theta^{k+\alpha-1} \exp(-(n + 1/\beta)\theta) \end{aligned}$$

I do not need to evaluate $g(\mathbf{x})$, since it does not depend on θ and is therefore not contributing to information about the parameters. The posterior is then $\Gamma(k + \alpha, 1/(n + 1/\beta))$.
Reply: You have illustrated that the family of gamma distributions is the conjugate family of distributions for the Poisson distribution. It is not unproblematic though to choose values or prior distributions of α and β for a gamma prior, unless we have vast experience and/or access to a great

amount of previous data. A compromise between the flat and gamma priors is given by $\pi(\theta) \propto 1/\sqrt{\theta}$. This could be reasonable if we tend to believe more in lower than in higher values of θ. Furthermore, we do not have to choose specific values of parameters. What is the posterior distribution given this prior?
Student:

$$f(\mathbf{x}|\theta)\pi(\theta) \propto \frac{1}{\prod_{i=1}^{n} x_i!}\,\theta^{\sum_{i=1}^{n} x_i}\exp(-n\theta)\,\frac{1}{\sqrt{\theta}} = \frac{1}{\prod_{i=1}^{n} x_i!}\,\theta^{\sum_{i=1}^{n} x_i - 1/2}\exp(-n\theta)$$

I trust that this is enough, verifying that we once again arrive at a gamma distribution, this time with $\alpha = \sum_{i=1}^{n} x_i + 1/2$ and $\beta = 1/n$, a result close to our first case with the flat prior.
Reply: I agree! Actually, this prior belongs to the class of Jeffreys priors, where $\pi(\theta) \propto \sqrt{I(\theta)}$, and where $I(\theta)$ is the Fisher information number. Check that!
Student: All right, $I(\theta) = E((\frac{\partial \text{log} f(X|\theta)}{\partial\theta})^2)$, but since the Poisson distribution belongs to an exponential family of distributions, we also have that $I(\theta) = -E(\frac{\partial \text{log} f(X|\theta)}{\partial\theta^2})$. Here $\text{log} f(X|\theta) = X\log\theta - \text{log} X_i! - \theta$, which means that $\frac{\partial}{\partial\theta^2}\text{log} f(X|\theta) = -X/\theta^2$ and $I(\theta) = 1/\theta$, so $\sqrt{I(\theta)} = 1/\sqrt{\theta}\,!$
Reply: Good! Furthermore we can observe that the class of Jeffreys's priors belongs to a type of priors which are said to be noninformative, in the sense that we have invariance under reparameterization.
Student: You say that the prior does not need to be flat in order to be called noninformative? What does the invariance property signify here?
Reply: Yes and the invariance property means that if we let $\tau = u(\theta)$, where u is a one-to-one function, then $\pi(\theta) \propto \sqrt{I(\theta)}$ implies $\pi(\tau) \propto \sqrt{I(\tau)}$. Can you show this?
Student: The density for $\pi(\tau)$, according to the general transformation "formula", should be $\pi(\theta)|\frac{\partial\theta}{\partial\tau}|$ $(\theta = u^{-1}(\tau))$ and $I(\tau) = E((\frac{\partial log f(X|u^{-1}(\tau))}{\partial\tau})^2)$. A lot of brackets!
Reply: Well, you need all of them!
Student: Then, using the chain rule I get $I(\tau) = E((\frac{\partial log f(X|\theta)}{\delta\theta}\frac{\partial\theta}{\partial\tau})^2) = (\frac{\partial\theta}{\partial\tau})^2 I(\theta)$ and since $\pi(\theta) \propto \sqrt{I(\theta)}$, I am done!
Reply: Excellent. We have not yet discussed the issue of constructing credible intervals for θ, but before doing that, let us briefly look at the expected value $E(\Theta|\text{x})$ for the three posterior distributions that you have derived.
Student: Generally for a $\Gamma(\alpha, \beta)$ the expected value is $\alpha \cdot \beta$, so

$$\begin{aligned} \pi(\theta)\text{ flat} : E(\Theta|\mathbf{x}) &= \frac{\sum_{i=1}^{n} x_i + 1}{n} \\ \pi(\theta) \propto I(\theta) : E(\Theta|\mathbf{x}) &= \frac{\sum_{i=1}^{n} x_i + 1/2}{n} \\ \pi(\theta) \sim \Gamma(\alpha,\beta) : E(\Theta|\mathbf{x}) &= \frac{\sum_{i=1}^{n} x_i + \alpha}{n + 1/\beta} \end{aligned}$$

The influence of the parameters on the priors decreases with increasing values of n. This makes sense!
Reply: Yes indeed! We can also observe that it seems relevant to call the first two priors uninformative. Furthermore the expectations can be seen as functions of the prior mean and the maximum likelihood estimate $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$.
Student: That escaped me regarding the last expression, but let me try and rewrite it:

$$\frac{\sum_{i=1}^{n} x_i + \alpha}{n + 1/\beta} = \frac{\beta\sum_{i=1}^{n} x_i}{n\beta + 1} + \frac{\alpha\beta}{n\beta + 1} = \bar{x}\frac{n\beta}{n\beta + 1} + \alpha\beta\frac{1}{n\beta + 1}$$

This is elegant! We get a weighted average of the sample mean $\bar{x}$ and the prior mean $\alpha \cdot \beta$, where the first weight tends to 1 and the second weight to 0 when $n \to \infty$.
Reply: The conditional expectation $E(\Theta|\mathbf{x})$ is actually a Bayes's point estimate of θ. It is formally derived as the decision function $\delta(\mathbf{x})$ which minimizes the conditional expectation of the loss function $L(\theta, \delta(\mathbf{x})) = (\theta - \delta(\mathbf{x}))^2$. (Ideally we should perhaps call this a predicted value of Θ rather than an estimated value of θ.)
We are now finally prepared for the Bayesian interval estimates of θ. They are called credible intervals and are constructed as posterior prediction intervals for Θ.
Student: Then if the posterior distribution is known, we can find, say, a and b so that

$$P(a|\mathbf{x} < \Theta < b|\mathbf{x}) = 1 - p$$

(Intentionally I avoided the use of α at the right hand side!)
Reply: Yes, and it is important to acknowledge that a and b depend on $\mathbf{x}$.
Student: And on α and β for the gamma prior!
Reply: True! As usual we can further choose, for the sake of uniqueness and symmetry, a and b so that

$$P(\Theta < a|\mathbf{x}) = P(\Theta > b|\mathbf{x}) = \frac{p}{2}$$

Bayes's interval estimation is really more straightforward than Bayes's point estimation, where in the latter situation we have to specify a loss function. It is also worth pointing out that $\bar{x}$ is a sufficient statistic for the Poisson parameter θ, so we can condition on $\bar{x}$ instead of $\mathbf{x}$.
Now I think it is time to study the behavior of credible intervals based on the prior and posterior distributions we have considered here. This could be accomplished analytically or by a simulation study, where you can labor with different scenarios. There will be many degrees of freedom for you!
Student: I accept the challenge!
Reply: Good! We will meet again in a few days and discuss how to accomplish this.

3. THE SECOND MEETING

So, have you given this enough thought, do you think?
Student: Well, at first it seemed a bit strange to evaluate credible intervals, since what we get in the end is an interval with a prespecified probability of coverage. Could the length be something to consider?
Reply: Impolitely I will reply to your question with another. Did you consider comparisons with confidence intervals for a fixed parameter?
Student: That was tempting and I have been looking into what Casella and Berger have to say about this very case with the Poisson distribution parameter in their "Statistical Inference". They make comparisons between confidence and credible intervals and I guess the main conclusion is that a confidence interval evaluated as a credible interval can perform poorly and vice versa. Is comparing these two concepts like comparing apples with oranges?
Reply: Not even that! I have a colleague who says that it is like comparing an apple with a lorry! It is really not very fruitful to make such comparisons, since the underlying concepts differ so much.
Student: Evaluating cases where we use the "wrong" prior might be something?
Reply: Yes, indeed! You have so far looked at three possible priors and why not combine "wrong" pairs of priors and posteriors? It is of course problematic to discuss in terms of "right" and "wrong" in a Bayesian context, but it seems to be a relevant way to illustrate the influence of the prior on the posterior.
Student: In that case I should first compute $a|\mathbf{x}$ and $b|\mathbf{x}$ given whatever p and prior I have and then use these limits to compute the probability of coverage using a "wrong" posterior? I take it that since the first two priors we have discussed are similar, I may consider only two of them?
Reply: Yes, and do not forget to try different sample sizes to observe how the effect of a "wrong" prior changes.
Student: Another thing, I have been thinking about the derivation of the posterior for the flat prior, where we obtained a gamma posterior. When I generate a random number I want to use a uniform distribution with some upper bound for the support and then the posterior will not exactly be a gamma distribution?
Reply: No, but in this case we can approximate with a $\Gamma(\sum_{i=1}^{n} x_i + 1, 1/n)$. The density function for the true posterior will have an adjustment factor, which is quite close to 1 even for moderate sizes of n.
Student: I think I know what to do now, so I will get back to my computer and MATLAB and hopefully return with some interesting results!
Reply: Good luck, see you next week!

4. THE THIRD MEETING

Teacher: Now I am curious about your results!
Student: I find them quite interesting even though they are not dramatic.
Reply: Something undramatic can be important too!
Student: Maybe! Anyway, there are some nice functions in MATLAB for generating values from distributions and computing probabilities and quantiles, so programming this was not hard work. For the flat prior I have used a uniform distribution on $(0, 100)$ to cover a wide enough range of possible values of θ. For the ”wrong” posterior as a first step I chose different combinations of α and β in such a way that the expected value $\alpha \cdot \beta$ equalled that of the prior (in this case 50). I have used the same $p = 0.95$ throughout and the sample sizes $n = 10$, 50 and 100. 10 000 Poisson samples were generated for each setup.
I could perhaps show the results of these cases first before I move on describing the other scenarios?
Reply: Please do!
Student: So, for the sake of completeness I include the results for all my chosen sample sizes, even if the level of inclusion is close to 0.95 already for $n = 10$.

Table 1: Empirical inclusion probabilities for credible intervals of a Poisson mean given $p = 0.95$ for the uniform $(0, 100)$ prior combined with the ”wrong” gamma prior $\Gamma(\alpha, \beta)$, where $\alpha \cdot \beta = 50$.

$n = 10$	
$\alpha = 5$, $\beta = 10$:	.944
$\alpha = 2.5$, $\beta = 20$:	.949
$\alpha = 0.25$, $\beta = 200$:	.949
$n = 50$	
$\alpha = 5$, $\beta = 10$:	.948
$\alpha = 2.5$, $\beta = 20$:	.950
$\alpha = 0.25$, $\beta = 200$:	.950
$n = 100$	
$\alpha = 5$, $\beta = 10$:	.949
$\alpha = 2.5$, $\beta = 20$:	.950
$\alpha = 0.25$, $\beta = 200$:	.949

Reply: By an inclusion probability you mean

$$\int_{a|\mathbf{x}}^{b|\mathbf{x}} f(\theta|\mathbf{x}) d\theta,$$

where $f(\theta|\mathbf{x})$ is a posterior based on a "wrong" prior?
Student: Yes, I admit I have not defined it properly, but I think *inclusion* is a proper expression, since θ is supposed to be random. *Coverage* does not seem suitable here.
Reply: I agree. Actually, in design-based survey sampling theory the inclusion probability equals the probability that a unit in the population is included in the random sample.
Looking at your results, in these cases there is not much penalty for assuming the "wrong" prior. Maybe you thought that we should get gradually deteriorating results for decreasing values of α?
Student: Yes, since the skewness coefficient for a $\Gamma(\alpha, \beta)$ is $2/\sqrt{\alpha}$ I expected that, but then I realized that one should also consider variances: $100^2/12 \simeq 833$ for the "true" uniform prior and $\alpha \cdot \beta^2 \simeq 500$ for the first gamma combination ($\alpha = 5$, $\beta = 10$) and 1000 for the second.
Reply: Yes, that could very well be the reason why we get this pattern of inclusion probabilities. Did you try some "uglier" scenarios?
Student: Well, the next setup was letting $\alpha \cdot \beta = 25$ with the following results:

Table 2: Empirical inclusion probabilities for credible intervals of a Poisson mean given $p = 0.95$ for the uniform $(0, 100)$ prior combined with the "wrong" gamma prior $\Gamma(\alpha, \beta)$, where $\alpha \cdot \beta = 25$.

$n = 10$	
$\alpha = 5$, $\beta = 5$:	.937
$\alpha = 2.5$, $\beta = 10$:	.947
$\alpha = 0.25$, $\beta = 100$:	.949

Reply: Still high inclusion probabilities!
Student: Next I tried $\alpha \cdot \beta = 10$:

Table 3: Empirical inclusion probabilities for credible intervals of a Poisson mean given $p = 0.95$ *for the uniform* $(0, 100)$ *prior combined with the "wrong" gamma prior* $\Gamma(\alpha, \beta)$*, where* $\alpha \cdot \beta = 10$*.*

$n = 10$	
$\alpha = 5$, $\beta = 2$:	.849
$\alpha = 2.5$, $\beta = 4$:	.924
$\alpha = 0.25$, $\beta = 40$:	.948
$n = 50$	
$\alpha = 5$, $\beta = 2$:	.928
$\alpha = 2.5$, $\beta = 4$:	.950
$\alpha = 0.25$, $\beta = 40$:	.950

Finally something substantially lower than 0.95!
Reply: However, you have to be pretty far off the mark with respect to the expected value to get this effect and then only for $n = 10$. There is clearly some stability here.
Student: After this I turned to the opposite situation, where the "true" prior is a $\Gamma(\alpha, \beta)$, while "falsely" assuming a flat improper prior. I started with the same pairs of α and β values as for the previous situation. Here is what happened for $n = 10$:

Table 4: Empirical inclusion probabilities for credible intervals of a Poisson mean given $p = 0.95$ *for a* $\Gamma(\alpha, \beta)$ *prior combined with the "wrong" assumption of a flat improper gamma prior.*

$\alpha \cdot \beta = 50, n = 10$	
$\alpha = 5$, $\beta = 10$:	.948
$\alpha = 2.5$, $\beta = 20$:	.949
$\alpha = 0.25$, $\beta = 40$:	.916

There is a tendency of decreasing inclusion probabilities for decreasing values of α, so then I tested $\alpha = 0.25$, $\beta = 100$, which gave me the inclusion probability 0.924 and the combination $\alpha = 0.25$, $\beta = 40$ resulted in 0.917. Still the inclusion probabilities are higher than 90%, but for $\alpha = 0.1$, $\beta = 100$ the result was 0.780. The same combination with $n = 50$ yielded 0.808 and with $n = 100$ I got 0.820. Thus, this is an example of a situation where we need a lot of information from the sample to "correct" for a misspecified prior.

Reply: You have clearly shown that one can construct examples of "wrong" priors, which lead to unreliable credible interval limits in terms of reduced inclusion probabilities. However, as we have mentioned earlier, the discussion of "right" or "wrong" priors is conceptually problematic. In practice the prior is chosen using information from previous similar situations and/or your own belief which we consider to be subjective.
Student: I guess there is some degree of subjectivity in all types of approaches including the frequentistic.
Reply: Yes, you can say that already in the choice of the approach to be used you are actually being subjective.
Student: This has really been an interesting scientific yourney for me! It is amazing how much you can illustrate in terms of statistical methodology just using the Poisson distribution.
Reply: It is certainly a very useful distribution. I am glad you have appreciated these excursions to "Poissonland"!

4. SUMMARY

Once again the Poisson distribution and its parameter θ turn out to be very useful for the illustration of statistical inference theory. Particularly when working within the Bayesian approach where several technical steps are required, we need a distribution which is simple to work with, yet inhabiting interesting properties.This paper is aimed at the construction of credible intervals and their properties when assuming "wrong" priors in some simple situations. The results indicate that even for small sample sizes there is a quite substantial robustness against misspecified priors.

REFERENCES

Casella, G., and Berger, R.L. (2002). *Statistical Inference* (2nd ed.), Pacific Grove CA: Duxbury Press.

Hogg, R.V., McKean J.W., and Craig A.T. (2005). *Introduction to Mathematical Statistics* (6th ed.), Pearson - Prentice Hall.