Cross Validated

Why is the Jeffreys prior useful?

Asked 7 years, 1 month ago Active 1 year, 3 months ago Viewed 18k times

61

I understand that the Jeffreys prior is invariant under re-parameterization. However, what I don't understand is why this property is desired.

Why wouldn't you want the prior to change under a change of variables?

bayesian prior

29

edited Oct 23 '13 at 15:31
Scorchi - Reinstate Monica ♦
24.4k 7 58 195

asked Oct 8 '12 at 22:57
tskuzzy
793 2 7 13

3 Of possible interest: Why are Jeffreys priors considered noninformative?. – user10525 Oct 9 '12 at 8:03

5 Answers

30

Let me complete Zen's answer. I don't very like the notion of "representing ignorance". The important thing is not the Jeffreys prior but the Jeffreys *posterior*. This posterior aims to reflect as best as possible the information about the parameters brought by the data. The invariance property is naturally required for the two following points. Consider for instance the binomial model with unknown proportion parameter θ and odds parameter $\psi = \frac{\theta}{1-\theta}$.

1. The Jeffreys posterior on θ reflects as best as possible the information about θ brought by the data. There is a one-to-one correspondence between θ and ψ. Then, transforming the Jeffreys posterior on θ into a posterior on ψ (via the usual change-of-variables formula) should yield a distribution reflecting as best as possible the information about ψ. Thus this distribution should be the Jeffreys posterior about ψ. This is the invariance property.
2. An important point when drawing conclusions of a statistical analysis is *scientific communication*. Imagine you give the Jeffreys posterior on θ to a scientific colleague. But he/she is interested in ψ rather than θ. Then this is not a problem with the invariance property: he/she just has to apply the change-of-variables formula.

edited Jan 30 '15 at 6:01
Avraham
2,490 16 29

answered Oct 9 '12 at 6:45
Stéphane Laurent
15.3k 5 48 92

Ah this clears things up a little bit. But is there an intuitively good reason why the posterior for the odds parameter should be the same as the posterior for the proportion parameter? That seems rather unnatural to me. – tskuzzy Oct 9 '12 at 7:15

It is not the same ! One is induced by the other by the change-of-variables formula. There is a one-to-one correspondence between the two parameters. Then the posterior distribution on one of these parameters should induce the posterior distribution on the other. – Stéphane Laurent Oct 9 '12 at 7:21

2 (+1) Stéphane. The OP seems to be still confused when he says "... should be the same...". The two posteriors are not "the same", what happens is that, for example, in Stéphane's example, you have that $P\{1/3 \leq \theta \leq 2/3 \mid X = x\} = P\{1/2 \leq \psi \leq 2 \mid X = x\}$; if you don't have this kind of consistency using defaults (computed) priors, then your priors are a little nutty. – Zen Oct 9 '12 at 18:05

1 I think what is missing from this post is that when there is a lot of information in the data about a parameter, the particular prior used doesn't really matter. For example, a binomial proportion, whether we use a uniform, jeffreys or haldane prior makes very little difference unless the posterior is very broad. In this case its a bit of an academic argument as to which prior is "right" because no meaningful conclusions can be drawn anyway. The real value of a non-informative prior is in multiple dimensions, but this problem has not been solved - Jeffreys prior is bad here. – probabilityislogic Oct 10 '12 at 11:06

3 This theory is incomplete and depends on parameter ordering, choice of compact region, and the likelihood function. So it doesn't obey the likelihood principle for example. Also, it is difficult to apply to non-independent data. Further, Bernardo's theory is only complete for 1-d parameter problems. It is probably the best method available currently though. A good competitor is the transformation group approach of Jaynes. – probabilityislogic Oct 10 '12 at 22:39

41

Suppose that you and a friend are analyzing the same set of data using a normal model. You adopt the usual parameterization of the normal model using the mean and the variance as parameters, but your friend prefers to parameterize the normal model with the coefficient of variation and the precision as parameters (which is perfectly "legal"). If both of you use Jeffreys' priors, your posterior distribution will be your friend's posterior distribution properly transformed from his parameterization to yours. It is in this sense that the Jeffreys' prior is "invariant"

(By the way, "invariant" is a horrible word; what we really mean is that it is "covariant" in the same sense of tensor calculus/differential geometry, but, of course, this term already has a well established probabilistic meaning, so we can't use it.)

Why is this consistency property desired? Because, if Jeffreys' prior has any chance of representing ignorance about the value of the parameters in an absolute sense (actually, it doesn't, but for other reasons not related to "invariance"), and not ignorance relatively to a particular parameterization of the model, it must be the case that, no matter which parameterizations we arbitrarily choose to start with, our posteriors should "match" after transformation.

Jeffreys himself violated this "invariance" property routinely when constructing his priors.

This paper has some interesting discussions about this and related subjects.

edited May 24 '15 at 14:25

answered Oct 9 '12 at 0:33

Zen
18.7k 3 57 101

1 +1: Good answer. But, why doesn't the Jeffreys' prior represent ignorance about the value of the parameters? – Neil G Oct 9 '12 at 5:50

4 Because it is not even a distribution. It is paradoxical to claim that a distribution reflects ignorance. A distribution always reflects information. – Stéphane Laurent Oct 9 '12 at 6:21

2 Another reference: projecteuclid.org/... – Stéphane Laurent Oct 9 '12 at 6:22

@StéphaneLaurent: One must have *some* belief even in a state of total ignorance. Whatever your posterior is minus whatever likelihood is induced by your data is the belief that you are assuming in that state of ignorance. The intuitive principle that must be respected when deciding that belief is that it should be invariant under changes of labels (including reparametrization). I'm not sure, but I think that principle alone (in all its possible interpretations — maximum entropy, invariant reparametrization, etc.) always decides the belief. – Neil G Jan 30 '15 at 7:24

Therefore, when one says "a distribution reflects ignorance", one means that the distribution concords with this principle. – Neil G Jan 30 '15 at 7:25

12

To add some quotations to Zen's great answer: According to Jaynes, the Jeffreys prior is an example of the principle of transformation groups, which results from the principle of indifference:

> The essence of the principle is just: (1) we recognize that a probability assignment is a means of describing a certain state i knowledge. (2) If the available evidence gives us no reason to consider proposition A_1 either more or less likely than A_2, then the only honest-way we can describe that state of knowledge is to assign them equal probabilities: $p_1 = p_2$. Any other procedure would be inconsistent in the sense that, by a mere interchange of the labels $(1, 2)$ we could then generate a new problem in which our state of knowledge is the same but in which we are assigning different probabilities…

Now, to answer your question: “Why wouldn't you want the prior to change under a change of variables?”

According to Jaynes, the parametrization is another kind of arbitrary label, and one should not be able to “by a mere interchange of the labels generate a new problem in which our state of knowledge is the same but in which we are assigning different probabilities.”

edited Oct 23 '13 at 15:32
Scortchi - Reinstate Monica ♦
24.4k 7 58 195

answered Oct 9 '12 at 6:15
Neil G
11.1k 2 35 75

2 Jaynes seems somewhat mystic to me. – Stéphane Laurent Oct 9 '12 at 6:24

@StéphaneLaurent: Maybe I was too easily converted then! But, I found this very convincing: E. T. Jaynes, “Where do we stand on Maximum Entropy?,” in The Maximum Entropy Formalism, R. Levine and M. Tribus, Eds. Cambridge, MA, USA: The MIT Press, 1979, pp. 15–118. – Neil G Oct 9 '12 at 6:25

2 Xian received a mail praising Jaynes: ceremade.dauphine.fr/~xian/critic.html It's a pity if you don't read French, this mail is both frightening and funny. The writer seems to have gone crazy by thinking too much about Bayesian statistics ;) – Stéphane Laurent Oct 9 '12 at 6:34

1 @StéphaneLaurent: Reading it now. This is absolutely right: "si vous affirmez en page 508 "the nonrepeatability of most experiments" à quoi bon ensuite "looking for optimal fequentist procedures" en page 512? Si la plupart des problèmes ne peuvent donc pas être traités par les procédures fréquentistes, comment le "choix Bayésien", qui se veut être le paradigme pour tout problème inférentiel, n'est-ce pas, peut-il se baser sur une réconciliation avec le fréquentisme (p. 517-518)? Pourquoi ne pas dire une fois pour toute qu'une probabilité n'est jamais une fréquence!" – Neil G Jan 30 '15 at 7:11

1 Also: "Le Principe du Maximum d'Entropie est lui absolument fondamental étant donné qu'il est nécessaire et suffisant pour régler ces cas d'école et que par conséquent il procure dans ces cas la signification véritable des probabilités a priori. Quand on sait qu'il permet ensuite d'unifier Théorie de l'Information, Mécanique Statistique, Thermodynamique…" describes my position as well. However, unlike the writer I have no interest in devoting hours convincing others to accept what I find so natural. – Neil G Jan 30 '15 at 7:18

4

While often of interest, if only for setting a reference prior against which to gauge other priors, Jeffreys priors may be completely useless as for instance when they lead to improper posteriors: this is for instance the case with the simple two-component Gaussian mixture

$$p\mathcal{N}(\mu_0, \sigma_0^2) + (1-p)\mathcal{N}(\mu_1, \sigma_1^2)$$

with all parameters unknown. In this case, the posterior of the Jeffreys prior does not exist, no matter how many observations are available. (The proof is available in a recent paper I wrote with Clara Grazian.)

answered Mar 30 '16 at 9:24
Xi'an
65.6k 8 105 400

-2

Jeffreys prior is *useless*. This is because:

1. It just specifies the form of the distribution; it does not tell you what its parameters should be.
2. You are never completely ignorant - there is always something about the parameter that you know (e.g. often it cannot be infinity). Use it for your inference by defining a prior distribution. Don't lie to yourself by saying that you don't know anything.
3. "Invariance under transformation" is not a desirable property. Your likelihood changes under transformation (e.g. by the Jacobian). This does not create "new problems," *pace* Jaynes. Why shouldn't the prior be treated the same?

Just don't use it.

answered Aug 17 '18 at 6:40
nec
1

1 Eh? Likelihood is not a density and won’t change under reparametrization – innisfree Jan 18 at 6:41