6. Jeffreys Prior I: Definition

Jeffreys Prior I: Definition

Jeffreys' prior

- **Jeffreys prior:**

$$\pi_J(\theta) \propto \sqrt{\det I(\theta)}$$

where $I(\theta)$ is the   matrix of the statistical model associated with $X_1, \ldots, X_n$ in the frequentist approach (provided it exists).

(Caption will be displayed when you start playing the video.)

- Bernoulli experiment: $\pi_J(p) \propto \frac{1}{\sqrt{p(1-p)}}$, $p \in (0,1)$: the prior is Beta(,   ).
- Gaussian experiment: $\pi_J(\theta) \propto 1$, $\theta \in \mathbb{R}$ is an   prior.

0:00 / 0:00   1.50x

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Jeffreys Prior

Jeffreys Prior is an attempt to incorporate frequentist ideas of likelihood in the Bayesian framework, as well as an example of a *non-informative prior*. This prior depends on the statistical model used for the observation data and the likelihood function. Mathematically, it is the prior $\pi_J(\theta)$ that satisfies

$$\pi_J(\theta) \propto \sqrt{\det I(\theta)},$$

where $I(\theta)$ is the **Fisher Information matrix** of the statistical model associated with $X_1, \ldots, X_n$ in the frequentist approach, provided that it exists.

In the one-variable case, Jeffreys prior reduces to

$$\pi_J(\theta) \propto \sqrt{I(\theta)}.$$

The Fisher information matrix $I(\theta)$ here is treated as a *linear transformation* matrix which maps one coordinate space to another (the logic behind such a mapping would be explained soon). In linear transformation terms, taking the determinant represents the ratio of volumes of corresponding spaces between coordinate system, which explains the intuition behind the use of $\det I(\theta)$.

Recitation Note : Two newly recorded recitations on Jeffreys prior are now available in the tabs after homework 9 and after midterm 2. The concepts discussed may be helpful for the upcoming lecture exercises.

Fisher Information and MLE Interpretation

1/1 point (graded)

Let our parameter of interest be θ. As computing Jeffreys prior makes use of the Fisher information $I(\theta)$, it is somehow related to the frequentist MLE approach (which has variance $I(\theta)^{-1}$). This yields interpretations of Jeffreys prior in terms of frequentist notions of estimation, uncertainty, and information.

For each statement, fill in the blank with the appropriate choice (more / less), then choose the option that represents your answers in order.

1. The Jeffreys prior gives more weight to values of θ whose MLE estimate has ____ uncertainty.

2. As a result, the Jeffreys prior yields more weight to values of θ where the data has ____ information towards deciding the parameter.

3. The Fisher information can be taken as a proxy for how much, at a particular parameter value θ, would equivalent shifts to the parameter influence the data. Thus, Jeffreys prior gives more weight to regions where the potential outcomes are ____ sensitive to slight changes in θ.

- ( ) more, more, more
- ( ) more, more, less
- ( ) more, less, more
- ( ) more, less, less
- (•) less, more, more
- ( ) less, more, less
- ( ) less, less, more
- ( ) less, less, less

✔

Solution:

- The weight given to a parameter value θ is the square root of its Fisher information $I(\theta)$, so more weight is given when $I(\theta)$ is high. The Fisher information is also the reciprocal of the MLE variance, so when the Fisher information is high, the MLE variance is low and thus the MLE has less uncertainty. Combining, we get that the Jeffreys prior gives more weight to values of θ whose MLE estimate has **less** uncertainty.

- Continuing from the above reasoning, when the MLE estimate has less uncertainty and we are able to estimate it more precisely. This corresponds to the data giving **more** information about the parameter when the Jeffreys prior yields larger values.

- Again, Jeffreys prior gives more weight to regions with high Fisher informations. By the given interpretation for the Fisher information, this means that at these areas, a small change to θ will influence the data relatively more, or in other words, potential outcomes are **more** sensitive to slight changes in θ.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Area Interpretation of Jeffreys Prior

0/1 point (graded)
We start with a fixed one-parameter statistical model where we use the MLE as our estimate, and consider the case where the number of samples n gets large. For each potential estimate θ, we construct using the asymptotic MLE variance the 95% confidence interval $X(\theta)$ centered at θ. Then, we consider the area over the interval $X(\theta)$ under the curve based on the Jeffreys prior. This area is is ____.

- the same regardless of θ. ✔
- significantly larger for values of θ where $I(\theta)$ is large.
- significantly larger for values of θ where $I(\theta)$ is small.

✖

Solution:

We work in the asymptotic view, where $n \to \infty$. Regardless of n, the Jeffreys prior is fixed, so as the interval $X(\theta)$ gets small (which always happens as $n \to \infty$ by asymptotic normality we can assume that the weight of Jeffreys prior over the interval is constant at $\sqrt{I(\theta)}$

The width of the interval is approximately $2 \cdot 1.96\sqrt{\frac{I(\theta)^{-1}}{n}}$ as the MLE has asymptotic variance $I(\theta)^{-1}$. Hence, the area under the interval based on Jeffreys prior pdf is $\left(\sqrt{I(\theta)}\right)\left(3.92\sqrt{\frac{I(\theta)^{-1}}{n}}\right) = \frac{3.92}{\sqrt{n}}$, which is **the same regardless of** θ.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 5 Bayesian statistics:Lecture 18: Jeffreys Prior and Bayesian Confidence Interval / 6. Jeffreys Prior I: Definition

Add a Post

Show all posts ▼ by recent activity ▼

? Area Interpretation: WHY can we use MLE asymptotic variance for other thetas 2

[Staff] Typo in solution for Area Interpretation of Jeffreys Prior 2

? Area under what curve? 3