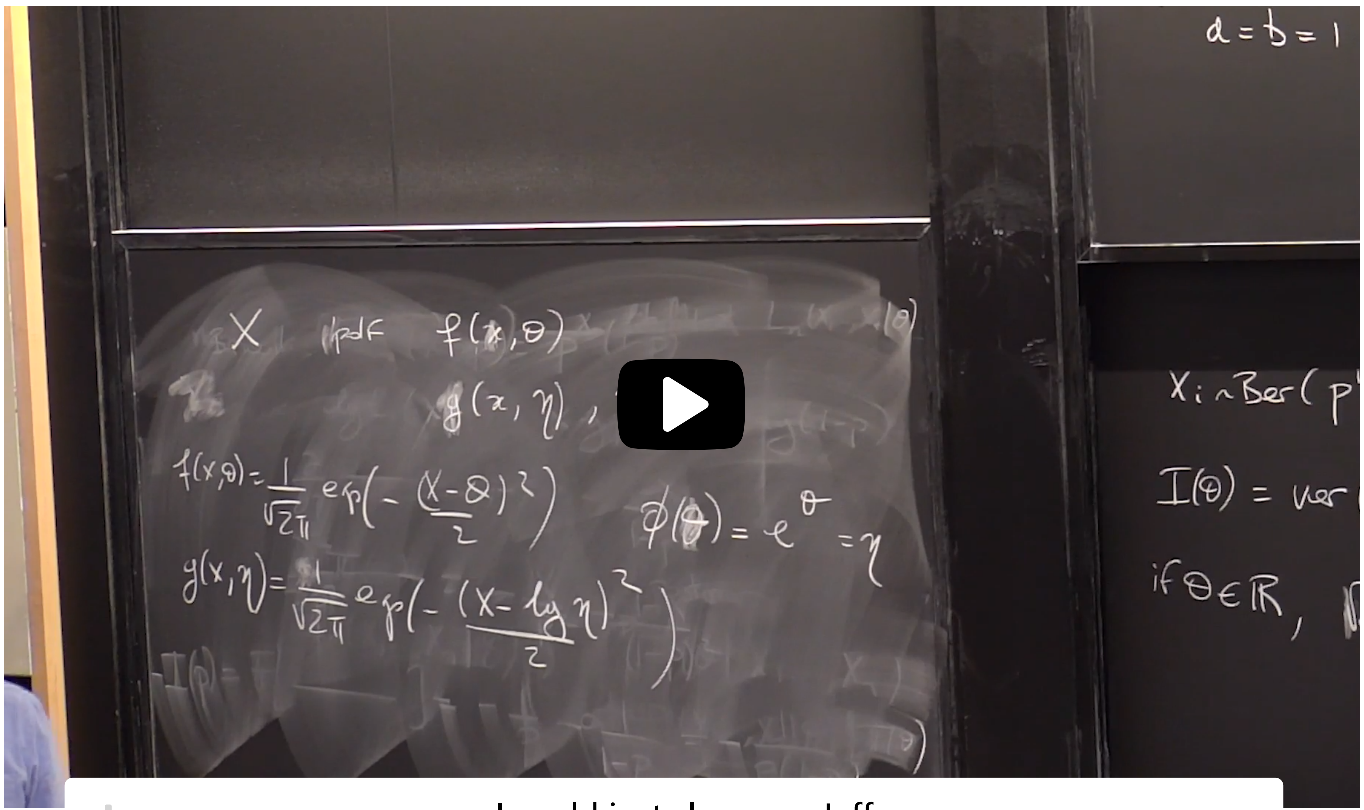
8. Jeffreys Prior III: Reparametrization Invariance

Jeffreys Prior III: Reparametrization Invariance



6:13 / 7:10 1.0x

computing the Fisher

information with respect to this new density,

so now I can compute L1 of x and eta.

I can take two derivatives, take the expectation,

and see what I'm getting.

If I start computing this Fisher information,

then I claim that the pdf of eta is actually--

which is just this transformation of theta--

is precisely this determinate.

End of transcript. Skip to the start.

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Reparametrization Invariance: Intuition

The lecture clip covered the **reparametrization invariance** property of Jeffreys prior. It states that if η is a reparametrization of θ (i.e. $\eta = \phi(\theta)$) for some one-to-one map ϕ), then the pdf $\tilde{\pi}(\cdot)$ of η satisfies

$$\tilde{\pi}(\eta) \propto \sqrt{\det \tilde{I}(\eta)}.$$

We examine the Jeffreys prior further. In the (typical) case where we have a single parameter, $\sqrt{\det \tilde{I}(\theta)}$ reduces to $\sqrt{\tilde{I}(\theta)}$. The Fisher information determines both the MLE asymptotic variance and Jeffreys prior, and as you've seen earlier is a measure of how *informative* the prior is towards the data. It in fact measures the how detectable marginal movements of θ are based on the observations.

This motivates the use of Jeffreys prior. The main motivation for using such a prior is because certain parametrizations may compress meaningful differences in θ into a small interval, whilst yielding large room for less impactful differences. In this case, a naive approach of using the uniform distribution would give an undue large weight to areas where modifying θ will not change the outcome much. Jeffreys prior directly adjusts for this through the Fisher information which is closely tied to MLE uncertainty.

This adjustment based on a quantitiative measure of uncertainty facilitates accurate conversion between parametrizations. Scaling based on the square root of the Fisher information allows us to abstract from an artificial view imposed by a particular parametrization into a universal measuring stick of parameter impact. The distribution given by Jeffreys prior is based on this universal measure, independent of our parametrization. As a result, regardless of the parametrization, Jeffreys prior would give the same distribution.

Now, it remains to explain why exactly the *square root* of the Fisher information was chosen. Recall that the asymptotic variance of the MLE is $I(\theta)^{-1}$. Then the uncertainty, in the same units as θ, is measured through the asymptotic standard deviation, which in turn is $I(\theta)^{-\frac{1}{2}}$. In the multidimensional case, the Fisher information approaches a multivariate Gaussian, where we have to take the square root of the asymptotic variance matrix in order to obtain an expression that's in the same units of the parameter v ector and thus quantifies uncertainty accordingly.

Reparametrization Invariance: Mechanics

1/1 point (graded)
Suppose a student claims that the reparametrization invariance principle allows us to do the following.

"Suppose that we have the Jeffreys prior for a statistical model using parameter θ, and we want to convert to parameter $\eta = \phi(\theta)$, where ϕ is an invertible function. Then we could simply substitute every occurrence of θ in the prior pdf with $\phi^{-1}(\eta)$ instead, and this would give us Jeffreys prior with parameter η."

Is the above approach correct? If not, what is/are the error(s)?

- [ ] The above approach is correct and would give us the correct Jeffreys prior with η as parameter.
- [ ] The above approach is incorrect as we are supposed to use $\phi(\eta)$ instead of $\phi^{-1}(\eta)$.
- [ ] The above approach is incorrect because the reparametrization invariance principle states that the Jeffreys prior is identical regardless of parameterization.
- [x] The above approach is incorrect because we have to also multiply by a factor of $\frac{d\theta}{d\eta} = \frac{1}{\phi'(\theta)}$ to obtain the correct Jeffreys prior. ✔
- [ ] The above approach is incorrect because the reparametrization invariance principle does not allow us to convert between parametrizations that have different Fisher information functions.

✔

$\phi(\theta) = \cancel{q}\,\theta^{\frac{1}{10}}$ $10 \cdot \cancel{q}\,\theta^{\frac{9}{10}} \to p^{10}$

Solution:

The error in the approach is that a crucial component of a distribution is the ***units*** in which it is written, which changes depending on the parametrization used. Therefore, if we are to rewrite the same distribution using a different parametrization, we have to convert between these units, which is done through the ratio $\boxed{\frac{d\theta}{d\eta}}$. As $\eta = \phi(\theta)$, $\frac{d\theta}{d\eta} = \frac{1}{\phi'(\theta)}$ is the correct interpreation of the ratio.

Hence, the fourth choice correctly notes the error. The other options do not correctly identify the sources of error and do not reflect proper use of Jeffreys prior or the reparametrization invariance property as per the given lecture.

Submit You have used 3 of 3 attempts

ℹ Answers are displayed within the problem

Reparametrization Invariance: Computation Example

2/3 points (graded)
We demonstrate the property of reparametrization invariance with a simple example on a Bernoulli statistical model. We start with the model $\mathsf{Ber}(q)$, which has parameter q. What is its Jeffreys prior? Express your answer as an un-normalized pdf $\pi(q)$ in proportionality notation such that $\pi(0.5) = 2$.

$\pi(q) \propto$ 1/sqrt(q*(1-q)) ✔ **Answer:** sqrt(1/(q*(1-q)))

$\frac{1}{\sqrt{q\cdot(1-q)}}$

Now, suppose that we write $q = p^{10}$ and thus wish to calculate the Jeffreys prior on the statistical model parametrized by p instead, i.e. $\mathsf{Ber}(p^{10})$. What is Jeffreys prior? Express your answer as an un-normalized pdf $\tilde{\pi}(p)$ in proportionality notation such that $\tilde{\pi}(2^{\frac{-1}{10}}) = 2^{\frac{1}{10}}$.

$\tilde{\pi}(p) \propto$ 1/(sqrt(p^10*(1-p^10))*1 ✖ **Answer:** sqrt((p^8)/(1-p^(10)))

$\frac{1}{\sqrt{p^{10}\cdot(1-p^{10})}\cdot 10\cdot p^9}$

STANDARD NOTATION

Convert the first form of Jeffreys prior (that is in terms of q) into the second form by writing q in terms of p and dq in terms of p and dp. Does it equal the expression for the Jeffreys prior you calculated in terms of p?

- [x] Yes ✔
- [ ] No

STANDARD NOTATION

Solution:

- As derived in lecture, from the model **Ber** (q) we calculate Jeffreys prior to be $\pi(q) \propto \sqrt{\frac{1}{q(1-q)}}$.

- The Fisher information $I(p)$ is defined to be $-\mathbb{E}\left[\frac{\partial^2}{\partial p^2}\log L(p|X)\right]$. Plugging in the expression $L(p|X) = (p^{10})^x(1-p^{10})^{1-x}$ for the likelihood, then evaluating, we get that

$$\begin{aligned} I(p) &= -\mathbb{E}\left[\frac{\partial^2}{\partial p^2}\log L(p|X)\right] \\ &= -\mathbb{E}\left[\frac{\partial^2}{\partial p^2}\left(x\log(p^{10}) + (1-x)\log(1-p^{10})\right)\right] \\ &= -\mathbb{E}\left[-\frac{10\left(p^{20}+p^{10}(9-11x)+x\right)}{p^2(p^{10}-1)^2}\right] \end{aligned}$$

Now noting that the expression is linear in x and using $\mathbb{E}[x] = p^{10}$, we get that

$$\begin{aligned} I(p) &= -\mathbb{E}\left[-\frac{10\left(p^{20}+p^{10}(9-11x)+x\right)}{p^2(p^{10}-1)^2}\right] \\ &= \frac{10\left(p^{20}+p^{10}(9-11p^{10})+p^{10}\right)}{p^2(p^{10}-1)^2} \\ &= \frac{10\left(10p^{10}-10p^{20}\right)}{p^2(p^{10}-1)^2} \\ &= \frac{100p^8}{1-p^{10}} \end{aligned}$$

Hence, $\tilde{\pi}(p) \propto \sqrt{I(p)} \propto \boxed{\sqrt{\frac{p^8}{1-p^{10}}}}$. It is easy to see that this satisfies the “un-normalization" condition $\tilde{\pi}(2^{\frac{-1}{10}}) = 2^{\frac{1}{10}}$, so this is our answer.

- Likewise, we can derive the expression $\tilde{\pi}(p) = \sqrt{\frac{p^8}{1-p^{10}}}$ right from $\boxed{\pi(q) \propto \sqrt{\frac{1}{q(1-q)}}}$. From $p^{10} = q$, differentiating both sides wrt p gives $\boxed{dq = (10p^9)\,dp}$. Hence, we substitute from $\sqrt{\frac{1}{q(1-q)}}dq$ using $dq = (10p^9)\,dp$ and $q = p^{10}$, to get:

积分换变量！

有点像chain rules

$$\begin{aligned} \sqrt{\frac{1}{q(1-q)}}dq &= \sqrt{\frac{1}{p^{10}(1-p^{10})}}\boxed{(10p^9)\,dp} \\ &\propto \sqrt{\frac{1}{p^{10}(1-p^{10})}}(p^9)\,dp \end{aligned}$$

$$= \sqrt{\frac{p^8}{1-p^{10}}},$$

which shows that these, indeed, are the same distribution under different parametrizations.

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

Discussion

Show Discussion

Topic: Unit 5 Bayesian statistics:Lecture 18: Jeffrey's Prior and Bayesian Confidence Interval / 8. Jeffreys Prior III: Reparametrization Invariance