Recitation 22: Reparametrization and Jeffreys Prior

Farrell Eldrian Wu

18.6501x Fall 2019

Contents

This set of recitation notes covers Recitation 22 on Jeffreys prior. The discussion here aims to reinforce the concepts covered in lecture by expanding on the intuition behind various topics in probability and statistics covered in this course. The mathematical arguments here will tend towards the intuitive rather than the rigorous side, while maintaining mathematical soundness. I hope that this will help improve your understanding of the Jeffreys prior and related concepts in statistics such as the Fisher information.

1 Reparameterizing a Distribution

We start by reviewing what it means to reparametrize a distribution, as this is a concept in probability central to the discussion on Jeffreys prior.

Suppose that we have a distribution $\pi(\theta)$ with a continuous PMF over the variable $\theta \in \mathbb{R}$. We reparametrize with a function $\phi : \mathbb{R} \to \mathbb{R}$ to the variable $\eta = \phi(\theta)$. Hence $\eta \in \mathbb{R}$ and we can define a distribution $\tilde{\pi}$ for η over $\mathbb{R}$. For simplicity, assume that ϕ is strictly monotone and continuously differentiable. We wish to compute the reparametrized distribution $\tilde{\pi}(\eta)$ in terms of the old distribution $\pi(\cdot)$ and the new variable η.

1.1 Reparametrization formula in one dimension

At first, the task does not seem very clear - what exactly does it mean to reparametrize a distribution? Let's consider a naive first attempt of just copying the value of the PMF at corresponding points, i.e. $\eta_0 = \phi(\theta_0) \to \tilde{\pi}(\eta_0) = \pi(\theta_0)$. We check whether it's even necessarily a probability distribution. Suppose we have $\phi(x) = 2x$. Then the distribution $\tilde{\pi}(\eta)$ will be the same as $\pi(\theta)$, just stretched by a factor of 2. Then this doubles the area under the PMF, which must be identically 1, so in this simple case of a linear transformation, we may not even get a probability distribution.

How can we fix this? If $\phi(x) = cx$ for some constant $c > 0$, then we can take the corresponding height from $\pi(\theta)$ then divide by c. For a given η, the corresponding θ is given by $\phi^{-1}(\eta)$, and thus the height is $\pi(\phi^{-1}(\eta)$. Dividing by c finally gives $\dfrac{\pi(\phi^{-1}(\eta))}{c}$. This ensures that the area under $\tilde{\pi}(\eta)$ is 1, while providing a direct link from $\tilde{\pi}(\eta)$ to the original distribution $\pi(\theta)$.

Study Question 1. What if $c < 0$ - does the formula still hold? If not, how can we modify it? What happens if $c = 0$?

Let's generalize this to monotone ϕ, not just linear functions. (For simplicity, we assume for now that ϕ is strictly increasing.) Suppose we have $\theta_0 < \theta_1$, $\eta_0 = \phi(\theta_0)$, $\eta_1 = \phi(\theta_1)$. We want the area under the curve in $\pi(\theta)$ from θ_0 to θ_1 to be the same as the area under the curve in $\tilde{\pi}(\eta)$ from η_0 to η_1. In integral form, this is

$$\int_{\theta_0}^{\theta_1} \pi(\theta)d\theta = \int_{\eta_0}^{\eta_1} \tilde{\pi}(\eta)d\eta. \tag{1}$$

Intuitively, this preserves interval probabilities after the change of variable, so this is a good first step.
Now how can we use this to compute $\tilde{\pi}(\eta)$? The equation (1) should hold for infinitesimal integrals, so consider a small step in θ, say from θ_0 to $\theta_0 + \Delta\theta$. The left hand side integral becomes

$$\int_{\theta_0}^{\theta_0+\Delta\theta} \pi(\theta)d\theta \tag{2}$$

As $\Delta\theta$ is infinitesimal, we can treat $\pi(\theta)$ to be approximately constant over the interval $[\theta_0, \theta_0+\Delta\theta]$, hence the integral in (2) evaluates to $\Delta\theta\ \pi(\theta_0)$. Similarly, the corresponding infinitesimal integral

$$\int_{\eta_0}^{\eta_0+\Delta\eta} \tilde{\pi}(\eta)d\eta \tag{3}$$

evaluates to $\Delta\eta\ \tilde{\pi}(\eta_0)$. Combining, we have

$$\Delta\theta\ \pi(\theta_0) = \int_{\theta_0}^{\theta_0+\Delta\theta} \pi(\theta)d\theta = \int_{\eta_0}^{\eta_0+\Delta\eta} \tilde{\pi}(\eta)d\eta = \Delta\eta\ \tilde{\pi}(\eta_0), \tag{4}$$

so

$$\tilde{\pi}(\eta_0) = \pi(\theta_0)\frac{\Delta\theta}{\Delta\eta}|_{\theta=\theta_0}. \tag{5}$$

What happens when ϕ is possibly decreasing? Then by our reasoning above, the (unsigned) area under the curve in the infinitesimal integral are $\pi(\theta)\,|\Delta\theta|$ and $\tilde{\pi}(\eta)\,|\Delta\eta|$. Equating them gives

$$\tilde{\pi}(\eta_0) = \pi(\theta_0)\left|\frac{\Delta\theta}{\Delta\eta}\right|_{\theta=\theta_0}. \tag{6}$$

We have determined the ratio between the PMF of the two distributions, so it remains to write our expression in terms of the variable η and the function ϕ. Let's start with the more complicated term, the ratio $\left|\frac{\Delta\theta}{\Delta\eta}\right|$. Our function ϕ is a function from θ to η, so the derivative $\phi'(\theta_0)$ is equal to $\left|\frac{\Delta\eta}{\Delta\theta}\right|_{\theta=\theta_0}$. Hence

$$\left|\frac{\Delta\theta}{\Delta\eta}\right|_{\theta=\theta_0} = \frac{1}{\phi'(\theta_0)}, \tag{7}$$

and plugging back to (6), then rewriting with the variables θ and η instead of θ_0 and η_0 we get

$$\tilde{\pi}(\eta) = \frac{\pi(\theta)}{|\phi'(\theta)|}. \tag{8}$$

We are almost there; the formula we want for $\tilde{\pi}(\eta)$ should only be in terms of η and ϕ, so the last step is to substitute $\theta = \phi^{-1}(\eta)$ into (8) to finally get

$$\boxed{\tilde{\pi}(\eta) = \frac{\pi(\phi^{-1}(\eta))}{|\phi'(\phi^{-1}(\eta))|}}. \tag{9}$$

Now we have a formula that we could plug π and ϕ into.

Study Question 2. Where does our derivation break down when we relax the assumption that ϕ is strictly monotone, say instead $\phi(\theta) = \theta^2$? What about if we allow ϕ to be weakly monotone, not necessarily strictly monotone?

1.2 Example: exponential distribution

Problem 1. Consider the the distribution $\pi(x) = e^{-x}$ $(x \in \mathbb{R}^+)$. Use the formula from the previous section to reparametrize into the variable y with the transformation function $y = \phi(x) = x^2$.

Solution 1. Adapting the formula from the previous section, we get

$$\tilde{\pi}(y) = \frac{\pi(\phi^{-1}(y))}{|\phi'(\phi^{-1}(y))|}. \tag{10}$$

Hence we need to compute $\phi'(\cdot)$ and $\phi^{-1}(\cdot)$, then afterwards we can just substitute the variables

in. $\phi^{-1}(y) = \sqrt{y}$, and $\phi'(x) = 2x$, so $\phi'(\phi^{-1}(y)) = 2\sqrt{y}$. Thus,

$$\tilde{\pi}(y) = \frac{\pi(\phi^{-1}(y))}{|\phi'(\phi^{-1}(y))|} = \frac{\pi(\sqrt{y})}{|2\sqrt{y}|} = \boxed{\frac{e^{\sqrt{y}}}{2\sqrt{y}}}. \tag{11}$$

Let's look into this calculation further by considering the CDF of $\pi(x)$ and $\tilde{\pi}(y)$. The CDF of $\pi(x)$ is

$$\int_{-\infty}^{x} e^{-x'} dx' = 1 - e^{-x}, \tag{12}$$

and the CDF of $\tilde{\pi}(y)$ is

$$\int_{-\infty}^{y} \frac{e^{\sqrt{y'}}}{2\sqrt{y'}} dy' = 1 - e^{-\sqrt{y}} \tag{13}$$

Notice that plugging $y = x^2$ into $1 - e^{-\sqrt{y}}$ in (13) gives exactly the expression $1 - e^{-x}$ in (12). The similar forms is not a coincidence. When we do a reparametrization, it is supposed to be the same distribution with a transformed variable. Thus if integrals have the same value across parametrizations, then the CDF, which is an integral, must also be the same after the change of variable.

Study Question 3. Repeat this example using the CDF interpretation: calculate the CDF of $\pi(x)$, rewrite x in terms of y, then differentiate with respect to y. How is this procedure related to the derivation of the reparametrization invariance formula in the previous section?

2 Jeffreys Prior

Recall from lecture that given a parametric statistical model parametrized by $\theta \in \mathbb{R}^d$, with Fisher information matrix $I(\theta) \in \mathbb{R}^{d \times d}$, the Jeffreys prior is defined to be the distribution

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

In this recitation, we focus on the one variable case. When $\theta \in \mathbb{R}$, $I(\theta)$ is just a single number, i.e. a 1×1 matrix, so its "determinant" is simply the number itself. Hence $\pi_J(\theta) \propto \sqrt{I(\theta)}$.

2.1 Intuition and MLE Interpretation

Jeffreys prior was designed to be a non-informative prior for the variable θ that corresponds to a statistical model parametrized by θ. The notion of being non-informative, in an informal sense of the term, is to not prefer any value of the parameter over another. One might then think a uniform prior is the way to go. But then, if we perform a reparametrization, the prior may no longer be uniform, so this breaks our first attempt of defining a non-informative prior.

The Jeffreys prior provides a resolution to this through its reparametrization invariance property, the proof of which is discussed in the next subsection. The definition of the Jeffreys prior in terms of the Fisher information relates the construction of the prior distribution to the likelihood function. In this subsection we discuss in more detail how this is accomplished, as well as how this is conceptually related to the frequentist concept of maximum likelihood estimation.

2.1.1 The choice of $\sqrt{I(\theta)}$

Why in particular do we choose $\sqrt{I(\theta)}$, not a simpler function, or even $I(\theta)$? We interpret $\sqrt{I(\theta)}$ in terms of the maximum likelihood estimator (MLE).

According to a theorem in in Unit 3, the asymptotic variance of the MLE is $I(\theta)^{-1}$, so $I(\theta)^{-\frac{1}{2}}$ can be interpreted as the standard error of a ML estimate. Given a fixed value for the parameter, the MLE will have an approximately Gaussian distribution with variance $I(\theta)^{-1}$. Thus when we construct confidence intervals of a certain confidence level, the width of these intervals will be a fixed multiple of the standard error. Hence, $I(\theta)^{-\frac{1}{2}}$ is in the same units as θ and thus represents the *radius* of uncertainty.

When we compare how much weight should be given to θ at different locations along the real line, we are somehow comparing how useful is the sample information at this particular location. One metric is by considering the radius of uncertainty, and then weighting it by the inverse of this radius. If it's more uncertain ($I(\theta)^{-\frac{1}{2}}$ large, hence $\sqrt{I(\theta)}$ small), we give a smaller weighting; if it's more certain ($I(\theta)^{-\frac{1}{2}}$ small, hence $\sqrt{I(\theta)}$ large), we give a larger weighting.

In summary, the reciprocal of the Jeffreys prior weighting is a concrete measure of the radius of uncertainty of the MLE.

2.1.2 Interpretation of $I(\theta)$

Recall the definition of the Fisher information

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln L(X_i|\theta)\right)^2\right] = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln L(X_i|\theta)\right].$$

We provide an interpretation of the Jeffreys prior for each of the two expressions for $I(\theta)$.

The first expression defines the Fisher information as the expected value of the squared derivative of the log likehood, at a particular θ. This is some metric of the effect of a marginal change of θ to the log likehood function as a whole, over the different possible values of X_i. Taking the expectation squared is also a common technique in statistics in order to account for both positive and negative changes.

We interpret this as the expected value of some actual change, by considering the impact of a marginal change $\Delta\theta$ to θ. Consider $I(\theta)(\Delta\theta)^2$, which by the definition of $I(\theta)$ is equal to $\mathbb{E}[(\Delta\ln L(X_i|\theta))^2]$, which can now be more easily seen as the expected value of the change to the log likelihood.

Let's keep this "amount of change", $\mathbb{E}[(\Delta\ln L(X_i|\theta))^2]$, constant across different values of θ. Then $\Delta\theta$ will be proportional to $I(\theta)^{-\frac{1}{2}}$. How is this related to information? Say if we need twice as much change in θ to achieve the same effect to the log likehood function at $\theta = \theta_0$ than at θ_1, then it's reasonable to claim that the amount of information (with respect to changes in θ) at $\theta = \theta_0$ is only half that at $\theta = \theta_1$. Hence $\sqrt{I(\theta)}$ is a measure of information compatible in scale with this "distance" interpretation.

Now we quickly discuss the second definition, which in words is the negative of the expectation of the second derivative of the log likelihood. The second derivative measures the curvature around θ. If the distribution is more curved, then it will be assigned a higher Fisher information and thus a higher Jeffreys prior weighting. This makes sense because having a quickly-changing likelihood function for different but close values of θ allows us to more accurately pin down the correct value of θ from the samples. Hence, $I(\theta)$ is greater where marginal changes to θ has a larger change to the likelihood function, which is similar to the interpretation for the first expression of $I(\theta)$.

Study Question 4. Consider the above discussion on the second expression for $I(\theta)$. If we care about how quick the likelihood function changes, why do we not use take the expectations of the first derivative instead? (Hint: Recall the proof in Unit 3 of why the two expressions for $I(\theta)$ are equal.) How does taking the second derivative still maintain this objective of evaluating sharp changes in the likelihood function?

2.1.3 Sensitivity and the Jeffreys prior weighting

In a statistical model $\theta \to \mathbb{P}_\theta$, different parametrizations may have different scales, and this is what leads to the Fisher information that relates different parametrizations. Jeffreys prior converts a parametric distribution into a uniform form by taking sensitivity into account. In a reparametrization from θ to η, some corresponding distances are squeezed while some may be stretched, but regardless of these distances in the parameter line, they still represent the same change in the likelihood function. If for example around θ_0 distances are stretched in the parametrization (i.e. $\phi'(\theta) > 1$), then a unit change in θ will correspond to a larger effect to the likelihood function than the same change in η.

To summarize, the weighting $\pi_J(\theta) \propto \sqrt{I(\theta)}$ gives higher weight to θ with a higher Fisher information, which is where

- marginal shifts have a relatively large effect to the sample values X_i, and
- the MLE of θ is more certain (as the asymptotic variance of the MLE is $I(\theta)^{-1}$).

2.2 Reparametrization Invariance Property

In this subsection, we state and prove the reparametrization invariance property of Jeffreys prior. We continue with our previous setup, where we have a model in θ and another in η, where the reparametrization from θ to η is given by $\eta = \phi(\theta)$. Denote the Fisher information for the model in θ to be $I_0(\theta)$ and the Fisher information for the model in η to be $I_1(\eta)$.

Theorem 1. (Reparametrization Invariance Property of Jeffreys Prior) Consider the Jeffreys prior distributions given by

$$\pi_J(\theta) \propto \sqrt{I_0(\theta)}$$

and

$$\tilde{\pi}_J(\eta) \propto \sqrt{I_1(\eta)},$$

where $I_0(\theta)$ and $I_1(\eta)$ are the Fisher information functions for the models parametrized by θ and η, respectively. Then the distribution $\tilde{\pi}_J(\eta)$ is the reparametrization of the distribution $\pi_J(\theta)$ after a change of variable $\eta = \phi(\theta)$.

Our approach is to characterize the relationship between $\pi_J(\theta)$ and $\tilde{\pi}_J(\eta)$ by examining the relationship between $I_0(\theta)$ and $I_1(\eta)$. Denote $L_0(X_i|\theta)$ and $L_1(X_i|\eta)$ to be the likelihood functions in the models parametrized by θ and η, respectively. Then writing out the formula for the Fisher information (using the first formula based on the square of the first derivative of the log likelihood) gives

$$I_0(\theta) = -\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln L_0(X_i|\theta)\right)^2\right] \tag{14}$$

and correspondingly,

$$I_1(\eta) = -\mathbb{E}\left[\left(\frac{\partial}{\partial\eta}\ln L_1(X_i|\eta)\right)^2\right] \tag{15}$$

Now, how can (14) and (15) be related? From the discussion in the previous subsection, we can relate them by the log likehood, which allows us to go back to the actual distributions of X_i, something which is not dependent on the choice of parametrization.

Let's write it out more concretely, using a similar style of reasoning as we did in the discussion on reparametrizing a distribution. Suppose that we have θ_0 and η_0 such that $\eta_0 = \phi(\theta_0)$. Take an infinitesimal change to θ, say $\theta_0 + \Delta\theta$, and write $\phi(\theta_0 + \Delta\theta) = \eta_0 + \Delta\eta$. Then, $\Delta\eta \approx \phi'(\theta_0)\Delta\theta$.

As $L_0(X_i|\theta)$ and $L_1(X_i|\eta)$ are likelihood functions for the same distribution (just under different parametrizations), we have

$$L_0(X_i|\theta) = L_1(X_i|\eta), \tag{16}$$

so

$$\ln L_0(X_i|\theta) = \ln L_1(X_i|\eta), \tag{17}$$

and similarly,

$$\ln L_0(X_i|\theta + \Delta\theta) = \ln L_1(X_i|\eta + \Delta\eta). \tag{18}$$

Hence subtracting (17) from (18) then writing the difference in the form of a derivative gives

$$\begin{aligned}\Delta\theta\frac{\partial}{\partial\theta}\ln L_0(X_i|\theta) &= \ln L_0(X_i|\theta + \Delta\theta) - \ln L_0(X_i|\theta) \\ &= \ln L_1(X_i|\eta + \Delta\eta) - \ln L_1(X_i|\eta) = \Delta\eta\frac{\partial}{\partial\eta}\ln L_1(X_i|\eta).\end{aligned} \tag{19}$$

Writing $\Delta\eta = \phi'(\theta)\Delta\theta$ (as our deltas are infinitesimal), and then cancelling out $\Delta\theta$ from both sides, we get

$$\frac{\partial}{\partial\theta}\ln L_0(X_i|\theta) = \phi'(\theta)\frac{\partial}{\partial\eta}\ln L_1(X_i|\eta) = \phi'(\phi^{-1}(\eta))\frac{\partial}{\partial\eta}\ln L_1(X_i|\eta) \tag{20}$$

Study Question 5. In the above calculations, we did not put absolute values around $\Delta\theta$ and $\Delta\eta$, unlike when we derived the reparametrization formula. Is our reasoning here still correct? If it is, what is the key difference between our use of delta expressions here and in the previous

section?

Substituting (20) into our expression for $I_1(\eta)$ (with the goal of relating this to $I_0(\theta)$, gives us

$$\begin{aligned} I_1(\eta) = -\mathbb{E}\left[\left(\frac{\partial}{\partial\eta}\ln L_1(X_i|\eta)\right)^2\right] &= -\mathbb{E}\left[\left(\frac{1}{\phi'(\phi^{-1}(\eta))}\frac{\partial}{\partial\theta}\ln L_0(X_i|\theta)\right)^2\right] \\ &= -\frac{1}{\phi'(\phi^{-1}(\eta))^2}\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln L_0(X_i|\theta)\right)^2\right] = \frac{I_0(\theta)}{\phi'(\phi^{-1}(\eta))^2}. \end{aligned} \tag{21}$$

Finally, we go back to the expression for the Jeffreys prior:

$$\tilde{\pi}(\eta) = \sqrt{I_1(\eta)} = \sqrt{\frac{I_0(\theta)}{\phi'(\phi^{-1}(\eta))^2}} = \frac{\pi(\theta)}{\sqrt{\phi'(\phi^{-1}(\eta))^2}} = \frac{\pi(\phi^{-1}(\eta))}{|\phi'(\phi^{-1}(\eta)|}, \tag{22}$$

which is precisely the reparametrization of $\pi(\theta)$ using the function ϕ in (9).

2.3 Example: Jeffreys prior for Ber(q^{10}) Model in Two Ways

Suppose we have a weighted coin that comes up heads with probability q. We flip the coin 10 times, then assign a binary random varible X to be 1 if all flips come up heads; otherwise, assign X to be 0. Then, $X \sim \text{Ber}(q^{10})$. Now, we want to define a non-informative prior on q based on this statistical model. For this, we use the Jeffreys prior.

We demonstrate the reparametrization invariance property of Jeffreys prior for the Ber(q^{10}) model by computing it in two ways. First, we calculate it directly from the definition, then second, we apply the reparametrization formula derived in the previous section to the Jeffreys prior for the Ber(p) model.

Problem 2. Compute the Jeffreys prior $\tilde{\pi}_J(q)$ for the Ber(q^{10}) statistical model directly from the definition

$$\tilde{\pi}_J(q) \propto \sqrt{I(q)}, \tag{23}$$

where $I(q)$ is the Fisher information for the Ber(q^{10}) statistical model.

Solution 2. Computing the Jeffreys prior boils down to calculating the Fisher information $I(q)$. As the Fisher Information is based on the likelihood, we first write out the likelihood $L(X_i|q)$ for the Ber(q^{10}) model.

From the definition of the Bernoulli model, X_i has PMF $L(X_i|q) = \begin{cases} q^{10}, & \text{for } x = 1 \\ 1 - q^{10}, & \text{for } x = 0. \end{cases}$

An equivalent form more suitable for algebraic manipulation and taking derivatives is

$$L(X_i|q) = (q^{10})^x(1 - q^{10})^{1-x} \tag{24}$$

Now we are ready to compute the Fisher information $I(q)$. Writing out the formula for $I(q)$ (we use the second formula based on the second derivative of the log likelihood) and then

substituting, we get

$$\begin{aligned} I(q) &= -\mathbb{E}\left[\frac{\partial^2}{\partial q^2}\ln L(X_i|q)\right] \\ &= -\mathbb{E}\left[\frac{\partial^2}{\partial q^2}\ln((q^{10})^x(1-q^{10})^{1-x})\right] \\ &= -\mathbb{E}\left[\frac{\partial^2}{\partial q^2}(10x\,\ln(q)+(1-x)\,\ln(1-q^{10}))\right] \\ &= -\mathbb{E}\left[\frac{10(-11xq^{10}+x+q^{20}+9q^{10})}{q^2(1-q^{10})^2}\right] \end{aligned} \tag{25}$$

This is an expectation of a linear "function" in x. Hence to evaluate this we simply note that $\mathbb{E}[x] = q^{10}$, replace every instance of x with q^{10}, then remove the outer expectation. Substituting, we thus get

$$\begin{aligned} I(q) &= -\mathbb{E}\left[\frac{10(-11xq^{10}+x+q^{20}+9q^{10})}{q^2(1-q^{10})^2}\right] \\ &= -\frac{10(-11(q^{10})q^{10}+(q^{10})+q^{20}+9q^{10})}{q^2(1-q^{10})^2} \\ &= \frac{100q^{10}-100q^{20}}{q^2(1-q^{10})^2} = \frac{100q^8}{1-q^{10}} \end{aligned} \tag{26}$$

Hence the Jeffreys prior is

$$\tilde{\pi}_J(q) \propto \sqrt{I(q)} = \sqrt{\frac{100q^8}{1-q^{10}}} \propto \boxed{\sqrt{\frac{q^8}{1-q^{10}}}}, \tag{27}$$

where the last step is a purely cosmetic simplification.

Problem 3. Given that the Jeffreys prior for the Ber(p) model is

$$\pi_J(p) \propto \frac{1}{p(1-p)}, \tag{28}$$

compute the Jeffreys prior $\tilde{\pi}_J(q)$ for the Ber(q^{10}) model by reparametrizing $\pi_J(p)$ with the change of variables function $\phi(q) = p^{\frac{1}{10}}$.

Solution 3. The reparametrization invariance principle states that the Jeffreys prior $\tilde{\pi}_J(q)$ is the same distribution as $\pi_J(p)$ reparametrized with $\phi(q) = p^{\frac{1}{10}}$.

To begin, we compute the functions

$$\phi'(x) = \frac{1}{10} p^{-\frac{9}{10}} \tag{29}$$

and

$$\phi^{-1}(x) = x^{10}. \tag{30}$$

Then, we use the reparametrization formula (9):

$$\tilde{\pi}(q) = \frac{\pi(\phi^{-1}(q))}{|\phi'(\phi^{-1}(q))|} = \frac{\pi(q^{10})}{|\phi'(q^{10})|} = \frac{\frac{1}{\sqrt{q^{10}(1-q^{10})}}}{\left|\frac{1}{10} q^{-9}\right|} = \frac{10q^9}{\sqrt{q^{10}(1-q^{10})}} = \sqrt{\frac{100q^8}{1-q^{10}}} \propto \boxed{\sqrt{\frac{q^8}{1-q^{10}}}} \tag{31}$$

As the distribution $\tilde{\pi}_J(q)$ is written in proportionality notation, all we need for two expressions to represent the same distribution is for each to be a constant multiple (not depending on the variable in which the distribution is over, but may depend on other parameters) of the other. Comparing the expressions for $\tilde{\pi}_J(q)$, we confirm that the two procedures indeed produce the same Jeffreys prior.

Study Question 6. Proportionality notation is typically used to avoid the need to write cumbersome constant factors that do not affect probability ratios. Define the *proportionality constant* to be the number that a distribution written in proportionality notation has to be multiplied to in order for it to integrate to 1.

Suppose we apply the reparametrization formula (9) derived in the previous section, to a distribution written in proportionality notation. What is the relationship between the proportionality constant in the old distribution and the proportionality constant in the reparametrized distribution?