3. Choosing a Prior

Choosing a Prior

9:32 / 9:32 1.0x

So it's the joint distribution of the xi.

So if I were to integral-- integrate this

with respect to the xi, I would certainly get 1.

But what I'm interested in is integrating with respect to p.

So there's no reason why this thing should integrate to 1.

And so the coefficient of proportionality

we have is just the integral of the log likelihood with respect

to p.

End of transcript. Skip to the start.

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Uniform Priors: True or False

1/1 point (graded)

Select from the following statements the **true** ones for uniform priors. (In this question, we also allow *improper* priors.)

- [ ] They can be defined only on parameter sets Θ with a finite number of possible values.
- [ ] They should integrate to 1 (or if the distribution is discrete; should sum to 1)
- [x] They reflect an equal belief in each possible hypothesis. ✔
- [x] The maximum a-posteriori and maximum likelihood estimators when using such a prior would always be the same. ✔

✔

Solution:

- The first choice is false. As discussed in the lecture, they can be defined on infinite sets or even non-discrete distributions with an uncountably infinite number of possible parameter values.
- The second choice is also false. If $\pi(\cdot)$ is improper, then it will definitely not integrate to 1 by definition.
- The third choice is correct. A uniform prior reflects an "equal" belief in each of the possible hypothesis.

- The last choice is also correct. Recall that the maximum-a-posterior estimator maximizes $\pi(\theta|X_1,\ldots,X_n)$ while the maximum likelihood estimator maximizes $L_n(X_1,\ldots,X_n|\theta)$, both taken as functions of θ. By Bayes' rule, we have that

$$\pi(\theta|X_1,\ldots,X_n) \propto L_n(X_1,\ldots,X_n|\theta)\,\pi(\theta) \propto L_n(X_1,\ldots,X_n|\theta),$$

where the first proportionality is due to Bayes' rule and the second proportionality is due to $\pi(\cdot)$ being uniform. The two statistics, when taken as a function of θ, are therefore identical up to a constant of proportionality. Hence, while the maximum values might be different, the value of θ attaining the maximum for both quantities are the same.

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

Beta Distribution: True or False

1/1 point (graded)
One specific prior discussed in the previous lecture is the Beta distribution, which was then demonstrated in a scenario with a Bernoulli statistical model. **Which of the follwing statements is/are true about the Beta distribution**, written as Beta (α,β) $\propto p^{\alpha-1}(1-p)^{\beta-1}$?

- [x] The Beta distribution is very suited to models where our parameter represents a probability due to its support being $[0,1]$. ✔
- [ ] The Beta distribution is very suited to models where our parameter represents a probability due to its maximum always being close to $\frac{1}{2}$.
- [x] The Beta distribution is very suited to models where our parameter represents a probability because multiplying it by p or $1-p$ simply involves incrementing the respective parameter. ✔

✔

Solution:

The first and third statements are correct.

- **The first statement is correct.** The Beta distribution indeed has support on the interval $[0,1]$, which is also the range of possible probabilities. Thus, using the Beta distribution to model possible parameters p would allow us to exactly cover the feasible range.
- **The second statement is incorrect.** The Beta family of distributions is very flexible and does not constrain us to symmetric shapes (which happens if we instead use a Gaussian prior). Indeed, if you recall the calculation of modes from the previous lecture, the mode can be at 0 or 1 for certain special cases. In the general case $\alpha>1, \beta>1$, the mode is at $\frac{\alpha-1}{\alpha+\beta-2}$, which could range anywhere in $(0,1)$ depending on α and β.
- **The third statement is correct.** Mathematically, it is easy to see that multiplying the PDF of a Beta distribution by p or $1-p$ increments the α or β coefficient, respectively, by 1. In practical terms, it is very common in statistical applications, as you've seen in the previous lecture, to multiply the likelihood function by either p or $1-p$ depending on the outcome of a binary trial.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Show Discussion

Topic: Unit 5 Bayesian statistics:Lecture 18: Jeffrey's Prior and Bayesian Confidence Interval / 3. Choosing a Prior