4. Bayesian Estimation

Setup:

In this problem, we will explore the intersection of Bayesian and frequentist inference. Let $X_1, X_2, \cdots, X_n \stackrel{\text{i.i.d}}{\sim} \mathsf{N}(0, \theta)$, for some unknown positive number θ, which is our parameter of interest. Suppose that we are unable to come up with a prior distribution for θ.

(a)

1.0/1 point (graded)

Compute the maximum likelihood estimator of θ. You may use the variables n, $\sum_{i=1}^n X_i$, and $\sum_{i=1}^n X_i^2$.

(Enter **Sigma_i(X_i)** for $\sum_{i=1}^n X_i$ and **Sigma_i(X_i^2)** for $\sum_{i=1}^n X_i^2$. Do not worry if the parser does not render properly; the grader works independently. If you wish to have proper rendering, enclose **Sigma_i(X_i)** by brackets.)

$\hat{\theta}^{\text{MLE}} =$ (Sigma_i(X_i^2))/n ✔ **Answer:** Sigma_i(X_i^2)/n

STANDARD NOTATION

Solution:

$$\hat{\theta} = S_n = \frac{1}{n}\sum_i (X_i)^2$$

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

(b)

2/2 points (graded)

Is the MLE $\hat{\theta}^{\text{MLE}}$ asymptotically normal?

- (•) It is asymptotically normal ✔
- ( ) It is **not** asymptotically normal

If it is asymptotically normal, what is its asymptotic variance $V(\theta)$? If it is not asymptotically normal, type in 0.

$V(\theta) =$ 2*theta^2 ✔ **Answer:** 2*theta^2

$2 \cdot \theta^2$

STANDARD NOTATION

Solution:

We can construct L_n and l as following.

$$L_n\left(x_1,\ldots,x_n|\theta\right)=\left(2\pi\theta\right)^{-n/2}\exp\{-\sum_i\frac{1}{2\theta}x_i^2\}$$

$$l=\log L_1=-\frac{1}{2}\log\left(2\pi\theta\right)-\frac{1}{2\theta}x_1^2$$

$$\frac{\partial l}{\partial\theta}=-\frac{1}{2\theta}+\frac{1}{2\theta^2}x_1^2$$

$$\frac{\partial^2 l}{\partial\theta^2}=\frac{1}{2\theta^2}-\frac{1}{\theta^3}x_1^2$$

$$I\left(\theta\right)=\frac{1}{2\theta^2}$$

Therefore, we can conclude that it is asymptotically normal, and its asymptotic variance is $2\theta^2$.

$$\sqrt{n}\left(\hat{\theta}-\theta\right)\xrightarrow[n\to\infty]{(d)}\mathcal{N}\left(0,2\theta^2\right)$$

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

(c)

1.0/2 points (graded)

Let's take a Bayesian approach here to arrive at an estimator.

Perform the following steps:

- Compute Jeffreys prior.
- Use Bayes formula to compute the posterior distribution.
- From the posterior distribution, compute the Bayesian estimator of θ. Recall that this is defined in lecture to be the mean of the distribution.

What is the Bayesian estimator $\hat{\theta}^{\text{Bayes}}$?

(Enter **Sigma_i(X_i)** for $\sum_{i=1}^n X_i$ and **Sigma_i(X_i^2)** for $\sum_{i=1}^n X_i^2$. Do not worry if the parser does not render properly; the grader works independently. If you wish to have proper rendering, enclose **Sigma_i(X_i)** and **Sigma_i(X_i^2)** by brackets.)

$\hat{\theta}^{\text{Bayes}}=$ (Sigma_i(X_i^2))/4 ✖ **Answer:** Sigma_i(X_i^2)/(n-2)

In this Bayesian problem, which, if any, of the prior or the posterior, is proper?

- The prior only.
- The posterior only. ✔

- [ ] Both the prior and the posterior.

- [ ] Neither the prior nor the posterior.

STANDARD NOTATION

Solution:

1. Compute Jeffreys prior. Is it proper?

$$\pi_j(\theta) \propto \sqrt{det\left(I\left(\theta\right)\right)} = \frac{1}{\sqrt{2\theta}}$$

Since $\frac{1}{\sqrt{2\theta}}$ integrates to infinity, the prior is improper.

2. Use Bayes formula in order to compute the posterior distribution.

$$\pi\left(\theta | X_1, \ldots, X_n\right) \propto \pi\left(\theta\right) L_n\left(x_1, \ldots, x_n | \theta\right) \propto \frac{\left(2\pi\theta\right)^{-n/2}}{\sqrt{2\theta}} exp\{-\sum_i \frac{x_i^2}{2\theta}\} \propto \theta^{-(n+2)/2} exp\{-\sum_i \frac{x_i^2}{2\theta}\}$$

The posterior distritubion is Inverse Gamma with parameters $\boxed{\alpha = \frac{n}{2}, \beta = \frac{1}{2}\sum_i X_i^2}$

Mean	$\frac{\beta}{\alpha - 1}$ for $\alpha > 1$

3. Compute the Bayesian estimator of θ associated with Jeffreys prior.

$$\boxed{\hat{\theta} = \int \theta\pi\left(\theta | X_1, \ldots, X_n\right) d\theta = \frac{1/2\sum X_i^2}{n/2 - 1} = \frac{\sum X_i^2}{n - 2}}$$

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

(d)

0 points possible (ungraded)
Consider the set of statements given below. Decide whether each of the statements below is true. If a statement is true, decide whether this reflects a Bayesian or frequentist property of the Bayesian estimator.

Note: This problem is about the Bayesian estimator that you obtained in (c).

A **frequentist property** refers to all estimator properties that were considered before this lecture and are used in the context where there is a fixed, true, parameter value, and we want our estimator to approximate this value.

On the other hand, a **Bayesian property** refers to properties that indicate that we weight the likelihood somehow, not just using the raw values of the likelihood. (This could involve no additional judgement on the value of the parameter, for example if we use Jeffreys prior.)

Which statements are true and reflects a Bayesian property of the Bayesian estimator in (c)?

- [ ] The Bayesian estimator is consistent.

- [ ] The Bayesian estimator is unbiased.

- [x] The Bayesian estimator gives in expectation a larger estimate if there are few observations, given a fixed θ, due to the nature of the prior used. ✔

- [ ] The Bayesian estimator does not assume any particular prior distribution that is independent of the conditional likelihood.

- [ ] The Bayesian estimator is asymptotically normal.

✔

Which statements are true and reflects a frequentist property of the Bayesian estimator in (c)?

- [ ] The Bayesian estimator is consistent. ✔

- [ ] The Bayesian estimator is unbiased.

- [ ] The Bayesian estimator gives in expectation a larger estimate if there are few observations, given a fixed θ, due to the nature of the prior used.

- [ ] The Bayesian estimator does not assume any particular prior distribution that is independent of the conditional likelihood. ✔

- [x] The Bayesian estimator is asymptotically normal. ✔ 这两个是频率方法的特性

✖

Solution:

X_i is drawn from the distribution $\mathsf{N}\,(0,\theta)$, so we compute

$$\mathbb{E}\left[X_i^2\right] = \mathbb{E}[X_i]^2 + \$\mathsf{Var}\,(X_i)\,\$ = 0^2 + \theta = \theta,$$

and

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i^2\right] = n\theta.$$

- The Bayesian estimator is consistent because

  因为有prior，所以不是无偏的

  $$\mathbb{E}\left[\hat{\theta}\right] = \frac{1}{n-2}\mathbb{E}\left[\sum_{i=1}^{n} X_i^2\right] = \frac{n}{n-2}\theta \to \theta$$

  as $n \to \infty$, as

  $$\boxed{\lim_{n\to\infty} \frac{n}{n-2} = 1.}$$

  Consistency is a frequentist property of an estimator as it is only desired when we assume a true value for the parameter.

- The Bayesian estiamtor is not unbiased, because as we calculated earlier,

  $$\mathbb{E}\left[\hat{\theta}\right] = \frac{n}{n-2}\theta \neq \theta.$$

- The expected value of the Bayesian estimator is $\frac{n}{n-2}\theta$, which is decreasing in n, so it is true that this is larger when there are only a few observations. Indeed, this is due to the prior used, $\frac{1}{\sqrt{2\theta}}$, which gives larger weight to smaller values of θ. This is a Bayesian property because the notion of having our observations matter more strongly when we have more observations is strongly related to the Bayesian concept of starting from a prior distribution then updating through our observations.

- This Bayesian set-up uses the Jeffreys prior, which depends completely on the conditional likelihood and is thus a non-informative prior. Despite being a prior, this reflects a frequentist property of the procedure (and thus the estimator), because we do not assume any distribution beyond what's contained in the model and hence in some sense have no "prior". In fact, Jeffreys prior bridges the ideological gap between frequentist and Bayesian statistics.

- It is true that the Bayesian estimator is asymptotically normal because it converges in distribution to $\mathbb{E}\left[X_i^2\right]$ which is an expectation of an average, which is known to be asymptotically normal by the Central Limit Theorem. Asymptotic normality again is a property that's desired in the frequentist approach, because we want the parameter to have a predictable frequency distribution over the true parameter

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Discussion

Show Discussion

Topic: Unit 5 Bayesian statistics:Homework 9: Bayesian Statistics / 4. Bayesian Estimation