Problem 2. Multiple Choice Questions: Linear Regression

(a)

2/2 points (graded)
Consider a Gaussian linear model $Y = aX + \epsilon$ in a Bayesian view. Consider the prior $\pi(a) = 1$ for all $a \in \mathbb{R}$. Determine whether each of the following statements is true or false.

$\pi(a)$ is a *uniform prior*.

- (selected) True ✔
- False ✔

Grading Note: This problem meant to ask whether $\pi(a)$ is **improper** instead of uniform. Since it is "uniform" but technically not a uniform prior because of unbounded support, credit is given to both answers.

$\pi(a)$ is a Jeffreys prior when we consider the likelihood $L(Y = y|A = a, X = x)$(where we assume x is known).

- (selected) True ✔
- False

Solution:

- As $\mathbb{R}$ is an unbounded set and $\pi(a) = 1$ is uniform over the set of possible parameters, the prior is improper.
- This is also a Jeffreys prior as

$$\log(\mathbb{P}[Y = y|A = a, X = x]) = \frac{-(y-ax)^2}{2} - \log(\sqrt{2\pi})$$

. Taking the second derivative in a it follows that Jeffreys prior is a uniform distribution, hence $\pi(a) = 1$ is the Jeffreys prior.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

(b)

2/3 points (graded)
Consider a linear regression model $\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\varepsilon}$ where

- $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is a random vector with $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$, $\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = I_n$, and no further assumptions are made about $\boldsymbol{\varepsilon}$
- $\mathbb{X}$ is an n by p deterministic matrix, and $\mathbb{X}^T\mathbb{X}$ is invertible.
- $\sigma > 0$ is an unknown constant.

Let $\hat{\boldsymbol{\beta}}$ denote the least squares estimator of β in this context. Determine whether each of the the following statements is true or false.

1. $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator for $\boldsymbol{\beta}$.

- (selected) True ✖
- False ✔

2. With the model written as $\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \sigma\boldsymbol{\varepsilon}$, $\hat{\boldsymbol{\beta}}$ has dimension $1 \times p$ (i.e. is a row vector of length p).

- True
- (selected) False ✔

3. $\hat{\beta}$ has a Gaussian distribution (even for small n).

- True
- (selected) False ✔

Solution:

1. The least squares estimator of $\hat{\beta}$ is only guaranteed to maximum likelihood estimator if ϵ is a Gaussian.
2. To answer the second question, X and β has to have matching dimensions to be multiplied, so β must be a column vector, not a row vector.
3. The least squares estimator of $\hat{\beta}$ is only guaranteed to have a Gaussian distribution if ϵ is a Gaussian. (Note however that $\hat{\beta}$ is an asymptotically normal estimator.)

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

(c)

1/1 point (graded)
Under the setup and assumptions as in part (b), $\mathbb{X}\hat{\boldsymbol{\beta}}$ is...
(Check all that apply.)

- [ ] Equal to $(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{Y}$
- [x] An unbiased estimator of $\mathbb{X}\boldsymbol{\beta}$ ✔
- [ ] A vector in $\mathbb{R}^p$

✔

Grading Note: Partial credit is given.

Solution:

- Note that $\hat{\boldsymbol{\beta}}$ is $(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{Y}$ and therefore the initial choice is false.
- The fourth choice is incorrect as $Xb\hat{e}ta$ is a vector in $\mathbb{R}^n$.

- Finally the third choice is correct as

$$\mathbb{E}\left[X\hat{\beta}\right] = \mathbb{E}\left[\mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{Y}\right] = \mathbb{E}\left[\mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\left(\mathbb{X}\boldsymbol{\beta}+\epsilon\right)\right] = \mathbb{X}\boldsymbol{\beta}.$$

Note that $\mathbb{X}\hat{\boldsymbol{\beta}}$ is the projection of Y onto the column space of X.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

(optional) (d)

0 points possible (ungraded)
Note: This ungraded problem is included here as an exercise for those with more linear algebra background.

The estimator $\hat{\sigma}^2 = \left\|Y - X\hat{\beta}\right\|_2^2$ for σ^2. (Check all that apply.)

- [ ] Satisfies $\mathbb{E}\left[\hat{\sigma}^2\right] = (n-p)\,\sigma^2$ ✔
- [x] Is unbiased
- [x] Satisfies $(n-p)\,\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-p}$
- [ ] None of the above

✖

Grading note: Partial credit is given.

Solution:

Note that if the initial choice is correct this immediately implies that the remaining choices are false. To prove this choice note that

$$\mathbb{E}\left[\left\|Y - X\hat{\beta}\right\|_2^2\right] = \mathbb{E}\left[\left\|Y - X(X^TX)^{-1}X^TY\right\|_2^2\right]$$

$$= \mathbb{E}\left[\left\|X\beta+\epsilon - X(X^TX)^{-1}X^TX\beta - X(X^TX)^{-1}X^T\epsilon\right\|_2^2\right] = \mathbb{E}\left[\left\|\left(I_n - X(X^TX)^{-1}X^T\right)\epsilon\right\|_2^2\right]$$

$$= tr\left[\mathbb{E}\left[\epsilon^T\left(I_n - X(X^TX)^{-1}X^T\right)^T\left(I_n - X(X^TX)^{-1}X^T\right)\epsilon\right]\right].$$

Since trace is cyclically invariant then

$$tr\left[\mathbb{E}\left[\epsilon^T\left(I_n - X(X^TX)^{-1}X^T\right)^T\left(I_n - X(X^TX)^{-1}X^T\right)\epsilon\right]\right]$$

$$= tr\left[\mathbb{E}\left[\left(I_n - X(X^TX)^{-1}X^T\right)^T\left(I_n - X(X^TX)^{-1}X^T\right)\epsilon\epsilon^T\right]\right] = tr\left[\left(I_n - X(X^TX)^{-1}X^T\right)^T\left(I_n - X(X^TX)^{-1}X^T\right)\sigma^2\right]$$

$$= tr\left[\left(I_n - X(X^TX)^{-1}X^T\right)\sigma^2\right] = \sigma^2\left(tr\left[I_n\right] - tr\left[X(X^TX)^{-1}X^T\right]\right) = \sigma^2\left(n - tr\left[(X^TX)^{-1}X^TX\right]\right) = \sigma^2\left(n-p\right)$$

and the result finally follows.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

(e)

2.0/2.0 points (graded)

Let $Y = a(X-b)^3 + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \theta^2)$ is independent of X. What is the regression function $f(x)$ of Y given X? (Check all that apply)

- [x] $f(x) = a(x-b)^3$ ✔
- [x] $f(x) = \mathbb{E}[Y|X=x]$ ✔
- [ ] $f(x) = 3a(x-b)^2$
- [ ] $f(x) = a + bx$

✔

Note: The problem statement should have stated "What is the regression function $f(x)$ of Y given $X = x$?" but this does not affect grading. Partial credit is given.

Solution:

A regression function of $f(x)$ of Y given $X = x$ is defined as $f(x) = \mathbb{E}[Y|X=x]$. In the model given it immediately follows that $f(x) = a(x-b)^3$ and therefore the first and second choices are correct.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Error and Bug Reports/Technical Issues

Show Discussion

Topic: Final exam:Final Exam / Problem 2. Multiple Choice Questions: Linear Regression