EBA35002 Fall 2022 Mock

Written exam

Jonas Moss

All subexercises are equally weighted.

1 Mathematical questions

In this exercise X_i are n iid observations from a normal distribution with unknown mean μ and unknown standard deviation σ.

1.a.

What is the asymptotic distribution of $\sqrt{n}(\overline{X}-\mu)$, where $\overline{x}$ denotes the mean of $x_1, x_2, \dots, x_n$?

> **Solution**
>
> By the central limit theorem, the asymptotic distribution of $\sqrt{n}(\overline{X}-\mu)$ is normal with standard deviation σ.

1.b.

Show that the maximum likelihood estimator of μ is $\overline{x}$.

> **Solution**
>
> Many resources about this online, e.g., this one.

1.c.

What is the Fisher information of μ, i.e., $I(\mu)$? (*Hint:* There are two ways to calculate this; one is considerably faster than the other.)

Solution

You can either calculate this directly, as shown in the video in the solution to the previous exercise, or you can use 1.a together with 1.b and the fact that asymptotic distributions are unique. Hence the Fisher information is $1/\sigma^2$.

1.d.

Show that the maximum likelihood estimator of σ^2 is $\widehat{\sigma^2} = \sum_{i=1}^n (X_i - \overline{X})^2/n$. Is this an unbiased estimator of σ^2?

Solution

(a) The likelihood is

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

and the loglikelihood l is

$$\begin{aligned}
\log\left[\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right] &= \sum_{i=1}^n \log\left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right] \\
&= -\left(\sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2}\right) - \frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2
\end{aligned}$$

Take the derivative with respect to σ^2, where we use that the derivative of x^{-1} is $-1/(x^{-2})$

$$\frac{\partial l}{\partial \sigma^2} = \frac{1}{2}\frac{1}{\sigma^4}\left(\sum_{i=1}^n (x_i-\mu)^2\right) - \frac{n}{2}\frac{1}{\sigma^2}.$$

Set all of this equal to 0 and obtain

$$\frac{1}{2}\frac{1}{\sigma^4}\left(\sum_{i=1}^n (x_i-\mu)^2\right) = \frac{n}{2}\frac{1}{\sigma^2}.$$

This is equivalent to

$$\left(\sum_{i=1}^n (x_i-\mu)^2\right)/n = \sigma^2.$$

Since we already know that $\overline{x}$ maximizes μ, independently of what σ^2 is, we have that

$$\left(\sum_{i=1}^n (x_i-\overline{x})^2\right)/n = \sigma^2$$

is the maximumn likelihood estimator of σ^2.

(b) No, the estimator is not unbiased, as $\sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)$ is the unbiased estimator of σ^2 (this is known from the previous course and this one.)

1.e.

Show that the Fisher information of σ^2 is $I(\sigma^2) = 1/(2\sigma^4)$.

Solution

The first derivative of the likelihood is

$$\frac{\partial l}{\partial \sigma^2} = \frac{1}{2}\frac{1}{\sigma^4}\left(\sum_{i=1}^{n}(x_i-\mu)^2\right) - \frac{n}{2}\frac{1}{\sigma^2}.$$

Put $\theta = \sigma^2$, so that he derivative of the likelihood with $n = 1$ is

$$\frac{\partial l}{\partial \theta} = \frac{1}{2}\frac{1}{\theta^2}\left(\sum_{i=1}^{n}(x_i-\mu)^2\right) - \frac{1}{2}\frac{1}{\theta}.$$

The second derivative, when $n = 1$, is

$$\frac{\partial^2 l}{\partial \theta^2} = -\frac{1}{2}\frac{2}{\theta^3}(x_i-\mu)^2 + \frac{1}{2}\frac{1}{\theta^2},$$

where we use that $(1/x^2)' = -2/x^3$ and $(1/x)' = -1/x^2$. The expectation of $(x_i - \mu)^2$ is $\sigma^2 = \theta$, hence

$$\begin{aligned} E\left[\frac{\partial^2 l}{\partial^2 \sigma^2}\right] &= \frac{1}{\theta^2} - \frac{1}{2}\frac{1}{\theta^2}. \\ &= \frac{1}{2}\frac{1}{\sigma^4}. \end{aligned}$$

1.f

Let θ be a k-dimensional vector parameter and $g : \mathbb{R}^k \to \mathbb{R}$ be a continuously differentiable function. Moreover, suppose $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma)$. What is the asymptotic distribution of $\sqrt{n}(g(\hat{\theta}) - g(\theta))$?

Solution

The *delta method (rule)* states that $\sqrt{n}(g(\hat{\theta}) - g(\theta))$ converges to a normal with variance $\nabla g(\theta)^T \Sigma \nabla g(\theta)$, where $\nabla g(\theta)$ is the gradient of g at θ.

1.g

Let $g(x, y) = x/\sqrt{y}$. Find the partial derivatives of g, i.e.,

$$\frac{\partial g}{\partial x}, \quad \frac{\partial g}{\partial y}.$$

Solution

Using that $\frac{d}{dx} x^p = px^{p-1}$, we find that

$$\frac{\partial g}{\partial x} = \frac{1}{\sqrt{y}}, \quad \frac{\partial g}{\partial y} = -\frac{x}{2y^{3/2}}$$

1.h

We want to do inference on μ/σ, sometimes called the *effect size*. What is the asymptotic distribution of $\overline{x}/\sqrt{\widehat{\sigma^2}}$? You may use that the Fisher information of (μ, σ) is a diagonal matrix with zero off-diagonal entries, i.e.,

$$\begin{bmatrix} a & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}$$

where a is the Fisher information you found in exercise 1.c.

Solution

We use the delta method along with $a = 1/\sigma^2$, found in exercise 1.c. By the delta method and the fact that the covariance matrix of $\theta = (\mu, \sigma^2)$ equals the inverse of the Fisher information matrix, we find that the variance equals:

$$\begin{bmatrix} 1/\sigma \\ -\mu/(2\sigma^3) \end{bmatrix}^T \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \begin{bmatrix} 1/\sigma \\ -\mu/(2\sigma^3) \end{bmatrix} = $$
$$1 + \frac{1}{2}\frac{\mu^2}{\sigma^2}$$

1.i

Use the information in the previous exercise to construct an approximate 95% confidence interval for μ/σ. If you weren't able to solve the previous exercise, explain how you would do it.

Solution

Suppose that $\sqrt{n}(\hat{\theta}-\theta) \to N(0,\tau)$. Using the Z-interval construction (see exercises for 9), a $(1-\alpha)$% confidence interval for θ is

$$CI(\theta; 1-\alpha) = [\hat{\theta} - \Phi^{-1}(1-\alpha/2)\sqrt{\hat{\tau}}/\sqrt{n}, \hat{\theta} + \Phi^{-1}(1-\alpha/2)\sqrt{\hat{\tau}}/\sqrt{n}],$$

where Φ^{-1} is the quantile function of the normal distribution, called ppf in Scipy, and $\hat{\tau}$ is an estimator of τ. Recall that

$$\Phi^{-1}(1-\alpha/2) \approx 1.96$$

when $\alpha = 0.95$
In our case $\tau = 1 + \frac{1}{2}\frac{\mu^2}{\sigma^2}$, and its estimator is $\hat{\tau} = 1 + \frac{1}{2}\frac{\hat{\mu}^2}{\widehat{\sigma^2}}$ hence the confidence interval is

$$\hat{\mu}/\widehat{\sigma^2} \pm 1.96/\sqrt{n}\left(1 + \frac{1}{2}\frac{\hat{\mu}^2}{\widehat{\sigma^2}}\right)$$

1.j

Using the results in the previous exercise, construct a confidence interval for μ/σ when $\overline{X} = 1$ and $\overline{X^2} = 2$.

Solution

Using the fact that the unbiased sample variance equals $\widehat{\sigma^2} = \overline{x^2} - \overline{x}^2 = 1$ and that $\hat{\mu} = \overline{x} = 1$, we find that

$$\hat{\mu}/\widehat{\sigma^2} \pm 1.96/\sqrt{n}\left(1 + \frac{1}{2}\frac{\hat{\mu}^2}{\widehat{\sigma^2}}\right)$$

equals

$$1 \pm \frac{3}{2} \cdot 1.96/\sqrt{n}$$

2 Regression questions

2.a

Suppose we have a regression model with a continuous response and one continuous covariate. What is the relationship between the R^2 and the correlation coefficient?

> **Solution**
>
> The absolute value of correlation coefficient is the square root of the R^2 in this case. The sign of the correlation equals the sign of the slope of the regression $y = a + bx$.

2.b

```python
import statsmodels.api as sm
import statsmodels.formula.api as smf
import numpy as np
cpssw8 = sm.datasets.get_rdataset("CPSSW8", "AER").data
model = smf.ols("np.log(earnings) ~ age - 1", data = cpssw8).fit()
```

Below we calculate the R^2:

```python
y = np.log(cpssw8.earnings)
y_mean = y.mean()
1 - np.mean((y - model.predict())**2) / np.mean((y - y_mean)**2)
```

With output `-1.084`! How is this possible? How can you guarantee this never happens? Change the formula to make sure that the R^2 is non-negative.

> **Solution**
>
> The R^2 reported above is an estimator of
>
> $$1 - \frac{\min_\beta E[(Y - \beta^T X)^2]}{\min_\mu E[(Y - \mu)^2]}$$
>
> But this model does not contain an intercept α, hence there is no guarantee that
>
> $$\min_\beta E[(Y - \beta^T X)^2] \leq \min_\mu E[(Y - \mu)^2].$$
>
> To make sure that this negative R^2 value never occurs, make sure to fit a model with an intercept. To do this, use `"np.log(earnings) ~ age"` instead.

2.c

Suppose you have 3 categorical covariates `a,b,c` with 3, 7 and 13 levels each. How many regression coefficients are there in the model `y~a*b*c`?

> **Solution**
>
> The number of covariates is $3 \cdot 7 \cdot 13$.

2.e

Mention three reasonable distance functions in linear regression. Which on is most popular? Name three reasons why it is the most popular.

> **Solution**
>
> Three reasonable distances are the quadratic, the absolute value distance, and the Huber distance. The quadaratic is the most popular since it is the easiest to compute, the easiest to work with mathematically, and since its resulting estimator is efficient when the errors are normal.

2.f

Alice is in big trouble! Her boss wants her to do a linear regression `satisfaction ~ age + gender`, where `satisfaction` is a number in $\{-3, -2, -1, 0, 1, 2, 3\}$ encoding customer satisfaction. But Alice has somehow managed to throw away her `satisfaction` data, replacing it with the variable `satisfied = 1 * (satisfaction >= 0)` instead. What should Alice do to fulfill her boss's wish, and why would it work?

> **Solution**
>
> Following the latent variable interpretation of binary regression models, she should use a regression `statisfied ~ age + gender` instead. When the errors are normal (they aren't in this case, since `satisfaction` is discrete), the Probit model is unambiguously the correct choice.

2.g

Alice shows Bob some limited output for four models she fitted:

Model Number	R squared	Adjusted R squared
1	0.017	0.007
2	0.018	-0.002
3	0.047	-0.026
4	0.161	0.035

Which of these models would you prefer, and why?

Solution

I prefer the final model, since it has the highest adjusted R squared.