1. Hypothesis Testing

Setup:

Suppose we have n observations $(\mathbf{X}_i, Y_i)$, where $Y_i \in \mathbb{R}$ is the dependent variable, $\mathbf{X}_i \in \mathbb{R}^p$ is the **column** $p \times 1$ vector of deterministic explanatory variables, and the relation between Y_i and $\mathbf{X}_i$ is given by

$$Y_i = \mathbf{X}_i^T \beta + \epsilon_i, \qquad i = 1, \ldots, n.$$

where ϵ_i are i.i.d. $\mathcal{N}(0, \sigma^2)$. As usual, let $\mathbb{X}$ denote the $n \times p$ matrix whose rows are $\mathbf{X}_i^T$.

Unless otherwise stated, assume $\mathbb{X}^T \mathbb{X} = \tau \mathbf{I}$ and that τ, σ^2 are known constants.

(a)

2/2 points (graded)

Recall that under reasonable assumptions (which is certainly satisfied in linear regression with Gaussian noise), the Fisher Information of a parameter θ given a family of distributions $\mathbf{P}_\theta$ can be computed via the following formula:

$$I(\theta) = -\sum_{i=1}^{n} H_\theta \ln f(Y_i; \theta)$$

Generating Speech Output

where H_θ denotes the Hessian differentiation operator with respect to θ. (Recall the definition in lecture 9).

In terms of $\mathbb{X}$, σ^2, compute the Fisher $I(\beta)$ information of β.

(Type **X** for $\mathbb{X}$, **trans(X)** for the transpose $\mathbf{X}^T$ of a matrix $\mathbb{X}$, and **X^(-1)** for the inverse $\mathbb{X}^{-1}$ of a matrix $\mathbb{X}$.)

$I(\beta) =$ (trans(X)*X)/sigma^2 ✔

Plugging in $\mathbb{X}^T\mathbb{X} = \tau\mathbf{I}$, then the Fisher Information simplifies to a scalar multiple of $\mathbf{I}$, so that it is a matrix of the form $\lambda\mathbf{I}$. Find the multiplicative constant λ, in terms of τ and σ.

$\lambda =$ tau/sigma^2 ✔

$\frac{\tau}{\sigma^2}$

STANDARD NOTATION

Submit You have used 1 of 3 attempts

✔ Correct (2/2 points)

(b)

3/3 points (graded)

Instructions: Fill in the blank in terms of σ and τ.

Based on the calculation of the Fisher Information (or by other means), we can conclude that the Maximum Likelihood Estimator $\hat{\beta}$ has entries $\hat{\beta}_1, \ldots, \hat{\beta}_p$ that are independent Gaussians, with variance:

Var $(\hat{\beta}_i) =$ sigma^2/tau ✔

Generating Speech Output

$\frac{\sigma^2}{\tau}$

Suppose we wish to test the hypotheses

$$H_0 : \beta_1 = \beta_2, \qquad H_1 : \beta_1 \neq \beta_2.$$

Based on the observation made above, a suitable test statistic is $T_n = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{\mathsf{Var}(\hat{\beta}_1 - \hat{\beta}_2)}}$.

Find the denominator $\sqrt{\mathsf{Var}\left(\hat{\beta}_1 - \hat{\beta}_2\right)}$ (including the square root) in terms of σ and τ.

$\sqrt{\mathsf{Var}\left(\hat{\beta}_1 - \hat{\beta}_2\right)} =$ sigma*sqrt(2/tau) ✔

What is the appropriate test at significance level $\alpha = 0.01$?

(Let q_α denote the standard normal α-quantile for each respective choice below.)

- ○ $\psi = \mathbf{1}\left(T_n > q_{0.01}\right)$
- ○ $\psi = \mathbf{1}\left(T_n > q_{0.005}\right)$
- ○ $\psi = \mathbf{1}\left(|T_n| > q_{0.01}\right)$
- ● $\psi = \mathbf{1}\left(|T_n| > q_{0.005}\right)$

✔

STANDARD NOTATION

Submit You have used 1 of 3 attempts

Generating Speech Output

✔ Correct (3/3 points)

(c)

2/2 points (graded)

Suppose we instead wish to test the hypotheses $H_0 : (\beta_1, \beta_2, \beta_3) = (0,0,0)$, $H_1 : (\beta_1, \beta_2, \beta_3) \neq (0,0,0)$.

Let γ be some appropriate value corresponding to the significance level, to be determined later. Choose all ψ that correctly represents the Bonferroni Test of H_0 against H_1.

- [ ] $\psi = \mathbf{1}\left\{\frac{|\hat\beta_1+\hat\beta_2+\hat\beta_3|}{3} > q_\gamma\right\}$
- [x] $\psi = \mathbf{1}\left\{\frac{\max(|\hat\beta_1|,|\hat\beta_2|,|\hat\beta_3|)}{\sqrt{\sigma^2/\tau}} > q_\gamma\right\}$
- [ ] $\psi = \prod_{i=1}^{3} \mathbf{1}\left\{\frac{|\beta_i|}{\sqrt{\sigma^2/\tau}} > q_\gamma\right\}$
- [ ] $\psi = \mathbf{1}\left\{|\hat\beta_1 - \hat\beta_2 - \hat\beta_3| > q_\gamma\right\}$
- [ ] $\psi = \mathbf{1}\left\{\left(|\hat\beta_1/\sqrt{\sigma^2/\tau}| > q_\gamma\right) \text{ and } \left(|\hat\beta_2/\sqrt{\sigma^2/\tau}| > q_\gamma\right) \text{ and } \left(|\hat\beta_3/\sqrt{\sigma^2/\tau}| > q_\gamma\right)\right\}$
- [x] $\psi = \mathbf{1}\left\{\left(|\hat\beta_1/\sqrt{\sigma^2/\tau}| > q_\gamma\right) \text{ or } \left(|\hat\beta_2/\sqrt{\sigma^2/\tau}| > q_\gamma\right) \text{ or } \left(|\hat\beta_3/\sqrt{\sigma^2/\tau}| > q_\gamma\right)\right\}$

✔

In the Bonferroni test of significance level $\alpha = 0.01$ for testing this particular H_0 against H_1, what is the numerical value of γ? Input a fraction or round to the nearest 10^{-5}, if necessary.

Generating Speech Output

0.001666667

Submit You have used 1 of 3 attempts

Correct (2/2 points)

(d)

5/5 points (graded)

Instructions: For an arbitrary significance level $\alpha \in (0,1)$, compute an order $(1-\alpha)$ confidence interval for $\beta_1 = 0$ by filling in the blanks. (In other words, find a confidence interval with confidence level $1-\alpha$.) Unless otherwise specified, express your answers in terms of σ^2, τ, α and the quantile q.

(Type $q(\alpha)$ to denote q_α, the $1-\alpha$-th quantile of the standard Gaussian.)

- The random variable $\hat{\beta}_1 - \beta_1$ is a Gaussian RV, with a variance that we computed earlier. Find the value of C such that $\mathbf{P}\left(-C \leq \hat{\beta}_1 - \beta_1 \leq C\right) = 1-\alpha$.

  $C =$ q(alpha/2)*sigma/sqrt(tau)

  This gives us the confidence interval $I = [\hat{\beta}_1 - C, \hat{\beta}_1 + C]$.

- Revisit part (b). If $\mathbb{X}^T\mathbb{X}$ were not diagonal, then in terms of σ and $\mathbb{X}$, the covariance matrix of $\hat{\beta}$ is

  $\Sigma_{\hat{\beta}} =$ sigma^2*(trans(X)*X)^(-1)

- The variance of $\hat{\beta}_1$ can be expressed in terms of a particular (i,j) entry of this matrix (the answer to the previous part), where the row-column ordered pair (i,j) is:

  $i =$ 1 , $j =$ 1

Generating Speech Output

Let $\delta^2 = \mathrm{Var}\,(\hat{\beta}_1)$ be this matrix entry. The new value of C becomes (in terms of δ, and q):

New value of C : q(alpha/2)*delta

STANDARD NOTATION

Submit You have used 1 of 4 attempts

Correct (5/5 points)

Discussion

Hide Discussion

Topic: Unit 6 Linear Regression:Homework 10 Linear regression / 1. Hypothesis Testing

Add a Post

All Posts

[STAFF] Is it just me or is the jump from lectures to homework this week more substantial than ever?

question posted 3 days ago by **DriftingWoods**

Prof did a nice tidy pulling out of X^T in front for LSE but did I miss / am I failing to recall a bunch of other matrix calculus that happened? Sums of Hessians of log of functions(not likelihood?) for a Fisher information? Fisher equals negative of the expectation of the Hessian of the log likelihood doesn't it? I'm completely at a loss even for the first question. This seems to be the case for a lot of the homework. 3 chances for 4 multiple choice options mean scores won't be much of an issue but I want to understand the material.

This post is visible to everyone.

Add a Response

3 responses

Iang Park

Generating Speech Output ago

Doing the matrix calculus for this problem would be substantial and prone to calculation errors. I would suggest to look for an easier way relating the Fisher Information Matrix to a quantity that we already derived and know. However, taking this approach, it is essential to be able to justify why this can be done.

Thank you - I have some better intuition as to what is going on now.

posted 2 days ago by **DriftingWoods**

Add a comment

Erocha (Community TA)
2 days ago

Note that the Fisher information is just the regular one we saw in class. The difference, I believe, is that you we have a deterministic design, so we have to consider all observations. It is like taking the log-likelihood for all observations (what is the distribution of which one of them, can you tell?). Moreover, note that the Hessian is linear. In addition, you may find useful to consider the sum of rank one matrices. Prof. Gilbert Strang actually has a lecture on that on OCW.

I had the same puzzles as you @DriftingWoods.

I thought it must be that here we had to use all the observations, whereas before when calculating Fisher Information we just considered one observation.

I do not understand, however, why the deterministic design forces us to use all the observations.

posted a day ago by **mbh038**

Note the summation in the definition above. It is including all observations. As you said, previously, the Fisher information had been defined for one observation. So my intuition for understanding the expression above is that before you treated X as fixed. Here, $\mathbb{X}$ is what is fixed, and you can only define $\hat{\beta}$ with this full matrix.

posted a day ago by **Erocha** (Community TA)

Generating Speech Output

Makes sense, thank you.

posted a day ago by **mbh038**

Add a comment

sandipan_dey
about an hour ago

I thought it's a standard practice to have X as a $n \times p$ matrix (n rows with p features, with each X_i as **row** vectors, NOT **column** vectors) and the the covariance matrix $X^T X$ to be as a $p \times p$ matrix (as in **PCA** we do the eigen decomposition of $X^T.X$) , as opposed to the outer-product matrix $X.X^T$. Is the convention (w.r.t, inner/outer products and the definition of the covariance matrix) the other way (it seems so, since we are always using $X.X^T$ as the covariance matrix in this course, with each X_i as column vectors)?

Add a comment

Preview

Submit

Showing all responses

Add a response:

Generating Speech Output

Preview

Submit

Generating Speech Output