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)

2.0/2 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:

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

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 ✔ **Answer:** 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 ✔ **Answer:** tau/sigma^2

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

STANDARD NOTATION

Solution:

Notice that Y is a Gaussian random variable, so plug the Gaussian pdf directly into the suggested formula. Let $\ell(y|\beta)$ denote the log-likelihood function for a single observation x:

$$\begin{aligned}\ell(y|\beta) &= \ln \frac{1}{2\pi\sigma^2} - \frac{1}{2\sigma^2}(y - \mathbf{x}^T\beta)^2 \\ &= \ln \frac{1}{2\pi\sigma^2} - \frac{1}{2\sigma^2}(y^2 - 2y\mathbf{x}^T\beta + \beta^T\mathbf{x}\mathbf{x}^T\beta)\end{aligned}$$

$$\Longrightarrow \nabla_\beta \ell(y|\beta) = -\frac{1}{2\sigma^2}(-2y\mathbf{x} + 2\mathbf{x}\mathbf{x}^T\beta)$$

$$= \frac{1}{\sigma^2}(y\mathbf{x} - \mathbf{x}\mathbf{x}^T\beta)$$

$$\implies H_\beta \ell(y|\beta) = \frac{-\mathbf{x}\mathbf{x}^T}{\sigma^2}$$

Therefore, $I(\beta)$ is the sum

$$\begin{aligned} I(\beta) &= -\sum_{i=1}^{n} \mathbb{E}_Y\left[-\frac{1}{\sigma^2}\mathbf{X}_i\mathbf{X}_i^T\right] \\ &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T \\ &= \frac{1}{\sigma^2}\mathbb{X}^T\mathbb{X} \end{aligned}$$

The outer product $\mathbf{X}_i\mathbf{X}_i^T$ is a matrix, and summing over them gives the matrix $\mathbb{X}^T\mathbb{X}$. The transposes are switched due to the convention that the rows of $\mathbb{X}$ are $\mathbf{X}_i^T$, since $\mathbf{X}_i$ are column vectors.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

(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:

$\mathsf{Var}(\hat{\beta}_i) =$ sigma^2/tau ✔ **Answer:** sigma^2/tau

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

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

$\sqrt{\mathsf{Var}(\hat{\beta}_1 - \hat{\beta}_2)} =$ sqrt(2*sigma^2/tau) ✔ **Answer:** sqrt(2*sigma^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.)

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

$\psi = \mathbf{1}\left(|T_n| > q_{0.005}\right)$ ✔

STANDARD NOTATION

Solution:

Recall that the Fisher Information matrix is the inverse of the Covariance of the Maximum likelihood estimator, $\hat{\beta}$. This tells us that the $\hat{\beta}$ has covariance $\frac{\sigma^2}{\tau} I$, which also means that the coordinates of $\hat{\beta}$ are i.i.d.

To design the test, observe that the statement $\beta_1 = \beta_2$ should be re-written as $\beta_1 - \beta_2 = 0$. The difference of two i.i.d. Gaussians is a Gaussian, with twice the variance. Thus, the correct answer for the denominator in the test statistic T_n is $\sqrt{2\sigma^2/\tau}$.

Finally, the test T_n suggests that we are trying to determine whether $\beta_1 - \beta_2 = 0$, by calculating/re-scaling $\hat{\beta}_1 - \hat{\beta}_2$. Intuitively, the null hypothesis should be rejected if $\hat{\beta}_1 - \hat{\beta}_2$ happens to be large (far away from zero). Therefore, we ought to apply the two-sided test $\psi = \mathbf{1}\left(|T_n| > q_{0.005}\right)$.

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

(c)

1/2 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] $\psi = \mathbf{1}\left\{\frac{\max(|\hat{\beta}_1|,|\hat{\beta}_2|,|\hat{\beta}_3|)}{\sqrt{\sigma^2/\tau}} > q_\gamma\right\}$ ✔
- [ ] $\psi = \prod_{i=1}^3 \mathbf{1}\left\{\frac{|\beta_i|}{\sqrt{\sigma^2/\tau}} > q_\gamma\right\}$
- [ ] $\psi = \mathbf{1}\left\{|\hat{\beta}_1 - \hat{\beta}_2 - \hat{\beta}_3| > q_\gamma\right\}$
- [x] $\psi = \mathbf{1}\left\{\left(|\hat{\beta}_1/\sqrt{\sigma^2/\tau}| > q_\gamma\right) \text{ and } \left(|\hat{\beta}_2/\sqrt{\sigma^2/\tau}| > q_\gamma\right) \text{ and } \left(|\hat{\beta}_3/\sqrt{\sigma^2/\tau}| > q_\gamma\right)\right\}$
- [ ] $\psi = \mathbf{1}\left\{\left(|\hat{\beta}_1/\sqrt{\sigma^2/\tau}| > q_\gamma\right) \text{ or } \left(|\hat{\beta}_2/\sqrt{\sigma^2/\tau}| > q_\gamma\right) \text{ or } \left(|\hat{\beta}_3/\sqrt{\sigma^2/\tau}| > q_\gamma\right)\right\}$ ✔

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.

0.01/6 ✔ **Answer:** 0.01/6

Solution:

The Bonferroni test is used in this setting, because we are simultaneously testing three hypotheses: $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$. If even one of these is false, then we would hope that H_0 is rejected. The correct choice here is $\psi = \mathbf{1}\{\frac{\max(|\hat{\beta}_1|,|\hat{\beta}_2|,|\hat{\beta}_3|)}{\sqrt{\sigma^2/\tau}} > q_\gamma\}$, which is equal to the logical or (∨) of the three conditions $\beta_i/\sqrt{\sigma^2/\tau} > q_\gamma$. In contrast, the product formula is the same as the logical and (and), which rejects only if all three mini-tests reject simultaneously. That is not what we want from our test.

When testing at significance level $\alpha = 0.01$, notice that all three tests on $\beta_1, \beta_2, \beta_3$ must be performed individually at level $\alpha/3$. Since each of these three tests is a two-sided test (e.g. $|\hat{\beta}_1| > q_\gamma$), the quartile we are looking for is the $\alpha/6$-quartile, or $\gamma = \frac{0.01}{6} \approx 0.00167$.

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

(d)

5.0/5 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 $\mathbf{P}(-C \le \hat{\beta}_1 - \beta_1 \le C) = 1 - \alpha$.

  $C =$ q(alpha/2)*sqrt(sigma^2/tau) ✔ **Answer:** sqrt(sigma^2/tau)*q(alpha/2)

  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 ✔ **Answer:** 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 , ✔ Answer: 1 $j =$ 1 ✔ **Answer:** 1

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

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

STANDARD NOTATION

Solution:

- Since $\hat{\beta}_1$ is gaussian with mean β_1 and variance σ^2/τ, we will take (as suggested by the provided inequality) a quantile corresponding to a two-sided test. The quantile is $q_{\alpha/2}$, scaled by the variance, so that $C = \sqrt{\frac{\sigma^2}{\tau}} q_{\alpha/2}$.
- If $\mathbb{X}^T\mathbb{X}$ were not diagonal, then the covariance matrix of $\hat{\beta}$ is $\sigma^2(\mathbb{X}^T\mathbb{X})^{-1}$.
- By definition, $\mathbf{Var}(\hat{\beta}_1)$ is the first diagonal entry of the covariance matrix, $\sigma^2(\mathbb{X}^T\mathbb{X})^{-1}$. The entry that should be entered is the ordered pair "**(1,1)**".
- Instead of the answer from the first part of this problem, we scale by the new standard deviation δ: $C = \delta q_{\alpha/2}$.

Submit You have used 2 of 4 attempts

Answers are displayed within the problem