MATH 343 / 643 Homework #2

Osman Khan

Wednesday 10^{th} April, 2024

Problem 1

This problem is about OLS estimation in regression. You can assume that

$$\begin{array}{lcl}
\boldsymbol{X} & := & [\mathbf{1}_n \mid \boldsymbol{x}_{\cdot 1} \mid \ldots \mid \boldsymbol{x}_{\cdot p}] \ \text{ with column indices } 0, 1, \ldots, p \ \text{ and row indices } 1, 2, \ldots, n \\
\boldsymbol{H} & := & \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T \\
\boldsymbol{Y} & = & \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mathcal{E}} \\
\boldsymbol{B} & := & \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y} \\
\hat{\boldsymbol{Y}} & = & \boldsymbol{H}\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} \\
\boldsymbol{E} & := & \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y}
\end{array}$$

where the entries of $\boldsymbol{X}$ are assumed fixed and known and the entries of $\boldsymbol{\beta}$ are the unknown parameter).

(a) [easy] When we "do inference" for the linear model, what is the parameter vector?

$\vec{\beta}$

(b) [easy] When we "do inference" for the linear model, what are considered the fixed and known quantities?

X

(c) [easy] When we "do inference" for the linear model, what are considered the random quantities? And what is the notation for their corresponding realizations?

$\mathbf{Y_i's}, \mathbf{\hat{Y_i}'s}$

(d) [easy] What is the "core assumption" in which the classic linear model inference follows?

i) ϵ_i is a realization from normal, ii) mean centered, iii) with homoskedasticity.

(e) [easy] From the core assumption, derive the distribution of $\boldsymbol{B}$.

$\vec{B} \sim N_{p+1}(\vec{\beta}, \sigma^2(X^TX)^{-1})$

(f) [easy] From this result, derive the distribution of B_j.

$B_j \sim N(\beta_j, \sigma^2(X^T X)^{-1}_{j,j})$

(g) [easy] From this result, derive the distribution of B_j standardized.

$\frac{B_j - \beta_j}{\sqrt{(\sigma^2(X^T X)^{-1}_{j,j})}} \sim N(0, 1)$

(h) [easy] from the core assumption, derive the distribution of $\hat{\boldsymbol{Y}}$.

$\vec{\hat{Y}} \sim N_n(X\vec{\beta}, \sigma^2 H)$

(i) [easy] From this result, derive the distribution of $\hat{Y}_i$.

$\hat{Y}_i \sim N(\vec{x_{\cdot i}}\vec{\beta}, \sigma^2 H_{i,i})$

(j) [easy] From this result, derive the distribution of $\hat{Y}_i$ standardized.

$\frac{\hat{Y}_i - \vec{x_{\cdot i}}}{\sqrt{\sigma^2 H_{i,i}}} \sim N(0, 1)$

(k) [easy] from the core assumption, derive the distribution of $\boldsymbol{E}$.

$\vec{E} \sim N_n(\vec{0}, \sigma^2(I - H))$

(l) [easy] From this result, derive the distribution of E_i.

$E_i \sim N(0, \sigma^2(I - H)_{i,i})$

(m) [easy] From this result, derive the distribution of E_i standardized.

$\frac{E_i}{\sqrt{(I-H)_{i,i}}} \sim N(0, 1)$

(n) [easy] From the core assumption, show that $\frac{1}{\sigma^2}\boldsymbol{\mathcal{E}}^\top\boldsymbol{\mathcal{E}} \sim \chi^2_n$.

Let $\vec{Z} \sim N_n(\vec{O_n}, I_n)$, and $\vec{\boldsymbol{\mathcal{E}}} = \sigma\vec{Z}$ which
$= (\sigma I_n)Z \sim N_n(\vec{O_n}, (\sigma I_n)I_n(\sigma I_n)^T) = N_n(\vec{O_n}, \sigma^2 I_n)$ so
$\vec{Z} = \frac{1}{\sigma}\boldsymbol{\mathcal{E}}$ therefore
$(\frac{1}{\sigma}\boldsymbol{\mathcal{E}}^T)(\frac{1}{\sigma}\boldsymbol{\mathcal{E}}) = \vec{Z}^T\vec{Z} \sim \chi^2_n$, and we have proven by law of transitivity.

(o) [easy] Let $\boldsymbol{B}_1 = \boldsymbol{H}$ and let $\boldsymbol{B}_2 = \boldsymbol{I}_n - \boldsymbol{H}$. Justify the use of Cochran's theorem and then find the distributions of $\frac{1}{\sigma^2}\boldsymbol{\mathcal{E}}^\top\boldsymbol{B}_1\boldsymbol{\mathcal{E}}$ and $\frac{1}{\sigma^2}\boldsymbol{\mathcal{E}}^\top\boldsymbol{B}_2\boldsymbol{\mathcal{E}}$.

$(\frac{1}{\sigma}\boldsymbol{\mathcal{E}}^T)(\frac{1}{\sigma}\boldsymbol{\mathcal{E}}) \sim \chi^2_n$,
$\text{rank}[B_2] = n - (p+1)$, $\text{rank}[B_1] = n$,
$B_2 + B_1 = I_n$, so by Cochran's thm:
$\frac{1}{\sigma^2}\boldsymbol{\mathcal{E}}^T B_1\boldsymbol{\mathcal{E}} \sim \chi^2_{p+1}$ and
$\frac{1}{\sigma^2}\boldsymbol{\mathcal{E}}^T B_2\boldsymbol{\mathcal{E}} \sim \chi^2_{n-(p+1)}$

(p) [easy] Show that $\frac{1}{\sigma^2}\boldsymbol{\mathcal{E}}^\top \boldsymbol{B}_1\boldsymbol{\mathcal{E}} = \frac{1}{\sigma^2}\,||\boldsymbol{X}(\boldsymbol{B}-\boldsymbol{\beta})||^2$.

Knowing $H = B_1$ is idempotent and symmetric,
$\frac{1}{\sigma^2}\boldsymbol{\mathcal{E}}^T B_1 \boldsymbol{\mathcal{E}}$
$= \frac{1}{\sigma^2}\boldsymbol{\mathcal{E}}^T B_1 B_1 \boldsymbol{\mathcal{E}}$
$= \frac{1}{\sigma^2}\boldsymbol{\mathcal{E}}^T B_1^T B_1 \boldsymbol{\mathcal{E}}$
$= \frac{1}{\sigma^2}(B_1\boldsymbol{\mathcal{E}})^T B_1 \boldsymbol{\mathcal{E}}$
$= \frac{1}{\sigma^2}||B_1\vec{\boldsymbol{\mathcal{E}}}||^2$
$= \frac{1}{\sigma^2}||B_1(\vec{Y} - X\vec{\beta})||^2$
$= \frac{1}{\sigma^2}||B_1\vec{Y} - B_1X\vec{\beta})||^2$
$= \frac{1}{\sigma^2}||(XB - B_1X\vec{\beta})||^2$

$= \frac{1}{\sigma^2}||(XB - X(X^TX)^{-1}X^TX\vec{\beta})||^2$
$= \frac{1}{\sigma^2}||(XB - X\vec{\beta})||^2$
$= \frac{1}{\sigma^2}||X(B - \vec{\beta})||^2. \blacksquare$

(q) [harder] Why is the term $||\boldsymbol{X}(\boldsymbol{B}-\boldsymbol{\beta})||^2$ used to measure the model's "estimation error"?

Estimation error is incurred when we don't have enough n. B is an estimate for β. Therefore, as $n \to \infty, B \to \beta$, and the term approaches zero, making it the model's estimation error.

(r) [easy] Show that $\frac{1}{\sigma^2}\boldsymbol{\mathcal{E}}^\top \boldsymbol{B}_2\boldsymbol{\mathcal{E}} = \frac{1}{\sigma^2}\,||\boldsymbol{E}||^2$.

Knowing $I - H = B_2$ is idempotent and symmetric, in addition to H already being idempotent
$\frac{1}{\sigma^2}\boldsymbol{\mathcal{E}}^T B_2 \boldsymbol{\mathcal{E}}$
$\frac{1}{\sigma^2}\boldsymbol{\mathcal{E}}^T B_2 B_2 \boldsymbol{\mathcal{E}}$
$\frac{1}{\sigma^2}\boldsymbol{\mathcal{E}}^T B_2^T B_2 \boldsymbol{\mathcal{E}}$
$\frac{1}{\sigma^2}(B_2\boldsymbol{\mathcal{E}})^T B_2 \boldsymbol{\mathcal{E}}$
$\frac{1}{\sigma^2}||B_2\boldsymbol{\mathcal{E}}||^2$
$\frac{1}{\sigma^2}||B_2(\vec{Y} - X\vec{\beta})||^2$
$\frac{1}{\sigma^2}||B_2\vec{Y} - B_2X\vec{\beta}||^2$
$\frac{1}{\sigma^2}||(I-H)Y - (I-H)X\beta||^2$
$\frac{1}{\sigma^2}||Y - HY - X\beta + HX\beta||^2$
$\frac{1}{\sigma^2}||Y - HY - X\beta + H(Y - E)||^2$
$\frac{1}{\sigma^2}||Y - XB - X\beta + H\hat{Y}||^2$
$\frac{1}{\sigma^2}||Y - XB - X\beta + HHY||^2$
$\frac{1}{\sigma^2}||Y - XB - \beta + HY||^2$
$\frac{1}{\sigma^2}||Y - XB - X\beta + XB||^2$
$\frac{1}{\sigma^2}||Y - X\beta||^2$
$\frac{1}{\sigma^2}||E||^2$

(s) [harder] In what scenarios is $\boldsymbol{\mathcal{E}}^\top \boldsymbol{B}_1 \boldsymbol{\mathcal{E}} > \boldsymbol{\mathcal{E}}^\top \boldsymbol{B}_2 \boldsymbol{\mathcal{E}}$?

When our model performs worse than the null model.

(t) [harder] Draw an illustration of $\boldsymbol{\mathcal{E}}$ being orthogonally projected onto colsp $[\boldsymbol{X}]$ via projection matrix $\boldsymbol{H}$. Use the previous answers to denote the quantities of the projection and the error of the projection.

(u) [difficult] A good linear model has a large or a small projection of the error? Discuss.

A good linear model has a small projection of the error, as the closer the vector of the model is to the real vector, the smaller the projection of the error.

(v) [easy] Find $\mathbb{E}\left[\frac{1}{\sigma^2}\left|\left|\boldsymbol{E}\right|\right|^2\right]$.

The expectation of a chi-squared r.v. is its d.f, therefore: $n - p + 1$

(w) [easy] Show that $\frac{||\boldsymbol{E}||^2}{n-(p+1)}$ is an unbiased estimate of σ^2.

$\mathbb{E}[\frac{1}{\sigma^2}||\mathbf{E}||^2] = n - p + 1$, so

$\frac{1}{n-p+1}\mathbb{E}[||\mathbf{E}||^2] = \sigma^2 \rightarrow \mathbb{E}[\frac{1}{n-p+1}||\mathbf{E}||^2] = \sigma^2.\blacksquare$

(x) [easy] Prove that $\dfrac{\sqrt{n-(p+1)}(B_j - \beta_j)}{||\boldsymbol{E}||\sqrt{\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}_{j,j}}} \sim T_{n-(p+1)}$.

We know $\frac{B_j - \beta_j}{\sqrt{(\sigma^2(X^TX)^{-1}_{j,j})}} \sim N(0,1)$, and
$\frac{1}{\sigma^2}||\mathbf{E}||^2 \sim \chi^2_{n-(p+1)}$, therefore using 340 knowledge, we can say that:
$\dfrac{\frac{B_j-\beta_j}{\sqrt{(\frac{1}{\sigma^2}(X^TX)^{-1}_{j,j})}}}{\sqrt{\frac{\frac{1}{\sigma^2}||\mathbf{E}||^2}{n-(p+1)}}} = \frac{\sqrt{(n-(p+1))}(B_j-\beta_j)}{|E||\sqrt{((X^TX)^{-1}_{j,j})}} \sim T_{n-(p+1)}$

(y) [easy] Let $H_0 : \beta_j = 0$. Find the test statistic using the fact from the previous question. Let s_e denote $RMSE := \sqrt{MSE} := \sqrt{SSE/(n-(p+1)} = \sqrt{||e||^2/(n-(p+1)}$.

$\frac{b_j}{S_e\sqrt{((X^TX)^{-1}_{j,j})}}$

(z) [easy] Consider a new parameter of interest $\mu_\star := \mathbb{E}[Y_\star] = \boldsymbol{x}_\star\boldsymbol{\beta}$, this is the expected response for a unit with measurements given in row vector $\boldsymbol{x}_\star$ whose first entry is 1. Prove that $\dfrac{\hat{Y}_\star - \mu_\star}{\sigma\sqrt{\boldsymbol{x}_\star\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{x}_\star^\top}} \sim \mathcal{N}(0,\,1)$.

$\mathbf{E}[Y_*] = \mathbf{x}_*\beta$ so $h^*(\vec{x_*}) = \vec{x_*}\vec{\beta}$ so $\hat{\vec{Y}} \sim N_n(X\vec{\beta}, \sigma^2 H) \rightarrow \hat{\vec{Y_i}} \sim N(\vec{x_{i.}}\vec{\beta}, \sigma^2 H_{i,i})$, so as:
$\vec{Y_*} = \vec{x_*}\vec{B} \rightarrow \hat{Y_*} \sim N(\vec{x_*}\vec{\beta}, \sigma^2\vec{x_*}(X^TX)^{-1}\vec{x_*}^T)$, and
$\frac{\hat{Y_*}-\mu_*}{\sqrt{\sigma^2\vec{x_*}(X^TX)^{-1}\vec{x_*}^T}} \sim N(0,1).\blacksquare$

(aa) [easy] Prove that $\dfrac{\sqrt{n-(p+1)}(\hat{Y}_\star - \mu_\star)}{||\boldsymbol{E}|| \sqrt{\boldsymbol{x}_\star \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{x}_\star^\top}} \sim T_{n-(p+1)}$.

$\frac{\hat{Y}_* - \mu_*}{\sqrt{\sigma^2 \vec{x_*}(X^TX)^{-1}\vec{x_*}^T}} \sim N(0,1)$ independent of
$\frac{1}{\sigma^2}||\mathbf{E}||^2 \sim \chi^2_{n-(p+1)}$, so
$\frac{\frac{\hat{Y}_* - \mu_*}{\sqrt{\sigma^2 \vec{x_*}(X^TX)^{-1}\vec{x_*}^T}}}{\sqrt{\frac{\frac{1}{\sigma^2}||\mathbf{E}||^2}{n-(p+1)}}} = \frac{\hat{Y}_* - \mu_*}{\sqrt{\frac{||\mathbf{E}||^2}{n-(p+1)}}\sqrt{\vec{x_*}(X^TX)^{-1}\vec{x_*}^T}} \sim T_{n-(p+1)}$

(bb) [easy] Let $H_0 : \mu_\star = 17$. Find the test statistic using the fact from the previous question. Let s_e denote the $RMSE$.

$\frac{\hat{Y}_* - 17}{s_e\sqrt{\vec{x_*}(X^TX)^{-1}\vec{x_*}^T}} \sim T_{n-(p+1)}$

(cc) [easy] Consider a new parameter of interest $y_\star = \boldsymbol{x}_\star \boldsymbol{\beta} + \epsilon_\star$, this is the response for a unit with measurements given in row vector $\boldsymbol{x}_\star$ whose first entry is 1. Prove that $\dfrac{\hat{Y}_\star - y_\star}{\sigma\sqrt{1 + \boldsymbol{x}_\star \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{x}_\star^\top}} \sim \mathcal{N}(0,\, 1)$.

ϵ_i's are assumed iid and $\sim N(0,1)$, and so are ϵ_*. So
$Y_* - \hat{Y}_* = Y_* - \vec{x_*}\vec{B} = \vec{x_*}\vec{\beta} + \epsilon_* - \vec{x_*}\vec{B} \sim N(0, \sigma^2 + \sigma^2\vec{x_*}(X^TX)\vec{x_*}^T) = N(0, \sigma^2(1 + \vec{x_*}(X^TX)\vec{x_*}^T) \rightarrow \frac{Y_* - \hat{Y}_*}{\sqrt{\sigma^2(1+\vec{x_*}(X^TX)\vec{x_*}^T}} \sim N(0,1)$

(dd) [easy] Prove that $\dfrac{\sqrt{n-(p+1)}(\hat{Y}_\star - y_\star)}{||\boldsymbol{E}|| \sqrt{1 + \boldsymbol{x}_\star \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{x}_\star^\top}} \sim T_{n-(p+1)}$.

$\frac{Y_* - \hat{Y}_*}{\sqrt{\sigma^2(1+\vec{x_*}(X^TX)\vec{x_*}^T}} \sim N(0,1)$, independent of:
$\frac{1}{\sigma^2}||\mathbf{E}||^2 \sim \chi^2_{n-(p+1)}$, so

$\frac{\frac{Y_* - \hat{Y}_*}{\sqrt{\sigma^2(1+\vec{x_*}(X^TX)\vec{x_*}^T}}}{\sqrt{\frac{\frac{1}{\sigma^2}||\mathbf{E}||^2}{n-(p+1)}}} = \frac{\sqrt{n-(p+1)}Y_* - \hat{Y}_*}{||\mathbf{E}||\sqrt{1+\vec{x_*}(X^TX)\vec{x_*}^T}} \sim T_{n-(p+1)}$

(ee) [easy] Let $H_0 : y_\star = 37$. Find the test statistic using the fact from the previous question. Let s_e denote the $RMSE$.
$\frac{37 - \hat{y_*}}{s_e\sqrt{1+\vec{x_*}(X^TX)\vec{x_*}^T}}$

(ff) [difficult] Let $S \subseteq \{1, 2, \ldots, p\}$, let $k := |S|$ and let $A = \{0\} \cup S^C$, its complement with zero for the index of the intercept. For convenience, assume you rearrange the columns of the design matrix so that $\boldsymbol{X} = [\boldsymbol{X}_A \mid \boldsymbol{X}_S]$ and the first column is $\mathbf{1}_n$. Let $\boldsymbol{H}_A := \boldsymbol{X}_A(\boldsymbol{X}_A^\top \boldsymbol{X}_A)^{-1}\boldsymbol{X}_A^\top$. It is obvious that $\boldsymbol{H} - \boldsymbol{H}_A$ is symmetric as both $\boldsymbol{H}$ and $\boldsymbol{H}_A$ are symmetric. To prove that $\boldsymbol{H} - \boldsymbol{H}_A$ is an orthogonal projection matrix, prove that it is idempotent. Hint: use the Gram-Schmidt decomposition for both matrices and use block matrix format for $\boldsymbol{H}$.

(gg) [easy] Let $\hat{\boldsymbol{Y}}_A := \boldsymbol{H}_A \boldsymbol{Y}$, the orthogonal projection onto colsp $[\boldsymbol{X}_A]$. Prove that

$$\frac{(n-(p+1))\left|\left|\hat{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_A\right|\right|^2}{k\left|\left|\boldsymbol{E}\right|\right|^2} \sim F_{k,n-(p+1)}.$$

(hh) [difficult] Let $\hat{\boldsymbol{E}}_A := (\boldsymbol{I}_n - \boldsymbol{H}_A)\boldsymbol{Y}$, the orthogonal projection onto the colsp $[\boldsymbol{X}_{A_\perp}]$. Prove that $\left|\left|\hat{\boldsymbol{E}}_A\right|\right|^2 - \left|\left|\hat{\boldsymbol{E}}\right|\right|^2 = \left|\left|\hat{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_A\right|\right|^2$.

(ii) [easy] Combining the two previous problems, write the test statistic for $H_0 : \boldsymbol{\beta}_S = \boldsymbol{0}_k$ where β_S denotes the subvector of $\boldsymbol{\beta}$ with indices S. Use the notation $\Delta SSE := SSE_A - SSE$ and MSE.

(jj) [difficult] Prove that the square root of the test statistic in (ii) is the same as t-test statistic from (y) when $k = 1$.

(kk) [harder] The point of this exericse is to demonstrate that the estimator used for the omnibus / global / overall F-test is nothing but a special case of the main result from (gg). Let $S = \{1, 2, \ldots, p\}$ and thus $k = p$ and $A = \{0\}$. Using the result from (gg), show that $\dfrac{(n-(p+1))\left|\left|\hat{\boldsymbol{Y}} - \bar{y}\boldsymbol{1}_n\right|\right|^2}{p\left|\left|\boldsymbol{E}\right|\right|^2} \sim F_{p,n-(p+1)}$.

(ll) [easy] Prove that the omnibus / global / overall F-test statistic is $\hat{\hat{F}} = MSR/MSE$ by using the result from (kk).

(mm) [difficult] [MA] Prove that the distribution that realizes the R^2 metric (the proportion of response variance explained by the model) is distributed as Beta $\left(\frac{p}{2}, \frac{n-(p+1)}{2}\right)$. This amounts to proving a fact found at the bottom of the F distribution's Wikipedia page .

(nn) [easy] Prove that the maximum likelihood estimate for $\boldsymbol{\beta}$ is $\boldsymbol{b}$, the OLS estimator.

$\beta^{MLE} = \vec{b}$

(oo) [harder] Prove that the maximum likelihood estimate for σ^2 is SSE/n.

${\sigma^2}^{MLE} = \frac{SSE}{n}$

(pp) [harder] Find the bias of the maximum likelihood estimator for σ^2 using your answers from (w) and (oo).

$-(p+1)SSE$

Problem 2

This problem is about two types of Bayesian estimation of the slope parameters in linear regression which lead to the ridge and lasso estimates.

(a) [easy] Write the prior assumption about $\boldsymbol{\beta}$ which yields the ridge estimates.

Jeffrey's Prior, that $\vec{\beta}$ is made of iid r.v's, which are distributed $\sim N(0, \tau^2)$

(b) [easy] Using the prior and core assumption (which implies a likelihood function for $\boldsymbol{B}$), derive the ridge estimates.

$\vec{b}_{RIDGE} = (X^T X + \lambda_{p+1}^T)^{-1} X^T \vec{y}$

(c) [easy] Write the prior assumption about $\boldsymbol{\beta}$ which yields the lasso estimates.

Laplace's Prior, that $\vec{\beta}$ is made of iid r.v's, which are distributed $\sim Laplace(0, \tau^2) := \frac{1}{2\tau^2} \boldsymbol{e}^{\frac{-|\beta|}{\tau^2}}$

(d) [easy] Using the prior and core assumption (which implies a likelihood function for $\boldsymbol{B}$), derive the lasso estimates to the point where you need to use a computer to run the optimization.

$\vec{b}_{LASSO} = argmin\{||\vec{y} - X\vec{\beta}||^2 + \frac{2\sigma^2}{\tau^2} \sum_{j=0}^{p} |\beta_j\} = argmin\{SSE + \lambda \sum_{j=0}^{p} |\beta_j|\}, where \lambda = \frac{2\sigma^2}{\tau^2}$

(e) [easy] Both ridge and lasso shrink the estimate of $\boldsymbol{\beta}$ towards what vector value?

The zero vector: $\vec{0_{p+1}}$

(f) [easy] Describe what the prestep called "variable selection" is within the modeling enterprise.

Variable selection is when we chose certain p's and build our model based on those p's.

(g) [easy] Describe why Lasso estimation has the added bonus of being able to perform variable selection and ridge does not.

Lasso picks specific p when p is large. Ridge does not have this property.

Problem 3

This problem is about the specific robust regression methods we studied.

(a) [easy] If we only know that the errors $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are independent, what tried and true method can we employ to get asymptotically valid inference for $\boldsymbol{\beta}$?

Bootstrapping!

(b) [easy] If we know that the errors $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are iid with expectation zero and variance σ^2 for all values of $\boldsymbol{x}$ (i.e. the errors are "homoskedastic") but the errors are not necessarily normal, what is the asymptotic distribution of $\boldsymbol{B}$?

$\vec{B} \dot{\sim} N(\vec{\beta}, \sigma^2 (X^T X)^{-1})$

(c) [easy] If we know that the errors $\mathcal{E}_1, \ldots, \mathcal{E}_n \stackrel{ind}{\sim} \mathcal{N}(0, \sigma_i^2)$ which means the errors are "heteroskedastic", what is the asymptotic distribution of $\boldsymbol{B}$ using the Huber-White estimator?

$\vec{B} \dot{\sim} N_{p+1}(\vec{\beta}, (X^T X)^{-1} X^T \hat{D} (X^T X)^{-1})$ where $\hat{D} = diag(E_i^2, n)$

(d) [easy] If we know that the errors $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are independent with expectation zero and variance σ_i^2 which means the errors are "heteroskedastic", what is the asymptotic distribution of $\boldsymbol{B}$ using the Huber-White estimator?

$\vec{B} \dot{\sim} N_{p+1}(\vec{\beta}, (X^T X)^{-1} X^T \hat{D} (X^T X)^{-1})$ where $\hat{D} = diag(E_i^2, n)$

(e) [easy] Is the F-tests we derived under the core assumption valid in any of the four above scenarios? Yes/no

No. We need three conditions for it: Normality, Homoskedascity, and independence of errors. 1, 2, 4 are not necessarily normal, while 3 is not homoskedastic.

Problem 4

This problem is about inference for the generalized linear model (glm).

(a) [harder] Let $Y_i \stackrel{ind}{\sim} \text{Bernoulli}(\theta_i)$ for $i = 1, \ldots, n$ where $\theta_i = \phi(\boldsymbol{x}_i \boldsymbol{\beta})$ and $\boldsymbol{x}_i \in \mathbb{R}^{p+1}$ whose first entry is always 1. For the link function, use the complementary log-log (i.e. the standard Gumbel CDF). Write out the full likelihood below. No need to simplify or take the log.

(b) [harder] Given the assumptions in (a), write the likelihood ratio estimate for the omnibus test of $H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$.

(c) [harder] Let $Y_i \stackrel{ind}{\sim} \text{Poisson}(\theta_i)$ for $i = 1, \ldots, n$ where $\theta_i = e^{\boldsymbol{x}_i \boldsymbol{\beta}}$ and $\boldsymbol{x}_i \in \mathbb{R}^3$ whose first entry is always 1. Write out the likelihood ratio when testing $H_0 : \beta_2 = \beta_3 = 0$.

(d) [harder] Let $Y_i \stackrel{ind}{\sim} \text{Weibull}(k, \theta_i)$ for $i = 1, \ldots, n$ where $\theta_i = e^{\boldsymbol{x}_i \boldsymbol{\beta}}$ and $\boldsymbol{x}_i \in \mathbb{R}^3$ whose first entry is always 1. This uses the alternate parameterization so that $\mathbb{E}[Y_i] = \theta_i \Gamma(1 + 1/k)$. There is a censoring vector $\boldsymbol{c}$ which is 1 when censored on the right (meaning the real y_i is $\geq$ to the observed y_i) and 0 when not censored. Write out the likelihood ratio when testing $H_0 : \beta_2 = \beta_3 = 0$.

(e) [difficult] [MA] Let $Y_i \stackrel{ind}{\sim} \mathcal{N}(\theta_i, \sigma^2)$ for $i = 1, \ldots, n$ where $\theta_i = \boldsymbol{x}_i \boldsymbol{\beta}$. This is the regular linear model. However there is a censoring vector $\boldsymbol{c}$ which is 1 when censored on the right (meaning the real y_i is $\geq$ to the observed y_i) and 0 when not censored. This is called the Tobit model. Write the likelihood ratio estimate for the omnibus test of $H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$.