Statistics 600 Problem Set 1

Due Wednesday September 29th at midnight

1. Suppose we use least squares to perform multiple linear regression, yielding fitted values $\hat{y} = X\hat{\beta} \in \mathcal{R}^n$. Do these fitted values maximize the sample correlation coefficient with the response vector, i.e. does the following hold?

$$\widehat{\text{cor}}(X\hat{\beta}, y) = \max_b \widehat{\text{cor}}(Xb, y)$$

Justify your answer.

Solution: Since the question only involves fitted values, we can linearly transform the columns of X as follows. Supposing that $1_n \in \text{col}(X)$ we can take the first column of X to be 1_n, and the remaining columns of X to be centered. Let $\tilde{X}$ denote the $n \times p$ matrix consisting of columns 2 through $p+1$ of X. The coefficient vector b can be partitioned $b = (b_0, \tilde{b})$ where $b_0 \in \mathcal{R}$ and $\tilde{b} \in \mathcal{R}^p$. Note that for any such b, $Xb - \overline{Xb} = \tilde{X}\tilde{b}$. Then

$$\begin{aligned} \widehat{\text{cor}}(Xb, y) &= \widehat{\text{cor}}(\tilde{X}\tilde{b}, y) \\ &= \frac{y'\tilde{X}\tilde{b}}{\|y - \bar{y}\| \cdot (\tilde{b}'\tilde{X}'\tilde{X}\tilde{b})^{1/2}} \\ &\propto \frac{y'\tilde{X}\tilde{b}}{(\tilde{b}'\tilde{X}'\tilde{X}\tilde{b})^{1/2}}. \end{aligned}$$

Write $\tilde{X}'\tilde{X} = R'R$, where R is square and invertible (e.g. using the QR factorization $\tilde{X} = QR$, so $\tilde{X}'\tilde{X} = R'R$). Now change variables by setting $u = Rb$, and the expression above becomes

$$y'\tilde{X}R^{-1}u/\|u\|.$$

By the Cauchy-Schwarz inequality, the above is maximized by setting $u \propto R^{-T}\tilde{X}^T y$, and changing back to the original coordinates we get $\hat{\tilde{b}} \propto R^{-1}R^{-T}\tilde{X}^T y = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'y$. Since the first column of X is orthogonal to all columns of $\tilde{X}$, the least squares fit of y on X is equal to $\bar{y} + \tilde{X}\hat{\tilde{b}}$, which has the same correlation with y as does $\tilde{X}\hat{\tilde{b}}$. Thus the least squares fitted values maximize the correlation, as long as $1_n \in \mathrm{col}(X)$.

If 1_n is not in $\mathrm{col}(X)$, then let $P = I - 1_n 1_n'/n$ denote the centering matrix. Then,

$$\begin{aligned}\widehat{\mathrm{cor}}(Xb, y) &= \frac{y'(I-P)Xb}{\|y-\bar{y}\| \cdot (b'X'(I-P)Xb)^{1/2}} \\ &\propto \frac{y'(I-P)Xb}{(b'X'(I-P)Xb)^{1/2}}.\end{aligned}$$

Since 1 is not in $\mathrm{col}(X)$, $X'(I-P)X$ is non-singular, so can be written in the form $R'R$, and changing variables then using Cauchy-Schwarz as above, we obtain that the maximizer of the correlation is $\hat{b} = (X'(I-P)X)^{-1}X'(I-P)y$. It is easy to show by example that this choice of b, which is obtained by using OLS on the centered design matrix, has higher correlation with y then the usual least squares fit with un-centered columns. Thus, OLS does not maximize the correlation coefficient with y when no intercept is in the model.

2. Suppose we have a least squares problem with more variables than observations. That is, we observe a response vector $y \in \mathcal{R}^n$, and a design matrix $X \in \mathcal{R}^{n\times p}$ where $p \geq n$ and the rows of X are linearly independent.

   (a) Derive an expression for the vector $\hat{\beta}$ that minimizes $\|\beta\|^2$ subject to $X\beta = y$.

   Solution

   Using the QR decomposition, write $X' = QR$, and since X' is nonsingular, R is invertible and the equation $X\beta = y$ becomes $Q'\beta = g$, where $g = R^{-T}y$. Next we will show that $\hat{\beta} \in \mathrm{col}(Q) = \mathrm{row}(X)$. We can write $\beta = \theta + \gamma$, where $\theta \in \mathrm{col}(Q)$ and $\gamma \in \mathrm{col}(Q)^\perp$. Note that $Q'\beta = Q'\theta$, and $\|\beta\|^2 = \|\theta\|^2 + \|\gamma\|^2$. Thus for any choice of θ satisfying $Q'\theta = g$, $\|\beta\|^2$ will always be minimized by setting $\gamma = 0$. Since $\theta \in \mathrm{col}(Q)$, we can write $\theta = Q\eta$ for some

$\eta \in \mathcal{R}^n$, and we have $Q'Q\eta = \eta = g$, and $\theta = Qg$. Thus the solution is $\hat{\beta} = QR^{-T}y$.

(b) Under what conditions is $\hat{\beta}$ unbiased? You may take the usual generating model $y = X\beta + \epsilon$ with $E[\epsilon|X] = 0$.

Solution We can write

$$\begin{aligned}
\hat{\beta} &= QR^{-T}y \\
&= QR^{-T}(X\beta + \epsilon) \\
&= QR^{-T}(R'Q'\beta + \epsilon) \\
&= QQ'\beta + QR^{-T}\epsilon.
\end{aligned}$$

Thus $E[\hat{\beta}|X] = QQ'\beta$, which is equal to β under the condition that $\beta \in \mathrm{col}(Q) = \mathrm{col}(X')$.

(c) Derive an expression for $\mathrm{cov}[\hat{\beta}|X]$, under the generating model $y = X\beta + \epsilon$ with $E[\epsilon|X] = 0$ and $\mathrm{cov}[\epsilon|X] = \sigma^2 I$.

Solution

$$\begin{aligned}
\mathrm{cov}(\hat{\beta}|X) &= \mathrm{cov}[QR^{-T}y|X] \\
&= \mathrm{cov}[QR^{-T}[X\beta + \epsilon)|X] \\
&= \mathrm{cov}[QR^{-T}(R'Q'\beta + \epsilon)|X] \\
&= \mathrm{cov}[QQ'\beta + QR^{-T}\epsilon)|X] \\
&= \mathrm{cov}[QR^{-T}\epsilon)|X] \\
&= \sigma^2 QR^{-T}R^{-1}Q'.
\end{aligned}$$

(d) What is the value of $E\|\hat{y} - y\|^2$?

Solution

$$\hat{y} = XQR^{-T}y = R'Q'QR^{-T}y = y.$$

Thus $E\|\hat{y} - y\|^2 = 0$ – or $\hat{y}$ is always equal to Y.

(e) What is the value of $E\|\hat{y} - Ey\|^2/n$?

Solution

$$E\|\hat{y} - Ey\|^2/n = E\|y - Ey\|^2/n = \sigma^2.$$

(f) Suppose we observe a random vector $y^* \in \mathcal{R}^n$ that has the same distribution as y, but is independent of y. What is the value of $E\|y^* - \hat{y}\|^2/n$?

Solution

Write $y^* = X\beta + \epsilon^*$,

$$E\|X\beta + \epsilon^* - (X\beta + \epsilon)\|^2/n = E\|\epsilon - \epsilon^*\|^2/n = 2\sigma^2.$$

3. Suppose we observe data from a simple linear model $y = \alpha + \beta x + \epsilon$ where $x, y, \epsilon \in \mathcal{R}^n$, $\alpha, \beta \in \mathcal{R}$, n is an even integer, $E[\epsilon|X] = 0$ and $\mathrm{cov}[\epsilon|X] = \sigma^2 I$. Suppose x and y are partitioned as

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \qquad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

where y_1 and y_2 each have half the length of y, and x_1 and x_2 each have half the length of x. Let $\hat{\beta}_1$ and $\hat{\beta}_2$ denote the least squares estimates obtained by regressing y_1 on x_1 and y_2 on x_2, respectively, and let $\tilde{\beta} = (\hat{\beta}_1 + \hat{\beta}_2)/2$.

(a) If $\bar{x}_1 = \bar{x}_2 = \bar{x}$, state a condition such that $\tilde{\beta}$ has the same variance as the least squares estimate $\hat{\beta}$ obtained by regressing y on x (using all n observations). Then state whether when this condition holds, $\tilde{\beta}$ is the least squares estimate, or is a different estimate with the same variance.

Solution

Let $T_1 = \sum_{i=1}^{n/2} \epsilon_i(x_i - \bar{x}_1)$ and $T_2 = \sum_{i=n/2+1}^{n} \epsilon_i(x_i - \bar{x}_2)$, and let $S_1 = \sum_{i=1}^{n/2}(x_i - \bar{x}_1)^2$ and $S_2 = \sum_{i=n/2+1}^{n}(x_i - \bar{x}_2)^2$. Then

$$\hat{\beta}_1 = \beta + T_1/S_1,$$

$$\hat{\beta}_2 = \beta + T_2/S_2,$$

and

$$\tilde{\beta} = \beta + \frac{T_1}{2S_1} + \frac{T_2}{2S_2}.$$

Since $\mathrm{var}(T_j) = \sigma^2 S_j$ for $j = 1, 2$, it follows that

$$\mathrm{var}\tilde{\beta} = \frac{\sigma^2}{4S_1} + \frac{\sigma^2}{4S_2}.$$

The variance of the least squares estimate using all the data is

$$\sigma^2 / \sum_i (X_i - \bar{X})^2 = \sigma^2/(S_1 + S_2).$$

The two variances are equal if only if

$$(S_1 + S_2)^2 = 4S_1S_2,$$

which is easily seen to hold if and only if $S_1 = S_2$. This is the condition required for the variance of $\tilde{\beta}$ to equal the variance of $\hat{\beta}$, and it is easy to see that when $S_1 = S_2$, $\tilde{\beta} = \hat{\beta}$.

(b) Now consider the more general case where $\bar{x}_1$ and $\bar{x}_2$ may differ. Show that in this case $\mathrm{var}\,\tilde{\beta}$ is always at least as great as $\mathrm{var}\,\hat{\beta}$, and derive a concise expression for the difference between the two variances.

Solution

By the Gauss-Markov theorem, since $\tilde{\beta}$ is linear and unbiased, if $\tilde{\beta} \not\equiv \hat{\beta}$, then $\text{var}(\tilde{\beta})$ must be greater than $\text{var}(\hat{\beta})$.

We can show this directly as follows.

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \bar{x})^2 &= \sum_{i=1}^{n/2}(x_i - \bar{x}_1 + \bar{x}_1 - \bar{X}_2 + \bar{x}_2 - \bar{x})^2 + \\ &\quad \sum_{i=n/2+1}^{n}(x_i - \bar{x}_1 + \bar{x}_1 - \bar{x}_2 + \bar{x}_2 - \bar{x})^2\end{aligned}$$

Taking the first term,

$$\begin{aligned}&\sum_{i=1}^{n/2}(X_i - \bar{X}_1 + \bar{X}_1 - \bar{X}_2 + \bar{X}_2 - \bar{X})^2 \\ &= \sum_{i=1}^{n/2}(X_i - \bar{X}_1)^2 + (\bar{X}_1 - \bar{X}_2)^2 + (\bar{X}_2 - \bar{X})^2 \\ &\quad +(X_i - \bar{X}_1)(\bar{X}_1 - \bar{X}_2) + (X_i - \bar{X}_1)(\bar{X}_2 - \bar{X}) \\ &\quad +(\bar{X}_1 - \bar{X}_2)(\bar{X}_2 - \bar{X}) \\ &= \sum_{i=1}^{n/2}(X_i - \bar{X}_1)^2 + (\bar{X}_1 - \bar{X}_2)^2 + (\bar{X}_2 - \bar{X})^2 + (\bar{X}_1 - \bar{X}_2)(\bar{X}_2 - \bar{X}) \\ &= S_1 + n(\bar{X}_1 - \bar{X}_2)^2/2 + n(\bar{X}_2 - \bar{X})^2/2 + n(\bar{X}_1 - \bar{X}_2)(\bar{X}_2 - \bar{X})/2.\end{aligned}$$

We can apply a similar calculation to obtain

$$\begin{aligned}&\sum_{i=n/2+1}^{n}(X_i - \bar{X}_1 + \bar{X}_1 - \bar{X}_2 + \bar{X}_2 - \bar{X})^2 \\ &= S_2 + n(\bar{X}_1 - \bar{X}_2)^2/2 + n(\bar{X}_1 - \bar{X})^2/2 + n(\bar{X}_1 - \bar{X}_2)(\bar{X}_1 - \bar{X})/2.\end{aligned}$$

Since

$$(\bar{X}_1 - \bar{X}_2)(\bar{X}_2 - \bar{X}) + (\bar{X}_1 - \bar{X}_2)(\bar{X}_1 - \bar{X}) = 0,$$

we have

$$\sum_{i=1}^{n}(X_i-\bar{X})^2 = S_1+S_2+n(\bar{X}_1-\bar{X}_2)^2+n(\bar{X}_1-\bar{X})^2/2+n(\bar{X}_2-\bar{X})^2/2.$$

Thus the difference in variances is

$$\text{var}(\tilde{\beta}) - \text{var}(\hat{\beta}) = \sigma^2/4S_1 + \sigma^2/4S_2 - 1/(S_1+S_2+D)$$

where $D = n(\bar{X}_1 - \bar{X})^2/n + n(\bar{X}_2 - \bar{X})^2/2$.

The difference in variances simplifies to

$$\frac{(S_1-S_2)^2 + D(S_1+S_2)}{4S_1S_2(S_1+S_2+D)}.$$

4. Prove that the "horizontal residuals" in simple linear regression sum to zero in a least squares fit of y (the dependent variable) on x (the independent variable). The horizontal residuals are the signed horizontal displacements along the line segments connecting each data point x_i,y_i to the fitted line $\hat{\alpha} + \hat{\beta}x$.

   Solution

   To get the i^{th} horizontal residual, solve

   $$\hat{\alpha} + \hat{\beta}x = y_i$$

   to get $\hat{x}_i = (y_i - \hat{\alpha})/\hat{\beta}$, so the residual becomes $h_i \equiv x_i - (y_i - \hat{\alpha})/\hat{\beta}$. Now if we sum these values we get

   $$\begin{aligned}\sum_i h_i &= \sum_i x_i - (y_i - \bar{y} + \hat{\beta}\bar{x})/\hat{\beta} \\ &= \sum_i(-y_i + \bar{y} + \hat{\beta}(x_i - \bar{x}))/\hat{\beta}\end{aligned}$$

$$\begin{aligned} &= \sum_i (\bar{y} - y_i)/\hat{\beta} + \sum_i (x_i - \bar{x}) \\ &= 0. \end{aligned}$$

5. (a) Suppose that $F \in \mathcal{R}^d$ is a vector, and I is the $d \times d$ identity matrix. Derive explicit expressions for $(I + FF')^{-1}$ and $(I - FF')^{-1}$. Hint: the answers have the form $I + \lambda FF'$, for $\lambda \in \mathcal{R}$.

Solution

To determine the inverse of $I + FF'$, set

$$\begin{aligned} I &= (I + FF')(I + \lambda FF') \\ &= I + \lambda FF' + FF' + \lambda \|F\|^2 FF' \\ &= I + (\lambda + 1 + \lambda \|F\|^2) FF'. \end{aligned}$$

We must have $1 + \lambda(1 + \|F\|^2) = 0$, so $\lambda = -1/(1 + \|F\|^2)$.

To determine the inverse of $I - FF'$, set

$$\begin{aligned} I &= (I - FF')(I + \lambda FF') \\ &= I + \lambda FF' - FF' - \lambda \|F\|^2 FF' \\ &= I + (\lambda - 1 - \lambda \|F\|^2) FF'. \end{aligned}$$

We must have $-1 + \lambda(1 - \|F\|^2) = 0$, so $\lambda = 1/(1 - \|F\|^2)$.

(b) Suppose we have an orthogonal design matrix $X \in \mathcal{R}^{n \times p+1}$, and we are able to add one additional observation to the data set (i.e. add one row to X). This row, denoted x, must satisfy the constraint $\|x\|^2 = 1$. Describe how x should be chosen so as to minimize the maximum of the variances of $\hat{\beta}_0, \ldots, \hat{\beta}_p$.

Solution

Let $\hat{\beta}$ denote the slope estimates based on all $n + 1$ cases. Then $X'X = I + xx'$, so $\text{cov}(\hat{\beta}) = I - xx'/2$. Thus the variance of $\hat{\beta}_j$ is

$\sigma^2(1 - x_j^2/2)$. The maximum of these variances is determined by the smallest of the x_j^2. Thus we want to maximize $\min_j x_j^2$ subject to $\sum_j x_j^2 = 1$. The solution is to have $x_j = 1/\sqrt{p+1}$ for all j.

6. (a) Derive an expression for $\mathrm{cov}(y, \hat{y})$, i.e. the $n \times n$ matrix containing all population covariances between elements of y and elements of $\hat{y}$.

Solution

Let P denote the projection onto the columnspace of X. Then,

$$\begin{aligned} \mathrm{cov}(y, \hat{y}) &= \mathrm{cov}(y, Py) \\ &= \mathrm{cov}(\epsilon, P\epsilon) \\ &= E[\epsilon\epsilon']P \\ &= \sigma^2 P. \end{aligned}$$

(b) Derive an expression for the expected value of the sample covariance between the observed and fitted values, $E\widehat{\mathrm{cov}}(\hat{y}, y)$ – note that this is a scalar. Consider whether this covariance can or cannot be positive, negative, or zero.

Solution

Let P be the projection matrix onto $\mathrm{col}(X)$. Then,

$$\begin{aligned} \widehat{\mathrm{cov}}(\hat{y}, y) &= (Py)'(y - \bar{y})/n \\ &= (y - \bar{y} + \bar{y})'P(y - \bar{y})/n \\ &= (y - \bar{y})'P(y - \bar{y})/n + \bar{y}'P(y - \bar{y})/n. \end{aligned}$$

Here, $\bar{y}$ is interpreted as an n-vector in which all values are equal to the sample mean of the y_i. This can be written $\bar{y} = n^{-1}\mathbf{1}\mathbf{1}'y$, where $\mathbf{1}$ is a n-vector of 1's. Since there is an intercept in the model, $P\mathbf{1} = 1$, so the second summand above is equal to

$$\frac{n^{-1}\mathbf{1}\mathbf{1}'(y-\bar{y})}{n},$$

which is zero since $\mathbf{1}'(y-\bar{y}) = 0$. Thus

$$\widehat{\text{cov}}(\hat{y}, y) = \frac{(y-\bar{y})'P(y-\bar{y})}{n} \geq 0.$$

The covariance cannot be negative. It can only be zero if $y - \bar{y} \in \text{span}(X)^{\perp}$.

7. "Total least squares" (TLS) for one covariate aims to identify a line ℓ that minimizes

$$\sum_i d\left((X_i, Y_i), \ell\right)^2,$$

where $d(Q, \ell)$ is the minimum distance in $\mathcal{R}^2$ between the point Q and any point on the line ℓ.

(a) Parameterize ℓ in the form $\{(X, \alpha + \beta X) | X \in \mathcal{R}\}$, for scalars α and β. Write down expressions for $d(Q, \ell)$ and a loss function that can be minimized to identify α and β. Both expressions should be explicit functions of α and β.

Solution

To identify the point on ℓ that is closest to X_i, Y_i, we minimize

$$(X - X_i)^2 + (\alpha + \beta X - Y_i)^2$$

as a function of X. Setting the first derivative to zero yields

$$X = \frac{X_i - \alpha\beta + Y_i\beta}{1 + \beta^2},$$

and the second derivative is $2(1+\beta^2)$, so this is a global minimizer. The loss function is

$$(1+\beta^2)^{-1}\sum_i R_i^2,$$

where $R_i = Y_i - \alpha - \beta X_i$ is the usual OLS residual.

(b) Parameterize ℓ in the form $\{Z \in \mathcal{R}^2 | B'(Z-W) = 0\}$, for 2-vectors B and W with $\|B\| = 1$. Write down expressions for $d(Q, \ell)$ and a loss function that can be minimized to identify B and W (W can be any point on ℓ and is therefore not uniquely identified). Both expressions should be explicit functions of B and W.

Solution

Let $Q_i = (X_i, Y_i)$ be a data point. Let P_i be the point on ℓ that is closest to Q_i. Then $Q_i - P_i$ is parallel to B, so we can write $P_i = Q_i - \lambda B$ for some $\lambda \in \mathcal{R}$, and since P_i is on ℓ we must have $B'(P_i - W) = 0$. Combining these two equations we can identify $\lambda = B'(Q_i - W)$. Therefore the $d(Q_i, \ell)^2$ is

$$B'(Q_i - W)(Q_i - W)'B$$

so the loss function is

$$B'\left(\sum_i (Q_i - W)(Q_i - W)'\right)B.$$

(c) Based on your expression in part (b), show that the TLS solution passes through the center of the data $(\bar{X}, \bar{Y})$, and use this to define a minimizing value for W.

Solution

$$\begin{aligned}\sum_i (Q_i - W)(Q_i - W)' &= \sum_i (Q_i - \bar{Q} + \bar{Q} - W)(Q_i - \bar{Q} + \bar{Q} - W)' \\ &= \sum_i (Q_i - \bar{Q})(Q_i - \bar{Q})' + \sum_i (Q_i - \bar{Q})(\bar{Q} - W)' +\end{aligned}$$

$$\begin{aligned} & \sum_i(\bar{Q}-W)(Q_i-\bar{Q})' + n(\bar{Q}-W)(\bar{Q}-W)' \\ = & \sum_i(Q_i-\bar{Q})(Q_i-\bar{Q})' + n(\bar{Q}-W)(\bar{Q}-W)'. \end{aligned}$$

therefore the value of the loss function will either stay constant or be reduced if we set $W = \bar{Q}$, which guarantees that ℓ contains $\bar{Q}$.

(d) Building on (b) and (c), construct a quadratic form whose minimizing value subject to $\|B\| = 1$ solves the TLS problem for B.

Solution

The quadratic form is

$$B'\left(\sum_i(Q_i-\bar{Q})(Q_i-\bar{Q})'\right)B.$$

8. (a) Suppose we are fitting a simple linear regression model to a data set of size n. Let $V_n = \widehat{\text{var}}(x_1, \ldots, x_n)$. Determine the fastest rate at which $V_n \to 0$ for which we still have $\text{var}(\hat{\beta}_n) \to 0$.

Solution

Since

$$\text{var}(\hat{\beta}) = \frac{\sigma^2}{(n-1)V_n}$$

we need $nV_n \to \infty$ (or $V_n \to 0$ "slower than $1/n$").

(b) Suppose we are fitting a regression model with two explanatory variables, having the form $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$, and the covariates are asymptotically standardized so that $\bar{x}_1, \bar{x}_2 \to 0$, and $\widehat{\text{var}}(x_1), \widehat{\text{var}}(x_2) \to 1$. Let $r_n = \widehat{\text{cov}}(x_1, x_2)$. What is the fastest rate at which $r_n \to 1$ such that we will still have $\text{var}(\hat{\beta}_1), \text{var}(\hat{\beta}_2) \to 0$?

Solution

The variance of $\hat{\beta}_1$ (which is the same as the variance of $\hat{\beta}_2$ is

$$\frac{1}{n(1-r_n^2)} = \frac{1}{n(1-r_n)(1+r_n)}.$$

So we need $n(1-r_n) \to \infty$, or $1-r_n$ goes to zero "slower than rate $1/n$".

9. This exercise aims to illustrate the effect of outliers in least squares fitting. Suppose we observe data that follows a linear model with $p = 1$ covariate: $y = \alpha + \beta x + \epsilon$. Specifically, consider a triangular array of data y_{in}, x_{in}, where $i = 1, \ldots, n$. There is also a random indicator $\delta_{in} \in \{0, 1\}$, that we do not observe, such that $\text{var}(\epsilon_{in}|X, \delta_{in} = 1) = k_n\sigma^2$, and $\text{var}(\epsilon_{in}|X, \delta_{in} = 0) = \sigma^2$ (the errors are centered, so that $E[\epsilon|X, \delta] \equiv 0$). Suppose the x_i are sampled independently from a population with variance σ_x^2, and $P(\delta_{in} = 1) = p_n$. Derive conditions on k_n and p_n such that (i) $n \cdot \text{var}(\hat{\beta})$ has a finite limit, and (ii) $n \cdot \text{var}(\hat{\beta})$ has the same limit that would occur if $k_n \equiv 1$.

Solution

The least squares estimator can be written

$$\hat{\beta}_n = \beta + \sum_i \epsilon_{in}(x_{in} - \bar{x}_n) / \sum_i (x_{in} - \bar{x}_n)^2.$$

Since the variance of the error term can be expressed

$$\begin{aligned}
\text{var}(\epsilon_{in}) &= \text{var}E(\epsilon_{in}|\delta_{in}) + E\text{var}(\epsilon_{in}|\delta_{in}) \\
&= \sigma^2(p_n k_n + 1 - p_n),
\end{aligned}$$

the variance of the estimator is

$$\text{var}\hat{\beta}_n = \sigma^2(p_n k_n + 1 - p_n) / \sum_i (X_{in} - \bar{X}_n)^2.$$

Scaling by n,

$$n \times \mathrm{var}\hat{\beta}_n = \sigma^2(p_n k_n + 1 - p_n)/n^{-1}\sum_i (X_{in} - \bar{X}_n)^2 \sim \sigma^2(p_n k_n + 1 - p_n)/\sigma_x^2.$$

Thus for (i), we need $p_n k_n + 1 - p_n$ to have a limit, and for (ii), we need $p_n(k_n - 1) \to 0$. A reasonable interpretation of this is that the outliers will not prevent the variance of $\hat{\beta}$ from going to zero at the usual rate as long as $p_n k_n$ stays bounded. For example, if fraction $p_n = 0.1$ of the errors have $k_n = 10$ times greater variance ($\sqrt{10} \approx 3.2$ times greater standard deviation), then the variance of $\hat{\beta}_n$ will decrease at the usual rate. But if we want the limiting variance ($\lim_{n\to\infty} n\mathrm{var}\hat{\beta}_n$) to be the same as when no outliers are present, we would need p_n to be much smaller, say $p_n = 0.01$.