Outline of Solution to Homework 1

Problem 2 Suppose that data $\{Y_i : i = 1, \ldots, n\}$ are observed along with concomitant information $\{x_i : i = 1, \ldots, n\}$. In order to capture the variability in Y, a linear regression of Y on x is modeled:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i; \ \ i = 1, \ldots, n,$$

where the errors $\{\epsilon_i\}$ are assumed independent and identical distributed (i.i.d) $\mathrm{N}(0, \sigma^2)$; σ^2 unknown.

(a) Derive formulas for the ordinary least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

Solution. Let $RSS(\boldsymbol{\beta}; \boldsymbol{Y}, \boldsymbol{x}) \equiv \sum_{i=1}^n (Y_i - [\beta_0 + \beta_1 x_i])^2$ be the residual sum of squares. Want to pick β_0 and β_1 such that RSS is minimized w.r.t. β_0 and β_1.
The partial derivatives are given by

$$\frac{\partial RSS}{\partial \beta_0} = 2n(\bar{Y} - \beta_0 - \beta_1 \bar{x}),$$
$$\frac{\partial RSS}{\partial \beta_1} = 2\left(\sum_{i=1}^n Y_i x_i - \beta_0 n\bar{x} - \beta_1 \sum_{1=1}^n x_i^2\right).$$

Setting the first equation equal to zero gives

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x},$$

and plugging this result into the second equation gives

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

(b) Define the residual sum of squares and derive an unbiased estimator of σ^2 based on it; call it $\hat{\sigma}^2$. Prove that $\hat{\sigma}^2$ is unbiased.

Solution. The residual sum of squares was defined in (a) as

$$RSS(\boldsymbol{\beta}; \boldsymbol{Y}, \boldsymbol{x}) \equiv \sum_{i=1}^n (Y_i - [\beta_0 + \beta_1 x_i])^2.$$

The first step in deriving an unbiased estimator on σ^2 based on RSS is to find the expected value of RSS.

There are number of ways to find the expected value of RSS.

The most lengthy way, and the method that assumes the least, is to attack this directly. That is

$$\begin{aligned}(Y_i - \hat\beta_0 - \hat\beta_1 x_i)^2 &= ([Y_i - \beta_0 - \beta_1 x_i] - [\hat\beta_0 - \beta_0] - [\hat\beta_1 - \beta_1]x_i)^2 \\ &= (\epsilon_i - [\hat\beta_0 - \beta_0] - [\hat\beta_1 - \beta_1]x_i)^2 \\ &= \mathrm{var}(\epsilon_i) + \mathrm{var}(\hat\beta_0) + \mathrm{var}(\hat\beta_1)x_i^2 \\ &\quad - 2\mathrm{cov}(\epsilon_i, \hat\beta_0) - 2\mathrm{cov}(\epsilon_i, \hat\beta_1)x_i + \mathrm{cov}(\hat\beta_0, \hat\beta_1)x_i.\end{aligned}$$

Then find the different $\mathrm{var}(\cdot)$ and $\mathrm{cov}(\cdot,\cdot)$ terms needed by using the fact that both $\hat\beta_1$ and $\hat\beta_0$ can be expressed as $\hat\beta_0 = \sum_{i=1}^n a_i Y_i$ and $\hat\beta_1 = \sum_{i=1}^n b_i Y_i$, for some a's and b's, and using the following result (since the Y's are independent and with identical variance),

$$\mathrm{var}\left(\sum_{i=1}^n a_i Y_i\right) = \sigma^2 \sum_{i=1}^n a_i^2 \quad \text{and} \quad \mathrm{cov}\left(\sum_{i=1}^n a_i Y_i, \sum_{i=1}^n b_i Y_i\right) = \sigma^2 \sum_{i=1}^n a_i b_i.$$

An easier approach uses matrix notation and a little fact about the expected value of a quadratic form.

In matrix notation, we have

$$\hat{\boldsymbol{Y}} \equiv \boldsymbol{X}\hat{\boldsymbol{\beta}}, \quad \text{with } \hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y},$$

where $\boldsymbol{X} = [\mathbf{1}\ \boldsymbol{x}]$, a matrix of two columns. Then

$$\boldsymbol{e} \equiv \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = (\boldsymbol{I} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Y} = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y},$$

with $\boldsymbol{P} \equiv \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. It is easy to show that $\boldsymbol{P}$ is symmetric and idempotent ($\boldsymbol{P}^2 = \boldsymbol{P}$) and, as a result, $(\boldsymbol{I} - \boldsymbol{P})$ is also symmetric and idempotent. Then the residual sum of squares can be writen as a quadratic form in $\boldsymbol{Y}$,

$$RSS = \sum_{i=1}^n (Y_i - \hat\beta_0 - \hat\beta_1 x_i)^2 = \boldsymbol{e}'\boldsymbol{e} = \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}.$$

A general result for quadratic forms states that if $\boldsymbol{Y}$ is a random variable with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{V}$ then the expected value of the quadratic form $\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}$ is given by $\mathrm{tr}(\boldsymbol{A}\boldsymbol{V}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$. Using this, we obtain

$$E[RSS] = E[\boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}] = \sigma^2 \mathrm{tr}(\boldsymbol{I} - \boldsymbol{P}) + \boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{X}\boldsymbol{\beta}.$$

Now, $\text{tr}(\boldsymbol{P}) = \text{tr}(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}') = \text{tr}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}) = \text{tr}(\boldsymbol{I}_2) = 2$; that is, $\text{tr}(\boldsymbol{I} - \boldsymbol{P}) = n - 2$. Also, $(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{X} = 0$, therefore,

$$E[RSS] = \sigma^2(n-2).$$

This directly suggests defining $\hat{\sigma}^2 \equiv RSS/(n-2)$ which is then an unbiased estimator of σ^2.

(c) Give a size α test of the hypotheses $H_0: \ \beta_1 = 0$ versus $H_1: \ \beta_1 \neq 0$.

Solution. The variance of $\hat{\beta}_1$ is $\sigma^2/\sum_{i=1}^n x_i^2$ and can be estimated by $s^2(\hat{\beta}_1) = \hat{\sigma}^2/\sum_{i=1}^n x_i^2$, with $\hat{\sigma}^2$ given in (b). If H_0 holds, then

$$T \equiv \frac{\hat{\beta}_1}{s(\hat{\beta}_1)} \sim \text{Student's } t \text{ with } n-2 \text{ degrees of freedom.}$$

Let $t_{n-2}(1-\alpha/2)$ be the $1-\alpha/2$ quantile of a Student's t distribution with $n-2$ degrees of freedom. Then a size α test of H_0 versus H_1 is given by:

$$\text{Reject } H_0 \text{ if } |T| > t_{n-2}(1-\alpha/2).$$

(d) Give a $100(1-\alpha)\%$ confidence interval for β_1.

Solution. From (c) we see that a $100(1-\alpha)\%$ confidence interval for β_1 can be given by

$$\left(\hat{\beta}_1 - t_{n-2}(1-\alpha/2)s(\hat{\beta}_1),\ \hat{\beta}_1 + t_{n-2}(1-\alpha/2)s(\hat{\beta}_1)\right).$$

Problem 3 Suppose Y is Poission with mean λ. Show that

$$E[1/(Y+1)] = \{1 - \exp(-\lambda)\}/\lambda.$$

Solution. We have

$$E[1/(Y+1)] = \sum_{y=0}^{\infty}\left(\frac{1}{y+1}\right)\frac{e^{-\lambda}\lambda^y}{y!} = \lambda^{-1}\sum_{y=0}^{\infty}\frac{e^{-\lambda}\lambda^{y+1}}{(y+1)!}$$
$$= \lambda^{-1}\sum_{y=1}^{\infty}\frac{e^{-\lambda}\lambda^y}{y!} = \lambda^{-1}(1-e^{-\lambda}).$$

Problem 4 Let $Y_1, \ldots, Y_n$ be i.i.d. with density

$$f_\phi(y) = \phi^{-1} \exp(-y/\phi) I(y \geq 0),$$

where $\phi > 0$ and $I(A)$ is the indicator function equal to 1 if A is true and equal to 0 otherwise. Give the maximum likelihood estimate for ϕ.

Solution. Note that this is the exponential density with scale paramater ϕ. The likelihood is given by

$$L(\phi; \boldsymbol{y}) = \prod_{i=1}^{n} f_\phi(y_i) = \phi^{-n} \exp(n\bar{y}/\phi) I(y_{(1)} \geq 0),$$

where $y_{(1)} \equiv \min\{y_1, \ldots, y_n\}$ and $\bar{y} \equiv (1/n) \sum_1^n y_i$. Given that $y_{(1)} \geq 0$, the log-likelihood is given by

$$\ell(\phi; \boldsymbol{y}) = -n \log(\phi) - n\bar{y}/\phi.$$

Taking the derivative of $\ell(\cdot; \boldsymbol{y})$ with respect to ϕ and putting it equal to zero gives,

$$\frac{\partial}{\partial \phi} \ell(\phi; \boldsymbol{y}) = -\frac{n}{\phi} + \frac{n\bar{y}}{\phi^2} = 0,$$

with a solution $\hat{\phi} = \bar{y}$. By looking at the second derivative we see that this is indeed a maximum, since

$$\frac{\partial^2}{\partial^2 \phi} \ell(\phi; \boldsymbol{y})|_{\phi = \hat{\phi}} = -\frac{n}{\hat{\phi}^2} < 0.$$

Hence, $\hat{\phi} = \bar{y}$ is the MLE.