Stat 512
Homework key 7

December 4, 2015

REGULAR PROBLEMS

1. Let $\hat{\theta}_1, \ldots, \hat{\theta}_n$ be independent estimators of $\theta$with sampling distributions $\hat{\theta}_k \sim (\theta, V_k)$. Consider linear combination of these estimators

$$\hat{\theta} = \sum_{k=1}^{p} \omega_k \hat{\theta}_k.$$

   (a) Find the mean and variance of $\hat{\theta}$.

   (b) Specify necessary and sufficient conditions on $\overrightarrow{\omega} = (\omega_1, \ldots, \omega_p)$ such that $\hat{\theta}$ is unbiased for θ.

   (c) Find the optimal values of $\overrightarrow{\omega}$ such that $\hat{\theta}$ is the "best linear unbiased estimator" in that it is unbiased and has lower variance than any other unbiased estimator.

Ans:

(a)

$$E\hat{\theta} = \theta \sum_{k=1}^{p} \omega_k.$$

$$Var(\hat{\theta}) = \sum_{k=1}^{p} \omega_k^2 V_k.$$

(b)

$$E\hat{\theta} = \theta.$$

$$\Leftrightarrow \theta \sum_{k=1}^{p} \omega_k = \theta$$

$$\Leftrightarrow \sum_{k=1}^{p} \omega_k = 1.$$

Hence a necessary and sufficient condition for the unbiasedness of θ is that $\sum_{k=1}^{p} \omega_k = 1$.

(c) We need to minimize $Var(\hat{\theta})$ w.r.t. $\overrightarrow{\omega}$when $\sum_{k=1}^{p} \omega_k = 1$. Since we have a constrained optimization problem, we will be using Lagrange Multiplier method. Hence we will solve for $\overrightarrow{\omega}$ setting the partial derivatives of the following function (with respect to ω_k-s and the Lagrange multiplier λ) to 0.

$$f(\overrightarrow{\omega}) = \sum_{k=1}^{p} \omega_k^2 V_k - \lambda(\sum_{k=1}^{p} \omega_k - 1)$$

d.w.r.t.ω_k we get

$$0 = 2\omega_k - \lambda \Rightarrow \omega_k = \frac{\lambda}{2V_k} \tag{1}$$

d.w.r.t. λ we get

$$0 = \sum_{k=1}^{p} \omega_k - 1$$

$$\Rightarrow \sum_{k=1}^{p} \frac{\lambda}{2V_k} = 1$$

$$\Rightarrow \lambda = \frac{2}{\sum_{k=1}^{p} 1/V_k}$$

From (1) we get that for all k, ω_k will have the following expression:

$$\omega_k = \frac{1/V_k}{\sum_{k=1}^{p} 1/V_k}$$

2. Let $Y_1, \dots, Y_n$ be independent random variables having distribution $Y_i \sim (\mu_i, \sigma^2)$ with $\mu_i = \beta_0 + \beta_1 x_i$ for known constants $x_1, \dots, x_n$ satisfying $\exists i, j$ s.t. $x_i \neq x_j$. (This is of course the setting of simple linear regression.) One intuitive class of estimators that has been extensively studied is the "least squares estimators" (LSE) such that $\hat{\vec{\mu}}$ minimizes the "sum of squares"

$$\hat{\vec{\mu}} = \arg\min \sum_{i=1}^{n} (Y_i - \hat{\mu}_i)^2,$$

which in turn leads to LSE $\hat{\vec{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)$ minimizing

$$\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2.$$

(a) Find a closed form expression for slope LSE $\hat{\beta}_1$.

Ans: Define $L \equiv \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2$. To find a minimum of L, we need to find the critical point where the gradient of L satisfies $\nabla L = \vec{0}$. This implies that

$$\frac{\partial L}{\partial \beta_0} = \sum_{i=1}^{n} 2(Y_i - \beta_0 - \beta_1 x_i)(-1) = -2(n\bar{Y} - n\beta_0 - n\bar{x}\beta_1) \equiv 0,$$

so that $\beta_0 = \bar{Y} - \beta_1 \bar{x}$. Then, we find the remaining partial derivative and set it to zero,

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} 2(Y_i - \beta_0 - \beta_1 x_i)(-x_i) = -\sum_{i=1}^{n} x_i Y_i + n\bar{x}\beta_0 + \beta_1 \sum_{i=1}^{n} x_i^2 \equiv 0,$$

so that we find

$$\beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i Y_i - n\bar{x}\beta_0 = \sum_{i=1}^{n} x_i Y_i - n\bar{x}(\bar{y} - \beta_1 \bar{x}) = \sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y}) + n\bar{x}^2 \beta_1.$$

Rearranging terms to solve for β_1, we arrive at the solution

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

(b) Find the mean and variance of the sampling distribution for $\hat{\beta}_1$.

Ans: We note that we can write our slope estimate as

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

Then, we have that

$$E[\hat{\beta}_1] = E\left[\frac{\sum_{i=1}^{n} (x_i - \bar{x}) Y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right] = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) E[Y_i]}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})[\beta_0 + \beta_1 x_i]}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \beta_1.$$

For the variance, we have

$$Var(\hat{\beta}_1) = Var\left(\frac{\sum_{i=1}^{n} (x_i - \bar{x}) Y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 Var(Y_i)}{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right]^2} = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

(c) Find a closed form expression for intercept LSE $\hat{\beta}_0$.

Ans: From (a), we found that

$$\beta_0 = \bar{Y} - \beta_1 \bar{x}.$$

Substituting our solution for $\hat{\beta}_1$,

$$\hat{\beta}_0 = \bar{Y} - \frac{\sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \bar{x}.$$

(d) Find the mean and variance of the sampling distribution for $\hat{\beta}_0$.

Ans: Using the linearity of expectation and knowledge of the mean from part (b),

$$E[\hat{\beta}_0] = E\left[\bar{Y} - \hat{\beta}_1\bar{x}\right] = \beta_0 + \beta_1\bar{x} - \beta_1\bar{x} = \beta_0.$$

For the variance, we first note that $\bar{Y} = \vec{a}_n^T\vec{Y}_n = n^{-1}\vec{1}_n^T\vec{Y}_n$ and $\hat{\beta}_1 = \vec{w}_n^T\vec{Y}_n$, with $w_{n,i} = (x_i - \bar{x})/\sum(x_j - \bar{x})^2$. Since $\vec{a}$ is orthogonal to $\vec{w}$, we know that $Cov(\vec{a}_n^T\vec{Y}_n, \vec{w}_n^T\vec{Y}_n) = Cov(\bar{Y}, \hat{\beta}_1) = 0$. Then, using our variance found in part (b) and knowledge of the necessary covariance,

$$Var(\hat{\beta}_0) = Var\left(\bar{Y} - \bar{x}\hat{\beta}_1\right) = Var(\bar{Y}) + Var(\bar{x}\hat{\beta}_1) + 0 = \frac{\sigma^2}{n} + \bar{x}^2\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \sigma^2\frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i - \bar{x})^2}.$$

(e) Find the covariance of $\hat{\beta}_0$ and $\hat{\beta}_1$.

Ans: Using linearity properties of the covariance,

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = Cov(\bar{Y} - \bar{x}\hat{\beta}_1, \hat{\beta}_1) = Cov(\bar{Y}, \hat{\beta}_1) - \bar{x}Var(\hat{\beta}_1) = -\bar{x}\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

2. Let $Y_1, \ldots, Y_n$ be independent random variables having distribution $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$ with $\mu_i = \beta_0 + \beta_1 x_i$ for known constants $x_1, \ldots, x_n$ satisfying $\exists i, j$ s.t. $x_i \neq x_j$. (This is of course the setting of simple linear regression.) Another class of estimators that has been extensively studied is the "maximum likelihood estimator" (MLE) of regression parameter $\overrightarrow{\beta}$ found by maximizing the likelihood function

$$L(\overrightarrow{\beta}|\overrightarrow{Y}) = f_{\overrightarrow{Y}}(\overrightarrow{Y}|\overrightarrow{\beta}),$$

(that is, the likelihood is just the joint density of the data evaluated at the observed data and regarded as a function of the unknown parameter $\overrightarrow{\beta}$). with MLE $\overrightarrow{\beta}$ defined by

$$\hat{\overrightarrow{\beta}} = \text{argmax} L(\overrightarrow{\beta}|\overrightarrow{Y})$$

Note that when the support of $\overrightarrow{Y}$ is independent of $\overrightarrow{\beta}$, we most often find the MLE by considering the maximization of the log likelihood

$$\mathcal{L}(\overrightarrow{\beta}|\overrightarrow{Y}) = \log\left(f_{\overrightarrow{Y}}(\overrightarrow{Y}|\overrightarrow{\beta})\right)$$

.

(a) What is the likelihood function $L(\overrightarrow{\beta}|\overrightarrow{Y})$ for this problem?

Ans:

$$L(\overrightarrow{\beta}|\overrightarrow{Y}) = f_{\overrightarrow{Y}}(\overrightarrow{Y}|\overrightarrow{\beta}) = \prod_{i=1}^n \frac{e^{-\dfrac{(y_i - \mu_i)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}$$

$$= \frac{e^{-\dfrac{\sum_{i=1}^n (y_i - \mu_i)^2}{2\sigma^2}}}{(\sqrt{2\pi}\sigma)^n}$$

(b) What is the log-likelihood function $\mathcal{L}(\overrightarrow{\beta}|\overrightarrow{Y})$ for this problem?

Ans:

$$\mathcal{L}(\overrightarrow{\beta}|\overrightarrow{Y}) = \log(L(\overrightarrow{\beta}|\overrightarrow{Y})) = -\frac{\sum_{i=1}^n (y_i - \mu_i)^2}{2\sigma^2} - n\log(\sqrt{2\pi}\sigma).$$

(c) Find a closed form expression for slope MLE $\hat{\beta}_1$. We have

$$\mathcal{L}(\overrightarrow{\beta}|\overrightarrow{Y}) = -\frac{\sum_{i=1}^n (y_i - \mu_i)^2}{2\sigma^2} - n\log(\sqrt{2\pi}\sigma).$$

$\hat{\beta}_1, \hat{\beta}_0$ maximizes the above function. Hence $\dfrac{\partial}{\partial\overrightarrow{\beta}}\mathcal{L}(\overrightarrow{\beta}|\overrightarrow{Y})\Big|_{\hat{\overrightarrow{\beta}}} = 0.$

d.w.r.t. $\hat{\beta}_1$ and setting the derivative equal to 0 we get,

$$0 = \frac{\partial}{\partial\beta_1}\mathcal{L}(\overrightarrow{\beta}|\overrightarrow{Y})\Big|_{\hat{\beta}_1} = \frac{\sum_{i=1}^n x_i(y_i - \hat{\beta}_1 x_i - \hat{\beta}_0)}{2\sigma^2}.$$

This leads to

$$\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 + n\hat{\beta}_0\bar{x} - \sum_{i=1}^{n} x_i y_i = 0 \tag{2}$$

d.w.r.t. $\hat{\beta}_0$ and setting the derivative equal to 0 we get,

$$0 = \frac{\partial}{\partial \beta_2}\mathcal{L}(\overrightarrow{\beta}|\overrightarrow{Y})\Big|_{\hat{\overrightarrow{\beta}}} = \frac{\sum_{i=1}^{n}(y_i - \hat{\beta}_1 x_i - \hat{\beta}_0)}{2\sigma^2}.$$

This leads to

$$\hat{\beta}_0 + \hat{\beta}_1\bar{x} - \bar{y} = 0 \tag{3}$$

(2c) and (3) are known as normal equations. From (3) we also get that $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$. Replacing the value of $\hat{\beta}_0$ in we get that

$$\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 + n(\bar{y} - \hat{\beta}_1\bar{x})\bar{x} - \sum_{i=1}^{n} x_i y_i = 0$$
$$\Rightarrow \hat{\beta}_1 \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} y_i(x_i - \bar{x})$$

since

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \tag{4}$$

and

$$\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = \sum_{i=1}^{n} y_i(x_i - \bar{x}) \tag{5}$$

Therefore we get

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

We observe that this $\hat{\beta}_1$is the same $\hat{\beta}_1$we got in question 2(a).

(d) Find the sampling distribution for $\hat{\beta}_1$.

Ans: Since $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sum_{i=1}^{n} a_i y_i$, $\hat{\beta}_1$is a linear combination of normal random variable Y_i-s, it is also normal. We find its mean and variance from question 2(b).

$$\hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

(e) Find a closed form expression for slope MLE $\hat{\beta}_0$.

Ans: From (3) we get that $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = \bar{y} - \frac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\bar{x}$. This expression is same as that of question 2(c).

(f) Find the sampling distribution for $\hat{\beta}_0$.

Ans:

$$\hat{\beta}_0 = \sum_{i=1}^{n}\frac{1}{n}y_i - \frac{\sum_{i=1}^{n}\bar{x}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$
$$\Rightarrow \hat{\beta}_0 = \sum_{i=1}^{n} b_i y_i.$$

Hence, $\hat{\beta}_0$ also is a linear combination of normal random variable Y_i-s, therefore it is also normal. We find its mean and variance from question 2(d).

$$\hat{\beta}_0 \sim \mathcal{N}\left(\beta_0, \frac{\sigma^2\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

(g) Find the covariance of $\hat{\beta}_1$and $\hat{\beta}_0$.
Ans: The covariance will be same as the covariance we found in question 2 (e). Hence,

$$Cov(\hat{\beta}_1, \hat{\beta}_0) = \frac{-\bar{x}\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

(h) Find the joint sampling distribution for $\overrightarrow{\hat{\beta}}$.
Ans: From question 3(d) and 3(f), we see that $\hat{\beta}_1 = \sum_{i=1}^n a_i y_i$ and $\hat{\beta}_0 = \sum_{i=1}^n b_i y_i$. Therefore,

$$\overrightarrow{\beta} = \begin{bmatrix} a_1, \ldots, a_n \\ b_1, \ldots, b_n \end{bmatrix} \overrightarrow{Y} = A\overrightarrow{Y}.$$

Again, $\overrightarrow{Y} \sim MVN(\overrightarrow{\mu}, \sigma^2 I)$. Therefore $A\overrightarrow{Y}$ is also is a multivariate normal with mean $A\overrightarrow{\mu}$ and variance-covariance matrix $\sigma^2 AA^T$. However we have already deduced the mean and variance covariance matrix in the previous questions. Therefore,

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_0 \end{bmatrix} \sim BVN\left(\begin{bmatrix} \beta_1 \\ \beta_0 \end{bmatrix}, \begin{bmatrix} \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} & \frac{-\bar{x}\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \frac{-\bar{x}\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} & \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \end{bmatrix} \right)$$

4. Let $Y_1, \ldots, Y_n$ be independent, identically distributed random variables having log normal distribution $Y_i \sim \mathcal{LN}(\mu, \sigma^2)$ with σ^2 known.

(a) Find a method of moments estimators (MME) for μ and provide its asymptotic sampling distribution.

Ans: Recall that for the lognormal distribution that

$$E[Y_i] = e^{\mu + \frac{1}{2}\sigma^2}, \quad Var[Y_i] = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}.$$

We now define the mapping $g(t) = \log t - \frac{1}{2}\sigma^2$ with $g'(t) = 1/t$ for $t > 0$. A natural MME for μ is then defined by setting

$$\bar{Y}_n = \exp(\hat{\mu}_n + \frac{1}{2}\sigma^2)$$

and then solving to find that

$$\hat{\mu}_n = g(\bar{Y}_n) = \log \bar{Y}_n - \frac{1}{2}\sigma^2.$$

From the central limit theorem, we know that

$$\sqrt{n}(\bar{Y}_n - e^{\mu + \frac{1}{2}\sigma^2}) \to_d \mathcal{N}(0, (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}).$$

Applying the delta method to the MME defined by $g(\cdot)$,

$$\sqrt{n}(g(\bar{Y}_n) - g(e^{\mu + \frac{1}{2}\sigma^2})) \to_d \mathcal{N}(0, g'(e^{\mu + \frac{1}{2}\sigma^2})^2 (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}).$$

This limit expression simplifies to the desired asymptotic sampling distribution for this MME,

$$\sqrt{n}(\hat{\mu}_n - \mu) \to_d \mathcal{N}(0, e^{\sigma^2} - 1).$$

(b) Derive a maximum likelihood estimate for μ and provide its asymptotic sampling distribution.

Ans: We first write the likelihood for μ in the sample,

$$\begin{aligned} L(\mu | \vec{Y}_n) &= \prod_{i=1}^n f(Y_i) \\ &= \prod_{i=1}^n \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp\left(\frac{-(\log Y_i - \mu)^2}{2\sigma^2} \right) \\ &\propto \exp\left(-\sum_{i=1}^n \frac{(\log Y_i - \mu)^2}{2\sigma^2} \right). \end{aligned}$$

We then take the log-likelihood, denoted by ℓ, and differentiate with respect to μ to find the maximizing value

$$\begin{aligned}
\ell(\mu) &= -\sum_{i=1}^{n} \frac{(\log Y_i - \mu)^2}{2\sigma^2}; \\
\dot{\ell}(\mu) &= -\frac{1}{\sigma^2} \sum_{i=1}^{n} 2(\log Y_i - \mu)(-1) \\
&= \frac{1}{\sigma^2} \sum_{i=1}^{n} (\log Y_i - \mu) \equiv 0; \\
\Rightarrow \hat{\mu}_n^{MLE} &= \frac{1}{n} \sum_{i=1}^{n} \log Y_i.
\end{aligned}$$

Recall that $X \equiv \log Y \sim \mathcal{N}(\mu, \sigma^2)$ by the definition of the lognormal distribution. Hence,

$$\hat{\mu}_n^{MLE} = \bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$$

and the asymptotic sampling distribution is found simply to be

$$\sqrt{n}(\hat{\mu}_n^{MLE} - \mu) \to_d \mathcal{N}(0, \sigma^2).$$

(c) Derive which of the two above estimators is more precise.

Ans: Comparing the asymptotic variances, denoted V_{MME} and V_{MLE}, we have that

$$V_{MME} = e^{\sigma^2} - 1 \geq \sigma^2 = V_{MLE}$$

for any value of the known σ^2. (This inequality follows by considering the "racetrack principle" from calculus for functions $f(x) = e^x - 1$ and $g(x) = x$. At $x = 0$, $f(0) = e^0 - 1 = 0 = g(0)$. For all $x > 0$, $f'(x) = e^x \geq 1 = g'(x)$, hence $f(x) > g(x)$ for all $x > 0$.)

5. Let $Y_1, \ldots, Y_n$ be independent, identically distributed random variables having normal distribution $Y_i \sim \mathcal{N}(\mu, \sigma^2)$ with σ^2 known. Let $\theta = Pr(Y_i \leq 0)$ be a target of inference.

(a) Derive a maximum likelihood estimate for μ and provide its asymptotic sampling distribution.

Ans: It suffices to consider the MLE for β_0 in the linear normal model in problem 3, letting all the $x_i = 0$ and $\beta_1 = 0$. Then,

$$\hat{\mu}_n^{MLE} = \bar{Y}_n.$$

The (asymptotic) sampling distribution of $\hat{\mu}_n^{MLE}$ is

$$\sqrt{n}(\hat{\mu}_n^{MLE} - \mu) \to_d \mathcal{N}(0, \sigma^2).$$

(b) Derive a maximum likelihood estimate for θ and provide its asymptotic distribution.

Ans: We can write θ in terms of the parameters of the original distribution as follows:

$$\theta = Pr(Y_i \leq 0) = F_Y(0) = \Phi\left(\frac{-\mu}{\sigma}\right).$$

Invariance properties of the MLE found in section 7.2 of Casella and Berger's *Statistical Inference, 2nd ed.* lead immediately to the MLE of θ as

$$\hat{\theta}_n^{MLE} = \Phi\left(\frac{-\hat{\mu}_n^{MLE}}{\sigma}\right).$$

Applying the delta method with $g(t) = \Phi\left(\frac{-t}{\sigma}\right)$ and $g'(t) = \frac{-1}{\sigma}\phi\left(\frac{-t}{\sigma}\right)$, we find the asymptotic sampling distribution of $\hat{\theta}_n^{MLE}$ to be

$$\sqrt{n}(\hat{\theta}_n^{MLE} - \theta) \to_d \mathcal{N}\left(0, \phi\left(\frac{-\mu}{\sigma}\right)^2\right).$$

(c) Let $W_i = \mathbf{1}_{(-\infty,0]}(Y_i)$ be an indicator that Y_i is less than or equal to 0, and derive an estimator of θ based on the W_i's and provide its asymptotic sampling distribution.

Ans: We have that each $W_i \sim \mathcal{B}(1, \theta)$, with mass function $f(w) = \theta^w(1-\theta)^{1-w}$ for $w = 0, 1$. The likelihood is then

$$L(\theta) = \prod_{i=1}^n f(w_i) = \prod_{i=1}^n \theta^{w_i}(1-\theta)^{1-w_i} = \theta^{\sum_{i=1}^n w_i}(1-\theta)^{n-\sum_{i=1}^n w_i}.$$

The log-likelihood is just

$$\ell(\theta) = \log\theta \sum_{i=1}^n w_i + \log(1-\theta)(n - \sum_{i=1}^n w_i).$$

Then taking derivatives and solving the score equation,

$$\dot{\ell}(\theta) = \frac{1}{\theta} n\bar{W}_n - \frac{1}{1-\theta} n(1-\bar{W}_n) \equiv 0,$$

we find that the MLE of θ based on the W_i is $\tilde{\theta}_n^{MLE} = \bar{W}_n$. Its asymptotic sampling distribution follows simply from the central limit theorem:

$$\sqrt{n}(\bar{W}_n - \theta) \to_d \mathcal{N}(0, \theta(1-\theta)).$$

(d) Derive which of the two estimators above is more precise.

Ans: Recalling that

$$\theta = Pr(Y_i \leq 0) = F_Y(0) = \Phi\left(\frac{-\mu}{\sigma}\right),$$

we have the asymptotic variances of the MLE for Y and V to be, respectively,

$$V_Y = \phi\left(\frac{-\mu}{\sigma}\right)^2, \quad V_W = \Phi\left(\frac{-\mu}{\sigma}\right)\left(1 - \Phi\left(\frac{-\mu}{\sigma}\right)\right).$$

Intuition suggests that $V_Y \leq V_W$ since the dichotomization of our sample when defining $W_1, \ldots, W_n$ should result in a loss of information when we know the true distribution of Y_i is Gaussian. Hence, we are interested in showing

$$g(x) = \Phi(x)[1 - \Phi(x)] - \phi^2(x) \geq 0$$

for all $x \in (-\infty, \infty)$. Note that $g(x)$ is symmetric about $x = 0$, so we can restrict attention to $x < 0$. We also note that

$$\lim_{x \to -\infty} g(x) = 0, \quad g(0) = .25 - 1/(2\pi) = .1487.$$

So if $g(x)$ is strictly increasing from $-\infty$ to 0, we know $g(x)$ is positive everywhere. Taking the derivative, we find

$$\begin{aligned} g'(x) &= \phi(x)[1 - \Phi(x)] - \phi(x)\Phi(x) + 2x\phi^2(x) \\ &= \phi(x)[1 - 2\Phi(x) + 2x\phi(x)]. \end{aligned}$$

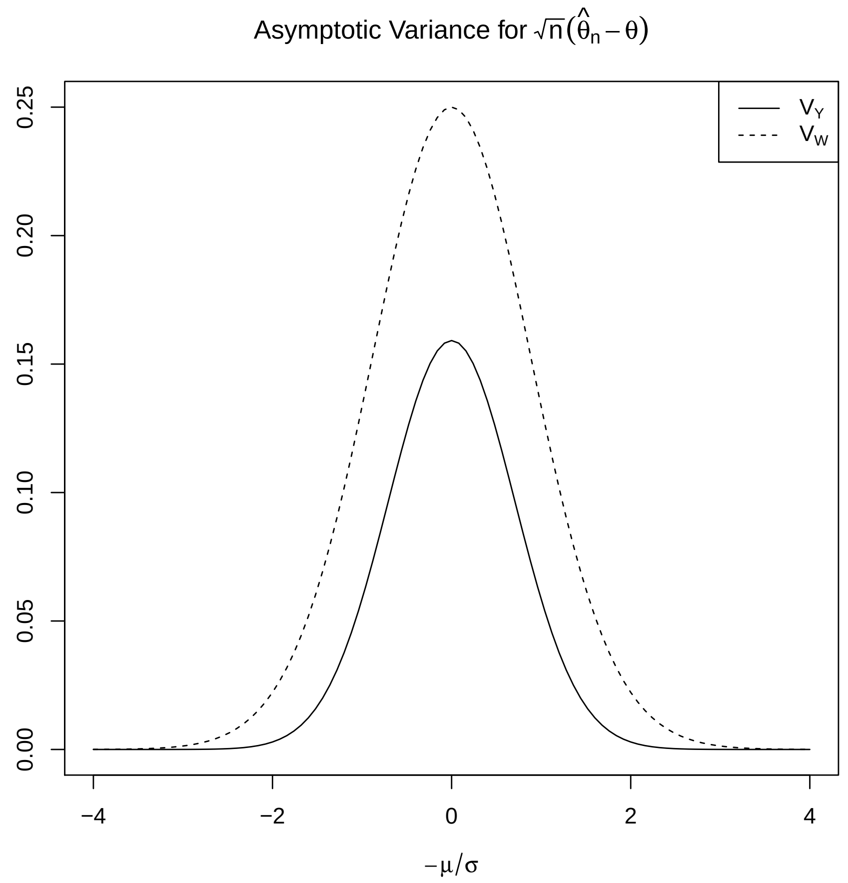
Because $\phi(x) > 0$, $g'(x)$ will be positive over $(-\infty, 0)$ if and only if

$$h(x) = 1 - 2\Phi(x) + 2x\phi(x) > 0, \quad \forall x \in (-\infty, 0).$$

But $h(-\infty) = 1$ and $h(0) = 0$, so if $h(x)$ is strictly decreasing over $(-\infty, 0)$ then we know $h(x)$ is positive. Taking the derivative

$$h'(x) = -2\phi(x) + 2\phi(x) - 2x^2\phi(x) = -2x^2\phi(x) < 0, \quad \forall x \in (-\infty, 0).$$

Hence, we have shown that $V_Y \leq V_W$, as suspected. We plot the two variances below. This plot (or one for $V_Y - V_W$, V_Y/V_W, etc.) was sufficient for partial credit, though an analytic solution like the one above was needed for full points. In practice, we may encounter scenarios when the graphical or numeric solution is the best we can do.


Asymptotic Variance for $\sqrt{n}(\hat{\theta}_n - \theta)$
V_Y
V_W
0.00
0.05
0.10
0.15
0.20
0.25
−4
−2
0
2
4
$-\mu/\sigma$

MORE INVOLVED PROBLEMS

1. Let Y1, . . . , Yn be independent random variables having distribution $Y_i \sim (\mu_i, \sigma^2)$ with $\mu_i = \beta_0 + \beta_1 x_i$ for known constants $x_1, \ldots, x_n$ satisfying $x_i \neq x_j$ for $i \neq j$. For $i \neq j$ define

$$\hat{\theta}_{ij} = \frac{(Y_i - Y_j)}{x_i - x_j}$$

(a) Find the mean and variance of the sampling distribution of $\hat{\theta}_{ij}$.

(b) Using the results of problem 1, find the values of a_{ij} such that

$$\hat{\theta} = \sum_{i=1}^{n} \sum_{j \neq i} a_{ij} \hat{\theta}_{ij}.$$

is the best linear unbiased estimator of $\theta = \beta_1$.

(c) Explicitly show that $\hat{\theta}$ is exactly the least squares estimate for β_1 as derived in problem 2 above.

Ans:

(a)

$$E\hat{\theta}_{ij} = \frac{\mu_i - \mu_j}{x_i - x_j} = \beta_1 \frac{x_i - x_j}{x_i - x_j} = \beta_1.$$

$$V(\hat{\theta}_{ij}) = \frac{V(Y_i) + V(Y_j)}{(x_i - x_j)^2} = \frac{2\sigma^2}{(x_i - x_j)^2}$$

(b) First we will express $\hat{\theta}$as a linear combination of y_i-s.

$$\hat{\theta} = \sum_{i=1}^{n} \sum_{j \neq i} a_{ij} \frac{Y_i - Y_j}{x_i - x_j}$$

$$= \sum_{i=1}^{n} \sum_{j \neq i} \frac{Y_i a_{ij}}{x_i - x_j} - \sum_{i=1}^{n} \sum_{j \neq i} \frac{Y_j a_{ij}}{x_i - x_j}$$

Now we interchange the sums w.r.t. i and j in the second term to get $\hat{\theta}$

$$= \sum_{i=1}^{n} \sum_{j \neq i} \frac{Y_i a_{ij}}{x_i - x_j} - \sum_{j=1}^{n} \sum_{i \neq j} \frac{Y_j a_{ij}}{x_i - x_j}$$

Now in the second term we interchange the notations of i and j. Therefore we get

$$\hat{\theta} = \sum_{i=1}^{n} \sum_{j \neq i} \frac{Y_i a_{ij}}{x_i - x_j} - \sum_{i=1}^{n} \sum_{j \neq i} \frac{Y_i a_{ji}}{x_j - x_i}$$

$$= \sum_{i=1}^{n} \sum_{i \neq j} \frac{Y_i(a_{ij} + a_{ji})}{x_i - x_j}$$

$$= \sum_{i=1}^{n} \sum_{j \neq i} \frac{Y_i(a_{ij} + a_{ji})}{x_i - x_j}$$

Therefore we see that the individual values of a_{ij} do not effect the value of $\hat{\theta}$. However, $\hat{\theta}$ is going to be dependent on $a_{ij} + a_{ji}$. We can atmost get an optimal value for $a_{ij} + a_{ji}$. Any values of a_{ij}, a_{ji} will be optimal as long as $a_{ij} + a_{ji}$ equals the optimal value. Let us take $a_{ij} = a_{ji}$. Define $\omega_i = \sum_{j \neq i} \dfrac{a_{ij}}{x_i - x_j}$. Therefore

$$\hat{\theta} = 2 \sum_{i=1}^{n} \omega_i Y_i.$$

For $\hat{\theta}$ to be unbiased we need

$$E\hat{\theta} = 2\sum_{i=1}^{n} \omega_i \mu_i = 2\sum_{i=1}^{n} \omega_i(\beta_0 + \beta_1 x_i) = 2\beta_0 \sum_{i=1}^{n} \omega_i + 2\beta_1 \sum_{i=1}^{n} \omega_i x_i = \beta_1.$$

Since the above holds for all β_0 and β_1, we must have

$$\sum_{i=1}^{n} \omega_i = 0 \tag{6}$$

$$\sum_{i=1}^{n} \omega_i x_i = \frac{1}{2} \tag{7}$$

$$Var(\hat{\theta}) = 4\sum_{i=1}^{n} \omega_i^2 \sigma^2 = 4\sigma^2 \sum_{i=1}^{n} \omega_i^2$$

We have to minimize $Var(\hat{\theta})$ with respect to the constraints in (6) and (7). We will minimize

$$f(\overrightarrow{\omega}) = \sum_{i=1}^{n} \omega_i^2 - \lambda_1(\sum_{i=1}^{n} \omega_i x_i - \frac{1}{2}) - \lambda_2 \sum_{i=1}^{n} \omega_i.$$

D.w.r.t. ω_i

$$2\omega_i - \lambda_1 x_i - \lambda_2 = 0 \ \forall i. \tag{8}$$

D.w.r.t. λ_1,

$$\sum_{i=1}^{n} x_i \omega_i = \frac{1}{2}. \tag{9}$$

D.w.r.t. λ_2,

$$\sum_{i=1}^{n} \omega_i = 0. \tag{10}$$

Summing the n equations in (8) and using (10),

$$2\sum_{i=1}^{n} \omega_i - \lambda_1 \sum_{i=1}^{n} x_i - n\lambda_2 = 0$$
$$\lambda_2 = -\lambda_1 \bar{x}$$

Replacing $\lambda_2 = -\lambda_1 \bar{x}$. Taking product of (8) with x_i and summing them up for all i we get

$$2\sum_{i=1}^{n} \omega_i x_i - \lambda_1 \sum_{i=1}^{n} x_i^2 - \lambda_2 \sum_{i=1}^{n} x_i = 0.$$

Since $\sum_{i=1}^{n} \omega_i x_i = \frac{1}{2}$,

$$1 - \lambda_1 \sum_{i=1}^{n} x_i^2 + n\lambda_1 \bar{x}\bar{x} = 0,$$
$$1 - \lambda_1(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2) = 0,$$
$$\lambda_1 = \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$
$$\lambda_2 = -\frac{\bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

since $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$. Replacing the value of λ_1 and λ_2 in (8),

$$2\omega_i - \frac{x_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} + \frac{\bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = 0$$
$$\omega_i = \frac{1}{2}\frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

We got the optimal values of ω_i. So now we can guess some values of a_{ij} which will satisfy

$$\sum_{j \neq i} \frac{a_{ij}}{x_i - x_j} = \omega_i = \frac{1}{2} \frac{x_i - \bar{x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{11}$$

We make a guess that $a_{ij} = \dfrac{(x_i - x_j)^2}{2n \sum_{i=1}^{n} (x_i - \bar{x})^2}$. We have to show that (11) holds. Now,

$$\begin{aligned}
\sum_{j \neq i} \frac{a_{ij}}{x_i - x_j} &= \sum_{j \neq i} \frac{x_i - x_j}{2n \sum_{i=1}^{n} (x_i - \bar{x})^2} \\
&= \frac{\sum_{j \neq i} (x_i - x_j)}{2n \sum_{i=1}^{n} (x_i - \bar{x})^2}
\end{aligned}$$

Notice that $\sum_{j \neq i} (x_i - x_j) = nx_i - x_i - n\bar{x} + x_i = n(x_i - \bar{x})$. Replacing this value in the above expression we get that,

$$\begin{aligned}
&\sum_{j \neq i} \frac{a_{ij}}{x_i - x_j} \\
&= \frac{n(x_i - \bar{x})}{2n \sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{1}{2} \frac{x_i - \bar{x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.
\end{aligned}$$

Therefore (11) is satisfied by this a_{ij}. Hence, $a_{ij} = a_{ji} = \dfrac{(x_i - x_j)^2}{2n \sum_{i=1}^{n} (x_i - \bar{x})^2}$. Also, the BLUE

$$\hat{\theta} = \frac{\sum_{i=1}^{n} \sum_{j \neq i} (x_i - x_j)(y_i - y_j)}{2n \sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)(y_i - y_j)}{2n \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

(c)

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)(y_i - y_j)}{2n \sum_{i=1}^{n} (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^{n} (\sum_{j=1}^{n} x_j y_j - n y_i \bar{x} - n x_i \bar{y} + n x_i y_i)}{2n \sum_{i=1}^{n} (x_i - \bar{x})^2} \\
&= \frac{n \sum_{i=1}^{n} x_i y_i - n^2 \bar{y} \bar{x} - n^2 \bar{x} \bar{y} + n \sum_{i=1}^{n} x_i y_i}{2n \sum_{i=1}^{n} (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^{n} x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
\end{aligned}$$

This is exactly the expression of LS estimator $\hat{\beta}_1$ in problem 2 above.