Midterm

Ec240a – Second Half, Fall 2023

Please read each question carefully. Start each question on a new bluebook page (or sheet of paper). The use of calculators and other computational aides is not allowed. Good luck!

[1] **[5 Points]** Please write your full name on this exam sheet and turn it in with your bluebook.

[2] **[25 Points]** Let $X \in \{0,1,2\}$ and $Y \in \{0,1,2\}$. The probability of the event $X = x$ and $Y = y$ for all possible combinations of x and y is given in the following table:

X\Y	0	1	2
0	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{3}{18}$
1	$\frac{3}{18}$	$\frac{1}{18}$	$\frac{1}{9}$
2	$\frac{1}{9}$	$\frac{3}{18}$	$\frac{1}{18}$

[a] **[5 Points]** Calculate $\mathbb{E}[Y]$ and $\mathbb{E}[Y|X=1]$. Are X and Y independent?

[b] **[5 Points]** Calculate $\mathbb{E}[X]$, $\mathbb{E}\left[X^2\right]$ and hence $\mathbb{V}(X)$.

[c] **[7 Points]** Calculate $\mathbb{C}(X,Y)$ and also the coefficient on X in $\mathbb{E}^*[Y|X]$.

[d] **[3 Points]** Calculate the intercept of $\mathbb{E}^*[Y|X]$.

[e] **[5 Points]** Repeat [a] to [d] above for the following joint distribution

X\Y	0	1	2
0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

[3] **[45 Points]** You have been hired by UNICEF to estimate the prevalence of childhood stunting (low height-for-age) across municipalities in a country where childhood malnutrition is commonplace. Let Y_{it} be the height-for-age Z score of individual $t = 1,\ldots,T$ in municipality $i = 1,\ldots,N$. In each municipality you draw T children at random and compute the average height-for-age Z score

$$\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}.$$

You assume that $Y_{it}|\,\theta_i \sim \mathcal{N}\left(\theta_i, \sigma^2\right)$ for $i = 1,\ldots,N$ and $t = 1,\ldots,T$ (and hence we have that $\bar{Y}_i|\,\theta_i \sim \mathcal{N}\left(\theta_i, \frac{\sigma^2}{T}\right)$.). In this model the expected height-for-age Z score, θ_i, varies across municipalities. Your goal is to estimate the municipality (population) means $\theta_1, \theta_2, \ldots, \theta_N$. Municipalities with low θ_i estimates will be slated to receive new anti-hunger and nutrition programs. Initially you may assume that σ^2 is known (in a healthy population of children $\sigma^2 \approx 1$ since height-for-age Z scores are calibrated to have unit variance in such a setting).

[a] **[5 Points]** Let $\|\mathbf{m}\| = \left[\sum_{i=1}^{N} m_i^2\right]^{1/2}$ denote the Euclidean norm of a vector. Let $\theta = (\theta_1,\ldots,\theta_N)'$. Show that

$$\mathbb{E}\left[\left\|\hat{\theta} - \theta\right\|^2\right] = \sum_{i=1}^{N} \mathbb{E}\left[\left(\hat{\theta}_i - \theta_i\right)^2\right],$$

with $\hat{\theta}$ some estimate – based upon the sample data $\mathbf{Y} = (Y_{11}, \ldots, Y_{1T}, \ldots, Y_{N1}, \ldots, Y_{NT})'$ – of θ. Explain why this measures *expected* estimation accuracy or *risk*? What is being averaged in the expectation? **[2 - 3 sentences].**

[b] **[8 Points]** Consider the following family of estimators for θ_i (for $i = 1, \ldots, N$):

$$\hat{\theta}_i = (1 - \lambda)\, \bar{Y}_i + \lambda\mu,$$

with μ the country-wide mean of Y_{it} (i.e., the expected height-for-age Z score of a randomly sampled child from the full country-wide population). You may assume that μ is known (perhaps from prior research). Assume that $0 \leq \lambda \leq 1$. Interpret this estimator? Why might the estimator with $\lambda = 0$ be sensible? How might you justify the estimator when $\lambda > 0$ **[3 - 4 sentences].**

[c] **[9 Points]** Show, for the family of estimates introduced in part [d], that

$$\mathbb{E}\left[\left\|\hat{\theta} - \theta\right\|^2\right] = (1 - \lambda)^2 \frac{N}{T}\sigma^2 + \lambda^2 \sum_{i=1}^{N} (\theta_i - \mu)^2 .$$

You hear, in the hallways of Evans, that "small λ means small bias" and "big λ means low variance". Explain? **[4 - 5 sentences].**

[d] **[5 Points]** Show that the risk-minimizing choice of λ, say λ^*, is

$$\lambda^* = \frac{N\sigma^2}{N\sigma^2 + \sum_{i=1}^{N} T\,(\theta_i - \mu)^2}.$$

Discuss why λ^* is declining in $\sum_{i=1}^{N} T\,(\theta_i - \mu)^2$. Provide some intuition for why this is optimal. **[3 - 4 sentences].**

[e] **[5 Points]** Show that

$$\sum_{i=1}^{N} \mathbb{E}\left[\left(\bar{Y}_i - \hat{\theta}_i\right)^2\right] = \sum_{i=1}^{N} \mathbb{E}\left[\left(\bar{Y}_i - \theta_i\right)^2\right] + \sum_{i=1}^{N} \mathbb{E}\left[\left(\hat{\theta}_i - \theta_i\right)^2\right] - \frac{2\sigma^2}{T} \mathrm{df}\left(\hat{\theta}\right)$$

with the degree-of-freedom of $\hat{\theta}$ (or *model complexity*) equal to

$$\mathrm{df}\left(\hat{\theta}\right) = \sum_{i=1}^{N} \frac{T}{\sigma^2} \mathbb{C}\left(\bar{Y}_i, \hat{\theta}_i\right).$$

We call the term to the left of the first equality above *apparent error*.

[f] **[5 Points]** Show that

$$\sum_{i=1}^{N} \mathbb{E}\left[\left(\bar{Y}_i - \theta_i\right)^2\right] = \frac{N}{T}\sigma^2$$

and also, for the family of estimates indexed by λ introduced in part [d] above, that

$$\mathrm{df}\left(\hat{\theta}\right) = N\,(1 - \lambda).$$

[g] **[4 Points]** Calculate apparent error and model complexity for $\hat{\theta}$ when $\lambda = 0$. Explain? **[2 - 3**

sentences].

[h] **[4 Points]** Calculate apparent error and model complexity for $\hat{\theta}$ when $\lambda = 1$. Explain? **[2 - 3 sentences].**

[4] **[25 Points]** Consider the following statistical model for the logarithm of daily city-wide sales of Bob Dylan's landmark *Christmas in the Heart* album:

$$\ln S = \alpha_0 + \beta_0 R + \gamma_0 P + U, \ \mathbb{E}\left[U \mid R, P\right] = 0,$$

where R is the number of times a song from the album is played on KALX on the given day, and P is the price of the album (which varies across your sample due to various (exogenous) record label promotions, holiday sales and so on). A friend estimates $\theta_0 = (\alpha_0, \beta_0, \gamma_0)'$ by the method of least squares. She claims that $\sqrt{N}\left(\hat{\theta} - \theta_0\right) \xrightarrow{D} \mathcal{N}\left(0, \Lambda_0\right)$ and reports the following:

$$\hat{\theta} = \begin{pmatrix} 1.0 \\ 0.01 \\ -0.51 \end{pmatrix}, \ \frac{\hat{\Lambda}}{N} = \begin{pmatrix} 0.25 & -0.002 & 0.010 \\ -0.002 & 0.01 & 0.005 \\ 0.010 & 0.005 & 0.03 \end{pmatrix}.$$

[a] **[2 Points]** Calculate a 95 confidence interval for β_0.

[b] **[5 Points]** Your friend would like to test the hypothesis that "for Bob Dylan one song on the radio is as good as cutting record price by \$1" (a phrase used by her record store boss). Explain why this corresponds to:

$$H_0 : \beta_0 = -\gamma_0$$
$$H_1 : \beta_0 \neq -\gamma_0$$

[c] **[5 Points]** We can re-write H_0 as

$$H_0 : C\theta = c$$

Provide the appropriate forms for C and c.

[d] **[5 Points]** How many restrictions on θ does H_0 imposes?

[e] **[5 Points]** Calculate the Wald statistics for H_0. Can we reject with size $\alpha = 0.05$?

[f] **[8 Points]** Now formalize and test the hypothesis that "for Bob Dylan one song on the radio is as good as cutting record price by \$3".