Second Midterm

Ec240a – Second Half, Fall 2016

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 $\mathbb{E}[X]$, $\mathbb{E}[X^2]$ 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] **[35 Points]** Consider the population of married men. Let Y denote log earnings for a generic random draw from this population, X his years of completed schooling and W the schooling of his spouse. Assume that the conditional mean of own log earnings given own and spouse's schooling is

$$\mathbb{E}[Y|X, W] = \alpha_0 + \beta_0 X + \gamma_0 W,$$

while the best linear predictor of spouse's schooling given own schooling is

$$\mathbb{E}^*[W|X] = \delta_0 + \zeta_0 X.$$

You may assume that the joint distribution of (W, X, Y) is such that these objects are well-defined. You may assume that all the slope coefficients in the two equations above are positive.

[a] **[2 Points]** Show that $\zeta_0 = \rho_{WX}\frac{\sigma_W}{\sigma_X}$, with ρ_{WX} the correlation of W with X, and σ_W and σ_X respectively the standard deviation of W and X. Further show that $\delta_0 = \mu_W - \rho_{WX}\frac{\sigma_W}{\sigma_X}\mu_X$ with μ_W and μ_X denoting the population means of W and X.

[b] **[3 Points]** Using your answers in [a] above, as well as the form of $\mathbb{E}[Y|W, X]$, provide an expression for $\mathbb{E}^*[Y|X]$.

[c] **[5 Points]** Consider another population of married men where $F_{Y|W,X}(y|W = w, X = x)$, $F_W(w)$ and $F_X(x)$ coincide with those for the population described above, but where $F_{W,X}(w, x)$ differs. Assume that in this alternative population $\rho_{WX} = 0$. Solve for $\mathbb{E}[Y|X, W]$, $\mathbb{E}^*[W|X]$ and $\mathbb{E}^*[Y|X]$. Use the notation established in parts [a] and [b] to formulate your answer.

[d] **[5 Points]** Assume that $F_W(w)$ and $F_X(x)$ are identical and that marriage is homogamous in terms of education so that $W = X$ for all couples (i.e., individuals choose partners with identical levels of education). Show that in this world $\rho_{WX} = 1$. Solve for $\mathbb{E}[Y|X, W]$, $\mathbb{E}^*[W|X]$ and $\mathbb{E}^*[Y|X]$. Use the notation established in parts [a] and [b] to formulate your answer.

[e] **[5 Points]** Compare the form of $\mathbb{E}^*[Y|X]$ in the original population with that in the two alternative populations of parts [c] and [d]. In which population does log earnings rise most steeply with years of schooling? Provide some intuition for your answer (5 sentences).

[f] **[5 Points]** Assume that schooling is binary valued, taking on the values 0,1. Let R_W be a 2×1 vector equal to $(1, 0)'$ if $W = 0$ and $(0, 1)'$ if $W = 1$. Let S_X be the analogous 2×1 vector defined using X. Let $T_{WX} = (R_W \otimes S_X)$ and

$$\mathbb{E}^*[Y|T_{WX}] = T'_{WX}\pi,$$

where a constant is *not* included and $\pi = (\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})'$. Show that

$$\pi_{jk} = \mathbb{E}[Y|W = j, X = k].$$

[g] **[5 Points]** Consider the null hypothesis that $\mathbb{E}[Y|W, X] = \alpha_0 + \beta_0 X + \gamma_0 W$. Maintaining this null find an explicit expression for each component of π in terms of α_0, β_0 and γ_0. Express this null in the form $C\pi = c$ for some matrix of constants C and vector of constants c.

[h] **[5 Points]** Let $W = 1$ if a wife has completed primary school and zero otherwise,

let $X = 1$ if a husband has completed primary school and zero otherwise. A least squares fit, loosely based on data from Brazil, of log husband's earnings on T_{WX} as defined in [f] using a random sample of size $N = 50,000$ yields point estimate of

$$\hat{\pi} = \begin{pmatrix} 5.50 \\ 6.00 \\ 5.00 \\ 7.00 \end{pmatrix}$$

with an estimated asymptotic variance-covariance matrix of

$$\hat{\Lambda} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 7 & 0 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Can you reject the null hypothesis (at the $\alpha = 0.05$ level) formulated in part [g] on the basis of this sample? For your reference the 0.95 quantiles of χ^2 random variables with parameters 1, 2 and 3 are, respectively, 3.84, 5.99 and 7.81.

[4] **[35 Points]** You observe a simple random sample of size N from the population

$$Y_0 \sim N\left(\mu, \frac{\sigma^2}{4}\right)$$

as well as a second, independent, simple random sample, also of size N, from the population

$$Y_1 \sim N\left(\mu, \sigma^2\right).$$

The value of σ^2 is *known*. Consider the family of estimates of μ

$$\hat{\mu}\left(c_0, c_1\right) = c_0 \bar{Y}_0 + c_1 \bar{Y}_1,$$

where $\bar{Y}_0 = \frac{1}{N} \sum_{i=1}^{N} Y_{0i}$ and $\bar{Y}_1 = \frac{1}{N} \sum_{i=1}^{N} Y_{1i}$.

[a] **[5 Points]** Show that mean squared error equal for a generic member of this family equals

$$\mathbb{E}\left[\left(\hat{\mu}\left(c_0, c_1\right) - \mu\right)^2\right] = \frac{c_0^2 \sigma^2}{4N} + \frac{c_1^2 \sigma^2}{N} + \left(1 - c_0 - c_1\right)^2 \mu^2. \tag{1}$$

[b] **[5 Points]** Show that the risk-minimizing choice of c_0 and c_1 are

$$\begin{pmatrix} c_0^* \\ c_1^* \end{pmatrix} = \frac{1}{\left(\frac{\sigma^2}{4N} + \mu^2\right)\left(\frac{\sigma^2}{4} + \mu^2\right) - \mu^2} \begin{pmatrix} \frac{\sigma^2}{N} \\ \frac{\sigma^2}{4N} \end{pmatrix}$$

[c] **[5 Points]** Show that $\mathbb{E}\left[\bar{Y}_0^2\right] = \frac{\sigma^2}{4N} + \mu^2$ and $\mathbb{E}\left[\bar{Y}_1^2\right] = \frac{\sigma^2}{N} + \mu^2$.

[d] **[5 Points]** Show that

$$\hat{R}\left(c_0, c_1\right) = \frac{c_0^2\sigma^2}{4N} + \frac{c_1^2\sigma^2}{N} + \left(1 - c_0 - c_1\right)^2 \frac{1}{2}\left\{\bar{Y}_0^2 + \bar{Y}_1^2 - \frac{5\sigma^2}{4N}\right\} \tag{2}$$

is an unbiased estimate of (1).

[e] **[5 Points]** Describe in *words* how one might use (2) to construct an implementable estimator. **[4 to 6 sentences]**.

[f] **[5 Points]** Let $0 \leq \lambda \leq 1$ and consider the alternative family of estimators

$$\hat{\mu}\left(\lambda\right) = \lambda\bar{Y}_0 + \left(1 - \lambda\right)\bar{Y}_1.$$

Prove that all estimators in this family are unbiased.

[g] **[5 Points]** Find the risk-minimizing choice of λ. Is the resulting estimator implementable?