ST2132 Tutorial 2 Solution 2210

1. Key result: Given x is observed, if c is used to predict Y, MSE is

$$\mathrm{E}[(Y-c)^2|x] = \mathrm{var}[Y|x] + \{\mathrm{E}[Y|x] - c\}^2$$

In particular, the best predictor is $\mathrm{E}[Y|x]$, and its MSE is $\mathrm{var}[Y|x]$.

Predictor	MSE
(a) $\mathrm{E}[X_2\|1] = \Pr(X_2 = 1\|X_1 = 1) = 0$	$\mathrm{var}[X_2\|1] = 0$
(b) $\mathrm{E}[X_2\|0] = \Pr(X_2 = 1\|X_1 = 0) = 1/3$	$\mathrm{var}[X_2\|0] = 1/3 \times (1 - 1/3) = 2/9$
(c) 0	$\mathrm{var}[X_2\|0] + (1/3 - 0)^2 = 2/9 + 1/9 = 1/3$
(d) $\mathrm{E}(X_1) = 1/4$	$\mathrm{var}[X_2\|1] + (0 - 1/4)^2 = 1/16$
(e) 1/4	$\mathrm{var}[X_2\|0] + (1/3 - 1/4)^2 = 2/9 + 1/144 \approx 0.23$

Notes:

(a) Given $X_1 = 0$, $X_2 = 0$, a constant.

(c) Yes, the outcome is 0 or 1, but the prediction is 1/3. More directly, MSE is $\mathrm{E}[(X_2 - 0)^2|0] = 1/3$.

2. (a) Since $\mathrm{E}(X_1) = K/N$,

$$\mathrm{E}(\mathrm{E}[X_2|X_1]) = \frac{K - K/N}{N-1} = \frac{K}{N}$$

(b)The numerator is

$$K(N-K) - K(1-X_1) - (N-K)X_1 + X_1(1-X_1)$$

Since X_1 is either 0 or 1, $X_1(1-X_1) = 0$, so the formula is obtained.

$$\mathrm{var}[X_2|X_1] = \frac{K(N-K) - K(1-X_1) - (N-K)X_1}{(N-1)^2}$$

(c) Combining $\mathrm{var}(\mathrm{E}[X_2|X_1]) = \mathrm{var}(X_1)/(N-1)^2$ and $\mathrm{var}(X_1) = K(N-K)/N^2$ gives the first formula.

The second formula follows from computing the expectation of (b), and $\mathrm{E}(1 - X_1) = (N-K)/N$.

(d) The sum is

$$\frac{K(N-K)}{(N-1)^2}\frac{1 + N^2 - 2N}{N^2} = \frac{K(N-K)}{N^2}$$

which is $\mathrm{var}(X_2)$, as expected.

3. (a) Integration by parts gives

$$\Gamma(\alpha+1) = \int_0^\infty x^\alpha e^{-x}\mathrm{d}x = \left[-x^\alpha e^{-x}\right]_0^\infty + \alpha\int_0^\infty x^{\alpha-1}e^{-x}\mathrm{d}x$$

Since $\alpha > 0$, $0^\alpha = 0$. Also $x^\alpha e^{-x} \to 0$ as $x \to \infty$. Hence the formula.

(b)

$$\mathrm{E}(X) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty x\,x^{\alpha-1}e^{-\lambda x}\mathrm{d}x = \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)} \times \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)}\int_0^\infty x^{(\alpha+1)-1}e^{-\lambda x}\mathrm{d}x = \frac{\alpha}{\lambda}$$

Similarly

$$\mathrm{E}(X^2) = \frac{\Gamma(\alpha+2)}{\lambda^2\Gamma(\alpha)} \times \frac{\lambda^{\alpha+2}}{\Gamma(\alpha+2)}\int_0^\infty x^{(\alpha+2)-1}e^{-\lambda x}\mathrm{d}x = \frac{(\alpha+1)\alpha}{\lambda^2}$$

$$\mathrm{var}(X) = \frac{(\alpha+1)\alpha}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}$$

4. $Y_n \overset{d}{=} Z_1^2 + \cdots + Z_n^2$, where Z_i's are IID standard normal RVs. $a = \mathrm{E}(Y_n) = n\mathrm{E}(Z_1^2) = n$. $\mathrm{var}(Y_n) = n\mathrm{var}(Z_1^2) = n(\mathrm{E}(Z_1^4) - 1^2) = 2n$. $b = \mathrm{SD}(Y_n) = \sqrt{2n}$.

5. (a) Predictor:

$$\mathrm{E}[Y|x] = \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X)$$

MSE is $\mathrm{var}[Y|x] = (1-\rho^2)\sigma_Y^2$.

(b)

$$\mathrm{E}[Y|X] = \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(X - \mu_X)$$

Since X is normal, $\mathrm{E}[Y|X]$ is also normal, with expectation

$$\mathrm{E}(\mathrm{E}[Y|X]) = \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}\left(E(X) - \mu_X\right) = \mu_Y$$

and variance

$$\mathrm{var}(\mathrm{E}[Y|X]) = \frac{\rho^2\sigma_Y^2}{\sigma_X^2} \times \sigma_X^2 = \rho^2\sigma_Y^2$$

$\mathrm{var}[Y|X]$ is the constant $(1-\rho^2)\sigma_Y^2$, which is also $\mathrm{E}(\mathrm{var}[Y|X])$. Hence, as expected,

$$\sigma_Y^2 = \mathrm{var}(\mathrm{E}[Y|X]) + \mathrm{E}(\mathrm{var}[Y|X])$$

[This is an instance of "analysis of variance": $\mathrm{var}(Y)$ = "variance explained by X" + "residual variance".]

(c) Mean MSE: $\mathrm{E}(\mathrm{var}[Y|X]) = (1-\rho^2)\sigma_Y^2$.

(d) From the variance matrix, $\mathrm{cov}(X, Y) = \rho\sigma_X\sigma_Y$. Hence the correlation is ρ.

(e) Rearrange terms in formula of $\mathrm{E}[Y|x]$.

[The formula expresses "regression to the mean". For example, in the Pearson data set, $\rho \approx 0.5$. Look at fathers who are 2 SDs above average, i.e., $(x - \mu_X)/\sigma_X = 2$. The mean height of their sons is only $(\mathrm{E}[Y|x] - \mu_Y)/\sigma_Y = 1$ SD above average.]