EE162 Discussion 5: Solutions
Jérôme Thai
March 6, 2014

1 Least Squares Estimate

Problem 1 (MMSE). In the LSE, the best guess $\hat{X}$ is a function of $Y$, i.e. of the form $g(Y)$ because the observed value of $Y$ is the only information we have about $X$. If the function $g(\cdot)$ can be arbitrary, it is the Minimum Mean Squares Estimate (MMSE) of $X$ given $Y$. We have:

\[ \hat{X}_{\text{MSE}} = \phi(Y) \quad \text{with} \quad \phi = \arg\min_g E[|X - g(Y)|^2] \]

Prove that: $\hat{X}_{\text{MSE}} = E[X|Y]$

Solution. From (a) of note 2, we have $X - E[X|Y] \perp g(Y) - E[X|Y]$ for any function $g(\cdot)$. Hence

\[ E[|X - g(Y)|^2] = E[(X - E[X|Y]) + (E[X|Y] - g(Y))^2] = E[|X - E[X|Y]|^2] + E[|g(Y) - E[X|Y]|^2] \]

because the inner product of the cross terms are zero (or by application of the Pythagorean theorem). Hence

\[ \min_{g(\cdot)} E[|X - g(Y)|^2] = E[|X - E[X|Y]|^2] + \min_{g(\cdot)} E[|g(Y) - E[X|Y]|^2] \]

which is minimized when $g(Y) = E[X|Y]$.

Problem 2 (LLSE). In the LSE, if we restrict $\hat{X}$ to be a linear function of $Y$, the best guess, denoted by $L[X|Y]$, is said to be the Linear Least Squares Estimate (LLSE). It is:

\[ L[X|Y] = \hat{X}_{\text{LLSE}} = a + bY \quad \text{with} \quad (a, b) = \arg\min_{(c, d)} E[|X - c - dY|^2] \]

Prove that: $L[X|Y] = E[X] + \frac{\text{cov}(X,Y)}{\text{var}(Y)}(Y - E[Y])$

Suppose we don’t know the joint distribution of $(X, Y)$, but we have $N$ observations: $(X_1, Y_1), \ldots, (X_N, Y_N)$, what would be a good estimate for $L[X|Y]$?

Solution. We have the following dependencies in $c$ and $d$:


The optimal values for $c$ and $d$, namely $a$ and $b$ respectively, must set the partial derivatives to zero

\[ 0 = \partial_c E[|X - a - bY|^2] = 2(a - E[X] + bE[Y]) \]
\[ 0 = \partial_d E[|X - a - bY|^2] = 2(bE[Y^2] - E[XY] + 2aE[Y]) \]

Solving for $a$ and $b$ gives:

\[ a + bY = E[X] + \frac{\text{cov}(X,Y)}{\text{var}(Y)}(Y - E[Y]) \]
Another approach consists in observing that

\[ E[X - a - bY] = 0 \quad \text{and} \quad E[(X - a - bY)Y] = 0 \quad \text{i.e.} \quad X - L[X|Y] \perp c + dY \quad \forall c, d \]

This can be seen by plugging in \( X - a - bY = X - E[X] - \frac{\text{cov}(X,Y)}{\text{var}(Y)}(Y - E[Y]) \), or as a direct consequence of the fact that the partial derivatives at \((a, b)\) are zero. The rest of the proof follows similarly as in problem 1 with \( c + dY \) instead of \( g(Y) \) and \( L[X|Y] = a + bY \) instead of \( E[X|Y] \).

2 Jointly Gaussian

Problem 3. Prove (d) of Note 5.

Solution. We note that \( Y - L[Y|Z] \perp Z \), and \( E[Y - L[Y|Z]] = 0 \) hence


It follows that \( Y - L[Y|Z] \) and \( Z \) are uncorrelated. Since they are both linear transformations of the Gaussian r.v. \( X = (Y, Z) \), then \( Y - L[Y|Z] \) and \( Z \) are jointly Gaussian. This implies that they are also independent, hence \( Y - L[Y|Z] \) and \( \phi(Z) \) are independent for all functions \( \phi(\cdot) \). Then \( Y - L[Y|Z] \) and \( \phi(Z) \) are uncorrelated for all functions \( \phi(\cdot) \). Since \( Y - L[Y|Z] \) is an unbiased estimator we have

\[ E[(Y - L[Y|Z])\phi(Z)] = \text{cov}(Y - L[Y|Z], \phi(Z)) + E[Y - L[Y|Z]]E[\phi(Z)] = 0 \]

hence \( Y - L[Y|Z] \perp \phi(Z), \quad \forall \phi(\cdot), \text{ so } L[Y|Z] = E[Y|Z] \).

Problem 4. Let \( X \) be the height of the father, \( Y \) the height of the son, in a sample of father-son pairs. Assume \( X \) and \( Y \) bivariate normal, as found by Karl Pearson around 1900. Assume \( E(X) = 68 \) (inches), \( E(Y) = 69 \), \( \sigma X = \sigma Y = 2 \), \( \rho = .5 \). (We expect to be positive because on the average, the taller the father, the taller the son.) Given \( X = 80 \) (6 feet 8 inches), what is the distribution of \( Y \)?

Solution. \( Y \) is normal with mean

\[ \mu_Y + \frac{\rho \sigma_Y}{\sigma_X}(x\mu_X) = 69 + .5(80 - 68) = 75 \]

which is 6 feet 3 inches. The variance of \( Y \) given \( X = 80 \) is

\[ \sigma_Y^2(1^2) = 4(3/4) = 3. \]

Thus the son will tend to be of above average height, but not as tall as the father. This phenomenon is often called regression, and the line \( y = \mu_Y + \frac{\sigma_Y}{\sigma_X}(x\mu_X) \) is called the regression line.