1 Least Squares Estimate

Note 1 (LSE). We know the joint distribution of the pair of random variables \((X, Y)\) and we want to estimate \(X\) from the observed value of \(Y\). A standard example for the estimation problem is finding a random variable \(\hat{X}\) such that \(E[|X - \hat{X}|^2]\) is minimized:

\[
\min_{\hat{X}} E[|X - \hat{X}|^2] \tag{1}
\]

The best guess \(\hat{X}\) (given by the above minimization problem) is is the Least Squares Estimate (LSE).

Note 2 (Norm in the \(L^2\) space). For square-integrable functions \(g\) of \((X, Y)\), i.e. \(E[|g(X, Y)|^2] < \infty\), we can define:

(a) A scalar product: \(g_1 \cdot g_2 = E[g_1(X, Y)g_2(X, Y)]\)
(b) A norm: \(\|g\|^2 = g \cdot g = E[|g(X, Y)|^2] = \int_{-\infty}^{+\infty} |g(x, y)|^2 f_{X,Y}(x, y) dx dy\)
(c) Orthogonality: \(g_1 \perp g_2 \iff E[g_1(X, Y)g_2(X, Y)] = 0\)

Note 3 (Conditional Expectation). The conditional expectation of \(g(X)\) given \(Y\) is defined by

\[
E[X|Y = y] = \int_{-\infty}^{+\infty} g(x) \frac{f_{X,Y}(x, y)}{f_Y(y)} dx \tag{2}
\]

It satisfies the following orthogonality properties:

(a) \(\forall \phi(\cdot), \quad X - E[X|Y] \perp \phi(Y)\)
(b) \(X - g(Y) \perp \phi(Y), \quad \forall \phi(\cdot)\quad \implies \quad g(Y) = E[X|Y]\)

Sketch of proof: For (a), we have from equation (2)

\[E[E[X|Y]\phi(Y)] = E[X\phi(Y)] \quad \text{hence} \quad E[(X - E[X|Y])\phi(Y)] = 0\]

For (b), we have

\[\forall \phi(\cdot), \quad X - g(Y) \perp \phi(Y) \quad \text{and} \quad X - E[X|Y] \perp \phi(Y)\]
\[\implies \forall \phi(\cdot), \quad (X - E[X|Y]) - (X - g(Y)) \perp \phi(Y)\]
\[\text{taking} \quad \phi(Y) = g(Y) - E[X|Y] \quad \implies \quad g(Y) - E[X|Y] \perp g(Y) - E[X|Y]\]

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:

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

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

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

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

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

2 Jointly Gaussian

Note 4. We say that the variables $(X_1, \cdots, X_n)$ are jointly Gaussian, denoted by $X := (X_1, \cdots, X_n) \sim N(\mu, \Sigma)$, if the joint distribution has density:

$$f_x(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right\}$$

For $X \sim N(\mu, \Sigma)$, we have: $Y = c + BX \quad \Rightarrow \quad Y \sim N(c + B\mu, B\Sigma B^T)$

Note 5. Two r.v. $Y, Z$ are jointly Gaussian if $(Y, Z)$ has density function of the form

$$f_{y,z}(y, z) = K \exp\left\{-\frac{1}{2}\begin{pmatrix} (y - \mu_y) & (z - \mu_z) \\ (y - \mu_y) & (z - \mu_z) \end{pmatrix}^T \begin{pmatrix} \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_{zz} \end{pmatrix} \begin{pmatrix} (y - \mu_y) & (z - \mu_z) \end{pmatrix}\right\}$$

We have the following properties for $Y, Z$ jointly Gaussian:

(a) $Y, Z$ are both produced by linear transformation of $X = (Y, Z) \sim N\left(\begin{pmatrix} \mu_y \\ \mu_z \end{pmatrix}, \begin{pmatrix} \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_{zz} \end{pmatrix}\right)$

(b) $Y|Z \sim N(\bar{\mu}, \bar{\Sigma})$ with $\bar{\mu} = \mu_y + \Sigma_{yz}\Sigma_z^{-1}(Z - \mu_z)$ and $\bar{\Sigma} = \Sigma_y - \Sigma_{yz}\Sigma_z^{-1}\Sigma_{zy}$

(c) $f_{y,z}(y, z) = f_y(y)f_z(z) \iff \Sigma_{zy} = \Sigma_{yz} = 0$, i.e., $Y \perp Z \iff \text{cov}(Y, Z) = 0$

(d) $L[Y|Z] = E[Y|Z]$

Note that $L[Y|Z] = a + BZ$ such that $(a, B) = \arg\min_{(c, D)} E[|Y - c - DZ|^2]$

Problem 3. Prove (d) of Note 5.

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 $Y$?