Least Squares Estimation

Goal: estimate r.v. X from observation Y

Analytical soln: $\hat{X}_{LS}(Y) = \mathbb{E}[X \mid Y]$ (only good in theory; not practical)

↳ in practice, we rely on linear estimation

Linear Estimation

Def: A linear estimator of X given $Y = (Y_1 \cdots Y_n)^T$

is of the form:

$$\hat{X}(Y) = \underbrace{a + \sum_i b_i Y_i}_{\text{affine } f^n \text{ of data } Y} \qquad a, b_1, \ldots b_n \in \mathbb{R}$$

Goal of linear estimation: Find best linear estimator

Define: the linear least-squares estimator is defined as:

$$\underbrace{\mathbb{L}[X \mid Y]}_{\text{best linear estimator}} := \arg\min_{\hat{X} \text{ linear est}} \mathbb{E}[\hat{X}(Y) - X]$$

aka BLE(X|Y)

or LLSE(X|Y)

Q/ How to compute $\mathbb{L}[X \mid Y]$?

Approach ① : Calculus

$$J(a, b_1, \ldots, b_n) := \mathbb{E}\Big(\underbrace{\sum b_i (Y_i - \mathbb{E}[Y_i]) + a - (X - \mathbb{E}[X])}_{\text{linear estimator}}\Big)^2$$

$$\frac{\partial}{\partial a} J = 0 \iff a = 0$$

$$\frac{\partial}{\partial b_i} J = 0 \implies \text{Cov}(X, Y_i) = \sum_{j=1}^{n} b_j \text{Cov}(Y_i, Y_j)$$

Collecting all n equations together, we get:

$$\Sigma_{XY} = b^T \Sigma_Y$$

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

Define covariance matrices:

$$\underbrace{\Sigma_{XY}}_{\text{row vector}} := \mathbb{E}[(X - \underbrace{\mu_X}_{\mathbb{E}[X]})(Y - \mu_Y)^T] = [\text{Cov}(X, Y_1) \cdots \text{Cov}(X, Y_n)]$$

$$\Sigma_Y := \mathbb{E}[(Y - \underbrace{\mu_Y}_{\mathbb{E}[Y] = \begin{bmatrix} \mathbb{E}[Y_1] \\ \vdots \\ \mathbb{E}[Y_n] \end{bmatrix}})(Y - \mu_Y)^T] = \begin{bmatrix} \text{Cov}(Y_1, Y_1) & \text{Cov}(Y_1, Y_2) & \cdots & \text{Cov}(Y_1, Y_n) \\ & \ddots & & \\ \text{Cov}(Y_n, Y_1) & & \cdots & \text{Cov}(Y_n, Y_n) \end{bmatrix}$$

Assuming Σ_Y is invertible (WLOG)

$\Rightarrow$ optimal b is given by $\boxed{b^T = \Sigma_{XY}\Sigma_Y^{-1}}$

Note: $\mathbb{L}[X|Y]$ depends only on mean & covariance of X & Y, and not P_{XY} in general is good news for practice.

$$\Rightarrow \boxed{\mathbb{L}[X|Y] = \mu_X + \Sigma_{XY}\Sigma_Y^{-1}(Y-\mu_Y)}$$

If X is a vector $X = [X_1 \cdots X_k]^T$ then this formula still works (only difference is the matrix will be rectangular instead of square)

$$\mathbb{E}[|\hat{X}(Y) - \hat{X}|^2] = \sum_{i=1}^{n} \mathbb{E}[|\hat{X}_i(Y) - X_i|^2]$$

Approach ② : Geometric/Linear Algebraic (more insightful than ①)

Let V be a vector space over a scalar field $\mathbb{R}$.

Let $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ be an inner product on V

ie 1) $\langle u, v\rangle = \langle v, u\rangle \quad \forall v, u \in V$

2) $\langle \alpha u + \beta v, w\rangle = \alpha\langle u, w\rangle + \beta\langle v, w\rangle \quad \forall \alpha, \beta \in \mathbb{R} \quad u, v, w \in V$

3) $\langle u, u\rangle \geq 0$ and $= 0$ iff $u = 0$

ex: $V = \mathbb{R}^3$, $\langle u, v\rangle = \sum_{i=1}^{3} u_i v_i p_i$ for some fixed $p_1 p_2 p_3 > 0$

finite-dim vector space

$$u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \quad v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

ex: $V = C_b([0,1])$ = set of cts. bounded fcns on domain $[0,1]$

infinite-dim vector space

ex inner prod: $\langle f, g\rangle = \int_0^1 f(x)g(x)p(x)\,dx \qquad p > 0 \in C_b(\mathbb{R})$

Every inner product induces a norm $\|\cdot\| : V \to [0, \infty)$ via:

$$\|v\| := \sqrt{\langle v, v\rangle} \qquad v \in V$$

Recall: norms satisfy 3 properties

① $\|\alpha v\| = |\alpha|\,\|v\| \qquad \alpha \in \mathbb{R}$

② $\|v\| \geq 0 \qquad \|v\| = 0 \iff v = 0$

③ $\|u + v\| \leq \|u\| + \|v\|$

$\mathcal{V}$ equipped with $\langle\cdot,\cdot\rangle$ is called a Hilbert space

if it's $\underbrace{\text{complete}}_{\text{we can take limits}}$ wrt $\|\cdot\|$

Def: vectors u, v in Hilbert space $\mathcal{V}$ are orthogonal if $\langle u,v\rangle = 0$.

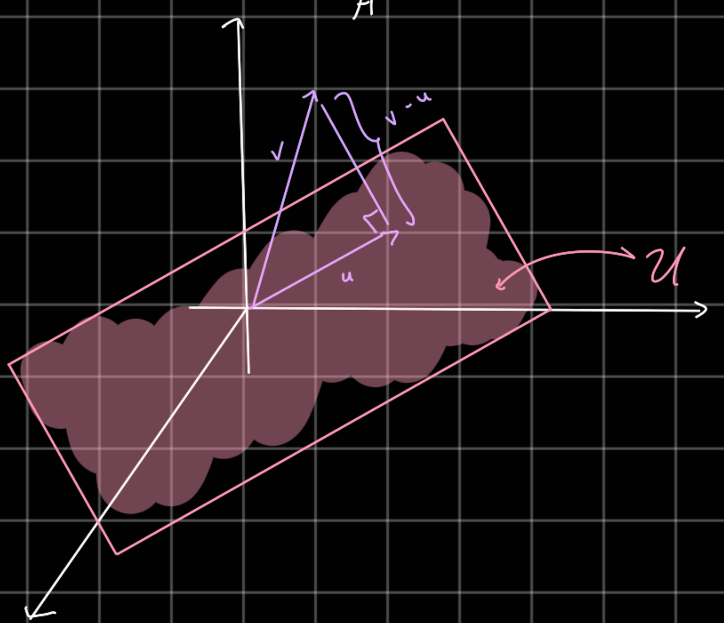
Hilbert Projection Theorem (HPT):

Let $\mathcal{H}$ be a Hilbert space & $\mathcal{U} \subseteq \mathcal{H}$ a closed linear subspace

For each $v \in \mathcal{H}$, there's a unique $u \in \mathcal{U}$ that's closest to v in $\|\cdot\|$.

ie. $\underset{u\in\mathcal{U}}{\text{argmin}}\ \|v-u\|$ exists & is unique

Moreover, $u \in \mathcal{U}$ is closest pt to v iff $\langle v-u, u'\rangle = 0 \quad \forall u' \in \mathcal{U}$ } ie error $(v-u)$ is orthogonal to $\mathcal{U}$



$\lrcorner$ = orthogonality btwn vectors in $\mathcal{H}$

Pythagorean Thm:

$$\|v\|^2 = \|u\|^2 + \|u-v\|^2$$

Pf:

$$\|u\|^2 + \|v-u\|^2 = \langle u,u\rangle + \langle v-u, v-u\rangle$$

$$= \langle u,u\rangle + \langle v-u, v\rangle - \langle v-u, u\rangle$$

$$= 2\cancel{\langle u, u-v\rangle}^{\,0\ \text{(orthogonality)}} + \langle v, v\rangle$$

$$= \langle v,v\rangle = \|v\|^2$$

Thm: Let (Ω, F, P) be a probability space. The collection of r.v's X on this space with $\mathbb{E}[X^2] < \infty$ form a Hilbert space wrt inner product $\langle X, Y\rangle := \mathbb{E}[XY]$.

Back to linear estimation:

For r.v.s $Y_1, \dots, Y_n$ w/ finite 2^{nd} moments, the space of rv's:

$$\mathcal{U} = \left\{ a + \sum b_i Y_i \ :\ a, b_1, \dots b_n \in \mathbb{R} \right\}$$

is a closed linear subspace of our Hilbert space of rv's.

By HPT,

$$\underset{u \in \mathcal{U}}{\operatorname{argmin}} \; \|X - \mathcal{U}\|^2 \quad \text{exists \& is unique}$$

$$= \underset{\hat{X}_{linear}}{\operatorname{argmin}} \; \mathbb{E}[\hat{X}(Y) - X]^2$$

$$= \mathbb{L}[X \mid Y]$$

Moreover, $\mathbb{L}[X|Y]$ characterized by:

$$\langle \mathbb{L}[X|Y] - X, L(Y) \rangle = 0 \quad \forall \text{ affine Fcns of } Y$$

$$\iff \mathbb{E}[(\mathbb{L}[X|Y] - X)(a + \textstyle\sum b_i Y_i)] = 0 \quad \forall a, b_1, \dots, b_n \in \mathbb{R}$$

$$\iff \underbrace{\mathbb{E}[\mathbb{L}[X|Y]] = \mathbb{E}[X]}_{\text{says that the BLE is unbiased}} \text{ and } \underbrace{\mathbb{E}[(\mathbb{L}[X|X] - X)Y^T] = 0}_{\text{says that the estimation error is uncorrelated with the observations}}$$

orthogonality principle (characterizes the best linear estimator)

ex: Use orthogonality principle to show

$$\mathbb{L}[X, Y] = \underbrace{\mu_x + \Sigma_{xy} \Sigma_y^{-1} (Y - \mu_y)}_{L}$$

$$\mathbb{E}[L] = \mathbb{E}[\mu_x + \Sigma_{xy} \Sigma_y^{-1} (Y - \mu_y)]$$

$$= \mu_x$$

$$= \mathbb{E}[X] \quad (\text{unbiased } \checkmark)$$

$$\mathbb{E}[(L - X)Y] = \mathbb{E}[\Sigma_{xy} \Sigma_y^{-1} (Y - \mu_y) - (X - \mu_x)(Y - \mu_y)^T]$$

$$= \Sigma_{xy} \Sigma_y^{-1} \Sigma_y - \Sigma_{xy}$$

$$= 0$$