

Math 4261: Mathematical Statistics I

Lecture 4: Conditioning and best estimates

Ref: Blitzstein–Hwang 7.3, 9.1–9.6

Fall 2026

Outline

Recap

Finishing the multivariate normal

Conditional expectation

Best estimates

Best linear estimates

Outline

Recap

Finishing the multivariate normal

Conditional expectation

Best estimates

Best linear estimates

Recap: Joint- Marginal-Conditional

We use f for either a PMF (discrete) or a PDF (continuous).

Joint: $f_{X,Y}(x,y)$ describes distribution of (X,Y) .

Marginalize: sum/integrate out the other variables

$$\text{Discrete: } f_X(x) = \sum_y f_{X,Y}(x,y), \quad f_Y(y) = \sum_x f_{X,Y}(x,y),$$

$$\text{Continuous: } f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y) dy, \quad f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y) dx.$$

Conditional: Dividing the joint law by the appropriate marginal

$$f_{Y|X}(y | x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad f_{X|Y}(x | y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Independence:

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \iff f_{Y|X}(y | x) = f_Y(y).$$

Recap: Covariance matrices

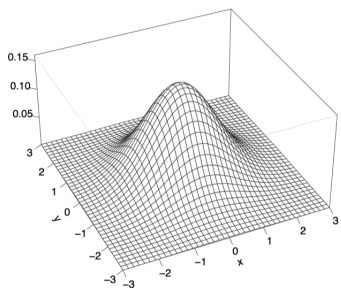
For a random vector $\mathbf{X} = (X_1, \dots, X_p)^\top$ with mean $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}]$,

$$\begin{aligned}\boldsymbol{\Sigma} &= \text{Cov}(\mathbf{X}) = \mathbb{E}\left[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top\right] \\ &= \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_p) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_p, X_1) & \text{Cov}(X_p, X_2) & \cdots & \text{Var}(X_p) \end{bmatrix}.\end{aligned}$$

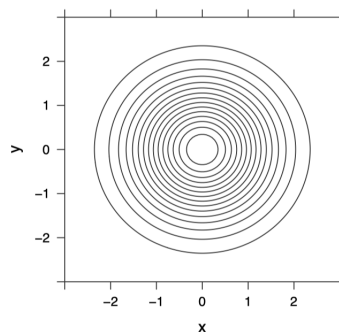
- ▶ $\boldsymbol{\Sigma}$ is symmetric and positive semidefinite.
- ▶ The covariance matrix tells us the **spread in every direction**: for every $\mathbf{a} \in \mathbb{R}^p$, $\text{Var}(\mathbf{a}^\top \mathbf{X}) = \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}$.
- ▶ A diagonal covariance matrix means the coordinates are **uncorrelated**; this does not generally imply independence.

Recap: Standard multivariate Gaussian

$$\mathbf{Z} = (Z_1, \dots, Z_p)^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p \times p}), \quad \text{i.e. } Z_1, \dots, Z_p \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1).$$



PDF surface



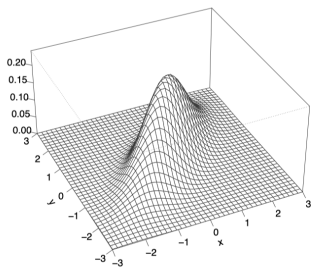
Circular contours

- ▶ PDF $f(z_1, \dots, z_p) = \frac{1}{(2\pi)^{p/2}} \exp \left\{ -\frac{1}{2} \|\mathbf{z}\|^2 \right\}$
- ▶ The distribution is **spherically symmetric**: rotations and reflections do not change it: $Q\mathbf{Z} \stackrel{d}{=} \mathbf{Z}$ for any orthonormal Q .
- ▶ $\mathbf{a}^\top \mathbf{Z} \sim \mathcal{N}(0, \|\mathbf{a}\|^2)$. If $\|\mathbf{a}\| = 1$ then $\mathbf{a}^\top \mathbf{Z} \sim \mathcal{N}(0, 1)$.

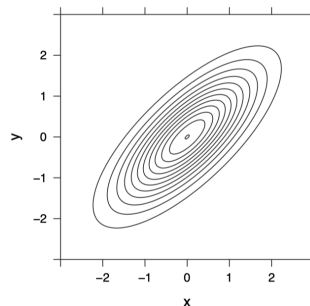
Recap: General multivariate Gaussian

A (general) multivariate Gaussian $\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ has PDF

$$f(\mathbf{z}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma})}} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})}$$



PDF surface



Elliptical contours

- ▶ $\boldsymbol{\mu}$ gives the center; $\boldsymbol{\Sigma}$ controls the spread, shape, and orientation.
- ▶ We'll see shortly that any multivariate Gaussian can be obtained by linearly transforming a standard one.
- ▶ Every linear function remains Gaussian.
- ▶ uncorrelated $\iff$ independent.

Why covariance matrix gives the shape?

For any unit vector $\mathbf{u}$, the variance in direction $\mathbf{u}$ is

$$\text{Var}(\mathbf{u}^\top \mathbf{X}) = \mathbf{u}^\top \Sigma \mathbf{u}.$$

Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ be the eigenvalues of Σ , with orthonormal eigenvectors $\mathbf{q}_1, \dots, \mathbf{q}_p$. Writing $\mathbf{u} = \sum_{i=1}^p c_i \mathbf{q}_i$, where $\sum_{i=1}^p c_i^2 = 1$, gives

$$\mathbf{u}^\top \Sigma \mathbf{u} = \sum_{i=1}^p \lambda_i c_i^2 \leq \lambda_1 \sum_{i=1}^p c_i^2 = \lambda_1.$$

Therefore:

- ▶ $\mathbf{q}_1$ is the direction with the **largest variance**, λ_1 .
- ▶ $\mathbf{q}_2$ has the next largest variance, λ_2 . And so on.
- ▶ The principal axes are the $\mathbf{q}_i$. The geometric stretch factors are related to λ_i .

Outline

Recap

Finishing the multivariate normal

Conditional expectation

Best estimates

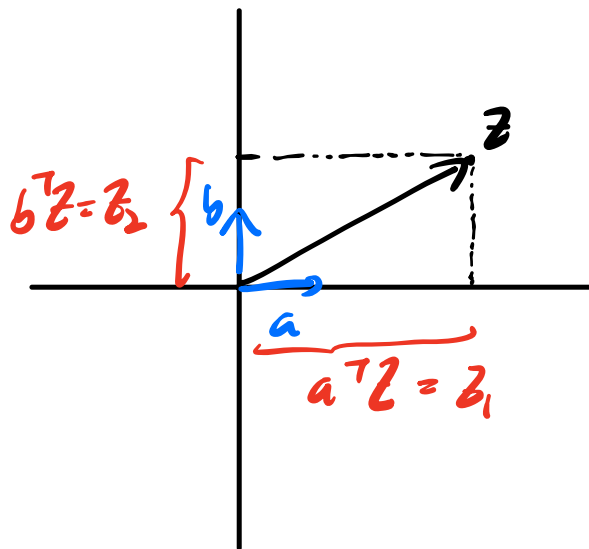
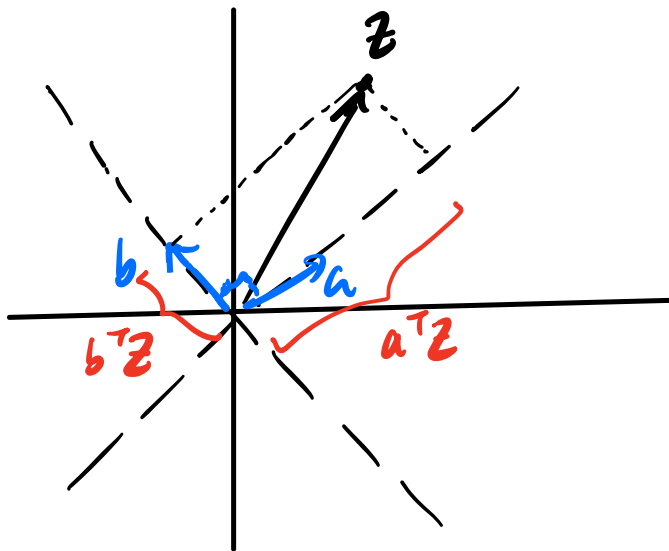
Best linear estimates

Consequences of symmetry—II

Let $\mathbf{Z} = (Z_1, \dots, Z_p) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p \times p})$ and let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p$ be **any two vectors having length 1** that are perpendicular to each other:

$$\mathbf{a}^\top \mathbf{b} = 0$$

Then $\mathbf{a}^\top \mathbf{Z}$ and $\mathbf{b}^\top \mathbf{Z}$ are **independent** $\mathcal{N}(0, 1)$ random variables because their joint distribution is the same as that of (Z_1, Z_2) .



Example: Independence of mean and residuals

Let

$$\mathbf{Z} = (Z_1, \dots, Z_p) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p \times p}).$$

What is the joint distribution of

$$\bar{Z} = \frac{1}{p} \sum_{j=1}^p Z_j \quad \text{and} \quad Z_1 - \bar{Z}?$$

Let $\mathbf{a} = \left(\frac{1}{p}, \dots, \frac{1}{p}\right)$, $\mathbf{b} = \left(1 - \frac{1}{p}, -\frac{1}{p}, \dots, -\frac{1}{p}\right)$

length of $\mathbf{a} = \frac{1}{\sqrt{p}}$ length of $\mathbf{b} = \sqrt{1 - \frac{1}{p}}$

Observe, crucially, $\mathbf{a}^T \mathbf{b} = 0$ so $\mathbf{a}, \mathbf{b}$ perpendicular

Hence $\bar{Z} = \mathbf{a}^T \mathbf{Z} \sim \mathcal{N}(0, \frac{1}{p})$
 $Z_1 - \bar{Z} = \mathbf{b}^T \mathbf{Z} \sim \mathcal{N}(0, 1 - \frac{1}{p})$

independent!

Properties of multivariate normal—1

Representation by standard MVN

Recall for univariate Gaussian that $X \sim \mathcal{N}(\mu, \sigma^2)$ can be represented as $X = \mu + \sigma Z$, $Z \sim \mathcal{N}(0, 1)$. Analogously:

Theorem

If $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in $\mathbb{R}^p$, then

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{V}\mathbf{Z}$$

for $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p \times p})$ and a matrix $\mathbf{V} \in \mathbb{R}^{p \times p}$ satisfying $\boldsymbol{\Sigma} = \mathbf{V}\mathbf{V}^\top$.

Conversely if $\mathbf{X} = \boldsymbol{\mu} + \mathbf{V}\mathbf{Z}$ where $\boldsymbol{\mu}$, $\mathbf{V}$ are any fixed vector/matrix and $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p \times p})$ then

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{where} \quad \boldsymbol{\Sigma} = \mathbf{V}\mathbf{V}^\top$$

(This gives an alternative definition of $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$).

Properties of multivariate normal—2

Linear transformations of a multivariate normal

Theorem

Suppose $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in $\mathbb{R}^p$. Let $\mathbf{a} \in \mathbb{R}^q$ and $\mathbf{B} \in \mathbb{R}^{q \times p}$ be fixed. Then

$$\mathbf{Y} = \mathbf{a} + \mathbf{B}\mathbf{X} \sim \mathcal{N}(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top).$$

We record a useful special case: when $\mathbf{b} \in \mathbb{R}^p$,

$$\mathbf{b}^\top \mathbf{X} \sim \mathcal{N}(\mathbf{b}^\top \boldsymbol{\mu}, \mathbf{b}^\top \boldsymbol{\Sigma} \mathbf{b}).$$

proof:

Write $\mathbf{X} = \boldsymbol{\mu} + \mathbf{V}\mathbf{Z}$, where $\mathbf{V}\mathbf{V}^\top = \boldsymbol{\Sigma}$,
 $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p \times p})$

Properties of multivariate normal—2

Then $Y = a + BX$

$$= a + B(\mu + VZ)$$

$$= \underbrace{(a + B\mu)}_{\text{deterministic mean part}} + \underbrace{BVZ}_{\text{randomness}}$$

If X has cov. Σ ,
 AX has cov
(AX)

$$\begin{aligned}\text{Cov } Y &= E BVZ (BVZ)^T = E BVZ Z^T V^T B^T \\ &= BV \underbrace{E[Z Z^T]}_{\text{Cov } Z = I} V^T B^T \\ &= B \Sigma B^T\end{aligned}$$

Then $Y \sim \mathcal{N}(a + B\mu, B \Sigma B^T)$

Outline

Recap

Finishing the multivariate normal

Conditional expectation

Best estimates

Best linear estimates

Motivation

Consider two r.v.'s X (observed) and Y (unobserved). The goal is to predict Y on the basis of X . Examples:

- ▶ Observe temperature in Atlanta, GA, predict temperature in Birmingham, AL (easy). Or predict temperature in NYC (harder).
- ▶ Observe past stock prices, predict tomorrow's price.

Let our estimate be $\widehat{Y} = \widehat{Y}(X)$ (function of X). A common measure of the quality of our estimate is

$$\text{Mean-square error (MSE)} = \mathbb{E}[(Y - \widehat{Y})^2]$$

It turns out that the best $\widehat{Y}$ (the one minimizing MSE) is the **conditional expectation** of Y given X !

Conditional expectation (given an event)

Continuous: Recall **conditional PDF** of Y given $X = x$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

The **conditional expectation** of Y given $X = x$ is

$$\mathbb{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x) \, dy$$

Discrete: Recall **conditional PMF** of Y given $X = x$ is the PMF

$$\mathbb{P}[Y = \bullet \mid X = x]$$

The **conditional expectation** of Y given $X = x$ is

$$\mathbb{E}[Y \mid X = x] = \sum_y y \cdot \mathbb{P}[Y = y \mid X = x].$$

Conditional expectation (given a r.v.)

Recall

$$\mathbb{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x) dy$$

This is a function of x . Let's call it $g(x)$. The notation

$$\mathbb{E}[Y \mid X]$$

is understood as the **random variable** $g(X)$.

- ▶ $g(X)$ does **not** refer to the (rather meaningless) $\mathbb{E}[Y \mid X = X]$.
- ▶ One must first figure out $g(x)$, then plug in X for x . E.g. $g(x) = x^2$, then $g(X) = X^2$.
- ▶ Somewhat confusingly, $\mathbb{E}[Y \mid X = x]$ and $\mathbb{E}[Y \mid X]$ are both referred to as conditional expectation.

Law of total expectation

A useful way to compute expectations:

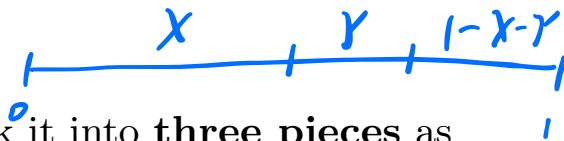
$$\mathbb{E}[Y] = \begin{cases} \sum_x \mathbb{E}[Y \mid X = x] \mathbb{P}[X = x] & \text{discrete} \\ \int \mathbb{E}[Y \mid X = x] \cdot f_X(x) dx & \text{continuous} \end{cases}$$

Succinctly (also called the tower property), this is written as

$$\boxed{\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]}.$$

Exercise: show that this follows directly from LOTUS.

Example: Stick breaking



(Revisited) Given a stick of unit length, break it into **three pieces** as follows.

1. Choose point X uniformly on the stick and break the stick into two pieces of length X and $1 - X$.
2. Choose point Y uniformly on the $1 - X$ piece and break it at Y .

This gives three pieces of length X , Y , and $1 - X - Y$.

One can show directly that $\mathbb{E}[Y] = 1/4$. Show this using Law of total expectation.

$$X \sim \text{Unif}(0, 1), \quad Y | X=x \sim \text{Unif}(0, x)$$

$$\mathbb{E}X = \frac{1}{2}$$

$$\mathbb{E}[Y | X=x] = \frac{x}{2} \quad \text{plugging in } X \text{ for } x,$$

$$\mathbb{E}[Y | X] = \frac{X}{2}.$$

LOTE:

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y | X]] = \mathbb{E}\left[\frac{X}{2}\right] = \frac{1}{4}.$$

Conditional variance

The **conditional variance** of Y given X is

$$\begin{aligned}\mathrm{Var}[Y \mid X] &= \mathbb{E}[(Y - \mathbb{E}[Y \mid X])^2 \mid X] \\ &= \mathbb{E}[Y^2 \mid X] - (\mathbb{E}[Y \mid X])^2\end{aligned}$$

- ▶ Just replace all instances of $\mathbb{E}[\cdot]$ with $\mathbb{E}[\cdot \mid X]$ in the (unconditional) variance $\mathrm{Var} Y$.
- ▶ Similar to $\mathbb{E}[Y \mid X]$, $\mathrm{Var}[Y \mid X]$ is a **random variable**. (It is a function of X).

Eve's law

For any r.v.'s X and Y :

$$\boxed{\text{Var } Y = \mathbb{E}[\text{Var}[Y \mid X]] + \text{Var}[\mathbb{E}[Y \mid X]]}$$

The ordering of $\mathbb{E}$ and Var spells **EVVE**, hence the name¹.

¹Also called the law of total variance.

Properties of conditional expectation

- ▶ **Dropping what's independent.** If X and Y independent then

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

- ▶ **Taking out what's known.** For any function h ,

$$\mathbb{E}[h(X) \cdot Y \mid X] = h(X) \cdot \mathbb{E}[Y \mid X]$$

- ▶ **Linearity.**

$$\begin{aligned}\mathbb{E}[Y_1 + Y_2 \mid X] &= \mathbb{E}[Y_1 \mid X] + \mathbb{E}[Y_2 \mid X] \\ \mathbb{E}[cY \mid X] &= c\mathbb{E}[Y \mid X]\end{aligned}$$

- ▶ **Law of total expectation**

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

Outline

Recap

Finishing the multivariate normal

Conditional expectation

Best estimates

Best linear estimates

Conditional expectation minimizes MSE

Recall the setup: two r.v.'s X (observed) and Y (unobserved).
Goal: predict Y on the basis of X . Denote our estimate of Y by

$$\delta(X) \quad (\text{function of } X)$$

A common measure of the quality of our estimate is

$$\text{Mean-square error (MSE)} = \mathbb{E} \left[(Y - \delta(X))^2 \right]$$

Warm up: no data available

Suppose there is no data X observed. We have to make a **blind guess** about Y . Goal becomes: find the constant c that minimizes the MSE

$$\mathbb{E}[(Y - c)^2]. \quad \text{Var } Z = \mathbb{E}[Z^2] - (\mathbb{E} Z)^2$$

The minimizer turns out to be the **mean**: $c_{\text{opt}} = \mathbb{E}[Y]$!

$$\begin{aligned} \text{MSE} &= \mathbb{E}[\overbrace{(Y - c)^2}^{Z^2}] = \underbrace{\text{Var}[Y - c]}_{\text{Var } Y} + \underbrace{(\mathbb{E}[Y - c])^2}_{= (\mathbb{E} Y - c)^2} \\ &\quad \text{(nothing to do with } c) \qquad \geq 0 \text{ minimized at } c = \mathbb{E} Y \end{aligned}$$

General case

Observe $X = x$. Find the best function $\delta(x)$ to estimate Y . That is, find the function (estimator) δ that minimizes the MSE

$$\mathbb{E}[(Y - \delta(X))^2].$$

Minimizer turns out to be the **conditional mean**!

$$\delta_{\text{opt}}(x) = \mathbb{E}[Y \mid X = x]$$

$$\text{MSE} = \mathbb{E}[(Y - \delta(X))^2] = \iint (y - \delta(x))^2 \underbrace{f_{X,Y}(x,y)}_{f_X(x) f_{Y|X}(y|x)} dx dy$$

$$= \int \left[\int (y - \delta(x))^2 f_{Y|X}(y|x) dy \right] f_X(x) dx$$

x here fixed so $\delta(x)$ is constant.

This is $\mathbb{E}[Y - \delta(x) \mid X=x]$

We have solved this problem before!

Outline

Recap

Finishing the multivariate normal

Conditional expectation

Best estimates

Best linear estimates

Motivation

The best estimator for predicting Y given X (in the MSE sense) is the conditional expectation $\mathbb{E}[Y \mid X]$.

A **linear estimator** uses

$$aX + b, \quad \text{for some } a, b \in \mathbb{R}$$

to predict Y .

Two reasons why linear is desirable:

- ▶ Linear rule is simple and interpretable
- ▶ Computing best rule $\mathbb{E}[Y \mid X]$ requires conditional or joint PDF, which might not be available.

It turns out for best linear estimate, we only need **mean, variance, and covariance!**

Finding the best linear estimator

Choose the **best coefficients** a and b in $\delta(X) = aX + b$.

The b that minimizes $\mathbb{E}[(Y - aX + b)^2]$ is

$$b = \mathbb{E}[Y - aX] = \mathbb{E}Y - a\mathbb{E}X$$

Then find minimizing a in

$$\text{MSE} = \mathbb{E}[(Y - aX - \mathbb{E}[Y - aX])^2]$$

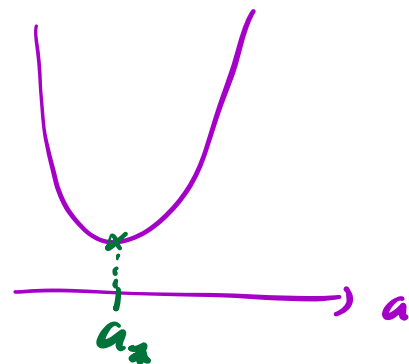
$$= \text{Var}[Y - aX]$$

$$= \text{Var} Y + a^2 \text{Var} X - 2a \text{Cov}[X, Y]$$

Quadratic function in a

Solve

$$a_{\text{ML}} = \frac{\text{Cov}[X, Y]}{\text{Var} X}$$



Best linear estimator

The best linear estimator $\delta(X) = aX + b$ has

$$\text{Best coefficients} = \begin{cases} a = \frac{\text{Cov}[X, Y]}{\text{Var } X} \\ b = \mathbb{E}Y - a\mathbb{E}X \end{cases}$$

Thus,

$$\text{Best linear estimator} = aX + b = \boxed{\mathbb{E}[Y] + \frac{\text{Cov}[X, Y]}{\text{Var } X}(X - \mathbb{E}X)}$$

This achieves the minimum MSE among all linear estimators:

$$(1 - \rho^2) \text{Var}[Y]$$

Usefulness of linear estimation

Best MSE among all linear estimators:

$$(1 - \rho^2) \text{Var}[Y]$$

- ▶ Linear estimator useful if X and Y are correlated ($\rho \neq 0$).
Perfect if $\rho = \pm 1$.
- ▶ Linear estimator useless if X and Y uncorrelated ($\rho = 0$).

Example

Let $X \sim \mathcal{N}(0, 1)$ and $Y = X^2$.

- ▶ One can compute $\rho = 0$. The best linear estimate will be $\mathbb{E}Y = 1$. Knowing X does not help.
- ▶ Best (non-linear) estimator is $\mathbb{E}[Y \mid X] = X^2 = Y$ which is perfect.

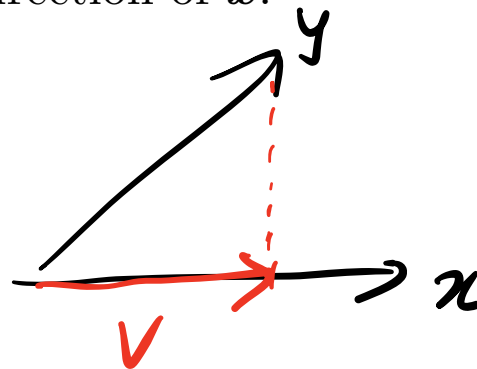
Projection interpretation

Best linear estimator:

$$\mathbb{E}[Y] + \frac{\text{Cov}[X, Y]}{\text{Var } X}(X - \mathbb{E}X)$$

Compare: projection of a vector y in the direction of x :

$$V = \frac{\langle x, y \rangle}{\|x\|^2} x$$



Dictionary with linear algebra:

$$\text{Cov} \longleftrightarrow \langle \cdot, \cdot \rangle$$

$$\text{Var} \longleftrightarrow \|\cdot\|^2$$

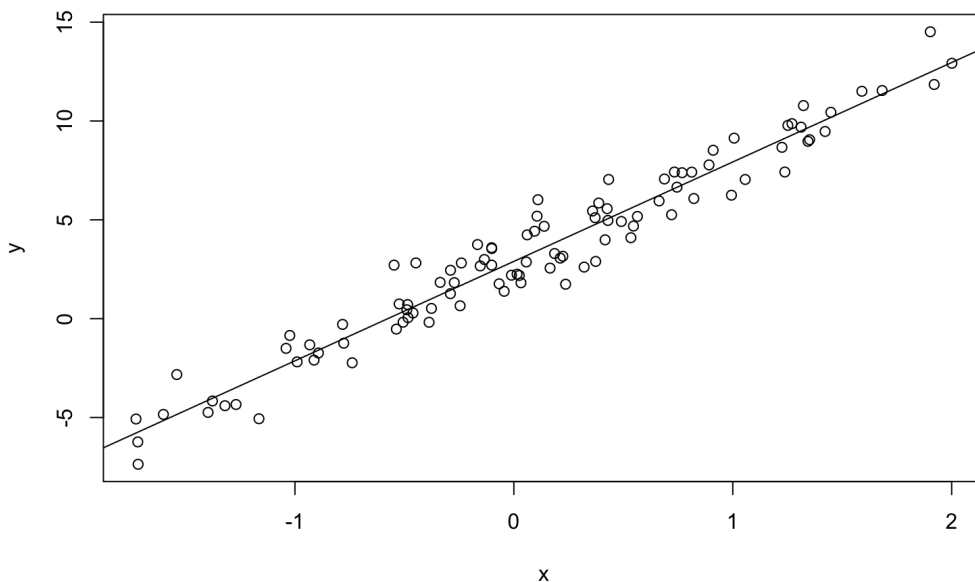
$$\rho \longleftrightarrow \text{Cosine of the angle}$$

$$\text{Uncorrelated} \longleftrightarrow \text{Orthogonal}$$

Visualization

Let $X \sim \mathcal{N}(0, 1)$ and $Y = 5X + 3 + \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$.

Generate 100 realizations of the data (X, Y) and fit a linear estimator using the **sample** mean, variance and covariance (will define these things later) to compute a and b



Example: Predicting die outcomes

Roll a fair die n times. Suppose we observe $X = \text{number of one's}$. Use a linear estimator to predict $Y = \text{number of six's}$.

[Hint: first compute $\text{Cov}[X, Y] = -n/36$.]

$$X \sim \text{Bin}(n, \frac{1}{6}), Y \sim \text{Bin}(n, \frac{1}{6})$$

$$\mathbb{E}X = \frac{n}{6} = \mathbb{E}Y, \quad \text{Var}X = \frac{5n}{6} = \text{Var}Y$$

Claim $\text{Cov}[X, Y] = -\frac{n}{36}$

Example: Predicting die outcomes

Best linear estimate

Intuition: if roll $n=60$ times and see $X = \# 1\text{'s} = 20$.

reasonable guess $Y = \frac{60-20}{5} = 8$

← 5 other numbers
2, 3, 4, 5, 6

$$EY + \frac{\text{Cov}[X, Y]}{\text{Var} X} (X - EX)$$

$$= \frac{\cancel{\rho} \cancel{\sqrt{\text{Var} X}}}{\cancel{\sqrt{\text{Var} Y}}} (X - EX)$$

$$= \frac{1}{6} + \left(-\frac{1}{5}\right) \left(X - \frac{1}{6}\right)$$

$$= \frac{1-X}{5}$$

$$\rho = \frac{\text{Cov}}{\sqrt{\text{Var} X} \sqrt{\text{Var} Y}}$$

$$\frac{\text{Cov}}{\text{Var} X} = \frac{\rho \sqrt{\text{Var} Y}}{\sqrt{\text{Var} X}}$$

Proof of claim

Decompose X, Y as sum of Bernoulli

$$X = \sum_{i=1}^n X_i, \quad X_i = \begin{cases} 1, & \text{ith toss 1} \\ 0, & \text{o/w} \end{cases}$$

$$Y = \sum_{i=1}^n Y_i, \quad Y_i = \begin{cases} 1, & \text{ith toss 6} \\ 0, & \text{o/w} \end{cases}$$

for $i \neq j$, X_i, Y_j independent

$$\text{Cov}[X, Y] = \text{Cov}\left[\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i\right] \quad \text{= 0 by independence}$$

$$= \sum_{i=1}^n \text{Cov}[X_i, Y_i] + \sum_{i \neq j} \text{Cov}[X_i, Y_j]$$

$$= \sum_{i=1}^n \underbrace{\mathbb{E}[X_i Y_i]}_{\substack{\text{only 1} \\ \text{if } i\text{th} \\ \text{toss both 1 and 6} \\ \text{— impossible}}} - \underbrace{\mathbb{E}[X_i]}_{\substack{\text{P[ith toss is 1]} \\ = \frac{1}{6}}} \underbrace{\mathbb{E}[Y_i]}_{\frac{1}{6}}$$

$$= \sum_{i=1}^n -\left(\frac{1}{6}\right)^2 = -\frac{n}{36}$$

$$\text{Also } \text{Var } X = \text{Var } Y = n \cdot \frac{1}{6} \cdot \frac{5}{6}$$

Thus

$$\rho(X, Y) = \frac{-\frac{n}{36}}{n \cdot \frac{1}{6} \cdot \frac{5}{6}} = -\frac{1}{5} = -0.2$$

Negative correlated, but only weakly,
as expected.