# Linear Regression Problem Sets

## Problem 1

A data set is constructed by taking 100 samples from a normal distribution with mean μ = 5 and standard deviation σ = 2 to construct a random variable Xi, i = 1, , 100. Then, a 2nd random variable Yi, i = 1, , 100 is constructed by taking the values of the corresponding Xi and adding one half of a third random variate drawn from a normal distribution with mean 5 and standard deviation 2 and thus a set of 100 pairs (Xi, Yi) is obtained. Find the parameters of a linear regression of Y on X (both by doing the numerical experiment and by calculating the result analytically).

## Solution

You generate

$$X_i \sim \mathcal{N}(5, 2^2)$$

and independently

$$Z_i \sim \mathcal{N}(5, 2^2)$$

then define

$$Y_i = X_i + \frac{1}{2} Z_i.$$

We want the regression of $Y$on $X$:

$$Y = mX + b$$

### Step 1: Slope $\boldsymbol{m}$

For linear regression,

$$m = \frac{\operatorname{Cov}(X, Y)}{\operatorname{Var}(X)}$$

Since $Y = X + \frac{1}{2} Z$,

$$\operatorname{Cov}(X, Y) = \operatorname{Cov}(X, X + \frac{1}{2} Z) = \operatorname{Var}(X) + \frac{1}{2} \operatorname{Cov}(X, Z)$$

Because $X$and $Z$are independent,

$$\mathrm{Cov}\,(X, Z) = 0$$

Also,

$$\mathrm{Var}\,(X) = 2^2 = 4$$

So,

$$\mathrm{Cov}\,(X, Y) = 4$$

Hence,

$$m = \frac{4}{4} = 1$$

**Step 2: Intercept *b***

$$b = \mathbb{E}[Y] - m\,\mathbb{E}[X]$$

Now,

$$\mathbb{E}[Y] = \mathbb{E}[X + \frac{1}{2}Z] = \mathbb{E}[X] + \frac{1}{2}\mathbb{E}[Z] = 5 + \frac{1}{2}(5) = 7.5$$

With $m = 1$and $\mathbb{E}[X] = 5$,

$$b = 7.5 - 1 \cdot 5 = 2.5$$

$$\boxed{m = 1, b = 2.5}$$

Thus, the regression line is:

$$\boxed{Y = X + 2.5}$$

## Problem 2

Repeated coin tossing of an (unfair) coin produces 100 heads up and 120 tails up. Find a maximum likelihood estimate for the probability that a coin toss will result in heads up.

## Solution

### Step 1: Define the random variable

Let

$$X_i = \begin{cases} 1, & \text{if the } i\text{-th toss is heads} \\ 0, & \text{if the } i\text{-th toss is tails} \end{cases}$$

Then $X_i$follows a Bernoulli distribution with parameter $p$:

$$P(X_i = x) = p^x(1-p)^{1-x}, x = 0,1$$

**Step 2: Write the likelihood function**

Suppose we observe $N = 220$ tosses in total, with:

- $N_u = 100$ heads
- $N_d = 120$ tails

The likelihood function is the probability of observing this data as a function of $p$:

$$L(p) = p^{N_u}(1-p)^{N_d}$$

So here:

$$L(p) = p^{100}(1-p)^{120}$$

**Step 3: Take the log-likelihood**

It is easier to maximize the logarithm of the likelihood:

$$\begin{aligned}&\ell(p) = \ln\ L(p)\\&\ell(p) = 100\ln\ p + 120\ln\ (1-p)\end{aligned}$$

**Step 4: Differentiate with respect to $p$**

$$\frac{d\ell}{dp} = \frac{100}{p} - \frac{120}{1-p}$$

**Step 5: Set the derivative equal to zero**

$$\frac{100}{p} - \frac{120}{1-p} = 0$$

Solve for $p$:

$$\begin{aligned}&\frac{100}{p} = \frac{120}{1-p}\\&100(1-p) = 120p\\&100 - 100p = 120p\\&100 = 220p\\&p = \frac{100}{220}\end{aligned}$$

$$\boxed{\hat{p} = \frac{100}{220} \approx 0.4545}$$

The maximum likelihood estimate of the probability of heads is the fraction of heads observed. Since only 100 out of 220 tosses were heads, the estimated probability is about 0.4545, meaning the coin is slightly biased toward tails.