Jackson Petty
PHYS 356
21 February 2018

ASSIGNMENT 2

Problem 1. Suppose you have bivariate data $(x_1, y_1), \ldots, (x_n, y_n)$. A common model is that there is a linear relationship between x and y, so in principle the data should lie exactly along a line. However since data have random noise and our model is probably not exact this will not be the case. What we can do is look for the line that best fits the data. To do this we will use a simple linear regression model. For bivariate data the simple linear regression model assumes that the x_i are not random but that for some values of the parameters a and b the value y_i is drawn from the random variable $Y_i \sim ax_i + b + \varepsilon_i$, where ε_i is a normal random variable with mean $\mu = 0$ and variance σ^2. We assume all of the random variables ε_i are independent.

(a) The distribution of Y_i depends on a, b, σ and x_i. Of these only a and b are not known. Give the formula for the likelihood function $f(y_i \mid a, b, x_i, \sigma)$ corresponding to one random value y_i.

(b) (i) Suppose we have data (1, 8), (3, 2), (5, 1). Based on our model write down the likelihood and log likelihood as functions of a, b, and σ.

(ii) For general data $(x_1, y_1), \ldots, (x_n, y_n)$ give the likelihood and log likelihood functions (again as functions of a, b, and σ).

(c) Assume σ is a constant, known value. For the data in part b(i) find the maximum likelihood estimates for a and b. Give confidence intervals for your estimates of a and b.

(d) Use python to plot the data and the regression line you found in 1c.

Solution to (a). If y_i is related to x_i through

$$y_i \sim ax_i + b - \varepsilon_i,$$

where $\varepsilon_i \sim \mathcal{N}(\mu, \sigma^2)$, let $\boldsymbol{\theta}$ be a tuple representing our parameters $\boldsymbol{\theta} = (x_i; a, b, \sigma)$. We know that $f(y_i \mid \boldsymbol{\theta})$ represents the probably density function for y_i given the parameters of $\boldsymbol{\theta}$. In general, we know that the likelihood function for a set of n observations y_i is given by

$$\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y}) = \prod_{i=1}^{n} f(y_i \mid \boldsymbol{\theta}), \tag{1}$$

and so for a single observation y_i, the likelihood function for one random value is simply

$$\mathcal{L}(\boldsymbol{\theta} \mid y_i) = f(y_i \mid \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{(y_i - [b + ax_i])^2/(2\sigma^2)}, \qquad \square$$

since if ε_i is normally distributed then $y_i \sim \mathcal{N}(ax_i + b, \sigma^2)$.

Solution to (b)(i). By equation (1), we know that

$$\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y}) = \prod_{i=1}^{n} f(y_i \mid \boldsymbol{\theta}).$$

Since our data is normally distributed, we know by part (a) the form that the likelihood function will take. Then for the particular set of obsernations, $\{(1, 8), (3, 2), (5, 1)\}$ the total likelihood function is then

$$\mathcal{L} = \left[\frac{1}{\sqrt{2\pi\sigma^2}} \right]^3 \cdot \exp\left(\frac{(8 - [b + a])^2 + (2 - [b + 3a])^2 + (1 - [b + 5a])^2}{2\sigma^2} \right)$$

$$= \left[\frac{1}{\sqrt{2\pi\sigma^2}} \right]^3 \cdot \exp\left(\frac{35a^2 + 18ab - 38a + 3b^2 - 22b + 69}{2\sigma^2} \right).$$

The log likelihood of this is then

$$\ln\mathcal{L} = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\left[(8 - [b + 1])^2 + (2 - [b + 3a])^2 + (1 - [b + 5a])^2 \right].$$

$$= \boxed{-\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\left[35a^2 + 18ab - 38a + 3b^2 - 22b + 69 \right].} \qquad \square$$

Solution to (ii). Let $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be the set of n observations based on the parameters $\boldsymbol{\theta}$. For a set of n iid observations, we know that the likelihood function for those observations is given by

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i \mid \boldsymbol{\theta}) = \boxed{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(y_i - [b + ax_i])^2/(2\sigma^2)}.}$$

The log likelihood transforms this into

$$\ln\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln \frac{1}{\sqrt{2\pi\sigma^2}} e^{(y_i - [b + ax_i])^2/(2\sigma^2)}$$

$$= \boxed{-\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - [b + ax_i])^2.} \qquad \square$$

Solution to (c). In order to maximize our likelihood estimates, we need to differentiate our likelihood function (or log likelihood) with respect to the subset of parameters $\boldsymbol{\theta}^* = (a, b)$ we wish to account for. So

$$\frac{\partial \ln \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})}{\partial \boldsymbol{\theta}^*} = 0 \implies \frac{\partial}{\partial \boldsymbol{\theta}}\left[-\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - [b + ax_i])^2\right] = 0,$$

Since all other terms are constant in a, b. Evaluating these derivatives separately, we see that

$$\frac{\partial \ln \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})}{\partial a} = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i \cdot (y_i - [b + ax_i]),$$

$$\frac{\partial \ln \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y})}{\partial b} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - [b + ax_i]).$$

When this system is minimized, we get that

$$\sum_i y_i - b - ax_i = 0, \quad \text{and} \quad \sum_i y_i x_i - bx_i - ax_i^2 = 0,$$

which we can simplify using the arithmetic means of x_i and y_i to

$$\bar{y} - b - a\bar{x} = 0, \quad \text{and} \quad \bar{y}\bar{x} - b\bar{x} - a\bar{x}^2 = 0.$$

This tells us that the maximum likelihood estimates for a and b are

$$\hat{b} = \bar{y} - \hat{a}\bar{x}, \quad \text{and} \quad \hat{a} = \frac{\text{Cov}(x, y)}{\text{Var}(x)} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}.$$

For the data given in part (i), we see that

$$\text{Cov}(x, y) \approx -4.67, \quad \text{Var}(x) \approx 14.33 \implies \boxed{\hat{a} = -3.07, \quad \hat{b} = 12.88.}$$

Since our data was initially drawn from a distribution approximating $\mathcal{N}(ax_i + b, \sigma^2)$, and we have drawn 3 samples, then the standard error for our estimates of a and b should be

$$\text{error} = \pm N \cdot \frac{\sigma}{\sqrt{3}},$$

where N is $\qquad \square$

Problem 2. Estimating parameters of a uniform distribution.

(a) Suppose we have data $1.2, 2.1, 1.3, 10.5, 5$ which we know is drawn independently from a uniform distribution $\sim \mathcal{U}(a, b)$. Give the maximum likelihood estimate for the parameters a and b. Recall that the pdf for $\mathcal{N}(a, b)$ is

$$f(x \mid a, b) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise.} \end{cases}$$

3

(b) Suppose we have data $x_1, x_2, \ldots, x_n$ which we know is drawn independently from a uniform distribution $\sim \mathcal{U}(a, b)$. Give the maximum likelihood estimate for the parameters a and b in mathematical form.

Solution to (a). We know that the interval $[a, b]$ must contain all our observations, or else they cannot have been drawn from the uniform distribution. Notice that $f \propto 1/(b-a)$, so to maximize the probability (and hence the likelihood), we must minimize $b - a$. This entails making our bounds as close as possible to the extrema of our data. If we set the upper bound b to be the largest point in our dataset, and a to be the smallest, then $b - a$ is as small as it can be. Then for the example data, our estimators for a and b should be

$$\boxed{\hat{a} = 1.2, \qquad \hat{b} = 10.5.}$$ $\qquad \square$

Solution to (b). As we've established, if $f(x \mid \boldsymbol{\theta})$ is the probably distribution for a single independent event, then for n samples $\mathbf{x}$ we see that for a fixed a, b our probability is

$$f(\mathbf{x} \mid \boldsymbol{\theta}) = \begin{cases} \left(\frac{1}{b-a}\right)^n, & a \leq x \leq b \\ 0, & \text{otherwise.} \end{cases}$$

This is maximized when $b - a$ is as small as possible, under the condition that $x_i \in [a, b]$ for all i. Then our system is maximized when

$$\boxed{\hat{a} = \min(x_1, \ldots, x_n), \qquad \hat{b} = \max(x_1, \ldots, x_n).}$$ $\qquad \square$