1 One-parameter families of random variables

Let $I \subset \mathbb{R}$ be an interval in $\mathbb{R}$. Let $Y|_x : (\Omega, p) \longrightarrow \mathbb{R}$ be a family of random variables, parametrized by $x \in I \subset \mathbb{R}$, defined on the probability set (Ω, p).

The *regression function* of the family $Y|_x$ is defined as follows:

$$I \longrightarrow \mathbb{R} : x \longmapsto E(Y|_x).$$

The *regression curve* of the family $Y|_x$ is the graph of the regression function.

For each fixed $x \in \mathbb{R}$, let $f_{Y|x}(y)$ denote the probability distribution of $Y|_x$.

2 The Simple Linear Model

A one-parameter family $\{\, Y|_x \,\}_{x \in I}$ of random variables is called a *simple linear model* if the following conditions are satisfied:

1. There exists $\beta_0, \beta_1, \sigma \in \mathbb{R}$, with $\sigma > 0$, such that $Y|_x \sim \mathcal{N}(\, \beta_0 + \beta_1 x \,,\, \sigma^2 \,)$, for each $x \in I$. In other words,

$$f_{Y|x}(y) \;=\; f_{Y|x}(y\,;\,\beta_0, \beta_1, \sigma) \;=\; \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{y-\beta_0-\beta_1 x}{\sigma}\right)^2}$$

2. For any $x_1, x_2 \in I$ with $x_1 \neq x_2$, we have that $Y|_{x_1}$ and $Y|_{x_2}$ are independent random variables.

3 The maximum likelihood estimators of the parameters β_0, β_1, and σ^2 of the Simple Linear Model

Suppose $\{\, Y|_x \,\}_{x \in I}$ is a simple linear model, with parameters β_0, β_1, and σ^2. Let $x_1, x_2, \ldots, x_n \in I$ be distinct. The *likelihood function* of the observations $Y_1 := Y|_{x_1}$, $Y_2 := Y|_{x_2}$, $\ldots$, $Y_n := Y|_{x_n}$ is defined as follows:

$$L(\, \beta_0, \beta_1, \sigma\,;\, y_1, \ldots, y_n \,) \;:=\; f_{Y|x_1}(y_1) \cdot f_{Y|x_2}(y_2) \cdots f_{Y|x_n}(y_n) \;=\; \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \cdot \prod_{i=1}^{n} e^{-\frac{1}{2}\left(\frac{y_i-\beta_0-\beta_1 x_i}{\sigma}\right)^2}$$

Theorem 3.1
The maximum likelihood estimators for β_0, β_1, and σ^2 of the simple linear model $\{\, Y|_x \,\}_{x \in I}$ based on the observations $Y_1 := Y|_{x_1}$, $Y_2 := Y|_{x_2}$, $\ldots$, $Y_n := Y|_{x_n}$ are given respectively by:

$$\widehat{\beta_1} \;=\; \frac{n\left(\displaystyle\sum_{i=1}^{n} x_i Y_i\right) - \left(\displaystyle\sum_{i=1}^{n} x_i\right)\left(\displaystyle\sum_{i=1}^{n} Y_i\right)}{n\left(\displaystyle\sum_{i=1}^{n} x_i^2\right) - \left(\displaystyle\sum_{i=1}^{n} x_i\right)^2} \;=\; \cdots \;=\; \frac{\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})\, Y_i}{\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\widehat{\beta_0} \;=\; \overline{Y} - \widehat{\beta}_1\, \overline{x} \;=\; \left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) - \widehat{\beta}_1\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right), \quad \text{where } \overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i,\ \overline{Y} := \frac{1}{n}\sum_{i=1}^{n} Y|_{x_i} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$\widehat{\sigma^2} \;=\; \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2$$

PROOF

$$-2\log L \;=\; -2\log L(\, \beta_0, \beta_1, \sigma\,;\, y_1, \ldots, y_n \,) \;=\; n \cdot \log(2\pi) + n \cdot \log\!\left(\sigma^2\right) + \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_i\right)^2$$

Hence, setting the partial derivatives of $-2\log L$ with respect to β_0, β_1, and σ^2 to zero yields:

$$
\begin{aligned}
0 &= \frac{\partial(-2\log L)}{\partial\beta_0} &= \frac{2}{\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)(-1) &= -\frac{2}{\sigma^2}\left\{\left(\sum_{i=1}^{n}y_i\right)-n\,\beta_0-\beta_1\left(\sum_{i=1}^{n}x_i\right)\right\} \\
0 &= \frac{\partial(-2\log L)}{\partial\beta_1} &= \frac{2}{\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)(-x_i) &= -\frac{2}{\sigma^2}\left\{\left(\sum_{i=1}^{n}x_i y_i\right)-\beta_0\left(\sum_{i=1}^{n}x_i\right)-\beta_1\left(\sum_{i=1}^{n}x_i^2\right)\right\} \\
0 &= \frac{\partial(-2\log L)}{\partial\sigma^2} &= \frac{n}{\sigma^2}-\frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2
\end{aligned}
$$

Then, $\left(\displaystyle\sum_{i=1}^{n}x_i\right)\times$ (first equation) $-\,n\times$ (second equation) yields:

$$
\beta_1\left\{n\left(\sum_{i=1}^{n}x_i^2\right)-\left(\sum_{i=1}^{n}x_i\right)^2\right\}-\left\{n\left(\sum_{i=1}^{n}x_i y_i\right)-\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)\right\} = 0,
$$

which gives the expression for $\widehat{\beta_1}$. Substituting this expression for $\widehat{\beta_1}$ into the first equation immediately yields the expression for $\widehat{\beta_0}$. Substituting the expressions for $\widehat{\beta_0}$ and $\widehat{\beta_1}$ into the third equation yields that for $\widehat{\sigma^2}$. □

Theorem 3.2 *The following are true:*

- $\widehat{\beta_0}$ *is normally distributed with*

$$
E\left(\widehat{\beta_0}\right)=\beta_0 \quad\text{and}\quad \mathrm{Var}\left(\widehat{\beta_0}\right)=\sigma^2\left(\frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}x_i^2}{\displaystyle\sum_{i=1}^{n}(x_i-\overline{x})^2}\right)
$$

- $\widehat{\beta_1}$ *is normally distributed with*

$$
E\left(\widehat{\beta_1}\right)=\beta_1 \quad\text{and}\quad \mathrm{Var}\left(\widehat{\beta_1}\right)=\sigma^2\left(\frac{1}{\displaystyle\sum_{i=1}^{n}(x_i-\overline{x})^2}\right)
$$

- $\widehat{\beta_1}$, $\widehat{\sigma^2}$, *and* $\overline{Y}:=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}Y_i$ *are mutually independent random variables.*

- ***Corollary*** $\widehat{\sigma^2}$ *and* $\widehat{Y}|_x := \widehat{\beta}_0+\widehat{\beta}_1 x=\overline{Y}+\widehat{\beta}_1(x-\overline{x})$ *are independent random variables.*

- $\widehat{Y}|_x := \widehat{\beta}_0+\widehat{\beta}_1\cdot x$ *is normally distributed with*

$$
E\left(\widehat{Y}|_x\right)=\beta_0+\beta_1\,x \quad\text{and}\quad \mathrm{Var}\left(\widehat{Y}|_x\right)=\sigma^2\left(\frac{1}{n}+\frac{(x-\overline{x})^2}{\displaystyle\sum_{i=1}^{n}(x_i-\overline{x})^2}\right)
$$

- $S^2 := \dfrac{n}{n-2}\cdot\widehat{\sigma^2} = \dfrac{1}{n-2}\displaystyle\sum_{i=1}^{n}\left(Y_i-\widehat{\beta}_0-\widehat{\beta}_1 x_i\right)^2$ *is an unbiased estimator for* σ^2, *and*

  $\left(\dfrac{n}{\sigma^2}\right)\cdot\widehat{\sigma^2} = \dfrac{(n-2)S^2}{\sigma^2} = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}\left(Y_i-\widehat{\beta}_0-\widehat{\beta}_1 x_i\right)^2$ *has a* χ^2*-distribution with* $n-2$ *degrees of freedom.*

References