

**Remark 7.** Example 4 illustrates that one should simplify the model as much as possible in data analysis. **How ?**

The data in Example 4 satisfy the Weibull distribution:

$\ln Y = \beta_o + \beta_1 x + \tau \ln Z$ , with  $Z \sim \text{Exp}(1)$ ,  $(\beta_o, \beta_1, \tau) = (0.1, 1.95, e^{0.07})$  ? Or  $(0, 2, 1)$  ?

The data fit Weibull() better than Exponential() (in CB), but one needs to check whether the data are indeed from Weibull (**how ?**) and try to simplify it.

The more the parameter, the less the efficiency, thus the SD increases if # of parameters increases (it is why *Log(scale)* is significant with p-value  $0.02 \times 2$ ).

	<i>weibull</i>	<i>gamma or exp</i>
$\beta_o$	0.12	0.11
$\beta_1$	0.044	0.042

The simplified model is  $\ln Y = \beta_o + \beta_1 x + \tau \ln Z$ , with  $(\beta_o, \beta_1, \tau) = (0, 2, 1)$  and  $Z \sim \text{Exp}(1)$ , and becomes  $\text{Exp}(\mu)$  with mean  $e^{2x}$ . **How to test it ?**

**Example 5.** (Simulation study). Generate data from  $Y \sim \text{Exp}(\mu)$ , where

$$S(t) = \exp(-\lambda t) = \exp(-\frac{t}{\mu}), t > 0,$$

$$Z = \frac{Y}{\mu} \sim \text{Exp}(1).$$

$$\ln Y = \ln \mu + \ln Z.$$

> library(MASS) ; library(survival) ; n=500 ; y=rexp(n,rate=5);

Then pretend we do not know it. Estimate  $E(Y)$  by the MLE based on Weibull, Exp or Gamma and by the LSE, using various codes:

non-regression:

fitdistr(y,"exponential"), fitdistr(y,"weibull"), fitdistr(y,"gamma")

regression and parametric:

survreg(Surv(y)~ 1),  
 survreg(Surv(y)~ 1, dist="exponential")  
 glm(y~1,family=Gamma(link=log))

semi-parametric:

lm(log(y)~1)  
 lm(y~1)

**Question:** What is the true value that is estimated ?

How many distinct values of the estimates ?

> (z=fitdistr(y,"exponential"))\$e)

rate  $\lambda = 1/\text{scale}$   $S(t) = \exp(-\lambda t)$

5.235007  $\hat{\lambda} = 1/\hat{\mu}$

> 1/z

0.1910217  $\hat{\mu}$

> (z=survreg(Surv(y)~1, dist="exponential"))\$co)  $Y = e^\alpha Z$

-1.655368  $\hat{\alpha}$

> exp(z)

0.1910217  $\hat{\mu} = e^{\hat{\alpha}}$

> predict(z,data.frame(x=0))

0.1910217  $\hat{\mu}$

> z=glm(y~1,family=Gamma(link=log),maxit=50)

> z[[1]] (= link  $\ln \mu = \alpha$ )