(mm), before breaking (x) and the cross–sectional area in square millimeters (mm^2) (y):

(5.28, 52.36) (5.40, 52.58) (4.65, 51.07) (4.76, 52.28) (5.55, 53.02)
(5.73, 52.10) (5.84, 52.61) (4.97, 52.21) (5.50, 52.39) (6.24, 53.77)

(a) Find the equation of the least squares regression line.

(b) Plot the points and the line on the same graph.

(c) Interpret your output.

6.5-10. The "golden ratio" is $\phi = (1+\sqrt{5})/2$. John Putz, a mathematician who was interested in music, analyzed Mozart's sonata movements, which are divided into two distinct sections, both of which are repeated in performance (see References). The length of the "Exposition" in measures is represented by a and the length of the "Development and Recapitulation" is represented by b. Putz's conjecture was that Mozart divided his movements close to the golden ratio. That is, Putz was interested in studying whether a scatter plot of $a+b$ against b not only would be linear, but also would actually fall along the line $y = \phi x$. Here are the data in tabular form, in which the first column identifies the piece and movement by the Köchel cataloging system:

Köchel	a	b	$a+b$	Köchel	a	b	$a+b$
279, I	38	62	100	279, II	28	46	74
279, III	56	102	158	280, I	56	88	144
280, II	24	36	60	280, III	77	113	190
281, I	40	69	109	281, II	46	60	106
282, I	15	18	33	282, III	39	63	102
283, I	53	67	120	283, II	14	23	37
283, III	102	171	273	284, I	51	76	127
309, I	58	97	155	311, I	39	73	112
310, I	49	84	133	330, I	58	92	150
330, III	68	103	171	332, I	93	136	229
332, III	90	155	245	333, I	63	102	165
333, II	31	50	81	457, I	74	93	167
533, I	102	137	239	533, II	46	76	122
545, I	28	45	73	547a, I	78	118	196
570, I	79	130	209				

(a) Make a scatter plot of the points $a+b$ against the points b. Is this plot linear?

(b) Find the equation of the least squares regression line. Superimpose it on the scatter plot.

(c) On the scatter plot, superimpose the line $y = \phi x$. Compare this line with the least squares regression line (graphically if you wish).

(d) Find the sample mean of the points $(a+b)/b$. Is the mean close to ϕ?

6.6* ASYMPTOTIC DISTRIBUTIONS OF MAXIMUM LIKELIHOOD ESTIMATORS

Let us consider a distribution of the continuous type with pdf $f(x;\theta)$ such that the parameter θ is not involved in the support of the distribution. Moreover, we want $f(x;\theta)$ to possess a number of mathematical properties that we do not list here. However, in particular, we want to be able to find the maximum likelihood estimator $\widehat{\theta}$ by solving

$$\frac{\partial[\ln L(\theta)]}{\partial\theta} = 0,$$

where here we use a partial derivative sign because $L(\theta)$ involves $x_1, x_2, \ldots, x_n$, too. That is,

$$\frac{\partial[\ln L(\widehat{\theta})]}{\partial\theta} = 0,$$

where now, with $\widehat{\theta}$ in this expression, $L(\widehat{\theta}) = f(X_1;\widehat{\theta})f(X_2;\widehat{\theta})\cdots f(X_n;\widehat{\theta})$. We can approximate the left-hand member of this latter equation by a linear function found from the first two terms of a Taylor's series expanded about θ, namely,

$$\frac{\partial[\ln L(\theta)]}{\partial\theta} + (\widehat{\theta} - \theta)\frac{\partial^2[\ln L(\theta)]}{\partial\theta^2} \approx 0,$$

when $L(\theta) = f(X_1;\theta)f(X_2;\theta)\cdots f(X_n;\theta)$.

Obviously, this approximation is good enough only if $\widehat{\theta}$ is close to θ, and an adequate mathematical proof involves those conditions, which we have not given here. (See Hogg, McKean, and Craig, 2013.) But a heuristic argument can be made by solving for $\widehat{\theta} - \theta$ to obtain

$$\widehat{\theta} - \theta = \frac{\dfrac{\partial[\ln L(\theta)]}{\partial\theta}}{-\dfrac{\partial^2[\ln L(\theta)]}{\partial\theta^2}}. \tag{6.6-1}$$

Recall that

$$\ln L(\theta) = \ln f(X_1;\theta) + \ln f(X_2;\theta) + \cdots + \ln f(X_n;\theta)$$

and

$$\frac{\partial \ln L(\theta)}{\partial\theta} = \sum_{i=1}^{n} \frac{\partial[\ln f(X_i;\theta)]}{\partial\theta}, \tag{6.6-2}$$

which is the numerator in Equation 6.6-1. However, Equation 6.6-2 gives the sum of the n independent and identically distributed random variables

$$Y_i = \frac{\partial[\ln f(X_i;\theta)]}{\partial\theta}, \quad i = 1, 2, \ldots, n,$$

and thus, by the central limit theorem, has an approximate normal distribution with mean (in the continuous case) equal to

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{\partial[\ln f(x;\theta)]}{\partial\theta} f(x;\theta)\,dx &= \int_{-\infty}^{\infty} \frac{\partial[f(x;\theta)]}{\partial\theta}\frac{f(x;\theta)}{f(x;\theta)}\,dx \\ &= \int_{-\infty}^{\infty} \frac{\partial[f(x;\theta)]}{\partial\theta}\,dx \\ &= \frac{\partial}{\partial\theta}\left[\int_{-\infty}^{\infty} f(x;\theta)\,dx\right] \\ &= \frac{\partial}{\partial\theta}[1] \\ &= 0.\end{aligned}$$

Clearly, we need a certain mathematical condition that makes it permissible to interchange the operations of integration and differentiation in those last steps. Of course, the integral of $f(x;\theta)$ is equal to 1 because it is a pdf.

Since we now know that the mean of each Y is

$$\int_{-\infty}^{\infty} \frac{\partial[\ln f(x;\theta)]}{\partial\theta} f(x;\theta)\,dx = 0,$$

let us take derivatives of each member of this equation with respect to θ, obtaining

$$\int_{-\infty}^{\infty} \left\{ \frac{\partial^2[\ln f(x;\theta)]}{\partial\theta^2} f(x;\theta) + \frac{\partial[\ln f(x;\theta)]}{\partial\theta}\frac{\partial[f(x;\theta)]}{\partial\theta} \right\} dx = 0.$$

However,

$$\frac{\partial[f(x;\theta)]}{\partial\theta} = \frac{\partial[\ln f(x;\theta)]}{\partial\theta} f(x;\theta);$$

so

$$\int_{-\infty}^{\infty} \left\{ \frac{\partial[\ln f(x;\theta)]}{\partial\theta} \right\}^2 f(x;\theta)\, dx = -\int_{-\infty}^{\infty} \frac{\partial^2[\ln f(x;\theta)]}{\partial\theta^2} f(x;\theta)\, dx.$$

Since $E(Y) = 0$, this last expression provides the variance of $Y = \partial[\ln f(X;\theta)]/\partial\theta$. Then the variance of the sum in Equation 6.6-2 is n times this value, namely,

$$-nE\left\{ \frac{\partial^2[\ln f(X;\theta)]}{\partial\theta^2} \right\}.$$

Let us rewrite Equation 6.6-1 as

$$\frac{\sqrt{n}\,(\widehat{\theta} - \theta)}{\left(\dfrac{1}{\sqrt{-E\{\partial^2[\ln f(X;\theta)]/\partial\theta^2\}}} \right)} = \frac{\left(\dfrac{\partial[\ln L(\theta)]/\partial\theta}{\sqrt{-nE\{\partial^2[\ln f(X;\theta)]/\partial\theta^2\}}} \right)}{\left(\dfrac{-\dfrac{1}{n}\dfrac{\partial^2[\ln L(\theta)]}{\partial\theta^2}}{E\{-\partial^2[\ln f(X;\theta)]/\partial\theta^2\}} \right)}. \tag{6.6-3}$$

Since it is the sum of n independent random variables (see Equation 6.6-2),

$$\partial[\ln f(X_i;\theta)]/\partial\theta, \qquad i = 1, 2, \ldots, n,$$

the numerator of the right-hand member of Equation 6.6-3 has an approximate $N(0,1)$ distribution, and the aforementioned unstated mathematical conditions require, in some sense, that

$$-\frac{1}{n}\frac{\partial^2[\ln L(\theta)]}{\partial\theta^2} \quad \text{converge to} \quad E\{-\partial^2[\ln f(X;\theta)]/\partial\theta^2\}.$$

Accordingly, the ratios given in Equation 6.6-3 must be approximately $N(0,1)$. That is, $\widehat{\theta}$ has an approximate normal distribution with mean θ and standard deviation

$$\frac{1}{\sqrt{-nE\{\partial^2[\ln f(X;\theta)]/\partial\theta^2\}}}.$$

Example 6.6-1 (Continuation of Example 6.4-1.) With the underlying exponential pdf

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, \qquad 0 < x < \infty, \qquad \theta \in \Omega = \{\theta : 0 < \theta < \infty\},$$

$\overline{X}$ is the maximum likelihood estimator. Since

$$\ln f(x;\theta) = -\ln\theta - \frac{x}{\theta}$$

and

$$\frac{\partial[\ln f(x;\theta)]}{\partial\theta} = -\frac{1}{\theta} + \frac{x}{\theta^2} \quad \text{and} \quad \frac{\partial^2[\ln f(x;\theta)]}{\partial\theta} = \frac{1}{\theta^2} - \frac{2x}{\theta^3},$$

we have

$$-E\left[\frac{1}{\theta^2} - \frac{2X}{\theta^3}\right] = -\frac{1}{\theta^2} + \frac{2\theta}{\theta^3} = \frac{1}{\theta^2},$$

because $E(X) = \theta$. That is, $\overline{X}$ has an approximate normal distribution with mean θ and standard deviation $\theta/\sqrt{n}$. Thus, the random interval $\overline{X} \pm 1.96(\theta/\sqrt{n})$ has an approximate probability of 0.95 that it covers θ. Substituting the observed $\overline{x}$ for θ, as well as for $\overline{X}$, we say that $\overline{x} \pm 1.96\overline{x}/\sqrt{n}$ is an approximate 95% confidence interval for θ. ■

While the development of the preceding result used a continuous-type distribution, the result holds for the discrete type also, as long as the support does not involve the parameter. This is illustrated in the next example.

Example 6.6-2 (Continuation of Exercise 6.4-3.) If the random sample arises from a Poisson distribution with pmf

$$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, 2, \ldots; \qquad \lambda \in \Omega = \{\lambda : 0 < \lambda < \infty\},$$

then the maximum likelihood estimator for λ is $\widehat{\lambda} = \overline{X}$. Now,

$$\ln f(x;\lambda) = x \ln \lambda - \lambda - \ln x!.$$

Also,

$$\frac{\partial[\ln f(x;\lambda)]}{\partial \lambda} = \frac{x}{\lambda} - 1 \qquad \text{and} \qquad \frac{\partial^2[\ln f(x;\lambda)]}{\partial \lambda^2} = -\frac{x}{\lambda^2}.$$

Thus,

$$-E\left(-\frac{X}{\lambda^2}\right) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda},$$

and $\widehat{\lambda} = \overline{X}$ has an approximate normal distribution with mean λ and standard deviation $\sqrt{\lambda/n}$. Finally, $\overline{x} \pm 1.645\sqrt{\overline{x}/n}$ serves as an approximate 90% confidence interval for λ. With the data in Exercise 6.4-3, $\overline{x} = 2.225$, and it follows that this interval ranges from 1.837 to 2.613. ■

It is interesting that there is another theorem which is somewhat related to the preceding result in that the variance of $\widehat{\theta}$ serves as a lower bound for the variance of every unbiased estimator of θ. Thus, we know that if a certain unbiased estimator has a variance equal to that lower bound, we cannot find a better one, and hence that estimator is the best in the sense of being the minimum-variance unbiased estimator. So, in the limit, the maximum likelihood estimator is this type of best estimator.

We describe this **Rao–Cramér inequality** here without proof. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution of the continuous type with pdf $f(x;\theta)$, $\theta \in \Omega = \{\theta : c < \theta < d\}$, where the support of X does not depend upon θ, so that we can differentiate, with respect to θ, under integral signs like that in the following integral:

$$\int_{-\infty}^{\infty} f(x;\theta)\, dx = 1.$$

If $Y = u(X_1, X_2, \ldots, X_n)$ is an unbiased estimator of θ, then

$$\operatorname{Var}(Y) \geq \frac{1}{n \int_{-\infty}^{\infty} \{[\partial \ln f(x;\theta)/\partial\theta]\}^2 f(x;\theta)\,dx}$$

$$= \frac{-1}{n \int_{-\infty}^{\infty} [\partial^2 \ln f(x;\theta)/\partial\theta^2]\, f(x;\theta)\,dx}.$$

Note that the integrals in the denominators are, respectively, the expectations

$$E\left\{\left[\frac{\partial \ln f(X;\theta)}{\partial\theta}\right]^2\right\} \quad \text{and} \quad E\left[\frac{\partial^2 \ln f(X;\theta)}{\partial\theta^2}\right];$$

sometimes one is easier to compute than the other. Note also that although the Rao–Cramér lower bound has been stated only for a continuous-type distribution, it is also true for a discrete-type distribution, with summations replacing integrals.

We have computed this lower bound for each of two distributions: exponential with mean θ and Poisson with mean λ. Those respective lower bounds were θ^2/n and λ/n. (See Examples 6.6-1 and 6.6-2.) Since, in each case, the variance of $\overline{X}$ equals the lower bound, then $\overline{X}$ is the minimum-variance unbiased estimator.

Let us consider another example.

Example 6.6-3 (Continuation of Exercise 6.4-7.) Let the pdf of X be given by

$$f(x;\theta) = \theta x^{\theta-1}, \qquad 0 < x < 1, \qquad \theta \in \Omega = \{\theta : 0 < \theta < \infty\}.$$

We then have

$$\ln f(x;\theta) = \ln\theta + (\theta - 1)\ln x,$$

$$\frac{\partial \ln f(x;\theta)}{\partial\theta} = \frac{1}{\theta} + \ln x,$$

and

$$\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2} = -\frac{1}{\theta^2}.$$

Since $E(-1/\theta^2) = -1/\theta^2$, the greatest lower bound of the variance of every unbiased estimator of θ is θ^2/n. Moreover, the maximum likelihood estimator $\widehat{\theta} = -n/\ln\prod_{i=1}^{n} X_i$ has an approximate normal distribution with mean θ and variance θ^2/n. Thus, in a limiting sense, $\widehat{\theta}$ is the minimum variance unbiased estimator of θ. ■

To measure the value of estimators, their variances are compared with the Rao–Cramér lower bound. The ratio of the Rao–Cramér lower bound to the actual variance of any unbiased estimator is called the **efficiency** of that estimator. An estimator with an efficiency of, say, 50%, means that $1/0.5 = 2$ times as many sample observations are needed to do as well in estimation as can be done with the minimum variance unbiased estimator (the 100% efficient estimator).