Consider the Bernoulli model with probability of success $\theta$ which was treated in Example 6.2.1. In the example we showed that $1/nI(\theta) = \theta(1-\theta)/n$. From Example 4.1.2 of Section 4.1, the mle of $\theta$ is $\overline{X}$. The mean and variance of a Bernoulli $(\theta)$ distribution are $\theta$ and $\theta(1-\theta)$, respectively. Hence the mean and variance of $\overline{X}$ are $\theta$ and $\theta(1-\theta)/n$, respectively. That is, in this case the variance of the mle has attained the Rao–Cramér lower bound.

We now make the following definitions.

**Definition 6.2.1** (Efficient Estimator). *Let $Y$ be an unbiased estimator of a parameter $\theta$ in the case of point estimation. The statistic $Y$ is called an* **efficient estimator** *of $\theta$ if and only if the variance of $Y$ attains the Rao–Cramér lower bound.*

**Definition 6.2.2** (Efficiency). *In cases in which we can differentiate with respect to a parameter under an integral or summation symbol, the ratio of the Rao–Cramér lower bound to the actual variance of any unbiased estimator of a parameter is called the* **efficiency** *of that estimator.*

**Example 6.2.3** (Poisson($\theta$) Distribution). Let $X_1, X_2, \ldots, X_n$ denote a random sample from a Poisson distribution that has the mean $\theta > 0$. It is known that $\overline{X}$ is an mle of $\theta$; we shall show that it is also an efficient estimator of $\theta$. We have

$$\begin{aligned} \frac{\partial \log f(x;\theta)}{\partial\theta} &= \frac{\partial}{\partial\theta}(x\log\theta - \theta - \log x!) \\ &= \frac{x}{\theta} - 1 = \frac{x-\theta}{\theta}. \end{aligned}$$

Accordingly,

$$E\left[\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right)^2\right] = \frac{E(X-\theta)^2}{\theta^2} = \frac{\sigma^2}{\theta^2} = \frac{\theta}{\theta^2} = \frac{1}{\theta}.$$

The Rao–Cramér lower bound in this case is $1/[n(1/\theta)] = \theta/n$. But $\theta/n$ is the variance of $\overline{X}$. Hence $\overline{X}$ is an efficient estimator of $\theta$. ■

**Example 6.2.4** (Beta($\theta$, 1) Distribution). Let $X_1, X_2, \ldots, X_n$ denote a random sample of size $n > 2$ from a distribution with pdf

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1} & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere,} \end{cases} \tag{6.2.14}$$

where the parameter space is $\Omega = (0, \infty)$. This is the beta distribution, (3.3.5), with parameters $\theta$ and 1, which we denote by beta$(\theta, 1)$. The derivative of the log of $f$ is

$$\frac{\partial \log f}{\partial\theta} = \log x + \frac{1}{\theta}. \tag{6.2.15}$$

From this we have $\partial^2 \log f/\partial\theta^2 = -\theta^{-2}$. Hence the information is $I(\theta) = \theta^{-2}$.



Next, we find the mle of $\theta$ and investigate its efficiency. The log of the likelihood function is

$$l(\theta) = \theta\sum_{i=1}^{n}\log x_i - \sum_{i=1}^{n}\log x_i + n\log\theta.$$

The first partial of $l(\theta)$ is

$$\frac{\partial l(\theta)}{\partial\theta} = \sum_{i=1}^{n}\log x_i + \frac{n}{\theta}. \tag{6.2.16}$$

Setting this to 0 and solving for $\theta$, the mle is $\widehat{\theta} = -n/\sum_{i=1}^{n}\log X_i$. To obtain the distribution of $\widehat{\theta}$, let $Y_i = -\log X_i$. A straight transformation argument shows that the distribution is $\Gamma(1, 1/\theta)$. Because the $X_i$s are independent, Theorem 3.3.2 shows that $W = \sum_{i=1}^{n} Y_i$ is $\Gamma(n, 1/\theta)$. Theorem 3.3.1 shows that

$$E[W^k] = \frac{(n+k-1)!}{\theta^k(n-1)!}, \tag{6.2.17}$$

for $k > -n$. So, in particular for $k = -1$, we get

$$E[\widehat{\theta}] = nE[W^{-1}] = \theta\frac{n}{n-1}.$$

Hence, $\widehat{\theta}$ is biased, but the bias vanishes as $n \to \infty$. Also, note that the estimator $[(n-1)/n]\widehat{\theta}$ is unbiased. For $k = -2$, we get

$$E[\widehat{\theta}^2] = n^2E[W^{-2}] = \theta^2\frac{n^2}{(n-1)(n-2)},$$

and, hence, after simplifying $E(\widehat{\theta}^2) - [E(\widehat{\theta})]^2$, we obtain

$$\mathrm{Var}(\widehat{\theta}) = \theta^2\frac{n^2}{(n-1)^2(n-2)}.$$

From this, we can obtain the variance of the unbiased estimator $[(n-1)/n]\widehat{\theta}$, i.e.,

$$\mathrm{Var}\left(\frac{n-1}{n}\widehat{\theta}\right) = \frac{\theta^2}{n-2}.$$

From above, the information is $I(\theta) = \theta^{-2}$ and, hence, the variance of an unbiased efficient estimator is $\theta^2/n$. Because $\frac{\theta^2}{n-2} > \frac{\theta^2}{n}$, the unbiased estimator $[(n-1)/n]\widehat{\theta}$ is not efficient. Notice, though, that its efficiency (as in Definition 6.2.2) converges to 1 as $n \to \infty$. Later in this section, we say that $[(n-1)/n]\widehat{\theta}$ is asymptotically efficient. ■

In the above examples, we were able to obtain the mles in closed form along with their distributions and, hence, moments. This is often not the case. Maximum likelihood estimators, however, have an asymptotic normal distribution. In fact, mles are asymptotically efficient. To prove these assertions, we need the additional regularity condition given by