You might have noticed that the likelihood function has appeared in many of the results and examples concerning sufficient statistics. There is a relationship between the maximum likelihood estimator $\hat{\theta}$ and a sufficient statistic for an unknown parameter θ, which is given next.

Theorem 2.6 Let $X_1, X_2, \ldots, X_n$ be a random sample from a population described by $f_X(x)$ with unknown parameter θ. If $\hat{\theta}$ is a maximum likelihood estimator of θ that exists uniquely and $Y = g(X_1, X_2, \ldots, X_n)$ is a sufficient statistic for θ, then $\hat{\theta}$ is a function of $Y = g(X_1, X_2, \ldots, X_n)$.

Proof The sufficient statistic $Y = g(X_1, X_2, \ldots, X_n)$ must satisfy

$$L(\theta) = f_X(x_1) f_X(x_2) \ldots f_X(x_n) = h_1(g(x_1, x_2, \ldots, x_n), \theta) \cdot h_2(x_1, x_2, \ldots, x_n)$$

for nonnegative functions h_1 and h_2 by the Fisher–Neyman factorization criteria. Selecting h_1 to be the probability mass function or probability density function of the sufficient statistic, the likelihood function $L(\theta)$ and $h_1(g(x_1, x_2, \ldots, x_n), \theta)$ are maximized at the same θ value, namely $\hat{\theta}$. Furthermore, the maximum likelihood estimator $\hat{\theta}$ must be a function of the sufficient statistic $Y = g(X_1, X_2, \ldots, X_n)$. □

So maximum likelihood estimators are functions of sufficient statistics. This result will be illustrated next for a random sample of n observations drawn from a population that has the Rayleigh distribution, which is a special case of the Weibull distribution with shape parameter 2.

Example 2.32 Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with probability density function

$$f(x) = 2\lambda^2 x e^{-(\lambda x)^2} \qquad x > 0,$$

where λ is a positive unknown parameter.

(a) Find the maximum likelihood estimator $\hat{\lambda}$.

(b) Find a sufficient statistic for λ.

(c) Show that the maximum likelihood estimator is a function of the sufficient statistic.

(a) The likelihood function is

$$L(\lambda) = \prod_{i=1}^{n} f(x_i) = 2^n \lambda^{2n} \left(\prod_{i=1}^{n} x_i \right) e^{-\lambda^2 \sum_{i=1}^{n} x_i^2}.$$

The log likelihood function is

$$\ln L(\lambda) = n \ln 2 + 2n \ln \lambda + \sum_{i=1}^{n} \ln x_i - \lambda^2 \sum_{i=1}^{n} x_i^2.$$

The score is

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = \frac{2n}{\lambda} - 2\lambda \sum_{i=1}^{n} x_i^2.$$

Equating the score to zero and solving for λ yields a closed-form maximum likelihood estimator

$$\hat{\lambda} = \sqrt{\frac{n}{\sum_{i=1}^{n} x_i^2}}.$$

The second partial derivative of the log likelihood function

$$\frac{\partial^2 \ln L(\lambda)}{\partial \lambda^2} = -\frac{2n}{\lambda^2} - 2\sum_{i=1}^{n} x_i^2$$

evaluates to $-4\sum_{i=1}^{n} x_i^2 < 0$ at the maximum likelihood estimator $\hat{\lambda}$, so the maximum likelihood estimator $\hat{\lambda}$ is the unique value that maximizes the log likelihood function.

(b) The likelihood function can be factored as

$$L(\lambda) = \underbrace{\lambda^{2n} e^{-\lambda^2 \sum_{i=1}^{n} x_i^2}}_{h_1} \cdot \underbrace{2^n \left(\prod_{i=1}^{n} x_i\right)}_{h_2},$$

so $Y = \sum_{i=1}^{n} X_i^2$ is a sufficient statistic by the Fisher–Neyman factorization criteria.

(c) The maximum likelihood estimator

$$\hat{\lambda} = \sqrt{\frac{n}{\sum_{i=1}^{n} x_i^2}}$$

is a function of the sufficient statistic $Y = \sum_{i=1}^{n} X_i^2$.

Sufficient statistics can be used in our search for high quality point estimates. The Rao–Blackwell theorem provides a mechanism for combining a sufficient statistic and an unbiased estimator to arrive at a second unbiased estimator with a smaller variance.

Theorem 2.7 (Rao–Blackwell theorem) Let $X_1, X_2, \ldots, X_n$ be a random sample from a population described by $f(x)$, a probability distribution with a single unknown parameter θ. Let $Y_1 = g_1(X_1, X_2, \ldots, X_n)$ be a sufficient statistic for θ and let $Y_2 = g_2(X_1, X_2, \ldots, X_n)$ be an unbiased estimator of θ. Then $E[Y_2 \,|\, Y_1]$ is an unbiased estimator of θ whose variance is less than or equal to the variance of Y_2 when the expectations exist.

Proof Recall from probability theory that for random variables X and Y,

$$E[X] = E\big[E[X \,|\, Y]\big]$$

and

$$V[X] = E\big[V[X \,|\, Y]\big] + V\big[E[X \,|\, Y]\big]$$

when the expectations exist. Letting Y_2 assume the role of X and Y_1 assume the role of Y, the statistic $E[Y_2 \,|\, Y_1]$ must be unbiased because

$$E\big[E[Y_2 \,|\, Y_1]\big] = E[Y_2] = \theta$$