Applications in statistical inference

References:
1. Bickel, D. and Doksum, K. *Mathematical Statistics*
2. Lehman, E. and Casella, G. *Theory of Point Estimation*

Let $X_1, X_2, \ldots$ be an i.i.d sequence such that $E(X) = \mu$. Then, since $\bar{X}_n \xrightarrow[n\to\infty]{P} \mu$, we have that $g(\bar{X}_n) \xrightarrow[n\to\infty]{P} g(\mu)$ for all continuous functions g.

Example. Let $X_1, X_2, \ldots$ be an i.i.d sequence such that $E(X) = \mu$ and $Var(X) = \sigma^2$. Then $\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 \xrightarrow[n\to\infty]{P} \sigma^2$.

Proof. Since $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow[n\to\infty]{P} E(X^2) = \mu^2 + \sigma^2$ and $\bar{X}^2 \xrightarrow[n\to\infty]{P} (E(X))^2 = \mu^2$, we obtain, from Slusky,

$$\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2 \xrightarrow[n\to\infty]{P} \{\mu^2 + \sigma^2\} - \mu^2 = \sigma^2.$$

Example. Let $X_1, X_2, \ldots$ be an i.i.d sequence such that $E(X) = \mu$ and $Var(X) = \sigma^2$. Then from CLT, we have,

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \xrightarrow[n\to\infty]{d} N(0,1).$$

From this we obtain that

$$\left(\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}\right)^2 = \frac{n\left(\bar{X}_n - \mu\right)^2}{\sigma^2} \xrightarrow[n\to\infty]{d} \chi_1^2,$$

since $g(x) = x^2$ is continuous.

Large Sample Optimality

It is desirable to see that statistical procedures have optimality properties as more and more data become available – estimators should be close to the true quantities, their errors become small and so on.

Definition. Let $X_1, X_2, \ldots$ be i.i.d P_θ. An estimator $T_n(X_1, X_2, \ldots, X_n)$ of $q(\theta)$ is said to be consistent (strong) if

$$T_n(X_1, X_2, \ldots, X_n) \xrightarrow[n\to\infty]{P} q(\theta)$$

(strong if convergence is a.s.).

Result. Suppose $q(\theta) = g(\mu_1(\theta), \dots, \mu_k(\theta))$ for $k \geq 1$, where $\mu_r(\theta) = E_\theta(X^r)$, $r = 1, 2, \dots$ and g is continuous. Let the sample moments be $\hat{\mu}_r(\theta) = \frac{1}{n}\sum_{j=1}^n X_j^r$. Then the method of moments estimate $T_n = g(\hat{\mu}_1(\theta), \dots, \hat{\mu}_k(\theta))$ is consistent.

This follows from the fact that

$$(\hat{\mu}_1(\theta), \dots, \hat{\mu}_k(\theta)) \xrightarrow[n\to\infty]{P} (\mu_1(\theta), \dots, \mu_k(\theta))$$

so that

$$g(\hat{\mu}_1(\theta), \dots, \hat{\mu}_k(\theta)) \xrightarrow[n\to\infty]{P} g(\mu_1(\theta), \dots, \mu_k(\theta)).$$

Note that unbiasedness does not imply consistency. For instance, consider i.i.d $X_1, \dots, X_n$ with $E(X) = \mu$ and $Var(X) = \sigma^2 < \infty$. We know then that $T_n = \bar{X}$ is both consistent and unbiased, whereas $U_n = X_1$ is unbiased but not consistent.

To establish the consistency of a given estimator, LLN may not be useful in some situations.

Example. Let $X_1, \dots, X_n$ be i.i.d $U(0, \theta)$. Then $E(X) = \theta/2$. A method of moments estimator is
$\hat{\theta}_1 = \frac{2}{n}\sum_{i=1}^n X_i$, which is consistent from WLLN. Note that

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{P} E(X) = \frac{\theta}{2}.$$

The MLE, however, is different, and is
$\hat{\theta}_2 = X_{(n)}$, a function of the minimal sufficient statistic. Is this consistent? MLE is generally consistent, but regularity conditions apply, so it is easier to prove it directly rather than checking those conditions. One can establish the consistency of $\hat{\theta}_2$ by finding its mean and variance (using the distribution of o.s.; $X_{(n)}/\theta$ is Beta(1, n)) and applying the Chebychev's inequality. Alternatively, since $0 < X_{(n)} < \theta$,

$$\begin{aligned} P(|X_{(n)} - \theta| > \epsilon) &= P(X_{(n)} - \theta < -\epsilon) + P(X_{(n)} - \theta > \epsilon) \\ &= P(X_{(n)} < \theta - \epsilon) = (P(X < \theta - \epsilon))^n \\ &= \left(\frac{\theta - \epsilon}{\theta}\right)^n = \left(1 - \frac{\epsilon}{\theta}\right)^n \\ &\xrightarrow[n\to\infty]{} 0 \end{aligned}$$

for any fixed $\epsilon > 0$. MLE is not always consistent as the following example shows.

Example (Neyman-Scott problem). This problem involves estimating the precision of a measuring device by measuring a large number of different quantities. Suppose two independent measurements each of $\mu_1, \mu_2, \ldots\ldots$ are made. In other words, let $\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_i \\ \mu_i \end{pmatrix}, \sigma^2 I_2\right)$, $i = 1, 2, \ldots$ be independent. Now there is no question of consistent estimators for μ_i since only two observations are available. σ^2 is important for calibration purposes. What is the MLE of σ^2? Note

$$
\begin{aligned}
&f\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \cdots, \begin{pmatrix} x_n \\ y_n \end{pmatrix}\right) = \\
&\qquad (2\pi)^{-n}(\sigma^2)^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(x_i-\mu_i)^2+(y_i-\mu_i)^2\right]\right).
\end{aligned}
$$

Fix σ^2. Since X_i and Y_i are i.i.d $N(\mu_i, \sigma^2)$,

$$\hat{\mu}_i = \frac{1}{2}(X_i + Y_i), i = 1, 2, \ldots \text{ independent of } \sigma^2 \; .$$

Therefore,

$$
\begin{aligned}
&\max_{\{\mu_i\},\sigma^2} L\left(\{\mu_i\}, \sigma^2, (\mathbf{x}, \mathbf{y})\right) = \max_{\sigma^2} L\left(\{\hat{\mu}_i\}, \sigma^2, (\mathbf{x}, \mathbf{y})\right) \\
&= \max_{\sigma^2} (\sigma^2)^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(x_i-\hat{\mu}_i)^2+(y_i-\hat{\mu}_i)^2\right]\right).
\end{aligned}
$$

Thus

$$
\begin{aligned}
\hat{\sigma^2} &= \frac{1}{2n}\sum_{i=1}^{n}\left[(x_i-\hat{\mu}_i)^2+(y_i-\hat{\mu}_i)^2\right] \\
&= \frac{1}{2n}\sum_{i=1}^{n} 2\left(x_i - \frac{1}{2}(x_i+y_i)\right)^2 = \frac{1}{4n}\sum_{i=1}^{n}(x_i-y_i)^2.
\end{aligned}
$$

Now note that $X_i - Y_i \sim N(0, 2\sigma^2)$ i.i.d., so

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2 \xrightarrow[n\to\infty]{P} E(X-Y)^2 = 2\sigma^2, \text{ and hence}$$

$$\frac{1}{4n}\sum_{i=1}^{n}(X_i - Y_i)^2 \xrightarrow[n\to\infty]{P} \frac{\sigma^2}{2}.$$

This shows that the MLE of σ^2 is inconsistent here. Consistent estimators are easily available, however. For example, $\frac{1}{2n}\sum_{i=1}^{n}(X_i - Y_i)^2$ is a consistent estimator. The problem with the MLE here is that it considers the problem of estimating the infinite sequence of μ_i in addition to σ^2 in the limit as n grows.

Example. $X_1, \ldots, X_n$ i.i.d Bernoulli(p). Estimate p. Then $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}_n$ is MLE, method of moments as well as UMVUE. By WLLN, $\hat{p} = \bar{X}_n \xrightarrow[n\to\infty]{P} p$, thus showing that it is consistent. What if we want to estimate $q(p) = p(1-p)$? Then $\hat{q}(p) = \hat{p}(1-\hat{p})$ is MLE or method of moments estimator. Since $q(x) = x(1-x)$ is a continuous function and $\hat{p} \xrightarrow[n\to\infty]{P} p$, we have that $\hat{q}(p) = \hat{p}(1-\hat{p})$ is consistent for $q(p) = p(1-p)$. How good is this estimator?