Parameter Estimation

Samples from a probability distribution $F(x)$ are: $x = [x_1, x_2, \ldots, x_n]^T$. The probability distribution has a parameter vector $\theta = [\theta_1, \theta_2, \ldots, \theta_m]^T$.

**Estimator:** Statistic used to estimate unknown $\theta$.

**Estimate:** Observed value of the estimator.

**Maximum Likelihood Estimator**

The likelihood for independent samples $x$ is defined as

$$L(x; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

The maximum likelihood estimator is defined as

$$\hat{\theta}_{ML} = \arg \max_{\theta} L(x; \theta)$$

To estimate the value of $\theta$ that maximizes $L$ or equivalently $\ln L$ we will set

$$\frac{\partial \ln L}{\partial \theta_i} = 0, \quad \text{for } i = 1, 2, \ldots, m$$

**Example 1.**

For Bernoulli distribution,

$$P(X = x) = p^x (1 - p)^{1-x}$$

Hence, among $n$ observations, the likelihood is defined as

$$L(x; p) = \prod_{i=1}^{n} p^{x_i} (1 - p)^{1-x_i}$$

$$= p^{\sum x_i} (1 - p)^{n - \sum x_i}$$

$$= p^{\bar{x}n} (1 - p)^{n(1 - \bar{x})}$$

The log-likelihood is

$$\ln L = n \bar{x} \ln p + n (1 - \bar{x}) \ln (1 - p)$$
Taking derivative with respect to the parameter $p$

$$\frac{d \ln L}{dp} = \frac{n \bar{x}}{p} - n \frac{(1 - \bar{x})}{1 - p} = 0$$

$$(1 - p)\bar{x} - (1 - \bar{x})p = 0$$

$$\Rightarrow \hat{p} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Hence, the ML estimator is $\hat{p} = \bar{x}$

**Example 2.**

For Poisson distribution

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Hence, among $n$ observations, the likelihood is defined as

$$L(x; \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{(-\lambda)}$$

$$= \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{(-n\lambda)}$$

The log-likelihood is

$$\ln L = n \bar{x} \ln \lambda - n \lambda - \sum_{i=1}^{n} \ln(x_i!)$$

Taking derivative with respect to the parameter $\lambda$

$$\frac{d \ln L}{d\lambda} = \frac{n \bar{x}}{\lambda} - n = 0$$

$$\Rightarrow \hat{\lambda} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Hence, the ML estimator is $\hat{\lambda} = \bar{x}$

**Example 3.**

For Gaussian distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right]$$

Hence, among $n$ observations, the likelihood is defined as

$$L(x; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x_i - \mu)^2}{2\sigma^2} \right]$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[ -\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} \right]$$

Subhayan De
The log-likelihood is
\[
\ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}
\]

Taking derivative with respect to the parameter \(\mu\)
\[
\frac{\partial \ln L}{\partial \mu} = - \sum_{i=1}^{n} \frac{(x_i - \mu)}{\sigma^2} = 0
\]
\[
\Rightarrow \hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}
\]

Hence, the ML estimator is \(\hat{\mu} = \bar{x}\).

Taking derivative with respect to the parameter \(\sigma^2\)
\[
\frac{\partial \ln L}{\partial (\sigma^2)} = - \frac{n}{2\sigma^2} + \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^4} = 0
\]
\[
\Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=0}^{n} (x_i - \hat{\mu})^2
\]

Hence, the ML estimator is \(\hat{\sigma}^2 = \frac{1}{n} \sum_{i=0}^{n} (x_i - \hat{\mu})^2\).

**Example 4.**

For Gamma distribution
\[
f(x) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x}
\]

Hence, among \(n\) observations, the likelihood is defined as
\[
L(x; \alpha, \lambda) = \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)} \lambda^\alpha x_i^{\alpha-1} e^{-\lambda x_i}
\]
\[
= \frac{1}{\Gamma(\alpha)^n} \lambda^{n\alpha} (\prod_{i=1}^{n} x_i^{\alpha-1}) e^{-\lambda \sum_{i=1}^{n} x_i}
\]

The log-likelihood is
\[
\ln L = (\alpha - 1) \sum_{i=1}^{n} \ln x_i - \lambda \sum_{i=1}^{n} x_i + (n\alpha) \ln \lambda - n \ln \Gamma(\alpha)
\]

Taking derivative with respect to the parameter \(\lambda\)
\[
\frac{\partial \ln L}{\partial \lambda} = - \sum_{i=1}^{n} x_i + \frac{n\alpha}{\lambda} = 0
\]
\[
\Rightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}
\]

Subhayan De
Hence, the ML estimator is $\hat{\lambda} = \frac{\hat{\alpha}}{n \sum_{i=1}^{n} x_i}$.

Taking derivative with respect to the parameter $\alpha$

$$\frac{\partial \ln L}{\partial \alpha} = \sum_{i=1}^{n} \ln x_i + n \ln \lambda - \frac{n \Gamma'(\alpha)}{\Gamma(\alpha)} = 0$$

$$\Rightarrow \ln \hat{\alpha} - \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = \ln \left( \frac{1}{n} \sum_{i=1}^{n} x_i \right) - \frac{1}{n} \sum_{i=1}^{n} \ln x_i$$

This is a nonlinear equation needed to be solved to get $\hat{\alpha}$.

**Example 5.**

If the observations \{0.3, 0.2, 0.5, 0.8, 0.9\} are obtained from a distribution with $f(x) = \theta x^{\theta - 1}$, $x \geq 0$ then estimate the value of $\theta$ using Maximul Likelihood method.

The likelihood is defined as

$$L(x; \theta) = \prod_{i=1}^{5} \theta x_i^{\theta - 1}$$

The log likelihood is

$$\ln L = 5 \ln \theta + (\theta - 1) \sum_{i=1}^{5} \ln x_i$$

Taking derivative of $\ln L$ with respect to $\theta$

$$\frac{\partial \ln L}{\partial \theta} = \frac{5}{\theta} + \sum_{i=1}^{5} \ln x_i = 0$$

$$\Rightarrow \hat{\theta} = \frac{5}{\sum_{i=1}^{5} \ln x_i} = 1.3038$$

**Example 6.**

For Uniform distribution in $(0, \theta)$

$$f(x) = \frac{1}{\theta}, \quad 0 < x < \theta$$

Hence, among $n$ observations, the likelihood is defined as

$$L(x; \theta) = \prod_{i=1}^{n} \frac{1}{\theta} = \frac{1}{\theta^n}$$

The log-likelihood is

$$\ln L = -n \ln \theta$$

This is maximized when $\theta$ is minimum but $\theta \geq \max(x_1, x_2, \ldots, x_n)$. Hence, the ML estimator is $\hat{\theta} = \max(x_1, x_2, \ldots, x_n)$.

Subhayan De
Interval Estimate

Let \( X_1, X_2, \ldots, X_n \) are samples from a Gaussian distribution with mean \( \mu \) and variance \( \sigma^2 \). The point estimator \( \bar{X} \) is Gaussian with mean \( \mu \) and variance \( \sigma^2/n \). Hence,

\[
P \left( -1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96 \right) = 0.95
\]

\[
P \left( \bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right) = 0.95
\]

Based on the observations, with 95\% we can say that the population mean \( \mu \) lies within the interval \( (\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}) \) — known as the 95 percent confidence interval estimate of \( \mu \).

In general, 100(1 - \( \alpha \)) percent two-sided confidence interval for \( \mu \) is \( (\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) \).

One-sided upper confidence interval for \( \mu \) is \( (\bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}}, +\infty) \).

One-sided lower confidence interval for \( \mu \) is \( (-\infty, \bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}}) \).

**Sample size:** If we want the 100(1 - \( \alpha \)) percent two-sided confidence interval for \( \mu \) to be within \( (\bar{x} \pm \Delta x) \) we need a sample size \( n = \left( \frac{2z_{\alpha/2}\sigma}{\Delta x} \right)^2 \).

**Quick reference:** 100(1 - \( \alpha \))\% two-sided confidence interval:

- 90\% confidence: \( \alpha = 0.10 \), \( z_{0.05} = 1.65 \)
- 95\% confidence: \( \alpha = 0.05 \), \( z_{0.025} = 1.96 \)
- 98\% confidence: \( \alpha = 0.02 \), \( z_{0.01} = 2.33 \)
- 99\% confidence: \( \alpha = 0.01 \), \( z_{0.005} = 2.58 \)

Similarly, the following Table shows a variety of cases for samples from a normal population:

Note that, \( s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \).

<table>
<thead>
<tr>
<th>Case</th>
<th>Parameter</th>
<th>Confidence interval</th>
<th>Lower interval</th>
<th>Upper interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma^2 ) known</td>
<td>( \mu )</td>
<td>( (\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) )</td>
<td>( (-\infty, \bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}}) )</td>
<td>( (\bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}}, +\infty) )</td>
</tr>
<tr>
<td>( \sigma^2 ) unknown</td>
<td>( \mu )</td>
<td>( (\bar{x} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}) )</td>
<td>( (-\infty, \bar{x} + t_{\alpha,n-1} \frac{s}{\sqrt{n}}) )</td>
<td>( (\bar{x} - t_{\alpha,n-1} \frac{s}{\sqrt{n}}, +\infty) )</td>
</tr>
<tr>
<td>( \mu ) unknown</td>
<td>( \sigma^2 )</td>
<td>( \left( \frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \right) )</td>
<td>( (0, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}) )</td>
<td>( \left( \frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, +\infty \right) )</td>
</tr>
</tbody>
</table>

**Example 6.**

Estimate the sample size needed for mean to be within ±0.25 where \( \sigma = 2 \) and a confidence interval of 95\%.

Subhayan De
Example 7.

The lifetime \( X \) of light bulbs are exponentially distributed. Based on observation of 81 light bulbs we obtain their average lifetime is 200 hours. Estimate the 95% confidence interval for the mean lifetime.

For exponentially distributed random variable \( X \),

\[ f(x) = \lambda e^{-\lambda x} \]

The mean of \( X \) is \( 1/\lambda \) and variance is \( 1/\lambda^2 \). For large number of samples \( n \), the sample mean is Gaussian with mean \( 1/\lambda \) and variance \( \frac{1}{n\lambda^2} \).

Hence, we can write

\[
P \left( -1.96 < \frac{\bar{X} - \frac{1}{\lambda}}{\frac{1}{\lambda\sqrt{n}}} < 1.96 \right) = 0.95
\]

\[
P \left( \frac{1}{\lambda} - 1.96 \frac{1}{\lambda\sqrt{n}} < \bar{X} < \frac{1}{\lambda} + 1.96 \frac{1}{\lambda\sqrt{n}} \right) = 0.95
\]

\[
P \left\{ \frac{1}{\lambda} \left( 1 - \frac{1.96}{\sqrt{n}} \right) < \bar{X} < \frac{1}{\lambda} \left( 1 + \frac{1.96}{\sqrt{n}} \right) \right\} = 0.95
\]

\[
P \left( \frac{\bar{X}}{1 + 1.96/\sqrt{n}} < \frac{1}{\lambda} < \frac{\bar{X}}{1 - 1.96/\sqrt{n}} \right) = 0.95
\]

Hence, the 95% confidence interval for the mean lifetime of the bulbs is \( \frac{200}{1 + 1.96/\sqrt{81}} < \frac{1}{\lambda} < \frac{200}{1 - 1.96/\sqrt{81}} \) or 164 < \( \frac{1}{\lambda} < 256 \).

Example 8.

For Poisson distributed random variable get the \( 100(1 - \alpha) \) confidence interval.

The p.m.f. is given by

\[ P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!} \]

The mean \( \mathbb{E}[X] = \lambda = \text{Var}(X) \). Hence, for large \( n \) \( \bar{X} \) is approximately Gaussian with mean \( \lambda \)

Subhayan De
and variance $\lambda/n$. This helps in writing

$$P \left( \bar{X} - 1.96 \sqrt{\frac{\lambda}{n}} < \lambda < \bar{X} + 1.96 \sqrt{\frac{\lambda}{n}} \right) = 0.95$$

$$P \left( |\bar{X} - \lambda| < 1.96 \sqrt{\frac{\lambda}{n}} \right) = 0.95$$

$$P \left\{ (\bar{X} - \lambda)^2 < \frac{(1.96)^2}{n} \lambda \right\} = 0.95$$

Therefore, the confidence interval is the two solutions of the following quadratic equation

$$(\bar{x} - \lambda)^2 = \frac{(1.96)^2}{n} \lambda$$