Maximum Likelihood Estimation (MLE)

Plan for today

In first 25 minutes

- Brief review
- MLE definition and motivation
- Simple examples with analytic form
- A slightly more complex example requiring a numerical solution

In next 25 minutes (will not be presented)

- Further examples
- Overview of properties of MLEs:
  - Invariance
  - Asymptotic consistency
  - Asymptotic normality (yields asymptotic confidence intervals)
  - Asymptotic efficiency (Cramer-Rao inequality)

Brief Review

- A random sample $\underline{X} = \Big(X_1, \ldots, X_n\Big)$ with postulated density: $f(x \mid \theta)$
- The joint n-dimensional density of the random sample: $\prod_{i=1}^{n} f(x_i \mid \theta)$.
- $\theta \in \Theta$ is the unknown parameter of interest
- An estimator, $\hat{\theta}(\underline{X})$ is a statistic, hopefully with some desirable properties:
    - Unbiased
    - Consistent
    - Low MSE
    - Easy to compute
    - More to come (later in the subject)...
- One general method of estimation: Method of moments
- We now move to another method, central in statistics, MLE...

MLE: Definition and Motivation

Definition

Likelihood function: $L(\theta \,|\, \underline{X}) = \prod_{i=1}^{n} f(X_i \,|\, \theta)$

Maximum Likelihood Estimator (MLE): $\hat{\theta}(\underline{X}) = \text{argmax}_{\theta \in \Theta} L(\theta \,|\, \underline{X})$

Observe:

- $L(\theta \,|\, \underline{X})$ is a "random function" of θ, determined by the values of $\underline{X}$
- $\hat{\theta}(\underline{X})$ is a statistic (random variable)
- We assume here: The MLE exists and is unique

Motivation

Given the observed data, $\underline{X}$, choose $\hat{\theta}$ that is the **most likely** value of θ

Key Strengths

Generic, asymptotic properties, computable for complex models

Note: Often easier to maximize the *log-likelihood*: $\ell(\theta \,|\, \underline{X}) = \log L(\theta \,|\, \underline{X})$

Example 1: Counts of Events

Reminder:

$$\hat{\theta}(\underline{X}) = \text{argmax}_{\theta \in \Theta} \prod_{i=1}^{n} f(X_i \mid \theta) = \text{argmax}_{\theta \in \Theta} \sum_{i=1}^{n} \log\left(f(X_i \mid \theta)\right)$$

Assume a Poisson distribution: $f(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$, so θ is λ

$$\hat{\lambda} = \text{argmax}_{\lambda \in (0,\infty)} \sum_{i=1}^{n} \log\left(e^{-\lambda}\frac{\lambda^{X_i}}{X_i!}\right)$$

Optimize the log-likelihood with respect to λ (solve $\frac{d\ell(\lambda|\underline{X})}{d\lambda} = 0$):

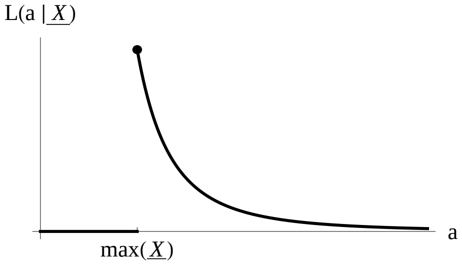
$$\text{MLE: } \hat{\lambda}(\underline{X}) = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$$

Example 2: Uniform Distribution on $[0, a]$

$f(x|a) = \frac{1}{a}\mathbf{1}_{\{x\in[0,a]\}}$, thus

$$L(a\,|\,\underline{X}) = \frac{1}{a^n}\prod_{i=1}^{n}\mathbf{1}_{\{X_i\in[0,a]\}} = \frac{1}{a^n}\mathbf{1}_{\{0\leq\min(\underline{X})\}}\ \mathbf{1}_{\{\max(\underline{X})\leq a\}}$$

Now for a given sample, $\underline{X}$, plot $L(a\,|\,\underline{X})$:



MLE: $\hat{a}(\underline{X}) = \max(\underline{X})$

Example 3: A Mixture Distribution

Data from a Consulting Project: Defects in Production

- Defect counts collected in 10 minute intervals over 36 hours, (216 data points). Many corrupt data points (known machine breakdowns etc...), actual data points, $n = 193$
- The Poisson distribution is a natural choice, but the data "says" otherwise:

$$\begin{aligned} \underline{x}_{\text{sorted}} &= \big(58, 45, 43, 33(2), 32, 27, 16, 15, 14, 10, 7, \\ &\quad 6(2), 5, 4(3), 3(6), 2(6), 1(31), 0(132)\big) \\ \overline{x} &= 2.19 \\ S^2 &= 58.97 \end{aligned}$$

- Thought: The machine is in "good" operating condition for a fraction p of the time and is in "bad" operating condition in the remaining time. How about this model:

$$f(x \mid p, \lambda_1, \lambda_2) = pe^{-\lambda_1} \frac{x^{\lambda_1}}{x!} + (1 - p)e^{-\lambda_2} \frac{x^{\lambda_2}}{x!}$$

- The parameter: $\theta = (p, \lambda_1, \lambda_2) \in [0, 1] \times (0, \infty) \times (0, \infty)$

Example 3 (cont.): A Mixture Distribution

MLE for the Poisson Mixture Model, $\theta = (p, \lambda_1, \lambda_2)$

$$f(x \mid p, \lambda_1, \lambda_2) = pe^{-\lambda_1}\frac{x^{\lambda_1}}{x!} + (1-p)e^{-\lambda_2}\frac{x^{\lambda_2}}{x!}$$

$$\ell(p, \lambda_1, \lambda_2) = \sum_{i=1}^{n} \log\left(pe^{-\lambda_1}\frac{X_i^{\lambda_1}}{X_i!} + (1-p)e^{-\lambda_2}\frac{X_i^{\lambda_2}}{X_i!}\right)$$

Maximization of $\ell(p, \lambda_1, \lambda_2)$ over $[0,1] \times (0,\infty) \times (0,\infty)$ is not trivial.... But with computational software it is doable. Here (for example) is Mathematica code:

In : NMaximize[{logLike[defectCounts], {0 < p, p < 1, 0 < λ1, 0 < λ2}}, {p, λ1, λ2}]

Out : {−318.515, {p → 0.942881, λ1 → 0.532122, λ2 → 29.5869}}

Moving on (2'nd 25 minutes)...

- Further examples
- Overview of properties of MLEs:
  - Invariance
  - Asymptotic consistency
  - Asymptotic normality (yields asymptotic confidence intervals)
  - Asymptotic efficiency (Cramer-Rao inequality)