PH 712 Probability and Statistical Inference

Part IV: Point Estimation I

Zhiyang Zhou (zhou67@uwm.edu, zhiyanggeezhou.github.io)

2025/Oct/29 10:48:12

Framework of statistical inference/learning

- Goal: infer/learn the distribution of RV X, say f_X, from a random sample $X_1, \ldots, X_n$
- Assumption: $f_X \approx \hat{f}_X$ (statistical model)
  - E.g., $\hat{f}_X = \mathcal{N}(\mu, \sigma^2)$, reducing the task to estimating (μ, σ)
- Point estimation: make the "best" guess about unknown parameter(s)
  - E.g., estimate (μ, σ) by $(\hat{\mu}, \hat{\sigma})$
- Hypothesis testing
  - E.g., confirm whether $\mu = 0$ by testing $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$
- Interval estimation: construct an interval likely to cover the unknown parameter
  - E.g., construct an interval, say (c_1, c_2), such that $c_1 < \mu < c_2$ with a high probability

Point estimation

- θ: the unknown parameter
  - A unknown scalar (i.e., we only consider cases with one unknown parameter)
- The generation of a guess on the value of θ based on the random sample $X_1, \ldots, X_n$
- Estimator: the generated guess, say $\hat{\theta}$
  - A statistic (why?) and hence an RV
  - E.g., sometimes, $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ (sample mean) is an estimator of certain parameter θ
- Estimate: plugging the realization of the random sample, say $x_1, \ldots, x_n$, into the estimator
  - A number (why?) and NOT randomized
  - E.g., $n^{-1} \sum_{i=1}^n x_i$ is an estimate of certain parameter θ

Maximum Likelihood (ML) Estimator (MLE)

- Θ: the set of allowed values of θ
- Likelihood function: an alias of joint pdf/pmf

$$L(\theta) = L(\theta; X_1, \ldots, X_n) = f_{X_1,\ldots,X_n}(X_1, \ldots, X_n \mid \theta), \quad \theta \in \Theta$$

  - $f_{X_1,\ldots,X_n}$: the joint pdf/pmf of $X_1, \ldots, X_n$

- Log-likelihood function: the natural logarithm of likelihood function

$$\ell(\theta) = \ln L(\theta), \quad \theta \in \Theta$$

- $\hat{\theta}_{\mathrm{ML}}$ is the MLE for θ if $\hat{\theta}_{\mathrm{ML}}$ is the maximizer of $L(\theta)$ (equiv. the maximizer of $\ell(\theta)$) with respect to θ constrained in Θ
  - In the math notation,

$$\hat{\theta}_{\mathrm{ML}} = \arg\max_{\theta \in \Theta} L(\theta) = \arg\max_{\theta \in \Theta} \ell(\theta)$$

  - That is to say, $L(\hat{\theta}_{\mathrm{ML}}) \geq L(\theta)$ and $\ell(\hat{\theta}_{\mathrm{ML}}) \geq \ell(\theta)$, for all $\theta \in \Theta$.
- Invariance property of MLE: if $\hat{\theta}_{\mathrm{ML}}$ is the MLE of θ, then $g(\hat{\theta}_{\mathrm{ML}})$ is the MLE of $g(\theta)$ for any given function $g(\cdot)$.

How to locate the ML estimator (MLE) constrained in Θ?

- If $L(\theta)$ (or equiv. $\ell(\theta)$) is monotonic with respect to $\theta \in \Theta$, then the MLE lies at one boundary point of Θ
- If $\ell(\theta)$ is non-monotonic but differentiable with respect to $\theta \in \Theta$, then
  1. Collect all the candidates including:
     - Stationary points, i.e., solutions to the equation $S(\theta) = 0$ subject to $\theta \in \Theta$
       * Where $S(\theta) = \ell'(\theta)$ is called the score/gradient
     - Boundary points of Θ
  2. Compare the values of log-likelihood or likelihood evaluated at all the above candidates

How to locate the ML estimate constrained in Θ?

- Reachable only when the realization of $X_1, \ldots, X_n$ are available
- Theoretical way: figuring out the MLE before plugging the realization of $X_1, \ldots, X_n$ into the MLE
- Numerical way: `R` function `optim()`

Example Lec4.1

- Suppose $X_1, \ldots, X_n$ is an iid sample following $\mathcal{N}(\mu, \sigma^2)$, i.e., $f_{X_i}(x \mid \theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$, $x \in \mathbb{R}$, with unknown μ and known $\sigma = 1$. The MLE of μ is $\hat{\mu}_{\mathrm{ML}} = \bar{X} = n^{-1}\sum_{i=1}^n X_i$.
  - If the realization of the sample is $1, \ldots, 10$, find the ML estimate of μ.

```r
sample = 1:10
ell = function(mu){
  n = length(sample)
  sigma = 1   # known
  -n/2*log(2*pi*sigma^2) - sum((sample - mu)^2)/(2*sigma^2)
}
optim(par = 0,
  lower = -Inf, upper = Inf,
  fn=ell, method="L-BFGS-B",
  control=list(fnscale=-1))$par
```

- Suppose $X_1, \ldots, X_n$ is an iid sample following $\mathcal{N}(\mu, \sigma^2)$, i.e., $f_{X_i}(x \mid \theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$, $x \in \mathbb{R}$, with known $\mu = 5$ and unknown $\sigma > 0$. The MLE of σ is $\hat{\sigma}_{\mathrm{ML}} = \sqrt{n^{-1}\sum_{i=1}^n (X_i - \mu)^2}$.
  - If the realization of the sample is $1, \ldots, 10$, find the ML estimate of σ.

```r
sample = 1:10
ell = function(sigma){
  n = length(sample)
  mu = 5  # known
  -n/2*log(2*pi*sigma^2) - sum((sample - mu)^2)/(2*sigma^2)
}
optim(par = 10,
  lower = 0.00001, upper = Inf,
  fn=ell, method="L-BFGS-B",
  control=list(fnscale=-1))$par
```

- Suppose $X_1, \dots, X_n$ is an iid sample following $p_{X_i}(x \mid \theta) = \theta^x (1-\theta)^{1-x} \mathbf{1}_{\{0,1\}}(x)$, $\theta \in [0, 1/2]$. The MLE of θ is $\hat{\theta}_{\text{ML}} = \min\{\bar{X}, 1/2\}$.
  - If the realization of the sample is $0, 1, 1, 1, 0$, find the ML estimate of θ.

```r
sample = c(0,1,1,1,0)
ell = function(theta){
  n = length(sample)
  sum(sample)*log(theta) + (n - sum(sample))*log(1 - theta)
}
res = optim(par = .15,
  lower = 0.00001, upper = 1/2,
  fn=ell, method="L-BFGS-B",
  control=list(fnscale=-1))
res$par
```

- Suppose $X_1, \dots, X_n$ is an iid sample following an exponential distribution, i.e., $f_X(x \mid \beta) = \beta^{-1} \exp(-x/\beta) \mathbf{1}_{(0,\infty)}(x)$, $\beta > 0$. The MLE of β is $\hat{\beta}_{\text{ML}} = \bar{X}$.
  - If the realization of the sample is $1, \dots, 10$, find the ML estimate of β.

- Suppose $X_1, \dots, X_n$ is an iid sample following a beta distribution, i.e., $f_X(x \mid \theta) = \theta x^{\theta-1} \mathbf{1}_{[0,1]}(x)$, $\theta > 0$. The MLE of θ is $\hat{\theta}_{\text{ML}} = -n / \sum_{i=1}^n \ln X_i$.
  - If the realization of the sample is $0.1, \dots, 0.9$, find the ML estimate of θ.

```r
sample = (1:9)/10
ell = function(theta){
  n = length(sample)
  n*log(theta) + (theta - 1)*log(prod(sample))
}
res = optim(par = 10,
  lower = 0.00001, upper = Inf,
  fn=ell, method="L-BFGS-B",
  control=list(fnscale=-1))
res$par
```

- The simplest linear model (or linear regression) is a collection of independent random variables $Y_1, \dots, Y_n$ such that

$$Y_i = \beta x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

where $x_1, \ldots, x_n$ are nonrandomized, and $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{iid}}{\sim} f_\varepsilon(t) = \sqrt{2\pi}\exp(-t^2/2)$ (i.e., $\mathcal{N}(0,1)$). The MLE of β is $\hat{\beta}_{\text{ML}} = \sum_i x_i Y_i / \sum_i x_i^2$.

- Suppose x-values are $1, \ldots, 10$. Correspondingly, observed Y-values are $2, \ldots, 11$. Find the ML estimate of β. (Hint: create the likelihood by noting that $Y_i \sim \mathcal{N}(\beta x_i, 1)$.)

```
x_vals = 1:10
y_vals = 2:11
ell = function(beta){
  n = length(x_vals)
  -n/2*log(2*pi)-sum((y_vals-x_vals*beta)^2)/2
}
res = optim(par = 0,
  lower = -Inf, upper = Inf,
  fn=ell, method="L-BFGS-B",
  control=list(fnscale=-1))
res$par
```