PH 712 Probability and Statistical Inference

Part VII: Point Estimation II

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

2025/Nov/10 16:57:48

Consistency of MLE

- For an iid sample, under certain conditions:
  - $\hat{\theta}_{\mathrm{ML}} \approx \theta$ as $n \to \infty$

Example of inconsistent MLE

There are independent $X_{i1}, X_{i2} \sim \mathcal{N}(\mu_i, \sigma^2)$, $i = 1, \ldots, n$. Then $\widehat{\sigma^2}_{\mathrm{ML}}$ is NOT consistent for σ^2.

Asymptotic efficiency of MLE

- For an iid sample, under certain conditions (more than those for consistency):
  - $\sqrt{n}(\hat{\theta}_{\mathrm{ML}} - \theta) \approx \mathcal{N}(0, I_1^{-1}(\theta))$ as $n \to \infty$
    - For an iid sample, $I_1(\theta) = n^{-1} I_n(\theta)$, no longer a function of n

Implications of consistency and asymptotic efficiency of MLE

- As $n \to \infty$,
  - $\mathrm{E}(\hat{\theta}_{\mathrm{ML}}) \approx \theta$ (asymptotically unbiased)
  - $\mathrm{var}(\hat{\theta}_{\mathrm{ML}}) \approx I_n^{-1}(\theta)$ (asymptotically attains CRLB)
  - $\mathrm{MSE}(\hat{\theta}_{\mathrm{ML}}) \approx I_n^{-1}(\theta)$
- Promoting the use of MLE in practice
- Approximating the distribution of MLE
  - As $n \to \infty$, $\hat{\theta}_{\mathrm{ML}} - \theta \approx \mathcal{N}(0, I_n^{-1}(\theta))$ or equiv. $\hat{\theta}_{\mathrm{ML}} \approx \mathcal{N}(\theta, I_n^{-1}(\theta))$
- Unknown Fisher information $I_n(\theta)$? Recall that $I_n(\theta)$ may be approximated by the observed Fisher information $\hat{I}_n(\hat{\theta}_{\mathrm{ML}})$ (as mentioned in Part V when introducing CRLB)
  - Observed Fisher information (i.e., the minus Hessian evaluated at $\hat{\theta}_{\mathrm{ML}}$): $\hat{I}_n(\hat{\theta}_{\mathrm{ML}}) = -\ell''(\hat{\theta}_{\mathrm{ML}})$

Example Lec7.1

- Suppose $X_1, \ldots, X_n$ is an iid sample following $p_{X_i}(x \mid \theta) = \theta^x (1-\theta)^{1-x} \mathbf{1}_{\{0,1\}}(x)$, $\theta \in [0, 1/2]$.
  - Observing the sample $0, 1, 1, 1, 0$, give the approximate distribution of MLE $\hat{\theta}_{\mathrm{ML}}$.

Ans: $\mathcal{N}(\theta, .01)$ (bootstrap) or $\mathcal{N}(\theta, .05)$ (CRLB).

```r
xs = c(0,1,1,1,0)
n = length(xs)
ell = function(theta, xs){
  sum(xs)*log(theta) + (n - sum(xs))*log(1 - theta)
}

# Way 1: bootstrap
set.seed(712)
B = 1e4
theta_hat_bs = numeric(B)
for (b in 1:B) {
  xbs = sample(xs, size=n, replace=TRUE)
  theta_hat_bs[b] = optim(
    par = .25, lower = .00001, upper = .5,
    fn = ell, xs = xbs,
    method="L-BFGS-B",
    control=list(fnscale=-1))$par
}
var_1 = var(theta_hat_bs); var_1

# Way 2: CRLB
result = optim(
  par = .25, lower = .00001, upper = 1/2,
  fn = ell, xs = xs, hessian = T,
  method="L-BFGS-B",
  control=list(fnscale=-1))
f_info = -result$hessian
var_2 = 1/f_info; var_2
```

- Suppose $X_1, \dots, X_n$ is an iid sample following an exponential distribution, i.e., $f_X(x \mid \beta) = \beta^{-1}\exp(-x/\beta)\mathbf{1}_{(0,\infty)}(x)$, $\beta > 0$.
  - Observing the sample $1, \dots, 10$, give the approximate distribution of MLE $\hat{\beta}_{\text{ML}}$.

```r
xs = 1:10
n = length(xs)
ell = function(beta, xs){
  -n*log(beta) - sum(xs)/beta
}

# Way 1: bootstrap
set.seed(712)
B = 1e4
theta_hat_bs = numeric(B)
for (b in 1:B) {
  xbs = sample(xs, size=n, replace=TRUE)
  theta_hat_bs[b] = optim(
    par = 10, lower = .00001, upper = Inf,
    fn = ell, xs = xbs,
    method="L-BFGS-B",
    control=list(fnscale=-1))$par
}
var_1 = var(theta_hat_bs); var_1
```

```r
# Way 2: CRLB
result = optim(
  par = 10, lower = .00001, upper = Inf,
  fn = ell, xs = xs, hessian = T,
  method="L-BFGS-B",
  control=list(fnscale=-1))
f_info = -result$hessian
var_2 = 1/f_info; var_2
```

- $X_1, \ldots, X_n \overset{\text{iid}}{\sim} p(x \mid \lambda) = \lambda^x \exp(-\lambda)/x!$, $x \in \mathbb{Z}^+ \cup \{0\}$, $\lambda > 0$.
  - Observing the sample $1, \ldots, 10$, give the approximate distribution of MLE $\hat{\lambda}_{\text{ML}}$.

Delta method

- Used to approximate the distribution of $h(T_n)$ when
  - T_n is asymptotically normally distributed
  - $h(\cdot)$ doesn't rely on n
- (Delta method) Suppose T_n is an estimator of θ. If $\sqrt{n}(T_n - \theta) \approx \mathcal{N}(0, \sigma^2)$, h is NOT a function of n, AND $h'(\theta) \neq 0$, then as $n \to \infty$,
  - $\mathrm{E}\{h(T_n)\} \approx h(\theta)$
  - $\mathrm{var}\{h(T_n)\} \approx \{h'(\theta)\}^2\sigma^2/n$
  - $\sqrt{n}\{h(T_n) - h(\theta)\} \approx \mathcal{N}(0, \{h'(\theta)\}^2\sigma^2)$

Example Lec7.2

- $X_1, \ldots, X_n \overset{\text{iid}}{\sim} p(x \mid \lambda) = \lambda^x \exp(-\lambda)/x!$, $x \in \mathbb{Z}^+ \cup \{0\}$, $\lambda > 0$. Suppose the sample is $1, \ldots, 10$.
  1. What is the ML estimate of $\theta = \Pr(X_i = 0)$?
  2. Give the approximate distribution of MLE of θ.