

PH 712 Probability and Statistical Inference

Part V: Evaluating Estimators I

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

2025/Oct/29 13:49:46

Bias

- Bias of $\hat{\theta}$: $\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$
- Unbiasedness: $E(\hat{\theta}) = \theta \Leftrightarrow \hat{\theta}$ is an unbiased estimator of θ

Mean squared error (MSE)

- $\text{MSE}(\hat{\theta}) = E\{(\hat{\theta} - \theta)^2\} = \text{Bias}^2(\hat{\theta}) + \text{var}(\hat{\theta})$
 - The lower the better
 - $\text{MSE}(\hat{\theta}) = \text{var}(\hat{\theta})$ for unbiased $\hat{\theta}$

Numerically approximate MSE: using the (nonparametric) bootstrap

- Implementation
 1. Suppose you observe x_1, \dots, x_n for an iid sample of size n .
 2. Set a seed to make your result reproducible.
 3. For b in $1 : B$, do steps a–b.
 - a. Generate a bootstrap sample $x_1^{(b)}, \dots, x_n^{(b)}$ by drawing a sample of size n with replacement from $\{x_1, \dots, x_n\}$.
 - b. Generate a new estimate $\hat{\theta}^{(b)}$ from $x_1^{(b)}, \dots, x_n^{(b)}$.
 4. $\text{MSE}(\hat{\theta}) \approx B^{-1} \sum_{b=1}^B (\hat{\theta}^{(b)} - \theta)^2$.
- A similar question: how to numerically approximate $\text{var}(\hat{\theta})$?
 - $\text{var}(\hat{\theta}) \approx B^{-1} \sum_{b=1}^B (\hat{\theta}^{(b)} - B^{-1} \sum_{b=1}^B \hat{\theta}^{(b)})^2$

Example Lec5.1

- Suppose X_1, \dots, X_n is an iid sample following $\mathcal{N}(\mu, \sigma^2)$, i.e., $f_{X_i}(x | \theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$, $x \in \mathbb{R}$, with unknown μ and known $\sigma = 1$. The MLE of μ is $\hat{\mu}_{\text{ML}} = \bar{X} = n^{-1} \sum_{i=1}^n X_i$.
 - Observing the sample $1, \dots, 10$, numerically check the MSE of $\hat{\mu}_{\text{ML}}$ for $\mu = 5$.

```
set.seed(712)
xs = 1:10
n = length(xs)
ell = function(mu, xs){
  sigma = 1 # known
  -n/2*log(2*pi*sigma^2) - sum((xs - mu)^2)/(2*sigma^2)
}
```

```

B = 1e4
mu_hat_bs = numeric(B)
for (b in 1:B) {
  xbs = sample(xs, size=n, replace=TRUE)
  mu_hat_bs[b] = optim(
    par = 0, lower = -Inf, upper = Inf,
    fn = ell, xs = xbs,
    method="L-BFGS-B",
    control=list(fnscale=-1))$par
}
mse = mean((mu_hat_bs - 5)^2)

```

-
- Suppose X_1, \dots, X_n is an iid sample following $\mathcal{N}(\mu, \sigma^2)$, i.e., $f_{X_i}(x | \theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$, $x \in \mathbb{R}$, with known $\mu = 5$ and unknown $\sigma > 0$. The MLE of σ is $\hat{\sigma}_{\text{ML}} = \sqrt{n^{-1} \sum_{i=1}^n (X_i - \mu)^2}$.
 – Observing the sample 1, ..., 10, numerically check the MSE of $\hat{\sigma}_{\text{ML}}$ for $\sigma = 1$.

```

set.seed(712)
xs = 1:10
n = length(xs)
ell = function(sigma, xs){
  mu = 5 # known
  -n/2*log(2*pi*sigma^2) - sum((xs - mu)^2)/(2*sigma^2)
}

B = 1e4
sigma_hat_bs = numeric(B)
for (b in 1:B) {
  xbs = sample(xs, size=n, replace=TRUE)
  sigma_hat_bs[b] = optim(
    par = 10, lower = .00001, upper = Inf,
    fn = ell, xs = xbs,
    method="L-BFGS-B",
    control=list(fnscale=-1))$par
}
mse = mean((sigma_hat_bs - 1)^2)

```

-
- Suppose X_1, \dots, X_n is an iid sample following $p_{X_i}(x | \theta) = \theta^x (1 - \theta)^{1-x} \mathbf{1}_{\{0,1\}}(x)$, $\theta \in [0, 1/2]$. The MLE of θ is $\hat{\theta}_{\text{ML}} = \min\{\bar{X}, 1/2\}$.
 – Observing the sample 0, 1, 1, 1, 0, numerically check the MSE of $\hat{\theta}_{\text{ML}}$ for $\theta = .5$.

```

set.seed(712)
xs = c(0,1,1,1,0)
n = length(xs)
ell = function(theta, xs){
  sum(xs)*log(theta) + (n - sum(xs))*log(1 - theta)
}

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
}
mse = mean((theta_hat_bs - .5)^2)

```

-
- Suppose X_1, \dots, X_n is an iid sample following an exponential distribution, i.e., $f_X(x | \beta) = \beta^{-1} \exp(-x/\beta) \mathbf{1}_{(0, \infty)}(x)$, $\beta > 0$. The MLE of β is $\hat{\beta}_{\text{ML}} = \bar{X}$.
 - Observing the sample $1, \dots, 10$, numerically check the MSE of $\hat{\beta}_{\text{ML}}$ for $\beta = 5.5$.

-
- Suppose X_1, \dots, X_n is an iid sample following a beta distribution, i.e., $f_X(x | \theta) = \theta x^{\theta-1} \mathbf{1}_{[0,1]}(x)$, $\theta > 0$. The MLE of θ is $\hat{\theta}_{\text{ML}} = -n / \sum_{i=1}^n \ln X_i$.
 - Observing the sample $.1, \dots, .9$, numerically check the MSE of $\hat{\theta}_{\text{ML}}$ for $\theta = 1$.

```

set.seed(712)
xs = (1:9)/10
n = length(xs)
ell = function(theta, xs){
  n*log(theta) + (theta - 1)*log(prod(xs))
}

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 = 1, lower = .00001, upper = Inf,
    fn = ell, xs = xbs,
    method="L-BFGS-B",
    control=list(fnscale=-1))$par
}
mse = mean((theta_hat_bs - 1)^2)

```

Cramér-Rao lower bound (CRLB)

- Score/gradient: the derivative of the log-likelihood function (with respect to θ); denoted by $\ell'(\theta)$.
- Hessian: the second-order derivative of the log-likelihood function (with respect to θ); denoted by $\ell''(\theta)$.
- Fisher information $I_n(\theta) = \text{var}\{\ell'(\theta)\} = \text{E}\{[\ell'(\theta)]^2\} = -\text{E}\{\ell''(\theta)\}$
 - In practice, θ is unknown $\Rightarrow I_n(\theta)$ is unknown and can be approximated by $-\ell''(\hat{\theta}_{\text{ML}})$ (the observed Fisher information)
 - $-\ell''(\hat{\theta}_{\text{ML}})$ may be approximated by negative `optim()$hessian`
- Under certain conditions, for any unbiased estimator $\hat{\theta}$ (i.e., $\text{E}(\hat{\theta}) = \theta$), $\text{var}(\hat{\theta}) \geq I_n^{-1}(\theta)$ (CRLB for unbiased estimators of θ)
- Efficiency: For an UNBIASED estimator of θ , say $\hat{\theta}$, the *efficiency* of $\hat{\theta}$ is the ratio of the CRLB to $\text{var}(\hat{\theta})$, i.e., $I_n^{-1}(\theta)/\text{var}(\hat{\theta})$ (typically capped by 1).
 - The higher efficiency the better.

- $\hat{\theta}$ is an efficient estimator of $\theta \iff E(\hat{\theta}) = \theta$ and its efficiency = 1.

Example Lec5.2

- Suppose X_1, \dots, X_n is an iid sample following $\mathcal{N}(\mu, \sigma^2)$, i.e., $f_{X_i}(x | \theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$, $x \in \mathbb{R}$, with unknown μ and known $\sigma = 1$.
 - Observing the sample 1, ..., 10, numerically give the CRLB of unbiased estimator of μ .

- Suppose X_1, \dots, X_n is an iid sample following $\mathcal{N}(\mu, \sigma^2)$, i.e., $f_{X_i}(x | \theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$, $x \in \mathbb{R}$, with known $\mu = 5$ and unknown $\sigma > 0$.
 - Observing the sample 1, ..., 10, numerically give the CRLB of unbiased estimator of σ .

- Suppose X_1, \dots, X_n is an iid sample following $p_{X_i}(x | \theta) = \theta^x(1 - \theta)^{1-x}\mathbf{1}_{\{0,1\}}(x)$, $\theta \in [0, 1/2]$.
 - Observing the sample 0, 1, 1, 1, 0, numerically give the CRLB of unbiased estimator of θ .

- Suppose X_1, \dots, X_n is an iid sample following an exponential distribution, i.e., $f_X(x | \beta) = \beta^{-1} \exp(-x/\beta)\mathbf{1}_{(0,\infty)}(x)$, $\beta > 0$.
 - Observing the sample 1, ..., 10, numerically give the CRLB of unbiased estimator of β .

- Suppose X_1, \dots, X_n is an iid sample following a beta distribution, i.e., $f_X(x | \theta) = \theta x^{\theta-1}\mathbf{1}_{[0,1]}(x)$, $\theta > 0$.
 - Observing the sample 1, ..., 10, numerically give the CRLB of unbiased estimator of θ .