PH 712 Probability and Statistical Inference

Part V: Properties of Estimators I: Finite-sample Propoties

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

2025/Nov/03 19:28:33

Bias

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

Mean squared error (MSE)

- $\text{MSE}(\hat{\theta}) = \text{E}\{(\hat{\theta} - \theta)^2\} = \text{Bias}^2(\hat{\theta}) + \text{var}(\hat{\theta})$
  - 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 $x_1, \ldots, x_n$ 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 $x_1^{(b)}, \ldots, x_n^{(b)}$ by drawing a sample of size n with replacement from $\{x_1, \ldots, x_n\}$.
     b. Generate a new estimate $\hat{\theta}^{(b)}$ from $x_1^{(b)}, \ldots, x_n^{(b)}$.
  4. $\text{MSE}(\hat{\theta}) \approx B^{-1} \sum_{b=1}^{B} (\hat{\theta}^{(b)} - \theta)^2$.
- Similar questions:
  - How to numerically approximate $\text{E}(\hat{\theta})$?
    * $\text{E}(\hat{\theta}) \approx B^{-1} \sum_{b=1} \hat{\theta}^{(b)}$
  - How to numerically approximate $\text{var}(\hat{\theta})$?
    * $\text{var}(\hat{\theta}) \approx B^{-1} \sum_{b=1}^{B} (\hat{\theta}^{(b)} - B^{-1} \sum_{b=1} \hat{\theta}^{(b)})^2 \approx (B-1)^{-1} \sum_{b=1}^{B} (\hat{\theta}^{(b)} - B^{-1} \sum_{b=1} \hat{\theta}^{(b)})^2$

Example Lec5.1

- 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 $\hat{\mu}_{\text{ML}} = \bar{X} = n^{-1} \sum_{i=1}^{n} X_i$.
  - Observing the sample $1, \ldots, 10$, 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
```

```r
  -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 $X_1, \ldots, X_n$ 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 $\hat{\sigma}_{\mathrm{ML}} = \sqrt{n^{-1}\sum_{i=1}^n (X_i - \mu)^2}$.
  - Observing the sample $1, \ldots, 10$, numerically check the MSE of $\hat{\sigma}_{\mathrm{ML}}$ for $\sigma = 1$.

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

```r
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 $X_1, \ldots, X_n$ 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 $\hat{\beta}_{\mathrm{ML}} = \bar{X}$.
  - Observing the sample $1, \ldots, 10$, numerically check the MSE of $\hat{\beta}_{\mathrm{ML}}$ for $\beta = 5.5$.

- Suppose $X_1, \ldots, X_n$ 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 $\hat{\theta}_{\mathrm{ML}} = -n/\sum_{i=1}^{n} \ln X_i$.
  - Observing the sample $.1, \ldots, .9$, numerically check the MSE of $\hat{\theta}_{\mathrm{ML}}$ 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 $I_n(\theta) = \mathrm{var}\{\ell'(\theta)\} = \mathrm{E}[\{\ell'(\theta)\}^2] = -\mathrm{E}\{\ell''(\theta)\}$
  - In practice, θ is unknown $\Rightarrow I_n(\theta)$ is unknown and can be approximated by $-\ell''(\hat{\theta}_{\mathrm{ML}})$ ($-\ell''(\hat{\theta}_{\mathrm{ML}})$ is the observed Fisher information)
  - $-\ell''(\hat{\theta}_{\mathrm{ML}})$ may be approximated by negative `optim()$hessian`
- Under certain conditions, for any unbiased estimator $\hat{\theta}$ (i.e., $\mathrm{E}(\hat{\theta}) = \theta$),

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{var}(\hat{\theta}) \geq I_n^{-1}(\theta)$$

  - $I_n^{-1}(\theta)$ serving as a lower bound on the MSE/variance of any unbiased estimator 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 $\theta \iff \text{E}(\hat{\theta}) = \theta$ and its efficiency $= 1 \iff \text{E}(\hat{\theta}) = \theta$ and $\text{var}(\hat{\theta}) =$CRLB.

Example Lec5.2

- 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 $\sigma = 1$.
  - Observing the sample $1, \ldots, 10$, numerically give the CRLB of unbiased estimator of μ.

```
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)
}
result = optim(
  par = 0, lower = -Inf, upper = Inf,
  fn = ell, xs = xs,
  method="L-BFGS-B", hessian = T,
  control=list(fnscale=-1))
f_info = -result$hessian
crlb = 1/f_info
```

- 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 $\sigma > 0$.
  - Observing the sample $1, \ldots, 10$, numerically give the CRLB of unbiased estimator of σ.

```
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)
}
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
crlb = 1/f_info
```

- 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 $0, 1, 1, 1, 0$, numerically give the CRLB of unbiased estimator of θ.

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

```
}
result = optim(
  par = .25, lower = 0, upper = 1/2,
  fn = ell, xs = xs, hessian = T,
  method="L-BFGS-B",
  control=list(fnscale=-1))
f_info = -result$hessian
crlb = 1/f_info
```

- Suppose $X_1, \ldots, X_n$ 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 $1, \ldots, 10$, numerically give the CRLB of unbiased estimator of β.

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

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

```
xs = (1:9)/10
n = length(xs)
ell = function(theta, xs){
  n*log(theta) + (theta - 1)*log(prod(xs))
}
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
crlb = 1/f_info
```