

All of Statistics - Chapter 9 Solutions

Jul 25, 2020

1.

Let \hat{F} be the empirical distribution. The method of moment estimators (MMEs) satisfy

$$\begin{aligned}E_{\hat{F}}X &= \frac{\tilde{\alpha}}{\tilde{\beta}} \\E_{\hat{F}}[X^2] &= \frac{\tilde{\alpha}(\tilde{\alpha} + 1)}{\tilde{\beta}^2}.\end{aligned}$$

Solving for $\tilde{\alpha}$ and $\tilde{\beta}$,

$$\begin{aligned}\tilde{\alpha} &= \frac{(E_{\hat{F}}X)^2}{\text{Var}_{\hat{F}}(X)} \\ \tilde{\beta} &= \frac{E_{\hat{F}}X}{\text{Var}_{\hat{F}}(X)}.\end{aligned}$$

2.

a)

The MMEs satisfy

$$\begin{aligned}E_{\hat{F}}X &= \frac{\tilde{a} + \tilde{b}}{2} \\ \text{Var}_{\hat{F}}(X) &= \frac{1}{12}(\tilde{b} - \tilde{a})^2.\end{aligned}$$

Define $\tilde{c} = \tilde{b} - \tilde{a}$. It follows that

$$\tilde{c} = 2\sqrt{3}\text{Std}_{\hat{F}}(X).$$

Since $\tilde{a} + \tilde{b} = 2\tilde{a} + \tilde{c}$ and $\tilde{a} + \tilde{b} = 2\tilde{b} - \tilde{c}$,

$$\tilde{a} = E_{\hat{F}}X - \sqrt{3}\text{Std}_{\hat{F}}(X)$$

$$\tilde{b} = E_{\hat{F}}X + \sqrt{3}\text{Std}_{\hat{F}}(X).$$

b)

The maximum likelihood estimators (MLEs) maximize

$$\mathcal{L}(a, b) = \prod_i f(X_i; a, b) = \frac{1}{(b-a)^n} \prod_i I_{(a,b)}(X_i).$$

The maximum occurs at $\hat{a} = \min_i X_i$ and $\hat{b} = \max_i X_i$.

c)

By equivariance, the MLE of $\tau = (a + b)/2$ is $\hat{\tau} = (\hat{a} + \hat{b})/2$.

d)

Estimator	Mean squared error (MSE)
MLE	0.015
Non-parametric plug-in estimate	0.033

The MSE of $\hat{\tau}$, the MLE of the mean, is computed using the code below.

```
import numpy as np

a = 1.
b = 3.
n = 10
n_sims = 10**6

samples = np.random.uniform(low=a, high=b, size=[n_sims, n])

a_mle = np.min(samples, axis=1) # Maximum likelihood estimator of a.
b_mle = np.max(samples, axis=1) # Maximum likelihood estimator of b.

tau = (a + b) / 2. # Mean of Uniform(a, b).
tau_mle = (a_mle + b_mle) / 2. # Maximum likelihood estimator of the mean.

mse = np.mean((tau_mle - tau)**2)
```

The non-parameteric plug-in estimate of τ is $\tilde{\tau} = E_{\hat{F}}X$. Since this estimator is unbiased, its MSE is

$$\text{Var}(\tilde{\tau}) = \frac{1}{n} \text{Var}(X_1) = \frac{1}{12n} (b-a)^2$$

3.

a)

Let Z be a standard normal random variable so that

$$0.95 = P(X < \tau) = P(Z < (\tau - \mu)/\sigma) = F_Z((\tau - \mu)/\sigma)$$

and hence

$$\tau = g(\mu, \sigma) \equiv F_Z^{-1}(0.95)\sigma + \mu.$$

The MLEs of the mean and standard deviation of the original distribution are $\hat{\mu} = E_{\hat{F}}X$ and $\hat{\sigma} = \text{Std}_{\hat{F}}(X)$. By equivariance, the MLE of τ is $\hat{\tau} = g(\hat{\mu}, \hat{\sigma})$.

b)

By the delta method,

$$\begin{aligned} \widehat{\text{se}}(\hat{\tau}) &= \sqrt{\nabla g^\top \hat{J}_n \nabla g} = \sqrt{\frac{1}{n} \begin{pmatrix} 1 & \\ & F_Z^{-1}(0.95) \end{pmatrix}^\top \begin{pmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{2} \end{pmatrix} \begin{pmatrix} 1 \\ F_Z^{-1}(0.95) \end{pmatrix}} \\ &= \sigma \sqrt{\frac{1}{n} + \frac{1}{2n} \left(F_Z^{-1}(0.95)\right)^2} \end{aligned}$$

A $1 - \alpha$ confidence interval is $\hat{\tau} \pm z_{\alpha/2} \cdot \widehat{\text{se}}(\hat{\tau})$.

c)

Estimator	SE
Delta Method	0.558
Parametric Bootstrap (100,000 samples)	0.557

Code for the delta method and parametric bootstrap are given below.

```
import numpy as np
from scipy.stats import norm

data = np.array([ 3.23, -2.50,  1.88, -0.68,  4.43,  0.17,
                 1.03, -0.07, -0.01,  0.76,  1.76,  3.18,
                 0.33, -0.31,  0.30, -0.61,  1.52,  5.43,
                 1.54,  2.28,  0.42,  2.33, -1.03,  4.00,
                 0.39])

se_delta_method = np.std(data) \
    * np.sqrt(1. / n_samples * (1. + 0.5 * norm.ppf(0.95)**2))

n_samples = data.size
n_sims = 10**5
samples = np.std(data) * np.random.randn(n_sims, n_samples) + np.mean(data)
tau_mles = np.std(samples, axis=1) * norm.ppf(0.95) + np.mean(samples, axis=1)
se_parametric_bootstrap = np.std(tau_mles)
```

4.

By Question 2 (b), the MLE of θ is $Y = \max_i X_i$. Its CDF is $F_Y = F_{X_1}^n$. Therefore, for any $\epsilon > 0$, $F_Y(\theta - \epsilon) < 1$ and $F_Y(\theta + \epsilon) = 1$. It follows that

$$\begin{aligned} P(|Y - \theta| > \epsilon) &= 1 - P(\theta - \epsilon \leq Y \leq \theta + \epsilon) = 1 - F_Y(\theta + \epsilon) + F_Y(\theta - \epsilon) \\ &= F_Y(\theta - \epsilon) = F_{X_1}(\theta - \epsilon)^n \end{aligned}$$

Taking a limit in n yields the desired result.

5.

Since $\lambda = EX_1$, the MME is the sample mean.

The MLE is also the sample mean. To see this, note that

$$\log f(X_i; \lambda) = \log \left(\frac{\lambda^{X_i} e^{-\lambda}}{X_i!} \right) = X_i \log \lambda - \lambda - \log(X_i!)$$

and hence

$$s(X_i; \lambda) \equiv \frac{d}{d\lambda} [\log f(X_i; \lambda)] = \frac{X_i}{\lambda} - 1.$$

Therefore, the derivative of the log likelihood is

$$\frac{d}{d\lambda} [\log \mathcal{L}(\lambda)] = \sum_i \frac{d}{d\lambda} [\log f(X_i; \lambda)] = \frac{1}{\lambda} E_{\hat{F}} X - n.$$

Setting this to zero and solving for the parameter yields the desired result.

The Fisher information is

$$I(\lambda) = \text{Var}(s(X_i; \lambda)) = \frac{1}{\lambda}.$$

6.

a)

$\hat{\theta} = E_{\hat{F}} X$ is the MLE of the mean. Let Z be a standard normal random variable. Since

$$\psi = P(Y_1 = 1) = P(X_1 > 0) = P(Z > -\theta) = F_Z(\theta),$$

the MLE of ψ is $\hat{\psi} = F_Z(\hat{\theta})$ by equivariance.

b)

An approximate 95% confidence interval (CI) for ψ is $\hat{\psi} \pm 2\text{se}(\hat{\psi})$ where, by the delta method,

$$\text{se}(\hat{\psi}) = \left| f_Z(\hat{\theta}) \right| \text{se}(\hat{\theta}) = f_Z(\hat{\theta}) \frac{1}{\sqrt{n}}$$

c)

By the law of large numbers (LLN), $\tilde{\psi}$ converges in probability to $EY_1 = P(Y_1 = 1) = \psi$.

d)

Similarly to Part (b),

$$\text{se}(\hat{\psi}) = f_Z(\theta) \frac{1}{\sqrt{n}}.$$

Moreover,

$$\text{se}(\tilde{\psi}) = \frac{1}{\sqrt{n}} \text{se}(Y_1) = \sqrt{\frac{\psi(1-\psi)}{n}}.$$

Therefore,

$$\text{ARE}(\tilde{\psi}, \hat{\psi}) = \frac{(f_Z(\theta))^2}{\psi(1-\psi)}$$

It's possible to show that the ARE achieves its maximum value of $2/\pi \approx 0.637$ at $\theta = 0$. Note that this quantity is necessarily less than one due to the asymptotic optimality of the MLE (Theorem 9.23).

e)

Suppose the distribution is not normal. Under sufficient regularity, the LLN guarantees the sample mean $\hat{\theta}$ to converge, in probability, to the true mean μ of the distribution. As such, $\hat{\psi} = F_Z(\hat{\theta})$ converges, in probability, to $F_Z(\mu)$.

7.

a)

The log-likelihood for drug $i = 1, 2$ is

$$s(X_i; p_i) = \log f(X_i; p_i) = \log \binom{n_i}{X_i} + X_i \log p_i + (n_i - X_i) \log (1 - p_i).$$

Taking derivatives,

$$\frac{d}{dp_i} [s(X_i; p_i)] = \frac{X_i}{p_i} + \frac{X_i - n}{1 - p_i} = \frac{X_i - np_i}{p_i(1 - p_i)}.$$

It follows that the MLE is $\hat{p}_i = X_i/n$. By equivariance the MLE of $\psi = p_1 - p_2$ is $\hat{\psi} = \hat{p}_1 - \hat{p}_2$.

b)

The Fisher information of drug $i = 1, 2$ is

$$\text{Var}(s(X_i; p_i)) = \text{Var}\left(\frac{X_i}{p_i} + \frac{X_i - n}{1 - p_i}\right) = \frac{n_i}{p_i(1 - p_i)}.$$

Since the two trials are independent, the complete Fisher information is

$$I(p_1, p_2) = \text{diag}\left(\frac{n_1}{p_1(1 - p_1)}, \frac{n_2}{p_2(1 - p_2)}\right).$$

c)

By the delta method,

$$\text{se}(\hat{\psi}) = \sqrt{\begin{pmatrix} +1 \\ -1 \end{pmatrix}^\top I(\hat{p}_1, \hat{p}_2)^{-1} \begin{pmatrix} +1 \\ -1 \end{pmatrix}} = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}.$$

d)

Method	90% CI Lower bound	90% CI Upper bound
Delta Method	-0.009	0.129
Parametric Bootstrap (100,000 samples)	-0.009	0.129

```
import numpy as np
from scipy.stats import norm
```

```
n = 200
x1 = 160
x2 = 148
n_sims = 10**5
```

```
p1_mle = x1/n
p2_mle = x2/n
psi_mle = p1_mle - p2_mle
ppf_0p95 = norm.ppf(0.95)
```

```
se_delta = np.sqrt(p1_mle * (1. - p1_mle) / n + p2_mle * (1. - p2_mle) / n)
```

```

print('90% CI delta method: [{:.3f}, {:.3f}]'.format(
    psi_mle - 1.645 * se_delta_method, psi_mle + 1.645 * se_delta_method))

samples1 = np.random.binomial(n, p1_mle, size=[n_sims])
samples2 = np.random.binomial(n, p2_mle, size=[n_sims])
psi_mles = samples1/n - samples2/n
se_parametric_bootstrap = np.std(psi_mles)
print('90% CI parametric bootstrap: [{:.3f}, {:.3f}]'.format(
    psi_mle - 1.645 * se_parametric_bootstrap,
    psi_mle + 1.645 * se_parametric_bootstrap))

```

8.

The log likelihood is

$$\mathcal{L}(\mu, \sigma) = \sum_i \log f(X_i; \mu, \sigma) = -\frac{n}{2} \log(2\pi) - n \log(\sigma) - \frac{1}{2\sigma^2} \sum_i (X_i - \mu)^2.$$

Taking derivatives,

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mu} &= \frac{\sum_i (X_i - \mu)}{\sigma^2} \\ \frac{\partial^2 \mathcal{L}}{\partial \mu^2} &= -\frac{n}{\sigma^2} \\ \frac{\partial \mathcal{L}}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_i (X_i - \mu)^2 \\ \frac{\partial \mathcal{L}}{\partial \mu \partial \sigma} &= -\frac{2 \sum_i (X_i - \mu)}{\sigma^2} \\ \frac{\partial^2 \mathcal{L}}{\partial \sigma^2} &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_i (X_i - \mu)^2. \end{aligned}$$

Taking expectations,

$$\begin{aligned} \mathbb{E} \left[\frac{\partial \mathcal{L}}{\partial \mu \partial \sigma} \right] &= 0 \\ \mathbb{E} \left[\frac{\partial^2 \mathcal{L}}{\partial \sigma^2} \right] &= -\frac{2n}{\sigma^2}. \end{aligned}$$

Therefore, the Fisher information is

$$I(\mu, \sigma) = -\frac{n}{\sigma^2} \begin{pmatrix} 1 & \\ & 2 \end{pmatrix}.$$

9.

a)

Results are given below.

Method	95% CI Lower bound	95% CI Upper bound
Delta Method	126.146	189.219
Parametric Bootstrap (100,000 samples)	126.076	189.288
Non-parametric Bootstrap (100,000 samples)	129.553	185.812

The MLE of θ is $\hat{\theta} = g(\hat{\mu}) \equiv \exp \hat{\mu}$ where $\hat{\mu} = E_{\hat{F}} X$ is the MLE of the mean. By the delta method,

$$\widehat{se}(\hat{\theta}) = \left| g'(\hat{\mu}) \right| se(\hat{\mu}) = \frac{\exp(\hat{\mu})}{\sqrt{n}}.$$

Therefore, a 95% CI for θ is $(1 \pm 2n^{-1/2})\exp(\hat{\mu})$.

Code for all three methods is given below.

```
import numpy as np

np.random.seed(1)
data = np.random.randn(100) + 5.
mu_mle = np.mean(data)

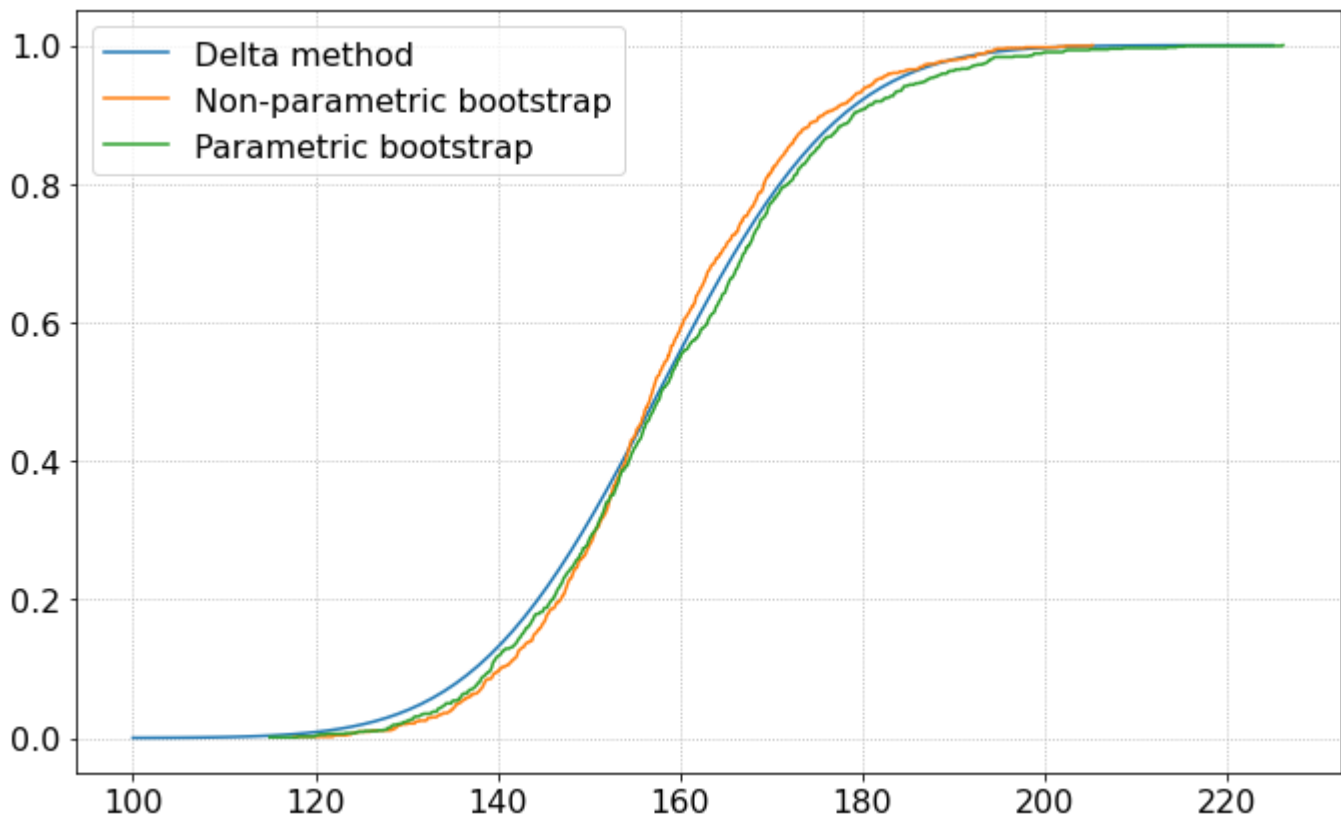
n_sims = 10**5
n_samples = data.size

print('95% CI delta method: [{:.3f}, {:.3f}]'.format(
    (1. - 2. / np.sqrt(n_samples)) * np.exp(mu_mle),
    (1. + 2. / np.sqrt(n_samples)) * np.exp(mu_mle)))

samples = np.random.randn(n_sims, n_samples) + mu_mle
theta_mles = np.exp(np.mean(samples, axis=1))
se_parametric_bootstrap = np.std(theta_mles)
print('95% CI parametric bootstrap: [{:.3f}, {:.3f}]'.format(
    np.exp(mu_mle) - 2. * se_parametric_bootstrap,
    np.exp(mu_mle) + 2. * se_parametric_bootstrap))

indices = np.random.randint(n_samples, size=[n_sims * n_samples])
samples = data[indices]
splits = np.split(samples, n_sims)
theta_mles = np.empty([n_sims])
for i, split in enumerate(splits):
    theta_mles[i] = np.exp(np.mean(split))
se_bootstrap = np.std(theta_mles)
print('95% CI non-parametric bootstrap: [{:.3f}, {:.3f}]'.format(
    np.exp(mu_mle) - 2. * se_bootstrap,
    np.exp(mu_mle) + 2. * se_bootstrap))
```

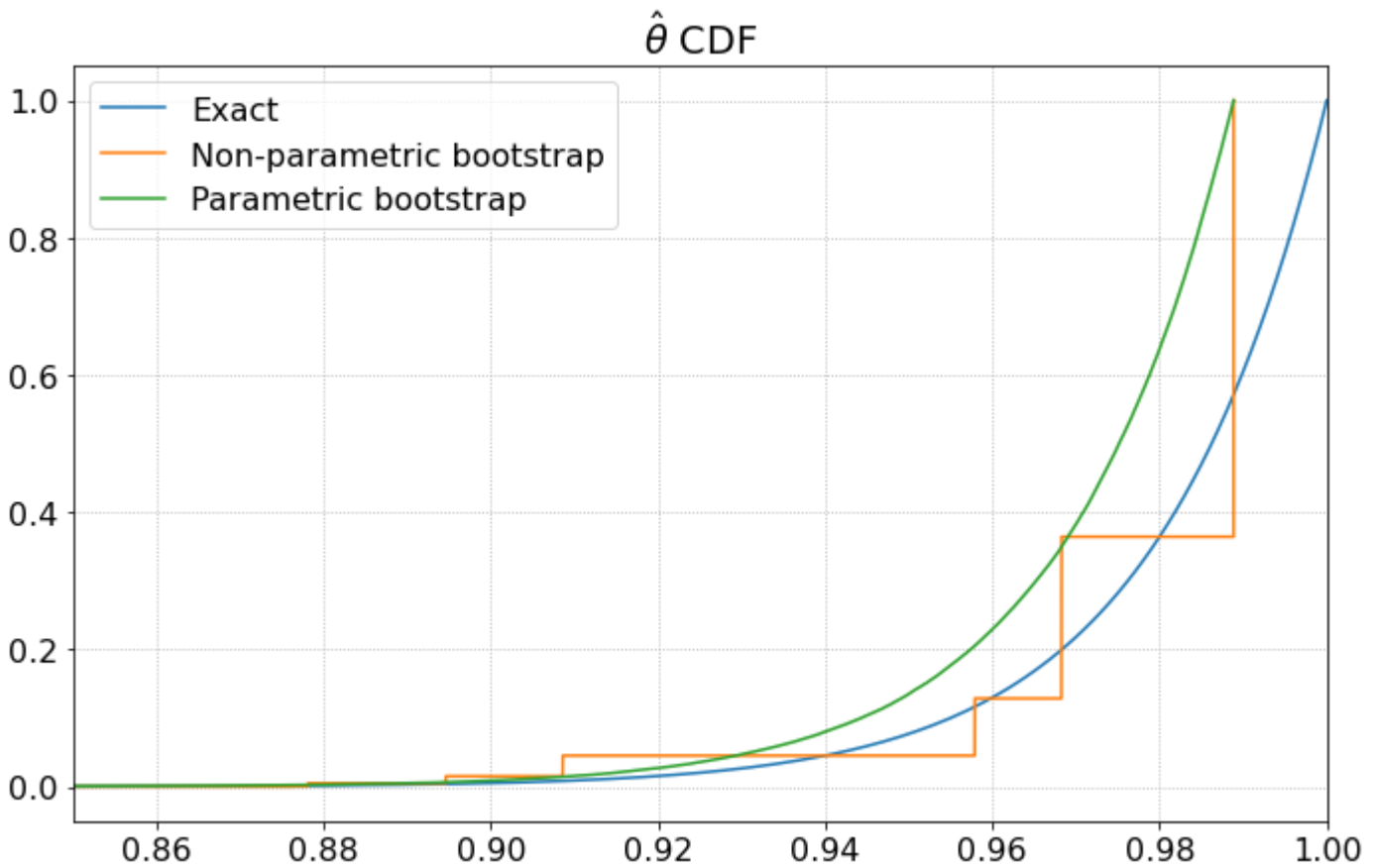
b)



10.

a)

The CDF of $\hat{\theta}$ is $F(y) = (y/\theta)^n$. This, along with CDFs of bootstrap estimators, are plotted below.



b)

The parametric bootstrap estimator of this parameter has a continuous distribution and hence zero probability of being equal to $\hat{\theta}$ exactly. Let $\hat{\theta}^* = \max_{1 \leq i \leq n} X_i^*$ be the non-parametric bootstrap estimator. Note that

$$P(\hat{\theta}^* = \hat{\theta}) = 1 - P(\hat{\theta}^* \neq X_{(n)}) = 1 - (1 - 1/n)^n \rightarrow 1 - 1/e \approx 0.632.$$

That is, the non-parametric bootstrap estimator has a good chance of being equal to the MLE. These phenomena are visible in the plot above.