Chapter 8

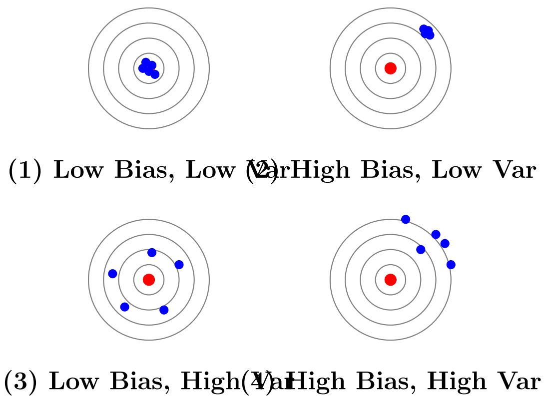
Point Estimation

Detailed Solutions

8.1 Basic Concept

8.1 The Goal of Estimation

Problem: Target Board Analogy.



(1) Low Bias, Low Var **(2) High Bias, Low Var**

(3) Low Bias, High Var **(4) High Bias, High Var**

Solution:

(a) **Estimator vs. Estimate:**

- **Point Estimator ($\hat{\Theta}$):** A formula or function of the random sample (e.g., $\bar{X} = \frac{1}{n}\sum X_i$). It is a **Random Variable**.
- **Estimate ($\hat{\theta}$):** A specific numerical value calculated from realized data (e.g., $\bar{x} = 5.2$). It is a **Constant**.

(b) **Consistency:** An estimator is consistent if, as sample size $n \to \infty$, the estimator converges in probability to the true parameter θ. $\bar{X}$ is consistent for μ because of the Law of Large Numbers (Variance $\sigma^2/n \to 0$).

(c) **Target Analogy:**

- **Bias (Accuracy):** How close the "center" of the shots is to the bullseye.

- **Variance (Precision):** How tightly clustered the shots are.
- **"Unbiased but Inefficient":** Corresponds to **Scenario (3)**. The average position is correct (on the bullseye), but the spread is wide (High Variance).

(a) Estimator is the RV. (c) Scenario (3) is Unbiased but Inefficient.

8.2 Properties of Estimators

Problem: Unbiasedness, Efficiency, MSE.
Solution:

(a) **Unbiasedness of $\bar{X}$:**

$$E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\cdot(n\mu) = \mu$$

Since $E[\bar{X}] = \mu$, it is unbiased.

(b) **Efficiency:** If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased, we choose the one with the **smaller variance**. The one with minimum variance is called the Minimum Variance Unbiased Estimator (MVUE).

(c) **MSE Proof:** Let $b = E[\hat{\theta}] - \theta$ be the bias.

$$\begin{aligned} MSE(\hat{\theta}) &= E[(\hat{\theta} - \theta)^2] \\ &= E[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)^2] \\ &= E[(\underbrace{\hat{\theta} - E[\hat{\theta}]}_{A} + \underbrace{b}_{B})^2] \\ &= E[A^2 + 2AB + B^2] = E[A^2] + 2bE[A] + b^2 \end{aligned}$$

$E[A] = E[\hat{\theta} - E[\hat{\theta}]] = 0$. $E[A^2] = E[(\hat{\theta} - E[\hat{\theta}])^2] = Var(\hat{\theta})$.

$$\therefore MSE = Var(\hat{\theta}) + b^2 = Var(\hat{\theta}) + [Bias(\hat{\theta})]^2$$

(c) $MSE = Variance + Bias^2$

8.2 Intermediate

8.3 MOM vs. MLE: Poisson Distribution

Problem: $X \sim Poisson(\lambda)$.
Solution:

(a) **MOM:** Population Mean $E[X] = \lambda$. Sample Mean $M_1 = \bar{X}$. Equating them:

$$\hat{\lambda}_{MOM} = \bar{X}$$

(b) **MLE:** Likelihood $L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod x_i!}$. Log-Likelihood $\ell(\lambda) = -n\lambda + (\sum x_i)\ln\lambda - \ln(\prod x_i!)$. Differentiate w.r.t λ:

$$\frac{d\ell}{d\lambda} = -n + \frac{\sum x_i}{\lambda} = 0$$

$$n = \frac{\sum x_i}{\lambda} \implies \hat{\lambda}_{MLE} = \frac{\sum x_i}{n} = \bar{X}$$

(c) **Invariance Property:** We want to estimate $g(\lambda) = e^{-\lambda}$. The MLE of a function is the function of the MLE.

$$\widehat{P(X=0)}_{MLE} = e^{-\hat{\lambda}_{MLE}} = e^{-\bar{X}}$$

(a) $\bar{X}$
(b) $\bar{X}$
(c) $e^{-\bar{X}}$

8.4 MLE for Pareto Distribution

Problem: $f(x) = \alpha x_m^{\alpha} x^{-(\alpha+1)}$. Known x_m.
Solution:

(a) **Likelihood:**

$$L(\alpha) = \prod_{i=1}^{n} \alpha x_m^{\alpha} x_i^{-(\alpha+1)} = \alpha^n x_m^{n\alpha} \left(\prod x_i\right)^{-(\alpha+1)}$$

(b) **MLE Derivation:**

$$\ell(\alpha) = n\ln\alpha + n\alpha\ln x_m - (\alpha+1)\sum \ln x_i$$

$$\frac{d\ell}{d\alpha} = \frac{n}{\alpha} + n\ln x_m - \sum \ln x_i = 0$$

$$\frac{n}{\alpha} = \sum \ln x_i - n\ln x_m = \sum(\ln x_i - \ln x_m) = \sum \ln(x_i/x_m)$$

$$\hat{\alpha}_{MLE} = \frac{n}{\sum_{i=1}^{n} \ln(x_i/x_m)}$$

$\hat{\alpha} = n/\sum \ln(x_i/x_m)$

8.5 Parameter Estimation for Power Function

Problem: $f(x) = \theta x^{\theta-1}, 0 < x < 1.$
Solution:

(a) **MOM:** $E[X] = \int_0^1 x(\theta x^{\theta-1})dx = \theta \int_0^1 x^\theta dx = \theta \left[\frac{x^{\theta+1}}{\theta+1}\right]_0^1 = \frac{\theta}{\theta+1}$.

Set $\bar{X} = \frac{\theta}{\theta+1} \implies \bar{X}(\theta+1) = \theta \implies \theta(1-\bar{X}) = \bar{X}$.

$$\hat{\theta}_{MOM} = \frac{\bar{X}}{1-\bar{X}}$$

(b) **MLE:** $L(\theta) = \prod \theta x_i^{\theta-1} = \theta^n (\prod x_i)^{\theta-1}$.

$\ell(\theta) = n \ln \theta + (\theta - 1) \sum \ln x_i$. $\frac{d\ell}{d\theta} = \frac{n}{\theta} + \sum \ln x_i = 0$.

$$\hat{\theta}_{MLE} = -\frac{n}{\sum \ln x_i}$$

(c) **Calculation:** Data: $0.5, 0.8, 0.9$. $\bar{X} = 2.2/3 \approx 0.733$. $\hat{\theta}_{MOM} = 0.733/(1-0.733) = 0.733/0.267 \approx 2.75$.

$\sum \ln x_i = \ln(0.5) + \ln(0.8) + \ln(0.9) = -0.693 - 0.223 - 0.105 = -1.021$. $\hat{\theta}_{MLE} = -3/(-1.021) \approx 2.94$.

(a) $\bar{X}/(1-\bar{X})$ (b) $-n/\sum \ln x_i$
(c) MOM ≈ 2.75, MLE ≈ 2.94

8.6 Rayleigh Distribution

Problem: $f(x) = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}$. Let $\theta = \sigma^2$.
Solution:

(a) **MLE:** $L(\theta) = \prod \frac{x_i}{\theta} e^{-x_i^2/2\theta} = \frac{\prod x_i}{\theta^n} e^{-\sum x_i^2/2\theta}$. $\ell(\theta) = \sum \ln x_i - n \ln \theta - \frac{1}{2\theta} \sum x_i^2$.

$$\frac{d\ell}{d\theta} = -\frac{n}{\theta} + \frac{1}{2\theta^2} \sum x_i^2 = 0$$

$$\frac{n}{\theta} = \frac{\sum x_i^2}{2\theta^2} \implies n = \frac{\sum x_i^2}{2\theta} \implies \hat{\theta}_{MLE} = \frac{1}{2n} \sum_{i=1}^{n} x_i^2$$

(b) **Unbiasedness:** $E[\hat{\theta}] = E[\frac{1}{2n} \sum X_i^2] = \frac{1}{2n} \sum E[X_i^2]$.

Given $E[X^2] = 2\sigma^2 = 2\theta$. $E[\hat{\theta}] = \frac{1}{2n} \cdot n(2\theta) = \theta$. Yes, it is Unbiased.

(a) $\hat{\sigma}^2 = \frac{1}{2n} \sum x_i^2$
(b) Yes, Unbiased.

8.7 Comparing Two Estimators

Problem: $\hat{\mu}_1 = \bar{X}$, $\hat{\mu}_2 = \frac{X_1 + X_n}{2}$.
Solution:

(a) **Unbiased:** $E[\hat{\mu}_1] = \mu$ (Proven before).

$E[\hat{\mu}_2] = \frac{1}{2}(E[X_1] + E[X_n]) = \frac{1}{2}(\mu + \mu) = \mu$. Both are unbiased.

(b) **Variance:**

$Var(\hat{\mu}_1) = \frac{\sigma^2}{n}$.

$Var(\hat{\mu}_2) = \frac{1}{4}(Var(X_1) + Var(X_n)) = \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{\sigma^2}{2}$.

(c) **Efficiency:** For $n > 2$, $\frac{\sigma^2}{n} < \frac{\sigma^2}{2}$. $\bar{X}$ has lower variance (MSE). Using all data provides more information and averages out noise better than using just two points.

(c) $\bar{X}$ is more efficient.

8.8 The Bias-Variance Trade-off

Problem: Estimator $\hat{\sigma}_c^2 = c\sum(X_i - \bar{X})^2$. Let $S_{xx} = \sum(X_i - \bar{X})^2$. Given $E[S_{xx}] = (n-1)\sigma^2$, $Var(S_{xx}) = 2(n-1)\sigma^4$.
Solution:

(a) **Derive MSE:**

$$\begin{aligned}\text{Bias} &= E[cS_{xx}] - \sigma^2\\ &= c(n-1)\sigma^2 - \sigma^2\\ &= \sigma^2\big[c(n-1) - 1\big],\\ \text{Var} &= c^2\,\text{Var}(S_{xx})\\ &= 2(n-1)c^2\sigma^4,\\ \text{MSE} &= \text{Var} + \text{Bias}^2\\ &= 2(n-1)c^2\sigma^4 + \sigma^4\big[c(n-1)-1\big]^2.\end{aligned}$$

$$\text{Factor out } \sigma^4: \qquad \text{MSE} = \sigma^4\left[2(n-1)c^2 + \big(c(n-1)-1\big)^2\right].$$

(b) **Minimize MSE:**

$$\text{Let } k = n-1.$$

$$\text{Minimize } \; 2kc^2 + (ck-1)^2.$$

$$\begin{aligned}f(c) &= 2kc^2 + (ck-1)^2\\ &= 2kc^2 + c^2k^2 - 2ck + 1\\ &= c^2(2k+k^2) - 2ck + 1.\end{aligned}$$

Differentiate with respect to c:

$$f'(c) = 2c(2k+k^2) - 2k.$$

Set $f'(c) = 0$:

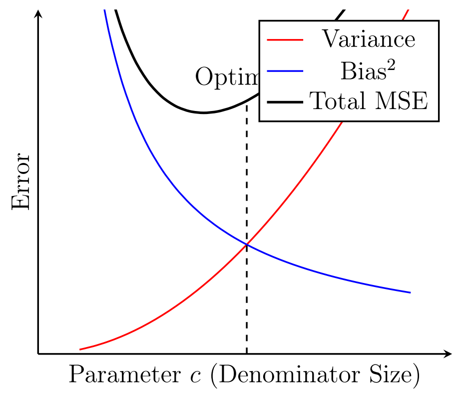
$$\begin{aligned}2c(2k+k^2) - 2k &= 0\\ c(2k+k^2) &= k\\ c = \frac{k}{k(2+k)} &= \frac{1}{k+2}.\end{aligned}$$

Substitute $k = n-1$:

$$\boxed{c = \frac{1}{(n-1)+2} = \frac{1}{n+1}}$$

(c) **Interpretation:** The standard unbiased estimator S^2 uses $c = \frac{1}{n-1}$. The MLE uses $c = \frac{1}{n}$. The minimum MSE estimator uses $c = \frac{1}{n+1}$. This shows that accepting some bias (by dividing by $n+1$) leads to a lower total error than the unbiased estimator.

(d) **Trade-off Graph:**



Minimizing Bias (Blue line $\rightarrow 0$) causes Variance (Red line) to explode, making Total Error high. The optimum balances both.

8.3 Challenge

8.9 MLE with Boundary Condition (Uniform)

Problem: $U(0, \theta)$.
Solution:

(a) **Likelihood:** $L(\theta) = \prod \frac{1}{\theta}\mathbb{I}(0 \le x_i \le \theta) = \frac{1}{\theta^n}\mathbb{I}(\max(x_i) \le \theta)$. It is non-zero only if $\theta \ge x_{(n)}$.

(b) **Differentiation Fails:** $L(\theta) = \theta^{-n}$ is a decreasing function. The derivative $-n\theta^{-n-1}$ is never zero. It suggests $\theta \to \infty$ minimizes likelihood (which is wrong) or $\theta \to 0$ maximizes it (but violates constraint).

(c) **Logic:** We want to maximize $\frac{1}{\theta^n}$, which means minimizing θ. The smallest valid θ is constrained by the data: θ must be at least as large as the maximum observation. Therefore, $\hat{\theta}_{MLE} = X_{(n)}$.

(d) **Bias Check:** $E[\hat{\theta}] = \frac{n}{n+1}\theta$. Bias $= \frac{n}{n+1}\theta - \theta = \frac{-\theta}{n+1}$. Unbiased Estimator: $\hat{\theta}_U = \frac{n+1}{n}X_{(n)}$.

(c) $\hat{\theta}_{MLE} = X_{(n)}$
(d) Unbiased $= \frac{n+1}{n}X_{(n)}$

8.10 Shifted Exponential

Problem: $f(x) = \lambda e^{-\lambda(x-\delta)}, x \ge \delta$.
Solution:

(a) **Estimate δ:** The likelihood contains term $\mathbb{I}(x_i \ge \delta) = \mathbb{I}(\min(x_i) \ge \delta)$. To maximize likelihood, we want $\lambda e^{-\lambda(x-\delta)}$ to be large $\to e^{\lambda\delta}$ large $\to \delta$ large. Constraint: $\delta \le \min(x_i)$. $\hat{\delta}_{MLE} = X_{(1)}$.

(b) **Estimate λ:** $\ln L = n \ln \lambda - \lambda \sum(x_i - \hat{\delta})$. Diff w.r.t λ: $\frac{n}{\lambda} - \sum(x_i - \hat{\delta}) = 0$. $\hat{\lambda} = \frac{n}{\sum(x_i - \hat{\delta})}$.

(a) $\hat{\delta} = \min(X_i)$
(b) $\hat{\lambda} = n / \sum(x_i - \min(x_i))$

8.11 Optimal Scaling (MSE Trade-off)

Problem: $N(0, \sigma^2)$, $\hat{\sigma}_c^2 = c \sum X_i^2$.
Solution:

(a) **Bias and Variance:** Recall $X_i^2/\sigma^2 \sim \chi_1^2$. Sum $\sim \chi_n^2$. $E[\sum X_i^2] = n\sigma^2$. $Var(\sum X_i^2) = 2n\sigma^4$. $Bias = c(n\sigma^2) - \sigma^2 = \sigma^2(cn - 1)$. $Var = c^2(2n\sigma^4)$.

(b) **Minimize MSE:** $MSE = 2nc^2\sigma^4 + \sigma^4(cn - 1)^2$. Diff w.r.t c: $4nc + 2(cn - 1)n = 0$. $2c + cn - 1 = 0 \implies c(n + 2) = 1$.

$$c = \frac{1}{n + 2}$$

(c) **Comparison:** Unbiased: $c = 1/n$. MLE: $c = 1/n$. MSE Optimal: $c = 1/(n + 2)$. Again, shrinking the estimator slightly reduces variance more than it adds squared bias.

$\boxed{\text{(b) } c = 1/(n + 2)}$

8.4 Application

8.12 The Bias-Variance Trade-off (Simulation)

Problem: Compare S^2 (n-1) vs MLE (n).
Solution:

```
import numpy as np
import matplotlib.pyplot as plt

true_var = 100
n = 5
num_sims = 10000
data = np.random.normal(0, np.sqrt(true_var), (num_sims, n))

# Estimators
var_unbiased = np.var(data, axis=1, ddof=1)
var_mle = np.var(data, axis=1, ddof=0)

# Metrics
print(f"Unbiased MSE: {np.mean((var_unbiased - true_var)**2):.2f}")
print(f"MLE MSE:      {np.mean((var_mle - true_var)**2):.2f}")

# Plot
plt.figure(figsize=(10, 6))
plt.hist(var_unbiased, bins=50, alpha=0.5, label='Unbiased (n-1)',
    density=True)
plt.hist(var_mle, bins=50, alpha=0.5, label='MLE (n)', density=True)
plt.axvline(true_var, color='r', ls='--', label='True Var')
plt.legend()
plt.show()
```

(a) Unbiased is centered at red line. MLE is shifted left (underestimates).
(b) MLE has narrower spread.
(c) MLE usually has lower MSE because $1/n$ shrinks variance enough to outweigh the small bias squared.

8.13 The German Tank Problem

Problem: $U(0, \theta)$. Max vs 2Mean vs Corrected.
Solution:

(a) **Negative Bias:** The sample maximum can never exceed the population maximum. It is always $\leq \theta$. So $E[\max] < \theta$.

(b) **Variance:** MOM ($2\bar{X}$) depends on the sum of variables, having variance $\propto \theta^2/n$. The Max depends on extreme values, which converge to θ much faster (variance $\propto \theta^2/n^2$). Max is much more precise.

(c) **Corrected Max:** It removes the bias of the Max estimator while keeping its low variance property. It is the MVUE (Minimum Variance Unbiased Estimator).

$\boxed{\text{Corrected Max is the best: Unbiased + Low Variance.}}$