Mathematical Statistics

Tutorial 4

1. Let $X_1, X_2, \ldots, X_n$ be an independent and identically distributed (i.i.d.) random sample from a population X.

   Assume that the population variance is $V(X) = \sigma^2 < \infty$ and that the fourth central moment, μ_4, exists and is finite:
$$\mu_4 = E\left[(X - E[X])^4\right] < \infty$$

   Then, the sample variance, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$, is **asymptotically normally distributed** with a mean of $V(X)$ and an asymptotic variance given by $\frac{1}{n}(\mu_4 - \sigma^4)$.

   Formally, the standardized sequence converges in distribution to a Normal distribution:
$$\sqrt{n}(S^2 - V(X)) \xrightarrow{d} \mathcal{N}\left(0, \mu_4 - (V(X))^2\right)$$

   where the asymptotic variance is often written as $\mu_4 - \sigma^4$.

2. Let $X_1, \ldots, X_n$ be a random sample from an Exponential distribution with mean parameter θ. The Probability Density Function (PDF) is given by $f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$ for $x > 0$. We know that the variance of the distribution is $q(\theta) = \mathrm{Var}(X) = \theta^2$.

   (a) Calculate the Fisher Information per observation, $I(\theta)$.

   (b) Find the MLE of θ ($\hat{\theta}_n$) and its asymptotic distribution.

   (c) We are interested in estimating the variance $q(\theta) = \mathrm{Var}(X)$. Propose the MLE estimator of $q(\theta)$ and find its asymptotic distribution.

   (d) Another way to estimate $q(\theta)$ is with the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$. Calculate the asymptotic distribution of S^2. If the sample size is large, which estimator for the variance would be preferred, and why?

3. Let $X_1, \ldots, X_n$ be a random sample from a Normal distribution $\mathcal{N}(\mu, \sigma^2)$. Derive the $(1-\alpha)100\%$ confidence interval for the population variance σ^2 in two cases: (a) μ is known, and (b) μ is unknown.

Problem 1

Let $X_1, X_2, \ldots, X_n$ be an independent and identically distributed (i.i.d.) random sample from a population X.

Assume that the population variance is $V(X) = \sigma^2 < \infty$ and that the fourth central moment, μ_4, exists and is finite:

$$\mu_4 = E\left[(X - E[X])^4\right] < \infty$$

Then, the sample variance, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$, is asymptotically normally distributed with a mean of $V(X)$ and an asymptotic variance given by $\frac{1}{n}(\mu_4 - \sigma^4)$.

Formally, the standardized sequence converges in distribution to a Normal distribution:

$$\sqrt{n}(S^2 - V(X)) \xrightarrow{d} \mathcal{N}\left(0, \mu_4 - (V(X))^2\right)$$

where the asymptotic variance is often written as $\mu_4 - \sigma^4$.

Solutions:

We use the algebraic identity for the sum of squares:

$$\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \sum_{i=1}^{n}(X_i - E[X])^2 - n(\bar{X}_n - E[X])^2$$

To find the asymptotic distribution of $\sqrt{n}(S^2 - \sigma^2)$, note that,

$$\begin{aligned}
\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 - \sigma^2\right) &= \sqrt{n}\left(\frac{1}{n}\left[\sum_{i=1}^{n}(X_i - E[X])^2 - n(\bar{X}_n - E[X])^2\right] - \sigma^2\right) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - E[X])^2 - \sigma^2\right) - \sqrt{n}(\bar{X}_n - E[X])^2 \\
&= A + B
\end{aligned}$$

Let $Y_i = (X_i - E[X])^2$. Since X_i are i.i.d., Y_i are also i.i.d. We find the mean and variance of Y_i:

- **Mean of Y_i**: $E[Y_i] = E[(X_i - E[X])^2] = V(X) = \sigma^2$.
- **Variance of Y_i**: $\mathrm{Var}(Y_i) = E[Y_i^2] - (E[Y_i])^2 = E[(X_i - E[X])^4] - (V(X))^2 = \mu_4 - \sigma^4$.

By the **Central Limit Theorem (CLT)** applied to the sample mean of Y_i, $\bar{Y}_n = \frac{1}{n}\sum Y_i$:

$$\sqrt{n}(\bar{Y}_n - E[Y_i]) = \sqrt{n}(\bar{Y}_n - \sigma^2) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - E[X])^2 - \sigma^2\right) \xrightarrow{d} \mathcal{N}(0, \mu_4 - \sigma^4) \quad \text{(Term A)}$$

We analyze the convergence of the Residual Term (Term B):

$$\text{Term B} = \sqrt{n}(\bar{X}_n - E[X])^2$$

From the CLT applied to $\bar{X}_n$:

$$\sqrt{n}(\bar{X}_n - E[X]) \xrightarrow{d} Z \sim \mathcal{N}(0, \sigma^2)$$

We rewrite Term B by factoring $\frac{1}{\sqrt{n}}$:

$$\sqrt{n}(\bar{X}_n - E[X])^2 = \frac{1}{\sqrt{n}}\left[\sqrt{n}(\bar{X}_n - E[X])\right]^2$$

The first part, $\frac{1}{\sqrt{n}}$, converges to 0 in probability. The second part, $\left[\sqrt{n}(\bar{X}_n - E[X])\right]^2$, converges in distribution to Z^2 (which is finite almost surely).

By the continuous mapping theorem and properties of convergence, the product of a sequence converging to zero in probability and a sequence converging in distribution is a sequence that converges to zero in probability:

$$\sqrt{n}(\bar{X}_n - E[X])^2 \xrightarrow{p} 0$$

Therefore,by **Slutsky's Theorem**, the term that converges to zero is negligible in the limit:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 - \sigma^2\right) = \underbrace{\sqrt{n}(\bar{Y}_n - \sigma^2)}_{\xrightarrow{d} \mathcal{N}(0,\mu_4 - \sigma^4)} - \underbrace{\sqrt{n}(\bar{X}_n - E[X])^2}_{\xrightarrow{p} 0} \overset{d}{=\!\!\rightarrow} \mathcal{N}\left(0, \mu_4 - \sigma^4\right)$$

Finally, since $S^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ and $\frac{n}{n-1} \to 1$, S^2 and $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ are asymptotically equivalent, sharing the same limiting distribution.

$$\sqrt{n}(S^2 - \sigma^2) \xrightarrow{d} \mathcal{N}\left(0, \mu_4 - \sigma^4\right)$$

Problem 2

Probability Density Function (PDF) is given by $f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$ for $x > 0$. We know that the variance of the distribution is $q(\theta) = \mathrm{Var}(X) = \theta^2$.

(a) Calculate the Fisher Information per observation, $I(\theta)$.

(b) Find the MLE of θ ($\hat{\theta}_n$) and its asymptotic distribution.

(c) We are interested in estimating the variance $q(\theta) = \mathrm{Var}(X)$. Propose the MLE estimator of $q(\theta)$ and find its asymptotic distribution.

(d) Another way to estimate $q(\theta)$ is with the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$. Calculate the asymptotic distribution of $\sqrt{n}(S^2 - \sigma^2)$.

(e) If the sample size is large, which estimator for the variance would be preferred, and why?

Solution

The Fisher Information per observation is $I(\theta) = -E\left[\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\right]$.

1. Log-Likelihood for one observation:

$$\log f(X;\theta) = -\log(\theta) - \frac{X}{\theta}$$

2. Second Derivative:

$$\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2} = \frac{\partial}{\partial\theta}\left(-\frac{1}{\theta} + \frac{X}{\theta^2}\right) = \frac{1}{\theta^2} - \frac{2X}{\theta^3}$$

3. Fisher Information: Using $E[X] = \theta$:

$$I(\theta) = -E\left[\frac{1}{\theta^2} - \frac{2X}{\theta^3}\right] = -\left(\frac{1}{\theta^2} - \frac{2E[X]}{\theta^3}\right) = -\left(\frac{1}{\theta^2} - \frac{2\theta}{\theta^3}\right) = \frac{1}{\theta^2}$$

$$I(\theta) = \boldsymbol{\frac{1}{\theta^2}}$$

MLE of θ ($\hat{\theta}_n$): The MLE for the Exponential mean parameter is the sample mean:

$$\hat{\theta}_n = \bar{\mathbf{X}}_\mathbf{n}$$

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right)$$

Substituting $I(\theta) = 1/\theta^2$:

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{\theta^2}\right)$$

We are estimating $q(\theta) = \theta^2$, by the invariance property of MLE, $\hat{q}_n = q(\hat{\theta}_n) = (\bar{X}_n)^2$. We use the Delta Method with $q(\theta) = \theta^2$ and $q'(\theta) = 2\theta$. The Asymptotic Variance (AV) is

$$\mathrm{AV}(\bar{X}_n^2) = \frac{(2\theta)^2}{1/\theta^2} = 4\theta^4$$

The asymptotic distribution is:

$$\sqrt{n}(\bar{X}_n^2 - \theta^2) \xrightarrow{d} \mathcal{N}\left(0, 4\theta^4\right)$$

Asymptotic Distribution of $\sqrt{n}(S^2 - \sigma^2)$

1. **Calculation of Asymptotic Variance (AV):**

   - Variance: $\sigma^2 = \theta^2 \implies \sigma^4 = \theta^4$.
   - Fourth Central Moment for Exp(θ): $\mu_4 = 9\theta^4$.

$$\text{AV}(S^2) = \mu_4 - \sigma^4 = 9\theta^4 - \theta^4 = 8\theta^4$$

$$\sqrt{n}(S^2 - \sigma^2) \xrightarrow{d} \mathcal{N}\left(0, 8\theta^4\right)$$

We compare the Asymptotic Variances:

Estimator	**Asymptotic Variance (AV)**
$\bar{X}_n^2$	$4\theta^4$
S^2	$8\theta^4$

Since $\text{AV}(\bar{X}_n^2) < \text{AV}(S^2)$, the estimator $\bar{X}_n^2$ is **asymptotically more efficient**. For a large sample, the estimator $\mathbf{\bar{X}_n^2}$ would be preferred, as it achieves a smaller asymptotic variance and is thus more precise.

Problem 3

Let $X_1, \dots, X_n$ be a random sample from a Normal distribution $\mathcal{N}(\mu, \sigma^2)$. Derive the $(1-\alpha)100\%$ confidence interval for the population variance σ^2 in two cases: (a) μ is known, and (b) μ is unknown.

Solution

(a) Case 1: Mean μ is Known

- Pivot Quantity: $Y_1 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2_n$.
- Critical Values: We use the χ^2 quantiles $\chi^2_{\alpha/2,n}$ (lower) and $\chi^2_{1-\alpha/2,n}$ (upper, standard notation):

$$P\left(\chi^2_{\alpha/2,n} < Y_1 < \chi^2_{1-\alpha/2,n}\right) = 1-\alpha$$

- Invert to find σ^2:

$$\mathrm{CI}(\sigma^2) = \left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2,n}}, \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2,n}}\right]$$

(b) Case 2: Mean μ is Unknown

- Pivot Quantity: $Y_2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.
- Critical Values: We use the χ^2 quantiles with $df = n-1$:

$$P\left(\chi^2_{\alpha/2,n-1} < Y_2 < \chi^2_{1-\alpha/2,n-1}\right) = 1-\alpha$$

- Invert to find σ^2:

$$\mathrm{CI}(\sigma^2) = \left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}\right]$$