OS02 - Written Exam 2 - Theory of Estimation - Solutions

Problem 1 (≈ 6 points)**.** *The covariance between two jointly distributed (under $P_{\xi,\eta}$) random variables ξ and η is defined as follows*

$$\gamma = \mathbb{E}[(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta)].$$

The independent random vectors

$$\begin{pmatrix} \xi_1 \\ \eta_1 \end{pmatrix}, \begin{pmatrix} \xi_2 \\ \eta_2 \end{pmatrix}, \ldots, \begin{pmatrix} \xi_n \\ \eta_n \end{pmatrix}$$

are assumed to come from the distribution $P_{\xi,\eta}$. It is also assumed that $\mathbb{E}|\xi| < \infty$, $\mathbb{E}|\eta| < \infty$ and $\mathbb{E}|\xi\eta| < \infty$. The sampling covariance is defined as follows

$$\widehat{\gamma} = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{\xi})(\eta_i - \overline{\eta}), \quad \overline{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i, \;\; \overline{\eta} = \frac{1}{n}\sum_{i=1}^{n}\eta_i.$$

1. *Prove that $\widehat{\gamma}$ is a consistent estimator of γ (by using theorems and definitions).*

1. Solution : First of all, let us re-write the theoretical covariance γ and sampling covariance $\widehat{\gamma}$ in the following manner

$$\gamma = \mathbb{E}[(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta)] = \mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\cdot\mathbb{E}(\eta),$$

$$\widehat{\gamma} = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{\xi})(\eta_i - \overline{\eta}) = \frac{1}{n}\sum_{i=1}^{n}\xi_i\eta_i - \overline{\xi}\cdot\overline{\eta} = \frac{1}{n}\sum_{i=1}^{n}\xi_i\eta_i - \frac{1}{n}\sum_{i=1}^{n}\xi_i \cdot \frac{1}{n}\sum_{i=1}^{n}\eta_i.$$

Second, as it follows from the law of large numbers (assuming that $\mathbb{E}|\xi| < \infty$, $\mathbb{E}|\eta| < \infty$ and $\mathbb{E}|\xi\eta| < \infty$), if $n \to \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i \underset{\mathbb{P}}{\to} \mathbb{E}(\xi), \quad \frac{1}{n}\sum_{i=1}^{n}\eta_i \underset{\mathbb{P}}{\to} \mathbb{E}(\eta), \text{ and } \frac{1}{n}\sum_{i=1}^{n}\xi_i\eta_i \underset{\mathbb{P}}{\to} \mathbb{E}(\xi\cdot\eta).$$

Third, let us define the following continuous real function f of three real variables :

$$(x, y, z) \mapsto f(x, y, z) = x - y\cdot z.$$

Finally, by using the continuity theorem, we get (when $n \to \infty$)

$$\widehat{\gamma} = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{\xi})(\eta_i - \overline{\eta}) = f\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\eta_i, \frac{1}{n}\sum_{i=1}^{n}\xi_i, \frac{1}{n}\sum_{i=1}^{n}\eta_i\right) \underset{\mathbb{P}}{\to} f\left(\mathbb{E}(\xi\eta), \mathbb{E}(\xi), \mathbb{E}(\eta)\right) = \gamma.$$

Hence, as it follows from the definition, $\widehat{\gamma}$ is a consistent estimator of γ.

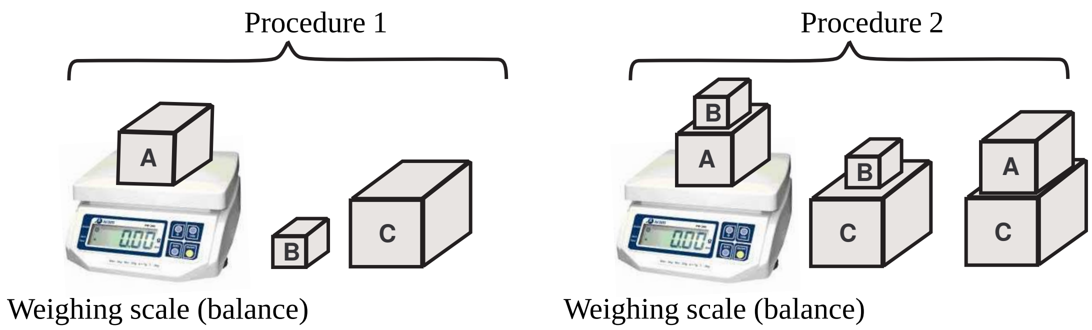

FIGURE 1 – Two different weighing procedures.

Problem 2 ($\approx$ 7 points)**.** *Let us consider three objects A, B and C. The true masses of these objects are denoted by α, β and γ, respectively. We have a weighing scale (balance), see Figure 1. We wish to compare two different weighing procedures to estimate the unknown masses α, β and γ :*

1. *individual weighing : we repeat this procedure 3 times (for A, B and C). Only one object (A or B or C) is installed on the balance at once.*
2. *pair weighing : we repeat this procedure 3 times (for A+B, B+C and A+C). Two objects (A+B or B+C or A+C) are installed on the balance at once.*

These procedures are illustrated in Figure 1. Let us consider the following additive measurement model of balance :

$$y = x + \xi, \quad \xi \sim \mathcal{N}(0, \sigma^2),$$

where x is the true mass of object, y is the measured (estimated) value of x and ξ is measurement noise. It is assumed that the variance of noise σ^2 is constant (and independent of the true mass).

1. *Write down the measurement models for both weighing procedures.*
2. *Find the estimators $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\gamma}$ for the unknown masses α, β and γ, respectively, by the method of moments for both weighing procedures. The first order moment can be used (see the hint on page 3).*
3. *Calculate the variances of the estimators $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\gamma}$ for each procedure.*
4. *Compare the variances of the first procedure with the variances of the second procedure. Which procedure is more precise (i.e. which variances are smaller) ?*

Solution :

1. (1 point) The measurement model of procedure 1 :

$$\begin{cases} y_1 &= \alpha + \xi_1 \\ y_2 &= \beta + \xi_2 \\ y_3 &= \gamma + \xi_3 \end{cases}, \quad \xi_i \sim \mathcal{N}(0, \sigma^2).$$

The measurement model of procedure 2 :

$$\left\{\begin{array}{lllllll} y_1 & = & \alpha & + & \beta & + & \xi_1 \\ y_2 & = & \beta & + & \gamma & + & \xi_2 \\ y_3 & = & \alpha & + & \gamma & + & \xi_3 \end{array}\right. , \ \xi_i \sim \mathcal{N}(0,\sigma^2).$$

2. (2.5 points) The method of moments estimator for procedure 1 :

$$\left\{\begin{array}{lll} \mathbb{E}(y_1) & = & \alpha \\ \mathbb{E}(y_2) & = & \beta \\ \mathbb{E}(y_3) & = & \gamma \end{array}\right. \quad \Rightarrow \quad \left\{\begin{array}{lll} \widehat{\alpha} & = & y_1 \\ \widehat{\beta} & = & y_2 \\ \widehat{\gamma} & = & y_3 \end{array}\right. .$$

The method of moments estimator for procedure 2 :

$$\left\{\begin{array}{lllll} \mathbb{E}(y_1) & = & \alpha & + & \beta \\ \mathbb{E}(y_2) & = & \beta & + & \gamma \\ \mathbb{E}(y_3) & = & \alpha & + & \gamma \end{array}\right. \quad \Rightarrow \quad \left\{\begin{array}{lllll} \widehat{\alpha} & + & \widehat{\beta} & = & y_1 \\ \widehat{\beta} & + & \widehat{\gamma} & = & y_2 \\ \widehat{\alpha} & + & \widehat{\gamma} & = & y_3 \end{array}\right.$$

or

$$\widehat{\theta} = B^{-1}Y \ \Leftrightarrow \ \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

3. (2.5 points) The variances of the estimators for procedure 1 are :

$$\left\{\begin{array}{lllll} \widehat{\alpha} & = & \alpha & + & \xi_1 \\ \widehat{\beta} & = & \beta & + & \xi_2 \\ \widehat{\gamma} & = & \gamma & + & \xi_3 \end{array}\right. , \ \xi_i \sim \mathcal{N}(0,\sigma^2) \ \Rightarrow \ \mathrm{var}\left(\widehat{\alpha}\right) = \mathrm{var}\left(\widehat{\beta}\right) = \mathrm{var}\left(\widehat{\gamma}\right) = \sigma^2.$$

The expectation of the estimator for procedure 2 is

$$\mathbb{E}\left(\widehat{\theta}\right) = \mathbb{E}\left[B^{-1}\left(B\theta + \xi\right)\right] = \theta + \mathbb{E}\left[B^{-1}\xi\right] = \theta$$

and the variance-covariance matrix of the estimator is

$$\begin{aligned} \mathrm{cov}\left(\widehat{\theta}\right) &= \mathbb{E}\left[\left(\widehat{\theta}-\theta\right)\left(\widehat{\theta}-\theta\right)^T\right] = \mathbb{E}\left[B^{-1}\xi\xi^T B^{-T}\right] = \sigma^2 B^{-1}B^{-T} \\ &= \sigma^2 \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} = \sigma^2 \begin{pmatrix} \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{pmatrix}. \end{aligned}$$

Hence, the variances of the estimators for procedure 2 are

$$\mathrm{var}\left(\widehat{\alpha}\right) = \mathrm{var}\left(\widehat{\beta}\right) = \mathrm{var}\left(\widehat{\gamma}\right) = \frac{3\sigma^2}{4}.$$

4. (1 point) The variances of procedure 2 are smaller : $\frac{3\sigma^2}{4} < \sigma^2$.

Problem 3 (≈ 7 points)**.** *The independent observations* $(\xi_1, \ldots, \xi_n)$ *are assumed to come from the Poisson distribution* $\Pi(\lambda)$, *where* $\lambda > 0$ *is a parameter. But we wish to estimate another parameter* θ, *which is defined as the following function of* λ *:*

$$\theta = \theta(\lambda) = \lambda e^{-\lambda},$$

by using the same observations $(\xi_1, \ldots, \xi_n)$. *Let us consider the two following heuristic estimators :*

$$\widehat{\theta}_1 = \overline{\xi} e^{-\overline{\xi}}, \quad \textit{where} \quad \overline{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i,$$

$$\widehat{\theta}_2 = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_{\{\xi_j=1\}}, \ \textit{where} \ \mathbb{I}_{\{\xi_j=1\}} = \begin{cases} 1 & \textit{si} \quad \xi_j = 1 \\ 0 & \textit{si} \quad \xi_j \neq 1 \end{cases} \quad \textit{is the indicator function of the event} \ \{\xi_j = 1\}.$$

1. *Is the estimator* $\widehat{\theta}_1$ *biased ? If the answer is "yes", then calculate the mean error* $\lambda \to b_1(\lambda)$. *Is the estimator* $\widehat{\theta}_1$ *asymptotically biased ?*
2. *Is the estimator* $\widehat{\theta}_2$ *biased ? If the answer is "yes", then calculate the mean error* $\lambda \to b_2(\lambda)$. *Is the estimator* $\widehat{\theta}_2$ *asymptotically biased ?*
3. *Can* $\widehat{\theta}_2$ *be interpreted as a method of moments estimator ? If the answer is "yes", how to choose the function* $x \mapsto g(x)$ *?*
4. *Are the estimators* $\widehat{\theta}_1$ *and* $\widehat{\theta}_2$ *consistent ?*

   Hint. *If* $\xi_i \sim \Pi(\lambda)$, *then* $\sum_{i=1}^{n} \xi_i \sim \Pi(n\lambda)$.

   Hint.

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad \textit{for} \quad |x| < \infty.$$

Solution :

1. (3 points) Calculate the expectation of $\widehat{\theta}_1 = \overline{\xi} e^{-\overline{\xi}}$:

$$\mathbb{E}\left(\widehat{\theta}_1\right) = \mathbb{E}\left(\overline{\xi} e^{-\overline{\xi}}\right) = \mathbb{E}\left(\frac{\zeta}{n} e^{-\frac{\zeta}{n}}\right),$$

where $\zeta = n\overline{\xi}$. If $\xi_i \sim \Pi(\lambda)$, then $\zeta = n\overline{\xi} = \sum_{i=1}^{n} \xi_i \sim \Pi(n\lambda)$. Therefore (taking into account the hints)

$$\begin{aligned} \mathbb{E}\left(\widehat{\theta}_1\right) &= \sum_{k=0}^{\infty} \frac{k}{n} e^{-\frac{k}{n}} \mathbb{P}\left(\zeta = k\right) = \sum_{k=0}^{\infty} \frac{k}{n} e^{-\frac{k}{n}} \frac{(n\lambda)^k}{k!} e^{-n\lambda} = \frac{e^{-n\lambda}}{n} \sum_{k=1}^{\infty} \frac{(n\lambda e^{-\frac{1}{n}})^k}{(k-1)!} \\ &= (n\lambda e^{-\frac{1}{n}}) \frac{e^{-n\lambda}}{n} \sum_{k-1=0}^{\infty} \frac{(n\lambda e^{-\frac{1}{n}})^{k-1}}{(k-1)!} = (n\lambda e^{-\frac{1}{n}}) \frac{e^{-n\lambda}}{n} e^{(n\lambda e^{-\frac{1}{n}})} = \lambda e^{n\lambda\left(e^{-\frac{1}{n}} - 1\right) - \frac{1}{n}}. \end{aligned}$$

Finally, we get

$$\mathbb{E}\left(\widehat{\theta}_1\right) = \lambda e^{n\lambda\left(e^{-\frac{1}{n}} - 1\right) - \frac{1}{n}} \neq \theta(\lambda) = \lambda e^{-\lambda},$$

hence, the estimator $\widehat{\theta}_1$ is biased and the mean error is $\lambda \to b_1(\lambda) = \lambda e^{n\lambda\left(e^{-\frac{1}{n}}-1\right)-\frac{1}{n}} - \lambda e^{-\lambda}$. By using the hint, we get after simple algebra that

$$\lim_{n\to\infty} \lambda e^{n\lambda\left(\sum_{k=0}^{\infty}\frac{(-1/n)^k}{k!}-1\right)-\frac{1}{n}} = \lambda e^{-\lambda}.$$

It implies that the estimator $\widehat{\theta}_1$ is asymptotically unbiased.

2. (1 point) Calculate the expectation of $\widehat{\theta}_2 = \frac{1}{n}\sum_{j=1}^{n}\mathbb{I}_{\{\xi_j=1\}}$:

$$\mathbb{E}\left(\widehat{\theta}_2\right) = \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n}\mathbb{I}_{\{\xi_j=1\}}\right) = n\frac{1}{n}\mathbb{E}\left(\mathbb{I}_{\{\xi_j=1\}}\right) = \mathbb{P}(\xi_i = 1) = \lambda e^{-\lambda} = \theta(\lambda).$$

Hence, the estimator $\widehat{\theta}_2$ is unbiased.

3. (1 point) Yes, the estimator $\widehat{\theta}_2$ can be interpreted as a method of moments estimator. The function g is defined as follows $x \mapsto g(x) = \mathbb{I}_{\{x=1\}}$.

4. (2 points) As it follows from the law of large numbers, if $n \to \infty$, then $\overline{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i \underset{\mathbb{P}}{\to} \mathbb{E}(\xi) = \lambda$. Taking into account the continuity theorem with the continuous function $f(x) = xe^{-x}$, we get

$$\widehat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n}\xi_i e^{-\frac{1}{n}\sum_{i=1}^{n}\xi_i} \underset{\mathbb{P}}{\to} \lambda e^{-\lambda} = \theta(\lambda).$$

The estimator $\widehat{\theta}_1$ is consistent.

Let us define a sequence of independent binary random variables $\mathbb{I}_{\{\xi_1=1\}}, \mathbb{I}_{\{\xi_2=1\}}, \ldots, \mathbb{I}_{\{\xi_n=1\}}$ ($\mathbb{I}_{\{\xi_i=1\}}$ obeys a Bernoulli distribution). As it follows from the law of large numbers, if $n \to \infty$, then

$$\widehat{\theta}_2 = \frac{1}{n}\sum_{j=1}^{n}\mathbb{I}_{\{\xi_j=1\}} \underset{\mathbb{P}}{\to} \mathbb{E}\left(\mathbb{I}_{\{\xi_i=1\}}\right) = \lambda e^{-\lambda} = \theta(\lambda).$$

The estimator $\widehat{\theta}_2$ is consistent.