RBOT101: MATHEMATICAL FOUNDATIONS OF ROBOTICS

SUMMER 2021

Parameter Estimation

Benjamin Gravell

The University of Texas at Dallas

Department of Mechanical Engineering

P4-1.

Taking expectation of the estimator expression for $\hat{\sigma}_X^2(n)$ with respect to the random variables X_i and expanding the square we have

$$
\begin{aligned}
\mathbb{E}[\hat{\sigma}_X^2(n)] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2\right] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(X_i-\frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\right] && (1)\\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(X_i-\frac{1}{n}\sum_{j=1}^{n}X_j\right)\left(X_i-\frac{1}{n}\sum_{j=1}^{n}X_j\right)\right] && (2)\\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i^2-\frac{2}{n}X_i\sum_{j=1}^{n}X_j+\frac{1}{n^2}\left(\sum_{k=1}^{n}X_k\right)\left(\sum_{\ell=1}^{n}X_\ell\right)\right] && (3)\\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i^2-\frac{2}{n}X_i^2-\frac{2}{n}\sum_{i\neq j}X_iX_j+\frac{1}{n^2}\sum_{k=1}^{n}X_k+\frac{1}{n^2}\sum_{k\neq\ell}\sum_{\ell=1}^{n}X_kX_\ell\right] && (4)
\end{aligned}
$$

where the last three summations within the brackets have $n-1$, n, and $n(n-1)$ terms respectively. Substituting the second moment expressions from part 2. we obtain

$$
\begin{aligned}
\mathbb{E}[\hat{\sigma}_X^2(n)] &= \frac{1}{n}\sum_{i=1}^{n}\left[(\sigma^2+\mu_X^2)-\frac{2}{n}(\sigma^2+\mu_X^2)-\frac{2}{n}\sum_{i\neq j}\mu_X^2+\frac{1}{n^2}\sum_{k=1}^{n}(\sigma^2+\mu_X^2)+\frac{1}{n^2}\sum_{k\neq\ell}\sum_{\ell=1}^{n}\mu_X^2\right] && (5)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\left(1-\frac{2}{n}+\frac{n}{n^2}\right)(\sigma^2+\mu_X^2)+\left(-\frac{2(n-1)}{n}+\frac{n(n-1)}{n^2}\right)\mu_X^2\right] && (6)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\left(1-\frac{1}{n}\right)(\sigma^2+\mu_X^2)+\left(-\frac{2(n-1)}{n}+\frac{(n-1)}{n}\right)\mu_X^2\right] && (7)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\left(1-\frac{1}{n}\right)(\sigma^2+\mu_X^2)-\left(\frac{n-1}{n}\right)\mu_X^2\right] && (8)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{n-1}{n}\right)(\sigma^2+\mu_X^2)-\left(\frac{n-1}{n}\right)\mu_X^2\right] && (9)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{n-1}{n}\right)\sigma^2 = \frac{1}{n}\cdot n\cdot\left(\frac{n-1}{n}\right)\sigma^2 && (10)\\
&= \left(\frac{n-1}{n}\right)\sigma^2 && (11)
\end{aligned}
$$

1. Thus $\hat{\sigma}^2_X(n)$ is a **biased** estimate of σ^2 since $\mathbb{E}[\hat{\sigma}^2_X(n)] \neq \sigma^2$.
2. Specifically, the current estimation scheme produces underestimates of the variance scaled by a factor of $(n-1)/n$.
3. However $\hat{\sigma}^2_X(n)$ is asymptotically unbiased, i.e. $\lim_{n\to\infty} \mathbb{E}[\hat{\sigma}^2_X(n)] = \sigma^2_X$.
4. To remedy this all we have to do is invert the scaling of the estimated variance according to the factor identified i.e. use the modified estimator

$$\hat{\sigma}^2_X(n) := \frac{1}{\frac{n-1}{n}} \cdot \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu}_X(n))^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu}_X(n))^2 \tag{12}$$

which is known as the **unbiased sample variance** using **Bessel's correction**.

P4-2.

Starting from the log-likelihood

$$\log L(\mu,\sigma) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{13}$$

the stationary point is found as

$$0 = \frac{\partial \log L(\mu,\sigma)}{\partial \mu}\Big|_{\mu^*,\sigma^*} \tag{14}$$

$$0 = \frac{\partial \log L(\mu,\sigma)}{\partial \sigma}\Big|_{\mu^*,\sigma^*}. \tag{15}$$

These become the system of equations

$$\sum_{i=1}^{n}(x_i-\mu^*) = 0 \tag{16}$$

$$-\frac{n}{\sigma^*} + \frac{1}{\sigma^{*3}}\sum_{i=1}^{n}(x_i-\mu^*)^2 = 0 \tag{17}$$

The first equation implies the maximally likely mean is the sample mean

$$\mu^* = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{18}$$

The second equation implies, by substituting μ^*, the maximally likely variance is the sample variance

$$\sigma^{*2} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu^*)^2 \tag{19}$$

The MLE mean and variance for a Gaussian are the sample mean and variance

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{20}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\hat{\mu})^2 \tag{21}$$

The MLE variance is biased, as shown in the previous problem.

P4-3.

The problem follows the linear measurement model $Y = H\theta + N$ with

$$Y = \begin{bmatrix} 6.2 \\ 7.8 \\ 2.2 \end{bmatrix}, \quad H = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}, \quad N = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} \tag{22}$$

We have

$$H^\top H = \begin{bmatrix} 3 & 4 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} = (3)(3) + (4)(4) + (1)(1) = 26, \tag{23}$$

$$H^\top y = \begin{bmatrix} 3 & 4 & 1 \end{bmatrix} \begin{bmatrix} 6.2 \\ 7.8 \\ 2.2 \end{bmatrix} = (3)(6.2) + (4)(7.8) + (1)(2.2) = 52. \tag{24}$$

So the least-squares estimate is

$$\boxed{\hat{\theta} = (H^\top H)^{-1} H^\top y = \frac{52}{26} = 2.} \tag{25}$$

P4-4.

We are concerned with the probability

$$\mathbb{P}[a \le S_n \le b]. \tag{26}$$

To use the CLT we need to rewrite this probability in terms of a standardized sum random variable

$$Z_n = \sum_{i=1}^{n} \frac{X_i - \mu_i}{s_n} \tag{27}$$

where $s_n^2 = \sum_{i=1}^n \sigma_i^2$. Doing this,

$$\begin{aligned}
\mathbb{P}[a \le S_n \le b] &= \mathbb{P}[a \le \sum_{i=1}^{n} X_i \le b] && (28)\\
&= \mathbb{P}[a - \sum_{i=1}^{n} \mu_i \le \sum_{i=1}^{n} (X_i - \mu_i) \le b - \sum_{i=1}^{n} \mu_i] && (29)\\
&= \mathbb{P}\left[\frac{a - \sum_{i=1}^n}{s_n} \le \sum_{i=1}^{n} \frac{X_i - \mu_i}{s_n} \le \frac{b - \sum_{i=1}^n}{s_n}\right] && (30)\\
&= \mathbb{P}\left[a' \le Z_n \le b'\right] && (31)
\end{aligned}$$

where the shifted and scaled limits are

$$a' = \frac{a - \sum_{i=1}^n \mu_i}{s_n}, \qquad b' = \frac{b - \sum_{i=1}^n \mu_i}{s_n}. \tag{32}$$

By the CLT, Z_n follows a nearly standard normal distribution, so

$$\mathbb{P}[a \le S_n \le b] = \mathbb{P}\left[a' \le Z_n \le b'\right] \approx \Phi(b') - \Phi(a') \tag{33}$$

where Φ_Z is the standard normal cdf.

Using the given problem data, we compute

$$\mu = 1/2, \quad \sigma^2 = 1/12, \tag{34}$$

and

$$s_n = \sqrt{\sum_{i=1}^{n} \sigma_i^2} = \sqrt{n\sigma^2} = \sqrt{100/12} \approx 2.887, \tag{35}$$

and

$$a' = \frac{a - \sum_{i=1}^{n} \mu_i}{s_n} = \frac{45 - \sum_{i=1}^{n} 1/2}{2.887} = \frac{-5}{2.887} = -1.732, \tag{36}$$

$$b' = \frac{b - \sum_{i=1}^{n} \mu_i}{s_n} = \frac{52.5 - \sum_{i=1}^{n} 1/2}{2.887} = \frac{2.5}{2.887} = 0.866. \tag{37}$$

Using `scipy.stats.norm` or a table for the standard normal cdf we obtain

$$\mathbb{P}[a \leq S_n \leq b] \approx \Phi(b') - \Phi(a') \approx 0.8068 - 0.0416 = 0.7652 \tag{38}$$