Los saltamontes de la región africana de Asyut se caracterizan por tener una longitud media de 2 cm, pudiendo admitirse una distribución normal para la longitud de tales ortópteros. Elegida una muestra aleatoria de 20 de ellos, sus longitudes en cm fueron las siguientes:

$$1'90, 1'85, 2'01, 1'95, 2'05, 2'00, 1'97, 2'02, 1'89, 2'01,$$

$$2'05, 1'95, 1'87, 2'05, 1'97, 1'85, 2'02, 1'95, 1'93, 2'05$$

Utilizando estos datos, se pide:

a) **Determinar la estimación de máxima verosimilitud de la desviación típica poblacional σ.**

b) **Calcular la probabilidad de que el estimador de máxima verosimilitud de σ subestime el verdadero valor de dicho parámetro.**

Si llamamos X a la *longitud de los saltamontes en estudio*, del enunciado se desprende que puede admitirse para X una distribución $N(2, \sigma)$, siendo la estimación de σ el objetivo del problema.

a) Para determinar el estimador de máxima verosimilitud de σ comenzaremos calculando la función de verosimilitud (CB-sección 5.2)

$$L(\sigma) = f_\sigma(x_1, ..., x_n) = \prod_{i=1}^{n} f_\sigma(x_i) = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - 2)^2 \right\}$$

de logaritmo neperiano

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

Su derivada igualada a cero (ecuación de verosimilitud) será

$$\frac{\partial}{\partial \sigma} \log L(\sigma) = -\frac{n}{\sigma} + \frac{4\sigma \sum_{i=1}^{n} (x_i - 2)^2}{4\sigma^4} = 0$$

de donde se obtiene el estimador de máxima verosimilitud

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - 2)^2}.$$

(Hacemos aquí la observación de que los estimadores de máxima verosimilitud respetan las transformaciones biyectivas; es decir, que si T es el estimador de máxima verosimilitud para θ, entonces $g(T)$ es el estimador de máxima verosimilitud para $g(\theta)$ siempre que g sea una función biyectiva. En este sentido, podía haberse determinado el estimador de máxima verosimilitud para la varianza poblacional σ^2 —más habitual y ya determinado en CB-sección 5.2— y extrayendo la raíz cuadrada obtendríamos el de la desviación típica σ, al ser la función $g(x) = \sqrt{x}$ una función biyectiva.)

A partir de los datos del enunciado obtenemos que la estimación de máxima verosimilitud será,

$$\hat{\sigma} = \sqrt{\frac{1}{20}\sum_{i=1}^{20}(X_i - 2)^2} = \sqrt{0'00546} = 0'0739.$$

b) La probabilidad pedida será

$$\begin{aligned} P\{\hat{\sigma} < \sigma\} &= P\left\{\frac{1}{20}\sum_{i=1}^{20}(X_i - 2)^2 < \sigma^2\right\} \\ &= P\left\{\frac{1}{\sigma^2}\sum_{i=1}^{20}(X_i - 2)^2 < 20\right\} \\ &= P\{\chi^2_{20} < 20\} = 0'5292 \end{aligned}$$

al tener $\sum_{i=1}^{20}(X_i - 2)^2/\sigma^2$ una distribución χ^2_{20} por estar ante un caso de estimación de la varianza de una población normal de media conocida (CB-sección 5.6), y en donde la última probabilidad se ha calculado por interpolación a partir de la tabla 4: De ella se obtiene que es $P\{\chi^2_{20} < 16'27\} = 0'3$ y $P\{\chi^2_{20} < 22'78\} = 0'7$, por lo que para un aumento de abscisa de $22'78 - 16'27 = 6'51$ se obtiene un aumento de probabilidad de $0'4$; para un aumento de abscisa de $20 - 16'27 = 3'73$ se obtendrá un aumento de probabilidad de $0'4 \times 3'73/6'51 = 0'2292$. Por tanto, $P\{\chi^2_{20} < 20\} = 0'3 + 0'2292 = 0'5292$.