Worms

1. mean: $\bar{X} = \frac{5+25+15+10+7+12+16+20}{8} = 13.75$

variance:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 =$$

$$= \frac{(-8.75)^2 + (11.25)^2 + (1.25)^2 + (-3.75)^2 + (-6.75)^2 + (-1.75)^2 + (2.25)^2}{7} =$$

$$= \frac{310.5}{7} = 44.357^{\sim}$$

2. mean $E[\bar{x}] = E[\frac{1}{n}\sum_{i=1}^{n} x_i]$

variance Mean is known so:

$$Var[\bar{x}] = \sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n} = \frac{310.5}{8} = 38.8125$$

3. Sample mean is 13.75, we know approximately that $(\bar{x} - \mu) \sim N(0, \frac{\sigma^2}{n})$ (CLT) and

$$\sigma^2 \approx S^2 = 44.357 \cong 6.66^2$$

additionally: $P(\bar{x} - 2\frac{\hat{\sigma}}{\sqrt{n}} \leq \mu \leq \bar{x} + 2\frac{\hat{\sigma}}{\sqrt{n}}) \approx 0.95$

$\Rightarrow$ the interval is:

$$(13.75 - 4.7093, 13.75 + 4.7093) =$$
$$= (9.04, 18.45)$$

Blood Samples - 70, 70, 85

$$\bar{x} = \frac{225}{3} = 75 \Rightarrow S^2 = \frac{\sum_{i=1}^{3}(x_i - \bar{x})^2}{n-1} = 75 = (\sqrt{75})^2$$

$$\hat{\sigma} = \sqrt{75}$$

Given that samples distribute normally:

$$P\left(\bar{x} - 2\frac{\hat{\sigma}}{\sqrt{n}} \le \mu \le \bar{x} + 2\frac{\hat{\sigma}}{\sqrt{n}}\right) \approx 0.95$$

$\Rightarrow$ interval would be: $(75-10, 75+10) = (65, 85)$

Waiting Time

Samples - 1, 1, 2, 3, 5, 8, 10, 21, 38, 81 mins

$$X \sim \exp(\lambda) \Rightarrow \mu = \frac{1}{\lambda} \qquad Var(X) = \frac{1}{\lambda^2}, \lambda > 0$$

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{170}{10} = 17 \Rightarrow S^2 = \frac{\sum_{i=1}^{10}(x_i - 17)^2}{9} = \frac{5760}{9} = 640$$

1. $\bar{x} = \frac{1}{\lambda} \rightarrow \hat{\lambda} = \frac{1}{\bar{x}} = \frac{1}{17}$

2. $S^2 = \frac{1}{\lambda^2} \Rightarrow \hat{\lambda} = \pm\sqrt{\frac{1}{S^2}} = \frac{1}{\sqrt{640}} = \frac{1}{8\sqrt{10}}$

The difference can be explained by low # of samples as I see it:

Even though we used unbiased estimators the # samples is very low which causes unexpected and inaccurate results when using method of moments (if # samples is high + LLN $\Rightarrow$ values should converge)

Scanned with CamScanner

City sizes

$X_{min} = 9000$ $\qquad$ $\bar{X} = 100{,}000$

$X \sim Pareto(X_{min}, \alpha = ?)$

1. $E[X] = \begin{cases} \infty & \alpha \leq 1 \\ \frac{\alpha X_{min}}{\alpha - 1} & \alpha > 1 \end{cases}$

$$\Rightarrow \bar{X} = \frac{\hat{\alpha} X_{min}}{\hat{\alpha} - 1}$$

$$100000(\hat{\alpha} - 1) = 9000\hat{\alpha}$$

$$\hat{\alpha} = \frac{100}{91} \approx 1.099$$

2. $P(X > 1{,}000{,}000) \approx \left(\frac{9000}{1{,}000{,}000}\right)^{\hat{\alpha}} \cong 5.648 \cdot 10^{-3}$

$\Rightarrow 0.5648\ \%$

Since there are 1000 towns we can expect 6 towns (5.648~ towns) with at least 1 M citizens.

3. Quantile 99.9

$$\Rightarrow \hat{Q}_{99.9} = X_{min}(1 - 0.999)^{-\frac{1}{\hat{\alpha}}} =$$

$$= 9000(1/1000)^{-\frac{91}{100}} \approx 4{,}833{,}286$$

Uniform Moments:

1. $f(x) = \frac{1}{b-a} \qquad a<x<b$

first moment:

$$E[X] = \int_a^b x f(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^2}{2}\Big|_a^b =$$

$$= \frac{b^2-a^2}{2(b-a)} = \frac{(b+a)\cancel{(b-a)}}{2\cancel{(b-a)}} = \frac{a+b}{2} \qquad \#1$$

Second moment:

$$E[X^2] = \int_a^b x^2 f(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^3}{3}\Big|_a^b =$$

$$= \frac{b^3-a^3}{3(b-a)} = \frac{\cancel{(b-a)}(a^2+b^2+ab)}{3\cancel{(b-a)}} = \frac{a^2+b^2+ab}{3} \qquad \#2$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \quad \#3 \qquad \bar{X}_{(2)} = \frac{1}{n}\sum_{i=1}^n X_i^2 \quad \#4$$

$\#1+\#3 \Rightarrow E[X]=\bar{X} \Rightarrow \frac{a+b}{2} = \bar{X} \Rightarrow a = 2\bar{X}-b \qquad \#5$

$\#2+\#4 \Rightarrow E[X^2] = \bar{X}_{(2)} \Rightarrow \frac{a^2+b^2+ab}{3} = \bar{X}_{(2)} \qquad \#6$

$$\xRightarrow[\#5+\#6]{} (2\bar{X}-b)^2 + b^2 + (2\bar{X}-b)b = 3\bar{X}_{(2)}$$

$$b^2 + 4\bar{X}^2 - 4b\bar{X} + \cancel{b^2} - \cancel{b^2} + 2b\bar{X} = 3\bar{X}_{(2)}$$

$$b^2 - 2b\bar{X} + 4\bar{X}^2 - 3\bar{X}_{(2)} = 0$$

$$\xRightarrow[\text{Quadratic formula}]{} \hat{b} = \frac{+\cancel{2}\bar{X} \pm \sqrt{(\cancel{2}\bar{X})^2 - 4(4\bar{X}^2 - 3\bar{X}_{(2)})}}{\cancel{2}}$$

$$\hat{b} = \bar{X} + \sqrt{3(\bar{X}_{(2)} - \bar{X}^2)}$$

$$\xRightarrow[\#5]{} \hat{a} = 2\bar{X} - \hat{b} = X - \sqrt{3\underbrace{(\bar{X}_{(2)} - \bar{X}^2)}_{\tilde{S}^2 \text{ (Variance of samples)}}}$$

$$\Rightarrow \hat{a} = \bar{X} - \sqrt{3}\cdot\tilde{S}$$

$$\hat{b} = \bar{X} + \sqrt{3}\,\tilde{S}$$

→ std of samples

2. The estimators are consistent but biased.

Consistent - We have explicit formulas including $\bar{X}, \tilde{S}$ estimators that should converge to the actual μ, σ due to LLN as the (number of samples) $\longrightarrow \infty$

Biased - Since we know that the variance $\tilde{S}^2 = (\bar{X}_{(2)} - \bar{X}^2)$ is a biased estimator $\to \tilde{S}$ is biased $\longrightarrow \hat{a}, \hat{b}$ are biased.

3. These estimators can produce unexpected results for low number of samples, for example:

Samples - $\{0,0,0,0,1\}$

we get: $\hat{a} = \frac{1}{5} - \frac{2\sqrt{3}}{5} \cong -0.493$

$\hat{b} = \frac{1}{5} + \frac{2\sqrt{3}}{5} \cong 0.893$

So $\hat{b} < 1$ but its clearly impossible since 1 is in our samples set (rough by-hand estimation for b can be better here easily)

4. $\bar{X} = \dfrac{0.8 + 0.97 + \ldots + 0.88}{5} = 0.73$

$$\sigma^2 = \frac{(0.8-0.73)^2 + \ldots + (0.88-0.73)^2}{5} = \frac{0.263}{5} = 0.0526$$

$$\to \sigma = \sqrt{0.0526} = \tilde{S}$$

$$\Rightarrow \hat{a} = 0.73 - \sqrt{3} \cdot \sqrt{0.0526} = 0.3327$$

$$\hat{b} = 0.73 + \sqrt{3} \cdot \sqrt{0.0526} = 1.1272 \sim$$

Comment: Similar to my previous answer, method of moments yielded an odd estimation for a ($\hat{a} > 0.33$ but 0.31 is in the samples set), due to the low number of samples