1. **Problem**
   A polling firm dials phone numbers in the London area at random. Among the 100 people interviewed, **32** were over 65 years old. This is not surprising because in the most recent census, **25%** of London's residents were over 65.

   Are the numbers in bold parameters or statistics?

   (a) 32 is a parameter, 25% is a statistic
   (b) 32 is a statistic, 25% is a parameter
   (c) Both are parameters
   (d) Both are statistics

   Solution
   A **parameter** is a number associated with the population. A **statistic** is a number associated with the sample. Therefore 32 is a statistic and 25% is a parameter.

2. **Problem**
   In a meat factory, a machine is packaging minced meat in boxes. We are interested in the average weight of meat in each box. What is the parameter space?

   (a) $[0, 1]$
   (b) $[0, \infty)$
   (c) $\{0, 1, 2, \dots\}$
   (d) $(-\infty, \infty)$

   Solution
   The **parameter space** is the set of all possible values of the parameter. In our case the parameter is the average weight of meat in the package which can be any number from 0 to infinity.

3. **Problem**
   An airline is monitoring its flight schedule for a period of one month. Suppose there are 100 flights in a month. What is the sample space?

   (a) $[0, 100]$
   (b) $[0, \infty)$
   (c) $\{0, 1, 2, \dots\}$
   (d) $\{0, 1, 2, \dots, 100\}$

   Solution
   The **sample space** is the set of all possible values we can observe. This cannot be less than 0 and no more than 100 and it must be whole numbers because we are counting so $\{0, 1, 2, \dots, 100\}$.

4. **Problem**
   Let $X_1, \dots, X_n$ be a random sample from the Exponential(μ) distribution. The parameter μ denotes the mean of the distribution. The asymptotic distribution of $\bar{X}$ according to the central limit theorem is

   (a) N(0,1)
   (b) N(μ,σ^2/n)
   (c) N(μ,μ^2/n)
   (d) N(μ,μ/n)

Solution
The mean of the exponential distribution is μ and the variance is μ^2. Therefore, according to the CLT, the asymptotic distribution of $\bar{X}$ is N(μ,μ^2/n).

5. **Problem**
Let $X_1, \ldots, X_n$ be a random sample from a distribution with moment generating function (mgf) $M(t)$. What is the correct formula for the mgf of the statistic $T = \sum X_i$?

   (a) $M(t)$
   (b) $M(t/n)$
   (c) $M(t)^n$
   (d) $M(t/n)^n$

   Solution
By the definition of mgf, $M(t) = E[\exp(tX)]$ so $E[\exp(tT)] = E[\exp(t(X_1 + \ldots + X_n))] = E[\exp(tX_1)\ldots\exp(tX_n)] = E[\exp(tX_1)]\ldots E[\exp(tX_n)] = M(t)\ldots M(t) = M(t)^n$.

6. **Problem**
The following is a random sample of size 4 from the Pareto distribution with parameter θ

$$3.28,\ 2.33,\ 1.14,\ 3.97$$

   The pdf is

$$f(x|\theta) = \frac{\theta}{x^{\theta+1}},\ \theta > 1,\ x > 1$$

   Use these data to estimate θ using the method of moments.
Solution
The mean of the distribution is

$$\begin{aligned} \mu &= \int_1^\infty x\frac{\theta}{x^{\theta+1}}dx \\ &= \int_1^\infty \frac{\theta}{x^\theta}dx \\ &= \int_1^\infty \theta \times x^{-\theta}dx \\ &= -\frac{\theta}{\theta-1} \times x^{-\theta+1}\Big]_1^\infty \\ &= \frac{\theta}{\theta-1}. \end{aligned}$$

   We set

$$\begin{aligned} \frac{\theta}{\theta-1} &= \bar{x} \\ \Rightarrow \theta &= \bar{x}(\theta-1) \\ \Rightarrow \bar{x} &= (\bar{x}-1)\theta \\ \Rightarrow \theta &= \frac{\bar{x}}{\bar{x}-1} \end{aligned}$$

   Substituting the data we have

$$\begin{aligned} \bar{x} &= (3.28 + 2.33 + 1.14 + 3.97)/4 = 2.68 \\ \hat{\theta} &= \frac{2.68}{2.68-1} = 1.595 \end{aligned}$$

7. **Problem**
The following is a random sample of size 5 from a $N(\mu, \sigma^2)$ population

3.2, 6.7, 3.2, 3.9, 3

The pdf is

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}.$$

Use these data to estimate the mean μ and the variance σ^2 using the maximum likelihood method.

(a) What is your estimate for the mean μ?

(b) What is your estimate for the variance σ^2?

Solution
The logarithm of the pdf is

$$\log f(x|\mu, \sigma^2) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2,$$

so the log-likelihood is

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

To estimate μ and σ^2, we differentiate with respect to both parameters and set the derivative equal to 0. This gives us a system of 2 equations to solve for μ and σ^2.

From the derivative with respect to μ we have

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

From the derivative with respect to σ^2 we have

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

From equation (1) we have

$$\begin{aligned}
&\sum_{i=1}^{n}(x_i-\mu) = 0\\
\Rightarrow\quad &\sum_{i=1}^{n}x_i - n\mu = 0\\
\Rightarrow\quad &\mu = \sum_{i=1}^{n}x_i/n.
\end{aligned}$$

So the MLE for μ is $\bar{x}$. This is unbiased as $\bar{x}$ is always unbiased for the mean of the population.

From equation (2), by substituting $\mu = \bar{x}$, we have

$$
\begin{aligned}
& -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 = 0 \\
\Rightarrow \quad & -n + \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 = 0 \\
\Rightarrow \quad & \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 = n \\
\Rightarrow \quad & \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2
\end{aligned}
$$

This MLE is not exactly S^2 because S^2 is divided by $(n-1)$, not by n. Because S^2 is unbiased, then the MLE must be biased.

Substituting the data to these formulae, we have

$$
\begin{aligned}
\hat{\mu} &= (3.2 + 6.7 + 3.2 + 3.9 + 3)/5 = 4 \\
\hat{\sigma}^2 &= ((-0.8)^2 + (2.7)^2 + (-0.8)^2 + (-0.1)^2 + (-1)^2)/5 = 1.916
\end{aligned}
$$

(a) The estimate for the mean is 4.

(b) The estimate for the variance is 1.916.

8. **Problem**

The following is a random sample of size 5 from the geometric distribution with parameter θ

3, 0, 1, 1, 2

The pmf is

$$f(x|\theta) = \theta \times (1-\theta)^x, \; \theta \in (0,1), \; x = 0, 1, 2, \ldots$$

Use these data to estimate θ using the maximum likelihood method.

Solution

The logarithm of the pdf is

$$\log f(x|\theta) = \log \theta + x \log(1-\theta),$$

so the log-likelihood is

$$\ell(\theta) = n \log \theta + \log(1-\theta) \sum_{i=1}^{n} x_i.$$

To estimate θ, we differentiate the log-likelihood with respect to θ and set the derivative equal to 0. This gives us an equation which we need to solve to find θ.

The derivative is

$$\begin{aligned}
& \frac{n}{\theta} - \frac{1}{1-\theta}\sum_{i=1}^{n} x_i = 0 \\
\Rightarrow\quad & \frac{n}{\theta} = \frac{1}{1-\theta}\sum_{i=1}^{n} x_i \\
\Rightarrow\quad & n(1-\theta) = \theta\sum_{i=1}^{n} x_i \\
\Rightarrow\quad & n = \theta(\sum_{i=1}^{n} x_i + n) \\
\Rightarrow\quad & \frac{n}{\sum_{i=1}^{n} x_i + n} = \theta \\
\Rightarrow\quad & \theta = \frac{1}{\bar{x}+1}
\end{aligned}$$

Substituting the data to these formulae, we have

$$\begin{aligned}
\bar{x} &= (3+0+1+1+2)/5 = 1.4 \\
\hat{\theta} &= \frac{1}{1.4+1} = 0.417
\end{aligned}$$

9. **Problem**
Let $X \sim N(6.3, 4.41)$. Compute the probability $P(X \leq 7)$ using the normal distribution table. (Give 4 decimal points in your answer.)

Solution
If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

Therefore $P(X < 7) = P(Z < \frac{7-6.3}{2.1}) = P(Z < 0.33)$.

We look for the number 0.33 in the margin of the normal distribution table and read the corresponding probability: 0.6293.

10. **Problem**
Let $X \sim N(9.7, 18.49)$. Compute the probability $P(X \leq 6.4)$ using the normal distribution table. (Give 4 decimal points in your answer.)

Solution
If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

Therefore $P(X < 6.4) = P(Z < \frac{6.4-9.7}{4.3}) = P(Z < -0.77)$.

Because the number -0.77 is negative, it can't be found in the margin of the normal distribution table.

Because of the symmetry of $N(0,1)$ around 0, we have the property $\Phi(-z) = 1 - \Phi(z)$ so instead we look for 0.77 and read the corresponding probability 0.7794. Then, the probability $P(X \leq 6.4) = 0.2206$.

11. **Problem**
Let $X \sim N(5.7, 3.24)$. Compute the probability $P(3.702 < X < 7.698)$ using the normal distribution table. (Give 4 decimal points in your answer.)

Solution
If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

Therefore $P(X < 7.698) = P(Z < \frac{7.698-5.7}{1.8}) = P(Z < 1.11)$.

We look for the number 1.11 in the margin of the normal distribution table and read the corresponding probability: 0.8665.

We do the same for the lower bound: $P(X < 3.702) = P(Z < \frac{3.702-5.7}{1.8}) = P(Z < -1.11)$.

We look for the number -1.11 in the margin of the normal distribution table and read the corresponding probability: 0.1335.

Therefore, the probability within is $0.8665 - 0.1335 = 0.733$.

12. **Problem**
Let $X \sim N(2.9, 0.81)$. Find x corresponding to probability $P(X \leq x) = 0.7389$ using the normal distribution table. (Give 2 decimal points in your answer.)

Solution
If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Therefore $P(X < x) = P(Z < \frac{x-2.9}{0.9}) = P(Z < z)$,
where $z = \frac{x-2.9}{0.9}$.

To have $P(Z < z) = 0.7389$ we must choose z appropriately.

We look for the number 0.7389 in the body of the normal distribution table and read the corresponding quantile from the margin: $z = 0.64$. Then, rearranging $\frac{x-2.9}{0.9} = 0.64$ we get $x = 2.9 + 0.64 \times 0.9 = 3.48$

13. **Problem**
Let $X \sim N(2.9, 0.81)$. Find x corresponding to probability $P(X \leq x) = 0.3409$ using the normal distribution table. (Give 2 decimal points in your answer.)

Solution
If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Therefore $P(X < x) = P(Z < \frac{x-2.9}{0.9}) = P(Z < z)$,
where $z = \frac{x-2.9}{0.9}$.

To have $P(Z < z) = 0.3409$ we must choose z appropriately.

Because $0.3409 < 0.5$ the number 0.3409 cannot be found in the normal distribution table.

Because of the symmetry of $N(0, 1)$ around 0, we have the property $\Phi(-z) = 1 - \Phi(z)$ so if $\Phi(z) = 0.3409 \Rightarrow 1 - \Phi(-z) = 1 - 0.3409 = 0.6591$.

We seek for 0.6591 in the body of the table and read the corresponding quantile from the margin: $-z = 0.41 \Rightarrow z = -0.41$. Then, rearranging $\frac{x-2.9}{0.9} = -0.41$ we get $x = 2.9 + -0.41 \times 0.9 = 2.53$

14. **Problem**
Let $X \sim$ Gamma(a, θ) with shape $a = 3$ and scale $\theta = 2$. Use Python to find the number x such that $P(X \leq x) = 0.43$. (Round your answer to 3 decimal places.)

Solution
Use the following code
from scipy.stats import gamma
gamma.ppf(0.43, a = 3, scale = 2)

15. **Problem**
The daily expenses of summer tourists in Bath are analysed. A survey with 60 tourists is conducted. This shows that the tourists spend on average 132.3 pounds. The variance σ^2 is equal to 136.8.

Determine a 95% confidence interval for the average daily expenses of a tourist.

(a) What is the lower confidence bound?

(b) What is the upper confidence bound?

Solution
For a 95% confidence interval we find that $\alpha = 0.05$ and $z_{1-\alpha/2} = 1.96$.

Then, the 95% confidence interval for the average expenses μ is given by:

$$\begin{aligned}
& \left[\bar{x} \;-\; 1.96\sqrt{\frac{\sigma^2}{n}},\; \bar{x} \;+\; 1.96\sqrt{\frac{\sigma^2}{n}}\right] \\
= & \left[132.3 \;-\; 1.96\sqrt{\frac{136.8}{60}},\; 132.3 \;+\; 1.96\sqrt{\frac{136.8}{60}}\right] \\
= & \; [129.34,\; 135.26]\,.
\end{aligned}$$

(a) The lower confidence bound is 129.34.

(b) The upper confidence bound is 135.26.

16. **Problem**
The weekly expenses for groceries for a family of four were monitored for 11 weeks. On average the family spent 132.7 pounds. The sample variance S^2 was equal to 191.3.

Determine a 95% confidence interval for the average weekly expenses of the family.

(a) What is the lower confidence bound?

(b) What is the upper confidence bound?

Solution
We are not told what the variance is so we have to use quantiles from the t distribution.

For a 95% confidence interval we find that $\alpha = 0.05$ and $t_{1-\alpha/2} = 2.228$ at $n-1 = 10$ degrees of freedom.

Then, the 95% confidence interval for the average expenses μ is given by:

$$\begin{aligned}
& \left[\bar{x} \;-t_{1-\alpha/2}\sqrt{\frac{S^2}{n}},\; \bar{x} \;+t_{1-\alpha/2}\sqrt{\frac{S^2}{n}}\right] \\
= & \left[132.7 \;-\; 2.228\sqrt{\frac{191.3}{11}},\; 132.7 \;+\; 2.228\sqrt{\frac{191.3}{11}}\right] \\
= & \; [123.409,\; 141.991]\,.
\end{aligned}$$

(a) The lower confidence bound is 123.409.

(b) The upper confidence bound is 141.991.