<u>Solution</u>

Using the hint, x_i below are central variables, i.e., $E(x_i) = 0$.

<u>$E(m_2)$</u>

$$\begin{aligned}
m_2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - m_1)^2 \\
&= \frac{1}{n}\left\{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right\} \\
&= \frac{1}{n}\left\{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 + \sum_{i\neq j} x_i x_j\right)\right\} \\
&= \frac{1}{n}\left\{\frac{n-1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i\neq j} x_i x_j\right\} \\
E(m_2) &= \frac{n-1}{n}\mu_2
\end{aligned}$$

<u>$Var(m_2)$</u>

$$\begin{aligned}
m_2^2 &= \frac{1}{n^2}\left\{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right\}^2 \\
&= \frac{1}{n^2}\left\{\left(\sum_{i=1}^{n} x_i^2\right)^2 - \frac{2}{n}\sum_{i=1}^{n} x_i^2\left(\sum_{i=1}^{n} x_i\right)^2 + \frac{1}{n^2}\left(\sum_{i=1}^{n} x_i\right)^4\right\}
\end{aligned}$$

$$\begin{aligned}
\left(\sum_{i=1}^{n} x_i^2\right)^2 &= \sum_{i=1}^{n} x_i^4 + \sum_{i\neq j} x_i^2 x_j^2 \\
E\left[\left(\sum_{i=1}^{n} x_i^2\right)^2\right] &= n\mu_4 + n(n-1)\mu_2^2
\end{aligned}$$

$$\begin{aligned}
\sum_{i=1}^{n} x_i^2\left(\sum_{i=1}^{n} x_i\right)^2 &= \sum_{i=1}^{n} x_i^4 + \sum_{i\neq j} x_i^2 x_j^2 + 2\sum_{i\neq j} x_i^3 x_j + \sum_{i\neq j\neq k} x_i^2 x_j x_k \\
E\left[\sum_{i=1}^{n} x_i^2\left(\sum_{i=1}^{n} x_i\right)^2\right] &= n\mu_4 + n(n-1)\mu_2^2
\end{aligned}$$

$$\left(\sum_{i=1}^{n} x_i\right)^4 = \sum_{i=1}^{n} x_i^4 + 3\sum_{i\neq j} x_i^2 x_j^2 + 4\sum_{i\neq j} x_i^3 x_j + 6\sum_{i\neq j\neq k} x_i^2 x_j x_k + \sum_{i\neq j\neq k\neq l} x_i x_j x_k x_l$$

$$
\begin{aligned}
E\left[\left(\sum_{i=1}^{n} x_i\right)^4\right] &= n\mu_4 + 3n(n-1)\mu_2^2 \\
E(m_2^2) &= \frac{(n-1)\left\{(n-1)\mu_4 + (n^2-2n+3)\mu_2^2\right)\right\}}{n^3}
\end{aligned}
$$

$$
\begin{aligned}
Var(m_2^2) &= E(m_2^4) - E(m_2^2) \\
&= \frac{(n-1)\left\{(n-1)\mu_4 - (n-3)\mu_2^2\right\}}{n^3}
\end{aligned}
$$

Assignment 2: Solution

Q3 Since $X_i \overset{\text{i.i.d.}}{\sim} \exp(\theta)$, we have

$$\begin{aligned} M_{X_i}(t) &= \frac{\theta}{\theta - t} \\ M_{\sum_{i=1}^n X_i} &= \Pi_{i=1}^n M_{X_i} = \Pi_{i=1}^n \frac{\theta}{\theta - t} = (\frac{\theta}{\theta - t})^n \end{aligned}$$

which implies, $\sum_{i=1}^n X_i \sim$ Gamma(n,θ).

16. X_i are i.i.d. with p.d.f. $f(x) = \frac{3}{2}x^2, \qquad -1 < x < 1.$

$$E(X_i) = \int_{-1}^{1} xf(x)dx = \int_{-1}^{1} \frac{3}{2}x^3 dx = \left[\frac{3}{8}x^4\right]_{-1}^{1} = 0$$

$$E(X_i^2) = \int_{-1}^{1} x^2 f(x)dx = \int_{-1}^{1} \frac{3}{2}x^4 dx = \left[\frac{3}{10}x^5\right]_{-1}^{1} = \frac{3}{5}$$

$$\therefore\ Var(X_i) = E(X_i^2) - E(X_i)^2 = \frac{3}{5} - 0^2 = 0.6$$

Since $Y = \sum_{i=1}^{15} X_i \simeq N(15 \times 0, 15 \times \frac{3}{5}) = N(0, 9)$ by C.L.T.

Hence:

$$\begin{aligned} P(-0.3 \le X \le 1.5) &\approx P\left(\frac{-0.3-0}{\sqrt{9}} \le Z \le \frac{1.5-0}{\sqrt{9}}\right) \\ &= P(-0.1 \le Z \le 0.5) \\ &= P(Z \le 0.5) - P(Z \le -0.1) \\ &= 0.6915 - 0.4602 = 0.2313 \end{aligned}$$

18. (a) X_i are *i.i.d* with p.d.f. $f(x) = (\frac{1}{4})^{x-1}(\frac{3}{4}), \qquad x = 1, 2, 3, \ldots$
By table, $\theta = \frac{3}{4}$, so

$$E(X_i) = \frac{1}{\theta} = \frac{3}{4}, Var(X_i) = \frac{1-\theta}{\theta^2} = \frac{1-3/4}{(3/4)^2} = \frac{4}{9}$$

So by C.L.T., $\sum_{i=1}^{36} X_i \simeq N(36 \times \frac{4}{3}, 36 \times \frac{4}{9}) \sim N(48, 16)$

$$\begin{aligned} P(46 \le \sum_{i=1}^{36} X_i \le 49) &\approx P\left(\frac{45.5-48}{\sqrt{16}} \le Z \le \frac{49.5-48}{\sqrt{16}}\right) \\ &= P(-0.625 \le Z \le 0.375) \\ &= 0.3802 \end{aligned}$$

(b)

$$\begin{aligned} P(1.25 \le \bar{X} \le 1.5) &= P(1.25 \times 36 \le \sum_{i=1}^{36} X_i \le 1.5 \times 36) \\ &= P(45 \le \sum_{i=1}^{36} X_i \le 54) \\ &= P\left(\frac{44.5-48}{\sqrt{16}} \le Z \le \frac{54.5-48}{\sqrt{16}}\right) \\ &= P(-0.875 \le Z \le 1.625) \\ &= 0.7571 \end{aligned}$$

Q3 in the midterm of 2013/2014

Let $U_1, ..., U_n$ be a random sample from the U(0,1).

1. (**2 marks**) Let $X = -\log(U)$. Find the distribution of X.

2. (**6 marks**) Let $Y = \dfrac{1}{\prod_{i=1}^{n} U_i^{\frac{1}{n}}}$, where $U_1, ..., U_n$ be a random sample from the U(0,1) and n is very large. Using Central Limit Theorem and Delta method to find the approximate distribution of Y.

Solution:

1. Let $X \sim F_X(x)$ where $F_X(x)$ is the CDF of X.

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(-\log(U) \leq x) \\ &= P(\log(U) \geq -x) = P(U \geq \exp(-x)) = 1 - P(U \leq \exp(-x)) \\ &= 1 - \exp(-x) \end{aligned}$$

Therefore, $X \sim Exp(1)$

2. By (a), $\log(Y) = -\frac{1}{n}\sum_{i=1}^{n} \log(U_i)$, where $-\log(U_i) \sim_{iid} Exp(1)$.
Since $E(-\log(U_i)) = 1$ and $Var(-\log(U_i)) = 1$, by central limit theorem,

$$\sqrt{n}\Big[-\frac{1}{n}\sum_{i=1}^{n}\log(U_i) - 1\Big] \to_d N(0,1)$$

Since $Y = \exp\left\{-\frac{1}{n}\sum_{i=1}^{n}\log(U_i)\right\}$, by delta method,

$$\sqrt{n}(Y - \exp(1)) \to_d N(0, \exp(2))$$

Therefore, for large enough n, $Y \to N(e, \frac{e^2}{n})$

1. **(8 marks)** Let X_1, X_2 be random variables having the bivariate normal distribution with parameters μ_1, μ_2, σ_1, σ_2, ρ (correlation coefficient between X_1 and X_2), i.e.,

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_2\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right).$$

Set

$$Y_1 = \frac{X_1 - \mu_1}{\sigma_1} + \frac{X_2 - \mu_2}{\sigma_2}, \; Y_2 = \frac{X_1 - \mu_1}{\sigma_1} - \frac{X_2 - \mu_2}{\sigma_2}.$$

Find the probability density functions of Y_1 and Y_2. Are they independent?

<u>Ans.</u>

$$\begin{aligned} Z_1 &= \frac{X_1 - \mu_1}{\sigma_1} \sim N(0,1) \\ Z_2 &= \frac{X_2 - \mu_1}{\sigma_2} \sim N(0,1) \\ Cov(Z_1, Z_2) &= \rho \end{aligned}$$

Prove that Y_1 and Y_2 are independent, use any method below:

(a) Since

$$\begin{aligned} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} &= \begin{pmatrix} Z_1 + Z_2 \\ Z_1 - Z_2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \\ \Rightarrow \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} &= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \end{aligned}$$

We consider the terms inside exponential, i.e.,

$$\begin{aligned} &\frac{1}{2} \begin{pmatrix} Z_1 & Z_2 \end{pmatrix}^T \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \\ = &\frac{1}{2(1-\rho^2)} \begin{pmatrix} Z_1 & Z_2 \end{pmatrix}^T \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \\ = &\frac{1}{8(1-\rho^2)} \begin{pmatrix} Y_1 & Y_2 \end{pmatrix}^T \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^T \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \\ = &\frac{1}{8(1-\rho^2)} \begin{pmatrix} Y_1 & Y_2 \end{pmatrix}^T \begin{pmatrix} 2(1-\rho) & 0 \\ 0 & 2(1+\rho) \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \end{aligned}$$

$\Rightarrow$ Y_1 & Y_2 are independent.

(b) Y_1, Y_2 are random variables having the bivariate normal distribution as linear combination of a multivariate random vector has a multivariate normal distribution.

$$\begin{aligned} Y_1 &\sim N(0, 2(1+\rho) \\ Y_2 &\sim N(0, 2(1-\rho) \end{aligned}$$

$Cov(Y_1, Y_2) = 0 \Rightarrow Y_1$ & Y_2 are independent.