Problem Set 8 Solution

Problem 1

$$E[X_i] = 75$$

$$\sqrt{Var(X_i)} = 15$$

a) recall $M_n(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$P[74 \le M_n(X) \le 76] = 1 - Pr[|M_n(X) - \overset{75}{E[X]}| \ge \overset{a}{1}]$$

Tchebyshev Inequality

$$\ge 1 - \underbrace{\frac{Var[X]}{n\, \underset{=1}{a^2}}}_{Var[M_n(X)]} = 1 - \frac{(15)^2}{n} = 0.99$$

$$\Downarrow$$

$$\boxed{n \ge 22500}$$

b)

$$\Pr\left(74 \le M_n(X) \le 76\right) = F_{M_n(X)}(76) - F_{M_n(X)}(74)$$

but $M_n(X) \sim N\left(\underset{75}{\overset{\mu}{\shortparallel}}, \underset{\frac{Var(X)}{n} = \frac{225}{n}}{\overset{\sigma^2}{\shortparallel}}\right)$

$$Var\left(M_n(X)\right) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2}\, n\, Var(X) = \frac{Var(X)}{n}$$

$$\Pr\left(74 \le M_n(X) \le 76\right) = \Phi\left(\frac{76-\mu}{\sigma}\right) - \Phi\left(\frac{74-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{76-75}{\frac{15}{\sqrt{n}}}\right) - \Phi\left(\frac{74-75}{\frac{15}{\sqrt{n}}}\right)$$

$$= \Phi\left(\frac{\sqrt{n}}{15}\right) - \Phi\left(\frac{-\sqrt{n}}{15}\right)$$

$$= 2\,\Phi\left(\frac{\sqrt{n}}{15}\right) - 1 = 0.99$$

$$\Downarrow$$

$$\underline{n = 1521}$$

Problem 2

Note that we can write Y_k as

$$Y_k = \left(\frac{X_{2k-1} - X_{2k}}{2}\right)^2 + \left(\frac{X_{2k} - X_{2k-1}}{2}\right)^2 = \frac{(X_{2k} - X_{2k-1})^2}{2} \tag{1}$$

Hence,

$$E\left[Y_k\right] = \frac{1}{2}E\left[X_{2k}^2 - 2X_{2k}X_{2k-1} + X_{2k-1}^2\right] = E\left[X^2\right] - (E\left[X\right])^2 = \mathrm{Var}[X] \tag{2}$$

Next we observe that $Y_1, Y_2, \ldots$ is an iid random sequence. If this independence is not obvious, consider that Y_1 is a function of X_1 and X_2, Y_2 is a function of X_3 and X_4, and so on. Since $X_1, X_2, \ldots$ is an idd sequence, $Y_1, Y_2, \ldots$ is an iid sequence. Hence, $E[M_n(Y)] = E[Y] = \mathrm{Var}[X]$, implying $M_n(Y)$ is an unbiased estimator of $\mathrm{Var}[X]$. We can use Theorem 7.5 to prove that $M_n(Y)$ is consistent if we show that $\mathrm{Var}[Y]$ is finite. Since $\mathrm{Var}[Y] \leq E[Y^2]$, it is sufficient to prove that $E[Y^2] < \infty$. Note that

$$Y_k^2 = \frac{X_{2k}^4 - 4X_{2k}^3 X_{2k-1} + 6X_{2k}^2 X_{2k-1}^2 - 4X_{2k} X_{2k-1}^3 + X_{2k-1}^4}{4} \tag{3}$$

Taking expectations yields

$$E\left[Y_k^2\right] = \frac{1}{2}E\left[X^4\right] - 2E\left[X^3\right]E\left[X\right] + \frac{3}{2}\left(E\left[X^2\right]\right)^2 \tag{4}$$

Hence, if the first four moments of X are finite, then $\mathrm{Var}[Y] \leq E[Y^2] < \infty$. By Theorem 7.5, the sequence $M_n(Y)$ is consistent.

Problem 3

$$P_X(x) = \begin{cases} 0.1 & x = 0 \\ 0.9 & x = 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

(a) $E[X]$ is in fact the same as $P_X(1)$ because X is a Bernoulli random variable.

(b) We can use the Chebyshev inequality to find

$$P[|M_{90}(X) - P_X(1)| \geq .05] = P[|M_{90}(X) - E[X]| \geq .05] \leq \alpha \tag{2}$$

In particular, the Chebyshev inequality states that

$$\alpha = \frac{\sigma_X^2}{90(.05)^2} = \frac{.09}{90(.05)^2} = 0.4 \tag{3}$$

(c) Now we wish to find the value of n such that $P[|M_n(X) - P_X(1)| \geq .03] \leq .01$. From the Chebyshev inequality, we write

$$0.1 = \frac{\sigma_X^2}{n(.03)^2}. \tag{4}$$

Since $\sigma_X^2 = 0.09$, solving for n yields $n = 100$.

Problem 4

Since the relative frequency of the error event E is $\hat{P}_n(E) = M_n(X_E)$ and $E[M_n(X_E)] = P[E]$, we can use Theorem 7.12(a) to write

$$P\left[\left|\hat{P}_n(A) - P[E]\right| \geq c\right] \leq \frac{\text{Var}[X_E]}{nc^2}. \tag{1}$$

Note that $\text{Var}[X_E] = P[E](1 - P[E])$ since X_E is a Bernoulli $(p = P[E])$ random variable. Using the additional fact that $P[E] \leq \epsilon$ and the fairly trivial fact that $1 - P[E] \leq 1$, we can conclude that

$$\text{Var}[X_E] = P[E](1 - P[E]) \leq P[E] \leq \epsilon. \tag{2}$$

Thus

$$P\left[\left|\hat{P}_n(A) - P[E]\right| \geq c\right] \leq \frac{\text{Var}[X_E]}{nc^2} \leq \frac{\epsilon}{nc^2}. \tag{3}$$

Problem 5

Both questions can be answered using the following equation from Example 7.6:

$$P\left[\left|\hat{P}_n(A) - P[A]\right| \geq c\right] \leq \frac{P[A](1 - P[A])}{nc^2} \tag{1}$$

The unusual part of this problem is that we are given the true value of $P[A]$. Since $P[A] = 0.01$, we can write

$$P\left[\left|\hat{P}_n(A) - P[A]\right| \geq c\right] \leq \frac{0.0099}{nc^2} \tag{2}$$

(a) In this part, we meet the requirement by choosing $c = 0.001$ yielding

$$P\left[\left|\hat{P}_n(A) - P[A]\right| \geq 0.001\right] \leq \frac{9900}{n} \tag{3}$$

Thus to have confidence level 0.01, we require that $9900/n \leq 0.01$. This requires $n \geq 990{,}000$.

(b) In this case, we meet the requirement by choosing $c = 10^{-3}P[A] = 10^{-5}$. This implies

$$P\left[\left|\hat{P}_n(A) - P[A]\right| \geq c\right] \leq \frac{P[A](1 - P[A])}{nc^2} = \frac{0.0099}{n10^{-10}} = \frac{9.9 \times 10^7}{n} \tag{4}$$

The confidence level 0.01 is met if $9.9 \times 10^7/n = 0.01$ or $n = 9.9 \times 10^9$.

Problem 6

(1). $E[M_{10}] = E\left[\frac{1}{10}\sum_{i=1}^{10} X_i\right] = \frac{1}{10} E\left[\sum_{i=1}^{10} X_i\right] = \frac{1}{10}\sum_{i=1}^{10} E[X_i]$

$= \frac{10}{10} = 1 \quad \Rightarrow \boxed{E[M_{10}] = 1}$

. $Var[M_{10}] = Var\left[\frac{1}{10}\sum_{i=1}^{10} X_i\right] = \left(\frac{1}{10}\right)^2 Var\left[\sum_{i=1}^{10} X_i\right]$

$\underset{\uparrow}{=} \frac{1}{100}\sum_{i=1}^{10} Var[X_i] = \frac{1}{100}\cdot 10 = \frac{1}{10}$

by independence $Cov[X_i, X_j] = 0$ for all $i \neq j$

so $\boxed{\sigma^2_{M_{10}} = \frac{1}{10}}$ and $\boxed{\sigma_{M_{10}} = \frac{1}{\sqrt{10}}}$

(2) Recall Chebyshev's Inequality

$P[|Y - \mu_Y| \geq c] \leq \frac{\sigma_Y^2}{c^2}$

For $Y = M_{10}$

$\left.\begin{array}{l}\mu_Y = E[M_{10}] = 1 \\ c = 3\end{array}\right\} \quad P[|M_{10} - 1| \geq 3] \leq \frac{\sigma^2_{M_{10}}}{9} \Leftrightarrow$

$P\left[(M_{10} - 1) \geq 3 \text{ or } -(M_{10} - 1) \geq 3\right] \leq \frac{1}{10 \cdot 9}$

$P\left[M_{10} \geq 4 \text{ or } M_{10} \leq -2\right] \leq \frac{1}{90}$

(3) CLT approximation says that M_{10} is approximately Gaussian with mean 1 and std $\sigma_{M_{10}} = \frac{1}{\sqrt{10}}$

Hence $P[|M_{10} - 1| \geq 3\sigma_{M_{10}}] = 1 - P[|M_{10} - 1| < 3\sigma_{M_{10}}]$

$$= 1 - P\left[-3\sigma_{M_{10}} < M_{10} - 1 < 3\sigma_{M_{10}}\right]$$

$$= 1 - P\left[-3\sigma_{M_{10}} + 1 < M_{10} < 3\sigma_{M_{10}} + 1\right]$$

$$= 1 - P\left[-\frac{3}{\sqrt{10}} + 1 < M_{10} < \frac{3}{\sqrt{10}} + 1\right]$$

$$= 1 - P\left[0.05 < M_{10} < 1.95\right]$$

$$= 1 - \left[\Phi\left[\frac{1.95 - 1}{\frac{1}{\sqrt{10}}}\right] - \Phi\left[\frac{0.05 - 1}{\frac{1}{\sqrt{10}}}\right]\right]$$

$$= \underbrace{1 - \Phi\left[\underbrace{0.95\sqrt{10}}_{3}\right]} + \Phi\left[\underbrace{-0.95\sqrt{10}}_{-3}\right]$$

$$= Q(3) + \underbrace{1 - \Phi(3)}_{Q(3)} = 2\,Q(3)$$

p143 of Textbook

$$= 2 \; 1.35 \; 10^{-3} = \boxed{2.7 \; 10^{-3}}$$

Problem 7

Recall that $X_1, X_2 \ldots X_n$ are independent exponential random variables with mean value $\mu_X = 5$ so that for $x \geq 0$, $F_X(x) = 1 - e^{-x/5}$.

(a) Using Theorem 7.1, $\sigma^2_{M_n(x)} = \sigma^2_X/n$. Realizing that $\sigma^2_X = 25$, we obtain

$$\mathrm{Var}[M_9(X)] = \frac{\sigma^2_X}{9} = \frac{25}{9} \tag{1}$$

(b)

$$P[X_1 \geq 7] = 1 - P[X_1 \leq 7] = 1 - F_X(7) = 1 - (1 - e^{-7/5}) = e^{-7/5} \approx 0.247 \tag{2}$$

(c) First we express $P[M_9(X) > 7]$ in terms of $X_1, \ldots, X_9$.

$$P[M_9(X) > 7] = 1 - P[M_9(X) \leq 7] = 1 - P[(X_1 + \ldots + X_9) \leq 63] \tag{3}$$

Now the probability that $M_9(X) > 7$ can be approximated using the Central Limit Theorem (CLT).

$$P[M_9(X) > 7] = 1 - P[(X_1 + \ldots + X_9) \leq 63] \approx 1 - \Phi(\frac{63 - 9\mu_X}{\sqrt{9}\sigma_X}) = 1 - \Phi(6/5) \tag{4}$$

Consulting with Table 3.1 yields $P[M_9(X) > 7] \approx 0.1151$.

Problem 8

Form the sample mean

$$M_{12}(x) = \frac{69.71 + 77.31 + \cdots + 71.36}{12} = 71.55\ \Omega$$

Since we have 12 samples we can use "Gaussian" Approximatio

$$P[M_{12}(x) - d < \mu_x \leq M_n(x) + d] = 1 - 2Q\left(\frac{d\sqrt{12}}{\sigma_x}\right) = 0.95$$

so $1 - 2Q\left(\frac{d\sqrt{12}}{\sqrt{13.7}}\right) = 0.95$

$$Q\left(d\sqrt{\frac{12}{13.7}}\right) = \frac{0.05}{2} = 0.025$$

$$1 - \Phi(d\ 0.9359) = 0.025$$

$$\Phi(d\ 0.9359) = 0.975$$

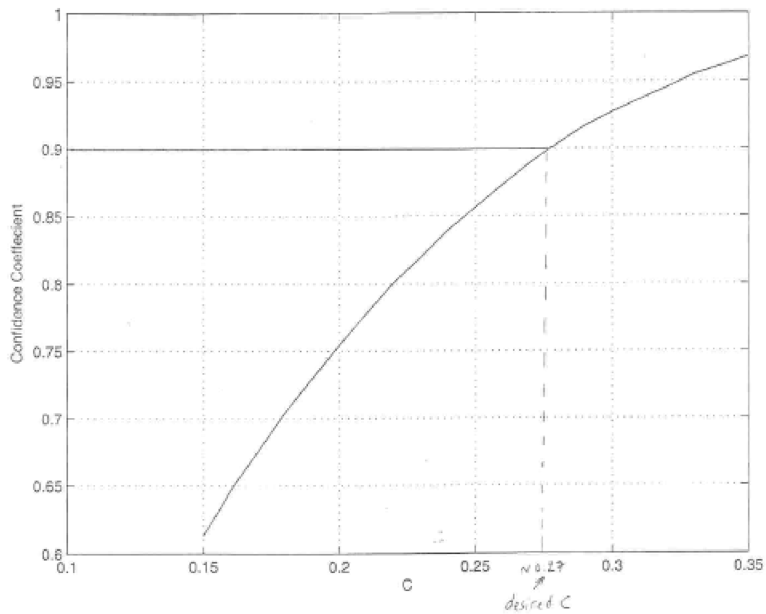
see p 142 of Textbook

$$d\ (0.9359) = 1.96 \Rightarrow d = 2.09$$

$$\boxed{P[69.46 \leq \mu_x \leq 73.64] = 95\%}$$

Problem 9

```
for i=1:10000
    sample_mean(i)=mean(rand(1,3));
end
C=0.15;
mean(0.5 > sample_mean - C & 0.5 < sample_mean + C)

% We get 0.6112

for i=1:10000
    sample_mean(i)=mean(rand(1,3));
end
C=0.35;
mean(0.5 > sample_mean - C & 0.5 < sample_mean + C)

% We get 0.9692

% So the desired C is between 0.15 and 0.35.
% Let us search for the desired C in that range

C=.15:0.01:0.35
for i=1:length(C)
    confidence(i)=mean(0.5 > sample_mean - C(i) & 0.5 < sample_mean + C(i));
end
plot(C,confidence)
grid;
xlabel('C');
ylabel('Confidence Coeffecient');

% So we see from the plot (below) that the desired C to get
% a confidence coeffecient of 90 % is  C ~ 0.27.
```



Problem 10

First we observe that the interval estimate can be expressed as

$$\left|\hat{P}_n(A) - P[A]\right| < 0.05. \tag{1}$$

Since $\hat{P}_n(A) = M_n(X_A)$ and $E[M_n(X_A)] = P[A]$, we can use Theorem 7.12(b) to write

$$P\left[\left|\hat{P}_n(A) - P[A]\right| < 0.05\right] \geq 1 - \frac{\text{Var}[X_A]}{n(0.05)^2}. \tag{2}$$

Note that $\text{Var}[X_A] = P[A](1 - P[A]) \leq 0.25$. Thus for confidence coefficient 0.9, we require that

$$1 - \frac{\text{Var}[X_A]}{n(0.05)^2} \geq 1 - \frac{0.25}{n(0.05)^2} \geq 0.9. \tag{3}$$

This implies $n \geq 1{,}000$ samples are needed.