Problem Set 8

Problem 1:

Let $X_1, X_2, \ldots$ denote an iid sequence of random variables, each with expected value 75 and standard deviation 15.

(a) How many samples n do we need to guarantee that the sample mean $M_n(X)$ is between 74 and 76 with probability 0.99?

(b) If each X_i has a Gaussian distribution, how many samples n' would we need to guarantee $M_{n'}(X)$ is between 74 and 76 with probability 0.99?

Problem 2:

Given the iid samples $X_1, X_2, \ldots$ of X, define the sequence $Y_1, Y_2, \ldots$ by

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

Note that each Y_k is an example of V_2', an estimate of the variance of X using two samples, given in Theorem 7.11. Show that if $E[X^k] < \infty$ for $k = 1, 2, 3, 4$, then the sample mean $M_n(Y)$ is a consistent, unbiased estimate of $\mathrm{Var}[X]$.

Problem 3:

$X_1, \ldots, X_n$ are n independent identically distributed samples of random variable X with PMF

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

(a) How is $E[X]$ related to $P_X(1)$?

(b) Use Chebyshev's inequality to find the confidence level α such that $M_{90}(X)$, the estimate based on 90 observations, is within 0.05 of $P_X(1)$. In other words, find α such that

$$P\left[\left|M_{90}(X) - P_X(1)\right| \geq 0.05\right] \leq \alpha.$$

(c) Use Chebyshev's inequality to find out how many samples n are necessary to have $M_n(X)$ within 0.03 of $P_X(1)$ with confidence level 0.1. In other words, find n such that

$$P\left[\left|M_n(X) - P_X(1)\right| \geq 0.03\right] \leq 0.1.$$

Problem 4:

In communication systems, the error probability $P[E]$ may be difficult to calculate; however it may be easy to derive an upper bound of the form $P[E] \leq \epsilon$. In this case, we may still want to estimate $P[E]$ using the relative frequency $\hat{P}_n(E)$ of E in n trials. In this case, show that

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

Problem 5:

When we perform an experiment, event A occurs with probability $P[A] = 0.01$. In this problem, we estimate $P[A]$ using $\hat{P}_n(A)$, the relative frequency of A over n independent trials.

(a) How many trials n are needed so that the interval estimate

$$\hat{P}_n(A) - 0.001 < P[A] < \hat{P}_n(A) + 0.001$$

has confidence coefficient $1 - \alpha = 0.99$?

(b) How many trials n are needed so that the probability $\hat{P}_n(A)$ differs from $P[A]$ by more than 0.1% is less than 0.01?

Problem 6:

Let $X_1, X_2, \cdots, X_10$ be independent identically distributed (i.i.d) continuous random variables with $E[X_i] = 1$ and $\text{Var}[X_i]=1$ for $i = 1, 2, \cdots, 10$.
(1) Find the expected value $E[M_{10}]$ and the variance $\sigma^2_{M_{10}}$ of the sample mean $M_{10} = \frac{1}{10}\sum_{i=1}^{10} X_i$.
(2) Provide an upper-bound on the probability that the random variable M_{10} exceeds 4 or is below -2.
(3) Use the central limit theorem approximation to obtain an "approximation" of the probability that the random variable M_{10} deviates from its mean by more than $3\sigma_{M_{10}}$.

Problem 7:

$X_1, X_2, \ldots, X_n$ is an iid sequence of exponential random variables each with expected value 5, i.e. $f_X(x) = \frac{1}{5}e^{-\frac{x}{5}} \cdot u(x)$.

a) What is $Var[M_9(X)]$, the variance of the sample mean based on nine trials?
b) What is $P[X_1 > 7]$, the probability that one outcome exceeds 7?
c) Estimate $P[M_9(X) > 7]$, the probability that the sample mean of nine trials exceeds 7? *Hint*: Use the central limit theorem.

Problem 8:

Let the random variable X represent measured resistance values in ohms. It is known that the variance of X is $\sigma_X^2 = 13.7\ \Omega^2$. Measurements of 12 sample resistors yields: 69.71, 77.31, 69.05, 68.66, 72.84, 75.31, 71.56, 73.33, 66.98, 71.72, 70.77, and 71.36 ohms. What are the confidence limits on the mean of X, μ_X if we require a confidence level of 95 %.

Problem 9:

Let us assume that we would like to design a 90 % confidence interval from sampling with sample size 3 from the uniform distribution between 0 and 1. In other words, we need to find C so that the interval [sample mean - C, sample mean + C] contains the true mean 90 % of the time.
(1) Use Matlab to test if $C = 0.15$ then $C = 0.35$ give the desired confidence interval.
(2) Deduce an approximate value for the correct value of C.

Problem 10:

In n independent experimental trials, the relative frequency of event A is $\hat{P}_n(A)$. How large should n be to ensure that the confidence interval estimate

$$\hat{P}_n(A) - 0.05 \leq P[A] \leq \hat{P}_n(A) + 0.05$$

has confidence coefficient 0.9?