In Class Problem Set 5

Problems:

1. In order to estimate f, the true fraction of smokers in a large population, Alvin selects n people at random. His estimator M_n is obtained by dividing S_n, the number of smokers in his sample, by n, i.e., $M_n = S_n/n$. Alvin chooses the sample size n to be the smallest possible number for which the Chebyshev inequality yields a guarantee that

$$\Pr(|M_n - f| \geq \epsilon) \leq \delta,$$

   where ϵ and δ are some prespecified tolerances. Determine how the value of n recommended by the Chebyshev inequality changes in the following cases.

   (a) The value of ϵ is reduced to half its original value.

   (b) The probability of δ is reduced to $\delta/2$.

2. Consider two sequences of random variables $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$, associated with the same experiment. Suppose that X_n converges to a and Y_n converges to b, in probability. Show that $X_n + Y_n$ converges to $a + b$, in probability.

3. A sequence X_n of random variables is said to converge to a number c in the mean square, if

$$\lim_{n\to\infty} \mathrm{E}\left[(X_n - c)^2\right] = 0.$$

   (a) Show that convergence in mean square implies convergence in probability

   (b) Give an example that shows that convergence in probability does not imply convergence in the mean square.

4. During each day, the probability that your computer's operating system crashes at least once is 5%, independent of every other day. You are interested in the probability of at least 45 crash-free days out of the next 50 days.

   (a) Find the probability of interest by using the normal approximation to the binomial.

   (b) Repeat part (a), this time using the Poisson approxmiation to the binomial.

5. The amount of time that a certain type of component functions before failing is a random variable with probability density function

$$f(x) = 2x \quad 0 < x < 1.$$

   Once the component fails, it is immediately replaced by another one of the same type. If we let X_i denote the lifetime of the ith component to be put in use, then $S_n = \sum_{i=1}^{n} X_i$ represents the time of the nth failure. The long-term rate at which failures occur, call it r, is defined by

$$r = \lim_{n\to\infty} \frac{n}{S_n}.$$

   Assuming that the random variables X_i, $i \geq 1$ are independent, determine r.

6. In order to estimate the probability of an event A, a sequence of Bernoulli trials is carried out and the relative frequency of A is observed. How large should n be in order to have a 0.95 probability that the relative frequency is within 0.05 of $p = \Pr(A)$?

   Hint: Let $X = \mathbf{1}_A$ be the indicator function of A. First, find the mean m and the variance σ^2 of $\mathbf{1}_A$ in terms of p. Note that since p is unknown then σ^2 is also unknown. However, you can easily find an upper bound for σ^2. Use the Chebyshev inequality to find a bound for

$$\Pr(|f_A(n) - \Pr(A)| < \epsilon)$$

   where $f_A(n)$ is the relative frequency of event A in a sequence of n indepdendent repetitions of a random experiment.