AIM 5002 Computational Statistics and Probability

(Spring 2021)

Assignment 5

Name:_________________________ Score:_______/6.5

Submit your assignment at CANVAS by uploading your file.
Due date: Saturday, 17th of the April, 2021 by 11:59 PM

---

1. A statistician wants to estimate the mean height h (in meters) of a population, based on n independent samples $X_1, \ldots, X_n$, chosen uniformly from the entire population. He uses the sample mean $M_n = (X_1 + \ldots + X_n)/n$ as the estimate of h, and a rough guess of 1.0 meters for the standard deviation of the samples X_i. (1 point)

   (a) How large should n be so that the standard deviation of M_n is at most 1 centimeter?
   $\text{var}(M_n) = \frac{var(X_1 + *** + X_n)}{n^2} = \frac{var(X_1) + *** + var(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} = \frac{1}{n}$
   $\text{var}(M_n) \leq 0.01^2 \rightarrow \frac{1}{n} \leq 0.01^2 \rightarrow n \geq 10{,}000$

   (b) How large should n be so that Chebyshev's inequality guarantees that the estimate is within 5 centimeters from h, with probability at least 0.99?
   $P(|M_n - h| \leq 0.05) \geq 0.99$
   where $h = E[M_n]$. Using $\text{var}(M_n) = 1/n$ and Chebyshev :
   $P(|M_n - h| \leq 0.05) = P(|M_n - E[M_n]| \leq 0.05) = 1 - P(|M_n - E[M_n]| \geq 0.05) \geq 1 - \frac{\left(\frac{1}{n}\right)}{0.05^2} \geq 0.99 \rightarrow n \geq 40{,}000$

2. A factory produces X_n gadgets on day n, where the X_n are independent and identically distributed random variables, with mean 5 and variance 9 (1.5 point).

   (a) Find an approximation to the probability that the total number of gadgets produced in 100 days is less than 440.
   Let $Y_n = X_1 + X_2 + \ldots + X_n$ total # of gadgets produced in n days
   $E[Y_n] = 5n$, $\text{var}(Y_n) = 9n$
   $P(Y_{100} < 440) = P(\frac{Y_{100} - 500}{30} < \frac{439.5 - 500}{30}) \approx \Phi(\frac{439.5 - 500}{30}) = \Phi(-2.02) = 1 - 0.9783 = 0.0217$

   (b) Find (approximately) the largest value of n such that
   $$P(X_1 + *** + X_n \geq 200 + 5n) \leq 0.05$$

$P(X_1 + *** + X_n \geq 200 + 5n) \leq 0.05 \rightarrow P(\frac{S_n - 5n}{3\sqrt{n}} \geq \frac{200}{3\sqrt{n}}) \leq 0.05$

$\rightarrow 1- \Phi\left(\frac{200}{3\sqrt{n}}\right) \leq 0.05 \rightarrow \Phi\left(\frac{200}{3\sqrt{n}}\right) \geq 0.95 \rightarrow \frac{200}{3\sqrt{n}} \geq 1.65 \rightarrow n \leq 1632$

(c) Let N be the first day on which the total number of gadgets produced exceeds 1000. Calculate an approximation to the probability that $N \geq 220$

$P(N \geq 220) = P(S_{219} \leq 1000)$

$= P(\frac{S_{219} - 5*219}{3\sqrt{219}} \leq \frac{1000-5*219}{3\sqrt{219}})$

$= 1- \Phi(2.14) = 1-0.9838 = 0.0162$

3. During rush hour. from 8 a.m. to 10 a.m., traffic accidents occur according to a Poisson process with a rate of 6 accidents per hour. Between 10 a.m. and 12 p.m., they occur as an independent Poisson process with a rate of 2 accidents per hour. What is the PMF of the total number of accidents between 8 a.m. and 12 p.m.? (0.5 point)

6*2 + 2*2 = 12 + 4 = 16

$e^{-16}\frac{16^k}{k!}$, k = 0, 1, …

4. A fisherman catches fish according to a Poisson process with rate λ=0.6 per hour. The fisherman will keep fishing for three hours. If he has caught at least one fish, he quits. Otherwise, he continues until he catches at least one fish. (2.5 point)

(a) Find the probability that he stays for more than three hours.

P(0, 3) = $e^{-0.6*3}$ = 0.165

(b) Find the probability that the total time he spends fishing is between three and five hours.

Zero arrival for 3 and at least one for 2 hours = P(0, 3)*{1 – P(0, 2)} = 0.165 * (1 - $e^{-0.6*2}$) = 0.165 * 0.6988 = 0.115303

(c) Find the probability that he catches at least two fish.

Must have fished for the first three hours → 1 – {P(0, 3) + P(1, 3)} = 1 – {$e^{-0.6*3}$ + 0.6*3$e^{-0.6*3}$} = 0.5372

(d) Find the expected number of fish that he catches.

N = # of fish caught after three hours.

E[# of fish caught] = E[# of fish caught during the first three hours] + E[N] = 3*0.6 + P(N=1) = 1.8 + P(0, 3) = 1.8 + 0.165 = 1.965

(e) Find the expected total fishing time, given that he has been fishing for five hours.

By the memorylessness of Poisson process, the future time is exponential with 1/0.6 → 5 + 1/0.6 = 6.667

5. A service station handles jobs of two types, A and B. (Multiple jobs can be processed simultaneously.) Arrivals of the two job types are independent Poisson processes with parameters $\lambda_A = 3$ and $\lambda_B = 4$ per minute, respectively. Type A jobs stay in the service station for exactly one minute. Each type B job stays in the service station for a random but integer amount of time which is geometrically distributed, with mean equal to 2, and independent of everything else. The service station started operating at some time in the remote past. (1 point)

   (a) What is the mean, variance, and PMF of the total number of jobs that arrive within a given three-minute interval?
   N = # of jobs that arrive in a given three-minute.
   E[N] = 3 $\lambda = 21$, $var(N) = 21$, and PMF:
   pN(n) = $\frac{21^n e^{-21}}{n!}$, n = 0, 1, 2, ***

   (b) We are told that during a 10-minute interval, exactly 10 new jobs arrived. What is the probability that exactly 3 of them are of type A?
   $\lambda_A / (\lambda_A + \lambda_B) = 3/7$
   $\binom{10}{3}\left(\frac{3}{7}\right)^3\left(\frac{4}{7}\right)^7 \approx 0.188$