EE 131A
Probability
Instructor: Lara Dolecek

Discussion Set 10 solutions
March 12, 2021
TAs: Lev Tauz, Debarnab Mitra

Chapter 7.1-7.3, 8 of *Probability, Statistics, and Random Processes* by A. Leon-Garcia

1. From past experience, it is known that the number of tickets purchased by a student standing in line at the ticket window for the football match of UCLA against USC follows a distribution that has mean $\mu = 2.4$ and standard deviation $\sigma = 2.0$. Suppose that few hours before the start of one of these matches there are 100 eager students standing in line to purchase tickets. If only 250 tickets remain, what is the probability that all 100 students will be able to purchase the tickets they desire?
Solution:
We are given that $\mu = 2.4, \sigma = 2, n = 100$. There are 250 tickets available, so the 100 students will be able to purchase the tickets they want if all together ask for less than 250 tickets. The probability for that is

$$P(T < 250) = P(z < \frac{250 - 100(2.4)}{\sqrt{100} \cdot 2}) = P(z < 0.5) = 1 - Q(0.5).$$

2. A student uses pens whose lifetime is an exponential random variable with mean 1 week. Use the central limit theorem to determine the minimum number of pens he should buy at the beginning of a 15-week semester, so that with probability .99 he does not run out of pens during the semester.
Solution:
Suppose that the student buys n pens and let X_i be the lifetime of the i^{th} pen. Let $S_n = \sum_{i=i}^{n} X_i$ be the total lifetime of the pens. We have,

$$E(S_n) = nE(X_i) = n$$

$$VAR(S_n) = nVAR(X_i) = n$$

. Assuming that S_n is approximately Gaussian we get

$$P(S_n > 15) = P(\frac{S_n - n}{\sqrt{n}} - \frac{15 - n}{\sqrt{n}}) = Q(\frac{15 - n}{\sqrt{n}})$$

We want smallest n such that $P(S_n > 15) = 0.99$. Thus

$$Q(\frac{15 - n}{\sqrt{n}}) = 0.99 \implies \frac{15 - n}{\sqrt{n}} = -2.3263 \implies n = 27.04.$$

Thus he should buy 28 pens.

3. *(CLT for a Poisson RV)* Suppose $X_1, X_2, \ldots X_n$ are n i.i.d RVs each having a Poisson distribution with parameter λ. Let $S_n = \sum_{i}^{n} X_i$. Note that the PMF of each X_i is given by

$$P(X_i = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0, 1, 2, \ldots$$

(a) Show that S_n is another Poisson random variable.
Hint. Use generating functions.
Solution:
We show that the sum S_n is a Poisson RV by checking the probability generating function (pgf) of S_n. The pgf of the Poisson random variable X_i is

$$\begin{aligned} G_{X_i}(z) &= E(z^{X_i}) \\ &= \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} z^k \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(z\lambda)^k}{k!} \\ &= e^{-\lambda} e^{z\lambda} \\ &= e^{\lambda(z-1)}. \end{aligned}$$

Hence,

$$\begin{aligned} G_{S_n}(z) &= E(z^{S_n}) \\ &= E(z^{\sum_i^n X_i}) \\ &= \prod_i^n E(z^{X_i}) \\ &= e^{n\lambda(z-1)} \end{aligned}$$

Thus, S_n is a Poisson RV with parameter $n\lambda$.

(b) Find the mean and variance of S_n.
Solution:
Since S_n is a Poisson RV with parameter $n\lambda$, $E(S_n) = n\lambda$ and $VAR(S_n) = n\lambda$.

(c) The number of messages arriving at a multiplexer is a Poisson RV with a rate of 15 messages per second. Use the central limit theorem to estimate the probability that more than 950 messages arrive in one minute.
Solution:
Let X_i be the number of message arriving in the i-th second. Then, X_i's are i.i.d Poisson RVs each with parameter 15. Let S be the number of messages arriving in one minute (60 seconds). Thus, $S = \sum_i^{60} X_i$. From part (a), S is a Poisson RV with parameter $60 \times 15 = 900$. Hence,

$$\begin{aligned} E(S) &= 900, \\ VAR(S) &= 900. \end{aligned}$$

Apply the central limit theorem on S, we obtain

$$\begin{aligned}
P(S > 950) &= P(\frac{S - E(S)}{\sqrt{VAR(S)}} > \frac{950 - E(S)}{\sqrt{VAR(S)}}) \\
&= P(\frac{S - E(S)}{\sqrt{VAR(S)}} > \frac{5}{3}) \\
&\approx Q(\frac{5}{3}) \\
&= 0.0478
\end{aligned}$$

4. *Chi-square test for testing the fit of a distribution to data*
The following histogram was obtained by counting the occurrence of the first digits in telephone numbers in one column of a telephone directory:

digit	0	1	2	3	4	5	6	7	8	9
observed	0	0	24	2	25	3	32	15	2	2

Test the goodness of fit of this data to a random variable that is uniformly distributed in the set $\{0, 1, \ldots, 9\}$ at a 1% significance level. Repeat for the case when the random variable is uniformly distributed in the set $\{2, 3, \ldots, 9\}$.
Solution: Here, the number of observations $n = 105$. For a random variable that is uniformly distributed in $\{0, 1, \ldots, 9\}$, $p_i = \frac{1}{10}, i = 0, 1, \ldots, 9$. Thus the expected number of observations $m_i = np_i = 105 \times \frac{1}{10} = 10.5$, $i = 0, 1, \ldots, 9$. Thus

$$D^2 = \sum_{i=0}^{9} \frac{(n_i - 10.5)^2}{10.5} = 130.33$$

Here $K = 10 \implies 9$ degrees of freedom. For significance level of 0.01, threshold $= 21.7$. Since $D^2 > 21.7$, we reject the hypothesis that the the data is uniformly distributed in $\{0, 1, \ldots, 9\}$.

For the case that the random variable is uniformly distributed in $\{2, 3, \ldots, 9\}$, we have $p_i = \frac{1}{8}, i = 2, 3, \ldots, 9$ and $p_i = 0$, $i = 0, 1$. Hence $m_i = \frac{105}{8} = 13.125, i = 2, 3, \ldots, 9$ and $m_i = 0$, $i = 0, 1$. Thus

$$D^2 = \sum_{i=2}^{10} \frac{(n_i - 13.125)^2}{13.125} = 83.26$$

Here again we have 9 degrees of freedom and hence the threshold $= 21.7$. Since $D^2 > 21.7$, we reject the hypothesis that the the data is uniformly distributed in $\{2, 3, \ldots, 9\}$.

5. It is known that 90% of the cabs in a city are yellow and 10% are green. A cab hits a pedestrian at night. One witness claims that the car involved in the accident was green. Based on previous record, this witness is 80% correct, meaning that when something happens he would make the right claim 80% of the time.

   (a) Describe the corresponding hypothesis testing problem.
   Solution:
   Let $H = 0$ denote the cab involved in the accident is yellow, and let $H = 1$ denote green.
   From the problem statement, we have

$$P[H = 0] = 90\%$$

$$P[H = 1] = 10\%$$

   Let Y denote the witness' claim: if $Y = 0$ then the witness claims yellow cab and if $Y = 1$ then the witness claims green cab. Then it is stated in the problem that

$$P[Y = 0|H = 0] = 80\%$$

$$P[Y = 1|H = 1] = 80\%$$

   The hypothesis testing problem can be described as follows: we need to make a decision based on the witness' claim that $Y = 1$ and our prior knowledge.

   (b) Based on this information, is it more likely that it was a green car or a yellow car?
   Solution:

   We first calculate $P[Y = 1]$,

$$\begin{aligned} P[Y = 1] &= P[H = 0]P[Y = 1|H = 0] + P[H = 1]P[Y = 1|H = 1] \\ &= 0.9 \times (1 - 0.8) + 0.1 \times 0.8 = 0.26 \end{aligned}$$

   Then, use Bayes rule, we have

$$P[H = 0|Y = 1] = \frac{P[H = 0]P[Y = 1|H = 0]}{P[Y = 1]} = \frac{0.9 \times 0.2}{0.26} \approx 0.69$$

$$P[H = 1|Y = 1] = \frac{P[H = 1]P[Y = 1|H = 1]}{P[Y = 1]} = \frac{0.1 \times 0.8}{0.26} \approx 0.31$$

   Therefore, it is more likely that a yellow car is involved in the accident.

6. The sum of a list of 48 real numbers is to be computed. Suppose that numbers are rounded off to the nearest integer so that each number has an error that is uniformly distributed in the interval (-0.5,0.5). Use the central limit theorem to estimate the probability that the absolute value of the total error in the sum of the 48 numbers exceeds 4.

Solution:
Let $X_1, X_2, \ldots X_{48}$ be the 48 rounding errors. Let $S = \sum_{i=1}^{48} X_i$ be the total error. Since X_i is uniformly distributed in the interval (-0.5,0.5), the PDF of X_i is

$$f_{X_i}(x) = \begin{cases} 1 & -0.5 < x < 0.5 \\ 0 & \text{otherwise.} \end{cases}$$

Hence,

$$E(X_i) = \int_{-0.5}^{0.5} x dx = 0.$$

$$\begin{aligned} VAR(X_i) &= E(X_i^2) - E(X_i)^2 \\ &= E(X_i^2) \\ &= \int_{-0.5}^{0.5} x^2 dx \\ &= \frac{1}{12}. \end{aligned}$$

Thus,

$$E(S) = E(\sum_{i=1}^{48} X_i) = 0,$$

$$\begin{aligned} VAR(S) &= \sum_{i=1}^{48} VAR(X_i) \\ &= 48 \times \frac{1}{12} \\ &= 4. \end{aligned}$$

Now we apply the central limit theorem to S:

$$\begin{aligned} P(|S| > 4) &= P(S > 4) + P(S < -4) \\ &= P(\frac{S - E(S)}{\sqrt{VAR(S)}} > \frac{4 - E(S)}{\sqrt{VAR(S)}}) + P(\frac{S - E(S)}{\sqrt{VAR(S)}} < \frac{-4 - E(S)}{\sqrt{VAR(S)}}) \\ &= P(\frac{S - E(S)}{\sqrt{VAR(S)}} > 2) + P(\frac{S - E(S)}{\sqrt{VAR(S)}} < -2) \\ &= Q(2) + 1 - Q(-2) \\ &= 0.0455. \end{aligned}$$