1. **[Geometric Random Variables]**
   During a bad economy, a graduating ECE student goes to career fair booths in the technology sector (e.g., Google, Apple, Qualcomm, Texas Instruments, Motorola, etc) - and his/her likelihood of receiving an off-campus interview invitation after a career fair booth visit depends on how well he/she did in ECE 313. Specifically, an A in 313 results in a probability $p = 0.95$ of obtaining an invitation, whereas a C in 313 results in a probability of $p = 0.15$ of an invitation.

   (a) Give the pmf for the random variable Y that denotes the number of career fair booth visits a student must make before his/her first invitation including the visit that results in the invitation. Express your answer in terms of p.
   Solution: $p_Y(k) = p(1-p)^{k-1}$ for $k \geq 1$.

   (b) On average, how many booth visits must an A student make before getting an off-campus interview invitation? How about a C student?
   Solution: $\mathrm{E}[Y] = \frac{1}{p} = \begin{cases} 1.0526 & \text{A student} \\ 6.6667 & \text{C student} \end{cases}.$

   (c) Assuming that each student visits 5 booths during a typical career fair, find the probability that an A student in 313 **will not** get an off-campus interview invitation. Similarly, find the probability that a C student in 313 **will** get an invitation during a typical career fair.
   Solution: P(an A student does not get an invitation in 5 trials)$=\sum_{k=6}^{\infty} p(1-p)^{k-1} = (1-p)^5 \sum_{k'=0}^{\infty} p(1-p)^{k'} = (1-p)^5 = (1-0.95)^5 = 3.125 \times 10^{-7}$.
   It can also be observed directly that P(an A student does not get an invitation in 5 trials) $= (1-p)^5$ by the independence of each trial.
   P(C gets an invitation in 5 trials)$=1-(1-p)^5 = 1-(1-0.15)^5 = 0.5563$.

2. **[Binomial and Poisson Distributions]**
   Suppose that 105 passengers hold reservations for a 100-passenger flight. The number of passengers who show up at the gate can be modeled as a binomial random variable X with parameters $(105, 0.9)$. (Hint: It is easy to find calculators for the binomial and Poisson distributions on line.)

   (a) On average, how many passengers show up at the gate?
   Solution: On average, $\mathsf{E}[X] = 105 \times 0.9 = 94.5$ passengers show up for the flight.

   (b) If $X \leq 100$, everyone who shows up gets to go. Find the value of $P\{X \leq 100\}$.
   Solution: $P\{X \leq 100\} = 1 - P\{X > 100\}$
   $= 1 - P\{X = 101\} - P\{X = 102\} - P\{X = 103\} - P\{X = 104\} - P\{X = 105\}$

$= 1 - \binom{105}{101}(0.9)^{101}(0.1)^4 - \binom{105}{102}(0.9)^{102}(0.1)^3 - \binom{105}{103}(0.9)^{103}(0.1)^2 - \binom{105}{104}(0.9)^{104}(0.1)^1 - \binom{105}{105}(0.9)^{105}(0.1)^0$

$= 1 - \binom{105}{4}(0.9)^{101}(0.1)^4 - \binom{105}{3}(0.9)^{102}(0.1)^3 - - \binom{105}{2}(0.9)^{103}(0.1)^2 - \binom{105}{1}(0.9)^{104}(0.1)^1 - \binom{105}{0}(0.9)^{105}(0.1)^0$

$= 0.9832\ldots$

(c) Explain why the number of *no-shows* can be modelled as a binomial random variable Y with parameters $(105, 0.1)$.

Solution: If X is a binomial random variable with parameters (n, p), then $Y = n - X$ is a binomial random variable with parameters $(n, 1-p)$. ALTERNATIVELY, Y is the number of successes in 105 independent trials, where a "success" means a no show, and the probability of success for each trial is 0.1.

(d) Notice that the probability that everyone who shows up gets to go can also be expressed as $P\{Y \geq 5\}$. Use the *Poisson approximation* to compute $P\{Y \geq 5\}$ and compare your answer to the exact answer that you found in part (b).

Solution: $P\{Y \geq 5\} = 1 - P\{Y = 0\} - P\{Y = 1\} - P\{Y = 2\} - P\{Y = 3\} - P\{Y = 4\}$ $= 1 - \exp(-10.5)\left[1 + \frac{10.5}{1!} + \frac{(10.5)^2}{2!} + \frac{(10.5)^3}{3!} + \frac{(10.5)^4}{4!}\right] = 0.9789$ which is close to 0.9832.

3. **[Web Site Hits]**
Suppose the number of hits a web site receives in any time interval is a Poisson random variable. A particular site gets on average 5 hits per second.

(a) What is the probability that there will be no hits in an interval of two seconds?

Solution: $X \sim \text{Poisson}(2 \times 5)$

$P(X = k) = e^{-10}\frac{10^k}{k!}$

$P(X = 0) = \frac{e^{-10}10^0}{0!} = e^{-10}$

(b) What is the probability that there is at least one hit in an interval of one second?

Solution: $\lambda = 5$, $X \sim \text{Poisson}(5)$, $P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{e^{-5}5^0}{0!} = 1 - e^{-5}$

4. **[Buffering in Video Streaming]**
Suppose that you are watching a video clip consisting of thirty frames per second from YouTube. Each frame consists of 1000 packets. When a frame is about to be played out, if five or more packets of the frame are lost (i.e., if you have received 995 or fewer packets of the frame), you will experience buffering during that frame. Each packet is lost with the probability θ, independently of other packets.

(a) Suppose that θ is unknown. If you have observed that 20 packets of a frame are lost, what is the maximum likelihood estimate $\hat{\theta}_{ML}(20)$ of probability θ?

Solution: From the solution to Example 2.8.1 in the lecture notes, the maximum likelihood estimate is $\hat{\theta}_{ML}(20) = \frac{20}{1000} = \frac{1}{50}$.

(b) Suppose $\theta = 0.001$. Using the Poisson approximation, find the probability that you experience buffering during a given frame.

Solution: Let X be the number of lost packets of the frame, which is a binomial random variable with parameters 1000 and 0.001. Let Y be a Poisson random variable with parameter $\lambda = 1000 \times 10^{-3} = 1$. By the Poisson approximation, the probability that five or more packets of the frame are lost can be approximated by

$$P(X \geq 5) \approx P(Y \geq 5) = 1 - P(Y < 5) = 1 - \sum_{k=0}^{4} \frac{e^{-\lambda}\lambda^k}{k!}$$

$$= 1 - e^{-1}(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!}) = 0.00366.$$

5. **[A random variable taking on values 0, 1, 2, 3, 4, 5]**
Let X denote a random variable with values in $\{0, 1, \ldots, 5\}$ with $E[X] = 2.8$.

(a) Using the Markov inequality, find an upper bound on $P\{X = 5\}$. (You'll need to make the bound as tight as possible in order to do part (b).)

Solution: Markov inequality:

$$P(X = 5) \;=\; P(X \geq 5) \;\leq\; \frac{E[X]}{5} \;=\; \frac{2.8}{5}$$

(b) Give a possible pmf for X so that $P\{X = 5\}$ is equal to the upper bound found in part (a).

Solution: $p_X(i) = \frac{2.8}{5} = 0.56$, if $i = 5$; $1 - \frac{2.8}{5} = 0.44$ if $i = 0$; 0 else.

6. **[Confidence Interval]**
Suppose a system operates with an error probability p. We want to estimate the error probability by $\widehat{p} = X/n$ where n is the number of times the system will be tested and X is the number of tests found to give an error.

(a) Let $n = 100$ and $p = 0.1$, and suppose α is such that $P\{|\widehat{p} - 0.1| \leq 0.5\} \geq \alpha$. Find α.

Solution: The number of errors, X, in n tests has the binomial distribution with parameters $n = 100$ and $p = 0.1$.

Using Chebychev inequality for confidence intervals involving Bernoulli r.v.s:

$p = 0.1$, $n = 100$

$$a\sqrt{\frac{p(1-p)}{n}} = 0.05 \implies a\sqrt{\frac{0.1 \times 0.9}{100}} = 0.05 \implies a = \frac{5}{3}$$

$$\alpha = 1 - \frac{1}{a^2} = 1 - \frac{9}{25} = 0.64$$

$\implies$ confidence level = 64%

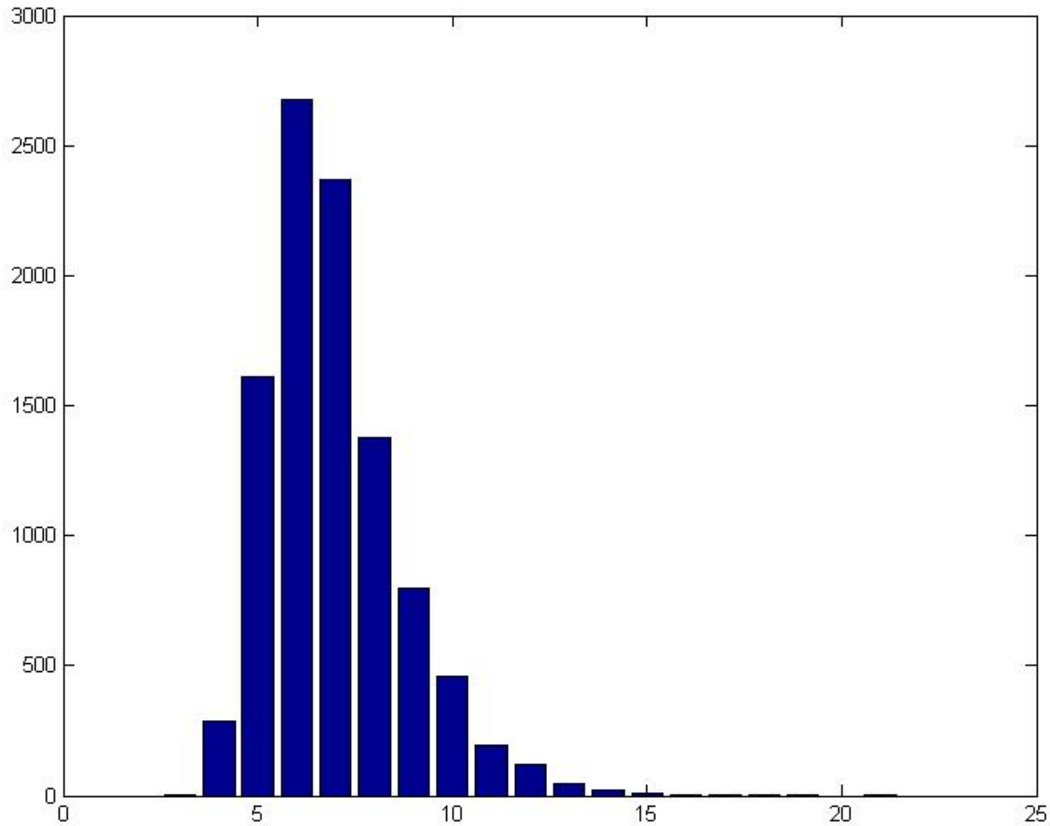
(b) Again assuming $p = 0.1$, find the number of tests required to have $\widehat{p}$ within 50% of its correct value with probability at least 90%. In other words, find n such that $P\{|\widehat{p} - 0.1| \leq 0.5\} \geq 0.9$.

Solution: $P[|\hat{p} - 0.1| < 0.05] \geq 0.9$
$p = 0.1$,
$1 - \frac{1}{a^2} = 0.9 \implies a^2 = 10$

$$a\sqrt{\frac{p(1-p)}{n}} = 0.05 \implies n = 360$$

7. **[Runs in Coin Flip Data]**
One experiment is to simulate 100 flips of a fair coin, and calculate the length of the longest run. (A run is a set of consecutive flips with the same outcomes. For example, the longest run for the sequence $HTTHTTTTHT$ has length four. Before starting the programming part of this exercise, try writing down a random looking string of H's and T's of length 100, and then determine the length of the longest run.) Simulate this experiment (using 100 flips per experiment) 10,000 times, and then make a histogram of the 10,000 values of the longest runs. What you should turn in: a copy of the histogram and the computer code you used to generate the sequences and to compute the lengths of longest runs.

Solution:


3000
2500
2000
1500
1000
500
0
0
5
10
15
20
25