Ma 3 - Problem Set 4

Marco Yang

2237027

1. (20 points) (a) Let $p_0, p_1, \ldots, p_n$ denote the probability mass function of the binomial distribution with parameters n and p. Let $q = 1-p$. Show that the binomial probabilities can be computed recursively by $p_0 = q^n$ and $p_k = (kq)^{-1} \cdot (n-k+1) \cdot p \cdot p_{k-1}$, for $k = 1, 2, \ldots, n$. Use this relation to find $P(X \leq 4)$ for $n = 9000$ and $p = 0.0005$.

Solution: The formula for the PMF of a binomial distribution with parameters n and p is

$$p_i = \binom{n}{i} \cdot p^i \cdot q^{n-i}.$$

Then, plugging in $i = 0$ gives us

$$p_0 = \binom{n}{0} \cdot p^0 \cdot q^{n-0} = q^n.$$

Plugging in $i = k-1$,

$$p_{k-1} = \binom{n}{k-1} \cdot p^{k-1} \cdot q^{n-(k-1)}.$$

Plugging in $i = k$, we derive

$$\begin{aligned}
p_k &= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\
&= \frac{n!}{k!(n-k)!} \cdot p \cdot p^{k-1} \cdot q^{-1} \cdot q^{n-(k-1)} \\
&= \frac{n!}{(k-1)!(n-(k-1)!)} \cdot \frac{n-(k-1)}{k} \cdot p \cdot p^{k-1} \cdot q^{-1} \cdot q^{n-(k-1)} \\
&= (kq)^{-1} \cdot (n-k+1) \cdot p \cdot \binom{n}{k-1} \cdot p^{k-1} \cdot q^{n-(k-1)} \\
&= (kq)^{-1} \cdot (n-k+1) \cdot p \cdot p_{k-1}.
\end{aligned}$$

For $P(X \leq 4), \lambda = 4.5$,

$$\begin{aligned}
p_0 &= 0.9995^{9000} \\
p_1 &= \frac{1}{1 \cdot 0.9995}(9000 - 1 + 1) \cdot 0.0005 \cdot p_0 \\
p_2 &= \frac{1}{2 \cdot 0.9995}(9000 - 2 + 1) \cdot 0.0005 \cdot p_1 \\
p_3 &= \frac{1}{3 \cdot 0.9995}(9000 - 3 + 1) \cdot 0.0005 \cdot p_2 \\
p_4 &= \frac{1}{4 \cdot 0.9995}(9000 - 4 + 1) \cdot 0.0005 \cdot p_3 \\
P(X \leq 4) &= \sum_{i=0}^{4} p_k \\
&= 0.5321.
\end{aligned}$$

(b) Show that the Poisson probablities $p_0, p_1, .., $ (i.e. the PMF) can be computed recursively by $p_0 = \exp(-\lambda)$ and $(\lambda/k) \cdot p_{k-1}$. Use this scheme to find $P(X \leq 4)$ for $\lambda = 4.5$ How does part b relate to a?

Solution: The formula for the PMF of the Poisson distribution is

$$p_i = \frac{e^{-\lambda}\lambda^i}{i!}.$$

Plugging in $i = 0$,

$$p_0 = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda}.$$

Plugging in $i = k - 1$,

$$p_{k-1} = \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!}.$$

Starting with the formula for p_k, we derive

$$\begin{aligned} p_k &= \frac{e^{-\lambda}\lambda^k}{k!} \\ &= \frac{e^{-\lambda} \cdot \lambda^{k-1}}{(k-1)!} \cdot \frac{\lambda}{k} \\ &= \frac{\lambda}{k} \cdot p_{k-1}. \end{aligned}$$

For $P(X \leq 4), \lambda = 4.5$,

$$\begin{aligned} p_0 &= \exp(-4.5) \\ p_1 &= -\frac{4.5}{1} \cdot p_0 \\ p_2 &= -\frac{4.5}{2} \cdot p_1 \\ p_3 &= -\frac{4.5}{3} \cdot p_2 \\ p_4 &= -\frac{4.5}{4} \cdot p_3 \\ P(X \leq 4) &= \sum_{i=0}^{4} p_k \\ &= 0.5321. \end{aligned}$$

The recursive formula of the Poisson distribution shows that it can be used to estimate the binomial distribution for larges values of n and small values of p since in the binomial distribution formula, $p^n \approx e^{-np} = e^{-\lambda}$ if n is large and p is small (can verify using Taylor expansion of e^{-np} and $p^n = e^{n \ln p}$ and $\ln p \to 1$ as $p \to 0$), $(kq)^{-1} \to \frac{1}{k}$ as $p \to 0 \iff q \to 1$, and the $(n - k + 1) \cdot p \to np = \lambda$ if $n \gg k$.

2. (15 points) The Cauchy cumulative distribution function is

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \cdot \tan^{-1}(x),\ -\infty < x < \infty.$$

(a) (5 points) Show that this is a cdf.

Solution: A valid cdf must be increasing, right-continuous, and converge to 0 and 1. The derivative of F w.r.t. x is

$$\frac{d}{dx}F(x) = \frac{d}{dx}\frac{1}{2} + \frac{1}{\pi} \cdot \tan^{-1}(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}.$$

Since $1 + x^2 > 0$, $\frac{1}{\pi} \cdot \frac{1}{1+x^2} > 0$ for all x, and F is increasing.

Since arctan is continuous and $\frac{1}{2}$ is continuous and F is continuous as well.

$$\begin{aligned}
\lim_{x\to-\infty} F(x) &= \lim_{x\to-\infty} \frac{1}{2} + \frac{1}{\pi} \cdot \tan^{-1}(x) \\
&= \frac{1}{2} + \frac{1}{\pi} \cdot -\frac{\pi}{2} \\
&= \frac{1}{2} - \frac{1}{2} \\
&= 0 \\
\lim_{x\to\infty} F(x) &= \lim_{x\to\infty} \frac{1}{2} + \frac{1}{\pi} \cdot \tan^{-1}(x) \\
&= \frac{1}{2} + \frac{1}{\pi} \cdot \frac{\pi}{2} \\
&= \frac{1}{2} + \frac{1}{2} \\
&= 1.
\end{aligned}$$

Thus, $F(x)$ is a cdf.

(b) (5 points) Find the density function.

Solution: Since the CDF is the accumulation (integral) of the density function, we can find the density by taking the derivative of F:

$$\begin{aligned}
p(x) &= \frac{d}{dx}F(x) \\
&= \frac{d}{dx}\frac{1}{2} + \frac{1}{\pi} \cdot \tan^{-1}(x) \\
&= \frac{1}{\pi} \cdot \frac{1}{1+x^2} \\
&= \frac{1}{\pi(1+x^2)}
\end{aligned}$$

.

(c) (5 points) Find x s.t. $P(X > x) = 0.1$.

Solution: We know that $P(X \leq x) = 1 - P(X > x)$. Thus, $P(X > x) = 0.1 \implies P(X \leq x) = 0.9$. Solving for this,

$$P(X \leq x) = 0.9 \implies \frac{1}{2} + \frac{1}{\pi} \cdot \tan^{-1}(x) = 0.9 \implies \tan^{-1}(x) = 0.4\pi \implies x = \tan(0.4\pi) = 3.078.$$

3. (15 points) Geometric R.V.

(a) (5 points) Find an expression for the cumulative distribution function of a geometric random variable.

Solution: Using the formula for the sum of a finite geometric series,

$$\begin{aligned} F(k) &= \sum_{i=0}^{k} q^k p \\ &= p \cdot \frac{1 - q^{k+1}}{1 - q} \\ &= 1 - q^{k+1}. \end{aligned}$$

The proof for the sum of the finite geometric series with ratio r is below:

$$\begin{aligned} S_k &= \sum_{i=0}^{k} q^k \\ S_k &= 1 + q + q^2 + \ldots + q^k \\ qS_k &= q + q^2 + q^3 + \ldots + q^{k+1} \\ S_k - qS_k &= 1 - q^{k+1} \\ S_k &= \frac{1 - q^{k+1}}{1 - q}. \end{aligned}$$

Intuitively, the formula for the CDF makes sense because it is equivalent to the complement of the probability that the non-favorable case happens at least k times in a row.

(b) (5 points) If X is a geometric random variable with $p = 0.5$, for what value of k is $P(X \leq k) = 0.99$?

Solution:

$$P(x \leq k) = 1 - q^{k+1} = 1 - 0.5^{k+1} = 0.99 \implies k = 5.64.$$

(c) (5 points) If X is a geometric random variable, show that

$$P(X > n + k - 1 | X > n - 1) = P(X > k).$$

The result above is an example of a "memoryless" stochastic phenomenom, i.e. if we know that X is larger than some value $n - 1$ then asking whether it is larger than $(n - 1) + k$ is the same as asking whether it is larger than k.

Solution:

Proof. Let p, q be the associate probabilities for the geometric r.v. Using the complement of the CDF, we have

$$P(X > k) = 1 - P(X \leq k) = q^{k+1}.$$

Applying Bayes' Theorem,

$$\begin{aligned} P(X > n+k-1 | X > n-1) &= \frac{P(X > n-1 | X > n+k-1) P(X > n+k-1)}{P(X > n-1)} \\ &= \frac{1 \cdot q^{n+k-1+1}}{q^{n-1+1}} \\ &= q^{k+1} \\ &= P(X > k). \end{aligned}$$

□

4. (15 points) Suppose that the liftime of an electronic component follows an exponential distribution with $\lambda = 0.1$.

   (a) (5 points) Find the probability that the lifetime is less than 10.

   Solution: The CDF of a exponential distribution is

   $$P(X \leq k) = 1 - e^{-\lambda k}.$$

   Thus,

   $$P(X \leq 10) = 1 - e^{-0.1 \cdot 10} = 0.6321.$$

   (b) (5 points) Find the probability that the lifetime is between 5 and 15.

   Solution: The probability that the lifetime is between 5 and 15 is the same as the probability that it's less than 15 but not less than 5.

   $$P(5 \leq X \leq 15) = P(X \leq 15) - P(X \leq 5) = 0.3834.$$

   (c) (5 points) Find t s.t. the probability that the lifetime is greater than t is 0.01.

   Solution:
   $$P(X > t) = 1 - (1 - e^{-0.1 \cdot t}) = 0.01 \implies t = 46.0517.$$

5. (10 points) Alice works from home and receives Slack messages from her boss at intervals of a Poisson process with $\lambda = 2$ per hour.

   (a) (5 points) If Alice takes a 10 minute walk around the block with her dog, what is the probability that she gets a message during that time?

Solution: Our adjusted lambda would be the rate of Slack messages per 10 minutes, which is

$$\lambda' = \frac{2}{\text{hr}} \cdot \frac{\text{hr}}{60 \text{ min}} \cdot 10 = \frac{1}{3}.$$

Thus, the probability of receiving at least one message in 10 minutes is

$$P(X > 0) = 1 - P(X = 0) = 1 - \frac{e^{-\lambda} \cdot \lambda^0}{0!} = 1 - e^{-\frac{1}{3}} = 0.2835.$$

(b) (5 points) How long can her walk be if she wishes the probability of receiving no message to be at most 0.5?

Solution: The adjusted lambda for t duration of time is $\lambda' = \lambda \cdot t = 2t/\text{hr}$. Thus, the probability of no messages within t time is

$$P(X = 0) = \frac{e^{-\lambda'} \cdot \lambda'^0}{0!} = e^{-2t}.$$

Our contraint is $P(X = 0) \leq 0.5$, so

$$e^{-2t} < 0.5 \implies t < \frac{\ln 2}{2} = 0.3656 \text{ hours}.$$

6. (15 points) If $X \sim N(0, \sigma^2)$, find the density of $Y = |X|$.

Solution: The PDF for $X \sim N(0, \sigma^2)$ is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Since we want the density of $Y = |X|$, the support is $x \geq 0$, with the probability density of all $x > 0$ being doubled since the probability of the negative counterpart is added. Thus, the PDF for $Y = |X|$ is

$$f'(x) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right), & x > 0 \\ \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right), & x = 0 \end{cases}.$$