Probability & Statistics for EECS: Homework #05

Due on Nov 12, 2023 at 23:59

Name: **Runkang Yang**
Student ID: 2022533080

Problem 1

1. (a)

   Since we can take the derivative of the CDF to obtain the PDF, we can integrate the PDF to obtain the CDF. Thus we have

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \int_{-\infty}^{x} f(x)dx \\ &= \int_{-\infty}^{x} \frac{1}{\pi(1+x^2)}dx \\ &= \frac{1}{\pi}\int_{-\infty}^{x} \frac{1}{1+x^2}dx \\ &= \frac{1}{\pi} arctanx|_{-\infty}^{x} \\ &= \frac{1}{\pi}(arctanx - (-\frac{\pi}{2})) \\ &= \frac{arctanx}{\pi} + \frac{1}{2} \end{aligned}$$

   (b)

$$\begin{aligned} F(x) &= \int_{1}^{x} f(x)dx \\ &= \int_{1}^{x} \frac{a}{x^{a+1}}dx \\ &= a\int_{1}^{x} x^{-(a+1)}dx \\ &= a \cdot (-\frac{1}{a})x^{-a}|_{1}^{x} \\ &= (-1) \cdot (\frac{1}{x^a} - 1) \\ &= 1 - \frac{1}{x^a} \end{aligned}$$

   So we can find the CDF of a Pareto r.v. with parameter a

$$F(x) = \begin{cases} 0 & x < 1 \\ 1 - \frac{1}{x^a} & x \geq 1 \end{cases}$$

   To check it is a valid CDF, first, we can find that
   $f(x) = F'(x) > 0$, which means that $F(x)$ is always increasing;
   if $x \to -\infty$, then $F(x) = 0$;
   if $x \to +\infty$, then $F(x) = 1$;
   if $x = 1$, then $F(1) = 0$, which means that $F(x)$ is left and right continuous at $x = 1$.
   Thus it is a valid CDF.

Problem 2

1. (a)

   The first success occurs after G failures have occured. Since each failure take Δt time, so we can get

   $$T = G\Delta t$$

   (b)

   First find $P(T > t)$

   $$\begin{aligned} P(T > t) &= P(G\Delta t > t) \\ &= P(G > \frac{t}{\Delta t}) \end{aligned}$$

   If $G > \frac{t}{\Delta t}$, it means that in the first $\lfloor \frac{t}{\Delta t} \rfloor$ times trials, all are failed, since the probability of each trail failed is $1 - \lambda\Delta t$, so the probability of all trails fail is $(1 - \lambda\Delta t)^{\lfloor \frac{t}{\Delta t} \rfloor}$, which is the value of $P(G > \frac{t}{\Delta t})$. Thus,

   $$P(T \leq t) = 1 - P(T > t) = 1 - (1 - \lambda\Delta t)^{\lfloor \frac{t}{\Delta t} \rfloor}$$

   (c)

   As $\Delta t \to 0$

   $$\begin{aligned} \lim_{\Delta t \to 0} P(T \leq t) &= \lim_{\Delta t \to 0} 1 - (1 - \lambda\Delta t)^{\lfloor \frac{t}{\Delta t} \rfloor} \\ &= 1 - \lim_{\Delta t \to 0} (1 - \lambda\Delta t)^{\lfloor \frac{t}{\Delta t} \rfloor} \\ &= 1 - \lim_{\Delta t \to 0} [(1 - \lambda\Delta t)^{\lfloor \frac{1}{\lambda\Delta t} \rfloor}]^{\lambda t} \\ &= 1 - e^{-\lambda t} \end{aligned}$$

   So as $\Delta t \to 0$, the CDF of T converges to the Expo(λ) CDF.

Problem 3

1. (a)

$$
\begin{aligned}
E(e^{-3x}) &= \sum_{k=0}^{+\infty} e^{-3k} \frac{(\lambda)^k}{k!} e^{-\lambda} \\
&= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(e^{-3}\lambda)^k}{k!} \\
&= e^{-\lambda} e^{e^{-3}\lambda} \\
&\neq e^{-3\lambda}
\end{aligned}
$$

So it is not unbiased for estimating θ.

(b)

$$
\begin{aligned}
E((-2)^X) &= \sum_{k=0}^{+\infty} (-2)^k \frac{(\lambda)^k}{k!} e^{-\lambda} \\
&= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(-2\lambda)^k}{k!} \\
&= e^{-\lambda} e^{-2\lambda} \\
&= e^{-3\lambda}
\end{aligned}
$$

So $g(X) = (-2)^X$ is an unbiased estimator for θ.

(c)

We can find that $g(X)$ can be negative, but θ is always non-negative, we can improve it by choosing the estimator $h(X) = max\{g(X), 0\}$ to get an non-negative value, and then we can get $|h(X) - \theta| \leq |g(X) - \theta|$, it can guarantee that $h(X)$ is always at least as good as $g(X)$and sometimes strictly better than $g(X)$.

Problem 4

1. (a)

Suppose that the sample of untagged elk X is of size k, it means that in the first m+k-1 samples, we capture k untagged elk, m-1 tagged elk and capture the mth tagged elk in the kth capture among the rest N-(m+k-1) elk. So the probability that the number of untagged elk in the new sample is k is

$$P(X = k) = \frac{\binom{n}{m-1}\binom{N-n}{k}}{\binom{N}{m+k-1}} \cdot \frac{n-m+1}{N-m-k+1}$$

About the total number of elk in the new sample Y, since we can find that $Y = X + m$, then we can get the PMF of Y by this

$$P(Y = y) = P(X + m = y)P(X = y - m) = \frac{\binom{n}{m-1}\binom{N-n}{y-m}}{\binom{N}{y-1}} \cdot \frac{n-m+1}{N-y+1}$$

(b)

Since we just focus on the number of untagged elk in the new sample X, then the continuing capture has nothing to do with X, so we can assume that even after getting m tagged elk, they continue to be captured until all N of them have been obtained to simplify our calculation.
From the hint, let's use I_j to indicate we capture the jth untagged elk, to find $P(I_j = 1)$, it means that the jth elk is captured before the first tagged elk is captured, which has nothing to do with other $m-1$ untagged elk, so the probability that we capture the jth elk among the 1 plus other n tagged elk is $\frac{1}{(n+1)}$. which means that $E(I_j) = P(I_j = 1) = \frac{1}{(n+1)}$.
On the other hand, we can find that $X_1 = I_1 + I_2 + ... + I_{N-n}$, so

$$E(X_1) = \sum_{j=0}^{n} I_j = (N-n)E(I_j) = \frac{N-n}{n+1}$$

Thus, we can get

$$E(X) = mE(X_1) = \frac{m(N-n)}{n+1}$$

Then

$$E(Y) = E(X + m) = m + \frac{m(N-n)}{n+1}$$

(c)

If the sampling is done with a fixed sample size equal to $E(Y)$ rather than sampling until exactly m tagged elk are obtained, then the problem is converted into the traditional ecological method capture-recapture, suppose that the expected number of tagged elk in the sample is M, then according to the formula of traditional capture-recapture, we have

$$\frac{M}{E(Y)} = \frac{n}{N}$$

then $M = \frac{mn(N+1)}{N(n+1)} = \frac{m(1+\frac{1}{N}}{1+\frac{1}{n}} < m$

Problem 5

1. (a)

During the exploration phase. we use R_i to denote the rank of the ith dish, so the expected sum of the ranks of the dishes in the exploration phase is $R_1 + R_2 + R_3 + ... + R_k$, and we can find that $E(R_i) = \frac{1+2+3+...+n}{n} = \frac{n+1}{2}$.

After the exploration phase. I will definitely choose the best dish (whose rank is X) in the rest $m-k$ exploitation phase, so the expected sum of the ranks of the dishes in the exploitation phase is $(m-k)X$.

So we can get the expected sum of the ranks of the dishes both in the exploration and exploitation phase is

$$E(R) = kE(R_i) + (m-k)E(X) = \frac{k(n+1)}{2} + (m-k)E(X)$$

(b)

Since we have already conducted k explorations, so the rank of the best dish is definitely greater than k, which means that the support of X is from k to n.

Suppose that the rank of the best dish that we find in the exploration phase is j, we can find the satisfied situation is to choose $k-1$ ranks from the $j-1$ ranks while the number of the total events is to choose k ranks from total n ranks, thus we can get

$$P(X = j) = \frac{\binom{j-1}{k-1}}{\binom{n}{k}}$$

(c)

$$E(X) = \sum_{j=k}^{n} jP(X = j) = \sum_{j=k}^{n} \frac{j\binom{j-1}{k-1}}{\binom{n}{k}} = \frac{k}{\binom{n}{k}} \sum_{j=k}^{n} \frac{j}{k} \cdot \binom{j-1}{k-1} = \frac{k}{\binom{n}{k}} \cdot \binom{n+1}{k+1} = \frac{k(n+1)}{k+1}$$

(d)

From (a) and (c), we can get that

$$E(R) = \frac{k(n+1)}{2} + (m-k)E(X) = \frac{k(n+1)}{2} + \frac{k(n+1)(m-k)}{k+1} = (n+1)(\frac{k}{2} + \frac{k(m-k)}{k+1})$$

Since k is a variable while n, m are constant, we can define

$$f(k) = \frac{k}{2} + \frac{k(m-k)}{k+1}$$

To find the value of k that maximizes f(k), we need to take the first derivative of f(k) and identify the point where the first derivative equals 0, with the derivative on the left of that point being greater than 0 and the derivative on the right being less than 0. By calculation, let

$f'(k) = \frac{m+1}{(k+1)^2} - \frac{1}{2} = 0 \to k = \sqrt{2(m+1)} - 1$(easy to verify f(k) obtains its maximum value when k is that value.)

To draw a conclusion, we get the optimal value of k is $\sqrt{2(m+1)} - 1$.

Problem 6

1.