Probability & Statistics for EECS: Homework #04

Due on Mar 31, 2024 at 23:59

Name: **Fei Pang**
Student ID: 2022533153

Problem 1

$$E(X) = \sum_{k=1}^{\infty} kP(X=k) = c\sum_{k=1}^{\infty} p^k = c\left(\sum_{k=0}^{\infty} p^k - p^0\right) = c\left(\frac{p}{(1-p)^2}\right)$$

We use formula $\mathrm{Var}(X) = E(X^2) - (E(X))^2$. To calculate $E(X^2)$:

$$E(X(X-1)) = \sum_{k=1}^{\infty} k(k-1)P(X=k) = \sum_{k=1}^{\infty} k(k-1)c\frac{p^k}{k} = c\sum_{k=1}^{\infty}(k-1)p^k$$

$$= cp^2\sum_{k=0}^{\infty} kp^{k-1} = cp^2\frac{d}{dp}\left(\sum_{k=0}^{\infty} p^k\right) = cp^2\frac{d}{dp}\left(\frac{1}{1-p}\right) = \frac{cp^2}{(1-p)^2}$$

To obtain $E(X^2)$:

$$E(X^2) = E(X(X-1)) + E(X) = \frac{cp^2}{(1-p)^2} + \frac{cp}{1-p} = \frac{cp^2 + cp(1-p)}{(1-p)^2}$$

Compute the variance:

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = \frac{cp^2 + cp(1-p)}{(1-p)^2} - \left(\frac{cp}{1-p}\right)^2$$

Problem 2

Let $N = w + b$, $p = w/N$, $q = 1 - p$.

(a) So $X \sim HGeom(w, b, n)$.

$$P(X=k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$$

$$E\binom{X}{2} = \binom{n}{2}\frac{w}{w+b}\frac{w-1}{w+b-1}$$

(b) By (a),

$$EX^2 - EX = E(X(X-1)) = n(n-1)p\frac{w-1}{N-1},$$

so

$$\begin{aligned}
Var(X) &= E(X^2) - (EX)^2 \\
&= n(n-1)p\frac{w-1}{N-1} + np - n^2p^2 \\
&= np\left(\frac{(n-1)(w-1)}{N-1} + 1 - np\right) \\
&= np\left(\frac{nw - w - n + N}{N-1} - \frac{nw}{N}\right) \\
&= np\left(\frac{Nnw - Nw - Nn + N^2 - Nnw + nw}{N(N-1)}\right) \\
&= np\left(\frac{(N-n)(N-w)}{N(N-1)}\right) \\
&= \frac{N-n}{N-1}npq.
\end{aligned}$$

Problem 3

N: # of toys needed to obtain all types of toys.
$N = N_1 + N_2 + \cdots + N_n$
$N_1 = 1$
$N_2 \sim Fs(\frac{n-1}{n}), N_3 \sim Fs(\frac{n-2}{n}) \cdots$ So $E(N_j) = \frac{n}{n-(j-1)}$.
$E(N) = E(N_1) + E(N_2) + \cdots + E(N_n) = n\sum_{j=1}^{n}\frac{1}{j}$.
since $Y \sim Fs(p)$:

$$E(Y) = \frac{1}{p}, Var(Y) = \frac{1-p}{p^2}$$

.

$$Var(N_j) = \frac{1-\frac{n-(j-1)}{n}}{[\frac{n-(j-1)}{n}]^2} = \frac{n(j-1)}{(n-(j-1))^2}$$

.

$$Var(N) = Var(N_1) + Var(N_2) + \cdots + Var(N_n) = n \cdot \sum_{j=1}^{n}\frac{j-1}{(n-(j-1))^2}$$

Problem 4

I_j : Type j toy occurs in m collected toys. $E(I_j) = 1-(1-p_j)^m$.
$N = I_1 + I_2 + \cdots + I_n$

$$E(N) = \sum_{i=1}^{n} E(1-(1-p_i)^m) = n - \sum_{i=1}^{n}(1-p_i)^m$$

$$E(N^2) = E\left((I_1+\cdots+I_n)^2\right) = \sum_{j=1}^{n}E(I_j) + 2\sum_{i<j}E(I_iI_j) = n - \sum_{i=1}^{n}(1-p_i)^m + 2\sum_{i<j}(1-(1-p_i)^m)(1-(1-p_j)^m)$$

So

$$\mathrm{Var}(N) = E(N^2) - E^2(N) = n - \sum_{i=1}^{n}(1-p_i)^m + 2\sum_{i<j}(1-(1-p_i)^m)(1-(1-p_j)^m) - \left(n - \sum_{i=1}^{n}(1-p_i)^m\right)^2$$

Problem 5

(a)

$$P(X \geq 24) = 1 - P(X \leq 23) = 0.493 < \frac{1}{2}$$

Thus, $m \geq 24$ cannot be the median. Similarly:

$$P(X \leq 22) = 1 - P(X \geq 23) < \frac{1}{2}$$

So $m = 23$ is the unique median.

Problem 5 continued on next page...

(b) Suppose that $X = k(1 \leq k \leq 366)$, it means that no birthday match among $k-1$ people. $Therefore I_1, I_2, \cdots, I_k = 1,\ I_{k+1}, I_{k+2}, \cdots, I_{366} = 0.\ Therefore I_1 + I_2 + \cdots + I_{366} = k\ Therefore X = I_1 + I_2 + \cdots + I_{366}$

$$E(X) = \sum_{j=1}^{366} E(I_j) = \sum_{j=1}^{366} P(I_j = 1) = \sum_{j=1}^{366} p_j$$

(c) For the set of keys [1, 4, 5, 10, 16, 17, 21], draw binary search tree of height 3, 4, 5, and 6. Use python to calculate:

```
EX = 0
for j in range(1, 367):
    pj = 1
    if j < 3:
        pj = 1
    if j > 2:
        for i in range(3, j + 1):
            pj = pj * (1 - (i - 2) / 365)
        EX = EX + pj
print(EX)
```

$$E(X) \approx 24.617$$

(d)

$$I_i^2 = I_i,\ I_i I_j = 1$$

$$X^2 = I_1 + \ldots + I_{366} + 2\sum_{j=3}^{366}\sum_{i=1}^{j-2} I_i I_j$$

$$= I_1 + \ldots + I_{366} + 2\sum_{j=3}^{366}(j-2)I_j$$

$$= \sum_{j=1}^{366}(2j-1)I_j$$

$$E(X^2) = \sum_{j=1}^{366}(2j-1)E(I_j)$$

$$= \sum_{j=1}^{366}(2j-1)P(I_j)$$

$$P(I_j) = E(I_j) = P(I_j = 1) = p_j$$

$$E(X^2) = \sum_{j=1}^{366}(2j-1)p_j$$

again, we use python:

```
        EX2 = 0
        for j in range(1, 367):
            pj = 1
            if j < 3:
                pj = 1
            if j > 2:
                for i in range(3, j+1):
                    pj = pj * (1 - (i - 2)/365)
            EX2 = EX2 + (2*j - 1) * pj
```

So

$$E(X^2) \approx 754.617$$

$$Var(X) = E(X^2) - (EX)^2 \approx 148.640$$