Week 3

Friday, 29 October 2021 16:31

1. There are 14 singers at The Voice 2021 getting ready for a public show and each needs to choose an unordered playlist of 3 different songs from 50 total songs. Each singer chooses a playlist of 3 songs; any subset of 3 songs out of 50 is equally likely. The singers choose songs independently of each other, i.e., it is possible to reuse the same song across different singers, but a singer cannot use a song more than once.
a) The music director needs to buy the rights to the songs the singers choose for any show. How many songs should director expect to buy the rights to for the show?
b) The rights to each song cost $100, and there is an overall processing fee of $40 to complete all the transactions. What is the expected cost to the music director?

a) Let X be the number of songs that the director needs to buy.
For $i = 1, 2, \dots, 50$, $X_i = 1$ is the song i is chosen by at least one singer and $X_i = 0$ otherwise.
Thus $E[X_i] = P(X_i = 1) = 1 - P(X_i = 0)$

$$P(X_i = 0) = P\left(\begin{matrix}\text{14 singers did not}\\ \text{choose the song } i\end{matrix}\right) = P\left(\begin{matrix}\text{1 singer did not}\\ \text{choose the song } i\end{matrix}\right)^{14}$$

$$= \left[\frac{\binom{49}{3}}{\binom{50}{3}}\right]^{14}$$

By the Linearity of Expectation: $E[X] = E[X_1] + \dots + E[X_{50}]$

$$= 50\, E[X_1]$$

$$= 50 \cdot \left[1 - \left(\frac{\binom{49}{3}}{\binom{50}{3}}\right)^{14}\right]$$

b) Total cost $= 100 \cdot X + 40 \overset{\text{LoE}}{\Longrightarrow} E[100X + 40] = 100\, E[X] + 40$

$$= 100 \cdot 50 \cdot \left[1 - \left(\frac{\binom{49}{3}}{\binom{50}{3}}\right)^{14}\right] + 40$$

2. Suppose we have two coins. Coin $C1$ comes up heads with probability 0.3 and coin $C2$ comes up heads with probability 0.9. We repeat this process 3 times:
- Choose a coin with equal probability.
- Flip that coin once.

Suppose X is the number of heads after 3 flips.
a) What is $E[X]$?
b) What is $Var(X)$?
c) Based on the number of heads we get, we earn $Y = 1/(X+1)$, dollars. What is $E[Y]$?

a) By the Law of Total Probability:

$$P(H) = P(H|C_1)\,P(C_1) + P(H|C_2)\,P(C_2)$$
$$= (.3)(.5) + (.9)(.5) = 0.6$$

Let $X_i = \begin{cases} 1 & \text{if the i-th flip was head} \\ 0 & \text{if the i-th flip was tail} \end{cases}$, $i = 1, 2, \dots, n$

$\Omega_X = \{0, 1, 2, 3\}$ is the # heads we can possibly obtain

Then $p_X(x) = \begin{cases} (0.4)^3 & , k = 0 \\ \binom{3}{1}(0.6)(0.4)^2 & , k = 1 \\ \binom{3}{2}(0.6)^2(0.4) & , k = 2 \end{cases}$

$1 \quad (0.6)^3 \quad , k = 3$

$\Rightarrow E[X] = 0 \cdot (0.4)^3 + 1\binom{3}{1}(0.6)(0.4)^2 + \dots + 3 \cdot (0.6)^3$
$= \boxed{1.8}$

b) $E[X^2] = 0^2 \cdot (0.4)^3 + \dots + 3^2 \cdot (0.6)^3 = 3.96$
$\Rightarrow Var[X] = E[X^2] - E[X]^2 = 3.96 - 1.8^2 = \boxed{0.72}$

c, $E\left[\frac{1}{X+1}\right] = \sum_x \left(\frac{1}{x+1}\right) p_X(x)$

$= \frac{1}{0+1}(0.4)^2 + \dots + \frac{1}{3+1} \cdot (0.6)^3$

$= \boxed{0.406}$

3) To determine whether a community of 1000 people containing Covid19 cases, we have their blood tested. However, rather than testing each individual separately (1000 tests is quite costly), it is decided to use a pooled testing strategy:

- **Phase 1**: First, place 1000 people into groups of 5. The blood samples of the 5 people in each group will be pooled and analysed together. If the test is positive (at least one person in the pool has Covid19 virus), continue to Phase 2. Otherwise, we can send the group home. Totally, 200 of these pooled tests are performed.
- Phase 2: Individually test each of the 5 people in the group. 5 of these individual tests are performed per group in Phase 2. Suppose that the probability that a person has Covid19 is 5% for all people, independently of others, and that the test has a 100% true positive rate and 0% false positive rate (note that this is unrealistic).

Using this strategy, compute the expected total number of blood tests (individual and pooled) that we will have to do across Phases 1 and 2.

- Let X_i be the number of tests needed for each group in the Phase 1
  $\Rightarrow 1 \le i \le 200$

$$X_i = \begin{cases} 1 & \text{if all 5 people in Group } i \text{ are neg} \\ 0 & \text{if anyone in group } i \text{ is pos.} \end{cases}$$

Thus $p_X(x) = \begin{cases} 0.95^5 & \text{if } i=1 \\ 1-(0.95)^5 & \text{if } i=0 \end{cases}$

$\Rightarrow E[X] = 1 \cdot (0.95)^2 + 0 \approx 2.13$ is the expected # tests for each group.

$\xrightarrow{\text{LoE}} E[Y = 200X] = 200E[X] = \boxed{428}$

4) There are 100 shoelaces in a box. At each stage, you pick two random ends and tie them together. Either this results in a longer shoelace (if the two ends came from different pieces), or it results in a loop (if the two ends came from the same piece). What are the expected number of steps until everything is in loops, and the expected number of loops after everything is in loops?

- Initially, there are 100 shoelaces = 200 ends. After each trial, the # ends decrease by 2, whether a longer shoelace or a loop is created $\Rightarrow$ the number of steps taken is always $= \boxed{100}$
- As any end is chosen equally likely, suppose at a time where there are n unlooped shoelaces $= 2 \cdot n$ ends, the prob of forming a new loop is: $\frac{n}{\binom{2n}{2}} = \frac{1}{2n-1}$

Thus the expect # loops is: $\boxed{\sum_{n=1}^{100} \frac{1}{2n-1}}$

5 Suppose we have a hash function h: $\mathcal{U} \to \{0,1, \ldots , m-1\}$ which maps from a universe $\mathcal{U}$ of strings (with length < 100) into m buckets, with each string independently and equally likely to be hashed into any bucket. We want to insert n strings $s1, \ldots , sn$ into our hash table.

a) Let $X1$ = h($s1$) be the index of the bucket that string $s1$ hashes into. What distribution does $X1$ have?
b) What is the probability that two particular strings $s1$ and sn hash to the same bucket?
c) If $Y1$ is the number of strings in the first bucket after inserting all n strings, what distribution does ~~Y have? What is the~~ probability that the first bucket is empty?
d) What is the expected number of empty buckets?

a) Any string is equally likely to hash into any bucket, thus

$$\boxed{X_1 \sim \mathcal{U}(0, m-1)}$$

b) $P(X_1 = X_2) = \frac{1}{m}$, as $\begin{cases} P(X_1) = \frac{1}{m} \text{ and } P(X_2) = \frac{1}{m} \\ \text{any strings are independently and equally likely to be hashed into any bucket} \end{cases}$

c) + The prob that each string is in the bucket is p, and $(1-p)$ otherwise
$\frac{1}{m}$ "

Thus $\boxed{Y_1 \sim \mathcal{B}(n, \frac{1}{m})}$

+ $P[Y_1 = 0] = \boxed{\left(1 - \frac{1}{m}\right)^n}$

d) Let $Z_i = \begin{cases} 1 & \text{if the buck } i \text{ is empty} \\ 0 & \text{otherwise} \end{cases}$ $(0 \le i \le m-1)$

Then $P[Z_i] = \left(1 - \frac{1}{m}\right)^n$

As all strings are equally likely to be hashed into any bucket, we have:

$$E[Z] = \sum_{i=0}^{m-1} \left(1 - \frac{1}{m}\right)^n$$

6 Suppose you are working at a technology company ABC, and you are unfortunately on-call for your team the entire year (that means, you are the person that other stakeholders may ping in the middle of the night to debug production issues). There are 5 Software Engineers on your team (including yourself), and each person ~~independently introd~~uces on average 0.1 bugs per work-week (Mon-Fri).

a) What is the probability of having a bug-free work-week?
b) What is the probability of having a ~~bug-free~~ day? W~~hat's~~ the relationship between your answer to this part ~~and~~ the previous part?
c) What is the probability that in a (52-week) year, that there are at least 40 bug-free weeks?
d) Suppose it's the first Monday of the year. When would you expect the first day where you had to debug (at least) one issue (in number of work-days from today)?
e) Suppose it's the first Monday of the year. What is the probability that your tenth day of debugging happens in February or ~~later (>20~~ work-days ~~from now~~)?

a) let X be the number of bug-free work-week

Then X follows the Poisson distribution.

Each person independently introduces on average 0.1 bugs per work week $\Rightarrow$ The team introduces 0.5 bugs in total

$\Rightarrow \lambda = 0.5$

Thus: $p(x=0) = \frac{0.5^0}{0!} e^{-0.5} = \boxed{0.606} = P_1$

b) $\lambda = 0.5$ for a work-week $\Rightarrow \lambda = 0.1$ for a work day.

Thus:

$P_2 = P(x=0) = \frac{0.1}{0!} e = \boxed{0.405}$

Relationship: $\boxed{P_2 = P_1^{1/5}}$

c) $P(x \geq 40) = \sum_{i=40}^{52} \binom{i}{52} p^i (1-p)^{52-i}$

$\approx \boxed{9.909 \times 10^{-3}}$

d)

X	1	2	...	n
$P(X=x)$	P_d	$(1-P_d)P_d$	...	$(1-P_d)^{n-1} P_d$

$P_d = (1-0.905)$

X = the day that the first bug appear.

Thus $E[X] = \sum_{k=1}^{n} k\, P_d (1-P_d)^{k-1}$

$= P_d \underbrace{\sum_{k=1}^{\infty} k (1-P_d)^{k-1}}_{S_n}$

Thus $S_n = \int_1^{\infty} S_n\, d.P_d = \int_1^{\infty} \sum_{i=1}^{\infty} i (1-P_d)^{i-1} dP_d$

$= \sum_{i=1}^{\infty} \int_1^{\infty} i (1-P_d)^{i-1} dP_d$

$= -\sum_{i=1}^{\infty} (1-P_d)^i = -\frac{0-(1-P_d)}{-1+(1-P_d)} = -\frac{1}{P_d} + 1$

$\Rightarrow S_n = \frac{1}{P_d^2} \Rightarrow \boxed{E_X = \frac{1}{P_d}} \approx 10.526$

$\Rightarrow$ The first day to debug is after $\boxed{11}$ work-days

e) $P(W > 20) = \sum_{i=21}^{\infty} \binom{9}{i-1} P_d^{10} \cdot (1-P_d)^{i-10}$

$= \boxed{\sum_{i=21}^{\infty} \binom{9}{i-1} (0.905)^{i-10} (0.095)^{10}}$