# Problem Set 10

1. MIT students sometimes delay laundry for a few days. Assume all random values described below are mutually independent.

a. A busy student must complete 3 problem sets before doing laundry. Each problem set requires 1 day with probability 2/3 and 2 days with probability 1/3. Let B be the number of days a busy student delays laundry. What is E[B]?

$P(P=1) = \frac{2}{3}, \quad P(P=2) = \frac{1}{3}$

$E[P] = \sum_{p_i} (P=p_i)p_i = 1(\frac{2}{3}) + 2(\frac{1}{3}) = \frac{4}{3}$

$E[B] = 3E[P] = 3(\frac{4}{3}) = \boxed{4}$

b. A 'relaxed' student rolls a fair, 6-sided die in the morning. If he rolls a 1, then he does his laundry immediately (with 0 days of delay). Otherwise, he delays for one day and repeats the experiment the following morning. Let R be the number of days a relaxed student delays laundry. What is E[R]?

$E[R] = E[\#\text{rolls to get a } 1] - 1$ ← since he does laundry immedictely

$= \frac{1}{p} - 1 = 6 - 1 = \boxed{5}$

mean time to failure

$p(D=1) = \frac{1}{6}$

C. Before doing laundry, an unlucky student must recover from illness for a number of days equal to the product of the numbers rolled on two fair, 6-sided dice. Let U be the expected number of days an unlucky delays laundry. What is $E[U]$?

Let $D_i$ be the number rolled on fair die $i$.

$U = D_1 D_2$

$E[U] = E[D_1 D_2] = E[D_1]E[D_2] = (3.5)(3.5) = \boxed{12.25}$

$E[D] = \frac{1}{6}\left(\sum_{i=1}^{6} i\right) = \frac{1+2+3+4+5+6}{6} = 3.5$

d. A student is busy with a probability 1/2, relaxed with probability 1/3, and unlucky with probability 1/6. Let $D$ be the number of days the student delays laundry. What is $E[D]$?

$$E[D] = P(S=b)E[B] + P(S=r)E[R] + P(S=u)E[U]$$
$$= \boxed{\frac{1}{2}(4) + \frac{1}{3}(5) + \frac{1}{6}\left(\frac{49}{4}\right)}$$

2. There are about $2.5 \cdot 10^8$ people in the US who might use a phone. Assume that each person is on the phone during each minute with probability $p = 0.01$.

a. What is the expected number of people on the phone at a given moment?

Let P be a RV representing if a given person is on the phone at a given moment.

$P(P=0) = 0.99$ $\quad P(P=1) = 0.01$ $\quad E[P] = 0.01 = 10^{-3}$

Let N be the number of people on a phone at a given moment.

$E[N] = (\#\text{ people})\, E[P] = (2.5 \cdot 10^8)\, 10^{-3} = \boxed{2.5 \cdot 10^5}$

b. Suppose that we construct a phone network whose capacity is a mere <u>one percent</u> above the expectation. Upper bound the probability that the network is overloaded in a given minute.

min fraction off the phone

We want to find $\mathrm{CDF}((1-0.0101)2.5\cdot 10^8)$ for a binomial distribution

from the notes $\mathrm{CDF}(\alpha n) \leq \left(\frac{1-\alpha}{1-\alpha/p}\right)\frac{2^{nH(\alpha)}}{\sqrt{2\pi\alpha(1-\alpha)n}}\, p^{\alpha n}(1-p)^{(1-\alpha)n}$

$$\approx \boxed{e^{-120}}$$

Where $n = 2.5\cdot 10^8$

$p = 0.99$ (off phone)

$\alpha = 0.9899 = (1-0.0101)$

C. What is the expected number of minutes (approximately) until the system is overloaded for the first time?

(Mean time to failure) $= \frac{1}{p} = \frac{1}{e^{-120}} = \boxed{e^{120} \text{ minutes}}$

↑ from b.

3. We are given a set of $n$ distinct positive integers. We then determine the maximum of these numbers by the following procedure:
Randomly arrange the numbers in a sequence
Let the "current maximum" initially be the first number in the sequence and the "current element" be the second element of the sequence. If the current element is greater than the current maximum, perform an "update": that is, change the current maximum to the current element. Either way, change the current element to be the next element of the sequence. Repeat this process until there is no next element.
Prove that the expected number of updates is $\sim \ln(n)$

Let $M_i$ be the indicator variable for the event that the ith element of the sequence is bigger than all the previous elements.
While iterating through a sequence, $M_i = 1 \Rightarrow$ update, $M_i = 0 \Rightarrow$ no update
$U = \sum_{i=1}^{n} M_i$
$P(M_i = 1) = \frac{1}{(i)}$ since the integers should be uniformly distributed
$E(U) = E(\sum_{i=1}^{n} M_i) = \sum_{i=1}^{n} E[M_i] = \sum_{i=1}^{n} \frac{1}{i} \sim \ln(n)$ since this is a harmonic series.

4. In a certain card game, each card has a point value.

- Numbered cards are worth 5 points each (2-9)
- 10s & face cards are worth 10 points each
- Aces are worth 15 points each.

a. Suppose that you thoroughly shuffle a 52-card deck. What is the expected total point value of the three cards on top of the deck after the shuffle?

$X = \sum_{i=1}^{3} C_i$, $C_i$ is the ith card in a shuffled deck

$P(C=5) = \frac{4(8)}{52} = \frac{8}{13}$

$P(C=10) = \frac{4(4)}{52} = \frac{4}{13}$

$P(C=15) = \frac{1(4)}{52} = \frac{1}{13}$

$E[C] = 5\left(\frac{8}{13}\right) + 10\left(\frac{4}{13}\right) + 15\left(\frac{1}{13}\right) = \frac{95}{13}$

$E[X] = 3E[C] = \boxed{\frac{285}{13}}$

all drawn at once

b. Suppose that you throw out all the red cards and shuffle the remaining 26-card, all-black deck. Now what is the expected total point value of the top 3 cards?

The same as a. Our solution did not consider conditional probabilities of different combinations of hands, so it is unchanged if the probabilities stay the same (they do).

5. A true story from World War II:

The army needs to identify soldiers with a disease called "klep". There is a way to test blood to determine whether it came from someone with klep. The straightforward approach is to test each soldier individually. This requires $n$ tests, where $n$ is the number of soldiers. A better approach is the following: group the soldiers into groups of $k$. Blend the blood samples of each group and apply the test once to each blended sample. If the group-blend doesn't have klep, we are done with that group after one test. If the group-blend fails the test, then someone in the group has klep, and we individually test all the soldiers in the group. Assume each soldier has klep with probability, $p$, independently of all other soldiers.

a. What is the expected number of tests as a function of $n, p,$ and $k$? (Assume $n$ is divisible by $k$ for simplicity)

Let $G_i$ be a RV representing the number of tests performed on the $i$th group of $k$ people.

$$P(G_i=1) = \prod_{j=1}^{k} P(\text{person}_j \text{ doesn't have klep})$$
$$= \prod_{j=1}^{k}(1-p) = (1-p)^k$$
$$P(G_i=k+1) = 1-(1-p)^k$$
$$E[G_i] = 1(1-p)^k + (1+k)(1-(1-p)^k)$$
$$= k+1-k(1-p)^k$$

Let $T$ be a RV representing # total tests performed

$$E[T] = E\left[\sum_{i=1}^{n/k} G_i\right] \qquad T = \sum_{i=1}^{n/k} G_i$$
$$= \sum_{i=1}^{n/k} E[G_i]$$
$$= \frac{n}{k}(k+1-k(1-p)^k)$$
$$= \boxed{n\left(\frac{1}{k}+1-(1-p)^k\right)}$$

b. How should $k$ be chosen to minimize the expected number of tests performed, and what is the resulting expectation.

Minimize by finding $k \ni \frac{d}{dk}(E[T]) = 0$

$\frac{d}{dk}\left(n(1-(1-p)^k + \frac{1}{k}\right) = n(-\ln(1-p)(1-p)^k - \frac{1}{k^2}) = 0$

for $p << \frac{1}{k}$, $(1-p)^k \sim 1 \wedge \ln(1-p) \sim -p$

$\frac{1}{k^2} = p \Rightarrow \boxed{k \approx \frac{1}{\sqrt{p}}}$

$E[T(k \approx \frac{1}{\sqrt{p}})] \approx n(1-1+\sqrt{p}) = \boxed{n\sqrt{p}}$

C. What fraction of the work does the grouping method expect to save over the straightforward approach in a million-strong army where 1% have klep?

Using the approximations from part b.

(fraction of work saved) $= 1 - \frac{E[T]}{n} = 1 - \frac{\sqrt{p}\,n}{n} = 1 - \sqrt{p} = 1 - \sqrt{0.01} = \boxed{0.9}$

6. The hat-check staff has had a long day, and at the end of the party they decide to return people's hats at random. Suppose that $n$ people have their hats returned at random. We have previously shown that the expected number of people who get their own hat back is $1$, irrespective of the total number of people. In this problem we will calculate the variance in the number of people who get their hat back.
Let $X_i = 1$ if the ith person gets their hat back and $0$ otherwise. Let $S_n ::= \sum_{i=1}^{n} X_i$, so $S_n$ is the total number of people who get their own hat back. Show that

a. $E[X_i^2] = \frac{1}{n}$

$P(X_i = 1) = \frac{1}{n} \quad P(X_i = 0) = 1 - \frac{1}{n}$
$P(X_i^2 = 1^2) = \frac{1}{n} \quad P(X_i^2 = 0^2) = 1 - \frac{1}{n}$
$E[X_i^2] = 1P(X_i^2 = 1) + 0\,P(X_i^2 = 0) = \frac{1}{n}$ //

b. $E[X_iX_j] = \frac{1}{n(n-1)}$ for $i \neq j$

$X_iX_j = 0$ iff $X_i = X_j = 1$

$P(X_iX_j = 1) = P(X_i = 1 \wedge X_j = 1)$

$= P(X_i = 1)(P(X_i = 1 | X_j = 1))$

$= \frac{1}{n}\left(\frac{1}{n-1}\right) \leftarrow$ since 1 hat is already correctly assigned

else $X_iX_j = 0$

So $E[X_iX_j] = 1(P(X_iX_j = 1)) = \frac{1}{n(n-1)}$ //

c. $E[S_n^2] = 2$ Hint: use (a) & (b)

$$E[S_n^2] = E[(\sum_{i=1}^{n} X_i)^2]$$

$$= E[\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} \sum_{j \neq i}^{n} X_i X_j]$$

$$= \sum_{i=1}^{n} E[X_i^2] + \sum_{i=1}^{n} \sum_{j \neq i} E[X_i X_j]$$

$$= nE[X_i^2] + n(n-1)E[X_i X_j]$$

$$= n\frac{1}{n} + \frac{n(n-1)}{n(n-1)} = 2 \;//$$

d. $Var[S_n] = 1$

$$Var[S_n] = E[S_n^2] - (E[S_n])^2$$
$$= 2 - (1)^2 \quad \text{(since we expect 1 person to get their hat back)}$$
$$= 1$$

e. Explain why you cannot use the variance of sums formula to calculate $Var[S_n]$.

The $X_i$'s are not independent.

f. Using Chebyshev's Inequality, show that $P(S_n \geq 11) \leq 0.01$ for any $n \leq 11$

Chebyshev's Inequality tells us

$$P(S_n - E[S_n] \geq 10) \leq \frac{Var[S_n]}{10^2}$$

$$P(S_n \geq 10 + E[S_n]) \leq \frac{1}{10^2}$$

$$P(S_n \geq 11) \leq 0.01 \,//$$

7. Let $R_1$ and $R_2$ be independent random variables, and $f_1$ and $f_2$ be functions such that domain$(f_i)$ = codomain$(R_i)$ for $i=1,2$. Prove that $f_1(R_1)$ and $f_2(R_2)$ are independent random variables.

<u>Proof</u> Consider $(Y_1, Y_2) \in$ (domain $f_1$, domain $f_2$)

$P(Y_1 = f_1(R_1) \wedge Y_2 = f_2(R_2))$

$= P(R_1 \in f_1^{-1}(Y_1) \wedge R_2 \in f_2^{-1}(Y_2))$

$= \sum_{x_1 \in f_1^{-1}(Y_1) \wedge x_2 \in f_2^{-1}(Y_2)} P(R_1 = x_1 \wedge R_2 = x_2)$ since $f_i(R_i) = Y_i$ is the disjoint union of all the events $R_i = r$ for $r \in f_i^{-1}(Y_i)$

$= \sum_{x_1 \in f_1^{-1}(Y_1) \wedge x_2 \in f_2^{-1}(Y_2)} P(R_1 = x_1) P(R_2 = x_2)$ since $R_1, R_2$ are independent

& $(f_1(R_1) = a \wedge f_2(R_2) = b)$ is the disjoint union of events $R_1 = x_1$ and $R_2 = x_2$ for $(x_1, x_2) \in f_1^{-1}(a) \times f_2^{-1}(b)$.

$= \sum_{R_1 \in f_1^{-1}(Y_1)} P(R_1) \sum_{R_2 \in f_2^{-1}(Y_2)} P(R_2)$

$= P(R_1 \in f_1^{-1}(R_1)) P(R_2 \in f_2^{-1}(R_2))$

$= P(Y_1 = f_1(R_1)) P(Y_2 = f_2(R_2))$

by the def of independent, $f_1(R_1), f_2(R_2)$ are independent //

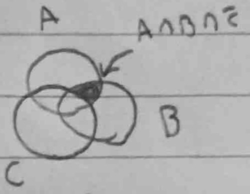
8. Let $A, B, C$ be events, and let $I_A, I_B, I_C$ be the corresponding indicator variables. Prove that $A, B, C$ are mutually independent iff the random variables $I_A, I_B, I_C$ are mutually independent.

Proof

$\rightarrow$

Suppose RVs $A, B, C$ are mutually independent

$P(I_A=1 \cap I_B=1 \cap I_C=0)$




$= P(A \cap B \cap \bar{C})$

$= P(A \cap B) - P(A \cap B \cap C)$

$= P(A)P(B) - P(A)P(B)P(C)$ by independence

$= P(A)P(B)(1 - P(C))$

$= P(A)P(B)P(\bar{C})$

$= P(I_A=1)P(I_B=1)P(I_C=0)$

this process can be repeated for all triplets of $(I_A, I_B, I_C)$ values to show

$A, B, C$ $\underset{\text{mutually}}{\wedge}$ independent $\Rightarrow$ $I_A, I_B, I_C$ $\underset{\text{mutually}}{\wedge}$ independent

$\leftarrow$

Suppose $I_A, I_B, I_C$ are independent

$P(A \cap B \cap C) = P(I_A=1 \cap I_B=1 \cap I_C=1)$

$= P(I_A=1)P(I_B=1)P(I_C=1)$ independence

$= P(A)P(B)P(C)$

Consider an arbitrary pair $(X, Y)$ from $\{A, B, C\}$, let $Z$ be the non-pair value

$P(X \cap Y) = P(X \cap Y \cap Z) + P(X \cap Y \cap \bar{Z}) = P(I_X=1 \cap I_Y=1 \cap I_Z=1) + P(I_X=1 \cap I_Y=1 \cap I_Z=0)$

$= P(I_X=1)P(I_Y=1)P(I_Z=1) + P(I_X=1)P(I_Y=1)P(I_Z=0)$

$= P(I_X=1)P(I_Y=1)(P(I_Z=1) + P(I_Z=0)) = P(X)P(Y)(P(Z) + P(\bar{Z}))$

$= P(X)P(Y)$, so $I_A, I_B, I_C$ mutually independent $\Rightarrow$ $A, B, C$ mutually independent

Thus, $A, B, C$ mutually independent iff $I_A, I_B, I_C$ mutually independent //