Massachusetts Institute of Technology
6.042J/18.062J: Mathematics for Computer Science
Professors David Karger and Nancy Lynch

Handout 44
April 28, 2000

Mini-Quiz 12

1. Write your name:

2. (Dice game handout, Ex. 1) Suppose a game involves tossing a fair coin until a head first arises. What is the probability that a head first arises on an even-numbered throw?

   Solution.

   The probability that a head first arises on an even-numbered throw can be easily broken into two cases based on the first flip. Half the time, a head comes up on the first flip and thus the first head does not appear on an even toss. The other half of the time, the first head will be on an even toss only if it is on a odd toss in the remaining sequence of tosses. Thus:

$$\begin{aligned} p &= \frac{1}{2}(1-p) \\ p &= \frac{1}{2} - \frac{1}{2}p \\ \frac{3}{2}p &= \frac{1}{2} \\ p &= \frac{1}{3} \end{aligned}$$

Tutorial 12 Problems

Problem 1 The St. Petersburg Casino offers the following game: the gambler bets a fixed wager, and then the dealer flips a fair coin (dealers do not flip coins in U.S. casinos, but they do in St. Petersburg) until it comes up heads. The gambler receives \$1 if the coin shows heads at the first toss, \$2 if it shows the first head at second toss, and in general $\$2^{k-1}$ if the dealer tosses the coin k times to get the first head.

(a) Suppose the fixed wager is \$10. What is the expected amount of money that the gambler will win in this game? Suppose the fixed wager is \$10,000?

Solution.

Let V be the random variable corresponding to the amount of money that the gambler is paid by the dealer. The distribution of V is as follows: for any $n \geq 1$,

$$\Pr\left[\, V = 2^{n-1} \,\right] = 2^{-n}$$

because the event that the money paid is 2^{n-1} is the same as the event that the dealer tosses $n-1$ tails followed by a head. The average of V is

$$\mathrm{E}\left[V\right] = \sum_{n=1}^{\infty}[2^{-n}2^{n-1}] = \sum_{n=1}^{\infty} 1/2 = \infty$$

So whatever the fixed wager, the gambler expects to win an infinite amount.

(b) What is the probability that the gambler does not lose money in a game when the fixed wager is \$10,000?

Solution.

The probability of recovering the money of the wager is only

$$\begin{aligned}
\Pr\left[\, V \geq 10,000 \,\right] &= \Pr\left[\, V \geq 2^{14} \,\right] \\
&= \Pr\left[\,\text{dealer flips at least 14 consecutive tails}\,\right] \\
&= 2^{-14} = \frac{1}{16,384}
\end{aligned}$$

(c) In reality, it would not be reasonable for the gambler to earn money by repeatedly playing the game with the fixed wager of \$10,000. Why?

Solution.

The model ignores the fact the both the casino and the player can lose only bounded amounts of money before going bankrupt. Truncating the game when either party goes bankrupt can completely change the expected winnings.

For example, since the probability of the gambler coming out ahead by any amount in a single game with the \$10,000 entry fee is 2^{-14}, the expected number of games that must be played before this happens is 2^{14}, so if the player has less than $\$2^{14} \cdot 10^5 = \$1,638,400,000$, he can expect to go bankrupt before coming out ahead in even one game.

Problem 2

(a) Suppose we flip a fair coin until two heads in a row come up. What is the expected number, F, of flips we perform? *Hint*: Let F_{T} be the expected number of further flips until two heads comes up, given that the previous flip was T, and likewise let F_{H} be the expected number of further flips until two heads comes up given that the previous flip was H. Argue that F_{T} will equal 1 plus half of F_{T} and half of F_{H}.

Solution.

Clearly $F = F_{\mathrm{T}}$. Assuming the hint

$$F_{\mathrm{T}} = 1 + \frac{1}{2}F_{\mathrm{T}} + \frac{1}{2}F_{\mathrm{H}}$$

Similarly,

$$F_{\mathrm{H}} = 1 + \frac{1}{2}F_{\mathrm{T}} + \frac{1}{2}(0)$$

Solving for F_{T} yields $F = 6$.

This can also be solved using a single equation formed by examining that cases for the first and second flips, but it is not as clean.

(b) Suppose we flip a fair coin until a head followed by a tail come up. What is the expected number, G, of flips we perform?

Solution.

Clearly $G = G_{\mathrm{T}}$. As above

$$\begin{aligned} G_{\mathrm{T}} &= 1 + \frac{1}{2}G_{\mathrm{T}} + \frac{1}{2}G_{\mathrm{H}} \\ G_{\mathrm{H}} &= 1 + \frac{1}{2}(0) + \frac{1}{2}G_{\mathrm{H}} \end{aligned}$$

So $G = G_{\mathrm{T}} = 4$.

(c) Suppose we now play a game: flip a fair coin until either HH or HT first occurs. You win if HT comes up first, lose if HH comes up first. What odds should you offer an opponent to make this a fair game?

Solution.

Even money odds. Although the expected time for HH is larger than for HT, the game of waiting for one or the other to come up first is fair!

To prove this, let W be the probability of winning. Clearly,

$$W = (1/2)\Pr[W \mid \text{first roll is } T] + (1/2)\Pr[W \mid \text{first roll is } H] = (1/2)W + (1/2)(1/2)$$

so $W = 1/2$.

Problem 3 The Coupon Collector Problem

We are trying to collect baseball cards. Assume there are n different cards. We can buy them one at a time. When we buy a card, it is equally likely that we get any one of the n cards, and purchases are mutually independent. We want to study how many cards we need to buy before we get one each of the n different kinds.

(a) Suppose we already have k different cards. What is the expected number of cards we need to buy to get a new card? *Hint*: Let X_k be the random variable equal to the number of purchases from the first time we accumulated k cards until we managed to purchase a $(k+1)$st new card.

Solution.

After getting k different cards, a new purchase has probability $(n-k)/n$ of getting a new card. We can view each new purchase as Bernoulli trial with probability of success $p = (n-k)/n$. Hence the expected number of trials to success is $1/p = n/(n-k)$. That is,

$$E(X_k) = \frac{n}{n-k}$$

(b) What is the expected number of cards we need to buy to get one of each kind? Express your answer as a simple expression involving a Harmonic number H_k for some k. *Hint*: Consider $\sum_{i=0}^{n-1} X_i$.

Solution.

The number of purchases we need is $\sum_{i=0}^{n-1} X_i$. Note that $X_0 = 1$. By linearity of expectation,

$$
\begin{aligned}
E(1+\sum_{i=1}^{n-1} X_i) &= \sum_{i=0}^{n-1} E(X_i) \\
&= \sum_{i=0}^{n-1} \frac{n}{n-i} \\
&= n\sum_{i=0}^{n-1} \frac{1}{n-i} \\
&= n\sum_{i=1}^{n} \frac{1}{i} \\
&= n\mathrm{H}_n \sim n\ln n + \gamma n + \Theta(1)
\end{aligned}
$$

Where γ is Euler's constant.

(c) Suppose we buy $n \ln n + cn$ cards, where $c > 0$ is a real number. Prove that the probability that we don't have the first card is at most $\frac{1}{ne^c}$. *Hint*: $1 - x \leq e^{-x}$ for $x < 1$.

Solution.

Let A_1 be the event that we don't have the first card in $n \ln n + cn$ trials. In any single trial the probability that we don't get A_1 is $1 - 1/n$. Since the trials are independent,

$$
\begin{aligned}
\Pr(A_1) &= (1 - \frac{1}{n})^{n\ln n + cn} \\
&\leq e^{-\frac{1}{n}(n\ln n + cn)} \\
&= e^{-(\ln n + c)} = \frac{1}{ne^c}
\end{aligned}
$$

(d) Suppose we buy $n \ln n + cn$ cards. Prove that the probability that we don't have all n cards is at most e^{-c}.

Solution.

Let A_i be the event that we don't have the i^{th} card after $n \ln n + cn$ trials.

Then the probability that we don't have all the cards is

$$\Pr(A_1 \cup \ldots \cup A_n) \leq \sum_{i=1}^{n} \Pr(A_i)$$

by Boole's inequality. But this is

$$n\frac{e^{-c}}{n} = e^{-c}$$

(e) Use the previous result to estimate the number of times a cards must be drawn from a standard deck of 52 in order to have 99% confidence of drawing each possible card at least once. Each drawn card is replaced, and the deck is freshly shuffled, between draws.

Solution.

Note that this is a very loose bound, since it uses Boole's inequality above.

Using the previous part, we can solve for c in

$$\begin{aligned} e^{-c} &\leq 0.01 \\ -c &\leq ln(0.01) \\ c &\geq -ln(0.01) \\ c &\geq 4.605 \end{aligned}$$

so, given this, we should buy $52ln(52) + (4.605)52$ cards, or 445 cards.

It is possible to simply calculate the odds for k cards directly, instead of using the last part. We then get:

$$\begin{aligned} n\left(1 - \frac{1}{n}\right)^k &\leq 0.01 \\ \left(1 - \frac{1}{52}\right)^k &\leq \frac{1}{5200} \\ \left(\frac{51}{52}\right)^k &\leq \frac{1}{5200} \\ ln\left[\left(\frac{51}{52}\right)^k\right] &\leq ln\left[\frac{1}{5200}\right] \\ k \cdot ln\left[\frac{51}{52}\right] &\leq -ln\,[5200] \\ k &\geq \frac{ln(5200)}{ln(52) - ln(51)} \\ k &\geq 440.64 \end{aligned}$$

It should not be surprising that this is a somewhat tighter result – the previous method relied on a further, looser bound.