CS 70 Discrete Mathematics and Probability Theory
Fall 2012 Vazirani **Section 11**

1 Indicator Variables.

Definition. *An* ***indicator variable*** *for an event A is a random variable $\mathbf{1}_A : \Omega \to \{0,1\}$, such that for all $\omega \in \Omega$:*

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}.$$

Let $\mathbf{1}_A$ be the indicator r.v. for an event A.

(a) Show that $\mathbb{E}[\mathbf{1}_A] = \Pr[\mathbf{1}_A = 1] = \Pr[A]$.

–Solution–

$$\begin{aligned} \mathbb{E}[\mathbf{1}_A] &= 0 \cdot \Pr[\mathbf{1}_A = 0] + 1 \cdot \Pr[\mathbf{1}_A = 1] \\ &= \Pr[\mathbf{1}_A = 1] \\ &= \Pr[\{\omega | \mathbf{1}_A(\omega) = 1\}] \\ &= \Pr[A]. \end{aligned}$$

(b) Suppose we toss a fair coin n times, landing $Y \in \{0, 1, \dots, n\}$ heads. If X is the indicator variable for the first coin toss landing heads, calculate $\mathbb{E}[X]$.

–Solution–

Let A be the event for the first coin toss landing heads–X is the indicator r.v. for A. Using the previous part:

$$\mathbb{E}[X] = \Pr[A] = \frac{1}{2}$$

(c) Calculate $\mathbb{E}[Y]$. Your answer should be a simple function of n.

[**Hint**: Express Y in terms of indicator variables and use linearity of expectation.]

–Solution–

Let X_k be the event for the k-th coin to land heads. Clearly $Y = \Sigma_{k=1}^{n} X_k$. Therefore

$$\mathbb{E}[Y] = \Sigma_{k=1}^{n} \mathbb{E}[X_k] = \frac{n}{2}$$

(d) Use the previous part to prove the following identity:

$$\sum_{i=1}^{n} i \cdot \binom{n}{i} = n2^{n-1}.$$

–Solution–

Both sides of the equation are $2^n \cdot \mathbb{E}[Y]$. The left-hand side is

$$\begin{aligned} 2^n \cdot \mathbb{E}[Y] &= 2^n \cdot \Sigma_{i=0}^{n} i \cdot \Pr[\text{the coin lands heads exactly i times}] \\ &= 2^n \cdot \Sigma_{i=0}^{n} i \cdot \frac{\binom{n}{i}}{2^n} \\ &= \Sigma_{i=1}^{n} i \cdot \binom{n}{i} \end{aligned}$$

The right-hand side is obvious.

2 Monkey writing Shakespeare

A monkey types on a 26-letter keyboard, with all lowercase letters. Assume that the monkey chooses each character independently and uniformly at random.

(a) The monkey types a million six-letter words at random. What is the expected number of times the word "hamlet" is typed? Let H_i be the event that the ith word typed is "hamlet." Are H_i and H_j independent for $i \neq j$?

–Solution–

Let $\mathbf{1}_{H_k}$ be the indicator variable for H_k, and let $X = \Sigma_{i=1}^{1{,}000{,}000} \mathbf{1}_{H_k}$—the number of times the word "hamlet" appears in total. Clearly,

$$\Pr[H_k] = \frac{1}{26^6}$$

And therefore

$$\mathbb{E}[X] = \Sigma_{i=1}^{1{,}000{,}000}\mathbb{E}[\mathbf{1}_{H_k}] = \frac{10^6}{26^6} = \left(\frac{5}{13}\right)^6$$

The events H_i and H_j are independent.

(b) Now the monkey types a million characters at random. What is the expected number of times the sequence "hamlet" appears? Letting H_i denote the event that the six-letter sequence that starts at the ith character is "hamlet," are H_i and H_j independent for $i \neq j$?

–Solution–

Clearly,

$$\Pr[H_k] = \begin{cases} 26^{-6} & k \leq 10^6 - 5 \\ 0 & \text{otherwise} \end{cases}$$

And therefore

$$\mathbb{E}[X] = \Sigma_{i=1}^{1{,}000{,}000}\mathbb{E}[\mathbf{1}_{H_k}] = (10^6 - 5) \cdot 26^{-6}$$

The events H_i and H_j are independent iff $|j - i| \geq 6$, because when H_i is occurs, $H_{i+1}, \ldots, H_{i+5}$ do not occur.

(c) Finally the monkey types a six-letter word at random. The monkey copies this word a million times to make a million-word text (meaning spaces between words are retained). What is the expected number of times the word "hamlet" appears in the text?

Letting H_i be the event that the ith word is "hamlet," are H_i and H_j independent for $i \neq j$?

–Solution–

Same result as part (a). The events H_i and H_j are never independent.

(d) Let random variable X be the number of times "hamlet" appears. Think about what the distribution of X looks like in each of the three cases (a)-(c) above. Are any of the three distributions the same?

–Solution–

Part (a) and (c) have the same expectation, but a completely different distribution. In part (c), Y is either 10^6 or 0, in (b) Y can be at most $10^6 - 5$, in part (a) Y can have any value between 0 and 10^6.

3 Number of Runs

We toss a fair coin n times. Runs are consecutive tosses with the same result. For instance, the toss sequence HHHTTHTH has 5 runs. What is the expected number of runs?

–Solution–

The first toss always starts a new run, while every other toss has a $\frac{1}{2}$ probability to start a new run. Let $\mathbf{1}_k$ be the indicator variable for whether toss k starts a new run.Let $X = \Sigma_{k=1}^{n} 1_k$ —the number of runs.

$$\mathbb{E}[X] = \Sigma_{k=1}^{n} \mathbb{E}[1_k] = 1 + (n-1) \cdot \frac{1}{2} = \frac{n+1}{2}$$

4 A Winning Strategy

Suppose you are at a casino and betting on a game in which you have probability p of winning in each round. Your strategy is to bet 1 in the first round and then in each subsequent round, double the amount of money you bet in the previous round. So you would bet 1, 2, 4, 8, and so on. You stop as soon as you earn a profit.

(a) Suppose you have unlimited money. What is the expected amount of money you will earn?

–Solution–

If you win in the k-th round, the amount won is 2^k, while the previous losses are $2^k - 1$. The total amount won is exactly 1. Let A_k be the event in which you win in round k, and let $\mathbf{1}_{A_k}$ be the indicator variable for A_k. We claim that the r.v.

$$X = \Sigma_{k=1}^{\infty} \mathbf{1}_{A_k}$$

Is the amount won in each game. If you win in round k, the amount won would be 1. The events A_k are all disjoint, so $X = 1$ on each such outcome.

$$\begin{aligned}
\mathbb{E}[X] &= \Sigma_{k=1}^{\infty} \mathbb{E}[\mathbf{1}_{A_k}] \\
&= \Sigma_{k=1}^{\infty} \Pr[A_k] \\
&= \Sigma_{k=1}^{\infty} (1-p)^{k-1} p \\
&= 1
\end{aligned}$$

(b) What is the expected number of rounds you will play?

–Solution–

The number of rounds played is distributed geometrically with parameter p. The mean is known to be $\frac{1}{p}$.

(c) What is the expected amount of money lost before your first win?

–Solution–

Let X be the total amount lost before the first win.

$$\begin{aligned}
\mathbb{E}[X] &= \Sigma_{k=1}^{\infty} 2^{k-1} \cdot \Pr\left[X = 2^{k-1}\right] \\
&= \Sigma_{k=1}^{\infty} 2^{k-1} (1-p)^{k-1} p \\
&= p \Sigma_{k=0}^{\infty} (2(1-p))^{k} \\
&= \frac{p}{2p-1}
\end{aligned}$$

Note that the sum is finite (and so is the expectation), iff $p > \frac{1}{2}$.