15-359: Probability and Computing

Assignment 2 | Due: February 3, 2012

You may write the exercises in less detail than the other questions.

Problem 1: Cell block (exercise) (5 pts.)

Calculate $E(\frac{X}{Y}|X^2 + Y \leq 10)$ from the joint probability of X and Y:

	X = 1	X = 2	X = 4
Y = 3	1/12	2/12	1/12
Y = 8	3/12	1/12	1/12
Y = 12	1/12	0/12	2/12

Problem 2: Friend of a friend (exercise) (5 pts.) We say the events A and B are conditionally independent given an event C if

$$P(A \cap B|C) = P(A|C) \cdot P(B|C)$$

Show that

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

Safely assume that all probabilities are greater than 0.

Problem 3: Expecting something different? (exercise) (5 pts.) Let X be a non-negative, discrete, integer-valued random variable. Prove:

$$E(X) = \sum_{x=0}^{\infty} P(X > x)$$

Problem 4: Big data (exercise) (10 pts.) You are told that the average file size in a database is $6K$.

A. Explain why it follows (from the definition of expectation) that fewer than half of the files can have size $> 12K$.

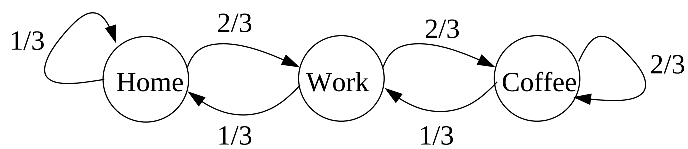
B. You are now given the additional information that the minimum file size is $3K$. Derive a tighter upper bound on the percentage of files which have size $> 12K$.

Problem 5: Making a stack of coins out of fish (10 pts.)

Prove that the Binomial(n, p) distribution is well-approximated by the Poisson(np) distribution when n is large and p is small. Hint: Start with the probability mass function for the Binomial(n, p) distribution. Set $p = \lambda/n$. Expand out all the terms. Take limits and show you get a Poisson(λ) distribution, where $\lambda = np$.

Problem 6: Coffee-theorem automata (20 pts.)

A student relocates every hour according to the probabilistic process shown below. If the student is home, she will go to work in the next hour with probability $\frac{2}{3}$ and will stay home in the next hour with probability $\frac{1}{3}$. The student can't stand to work more than an hour, however, so after working for an hour, she will either go get coffee (with probability $\frac{2}{3}$) or will go back home (with probability $\frac{1}{3}$). The student loves the coffee shop, so every hour she is there, she has probability $\frac{2}{3}$ of staying for another hour.



Assuming that the student is currently at work, what is the expected time (in hours) until she goes home?

Problem 7: Expecting to be astonished (25 pts.)

For any probability $p \in [0, 1]$, let $a(p) = \log_2(1/p)$ be the astonishment function, named thusly due to its following properties:

1. $a(1) = 0$ (something that happens with probability 1 isn't astonishing.)
2. $a(p_1) > a(p_2)$ if $p_1 < p_2$ (it's more astonishing when more unlikely outcomes happen.)
3. $a(p)$ is continuous in p (changing an outcome's probability by a little bit should change our astonishment by only a little bit.)
4. $a(p_1 \cdot p_2) = a(p_1) + a(p_2)$ (a product of probabilities corresponds to two independent events, which astonish us separately.)

Let X be a random variable taking n possible values $x_1, \ldots, x_n$. The excitement of X is the expected astonishment of X (how astonished we are to learn X, on average):

$$C(X) = \sum_{i=1}^{n} P(X = x_i) a(P(X = x_i))$$

(It's weird that $P(X = x_i)$ appears twice, but keep in mind that a takes a probability. Excitement is really a property of the *distribution* of X.)

A. Show $C(X) \geq 0$, with equality holding if and only if X is a constant random variable.

B. Let Y be another random variable taking values $y_1, \ldots, y_m$. The joint excitement of X and Y is

$$C(X,Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(X = x_i, Y = y_j) a(P(X = x_i, Y = y_j))$$

and the conditional excitement of X given Y is

$$C(X|Y) = \sum_{j=1}^{m} \sum_{i=1}^{n} P(X = x_i, Y = y_j) a(P(X = x_i | Y = y_j))$$

Show that $C(X,Y) = C(X|Y) + C(Y)$ for all Y.

C. Let's say Ω is our (finite) sample space. Let's say $p(x)$ is the PMF of X, i.e. $p(x) = P(X = x)$. Why is $1/p(X)$ a discrete random variable? Show that $C(X) = E(\log_2 1/p(X))$.

D. What's the maximum possible value of $C(X)$? What kind of X achieves this? (Hint: use Jensen's inequality for discrete random variables.)

Problem 8: You may call a k-clause a Klaus (20 pts.) Recall that a k-CNF boolean formula defined upon a set of boolean variables is an 'and' of clauses, each of which is an 'or' of length k. For example, here's a 3-CNF defined upon some boolean x_i:

$$(x_1 \vee x_3 \vee \neg x_8) \wedge (x_2 \vee \neg x_1 \vee x_5) \wedge (x_8 \vee x_5 \vee x_1) \wedge (\neg x_2 \vee x_5 \vee x_3)$$

Suppose you are given a k-CNF in which

- no variable is negated, and
- there are fewer than 2^{k-1} clauses.

Prove there is a 'mixed' assignment of the variables wherein each clause includes at least one true variable and at least one false variable.

Hint: this problem has nothing to do with satisfiability. Employ a proof strategy called the probabilistic method. To prove some claim C, create your own experiment (i.e. sample space and probability measure). Define a 'good event' which implies C. Show that this event has probability greater than zero. Conclude that C is true.

Problem 9: Shearing off projections (extra credit) (15 pts.) Suppose A is a set of n points in $\mathbb{R}^3$. Consider the number of two-dimensional projections

$$\begin{aligned} c_x &= |\{(y,z) : (x,y,z) \in A\}| \\ c_y &= |\{(x,z) : (x,y,z) \in A\}| \\ c_z &= |\{(x,y) : (x,y,z) \in A\}| \end{aligned}$$

Prove that $n^2 \leq c_x \cdot c_y \cdot c_z$ by (hint) picking a point from A in an exciting fashion and conditioning a bunch of times.