Discrete Math in CS	**Problem Set 8**
CS 280 Fall 2000	**Due Friday, December 1, 2000**

Reading: Rosen, Sections 4.4-4.5

---

(1) On the last problem set we talked about network fault-tolerance — how resilient is a network to the failure of nodes or edges?

Here's a simple probabilistic model we can consider. Let G be a connected graph, and consider the following random experiment. Each node in G *fails* with probability $\frac{1}{2}$. We then let H be the subgraph obtained from G by deleting all nodes that have failed, and all edges incident to these nodes. We say that G *partitions* if the subgraph H contains at least one node, and H is not connected. (So G has *not partitioned* if H is connected, or if all nodes of G have failed.)

Suppose that G is equal to a cycle on four nodes. Determine the probability that G partitions in the experiment above.

In particular, describe the underlying sample space, the event in question, and its probability. (You do not need to write out the entire sample space or the event; it is enough to give a precise mathematical formulation of it.)

(2) In class, we've used the problem of load-balancing jobs on processors as an on-going application of probabilistic analysis. Suppose we have n identical computing servers, and n jobs that need to be performed. However, because we need to route the jobs quickly, with limited knowledge of the global state of the system, we simply assign each job independently and uniformly at random to one of the servers. That is, the jobs arrive in sequence, and each is assigned uniformly at random to one of the servers, independent of the assignment of previous jobs.

Determine the probability that each server is assigned exactly one job, by giving a formula for this quantity in terms of n. In particular, describe the underlying sample space, the event in question, and its probability. (You do not need to write out the entire sample space or the event; it is enough to give a precise mathematical formulation of it.)

Also, show that this probability converges to 0 as n increases.

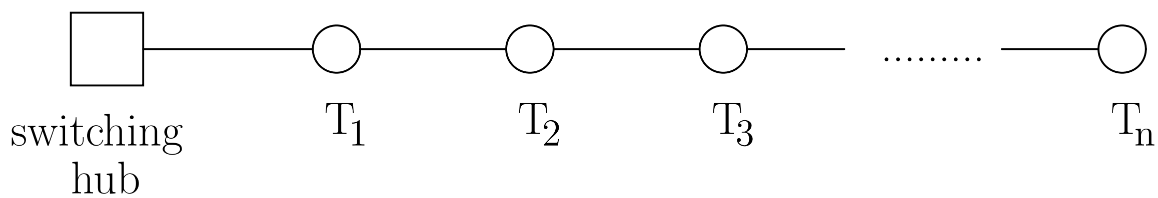
---

(3) Consider a county in which $100,000$ people vote in an election. There are only two candidates on the ballot: a Democratic candidate (denoted D) and a Republican candidate (denoted R). As it happens, this county is heavily Democratic, so $80,000$ people go to the polls with the intention of voting for D and $20,000$ go to the polls with the intention of voting for R.

However, the layout of the ballot is a little confusing, so each voter, independently and with probability $\frac{1}{100}$, votes for the wrong candidate; i.e. the one that he or she *didn't* intend to vote for. (Remember that in this election, there are only two candidates on the ballot.)

Let X denote the random variable equal to the number of votes received by the Democratic candidate D, when the voting is conducted with this process of error. Determine the expected value of X, and give an explanation of your derivation of this value.

(4) Out in a rural part of the state somewhere, n small towns have decided to get connected to a large Internet switching hub via a high-volume fiber-optic cable.



The towns are labeled $T_1, T_2, \ldots, T_n$, and they are all arranged on a single long highway, so that town T_i is i miles from the switching hub.

Now, this cable is quite expensive; it costs k dollars per mile, resulting in an overall cost of kn dollars for the whole cable. The towns get together and discuss how to divide up the cost of the cable.

First, one of the towns way out at the far end of the highway makes the following proposal.

> **Proposal A.** Divide the cost evenly among all towns, so each pays k dollars.

There's some sense in which Proposal A is fair, since it's like each town is paying for the mile of cable directly leading up to it.

But one of the towns very close to the switching hub objects, pointing out that the far-away towns are actually benefitting from a large section of the cable, whereas the close-in towns only benefit from a short section of it. So they make the following counter-proposal.

> **Proposal B.** Divide the cost so that the contribution of town T_i is proportional to i, its distance from the switching hub.

One of the other towns very close to the switching hub points out that there's another way to do a non-proportional division which is also natural. This is based on conceptually dividing the cable into n equal-length "edges" $e_1, \ldots, e_n$, where the first edge e_1 runs from the switching hub to T_1, and the i^{th} edge e_i $(i > 1)$ runs from T_{i-1} to T_i. Now we observe that while all the towns benefit from e_1, only the last town benefits from e_n. So they suggest

> **Proposal C.** Divide the cost separately for each edge e_i. The cost of e_i should be shared equally by the towns $T_i, T_{i+1}, \ldots, T_n$, since these are the towns "downstream" of e_i.

So now the towns have many different options; which is the fairest? To resolve this they turn to the work of Lloyd Shapley, one of the most famous mathematical economists of the

20th century; he proposed what is now called the *Shapley value* as a general mechanism for sharing costs or benefits among several parties. It can be viewed as determining the "marginal contribution" of each party, *assuming the parties arrive in a random order.*

Here's how it would work concretely in our setting. Consider an ordering $\mathcal{O}$ of the towns, and suppose that the towns "arrive" in this order. The *marginal cost of town T_i in order $\mathcal{O}$* is determined as follows. If T_i is first in the order $\mathcal{O}$, then T_i pays ki, the cost of running the cable all the way from the switching hub to T_i. Otherwise, look at the set of towns that come before T_i in the order $\mathcal{O}$, and let T_j be the farthest among these towns from the switching hub. When T_i arrives, we assume the cable already reaches out to T_j but no farther. So if $j > i$ (T_j is farther out than T_i), then the marginal cost of T_i is 0, since the cable already runs past T_i on its way out to T_j. On the other hand, if $j < i$, then the marginal cost of T_i is $k(i - j)$: the cost of extending the cable from T_j out to T_i.

(For example, suppose $n = 3$ and the towns arrive in the order T_1, T_3, T_2. First T_1 pays k when it arrives. Then, when T_3 arrives, it only has to pay $2k$ to extend the cable from T_1. Finally, when T_2 arrives, it doesn't have to pay anything since the cable already runs past it out to T_3.)

Now, let X_i be the random variable equal to the marginal cost of town T_i when the order $\mathcal{O}$ is selected uniformly at random from all permutations of the towns. Under the rules of the Shapley value, the amount that T_i should contribute to the overall cost of the cable is the expected value of X_i.

The question is: Which of the above three proposals, if any, gives the same division of costs as the Shapley value cost-sharing mechanism? Give a proof for your answer.

---

(5) Suppose you're designing strategies for selling items on the popular auction Web site `280bay.com`. Unlike other auction sites, `280bay.com` uses a *one-pass auction*, in which each bid must be immediately (and irrevocably) accepted or refused. Specifically,

- First, a seller puts up an item for sale.
- Then buyers appear in sequence.
- When buyer i appears, he or she makes a bid $b_i > 0$.
- The seller must decide immediately whether to accept the bid or not. If the seller accepts the bid, the item is sold and all future buyers are turned away. If the seller rejects the bid, buyer i departs and the bid is withdrawn; and only then does the seller see any future buyers.

Suppose an item is offered for sale, and there are n buyers, each with a distinct bid. Suppose further that the buyers appear in a random order, and that the seller knows the number n of buyers. We'd like to design a *strategy* whereby the seller has a reasonable chance of accepting the highest of the n bids. By a "strategy," we mean a rule by which the seller decides whether to accept each presented bid, based only on the value of n and the sequence of bids seen so far.

For example, the seller could always accept the first bid presented. This results in the seller accepting the highest of the n bids with probability only $1/n$, since it requires the highest bid to be the first one presented.

Give a strategy under which the seller accepts the highest of the n bids with probability at least $1/4$, regardless of the value of n. (For simplicity, you are allowed to assume that n is an even number.) Prove that your strategy achieves this probabilistic guarantee.