In-Class Problems Week 14, Fri.

Problem 1. A prison warden has two holding cells, Cell 1 and Cell 2, for his prisoners. Unfortunately, many pairs of these prisoners are *incompatible*, and there will be some trouble if an incompatible pair are in the same cell. The warden would like to minimize trouble by not having too many incompatible pairs in the same cell. Unfortunately, the warden has no idea how to split the up prisoners, and so he decides to go with a random assignment: he will assign prisoners to one cell or the other by successive (independent) flips of a fair coin.

For any two incompatible prisoners, a, b, let $T_{a,b}$ be 1 if a and b are placed in the same cell, and 0 otherwise.

(a) What is the expected value of $T_{a,b}$?

Suppose there are n incompatible sets of prisoners, a, b. The total trouble, T, of an assignment of prisoners to cells is the sum of $T_{a,b}$ where the sum is over the n sets of incompatible prisoners $a \neq b$. So T is the total number of incompatible sets of two prisoners that are in the same cell. The warden hopes to minimize the total trouble, T.

(b) What is the expected value of T?

(c) Explain why there must be a split of the prisoners between cells that separates at least half the incompatible pairs.

(d) Suppose a, b, c are different prisoners, where a and b are incompatible, and a and c are also incompatible. Explain why $T_{a,b}$ is independent of $T_{a,c}$. Conclude that set of all the $T_{a,b}$'s is pairwise independent.

(e) Are the $T_{a,b}$'s mutually independent?

(f) What is the variance of T?

(g) Suppose among 1000 prisoners, about a fifth of the pairs, say 100,000 pairs, turn out to be incompatible. Show that there is at most a 10% chance that the warden's random assignment differs by more than 1% from the expected number of incompatible pairs in the same cell. *Hint:* Chebyshev.

Problem 2. Now we look at the situation in the previous problem in more detail. Suppose there are levels of conflict between incompatible prisoners, where the conflict level of two prisoners who may hate each other's guts is 1 if they wouldn't actually touch each other, 2 if they might hurt each other but wouldn't cause a trip to the hospital, and 3 if having them in the same cell would be *really bad*. Suppose we model the situation by assuming that a random conflict level $w_{a,b}$ equal to 1,2, or 3 is assigned to every two incompatible prisoners, a, b, uniformly and independently of all other conflict levels.

So $T_{a,b}w_{a,b}$ is 0 if a, b are placed in different cells and is $w_{a,b}$ otherwise. Define the *total conflict* to be

$$C ::= \sum_{a,b \text{ incompatible}} T_{a,b}w_{a,b},$$

that is, the sum of the levels of conflicting pairs in which the members are assigned to the same cell. We would like the total conflict to be small.

(a) What is the expected value of $w_{a,b}$ and $T_{a,b}w_{a.b}$?

(b) What is the variance of $w_{a,b}$ and $T_{a,b}w_{a,b}$?

(c) What is the expected value of C?

(d) Are the $T_{a,b}w_{a,b}$'s pairwise independent? ...mutually independent?

(e) What is the variance of C?

(f) What does Chebyshev's inequality give for a bound on $\Pr\{|C - \mathrm{E}[C]|\} > n/4$?

(g) Suppose someone complains about our modeling the situation as choosing a random conflict level of 1,2, or 3, and agrees only that conflict levels range between 1 and 3. So then the w's for different incompatible pairs may have different distributions, but we still assume they are independent. Could we still use Chebyshev's inequality to say something about the probability of deviating from the mean? *Hint:* : What is the maximum possible variance for a random variable with values between 0 and 3?

Problem 3. A recent Gallup poll found that 35% of the adult population of the United States believes that the theory of evolution is "well-supported by the evidence". Gallup polled 1928 people and claims a margin of error of 3 percentage points.

Let's check Gallup's claim. Suppose that there are m adult Americans, of whom pm believe in evolution; this means that $(1-p)m$ Americans do not believe in evolution. Gallup polls 1928 Americans selected uniformly and independently at random. Of these, 675 believe in evolution, leading to Gallup's estimate that the fraction of Americans who believe in evolution is within 0.03 of $675/1928 \approx 0.350$.

(a) Explain how to use the Binomial Sampling Theorem (available in the Appendix) to determine the confidence level with which Gallup can make his claim. You do not actually have to do the calculation, but are welcome to if you have the means.

(b) If we accept all of Gallup's polling data and calculations, can we conclude that there is a high probability that the number of adult Americans who believe in evolution is 35 ± 3 percent?

(c) Explaining Sampling to a Jury The calculation above revealed that, based on a poll of 1928 people, we can be highly confident that the per cent of people in the U.S. who believe in evolution is $35\% \pm 3\%$. Note that the actual population of the U.S. was never considered, because *it did not matter*!

Suppose you were going to serve as an expert witness in a trial. How would you explain to a jury why the number of people necessary to poll *does not depend on the population size*? (Begin by explaining why it is reasonable to model polling as independent coin tosses. Remember that juries do not understand algebra or equations; you might be ok using a little arithmetic.)

1 Appendix

1.1 Binomial Sampling

Theorem. *Let $K_1, K_2, \ldots,$ be a sequence of mutually independent 0-1-valued random variables with the same expectation, p, and let*

$$S_n ::= \sum_{i=1}^{n} K_i.$$

Then, for $1/2 > \epsilon > 0$,

$$\Pr\left\{\left|\frac{S_n}{n} - p\right| \geq \epsilon\right\} \leq \frac{1+2\epsilon}{2\epsilon} \cdot \frac{2^{-n(1-H((1/2)-\epsilon))}}{\sqrt{2\pi(1/4-\epsilon^2)n}}. \tag{1}$$

1.2 Scheme Code for Sampling Bounds

```
(define (pr n eps)
  (* (/ (+ 1 (* 2 eps)) (* 2 eps)
     (sqrt (* 2 pi (- 1/4 (* eps eps))))
     (expt 2 (* n (- 1 (h (- 1/2 eps)))))
     (sqrt n))))

(define (h a)
  (cond ((>= 0 a) 1)
        ((>= a 1) 1)
        (else (- (+ (* a (log2 a)) (* (- 1 a) (log2 (- 1 a))))))))

(define (log2 a) (/ (log a) (log 2)))

(define pi (* 4 (atan 1)))

(pr 1928 0.03)
;Value: 9.982587419699058e-3
```

1.3 Chebyshev's Theorem

Theorem (Chebyshev). *Let R be a random variable, and let x be a positive real number. Then*

$$\Pr\{|R - \mathrm{E}[R]| \geq x\} \leq \frac{\mathrm{Var}[R]}{x^2}. \tag{2}$$