Section 13

Wednesday, August 7

CS 70: Discrete Mathematics and Probability Theory, Summer 2013

1. **(How to Boost Your Confidence)** You're a doctor and you want to measure a patient's blood pressure, but your machine only gives an estimate with relative error ϵ and confidence parameter 1/4 ("75% confident"). In other words, the output is a random variable X such that $\Pr\left[|X-p| \leq \epsilon p\right] \geq 3/4$ (where p is the true blood pressure), and with the remaining probability the machine malfunctions and may output a very inaccurate number. You'd like to boost your confidence to $1-\delta$ (for some tiny $\delta > 0$) by designing a random variable Y for which $\Pr\left[|Y-p| \leq \epsilon p\right] \geq 1-\delta$.

   1a. A natural idea is to take a bunch of independent samples $X_1, \ldots, X_n$ and average them: $Y = \frac{1}{n}(X_1 + \cdots + X_n)$. Explain why this is not guaranteed to work.

   The following strategy does work: Take Y to be the median of $X_1, \ldots, X_n$. In other words, sort the estimates X_i and take the one in the middle. (You may assume n is odd so there is no tie.) Let's prove that if $n \geq 3/\delta$ then the confidence is boosted to at least $1-\delta$.

   1b. Let's say a number is "good" if it's within ϵp of p, and "bad" if it's outside this range. Show that if the value of Y is bad then at least half of the estimates X_i must be bad. In other words, define Z_i to be the indicator that X_i is bad, and define $Z = \sum_{i=1}^{n} Z_i$, and show that if Y is bad then $Z \geq n/2$.

   1c. Find the distribution, expectation, and variance of Z. You may assume $\Pr\left[|X_i - p| \leq \epsilon p\right]$ is exactly 3/4 if you like.

   1d. Use Chebyshev's inequality to show that $\Pr[Y \text{ is bad}] \leq \delta$, assuming $n \geq 3/\delta$.

2. **(Marginal Distributions)** You have random variables X and Y with joint distribution: $\Pr[X=1, Y=1] = \frac{1}{3}$, $\Pr[X=1, Y=2] = \frac{1}{4}$, $\Pr[X=2, Y=1] = \frac{1}{4}$, $\Pr[X=2, Y=2] = \frac{1}{6}$. What are the marginal distributions of X and Y?

3. **(Geometric Inference)** You want to buy a printer, and you are considering three models. Each time you print something, the printer will get jammed independently with some probability. One of the models has probability $p_1 = 0.05$ of jamming, one has probability $p_2 = 0.10$, and the other has probability $p_3 = 0.15$, but you do not know which is which so you buy a uniformly random printer. Your new printer works fine for a while but then gets jammed the k^{th} time you try to print something. What is the posterior distribution on which printer you bought? In other words, for $i = 1, 2, 3$ what is the probability you got the p_i printer given that the k^{th} time is the first jam?

4. **(Poisson Inference)** If raisins are dropped randomly when making bread, then the number of raisins in a bread loaf has a Poisson distribution. Bakery A is known to have parameter λ_A, and bakery B is known to have parameter λ_B.

   4a. You have a loaf that came from bakery A with probability $\frac{1}{2}$ or bakery B with probability $\frac{1}{2}$. Being bored at breakfast on a Saturday ~~morning~~ afternoon, you count the raisins and find there are j of them. What is the posterior probability the loaf came from bakery A?

   4b. After eating that loaf, you look at another loaf from the same bakery (though you still don't know for sure which bakery they came from). The new loaf has k raisins. What is the new posterior probability that the loaves came from bakery A?