In-Class Problems Week 14, Wed.

Problem 1. Among the 73 students who reported their birthdays to us this term, we found there were 5 pairs with the same birthday (there were no triples).

Use the Chebyshev Bound to prove that in a class of 73, the probability that there are between 2 and 12 pairs of students with the same birthday is at least 3/4. (Assume all birthdays are equally likely and a year has 365 days.)

Hint: Let $S_{i,j}$ be the indicator variable for the ith and jth students having the same birthday. Let $M = \sum_{1 \leq i < j \leq 73} S_{i,j}$ be the number of pairs with the same birthday. Calculate $\mathrm{E}[M]$ and $\mathrm{Var}[M]$.

Problem 2. For any random variable, R, with mean, μ, and standard deviation, σ, the Chebyshev Bound says that for any real number $x > 0$,

$$\Pr\{|R - \mu| \geq x\} \leq \left(\frac{\sigma}{x}\right)^2.$$

Show that for any real number, μ, and real numbers $x \geq \sigma > 0$, there is an R for which the Chebyshev Bound is tight, that is,

$$\Pr\{|R| \geq x\} = \left(\frac{\sigma}{x}\right)^2. \tag{1}$$

Hint: First assume $\mu = 0$ and let R only take values 0, $-x$, and x.

Problem 3. The proof of the Pairwise Independent Sampling Theorem you just saw in lecture (it's also in Notes 14) was given for a sequence $R_1, R_2, \ldots$ of pairwise independent random variables with the same mean and variance. The proof is repeated in the Appendix.

We can generalize the Theorem to sequences of pairwise independent random variables, possibly with *different* distributions, as long as all their variances are bounded by some constant.

Theorem (Generalized Pairwise Independent Sampling)**.** *Let $X_1, X_2, \ldots$ be a sequence of pairwise independent random variables such that* $\text{Var}[X_i] \leq b$ *for some $b \geq 0$ and all $i \geq 1$. Let*

$$A_n ::= \frac{X_1 + X_2 + \cdots + X_n}{n},$$
$$\mu_n ::= \text{E}[A_n].$$

Then for every $\epsilon > 0$,

$$\Pr\{|A_n - \mu_n| > \epsilon\} \leq \frac{b}{\epsilon^2} \cdot \frac{1}{n}. \tag{2}$$

(a) Prove the Generalized Pairwise Independent Sampling Theorem.

(b) Conclude that the following holds:

Corollary (Generalized Weak Law of Large Numbers)**.** *For every $\epsilon > 0$,*

$$\lim_{n \to \infty} \Pr\{|A_n - \mu_n| \leq \epsilon\} = 1.$$

Problem 4. (a) A computer program crashes at the end of each hour of use with probability p, if it has not crashed already. We know that the expected number of hours, $\text{E}[H]$, until the program crashes is $1/p$. What is the variance of the number of hours until the program crashes? *Hint:* From Notes 11:

$$\sum_{1}^{\infty} k^2 x^k = \frac{x(1+x)}{(1-x)^3}$$

(b) What is the Chebyshev bound on

$$\Pr\{|H - (1/p)| > x/p\}$$

where $x > 0$?

(c) Conclude from part (b) that for $a \geq 2$,

$$\Pr\{H > a/p\} \leq \frac{q}{(a-1)^2}$$

Hint: Check that $|H - (1/p)| > (a-1)/p$ iff $H > a/p$.

(d) What actually is

$$\Pr\{H > a/p\}?$$

Conclude that for any fixed $p > 0$, the probability that $H > a/p$ is an asymptotically smaller function of a than the Chebyshev bound of part (c).

Appendix

The *variance*, $\text{Var}[R]$, of a random variable, R, is:

$$\text{Var}[R] ::= \text{E}\left[(R - \text{E}[R])^2\right].$$

Variance can also be equivalently defined as:

$$\text{Var}[R] ::= \text{E}\left[R^2\right] - \text{E}^2[R].$$

Lemma. *For $a, b \in \mathbb{R}$,*

$$\text{Var}[aR + b] = a^2\,\text{Var}[R] \tag{3}$$

Theorem. *If $R_1, R_2, \ldots, R_n$ are* pairwise *independent random variables, then*

$$\text{Var}[R_1 + R_2 + \cdots + R_n] = \text{Var}[R_1] + \text{Var}[R_2] + \cdots + \text{Var}[R_n].$$

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

$$\Pr\{|R - \text{E}[R]| \geq x\} \leq \frac{\text{Var}[R]}{x^2}. \tag{4}$$

Theorem (Pairwise Independent Sampling)**.** *Let*

$$A_n ::= \frac{\sum_{i=1}^{n} R_i}{n}$$

where $R_1, \ldots, R_n$ are pairwise independent random variables with the same mean, μ, and deviation, σ. Then

$$\Pr\{|A_n - \mu| > x\} \leq \left(\frac{\sigma}{x}\right)^2 \cdot \frac{1}{n}. \tag{5}$$

Proof. By linearity of expectation,

$$\text{E}[A_n] = \frac{\text{E}\left[\sum_{i=1}^{n} R_i\right]}{n} = \frac{\sum_{i=1}^{n} \text{E}[R_i]}{n} = \frac{n\mu}{n} = \mu.$$

Since the R_i's are pairwise independent, their variances will also add, so

$$\begin{aligned}
\text{Var}[A_n] &= \left(\frac{1}{n}\right)^2 \text{Var}\left[\sum_{i=1}^{n} R_i\right] && \text{(by (3))} \\
&= \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \text{Var}[R_i] && \text{(additivity)} \\
&= \left(\frac{1}{n}\right)^2 n\sigma^2 \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

Now letting R be A_n in Chebyshev's Bound (4) yields (5), as required.

□