**Staff Solutions to In-Class Problems Week 14, Wed.**

**STAFF NOTE:** Chebyshev Ch.19.3, Pairwise Independent Sampling Ch.19.4

**Problem 1.**
Let $K_n$ be the complete graph with $n$ vertices. Each of the edges of the graph will be randomly assigned one of the colors red, green, or blue. The assignments of colors to edges are mutually independent, and the probability of an edge being assigned red is $r$, blue is $b$, and green is $g$ (so $r + b + g = 1$).

A set of three vertices in the graph is called a *triangle*. A triangle is *monochromatic* if the three edges connecting the vertices are all the same color.

(a) Let $m$ be the probability that any given triangle, $T$, is monochromatic. Write a simple formula for $m$ in terms of $r, b,$ and $g$.

**Solution.** $m = r^3 + b^3 + g^3$

(b) Let $I_T$ be the indicator variable for whether $T$ is monochromatic. Write simple formulas in terms of $m, r, b,$ and $g$ for $\mathbb{E}[I_T]$ and $\text{Var}[I_T]$.

**Solution.**
\[
\mathbb{E}[I_T] = m, \\
\text{Var}[I_T] = m(1 - m).
\]

(c) What is the probability that $T$ and $U$ are both monochromatic if they do not share an edge?...if they do share an edge?

**Solution.**
\[
m^2 \text{if they do not share an edge} \\
r^5 + b^5 + g^5 \text{if they do share an edge}.
\]

If $T$ and $U$ do not share an edge, then the three edges of $T$ match with probability $m$, and independently, the three edges of $U$ match with probability $m$, so both match with probability $m^2$. If they do share an edge, the five edges among them must all match one of colors $r, g, b$.

**Now assume** $r = b = g = \frac{1}{3}$.

(d) Show that $I_T$ and $I_U$ are independent random variables.
Solution. Since $I_T$ and $I_U$ are indicators for events, it suffices to verify that

$$\Pr[I_T = 1] \cdot \Pr[I_U = 1] = \Pr[I_T \cdot I_U = 1].$$

(1)

In the case that $T$ and $U$ do not share an edge, (1) follows immediately from parts (a) and (c). If they do share an edge, this follows because

$$m^2 = \left(3(1/3)^3\right)^2 = (1/3)^4 = 3(1/3)^5 = r^5 + b^5 + g^5.$$

(e) Let $M$ be the number of monochromatic triangles. Write simple formulas in terms of $n$ and $m$ for $\operatorname{Ex}[M]$ and $\operatorname{Var}[M]$.

Solution.

$$\operatorname{Ex}[M] = m \cdot (\# \text{triangles})$$

$$= m \left( \binom{n}{3} \right),$$

(2)

$$\operatorname{Var}[M] = \operatorname{Var}[I_T] \cdot (\# \text{triangles})$$

$$= m(1-m) \left( \binom{n}{3} \right) = (1-m) \operatorname{Ex}[M].$$

(3)

(f) Let $\mu := \operatorname{Ex}[M]$. Use Chebyshev’s Bound to prove that

$$\Pr[|M - \mu| > \sqrt{\mu \log \mu}] \leq \frac{1}{\log \mu}.$$

STAFF NOTE: Have students work out rough numbers for $n = 10^3$. THESE NUMBERS NEED TO BE CHECKED: $\mu \approx 10^9$, $\log \mu \approx 30$, so there is only a 3% chance that $M$ will differ from its expected value by more than 15%.

Solution. According to Chebyshev’s Bound:

$$\Pr[|M - \mu| > c\sigma] \leq \frac{1}{c^2}$$

So

$$\Pr[|M - \mu| > \sqrt{\mu \log \mu}] = \Pr[|M - \mu| > \sqrt{\log \mu} \sqrt{\mu}]$$

$$\leq \Pr[|M - \mu| > \sqrt{\log \mu} \sqrt{(1-m)\mu}]$$

$$= \Pr[|M - \mu| > \sqrt{\log \mu} \sigma]$$

$$\leq \frac{1}{\log \mu}$$

(by Chebyshev)
(g) Conclude that
\[
\lim_{n \to \infty} \Pr[|M - \mu| > \sqrt{\mu \log \mu}] = 0
\]

Solution. According to (2),
\[
\mu = m \left( \frac{n}{3} \right) = \Theta(n^3)
\]
so
\[
\frac{1}{\log \mu} = \Theta \left( \frac{1}{\log n^3} \right) = \Theta \left( \frac{1}{\log n} \right).
\]
Since \(1/(\log n) \to 0\) as \(n \to \infty\), the \(O()\) bound on the probability goes to 0 which means an upper bound on the limit is 0. Since the probability is nonnegative, the limit must be exactly 0.

Problem 2.
For any random variable, \(R\), with mean, \(\mu\), and standard deviation, \(\sigma\), 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, \(\mu\), and real numbers \(x \geq \sigma > 0\), there is an \(R\) for which the Chebyshev bound is tight, that is,
\[
\Pr[|R - \mu| \geq x] = \left( \frac{\sigma}{x} \right)^2.
\]
(4)

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

Solution. From the hint, we aim to find an \(R\) with \(\text{Ex}[R] = 0\) and \(\text{Var}[R] = \sigma^2\) that satisfies equation (4).

Using the further hint that \(R\) takes only values 0, \(-x\), \(x\), we have
\[
0 = \text{Ex}[R] = x \Pr[R = x] - x \Pr[R = -x] = x (\Pr[R = x] - \Pr[R = -x])
\]
since \(x > 0\). Also,
\[
\sigma^2 = \text{Ex}[R^2] = x^2 \Pr[R = -x] + x^2 \Pr[R = x] = 2x^2 \Pr[R = x].
\]
so
\[
\Pr[R = x] = \frac{\sigma^2}{2x^2}.
\]
This implies
\[
\Pr[R = 0] = 1 - 2 \Pr[R = x] = 1 - \left( \frac{\sigma}{x} \right)^2,
\]
which completely determines the distribution of \(R\). Moreover,
\[
\Pr[|R| \geq x] = \Pr[R = -x] + \Pr[R = x] = 2 \Pr[R = x] = \left( \frac{\sigma}{x} \right)^2
\]
which confirms (4).

Finally, given \(\mu, x,\) and \(\sigma\), define \(R' := R + \mu\). Since \(\text{Var}[R'] = \text{Var}[R]\), the random variable \(R'\) will have mean \(\mu\) and standard deviation \(\sigma\) and (4) will hold with \(R'\) in place of \(R\).
Problem 3.
The proof of the Pairwise Independent Sampling Theorem 19.4.1 was given for a sequence $R_1, R_2, \ldots$ of pairwise independent random variables with the same mean and variance.

The theorem generalizes straightforwardly 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

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

Then for every $\epsilon > 0$,

$$\Pr[|A_n - \mu_n| > \epsilon] \leq \frac{b}{\epsilon^2} \cdot \frac{1}{n}. \quad (6)$$

(a) Prove the Generalized Pairwise Independent Sampling Theorem.

**Solution.** Essentially identical to the proof of Theorem 19.4.1 in the text, except that $S_n/n$ gets replaced by $A_n$, the constant $b$ replaces $\text{Var}[G_i]$, and the equality before the first occurrence of $b$ gets replaced by an inequality ($\leq$).

**STAFF NOTE:** It is OK for students to refer to the proof in the text and simply comment on where changes must be made. Here’s the revised proof:

**Proof.** We first observe that

$$\text{Var}[A_n] \leq \frac{b}{n}. \quad (7)$$

because

$$\text{Var}[A_n] = \frac{1}{n^2} \text{Var}\left[\sum_{i=1}^{n} X_i\right] \quad \text{(def of } A_n \text{ & Square Multiple Rule, Theorem 19.3.4)}$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}[X_i] \quad \text{(pairwise independent additivity)}$$

$$\leq \frac{1}{n^2} \cdot nb \quad \text{(since } \text{Var}[X_i] \leq b)$$

$$= \frac{b}{n}. \quad \text{(since } \text{Var}[X_i] \leq b)$$

This is enough to apply Chebyshev’s Theorem and conclude:

$$\Pr\left[|A_n - \mu_n| \geq \epsilon\right] \leq \frac{\text{Var}[A_n]}{\epsilon^2}. \quad \text{(Chebyshev’s bound)}$$

$$\leq \frac{b/n}{\epsilon^2} \quad \text{(by } (7))$$

$$= \frac{b \cdot 1}{\epsilon^2 \cdot n}. \quad \text{(Chebyshev’s bound)}$$
(b) Conclude that the following holds:

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

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

**Solution.**

$$
\Pr(|A_n - \mu_n| \leq \epsilon) = 1 - \Pr(|A_n - \mu_n| > \epsilon)
\geq 1 - \frac{b}{\epsilon^2} \cdot \frac{1}{n}
$$
(by (6)),

and for any fixed $\epsilon$, this last term approaches 1 as $n$ approaches infinity.

**Problem 4.** **STAFF NOTE:** This problem is a direct application of Chebyshev’s bound using pairwise independent additivity of variance (not sampling). It comes pretty directly from Section 19.4.2. It’s OK for students to look that up.

Let $B_1, B_2, \ldots, B_n$ be mutually independent random variables with a uniform distribution on the integer interval $[1, d]$. Let $D$ equal to the number of events $[B_i = B_j]$ that happen where $i \neq j$. It was observed in Section 16.4 (and proved in Problem 18.2) that $\Pr[B_i = B_j] = 1/d$ for $i \neq j$ and that the events $[B_i = B_j]$ are pairwise independent.

Let $E_{i,j}$ be the indicator variable for the event $[B_i = B_j]$.

(a) What are $\Ex[E_{i,j}]$ and $\Var[E_{i,j}]$ for $i \neq j$?

**Solution.** For $i \neq j$,

$$
\Ex[E_{i,j}] = \frac{1}{d} \quad \text{(8)}
$$

$$
\Var[E_{i,j}] = \frac{1}{d} \left(1 - \frac{1}{d}\right) \quad \text{(9)}
$$

Equation 8 follows from the fact that $\Pr[B_i = B_j] = 1/d$ and that $\Ex[E_{i,j}] = \Pr[B_i = B_j]$ by Lemma 18.4.2.

Equation 9 follows from (8) and the formula for the variance of an indicator variable given in Lemma 19.3.2.

(b) What are $\Ex[D]$ and $\Var[D]$?

**STAFF NOTE:** Hint: $D := \sum_{1 \leq i < j \leq n} E_{i,j}$

**Solution.** By definition

$$
D = \sum_{1 \leq i < j \leq n} E_{i,j},
$$

so by linearity of expectation

$$
\Ex[D] = \Ex\left[\sum_{1 \leq i < j \leq n} E_{i,j}\right] = \sum_{1 \leq i < j \leq n} \Ex[E_{i,j}] = \left(\frac{n}{2}\right) \cdot \frac{1}{d}.
$$
\[
\text{Var}[D] = \text{Var} \left[ \sum_{1 \leq i < j \leq n} E_{i,j} \right]
\]
\[
= \sum_{1 \leq i < j \leq n} \text{Var}[E_{i,j}] \quad \text{(pairwise independent additivity)}
\]
\[
= \binom{n}{2} \cdot \frac{1}{d} \left(1 - \frac{1}{d}\right) \quad \text{(by part (a)).}
\]

(e) In a 6.01 class of 500 students, the youngest student was born 15 years ago and the oldest 35 years ago. Let \( D \) be the number of students in the class who were born on exactly the same date. What is the probability that \( 4 \leq D \leq 32 \)? (For simplicity, assume that the distribution of birthdays is uniform over the 7305 days in the two decade interval from 35 years ago to 15 years ago.)

**Solution.** For a class of \( n = 500 \) students with \( d = 7305 \) possible days, we have from part (b)

\[
\text{Ex}[D] = 500 \cdot 499/2 \cdot \frac{1}{7305} \approx 17.1
\]

and

\[
\text{Var}[D] = 500 \cdot 499/2 \cdot \frac{1}{7305} \cdot \left(1 - \frac{1}{7305}\right) < 17.1.
\]

So by Chebyshev’s Theorem

\[
\text{Pr}[|S - 17.1| \geq 14] < \frac{17.1}{14^2} < 0.09.
\]

We conclude that there is a better than 91% chance that there will be between 4 and 32 pairs of students born on the same date.