In-Class Problems Week 15, Mon.

Problem 1. The Pairwise Independent Sampling Theorem generalizes easily to sequences of pairwise independent random variables, possibly with *different* means and variances, as long as 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*

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

Then for every $\epsilon > 0$,

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

(a) Prove the Generalized Pairwise Independent Sampling Theorem. *Hint:* The proof of the Pairwise Independent Sampling Theorem from the Notes is repeated in the Appendix.

(b) Conclude

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

$$\lim_{n\to\infty} \Pr\{|A_n - \mu_n| > \epsilon\} = 0.$$

Problem 2. Write out a proof that

$$\text{Var}[aR] = a^2 \, \text{Var}[R].$$

Problem 3. Finish discussing the "Explain sampling to a jury question" from last Friday.

Copyright © 2005, Prof. Albert R. Meyer and Prof. Ronitt Rubinfeld.

1 Appendix

1.1 Chebyshev's Theorem

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

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

1.2 Pairwise Independent Sampling

Theorem (Pairwise Independent Linearity of Variance). *If $R_1, R_2, \ldots, R_n$ are* pairwise *independent random variables, then*

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

Theorem (Pairwise Independent Sampling). *Let*

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

where $G_1, \ldots, G_n$ are pairwise independent random variables with the same mean, μ, and deviation, σ. Then

$$\Pr\{|A_n - \mu| > x\} \leq \left(\frac{\sigma}{x}\right)^2 \cdot \frac{1}{n}. \tag{3}$$

Proof. By linearity of expectation,

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

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

$$\begin{aligned}
\mathrm{Var}[A_n] &= \left(\frac{1}{n}\right)^2 \mathrm{Var}\left[\sum_{i=1}^n G_i\right] && (\mathrm{Var}[aR] = a^2\,\mathrm{Var}[R]) \\
&= \left(\frac{1}{n}\right)^2 \sum_{i=1}^n \mathrm{Var}[G_i] && (\text{linearity of variance}) \\
&= \left(\frac{1}{n}\right)^2 n\sigma^2 \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

Now letting R be A_n in Chebyshev's Bound (2) yields (3), as required.

□