Mathematics for Computer Science

MIT 6.042J/18.062J

Independent Sampling Theorem

Weak Law of Large Numbers

A_n ::= avg of n indep RV's with mean μ

Theorem: For all $\delta > 0$

$$\lim_{n\to\infty} \Pr[\,|A_n - \mu| > \delta\,] = 0$$

Proof:

Weak Law of Large Numbers

A_n ::= avg of n indep RV's with mean μ, var σ^2

Theorem: For all $\delta > 0$

$$\lim_{n\to\infty} \Pr[\,|A_n - \mu| > \delta\,] = 0$$

Proof:

Repeated Trials

$$E[A_n] ::= E\left[\frac{R_1 + R_2 + \cdots + R_n}{n}\right]$$

$$= \frac{E[R_1] + E[R_2] + \cdots + E[R_n]}{n}$$

$$= \frac{n\mu}{n} = \mu$$

Weak Law of Large Numbers

So by Chebyshev

$$\Pr[|A_n - \mu| > \delta] \le \frac{\mathrm{Var}[A_n]}{\delta^2}$$

need only show

$\mathrm{Var}[A_n] \to 0$ as $n \to \infty$

$\mathrm{Var}[A_n]$

$$\mathrm{Var}[A_n] = \mathrm{Var}\left[\frac{R_1 + R_2 + \cdots + R_n}{n}\right]$$

$$= \frac{\mathrm{Var}[R_1] + \mathrm{Var}[R_2] + \cdots + \mathrm{Var}[R_n]}{n^2}$$

$$\text{QED} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \to 0$$

Analysis of the Proof

proof only used that $R_1, \ldots, R_n$ have

- same mean
- same variance
- & variances add

— which follows from pairwise independence

Pairwise Independent Sampling

Theorem:

Let $R_1, \ldots, R_n$ be pairwise independent random vars with the same finite mean μ and variance σ^2. Let $A_n ::= (R_1 + R_2 + \cdots + R_n)/n$. Then

$$\Pr[|A_n - \mu| > \delta] \le \frac{1}{n}\left(\frac{\sigma}{\delta}\right)^2$$

Pairwise Independent Sampling

The punchline:

we now know how big a sample is needed to estimate the mean of any* random variable within any* desired tolerance with any* desired probability

*variance $< \infty$, tolerance > 0, probability < 1