6.262: Discrete Stochastic Processes
Lecture 2: Laws of large numbers

Probability models are useful for situations that

- are repeatable
- have a known set of possible outcomes
- have a random outcome for same input

For any model, a model for a sequence or an n-tuple of IID repetitions is well defined.

Relative frequencies and sample averages (in the extended model) 'become deterministic' and can be compared with real-world values.

The laws of large numbers specify what 'become deterministic' means.

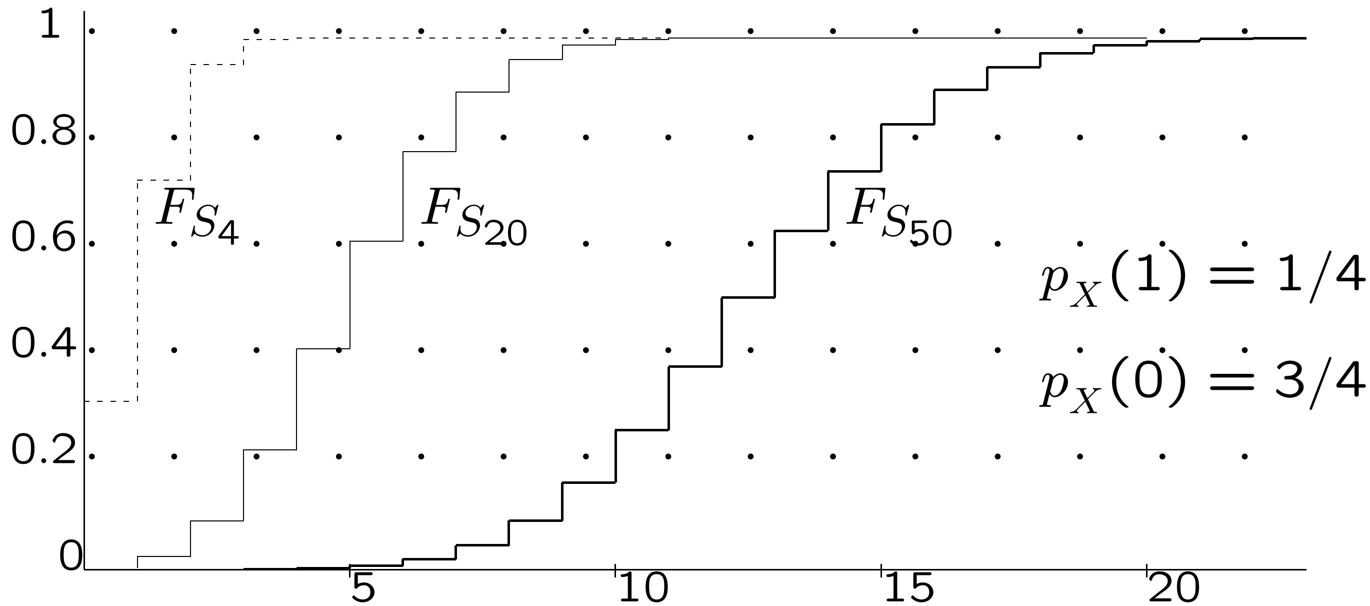
They only operate within the extended model, but provide our only truly experimental way to test the model.

Theory provides many many consistency checks and ways to avoid constant experimentation.

Understanding the determinism in large numbers is where most of the value of probability comes from.

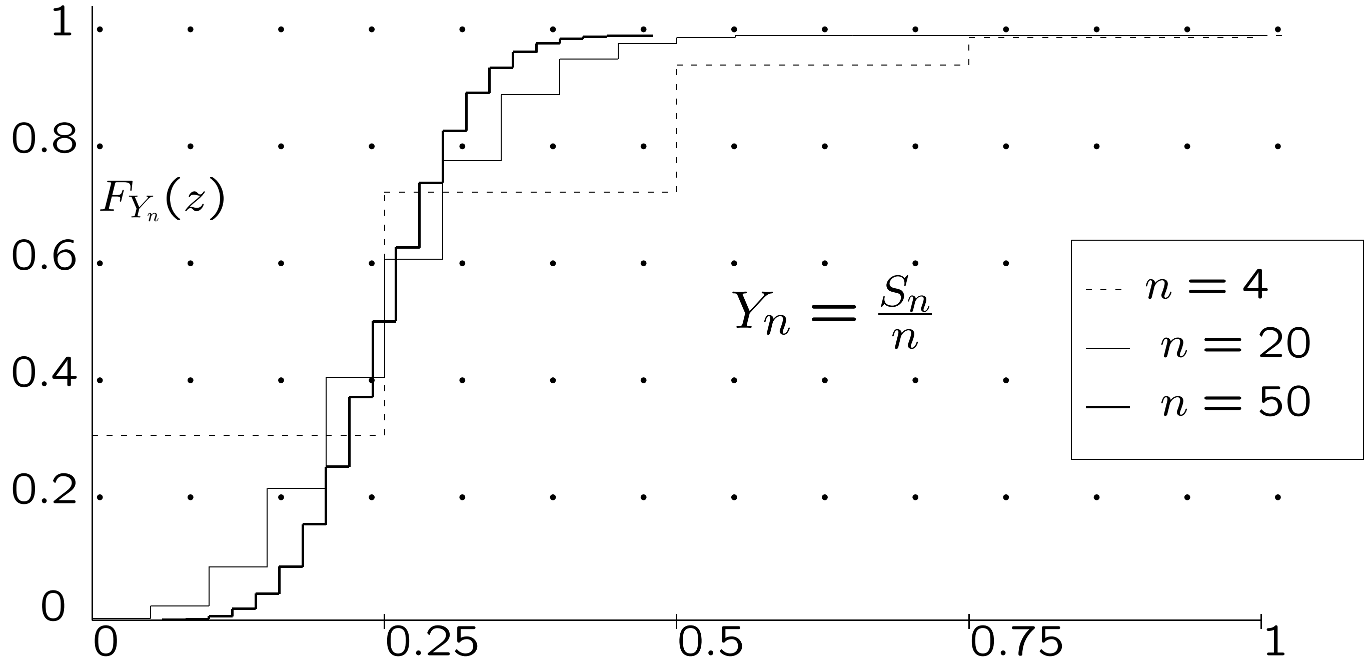
The weak law of large numbers

Let $X_1, X_2, \ldots, X_n$ **be IID rv's with mean** $\overline{X}$**, variance** σ^2**. Let** $S_n = X_1 + \cdots + X_n$**. Then** $\sigma^2_{S_n} = n\sigma^2$**.**



The center of the distribution varies with n **and the spread varies with** $\sqrt{n}$**.**

The sample average is S_n/n. It is a rv of mean $\overline{X}$ and variance σ^2/n.



The center of the distribution is $\overline{X}$ and the spread decreases with $1/\sqrt{n}$.

$$\mathbf{VAR}\left(\frac{S_n}{n}\right) = \mathbf{E}\left[\left(\frac{S_n}{n} - \overline{X}\right)^2\right] = \frac{\sigma^2}{n}. \tag{1}$$

$$\lim_{n\to\infty} \mathbf{E}\left[\left(\frac{S_n}{n} - \overline{X}\right)^2\right] = 0. \tag{2}$$

The rv S_n/n is said to converge to $\overline{X}$ in mean square. Note that (1) says more than (2), since it says the convergence is as $1/n$. But (2) establishes a standard form of convergence.

Neither (1) nor (2) directly provide any statement about the accuracy and the confidence level of S_n/n as an estimate of $\overline{X}$.

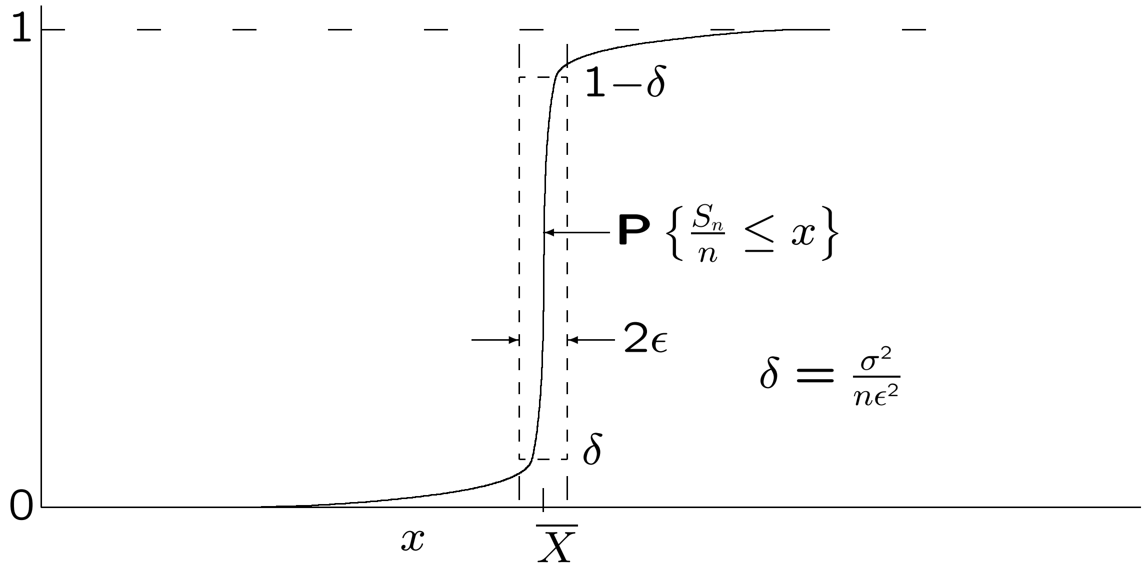
From the Chebyshev inequality,

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - \overline{X}\right| \geq \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2} \qquad \textbf{for any } \epsilon > 0 \qquad (3)$$

One can get an arbitrary accuracy of ϵ between sample average and mean with confidence $1 - \sigma^2/n\epsilon^2$, which can be made as close to 1 as we wish by increasing n. This gives us the weak law of large numbers:

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{S_n}{n} - \overline{X}\right| \geq \epsilon\right\} = 0 \qquad \textbf{for every } \epsilon > 0.$$

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - \overline{X}\right| \geq \epsilon\right\} \leq \delta \qquad \textbf{any } \delta > 0, \quad \textbf{large enough } n$$

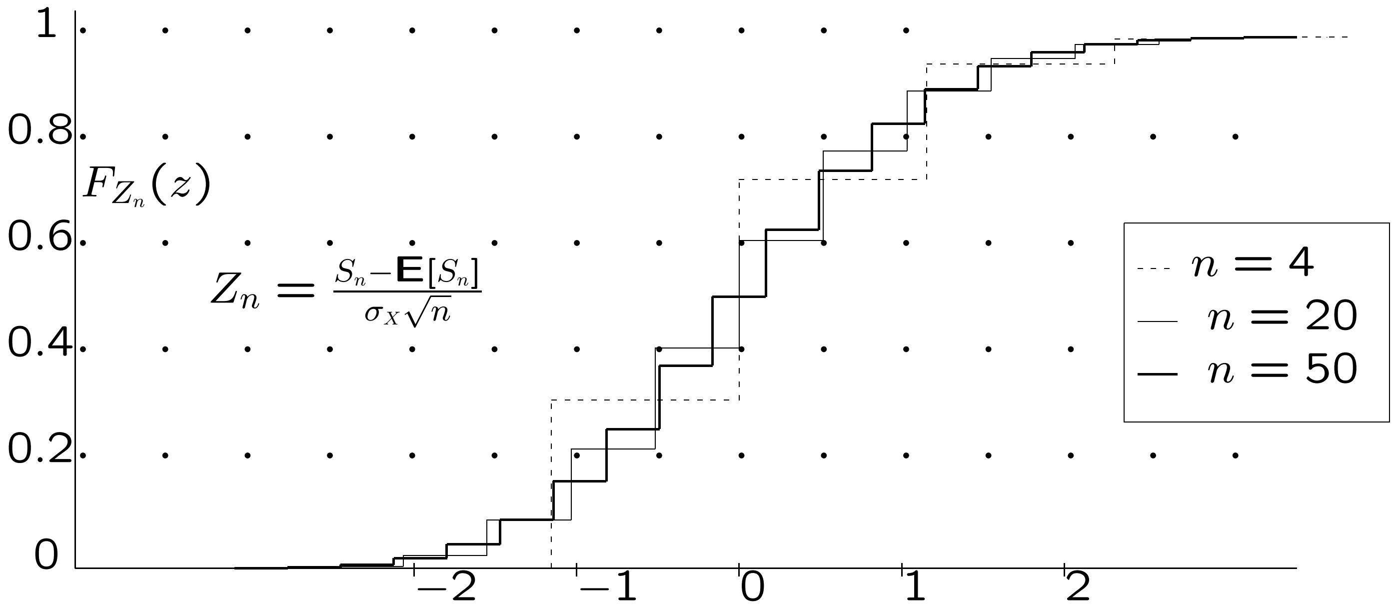


What this says is that P $\left\{\frac{S_n}{n} \leq x\right\}$ **is approaching a unit step at** $\overline{X}$ **as** $n \to \infty$**. For any fixed** ϵ**,** δ **goes to 0 as** $n \to \infty$**. Also,** δ **and** ϵ **can go to 0 as** $n \to \infty$**.**

This is called convergence in probability. It is also convergence in distribution

$$\lim_{n \to \infty} \mathbf{P}\left\{\frac{S_n}{n} \leq x\right\} = \mathbf{P}\left\{\overline{X} \leq x\right\} \quad \textbf{for all } x.$$

Recall that $S_n - n\overline{X}$ is a zero mean rv with variance $n\sigma^2$. Thus $\frac{S_n - n\overline{X}}{\sqrt{n}\sigma}$ is zero mean, unit variance.



Central limit theorem:

$$\lim_{n\to\infty}\left[\mathbf{P}\left\{\frac{S_n - n\overline{X}}{\sqrt{n}\,\sigma} \le y\right\}\right] = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)\,dx.$$

Example - Bernoulli process

$$N(t) = Y_1 + \cdots Y_t \qquad p_Y(1) = q, \quad p_Y(0) = 1 - q$$

The t**-tuple of** n **ones followed by** $t-n$ **zeros has probability** $q^n(1-q)^{t-n}$**. There are** $\binom{t}{n}$ t**-tuples.**

$$p_{N(t)}(n) = \binom{t}{n} q^n (1-q)^{t-n}$$

Using Stirlings approximation to the factorial, $t! \approx \sqrt{2\pi t}(t/e)^t$**, we get**

$$\begin{aligned} \binom{t}{n} &\approx \sqrt{\frac{t}{2\pi n(t-n)}} \left(\frac{t}{n}\right)^n \left(\frac{t}{t-n}\right)^{t-n} \\ p_{N(t)}(n) &\approx \sqrt{\frac{t}{2\pi n(t-n)}} \left(\frac{qt}{n}\right)^n \left(\frac{(1-q)t}{t-n}\right)^{t-n} \end{aligned}$$

For $n/t = q$**, right side is** $1/\sqrt{2\pi t q(1-q))}$

The focus here on theory, insight, conceptual understanding might be frustrating:

- For first year graduate students since it breaks undergraduate ”plug and chug” mode
- For the engineer since it breaks focus on bottom line-itis and requires conceptual rather than computational approach
- For the mathematician since insight is more important than generality or formalism.

Hold on —– It will get worse!

The strong law of large numbers

Version 1: For all $n > 1$, let $S_n = X_1 + \cdots + X_n$ where $X_1, X_2, \ldots$ are IID rv's with a finite mean $\overline{X}$. Then for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\left\{\bigcup_{m\geq n}\left\{\left|\frac{S_m}{m} - \overline{X}\right| > \epsilon\right\}\right\} = 0.$$

Contrast with weak law:

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{S_n}{n} - \overline{X}\right| \geq \epsilon\right\} = 0 \qquad \text{for every } \epsilon > 0.$$

The weak law probabilities are for n-tuples. It says that large enough n-tuples are sufficient.

The strong involves only (infinite) sequences. Looking at 10^{23}-tuples tells you nothing. It says that almost all sequences have relative frequencies that get within ϵ and stay there.

Why on earth should we care? The model becomes inappropriate long before the theorem takes effect.

There is an equivalent form of the theorem that says

$$\mathbf{P}\left\{\lim_{n\to\infty} \frac{S_n}{n} = \overline{X}\right\} = 1.$$

For each sample point in the set of sequences, S_n/n is a sequence of real numbers. This says that for 'almost all' sequences, that limit exists and equals $\overline{X}$.

This lets us treat stochastic processes as if they were sequences of numbers.

The strong law is very strange. Consider Bernoulli process. then $\overline{X}$ is $q = p_Y(1)$;.

For each q, there is a set of sequences of probability 1 (according to that probility measure).

This partitions the set of binary sequences into uncountably many sets of sequences, each of probability 1.

We outline a proof. This is, without a doubt, the most difficult theorem and proof in the course.

Let $E_m = \left\{ \left| S_m/m - \overline{X} \right| > \epsilon \right\}$ **and assume a MGF around** 0**.**

We will show that $\sum_m E_m < \infty$**.**

$$\mathbf{P}\{E_m\} = \mathbf{P}\left\{\frac{S_m}{m} - \overline{X} > \epsilon\right\} + \mathbf{P}\left\{\frac{S_m}{m} - \overline{X} < -\epsilon\right\}.$$

Both terms decrease geometrically with m**, so** $\sum_m \mathbf{P}\{E_m\} < \infty$**.**

Borel Cantelli lemma: If $\sum_m E_m < \infty$, then

$$\left[\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} E_m\right] = 0.$$

$\sum_{m=1}^{\infty} \mathbf{P}\{E_m\} < \infty$ **means that** $\sum_{m=1}^{n} \mathbf{P}\{E_m\}$ **converges to a finite limit as** $n \to \infty$**, and thus that**

$$\lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbf{P}\{E_m\} = 0.$$

$$\begin{aligned} \mathbf{P}\left\{\bigcap_n \bigcup_{m\geq n} E_m\right\} &\leq \mathbf{P}\left\{\bigcup_{m\geq n} E_m\right\} \qquad \textbf{for each } n \geq 0 \\ &\leq \sum_{m=n}^{\infty} \mathbf{P}\{E_m\} \end{aligned}$$