6.262: Discrete Stochastic Processes 2/9/11

Lecture 3: Laws of large numbers, convergence

Outline:

- Review of probability models
- Markov, Chebychev, Chernoff bounds
- Weak law of large numbers and convergence
- Central limit theorem and convergence
- Convergence with probability 1



## Review of probability models

Probability models are natural for real-world situations that are repeatable, using trials that

- have the same initial conditions
- are essentially isolated from each other
- have a fixed set of possible outcomes
- have essentially 'random' individual outcomes.

For any model, an extended 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 relative frequencies and sample averages in the repeated experiment.

The laws of large numbers (LLN's) specify what 'become deterministic' means.

They only operate within the extended model, but provide our only truly experimental way to compare the model with repeated trials of the real-world experiment.

Probability theory provides many many consistency checks and ways to avoid constant experimentation.

Common sense, knowledge of the real-world system, focus on critical issues, etc. often make repeated trials unnecessary.

The determinism in large numbers of trials underlies much of the value of probability.



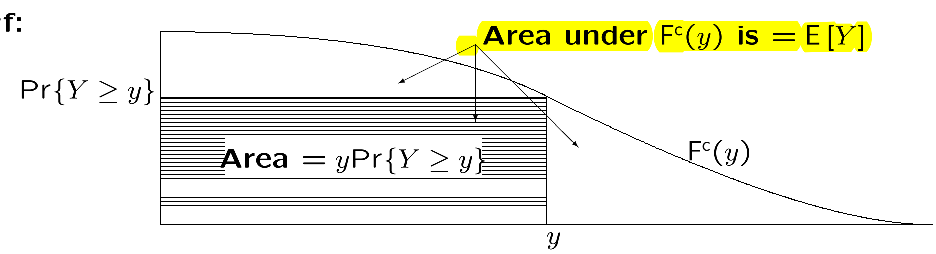
## Markov, Chebychev, Chernoff bounds

Inequalities, or bounds, play an unusually large role in probability. Part of the reason is their frequent use in limit theorems and part is an inherent imprecision in probability applications.

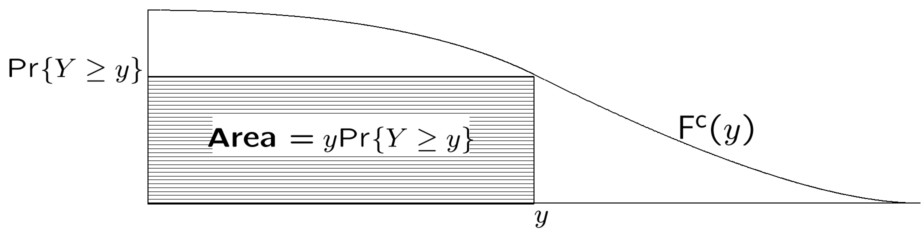
One of the simplest and most useful bounds is the Markov inequality: If $Y$ is a non-negative rv with an expectation $\mathsf{E}[Y]$, then for any real $y > 0$,

$$\Pr\{Y \geq y\} \leq \frac{\mathsf{E}[Y]}{y}$$

**Pf:**



**Markov inequality:** $\Pr\{Y \geq y\} \leq \frac{\mathsf{E}[Y]}{y}$




**Note that the Markov bound is usually very loose. It is tight (satisfied with equality) if $Y$ is binary with possible values 0 and $y$.**

**The Markov bound decreases very slowly (as $1/y$) with increasing $y$.**



**The Chebyshev inequality: If $Z$ has a mean $\mathsf{E}[Z] = \overline{Z}$ and a variance, $\sigma_Z^2$, then for any $\epsilon > 0$,**

$$\Pr\left\{|Z - \overline{Z}| \geq \epsilon\right\} \leq \frac{\sigma_Z^2}{\epsilon^2} \tag{1}$$

**Pf: Let $Y = (Z - \overline{Z})^2$. Then $\mathsf{E}[Y] = \sigma_Z^2$ and for any $y > 0$,**

$$\Pr\{Y \geq y\} \leq \sigma_Z^2/y; \qquad \Pr\left\{\sqrt{Y} \geq \sqrt{y}\right\} \leq \sigma_Z^2/y$$

**Now $\sqrt{Y} = |Z - \overline{Z}|$. Setting $\epsilon = \sqrt{y}$ yields (1).**

**Chebychev requires a variance, but decreases as $1/\epsilon^2$ with increasing distance $\epsilon$ from the mean.**

The **Chernoff bound**: For any $z > 0$ and any $r > 0$ such that the moment generating function $g_Z(r) = \mathsf{E}\left[e^{rZ}\right]$ exists,

$$\Pr\{Z \geq z\} \leq g_Z(r) \exp(-rz) \qquad (2)$$

Pf: Let $Y = e^{rZ}$. Then $\mathsf{E}[Y] = g_Z(r)$. For any $y > 0$, Markov says,

$$\Pr\{Y \geq y\} \leq g_Z(r)/y; \qquad \Pr\left\{e^{rZ} \geq e^{rz}\right\} \leq g_Z(r)/e^{rz},$$

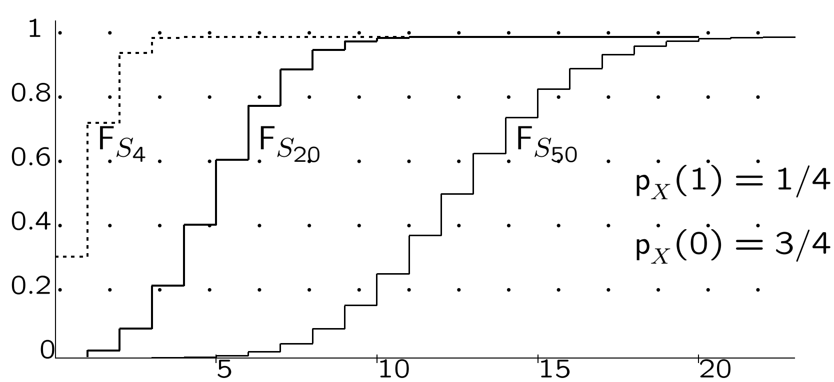
which is equivalent to (2).

This decreases exponentially with $z$ and is useful in studying large deviations from the mean.



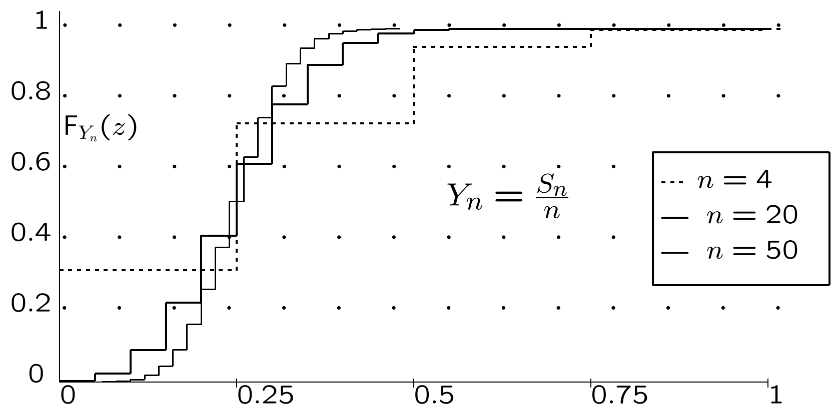
## The weak law of large numbers and convergence

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




The mean of the distribution varies with $n$ and the standard deviation varies with $\sqrt{n}$.

**The sample average is $S_n/n$, which is a rv of mean $\overline{X}$ and variance $\sigma^2/n$.**




**The mean of the distribution is $\overline{X}$ and the standard deviation decreases with $1/\sqrt{n}$.**



$$\textbf{VAR}\left(\frac{S_n}{n}\right) = \mathsf{E}\left[\left(\frac{S_n}{n} - \overline{X}\right)^2\right] = \frac{\sigma^2}{n}. \tag{3}$$

$$\lim_{n\to\infty} \mathsf{E}\left[\left(\frac{S_n}{n} - \overline{X}\right)^2\right] = 0. \tag{4}$$

**Note that (3) says more than (4), since it says the convergence is as $1/n$ and in fact it gives the variance explicitly. But (4) establishes a standard form of convergence of rv's called convergence in mean square.**

**Def: A sequence of rv's, $Y_1, Y_2, \dots$ converges in mean square to a rv $Y$ if**

$$\lim_{n\to\infty} \mathsf{E}\left[(Y_n - Y)^2\right] = 0$$

The fact that $S_n/n$ converges in mean square to $\overline{X}$ doesn't tell us directly what might be more interesting: what is the probabilility that $|S_n/n - \overline{X}|$ exceeds $\epsilon$ as a function of $\epsilon$ and $n$?

Applying Chebyshev to (3), however,

$$\Pr\left\{\left|\frac{S_n}{n} - \overline{X}\right| \geq \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2} \qquad \text{for every } \epsilon > 0 \qquad (5)$$

One can get an arbitrary accuracy of $\epsilon$ between sample average and mean with probability $1-\sigma^2/n\epsilon^2$, which can be made as close to 1 as we wish, by increasing $n$.

This gives us the weak law of large numbers (WLLN):

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



**WLLN:** $\lim_{n\to\infty} \Pr\left\{\left|\frac{S_n}{n} - \overline{X}\right| \geq \epsilon\right\} = 0 \qquad$ for every $\epsilon > 0$.

We have proven this under the assumption that $S_n = \sum_{n=1}^{n} X_n$ where $X_1, X_2, \dots,$ are IID with finite variance.

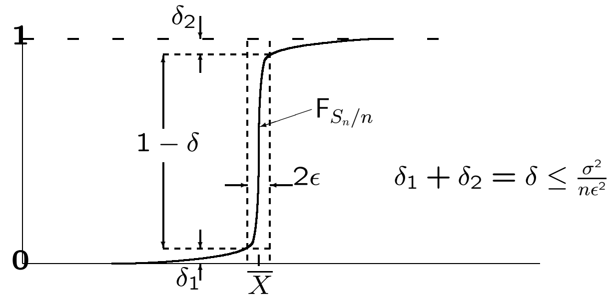
An equivalent statement (following from the definition of a limit of real numbers) is that for every $\delta > 0$,

$$\Pr\left\{\left|\frac{S_n}{n} - \overline{X}\right| \geq \epsilon\right\} \leq \delta \text{ for all large enough } n. \qquad (6)$$

Note that (6) tells us less about the speed of convergence than

$$\Pr\left\{\left|\frac{S_n}{n} - \overline{X}\right| \geq \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2}$$

But (6) holds without a variance (if $\mathsf{E}[|X|] < \infty$.)




**What this says is that $\Pr\left\{\frac{S_n}{n} \le x\right\}$ is approaching a unit step at $\overline{X}$ as $n \to \infty$. For any fixed $\epsilon$, $\delta$ goes to 0 as $n \to \infty$. If $\sigma_X < \infty$, then $\delta \to 0$ at least as $\sigma^2/n\epsilon^2$. Otherwise it might go to 0 more slowly.**

**Def: A sequence of rv's, $Y_1, Y_2, \ldots$ converges in probability to a rv $Y$ if for every $\epsilon > 0, \delta > 0$,**

$$\Pr\{|Y_n - Y| \ge \epsilon\} \le \delta \qquad \text{for all large enough } n$$

**This means that $\{S_n/n;\ n \ge 1\}$ converges to $\overline{X}$ in probability if $\mathsf{E}[|X|] < \infty$.**



**Review: We saw that if $\sigma_X$ exists and $X_1, X_2, \ldots$ are IID, then $\sigma_{S_n/n} = \sigma_X/\sqrt{n}$.**

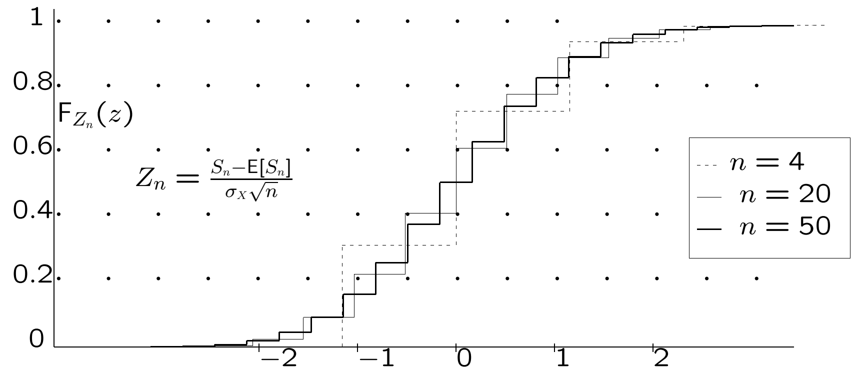
**Thus $S_n/n$ converges to $\overline{X}$ in mean square. Chebychev then shows that $S_n/n$ converges to $\overline{X}$ in probability.**

**In the same way, if $\{Y_n;\ n \ge 1\}$ converges to $Y$ in mean square, Chebychev show that it converges in probability.**

**That is, mean square convergence implies convergence in probability. The reverse is not true, since a variance is not required for the WLLN.**

**Finally, convergence in probability means that the distribution of $Y_n - Y$ approaches a unit step at 0.**

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:

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



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

Not only does $(S_n - n\overline{X})/\sqrt{n}\sigma_X$ have mean 0, variance 1 for all $n$, but it also becomes normal Gaussian.

We saw this for the Bernoulli case, but the general case is messy and the proof (by Fourier transforms) is not insightful.

The CLT applies to $\mathsf{F}_{S_n}$, not to the PMF or PDF.

Def: A sequence $Z_1, Z_2, \dots$ of rv's converges in distribution to $Z$ if $\lim_{n\to\infty} \mathsf{F}_{Z_n}(z) = \mathsf{F}_Z(z)$ for all $z$ where $\mathsf{F}_Z(z)$ is continuous.

The CLT says that $(S_n - n\overline{X})/\sqrt{n}\sigma_X$ converges in distribution to $\Phi$.

Convergence in distribution is almost a misnomer, since the rv's themselves do not necessarily become close to each other in any ordinary sense.

For example any sequence of IID rv's converge in distribution since they have the same distribution to start with.

Thm: Convergence in probability implies convergence in distribution.

Pf: Convergence of $\{Y_n;\, n \geq 1\}$ in probability means convergence to a unit step.

Thus convergence in mean square implies convergence in probability implies convergence in distribution.



Paradox: The CLT says something very strong about how $S_n/n$ converges to $\overline{X}$, but convergence in distribution is a very weak form of convergence.

Resolution: The rv's that converge in distribution in the CLT are $(S_n - n\overline{X})/\sqrt{n}\sigma_X$. Those that converge in probability to 0 are $(S_n - n\overline{X})/n$, a squashed version of $(S_n - n\overline{X})/\sqrt{n}\sigma_X$.

The CLT, for $0 < \sigma_X < \infty$, for example, says that $\lim_{n\to\infty} \Pr\Big\{(S_n - n\overline{X})/n \leq 0\Big\} = 1/2$. This can not be deduced from the WLLN.

## Convergence with probability 1

A rv is a far more complicated thing than a number. Thus it is not surprising that there are many types of convergence of a sequence of rv's.

A very important type is convergence with probability 1 (WP1). We introduce convergence WP1 here and discuss it more in Chap. 4.

The definition is deceptively simple.

**Def:** A sequence $Z_1, Z_2, \ldots,$ of rv's converges WP1 to a rv $Z$ if

$$\Pr\Big\{\omega \in \Omega : \lim_{n\to\infty} Z_n(\omega) = Z(\omega)\Big\} = 1$$



$$\Pr\Big\{\omega \in \Omega : \lim_{n\to\infty} Z_n(\omega) = Z(\omega)\Big\} = 1$$

In order to parse this, note that each sample point maps into a sequence of real numbers, $Z_1(\omega), Z_2(\omega), \ldots$.

Some of those sequences of real numbers have a limit, and in some cases, that limit is $Z(\omega)$. Convergence WP1 means that the set $\omega$ for which $Z_1(\omega), Z_2(\omega)$ has a limit, and that limit is $Z(\omega)$, is an event and that the probability of that event is 1.

One small piece of complexity that can be avoided here is looking at the sequence $\{Y_i = Z_i - Z;\ i \geq 1\}$ and asking if that sequence converges to 0 WP1.

**The strong law of large numbers (SLLN) says the following: Let $X_1, X_2, \dots$ be IID rv's with $\mathsf{E}[|X|] < \infty$. Then $\{S_n/n;\ n \geq 1\}$ converges to $\overline{X}$ WP1. In other words, all sample paths of $\{S_n/n;\ n \geq 1\}$ converge to $\overline{X}$ except for a set of probability 0.**

**These are the same conditions under which the WLLN holds. We will see, when we study renewal processes, that the SLLN is considerably easier to work with than the WLLN.**

**It will take some investment of time to feel at home with the SLLN, (and in particular to have a real sense about these sets of probability 1) and we put that off until chap 4.**