Laws of Large Numbers

concentration of measure: dist.

$\Rightarrow$ Statements about sample mean of many random variables, and how it converges to true mean, $S_n = X_1 + X_2 + \cdots + X_n$.

Prop: Suppose $X_1, X_2, \ldots$ is sequence of uncorrelated random variables s.t. each X_k has finite mean μ and variance less than or equal to C.

Then for any $\delta > 0$,

$$\Pr\left[\left| \frac{S_n}{n} - \mu \right| \geq \delta \right] \leq \frac{C}{n\delta^2}$$

(convergence in probability)

and so this $\to 0$ as $n \to \infty$.

Proof: The mean of S_n/n is

$$E[S_n/n] = E\left[\frac{\sum_{k=1}^{n} X_k}{n} \right] = \frac{\sum_{k=1}^{n} E[X_k]}{n} = \frac{n\mu}{n} = \mu.$$

The variance can be as follows:

$$\mathrm{var}\left(\frac{S_n}{n}\right) = \mathrm{var}\left(\frac{\sum_{k=1}^{n} X_k}{n}\right) = \frac{\sum_{k=1}^{n} \mathrm{var}(X_k)}{n^2} \leq \frac{nC}{n^2} = \frac{C}{n}.$$

now recall Chebyshev's ineq. if Z is random variable with finite mean μ and variance σ^2, then for any $d > 0$, $\Pr[|X - \mu| \geq \delta] \leq \frac{\sigma^2}{d^2}$.

so applying to $Z = \frac{S_n}{n}$, we get:

$$\Pr\left[\left| \frac{S_n}{n} - \mu \right| \geq \delta \right] \leq \frac{C}{n\delta^2}.$$

$\Rightarrow$ matlab simulation for $\frac{S_n}{n}$ vs. n and S_n vs. n for $\mathcal{U}(0,1)$ iid.

hand draw, Zenith radio telepathy experiments

Suppose $X_1, \dots, X_{100}$ are each r.v. with $\mu = 5$ and $\sigma^2 = 1$. Also, $|\mathrm{cov}(X_i, X_j)| \le 0.1$ if $i = j \pm 1$ and $\mathrm{cov}(X_i, X_j) = 0$ if $|i-j| \ge 2$.

Let $S_{100} = X_1 + \dots + X_{100}$

(a) show $\mathrm{var}(S_{100}) \le 120$.

(b) find upper bound on $\Pr\left[\left|\frac{S_{100}}{100} - 5\right| \ge 0.5\right]$.

(a) for any n, we know $\mathrm{var}(S_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{cov}(X_i, X_j)$.

$$= \sum_{i=1}^{n-1} \mathrm{cov}(X_i, X_{i+1}) + \sum_{i=1}^{n} \mathrm{cov}(X_i, X_i) + \sum_{i=2}^{n} \mathrm{cov}(X_i, X_{i-1})$$

$$\le (n-1)(0.1) + n + (n-1)(0.1) < 1.2n.$$

(b) from part (a), get that $\mathrm{var}\left(\frac{S_{100}}{100}\right) = \frac{1}{100^2} \mathrm{var}(S_{100}) \le 0.012$.

Also know $E\left[\frac{S_{100}}{100}\right] = \mu = 5$. So applying Chebyshev:

$$\Pr\left[\left|\frac{S_{100}}{100} - 5\right| \ge 0.5\right] \le \frac{0.012}{(0.5)^2} = 0.048.$$

Let $U_1, \dots, U_n$ be iid $\mathcal{E}(\lambda)$ with unknown λ. (has mean $\frac{1}{\lambda}$ and variance $\frac{1}{\lambda^2}$).

The ML estimator turns out to be $\hat{\lambda}_{ML} = \frac{n}{S_n}$, where $S_n = U_1 + \dots + U_n$.

Find the number of observations n s.t. $[(0.9)\hat{\lambda}_{ML}, (1.1)\hat{\lambda}_{ML}]$ is confidence interval for estimating λ with confidence level 96%.

Want $\Pr\left(\frac{0.9n}{S_n} \le \lambda \le \frac{1.1n}{S_n}\right] > 0.96$ or equivalently $\Pr\left[0.9 \le \frac{\lambda S_n}{n} \le 1.1\right] \ge 0.96$.

To use Chebyshev, note that $E\left[\frac{\lambda S_n}{n}\right] = E[\lambda U_1] = 1$.

$$\mathrm{var}\left[\frac{\lambda S_n}{n}\right] = \frac{\lambda^2}{n^2} \mathrm{var}(S_n) = \frac{\lambda^2 \mathrm{var}(U_1)}{n} = \frac{1}{n}.$$

So $\Pr\left[\left|\frac{\lambda S_n}{n} - 1\right| \ge \delta\right] \le \frac{1}{n\delta^2}$. Setting $\frac{1}{n\delta^2} = 0.04$ with $\delta = 0.1$ gives $n = \frac{1}{(0.1)^2(0.04)} = 2500$.

We know convergence of sample mean to mean using LLN; what about full distributional form of the sum?

What did De Moivre Laplace thm say?

<u>Central limit thm</u>. Suppose $X_1, X_2, \ldots$ are iid r.v. with mean μ and variance σ^2. Let $S_n = X_1 + \cdots X_n$. Then:

$$\lim_{n \to \infty} \Pr\left[\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le c \right] = \Phi(c).$$

Gaussian approx.

Ex. Suppose each day of the year, a stock:

$$\begin{cases} \text{increase by } 1\% & \text{w.p. } 0.5 \\ \text{stays} & \text{w.p. } 0.4 \\ \text{decrease by } 1\% & \text{w.p. } 0.1. \end{cases}$$

Consider value of 1 stock after 365 days.

(a) Probability stock at least triples.
(b) Probability stock at least quadruples.
(c) median value after one year.

Value after a year $Y = D_1 D_2 \cdots D_{365}$ where D_k is growth factor for day k. Convert to log to use CLT:

$$\ln Y = \sum_{k=1}^{365} \ln D_k.$$

$$\mu = E[\ln(D_k)] = 0.5 \ln 1.01 + 0.4 \ln 1 + 0.1 \ln .99 = 0.00397.$$

$$\sigma^2 = \text{var}(\ln(D_k)) = E[(\ln D_k)^2] - \mu^2 = 0.0000 4384$$

$\ln Y$ is approx Gaussian with mean $365\mu = 1.450$
~~mean~~ std $\sqrt{365\sigma^2} = 0.127$

[illegible]

[illegible]

(Weak [illegible]) [illegible]

Central limit theorem: Suppose $X_1, X_2, \ldots$ are i.i.d. r.v. with mean μ and variance σ^2. Let $S_n = X_1 + \cdots + X_n$. Then

$$\lim_{n\to\infty} \Pr\left\{ \frac{S_n - n\mu}{\sigma\sqrt{n}} \le x \right\} = \Phi(x)$$

Ex. [illegible]

Ex. Suppose a stock price each day [illegible]
{ increases by 1% [illegible]
{ decreases by 1% [illegible]

[illegible] after 1 year (~ 250 days) [illegible]

(a) [illegible]

(b) [illegible]

(c) [illegible]

[illegible] $Y = Y_0 Z_1 Z_2 \cdots Z_{250}$ where [illegible]

[illegible] (CLT)

$$\ln Y = \sum_{k=1}^{250} \ln Z_k$$

$E[\ln Z] = 0.5 \ln(1.01) + 0.5 \ln(0.99) =$ [illegible]

$Var(\ln Z) = E[(\ln Z)^2] - (E[\ln Z])^2 = 0.0000995$

$\ln Y$ is approximately normal with mean [illegible], standard deviation $\sqrt{250\sigma^2} = 0.157$

So $P[Y \geq c] = Pr[\ln(Y) \geq \ln(c)]$

$$= Pr\left[\frac{\ln(Y)-1.450}{0.127} \geq \frac{\ln(c)-1.450}{0.127}\right]$$

$$\approx Q\left(\frac{\ln(c)-1.450}{0.127}\right)$$

(a) In particular $Pr(Y \geq 3) \approx Q(-2.77) = 0.997$

(b) $Pr(Y \geq 4) \approx Q(-0.4965) = 0.69.$

(c) median value is value of c s.t. $Pr(Y \geq c) = 0.5$

which by Gaussian approx is $e^{1.450} = 4.26..$