Recall

- Limit theorems, convergence, etc

Let

common mean μ
common variance σ^2

$$M_n = \frac{X_1 + \cdots + X_n}{n}, \quad X_k \text{ iid}$$

Saw

$$\mathbb{E}(M_n) = \mu \ \forall n; \quad Var(M_n) = \frac{\sigma^2}{n} \ \forall n$$

Chebyshev Inequality:

$$\mathbb{P}(\{|X-\mu| \geq c\}) \leq \frac{Var(X)}{c^2} \quad \forall c$$

From this, it follows that

$$\mathbb{P}(\{|M_n - \mu| \geq \varepsilon\}) \leq \frac{\sigma^2}{n\varepsilon^2} \quad \forall \varepsilon > 0$$

Consequence

$$\mathbb{P}(\{|M_n - \mu| \geq \varepsilon\}) \to 0 \text{ as } n \to \infty \quad \forall \varepsilon > 0$$

This result is called the Weak Law of Large Numbers

← start of new material

Proof of Chebyshev:

given $c > 0$, define $g(X)$ by

$$g(X) = \begin{cases} 0 & |x-\mu| < c \\ c^2 & |x-\mu| \geq c \end{cases}$$

then

$$g(X) \leq |X-\mu|^2 \Rightarrow \mathbb{E}[g(X)] \leq \mathbb{E}(|X-\mu|^2) = \operatorname{Var}(X)$$

but

$$\mathbb{E}[g(X)] = 0 \cdot \mathbb{P}(|X-\mu| \leq c) + c^2\,\mathbb{P}(|X-\mu| \geq c)$$

So,

$$\mathbb{P}(|X-\mu| \geq c) \leq \frac{\operatorname{Var}(X)}{c^2}$$

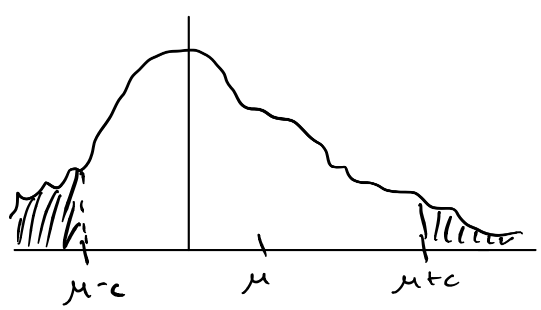
Another sometimes useful inequality

Markov Inequality: If X is a nonnegative-valued rv, then for every $c>0$,

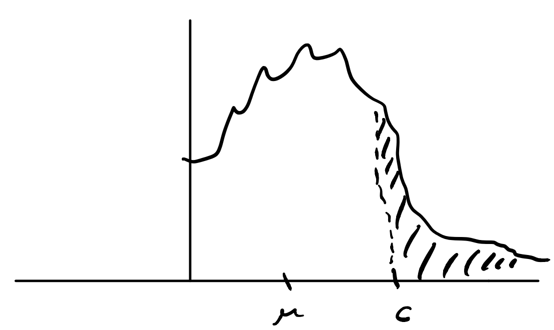
$$\mathbb{P}(\{X \geq c\}) \leq \frac{\mathbb{E}(X)}{c}$$

Think of these both as quantatative bounds on <u>tail</u> <u>probabilities</u>

Chebyshev



Markov



Example - Weak Law of Large Numbers

Have a coin with $\mathbb{P}(\{H\}) = p =$ unknown.

$$X_k = \begin{cases} 1, & H \text{ flip } k \\ 0, & T \text{ flip } k \end{cases} \quad ; \quad X_k \text{ independent}$$

$$M_n = \frac{X_1 + \cdots + X_n}{n} = \text{fraction of heads in first } n \text{ flips}$$

$$\mathbb{E}(M_n) = p = \mathbb{E}(X_k) \ \forall k$$

$$\text{Var}(M_n) = \frac{p(1-p)}{n} \quad (\text{recall } p(1-p) = \text{Var}(X_k) \ \forall k)$$

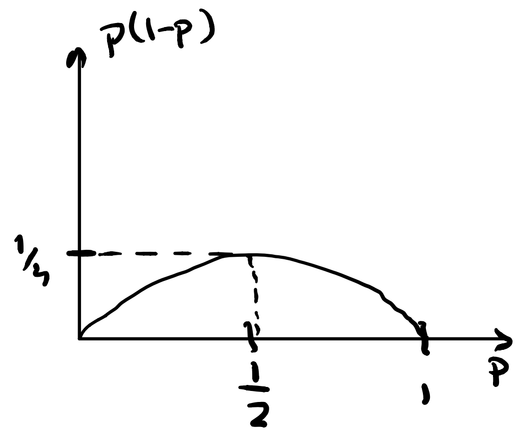
$\forall \varepsilon > 0$

$$\mathbb{P}(\{|M_n - p| \geq \varepsilon\}) \leq \frac{p(1-p)}{n\varepsilon^2}$$

Question: How large does n have to be to guarantee that M_n is within $\underbrace{1\% \text{ of } p}_{\text{accuracy}}$ with $\underbrace{\text{probability} \geq 95\%}_{\text{confidence}}$?

Thus, asking how large n has to be for

$$\mathbb{P}(|M_n - p| \geq \underbrace{0.01}_{\varepsilon}) \leq 0.05$$



Need n large enough so

$$\frac{p(1-p)}{10^{-4}\, n} \leq 0.05$$

Don't know p, but

$$p(1-p) \leq \frac{1}{4} \quad \forall p \in [0,1]$$

Calculating n,

$$\frac{1}{4n} \cdot \frac{1}{10^{-4}} \leq 0.05 \longrightarrow \underline{n \geq 50{,}000}$$

That was a conservative estimation - can do better...

Another consequence of Chebyshev:

$$\mathbb{P}\left(\{|X-\mu| \geq k\sigma\}\right) \leq \frac{1}{k^2}$$

"probability of being k std deviation away from mean $\leq \frac{1}{k^2}$"

The kind of convergences taking place in WLLN **converges in probability**

Definition: Given a sequence of rv's Y_n, $1 \leq n < \infty$, and a number a, say Y_n <u>converges in probability</u> to a as <u>$n \to \infty$</u> when $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\{|Y_n - a| \geq \varepsilon) = 0$$

WLLN just says "$M_n \to \mu$ in probability as $n \to \infty$"

Convergence in probability is a weakish kind of convergence; just one of many we encounter.

Another is, in the same context, say Y_n converges to a mean sequence when

$$\lim_{n\to\infty} \mathbb{E}\left[|Y_n - a|^2\right] = 0$$

Can show mean square convergence $\Rightarrow$ Convergence in probability

Example- Convergence in Probability is Weak

The sequence $Y_n \to 0$ in probability, but $\mathbb{E}(Y_n) = n \ \forall n$.

For each n, Y_n takes on 2 values: 0 an n^2 w/ respective probabilities $1-\frac{1}{n}$ and $\frac{1}{n}$.

Note $\mathbb{E}[Y_n] = \frac{n^2}{n} = n \ \forall n$

But, $\forall \varepsilon > 0$, $\mathbb{P}(Y_n \in [-\varepsilon, \varepsilon]) \to 1$ as $n \to \infty$

Central Limit Theorem

Recall that if X_k iid w/ common μ, σ^2

$$M_n = \frac{X_1 + \cdots + X_n}{n} \Rightarrow \mathbb{E}(M_n) = \mu \text{ and } Var(M_n) = \frac{\sigma^2}{n}$$

Form Z_n by renormalizing so $\mathbb{E}(Z_n) = 0$ and $Var(Z_n) = 1 \ \forall n$.

$$Z_n = \frac{X + \cdots + X_n - n\mu}{\sqrt{n}\,\sigma}$$

check that this works!

Fact: this is the Central Limit Theorem!

In this context, Z_n converges to, as $n \to \infty$, a Gaussian with mean $\mu = 0$ and variance $\sigma^2 = 1$.

Converging in the sense that for every z,

$$F_{Z_n}(z) \longleftrightarrow \Phi(z)$$

standard normal CDF

This is a super-general result - X_k's can be continuous, discrete, etc.

For example also justifies using Gaussian's to model "real-world noise", b/c such noise arises often a sum of lots of independent random effects.

People (often with lack of care) turn to CLT for better bounds on tail probabilities than Chebyshev gives us.

Re-vistit Coin Problem Above with CLT

$$\mathbb{P}(\{|M_n - p| \geq 0.01\}) \leq 0.05$$

because M_n distributed roughly symmetrically about μ for n large

$$\mathbb{P}(\{|M_n - \mu| \geq 0.01\}) \simeq 2\mathbb{P}(M_n - \mu \geq 0.01)$$

largest possible $p(1-p)$

$$\text{Set } Z_n = \frac{\sqrt{n}}{\sigma}(M_n - \mu) = 2\sqrt{n}(M_n - \mu) \; ; \quad \sigma = \frac{1}{4}$$

Interested in choosing n so

$$2\mathbb{P}(\{Z_n \geq 2(0.01)\sqrt{n}) < 0.05$$

$$\simeq 2(1 - \Phi(2(0.01)n)) \Rightarrow n \sim 9604 \text{ needed}$$