

## Recall

- limit theorems, convergence, etc

Let

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

common mean  $\mu$   
common variance  $\sigma^2$

Now

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

### Chebychev Inequality:

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

From this, it follows that

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

### Consequence

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

This result is called the Weak Law of Large Numbers

← start of new material

### Proof of Chebychev:

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

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

then

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

but

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

So,

$$\mathbb{P}(|x - \mu| > c) \leq \frac{\text{Var}(x)}{c^2}$$

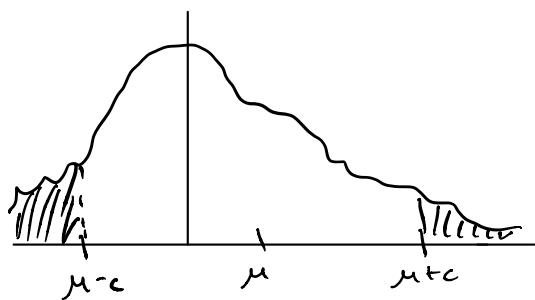
Another sometimes useful inequality

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

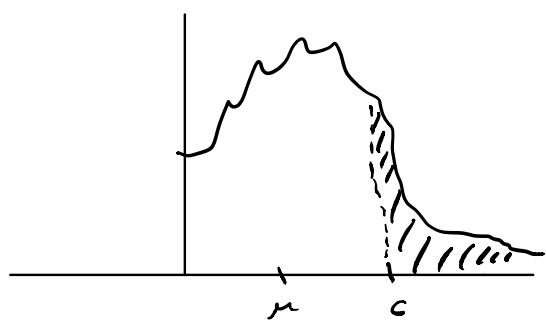
$$\mathbb{P}\{X \geq c\} \leq \frac{\mathbb{E}(X)}{c}$$

Think of these both as quantitative bounds on tail probabilities

Chesbyshev



Markov



## Example - Weak Law of Large Numbers

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

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

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

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

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

$$\forall \varepsilon > 0$$

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

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

Thus, asking how large  $n$  has to be for

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

Need  $n$  large enough so

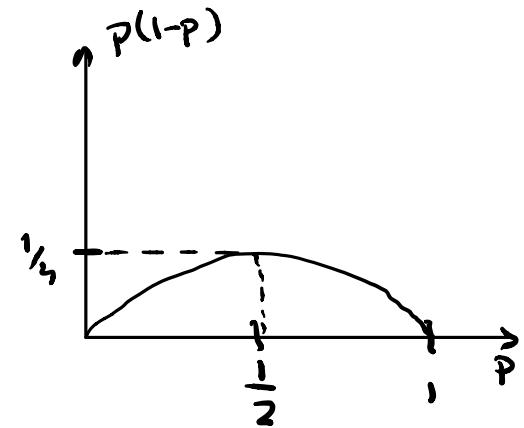
$$\frac{p(1-p)}{10^{-4} n} \leq 0.05$$

Don't know  $p$ , but

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

Calculating  $n$ ,

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



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

Another consequence of Chebyshev:

$$P(|X-\mu| > k\sigma) \leq \frac{1}{k^2}$$

"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 r.v's  $Y_n$ ,  $1 \leq n < \infty$ , and a number  $a$ , say  $Y_n$  converges in probability to  $a$  as  $n \rightarrow \infty$  when  $\forall \varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P(|Y_n - a| > \varepsilon) = 0$$

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

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

$$\lim_{n \rightarrow \infty} E[|Y_n - a|^2] = 0$$

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

## Example - Convergence in Probability is Weak

The sequence  $Y_n \rightarrow 0$  in probability, but  $\text{IE}(Y_n) = n \neq 0$ .

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

Note  $\text{IE}[Y_n] = \frac{n^2}{n} = n \neq 0$

But, if  $\varepsilon > 0$ ,  $\text{P}(Y_n \in [-\varepsilon, \varepsilon]) \rightarrow 1$  as  $n \rightarrow \infty$

## Central Limit Theorem

Recall that if  $X_k$  iid w/ common  $\mu, \sigma^2$

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

Form  $Z_n$  by renormalizing so  $\text{IE}(Z_n) = 0$  and  $\text{Var}(Z_n) = 1 \neq n$ .

$$Z_n = \frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} \quad \left. \begin{array}{l} \text{check that this} \\ \text{works!} \end{array} \right\}$$

Fact: this is the Central Limit Theorem!

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

Converging in the sense that for every  $z$ ,

$$F_{Z_n}(z) \longleftrightarrow \Phi(z) \quad \begin{array}{l} \text{standard normal CDF} \\ \swarrow \end{array}$$

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 Chebychev gives us.

Re-visit Coin Problem Above with CLT

$$P(|M_n - p| > 0.01) \leq 0.05 \quad \text{because } M_n \text{ distributed roughly symmetrically about } \mu \text{ for } n \text{ large}$$

$$P(|M_n - \mu| > 0.01) \approx 2P(M_n - \mu > 0.01)$$

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

Interested in choosing  $n$  so

$$\begin{aligned} 2P(Z_n > 2(0.01)\sqrt{n}) &< 0.05 \\ \approx 2(1 - \Phi(z(0.01)n)) &\Rightarrow n \sim 9604 \text{ needed} \end{aligned}$$