Last HW

Ch7: Problems: 7.32, 7.36, 7.38
Theoretical: 7.19

Ch8: Problems 8.1, 8.4, 8.5, 8.7
Theoretical: 8.1

Today: Start Limit Theorems

Let X be a random variable with some distribution (which may be unknown)

X represents the outcome of some experiment,

We can run several independent trials of this experiment

Generates a sequence of random variables $X_1, X_2, X_3, \ldots$

X_i = value on the ith trial

All random variables X_i are "essentially the same as X"

$X, X_1, X_2, X_3, \ldots$ all have same distribution

$X_1, X_2, X_3, \ldots$ are independent identically distributed

Particular values of $\overbrace{X_1, X_2, X_3, \dots}^{\text{samples}}$ are like a dataset we need to analyze.

How to find the mean $\mu = E[X]$?

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n} \quad \text{sample mean}$$

Why good: $E[\overline{X}] = E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i]$

each X_i has $E[X_i] = \mu$

$$= \frac{1}{n} n\mu = \mu$$

Q: How close is $\overline{X}$ to being μ?
Law of large numbers gives an answer.

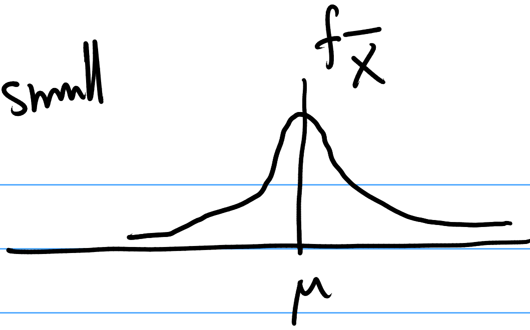
$$Var(\overline{X}) = \sum_{i=1}^{n} Var\left(\frac{X_i}{n}\right) = \sum_{i=1}^{n} \frac{1}{n^2} Var(X_i)$$

$$= \sum_{i=1}^{n} \frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n}$$

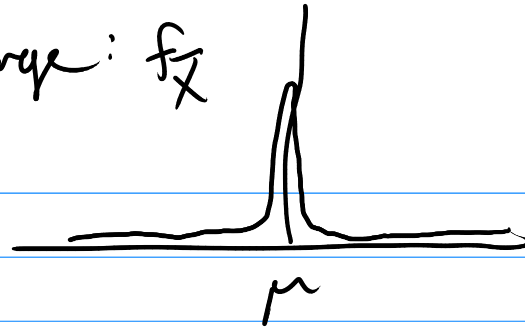
$\boxed{\sigma^2 = Var(X)}$ Variance of sample mean goes to 0 as $n \to \infty$

If you take $n = \#$ of samples to be very large, then it becomes very likely that $\overline{X}$ is close to μ

n small



n large:



Weak Law of Large numbers is a precise statement of this.

<u>Theorem</u> (Weak Law of Large number)

For any value of $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right| \geq \varepsilon\right\} = 0$$

$$\lim_{n\to\infty} P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right| < \varepsilon\right\} = 1$$

Even more quantitatively:

$$P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}$$

X has $\sigma^2 = 25$

Q: How large does n have to be?

$$P\left\{ \left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| \geq .1 \right\} \leq .05$$

$$P\left\{ \left| \frac{X_1 + \cdots + X_n}{n} - \mu \right| < .1 \right\} \geq .95$$

i.e. 95% confidence that sample mean is within .1 of the true value.

Need $\dfrac{\sigma^2}{n\varepsilon^2} \leq .05$

$$\frac{25}{n \cdot (.1)^2} \leq .05 \iff \frac{2500}{n} \leq .05 \iff n \geq 2500 \cdot 20 = 50000$$

Proof involves two lemmas

Lemma (Markov's inequality) Suppose X is a NONNEGATIVE Random variable

(i.e. $P\{X<0\}=0$)

Then: $P\{X \geq a\} \leq \frac{E[X]}{a}$ (true for any $a>0$)

Proof let $I = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases}$

$E[I] = P\{X \geq a\}$

Also: $\frac{X}{a} \geq I$: ~~indeed~~ if $X \geq a$ $\frac{X}{a} \geq 1 = I$

if $X < a$ $\frac{X}{a} \geq 0 = I$

(because $X \geq 0$)

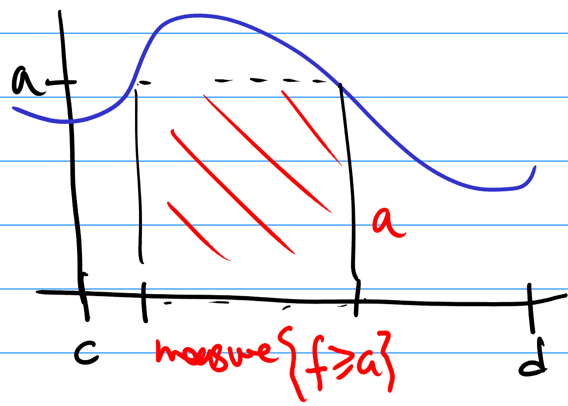
therefore $E\left[\frac{X}{a}\right] \geq E[I]$

$$\frac{E[X]}{a} \geq E[I] = P\{X \geq a\}$$

OPTIONAL

Analogy in calculus: $f(x)$ is a nonnegative function

then $\text{measure}\{f \geq a\} \leq \dfrac{\int_c^d f\,dx}{a}$



$\int_c^d f\,dx$ = Area under curve

$\geq$ Area of the rectangle

$= a \cdot \text{measure}\{f \geq a\}$

Corollary (Chebyshev's inequality) $(k>0)$

Apply Markov to $(X-\mu)^2$ $\quad E[(X-\mu)^2] = \sigma^2$

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

use $(X-\mu)^2 \geq k^2 \iff |X-\mu| \geq k$

get
$$\boxed{P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}}$$

Proof of Weak Law of Large numbers:

Apply Chebyshev to $\dfrac{X_1+\cdots+X_n}{n}$

$$E\left[\frac{X_1+\cdots+X_n}{n}\right]=\mu \qquad Var\left(\;\cdot\;\right)=\frac{\sigma^2}{n}$$

$$P\left\{\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|\geq\varepsilon\right\}\leq\frac{\sigma^2}{n}\frac{1}{\varepsilon^2}=\frac{\sigma^2}{n\varepsilon^2}$$

Example of Markov's inequality

X has mean $=50$ $\quad\sigma^2=25$

$$P\{X\geq 75\}\leq\frac{50}{75}=\frac{2}{3}$$