2.8 The Weak Law of Large Numbers

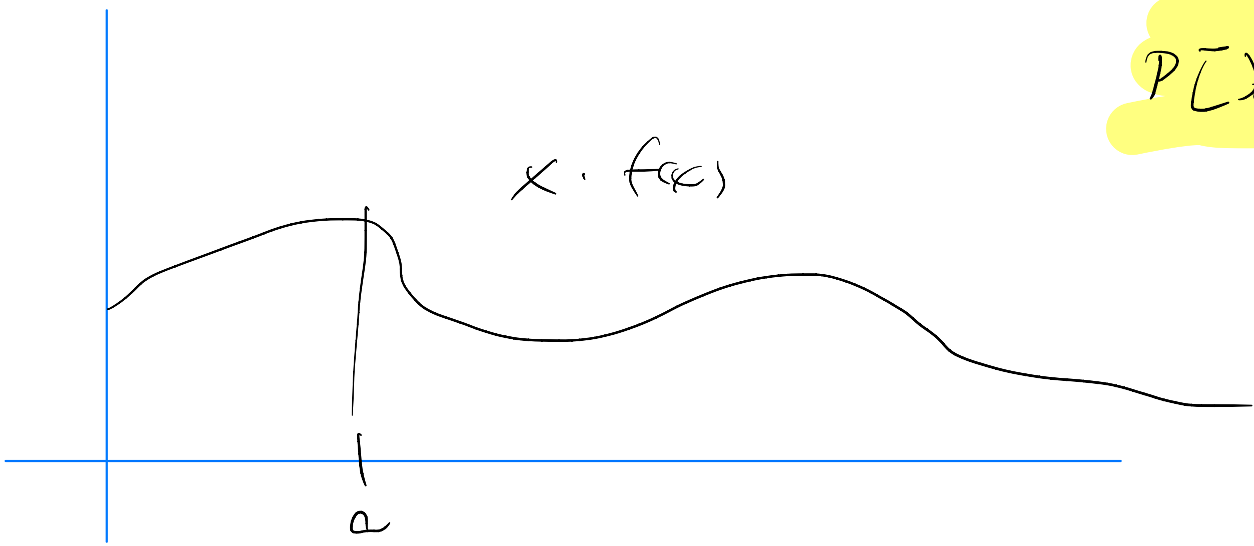
Markov's inequality

Let $X \geq 0$ s.th. $E[X]$ exists. Let also $a > 0$. Then

$$E[X] = \int_0^\infty x\, f(x)\, dx \geq \int_a^\infty x\, f(x)\, dx$$

$$\geq \int_a^\infty a\, f(x)\, dx = a \int_a^\infty f(x)\, dx = a\, P[X \geq a]$$

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



Apply Markov's Inequality:

$$Y := (X-\mu)^2 \quad , \quad a = k^2$$

Assume: $Var(X) = \sigma^2 \Rightarrow E[Y] = Var(X) = \sigma^2$

$$P[|X-\mu| \geq k]$$

$$= P[(X-\mu)^2 \geq k^2] = P[Y \geq k^2]$$

$$\leq \frac{E[Y]}{k^2} = \frac{\sigma^2}{k^2}$$

Tchebyshev's Inequality:

$$P[|X-\mu| \geq k] \leq \frac{\sigma^2}{k^2} \quad \text{if } Var(X) = \sigma^2$$

Example 52 Suppose the working time of a person is a RV X with $\mu = 40$ hrs.

1) How propable is it that the person will work more than 60 hrs?

Apply: $P[X \geq a] \leq \frac{E[X]}{a}$

$$P[X \geq 60] \leq \frac{\mu}{60} = \frac{40}{60} = \frac{2}{3}$$

Example 52 Suppose the working time of a person is a RV X with $\mu = 40$ hrs.

2) If $Var(X) = 16$, how probable is it the person will work between 32 and 48 hrs?

Apply: $P[\,|X-\mu| \geq k\,] \leq \frac{\sigma^2}{k^2}$

$$P[\,|X-40| \geq 8\,] \leq \frac{\sigma^2}{8^2} = \frac{16}{8^2} = \frac{1}{4}$$

$$\Rightarrow P[\,|X-40| \leq 8\,] \geq 1 - \frac{1}{4} = \frac{3}{4}$$

Example: Small Schools

Educational scientists found that among the schools that fare best in evaluations of teaching success, there are many more small schools than there are small schools among all schools.

(See statistics from North Carolina)

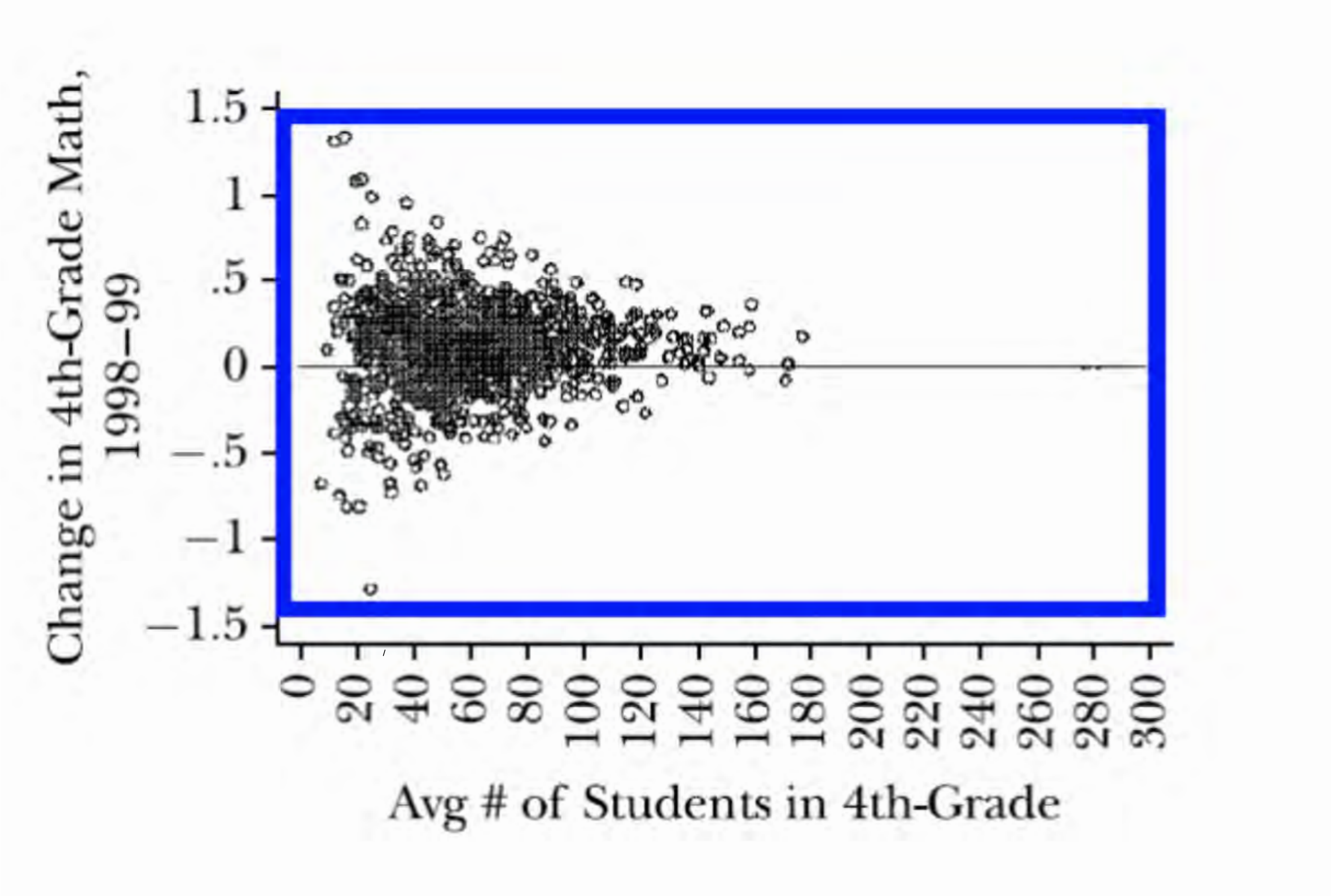
The Gates Foundation decided in the early 2000's to invest heavily in the establishment of small schools (e.g., by splitting larger schools into smaller ones)

Was that a good idea?

School Size	*Percentage Ever "Top 25" 1997–2000*
Smallest decile	27.7%
2nd	11.8
3rd	8.2
4th	3.6
5th	2.4
6th	3.6
7th	4.8
8th	7.1
9th	0
Largest decile	1.2
Total	7.0

Performance of small schools in North Carolina

Distribution of Performance wrt Student Numbers



What happens if we execute an experiment many times and take averages of the outcomes?

Let X be a RV, let $X_1, \dots, X_n$ be RVs that

1) have the same distribution as X
2) are independent

(i.i.d. RVs, i.e., independent identically distributed RVs)

Let $\sigma^2 = Var(X) = Var(X_i)$, $\mu = E[X] = E[X_i]$,

$$\overline{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

We have $\overline{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$. Then

$$E[\overline{X_n}] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} \cdot (n \cdot \mu) = \mu$$

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

idependence of the X_i

$$= \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{1}{n^2} \cdot (n \cdot \sigma^2) = \frac{\sigma^2}{n}$$

Chebyshev with $\overline{X_n}$ and $k = \varepsilon$

Probability of ε-outliers

$$P\left[\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right] = P[|\overline{X_n} - \mu| > \varepsilon]$$

$$\leq \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = \frac{1}{n} \cdot \frac{\sigma^2}{\varepsilon^2} \longrightarrow 0 \qquad (n \to \infty)$$

Theorem (Weak law of Large Numbers)

Let $X_1, \dots, X_n, \dots$ be i.i.d. RVs with mean μ and $Var(X) < \infty$.

Then for every $\varepsilon > 0$

$$\lim_{n \to \infty} P\left[\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| > \varepsilon \right] = 0$$

i.e., the probability of ε-outliers goes toward 0.

There is also a Strong Law of Large Numbers, which says

$$P\left[\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu \right] = 1$$

for i.i.d. RVs X_i provided $E[X_i] < \infty$.