Limit theorems and classical statistics

- $X, X_1, \ldots, X_n$ i.i.d.
  - Weak law of large numbers: $\dfrac{X_1 + \cdots + X_n}{n} \longrightarrow \mathbf{E}[X]$
  - Central limit theorem: $X_1 + \cdots + X_n \approx$ normal
- Estimating an unknown mean
  - accuracy of estimates
  - confidence intervals
- Classical statistics more generally
  - the philosophy
  - sample mean-based estimators
  - general methods (maximum likelihood)

Inequalitities, convergence, and the Weak Law of Large Numbers

- Inequalities
  - bound $\mathbf{P}(X \geq a)$ based on limited information about a distribution
  - Markov inequality (based on the mean)
  - Chebyshev inequality (based on the mean and variance)

- WLLN: $X, X_1, \ldots, X_n$ i.i.d.

$$\frac{X_1 + \cdots + X_n}{n} \longrightarrow \mathbf{E}[X]$$

  - application to polling

- Precise defn. of convergence
  - convergence "in probability"

The Markov inequality

- Use a bit of information about a distribution
  to learn something about probabilities of "extreme events"

- "If $X \geq 0$ and $\mathbf{E}[X]$ is small, then X is unlikely to be very large"

Markov inequality: If $X \geq 0$ and $a > 0$, then $\mathbf{P}(X \geq a) \leq \dfrac{\mathbf{E}[X]}{a}$

The Markov inequality

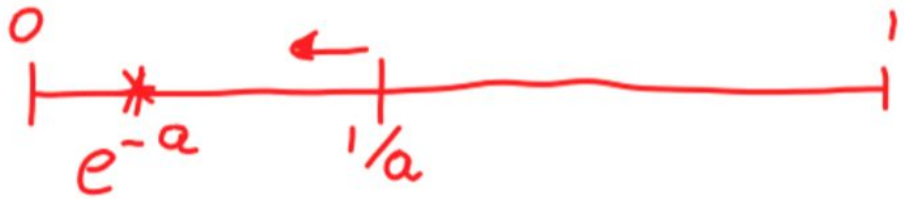
Markov inequality: If $X \geq 0$ and $a > 0$, then $\mathbf{P}(X \geq a) \leq \dfrac{\mathbf{E}[X]}{a}$

- **Example:** X is Exponential($\lambda = 1$): $\mathbf{P}(X \geq a) \leq$

The Markov inequality

Markov inequality: If $X \geq 0$ and $a > 0$, then $\mathbf{P}(X \geq a) \leq \dfrac{\mathbf{E}[X]}{a}$

- **Example:** X is Exponential($\lambda = 1$): $\mathbf{P}(X \geq a) \leq$



EXAMPLE 1. Suppose we have a 100 people attending an event. Suppose we ask:

- Is it possible that at least 95% of the people are younger than the average person in the group?

Here by "average" we mean "of mean age" (if you consider their median age, the following conclusions would not hold).

The answer is yes: one older person can pull up the average a lot (just imagine 99 one year olds being babysat by a 100 year old person).

Now consider the question:

- Is it possible that at least 50% of the people are older than twice the average age in the group?

Well, no, because if you take just those 50% whose age is more than double the average age and compute their average, you've already exceeded the average. Or think of it in terms of totals. If the average is μ, the the total age is 100 μ. Now if you have 50 people whose age is more than 2 μ, sum of their ages is already greater than $50 \times 2\ \mu = 100\ \mu$, which is impossible.

Similarly, you can't have a third of the people in the group who are older than 3 times the average age in the group. And that's exactly what Markov's inequality says.

The Chebyshev inequality

- Random variable X, with finite mean μ and variance σ^2
- "If the variance is small, then X is unlikely to be too far from the mean"

Chebyshev inequality: $\mathbf{P}\big(|X-\mu| \geq c\big) \leq \dfrac{\sigma^2}{c^2}$

Markov inequality: If $X \geq 0$ and $a > 0$, then $\mathbf{P}(X \geq a) \leq \dfrac{\mathbf{E}[X]}{a}$

The Chebyshev inequality

- Random variable X, with finite mean μ and variance σ^2
- "If the variance is small, then X is unlikely to be too far from the mean"

Chebyshev inequality: $\mathbf{P}\big(|X-\mu| \geq c\big) \leq \dfrac{\sigma^2}{c^2}$

Markov inequality: If $X \geq 0$ and $a > 0$, then $\mathbf{P}(X \geq a) \leq \dfrac{\mathbf{E}[X]}{a}$

$$P\big(|X-\mu| \geq c\big) = P\big((X-\mu)^2 \geq c^2\big) \leq \frac{E\big[(X-\mu)^2\big]}{c^2} = \frac{\sigma^2}{c^2}$$

The Chebyshev inequality

Chebyshev inequality: $\mathbf{P}\Big(|X-\mu| \geq c\Big) \leq \dfrac{\sigma^2}{c^2}$

$\mathbf{P}\Big(|X-\mu| \geq k\sigma\Big) \;\leq$

- **Example:** X is Exponential($\lambda = 1$): $\mathbf{P}(X \geq a) \leq \dfrac{1}{a}$ (Markov)

The Chebyshev inequality

Chebyshev inequality: $\mathbf{P}\big(|X-\mu| \geq c\big) \leq \dfrac{\sigma^2}{c^2}$ •

$$\mathbf{P}\big(|X-\mu| \geq k\sigma\big) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2} \qquad k=3 \qquad \leq \frac{1}{9}$$

- **Example:** X is Exponential($\lambda = 1$): $\mathbf{P}(X \geq a) \leq \dfrac{1}{a}$ (Markov)

$$P(X \geq a) = P(X-1 \geq a-1) \leq P(|X-1| \geq a-1) \leq \frac{1}{(a-1)^2} \sim \frac{1}{a^2}$$

Suppose that it is known that the number of items produced in a factory during a week is a random variable with mean 50.

(a) What can be said about the probability that this week's production will exceed 75?

(b) If the variance of a week's production is known to equal 25, then what can be said about the probability that this week's production will be between 40 and 60?

Solution. Let X be the number of items that will be produced in a week.

(a) By Markov's inequality,

$$P\{X > 75\} \leq \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$$

(b) By Chebyshev's inequality,

$$P\{|X - 50| \geq 10\} \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$$

Hence,

$$P\{|X - 50| < 10\} \geq 1 - \frac{1}{4} = \frac{3}{4}$$

so the probability that this week's production will be between 40 and 60 is at least .75. ■

If X is uniformly distributed over the interval (0, 10), then, since $E[X] = 5$ and $\text{Var}(X) = \frac{25}{3}$, it follows from Chebyshev's inequality that

$$P\{|X - 5| > 4\} \leq \frac{25}{3(16)} \approx .52$$

whereas the exact result is

$$P\{|X - 5| > 4\} = .20$$

Thus, although Chebyshev's inequality is correct, the upper bound that it provides is not particularly close to the actual probability.

Similarly, if X is a normal random variable with mean μ and variance σ^2, Chebyshev's inequality states that

$$P\{|X - \mu| > 2\sigma\} \leq \frac{1}{4}$$

whereas the actual probability is given by

$$P\{|X - \mu| > 2\sigma\} = P\left\{\left|\frac{X - \mu}{\sigma}\right| > 2\right\} = 2[1 - \Phi(2)] \approx .0456 \quad \blacksquare$$

The Weak Law of Large Numbers (WLLN)

- $X_1, X_2, \ldots$ i.i.d.; finite mean μ and variance σ^2

  Sample mean: $M_n = \dfrac{X_1 + \cdots + X_n}{n}$

- $\mathbf{E}[M_n] =$

- $\mathsf{Var}(M_n) =$

  $\mathbf{P}(|M_n - \mu| \geq \epsilon) \leq$

The Weak Law of Large Numbers (WLLN)

- $X_1, X_2, \ldots$ i.i.d.; finite mean μ and variance σ^2

  Sample mean: $M_n = \dfrac{X_1 + \cdots + X_n}{n}$ $\qquad \mu = E[X_i]$

- $\mathbf{E}[M_n] = \dfrac{E[X_1 + \cdots + X_n]}{n} = \dfrac{n\mu}{n} = \mu$

- $\mathrm{Var}(M_n) = \dfrac{\mathrm{Var}(X_1 + \cdots + X_n)}{n^2} = \dfrac{n\sigma^2}{n^2} = \dfrac{\sigma^2}{n}$

  $\mathbf{P}(|M_n - \mu| \geq \epsilon) \leq \dfrac{\mathrm{var}(M_n)}{\epsilon^2} = \dfrac{\sigma^2}{n\epsilon^2} \longrightarrow 0$ $\qquad n \to \infty$ $\qquad$ (fixed $\epsilon > 0$)

WLLN: For $\epsilon > 0$, $\mathbf{P}\Big(|M_n - \mu| \geq \epsilon\Big) = \mathbf{P}\Big(\Big|\dfrac{X_1 + \cdots + X_n}{n} - \mu\Big| \geq \epsilon\Big) \to 0$, as $n \to \infty$

Interpreting the WLLN

$$M_n = (X_1 + \cdots + X_n)/n$$

WLLN: For $\epsilon > 0$, $\mathbf{P}\big(|M_n - \mu| \geq \epsilon\big) = \mathbf{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0$, as $n \to \infty$

- One experiment
  - many measurements $X_i = \mu + W_i$
  - W_i: measurement noise; $\mathbf{E}[W_i] = 0$; independent W_i
  - **sample mean** M_n is unlikely to be far off from **true mean** μ

Interpreting the WLLN

$$M_n = (X_1 + \cdots + X_n)/n$$

WLLN: For $\epsilon > 0$, $\mathbf{P}\big(|M_n - \mu| \geq \epsilon\big) = \mathbf{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0$, as $n \to \infty$

- One experiment
  - many measurements $X_i = \mu + W_i$
  - W_i: measurement noise; $\mathbf{E}[W_i] = 0$; independent W_i
  - **sample mean** M_n is unlikely to be far off from **true mean** μ
- Many independent repetitions of the same experiment
  - event A, with $p = \mathbf{P}(A)$
  - X_i: indicator of event A

Interpreting the WLLN

$$M_n = (X_1 + \cdots + X_n)/n$$

WLLN: For $\epsilon > 0$, $\mathbf{P}\big(|M_n - \mu| \geq \epsilon\big) = \mathbf{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0$, as $n \to \infty$

- One experiment
  - many measurements $X_i = \mu + W_i$
  - W_i: measurement noise; $\mathbf{E}[W_i] = 0$; independent W_i
  - **sample mean** M_n is unlikely to be far off from **true mean** μ
- Many independent repetitions of the same experiment
  - event A, with $p = \mathbf{P}(A)$
  - X_i: indicator of event A
  - the sample mean M_n is the **empirical frequency** of event A
  - empirical frequency is unlikely to be far of from true probability p

The pollster's problem

- p: fraction of population that will vote "yes" in a referendum
- ith (randomly selected) person polled: $X_i = \begin{cases} 1, & \text{if yes,} \\ 0, & \text{if no.} \end{cases}$
- $M_n = (X_1 + \cdots + X_n)/n$: fraction of "yes" in our sample
- Would like "small error," e.g.: $|M_n - p| < 0.01$
- Try $n = 10,000$
- $\mathbf{P}\Big(|M_{10,000} - p| \geq 0.01\Big) \leq$

The pollster's problem

- p: fraction of population that will vote "yes" in a referendum
- ith (randomly selected) person polled: $X_i = \begin{cases} 1, & \text{if yes,} \\ 0, & \text{if no.} \end{cases}$

uniformly, independently

$E[X_i] = p$

- $M_n = (X_1 + \cdots + X_n)/n$: fraction of "yes" in our sample
- Would like "small error," e.g.: $|M_n - p| < 0.01$
- Try $n = 10,000$

- $\mathbf{P}\big(|M_{10,000} - p| \geq 0.01\big) \leq \dfrac{\sigma^2}{n\varepsilon^2} = \dfrac{p(1-p)}{10^4 \cdot 10^{-4}} \leq \dfrac{1}{4}$ ← want $\leq 5\%$

$\dfrac{1/4}{n\,10^{-4}} \leq \dfrac{5}{10^2} \iff n \geq \dfrac{10^6}{20} = 50,000$ ← will suffice