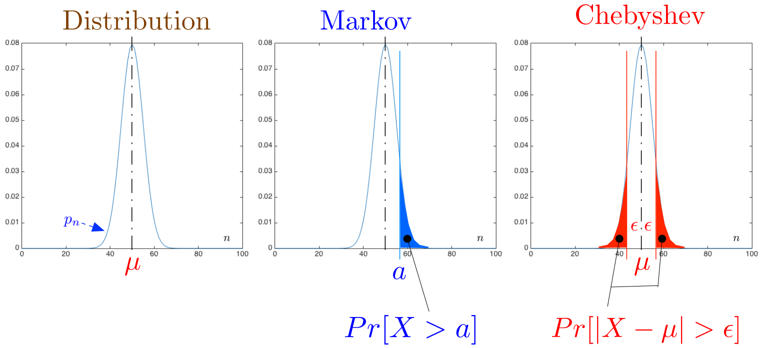
CS70: Lecture 33.

$\boxed{\text{WLLN, Confidence Intervals (CI): Chebyshev vs. CLT}}$

1. Review: Inequalities: Markov, Chebyshev
2. Law of Large Numbers
3. Review: CLT
4. Confidence Intervals: Chebyshev vs. CLT

Inequalities: An Overview



Markov Inequality

If X can only take non-negative values then

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

for all $a > 0$.

This inequality makes no assumptions on the existence of variance and so it can't be very strong for typical distributions. In fact, it is quite weak.

Chebyshev Inequality

If X is a random variable with finite mean and variance σ^2, then

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

for all $c > 0$.
Also, letting $c = k\sigma$:

$$P(|X - E[X]| \geq k\sigma) \leq \frac{1}{k^2}$$

Fraction of H's

Here is a classical application of Chebyshev's inequality.

How likely is it that the fraction of H's differs from 50%?

Let $X_m = 1$ if the m-th flip of a fair coin is H and $X_m = 0$ otherwise.

Define

$$M_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$

We want to estimate

$$Pr[|M_n - 0.5| \geq 0.1] = Pr[M_n \leq 0.4 \text{ or } M_n \geq 0.6].$$

By Chebyshev,

$$Pr[|M_n - 0.5| \geq 0.1] \leq \frac{var[M_n]}{(0.1)^2} = 100 var[M_n].$$

Now,

$$var[M_n] = \frac{1}{n^2}(var[X_1] + \cdots + var[X_n]) = \frac{1}{n} var[X_1] \leq \frac{1}{4n}.$$

$$Var(X_i) = p(1-p) \leq (.5)(.5) = \frac{1}{4}$$

Fraction of H's

$$M_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$

$$Pr[|M_n - 0.5| \geq 0.1] \leq \frac{25}{n}.$$

For $n = 1,000$, we find that this probability is less than 2.5%.

As $n \to \infty$, this probability goes to zero.

In fact, for any $\varepsilon > 0$, as $n \to \infty$, the probability that the fraction of Hs is within $\varepsilon > 0$ of 50% approaches 1:

$$Pr[|M_n - 0.5| \leq \varepsilon] \to 1.$$

This is an example of the (Weak) Law of Large Numbers.

We look at a general case next.

Weak Law of Large Numbers

We perform an experiment n times independently and

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

The fact that $var(M_n) \to 0$ at rate $\frac{1}{n}$ is great but what does that tell us about $P(|M_n - E[X_i|])$? How quickly does it go to zero?
Just use Chebyshev: $P(|X - E[X]| \geq c) \leq \frac{\sigma^2}{c^2}$

$$P(|M_n - E[X_i]) \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

for any $\epsilon > 0$.
This is a form of the Weak Law of Large Numbers.

Weak Law of Large Numbers

Theorem Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be pairwise independent with the same distribution and mean μ. Then, for all $\varepsilon > 0$,

$$Pr[|\frac{X_1 + \cdots + X_n}{n} - \mu| \geq \varepsilon] \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Proof:
Let $M_n = \frac{X_1 + \cdots + X_n}{n}$. Then

$$\begin{aligned} Pr[|M_n - \mu| \geq \varepsilon] &\leq \frac{var[M_n]}{\varepsilon^2} = \frac{var[X_1 + \cdots + X_n]}{n^2 \varepsilon^2} \\ &= \frac{n var[X_1]}{n^2 \varepsilon^2} = \frac{var[X_1]}{n \varepsilon^2} \rightarrow 0, \text{ as } n \rightarrow \infty. \end{aligned}$$

$\square$

What does the Weak Law Really Mean?

WLLN: $\lim_{n\to\infty} P(|M_n - \mu| \geq \epsilon) = 0$.
Just using the defn of limit: For any $\epsilon, \delta > 0$, there exists a number $n(\epsilon, \delta)$ such that

$$P(|M_n - \mu| \geq \epsilon) \leq \delta \quad \text{for all } n \geq n(\epsilon, \delta)$$

- δ:Confidence level
- ϵ: "Error"
- $n(\epsilon, \delta)$: threshold function for a given level of confidence and accuracy

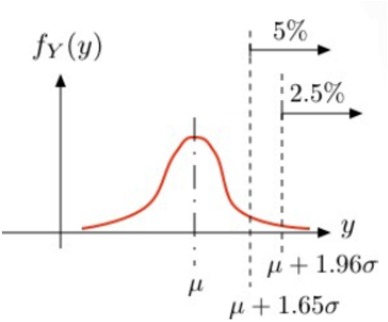
What this is saying is that if we compute M_n for large n then:
Almost Always, $|M_n - \mu| < \epsilon$.
We say that M_n **converges to** μ **in probability**.

Recap: Normal (Gaussian) Distribution.

For any μ and σ, a **normal** (aka **Gaussian**) random variable Y, which we write as $Y = \mathcal{N}(\mu, \sigma^2)$, has pdf

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2}.$$

Standard normal has $\mu = 0$ and $\sigma = 1$.



Note: $Pr[|Y - \mu| > 1.65\sigma] = 10\%$; $Pr[|Y - \mu| > 2\sigma] = 5\%$.

Recap: Central Limit Theorem

Central Limit Theorem
Let $X_1, X_2, \ldots$ be i.i.d. with $E[X_1] = \mu$ and $var(X_1) = \sigma^2$. Define

$$S_n := \frac{A_n - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then,

$$S_n \to \mathcal{N}(0,1), \text{as } n \to \infty.$$

That is,

$$Pr[S_n \le \alpha] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-x^2/2} dx.$$

$$E(S_n) = \frac{1}{\sigma/\sqrt{n}} (E(A_n) - \mu) = 0$$

$$Var(S_n) = \frac{1}{\sigma^2/n} Var(A_n) = 1.$$

Confidence Interval (CI) for Mean: CLT

Let $X_1, X_2, \dots$ be i.i.d. with mean μ and variance σ^2. Let

$$A_n = \frac{X_1 + \cdots + X_n}{n}.$$

The CLT states that

$$\frac{A_n - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \to \mathcal{N}(0,1) \text{ as } n \to \infty.$$

Thus, for $n \gg 1$, one has

$$Pr[-2 \le (\frac{A_n - \mu}{\sigma/\sqrt{n}}) \le 2] \approx 95\%.$$

Equivalently,

$$Pr[\mu \in [A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}]] \approx 95\%.$$

That is,

$$[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } \mu.$$

CI for Mean: CLT vs. Chebyshev

Let $X_1, X_2, \ldots$ be i.i.d. with mean μ and variance σ^2. Let

$$A_n = \frac{X_1 + \cdots + X_n}{n}.$$

The CLT states that

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \to \mathcal{N}(0,1) \text{ as } n \to \infty.$$

Also,

$$[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } \mu.$$

What would Chebyshev's bound give us?

$$[A_n - 4.5\frac{\sigma}{\sqrt{n}}, A_n + 4.5\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } \mu.(Why?)$$

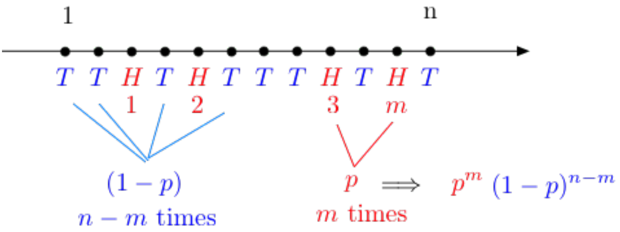
Thus, the CLT provides a smaller confidence interval.

Coins and CLT.

Let $X_1, X_2, \ldots$ be i.i.d. $B(p)$. Thus, $X_1 + \cdots + X_n = B(n,p)$.

Here, $\mu = p$ and $\sigma = \sqrt{p(1-p)}$. CLT states that

$$\frac{X_1 + \cdots + X_n - np}{\sqrt{p(1-p)n}} \to \mathcal{N}(0,1).$$



$\binom{n}{m}$ outcomes with m Hs and $n-m$ Ts

$$\implies \boxed{Pr[X = m] = \binom{n}{m} p^m (1-p)^{n-m}}$$

Coins and CLT.

Let $X_1, X_2, \ldots$ be i.i.d. $B(p)$. Thus, $X_1 + \cdots + X_n = B(n,p)$.
Here, $\mu = p$ and $\sigma = \sqrt{p(1-p)}$. CLT states that

$$\frac{X_1 + \cdots + X_n - np}{\sqrt{p(1-p)n}} \to \mathcal{N}(0,1)$$

and

$$[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } \mu$$

with $A_n = (X_1 + \cdots + X_n)/n$.
Hence,

$$[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } p.$$

Since $\sigma \leq 0.5$,

$$[A_n - 2\frac{0.5}{\sqrt{n}}, A_n + 2\frac{0.5}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } p.$$

Thus,

$$[A_n - \frac{1}{\sqrt{n}}, A_n + \frac{1}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } p.$$

Comparing Chebyshev and CLT: Polling

We ask n randomly sampled voters whether they support Bob.
$X_i = 1$ if the i^{th} voter says "yes" and $X_i = 0$ otherwise. The X_i are iid.
We want to be sure with prob ≥ 0.95 that $|M_{100} - p| \leq 0.1$. How many people should we ask?
Again, use the bound that $\mathrm{var}(X_i) \leq \frac{1}{4}$
By Chebyshev:

$$\frac{25}{n} \leq 0.05 \Rightarrow \boxed{n \geq 500}$$

By CLT:

$$2(1 - \phi(2 * 0.1 * \sqrt{n})) \leq 0.05$$

$$\phi(2 * 0.1 * \sqrt{n}) \geq 0.975$$

Since $\phi(1.96) = 0.975$:

$$\boxed{n \geq 96.04}$$

CLT much better than Chebyshev.

Summary

$\boxed{\text{Inequalities and Confidence Interals}}$

1. Inequalities: Markov and Chebyshev Tail Bounds
2. Weak Law of Large Numbers
3. Confidence Intervals: Chebyshev Bounds vs. CLT Approx.
4. CLT: X_n i.i.d. $\Longrightarrow \frac{A_n - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1)$
5. CI: $[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] =$ 95%-CI for μ.