The Central Limit Theorem, Confidence Intervals

Lec.25

August 4, 2020

CLT Fundamental Idea

For i.i.d. random variables X_i each with mean μ and variance σ^2, $S_n = \sum_{i=1}^{n} X_i$.

While the LLN tells us S_n/n is unlikely to be far from the true mean μ as $n \to \infty$, the Central Limit Theorem tells us that the *distribution* of $\frac{S_n}{n}$ approaches $\mathcal{N}(\mu, \frac{\sigma^2}{n})$.

Note that here the limiting distribution depends on the value of n; we can standardize $\frac{S_n}{n}$ so that the limiting distribution is the standard normal distribution and does not change with n.

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \frac{1}{n}\mathbb{E}[S_n] = \frac{1}{n} \cdot n \cdot \mu = \mu.$$

$$\mathrm{Var}\left[\frac{S_n}{n}\right] = \frac{1}{n^2}\mathrm{Var}[S_n] = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

CLT Definition

$$\frac{\frac{S_n}{n} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{S_n - n\mu}{\sigma \cdot \sqrt{n}}$$

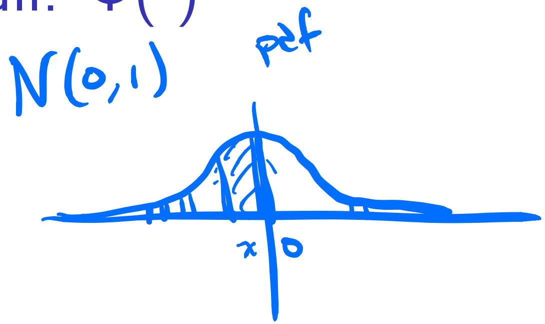
Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with common finite expectation $\mathbb{E}[X_i] = \mu$ and finite variance $\mathrm{Var}[X_i] = \sigma^2$. Let $S_n = \sum_{i=1}^{n} X_i$.

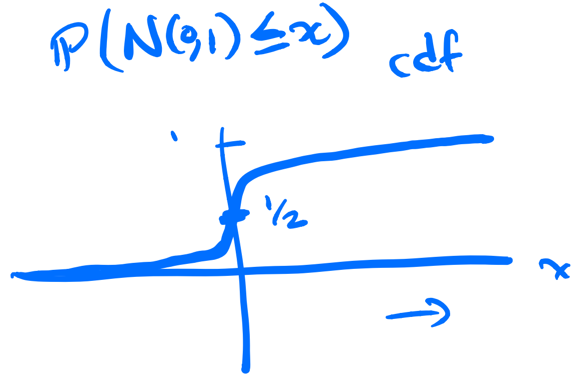
Then, the distribution of, $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ converges to $\mathcal{N}(0,1)$ as $n \to \infty$. In other words, for any constant $c \in \mathbb{R}$,

$$\underbrace{\Pr\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq c\right]}_{\text{cdf}} \to \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} e^{-x^2/2} dx}_{\text{cdf of standard normal } \mathcal{N}(0,1)} \quad \text{as } n \to \infty.$$

Don't need to know how to prove CLT, just need to know how to use it.

Recall: $\Phi(\cdot)$





$\Phi(\cdot)$ is the cdf of the standard normal random variable.

What the CLT states is that the cdf of the standardized sample mean of the X_i converges to $\Phi(\cdot)$ as $n \to \infty$, regardless of the distribution of the X_i as long as their mean and variance are finite.

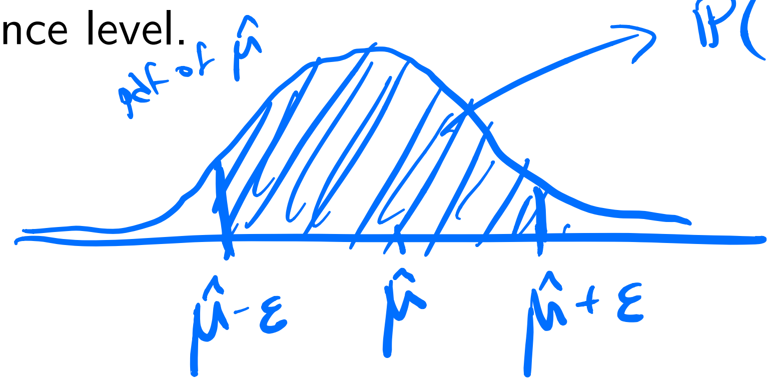
Confidence Intervals

$$\hat{\mu} = \frac{S_n}{n}$$

Provide a confidence level that the true parameter μ is with a certain range of the estimated parameter:

$$P(|\hat{\mu} - \mu| \le \epsilon) \ge 1 - \delta$$

We can think of ϵ as the error in our estimate, and $1 - \delta$ as our confidence level.



$\mathbb{P}(|\hat{\mu}-\mu| \le \varepsilon)$ want this to be greater than or equal to some confidence.

Example: Polling

You can poll people in a population as to whether or not they approve of the current president. X_i is 1 if person i approves of the current president, and 0 otherwise. We model $X_i \sim Bernoulli(\mu)$. You want to estimate μ, the underlying proportion of the population that approves of the current president. You want to know how many people you need to poll in order to be 95% confident that you are within 0.03 of the true proportion μ.

Let $\hat{\mu}_n = \frac{\sum_1^n X_i}{n}$ sample mean.

$\hat{\mu}_n = \frac{S_n}{n} = \frac{\sum_1^n X_i}{n}$

$$\mathbb{E}[\hat{\mu}_n] = \frac{1}{n} \cdot \mathbb{E}[S_n] = \frac{1}{n} \cdot n \cdot \mu = \mu.$$

$$Var[\hat{\mu}_n] = \frac{1}{n^2} \cdot Var(S_n) = \frac{n \cdot \mu(1-\mu)}{n^2}$$

Example: Polling

You can poll people in a population as to whether or not they approve of the current president. X_i is 1 if person i approves of the current president, and 0 otherwise. We model $X_i \sim \mathit{Bernoulli}(\mu)$. You want to estimate μ, the underlying proportion of the population that approves of the current president. You want to know how many people you need to poll in order to be 95% confident that you are within 0.03 of the true proportion μ.

Let $\hat{\mu}_n = \frac{\sum_{i=1}^n X_i}{n}$ be our *sample mean*.

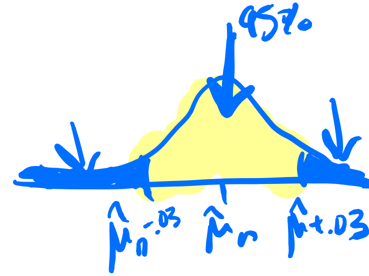
$$\mathbb{E}[\hat{\mu}_n] = \mathbb{E}[\frac{\sum_{i=1}^n X_i}{n}] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \frac{n\mu}{n} = \mu \tag{1}$$

$$\mathrm{Var}[\hat{\mu}_n] = \mathrm{Var}[\frac{\sum_{i=1}^n X_i}{n}] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}[X_i] \tag{2}$$

$$= \frac{n\mu \cdot (1-\mu)}{n^2} = \frac{\mu \cdot (1-\mu)}{n} \tag{3}$$

Example: Polling - Chebyshev

$$P(\underbrace{|\hat{\mu}_n - \mu|} \geq \underbrace{0.03}_{c}) \leq \overbrace{\underbrace{\frac{Var(\hat{\mu}_n)}{0.03^2}}}^{\leq 0.05}$$



$$\frac{\frac{\mu(1-\mu)}{n}}{0.03^2} \leq 0.05$$

$$\underline{1112} \cdot \frac{\mu(1-\mu)}{n} \leq 0.05$$

$$n \geq 1112 \cdot \mu(1-\mu) \cdot 20$$

In the worst case, $\underline{\mu(1-\mu)} = \frac{1}{4}$

$$n \geq 5560$$

Note: $X_i \sim Bern(\mu)$

$\Rightarrow Var(X_i) = \underline{\mu(1-\mu)}$

$0 \leq \mu \leq 1$

Example: Polling - Chebyshev

For 95% confidence, the sample mean can deviate from the true mean by 0.03 or more with 5% probability,

$$P(|\hat{\mu}_n - \mu| > 0.03) \leq \frac{\frac{\mu\cdot(1-\mu)}{n}}{0.03^2} \leq 0.05 \tag{4}$$

$$1112\frac{\mu(1-\mu)}{n} \leq 0.05 \tag{5}$$

In the worst case (worst means more people required), $\mu(1-\mu)$ is as large as possible. The max value of $\mu(1-\mu) = \frac{1}{4}$. So,

$$1112 * \frac{1}{4n} \leq 0.05 \tag{6}$$

$$\rightarrow 20 * 1112 \leq 4n \tag{7}$$

$$\rightarrow 5 * 1112 \leq n \tag{8}$$

$$\rightarrow 5560 \leq n \tag{9}$$

$$\tag{10}$$

Standard Normal Tails

Φ is the standard normal cdf.

$= \Phi(\alpha)$

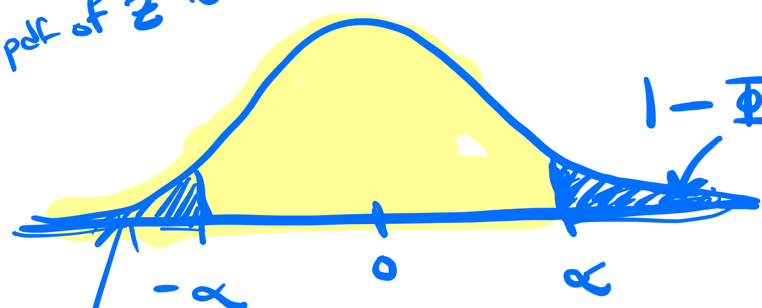
$= \mathbb{P}(Z \le \alpha)$

Let $Z \sim \mathcal{N}(0,1)$ and $\alpha \geq 0$. Then,

$$P(|Z| > \alpha) = 2(1 - \Phi(\alpha))$$

Proof:

pdf of $Z \sim \mathcal{N}(0,1)$



2 tails $\rightarrow 2(1-\Phi(\alpha))$

$1-\Phi(\alpha)$ (by symmetry)

Example: Polling - CLT

$\hat{\mu}_n = \frac{X_1 + \ldots + X_n}{n} = \frac{S_n}{n}$

$Z = \frac{\hat{\mu}_n - E[\hat{\mu}_n]}{\sqrt{Var[\hat{\mu}_n]}}$

$\frac{\frac{S_n}{n} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$

$$\boxed{Z = \frac{\hat{\mu}_n - \mu}{\sqrt{\frac{\mu\cdot(1-\mu)}{n}}} \to Z \sim \mathcal{N}(0, 1)}$$

5% or lower

95% or higher

$\hat{\mu}-0.03 \quad \hat{\mu} \quad \hat{\mu}+.03$

error

$P(|\hat{\mu}_n - \mu| > 0.03) = P\left(\underbrace{\frac{|\hat{\mu}_n - \mu|}{\sqrt{\frac{\mu(1-\mu)}{n}}}} > \frac{0.03}{\sqrt{\frac{\mu(1-\mu)}{n}}}\right)$

$= P\left(|Z| > \frac{0.03}{\sqrt{\frac{\mu(1-\mu)}{n}}}\right) \le 0.05$

So $P\left(|Z| > \frac{0.03\sqrt{n}}{\frac{1}{2}}\right) \le 0.05$

$\sqrt{\mu(1-\mu)}$ is at most $\sqrt{\frac{1}{4}} = \frac{1}{2}$

Let $\alpha = 0.06\sqrt{n}$, then

$P(|Z| > \alpha) \le 0.05$

$2(1 - \Phi(\alpha)) \le 0.05$

$\Rightarrow \Phi(\alpha) \ge 0.975$

$\alpha \ge 1.96$ (from table)

$0.06\sqrt{n} \ge 1.96$

$\boxed{n \ge 1068}$

Example: Polling - CLT

$$Z = \frac{\hat{\mu}_n - \mu}{\sqrt{\frac{\mu \cdot (1-\mu)}{n}}} \rightarrow Z \sim \mathcal{N}(0,1)$$

$$P(|\hat{\mu}_n - \mu| > 0.03) = P\left(\left|\frac{\hat{\mu}_n - \mu}{\sqrt{\frac{\mu \cdot (1-\mu)}{n}}}\right| > \frac{0.03}{\sqrt{\frac{\mu \cdot (1-\mu)}{n}}}\right) \tag{11}$$

$$= P\left(|Z| > \frac{0.03\sqrt{n}}{\sqrt{\mu \cdot (1-\mu)}}\right) \leq 0.05 \tag{12}$$

If we make the denominator larger, n will need to be larger in order to meet the confidence requirement. In the worst case, the denominator is $\frac{1}{2}$.

Example: Polling - CLT

So, in the worst case we need to satisfy:

$$P(|Z| > \frac{0.03\sqrt{n}}{1/2}) \leq 0.05$$
$$P(|Z| > 0.06\sqrt{n}) \leq 0.05$$

Let $\alpha = 0.06\sqrt{n}$, then

$$P(|Z| > \alpha) \leq 0.05 \tag{13}$$
$$2(1 - \Phi(\alpha)) \leq 0.05 \tag{14}$$
$$\Phi(\alpha) \geq 0.975 \tag{15}$$
$$\rightarrow \alpha = 1.96 \tag{16}$$
$$\rightarrow n \geq 1068 \tag{17}$$

Standard Normal CDF Table

Introduction to Probability, 2nd Ed, by D. Bertsekas and J. Tsitsiklis, Athena Scientific, 2008

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

The standard normal table. The entries in this table provide the numerical values of $\Phi(y) = \mathbf{P}(Y \le y)$, where Y is a standard normal random variable, for y between 0 and 3.49. For example, to find $\Phi(1.71)$, we look at the row corresponding to 1.7 and the column corresponding to 0.01, so that $\Phi(1.71) = .9564$. When y is negative, the value of $\Phi(y)$ can be found using the formula $\Phi(y) = 1 - \Phi(-y)$.

$.975 = \Phi(\alpha)$

$\to \alpha = 1.96$

$\Phi(\alpha) = .975$

2 min break.

Markov Chains Preview

$X_0, X_1, X_2, \ldots$ sequence of random variables

Think of X_n as the state of a system at time n.

↳ Assume that 'time' is discrete.
X_0 is the state @ time 0

↳ Each X_i will take on values in some finite state space $\mathcal{X}$.
Ex: If $\mathcal{X} = \{1, 2\}$
Then X_0 can either be 1 or 2.

Consider $X_0, X_1, \ldots X_n$.
want $\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots X_n = x_n)$.

easy:

Say X_i are independent
Then this joint
$= \mathbb{P}(X_0 = x_0) \cdot \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n)$

hard.

All the X_i have complex dependencies with one another.
Then we actually need access to the entire joint pmf
need $\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots X_n = x_n)$
for every $x_0, x_1, \ldots x_n$.

Markov Property

Think of X_n as the present/current state, and X_{n+1} as the future state. The Markov Property is

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, \ldots X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

$$\mathbb{P}(X_{n+1}=j \mid X_n = i, \ldots, X_0 = i) = \mathbb{P}(X_{n+1}=j \mid X_n$$

"The future is conditionally independent of the past given the current state"