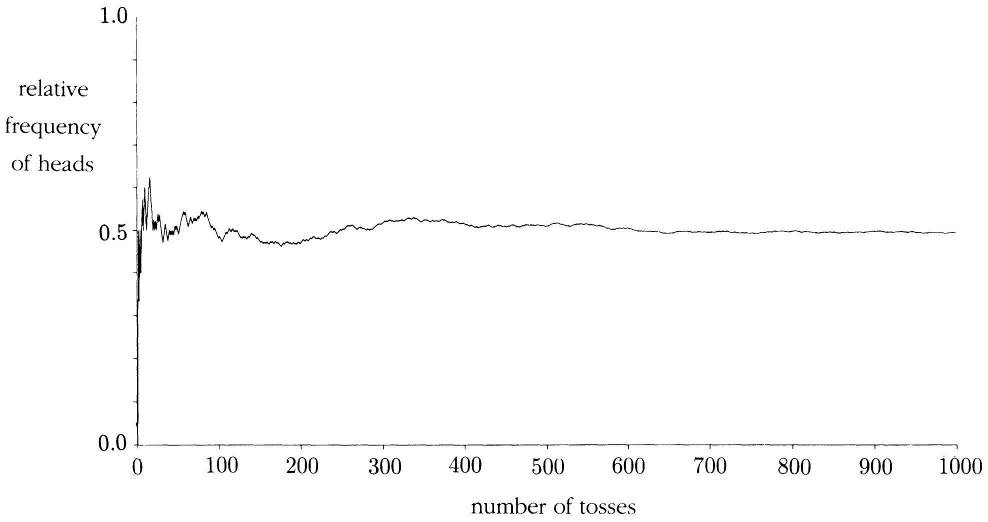
CS145: Probability & Computing

Lecture 10: Law of Large Numbers, Central Limit Theorem, Finite Sample Bounds

Figure credits:
*Bertsekas & Tsitsiklis, **Introduction to Probability**, 2008*
*Pitman, **Probability**, 1999*

CS145: Lecture 10 Outline

- Laws of Large Numbers
- Central Limit Theorem
- Finite Sample Bounds



Convergence in Probability

Convergence in Probability

Let $Y_1, Y_2, \ldots$ be a sequence of random variables (not necessarily independent), and let a be a real number. We say that the sequence Y_n **converges to a in probability**, if for every $\epsilon > 0$, we have

$$\lim_{n\to\infty} \mathbf{P}\big(|Y_n - a| \geq \epsilon\big) = 0.$$

"(almost all) of the PMF/PDF of Y_n eventually gets concentrated (arbitrarily) close to a"

Convergence of a Deterministic Sequence

Let $a_1, a_2, \ldots$ be a sequence of real numbers, and let a be another real number. We say that the sequence a_n converges to a, or $\lim_{n\to\infty} a_n = a$, if for every $\epsilon > 0$ there exists some n_0 such that

$$|a_n - a| \leq \epsilon, \qquad \text{for all } n \geq n_0.$$

"a_n eventually gets and stays (arbitrarily) close to a"

Convergence in Probability

Convergence in Probability

Let $Y_1, Y_2, \ldots$ be a sequence of random variables (not necessarily independent), and let a be a real number. We say that the sequence Y_n **converges to** a **in probability**, if for every $\epsilon > 0$, we have

$$\lim_{n\to\infty} \mathbf{P}\big(|Y_n - a| \geq \epsilon\big) = 0.$$

Example:

X_n is a sequence of independent uniform variables on $[0,1]$ and $Y_n = \min\{X_1, \ldots, X_n\}$

- We expect that Y_n converges to zero. To verify:

$$\begin{aligned}
\mathbf{P}\Big(|Y_n - 0| \geq \epsilon\Big) &= \mathbf{P}(X_1 \geq \epsilon, \ldots, X_n \geq \epsilon) \\
&= \mathbf{P}(X_1 \geq \epsilon) \cdots \mathbf{P}(X_n \geq \epsilon) \\
&= (1-\epsilon)^n.
\end{aligned}$$

The Weak Law of Large Numbers

Theorem

Let $x_1, \ldots, x_n$ *be independent, identically distributed random variables with finite mean,* $E[x_i] = \mu$. *For any* $\epsilon > 0$

$$Prob\{|\frac{1}{n}\sum_{i=1}^{n} x_i - \mu| \geq \epsilon\} \rightarrow 0$$

as $n \rightarrow \infty$.

The (Weak) Law of Large Numbers

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

$$M_n = \frac{X_1 + \cdots + X_n}{n}$$

sample mean or empirical mean

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

$$\mathrm{Var}[M_n] = \frac{\mathrm{var}(X_1 + \cdots + X_n)}{n^2} = \frac{\mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

- Chebyshev's inequality bounds distance between the true mean and the "empirical" or "sample" mean:

$$\mathbf{P}(|M_n - \mu| \geq \epsilon) \ \leq \ \frac{\mathrm{Var}(M_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

- *The empirical mean converges to the true mean in probability*

$$\lim_{n \to \infty} P(|M_n - \mu| \geq \epsilon) = 0$$

- True even if variance not finite, but proof more challenging.

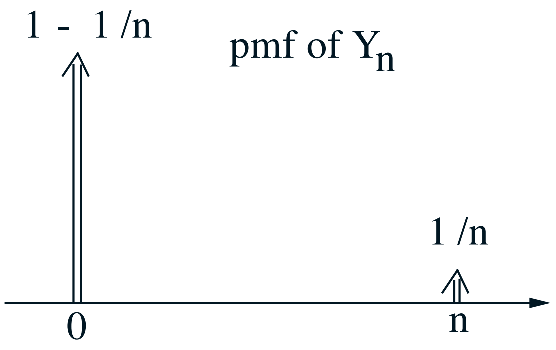
Why is it a "Weak" Law of Large Numbers?

Convergence in Probability

Let $Y_1, Y_2, \ldots$ be a sequence of random variables (not necessarily independent), and let a be a real number. We say that the sequence Y_n **converges to** a **in probability**, if for every $\epsilon > 0$, we have

$$\lim_{n\to\infty} \mathbf{P}\big(|Y_n - a| \geq \epsilon\big) = 0.$$

Example:



$$\text{For every } \epsilon > 0, \lim_{n\to\infty} P(|Y_n - 0| \geq \epsilon) = 0.$$

But even though Y_n converges in probability, occasionally it takes on very large values:

$$E[Y_n] = 1 \text{ for all } n.$$

The Strong Law of Large Numbers

Theorem

Let $x_1, \ldots, x_n$ be independent, identically distributed random variables with finite mean, $E[x_i] = \mu$.

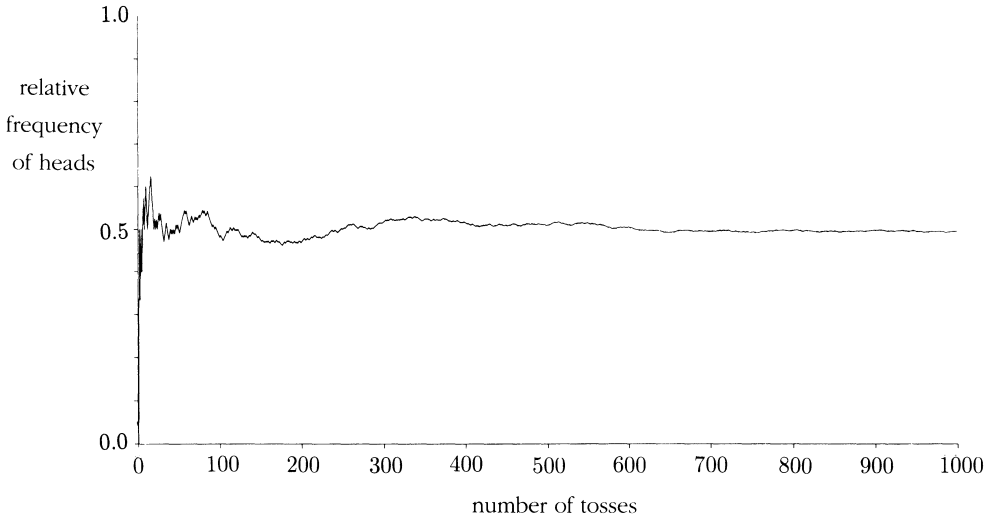
$$Prob\{\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i = \mu\} = 1.$$

- This stronger (but more technically challenging) notion of convergence rules out cases like the previous example
- For many practical scenarios, both forms of convergence hold, but convergence in probability is easier to show
- We focus exclusively on the weak law in this course

CS145: Lecture 10 Outline

- Laws of Large Numbers
- Central Limit Theorem
- Finite Sample Bounds

Convergence to the Mean



Convergence to the Mean

Let $X_1, \ldots, X_n$ with $Pr(X = -1) = Pr(X = 1) = 1/2$.

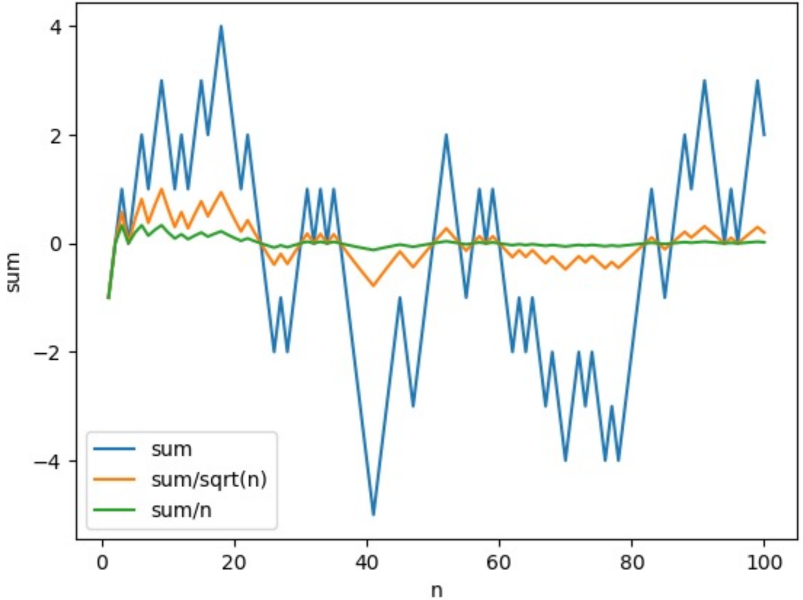
$E[X_i] = 0$. $Var[X_i] = \sigma[X_i] = 1$

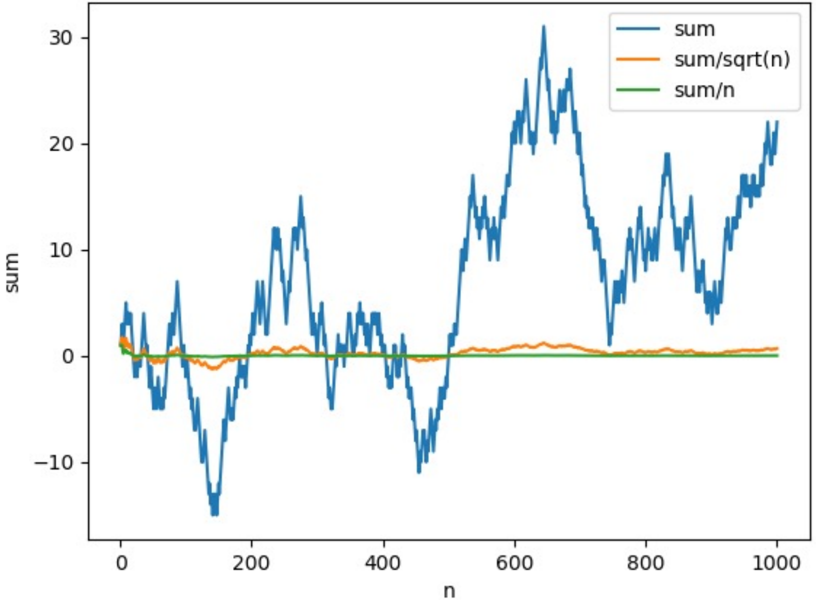
Sum: $X_n = \sum_{i=1}^{n} X_i$

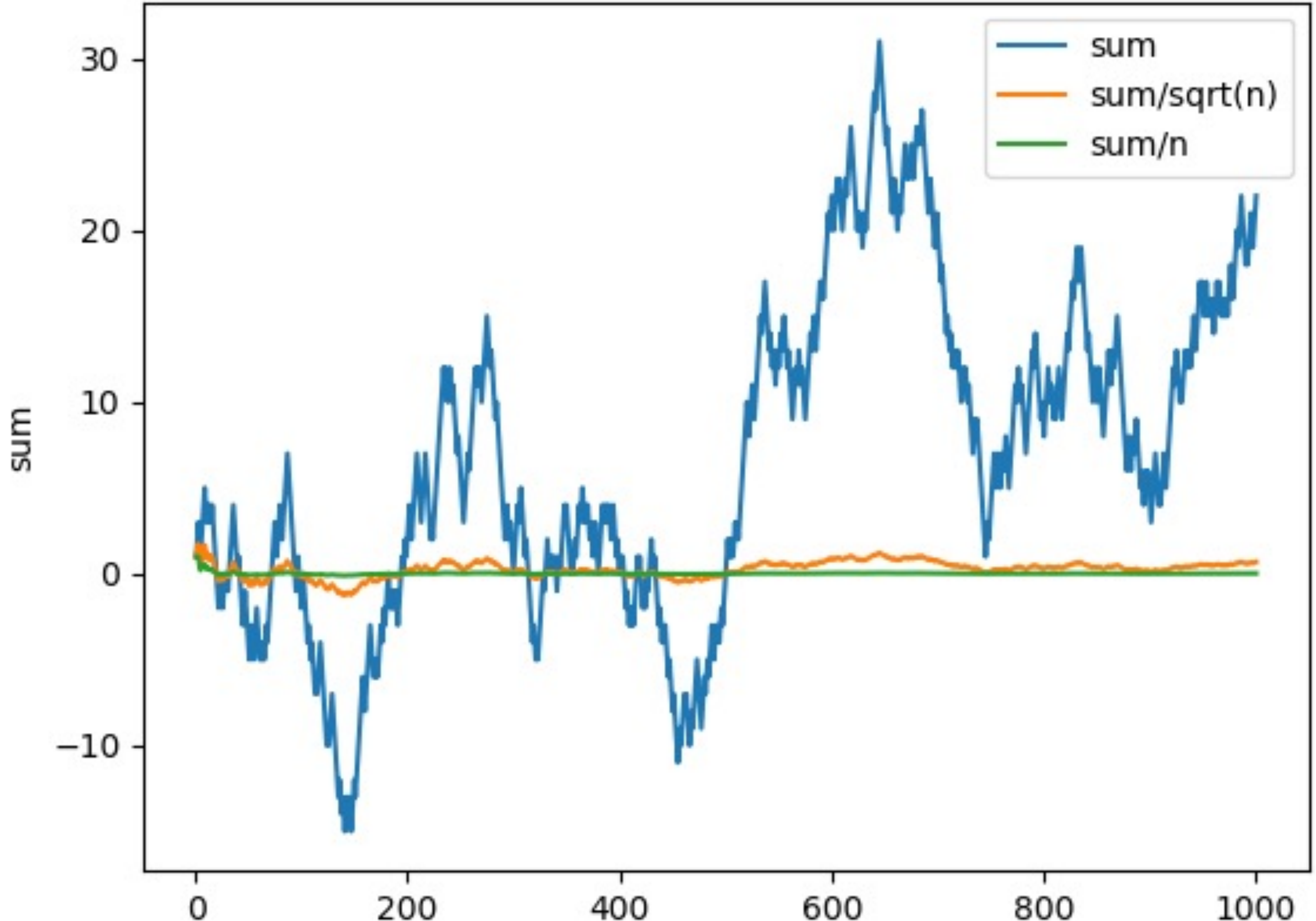
Sum/sqrt(n): $Y_n = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i}{\sigma[\sum_{i=1}^{n} X_i]}$

Sum/n: $Z_n = (\sum_{i=1}^{n} X_i)/n$

$$E[X_n] = E[Y_n] = E[Z_n] = 0$$
$$Var[X_n] = n$$
$$Var[Y_n] = 1$$
$$Var[Z_n] = 1/n$$






sum
sum/sqrt(n)
sum/n
sum
30
20
10
0
−10
0
200
400
600
800
1000

Scaling of the Sample Mean

- Sequence of *independent, identically distributed* random variables:

$$X_1, X_2, \ldots, X_n \qquad E[X_i] = \mu \qquad \mathrm{Var}[X_i] = \sigma^2 < \infty$$

- The variance of their *sum* increases with *n*:

$$S_n = \sum_{i=1}^n X_i \qquad E[S_n] = n\mu \qquad \mathrm{Var}[S_n] = n\sigma^2$$

- Law of Large Numbers: variance of the *empirical mean* decreases with *n*:

$$M_n = \frac{1}{n} S_n \qquad E[M_n] = \mu \qquad \mathrm{Var}[M_n] = \frac{\sigma^2}{n}$$

- *Standardized sum:* transform so mean and variance constant for all *n*

$$Z_n = \frac{S_n - E[S_n]}{\sqrt{\mathrm{Var}[S_n]}} = \frac{S_n - n\mu}{\sqrt{n}\sigma} \qquad E[Z_n] = 0 \qquad \mathrm{Var}[Z_n] = 1$$

What is the shape of the distribution of Z_n for large n?

Central Limit Theorem (CLT)

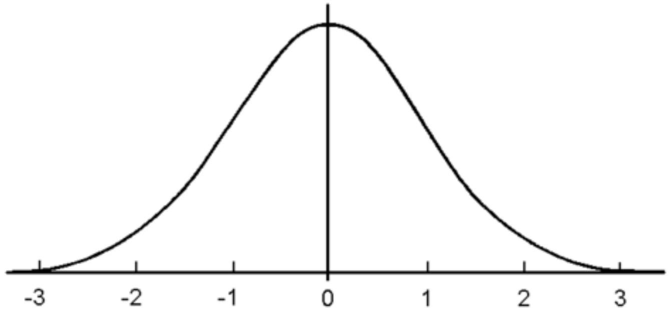
- "Standardized" $S_n = X_1 + \cdots + X_n$:

$$Z_n = \frac{S_n - \mathbf{E}[S_n]}{\sigma_{S_n}} = \frac{S_n - n\mathbf{E}[X]}{\sqrt{n}\,\sigma}$$

  - zero mean
  - unit variance

- Let Z be a standard normal r.v. (zero mean, unit variance)

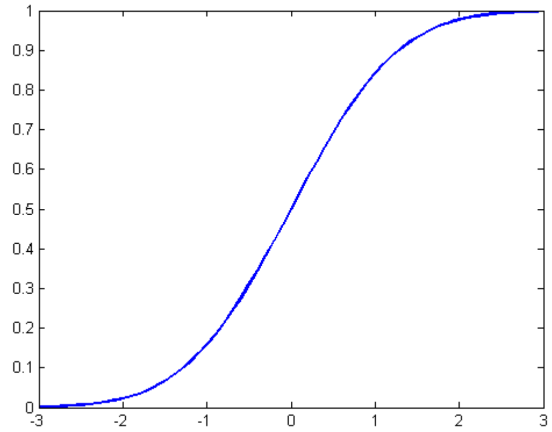
$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$



- **Theorem:** For every c:

$$\mathbf{P}(Z_n \le c) \to \mathbf{P}(Z \le c)$$

- $\mathbf{P}(Z \le c)$ is the standard normal CDF, $\Phi(c)$, available from the normal tables



Central Limit Theorem (CLT)

Usefulness

- universal; only means, variances matter
- accurate computational shortcut
- justification of normal models

What exactly does it say?

- CDF of Z_n converges to normal CDF
  - not a statement about convergence of PDFs or PMFs
- Treat Z_n as if normal
  - also treat S_n as if normal

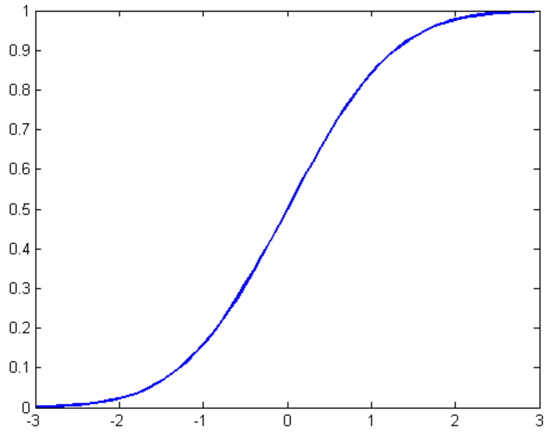
Can we use it when n is "moderate"?

- Yes, but no nice theorems to this effect
- Symmetry helps a lot

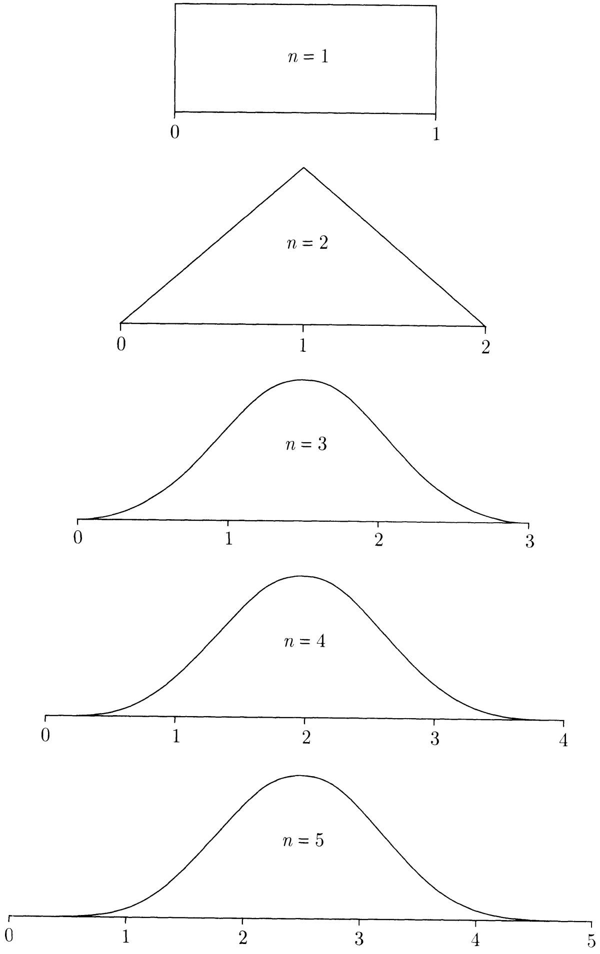
- **Theorem:** For every c:

$$\mathbf{P}(Z_n \leq c) \to \mathbf{P}(Z \leq c)$$

- $\mathbf{P}(Z \leq c)$ is the standard normal CDF, $\Phi(c)$, available from the normal tables



CLT: Uniform Random Variables



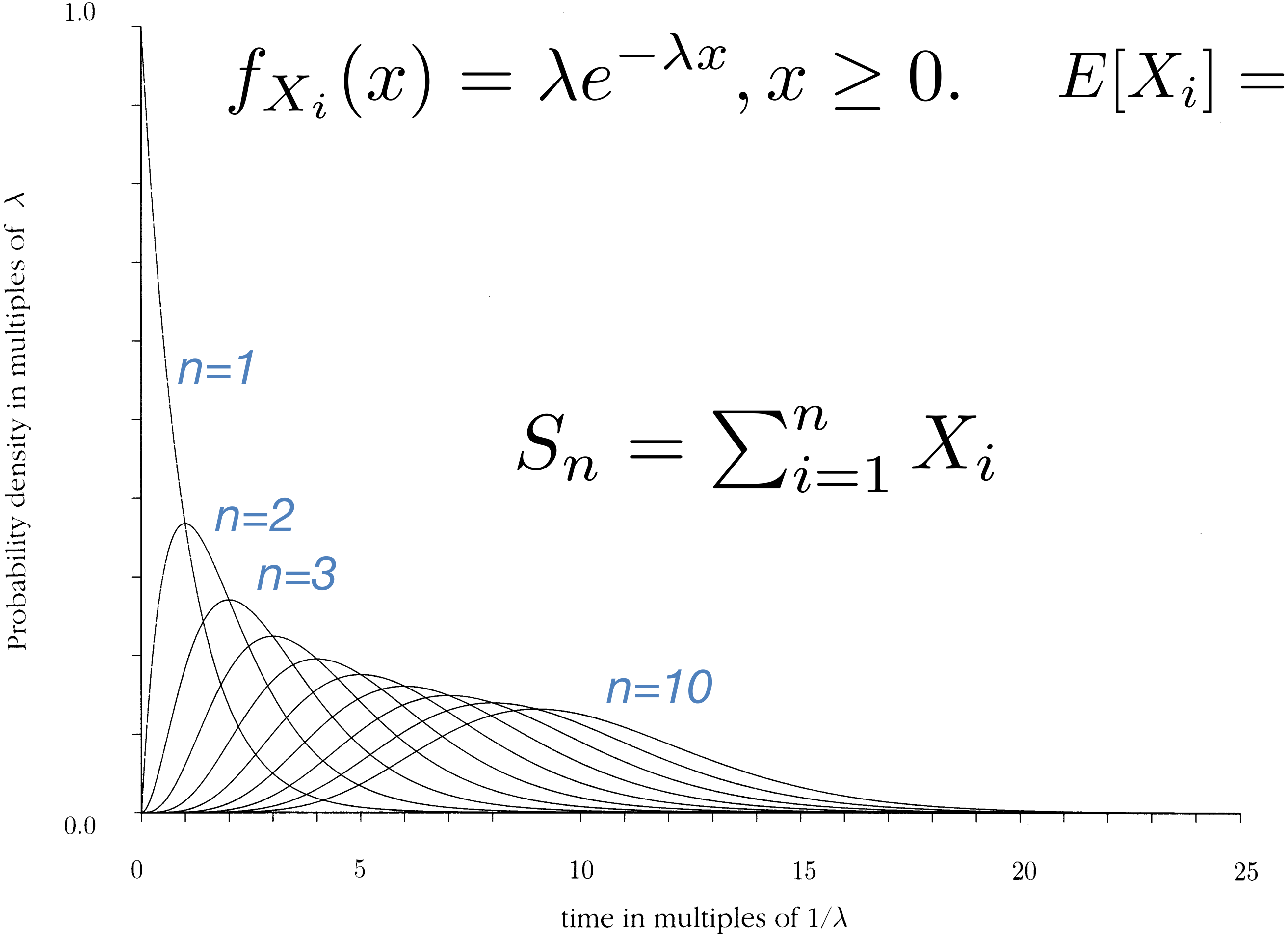
$$S_n = \sum_{i=1}^{n} X_i$$

$$f_{X_i}(x_i) = 1 \text{ if } 0 \le x_i \le 1, 0 \text{ otherwise.}$$

CLT: Exponential Random Variables

$$f_{X_i}(x) = \lambda e^{-\lambda x}, x \geq 0. \quad E[X_i] = \frac{1}{\lambda} \quad \mathrm{Var}[X_i] = \frac{1}{\lambda^2}$$

$$S_n = \sum_{i=1}^{n} X_i$$



Basic Central Limit Theorem

Theorem (DeMoivre-Laplace-Liapounoff)

Let $x_1, \ldots, x_n$ *be* n *independent, identically distributed random variables with mean* μ *and variance* σ^2. *Let* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, *then*

$$P\left(a \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \to \Phi(b) - \Phi(a)$$

as $n \to \infty$,

Proof of Central Limit Theorem

We need:

Lemma

Let $Z_1, Z_2, \ldots.$ be a sequence of random variables with distributions F_n and moment generating functions M_n. Let Z be a random variable with distribution F and moment generating function M. If $M_n(t) \to M(t)$ for all t then $F_n \to F$ for all X in which $F(X)$ is continuous.

By this lemma if $M_n \to e^{t^2/2}$ then Z has a $N(0,1)$ distribution.

Proof of Central Limit Theorem

Assume first that $x_1, ..., x_n$ such that for all i $E[x_i] = 0$, and $Var[x_i] = 1$.
The moment generation function of $x_i/\sqrt{n}$ is

$$E[e^{tx_i/\sqrt{n}}] = M(\frac{t}{\sqrt{n}}).$$

Thus,

$$E[e^{t\sum_{i=1}^{n} x_i/\sqrt{n}}] = (M(\frac{t}{\sqrt{n}}))^n.$$

Let $L(t) = \log M(t)$
$M(0) = 1$, $L(0) = 0$, $L'(0) = \frac{M'(0)}{M(0)} = E[x_i] = 0$.

$$L''(0) = \frac{M(0)M''(0) - (M'(0))^2}{(M(0))^2} = E[x_i^2] = 1$$

Proof of Central Limit Theorem

We need to show that that $(M(\frac{t}{\sqrt{n}}))^n \to e^{t^2/2}$ or $nL(t/\sqrt{n}) \to t^2/2$ as $n \to \infty$.

$$
\begin{aligned}
\lim_{n\to\infty} \frac{L(t/\sqrt{n})}{n^{-1}} &= \lim_{n\to\infty} \frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}} \\
&= \lim_{n\to\infty} \frac{L'(t/\sqrt{n})t}{2n^{-1/2}} \\
&= \lim_{n\to\infty} \frac{-L''(t/\sqrt{n})n^{-3/2}t^2}{-2n^{-3/2}} \\
&= \lim_{n\to\infty} L''(t/\sqrt{n})\frac{t^2}{2} \\
&= \frac{t^2}{2}
\end{aligned}
$$

Applying L'Hospital's rule (twice).

More General Version of CLT

Theorem

Let $x_1, ..., x_n$ *be* n *independent random variable, with* $E[x_i] = \mu_i$ *and* $Var[x_i] = \sigma_i^2$. *Assume that*

1. *For some value* M, $P(|x_i| < M) = 1$ *for all* i;
2. $\sum_{i=1}^n \sigma_i^2 \to \infty$;

then,

$$P\left(a \leq \frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \leq b\right) \to \Phi(b) - \Phi(a)$$

as $n \to \infty$.

Pollster's Problem: Chebyshev

- f: fraction of population that "..."
- 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
- Goal: 95% confidence of ≤1% error

$$\mathbf{P}(|M_n - f| \geq .01) \leq .05$$

- Use Chebyshev's inequality:

$$\begin{aligned} \mathbf{P}(|M_n - f| \geq .01) &\leq \frac{\sigma^2_{M_n}}{(0.01)^2} \\ &= \frac{\sigma^2_x}{n(0.01)^2} \leq \frac{1}{4n(0.01)^2} \end{aligned}$$

For any binary variable,

$$\mathrm{Var}(X_i) \leq \frac{1}{2^2}$$

- If $n = 50,000$,
  then $\mathbf{P}(|M_n - f| \geq .01) \leq .05$
  (conservative)

Pollster's Problem: CLT

- f: fraction of population that "..."
- ith (randomly selected) person polled:

$$X_i = \begin{cases} 1, & \text{if yes,} \\ 0, & \text{if no.} \end{cases}$$

- Event of interest: $|M_n - f| \geq .01$

$$\left| \frac{X_1 + \cdots + X_n - nf}{n} \right| \geq .01$$

$$\left| \frac{X_1 + \cdots + X_n - nf}{\sqrt{n}\sigma} \right| \geq \frac{.01\sqrt{n}}{\sigma}$$

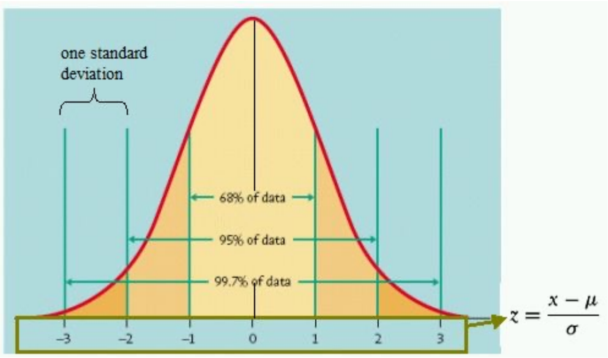
Find the smallest n such that
$P(|M_n - f| \geq .01) = Pr(|Z| \geq \frac{.01\sqrt{n}}{\sigma}) \leq 0.05$

- $M_n = (X_1 + \cdots + X_n)/n$
  fraction of "yes" in our sample
- Goal: 95% confidence of $\leq$1% error

$$\mathbf{P}(|M_n - f| \geq .01) \leq .05$$

For any binary variable,

$$\mathrm{Var}(X_i) \leq \frac{1}{2^2} \qquad \mathrm{Std}(X_i) \leq \frac{1}{2}$$



Pollster's Problem: CLT

Find the smallest n such that
$P(|M_n - f| \geq .01) = Pr(|Z| \geq \frac{.01\sqrt{n}}{\sigma}) \leq 0.05$

$$\begin{aligned} Pr(|Z| &\geq \frac{.01\sqrt{n}}{\sigma}) \\ &= Pr(Z \geq -\frac{.01\sqrt{n}}{\sigma}) + Pr(Z \geq \frac{.01\sqrt{n}}{\sigma}) \\ &= 2\left(1 - Pr(Z \leq \frac{.01\sqrt{n}}{\sigma})\right) \leq 0.05 \end{aligned}$$

We need $Pr(Z \leq \frac{.01\sqrt{n}}{\sigma}) \geq 0.975 \approx \Phi(1.96)$
$\frac{.01\sqrt{n}}{\sigma} \geq 1.96 \quad \Rightarrow \quad n \geq 10000$

Probability Content from -oo to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

How the Poll Was Conducted

The latest New York Times/CBS News poll is based on telephone interviews conducted Feb. 24-27 with 984 adults throughout the United States.

The sample of land-line telephone exchanges called was randomly selected by a computer from a complete list of more than 69,000 active residential exchanges across the country. The exchanges were chosen to ensure that each region of the country was represented in proportion to its population.

Within each exchange, random digits were added to form a complete telephone number, thus permitting access to listed and unlisted numbers alike. Within each household, one adult was designated by a random procedure to be the respondent for the survey.

To increase coverage, this land-line sample was supplemented by respondents reached through random dialing of cellphone numbers. The two samples were then combined.

Interviewers made multiple efforts to reach every phone number in the survey, calling back unanswered numbers on different days at different times of both day and evening.

The combined results have been weighted to adjust for variation in the sample relating to geographic region, sex, race, Hispanic origin, marital status, age, education and number of adults in the household. In addition, the land-line respondents were weighted to take account of the number of telephone lines into the residence, while the cellphone respondents were weighted according to whether they were reachable only by cellphone or also by land line.

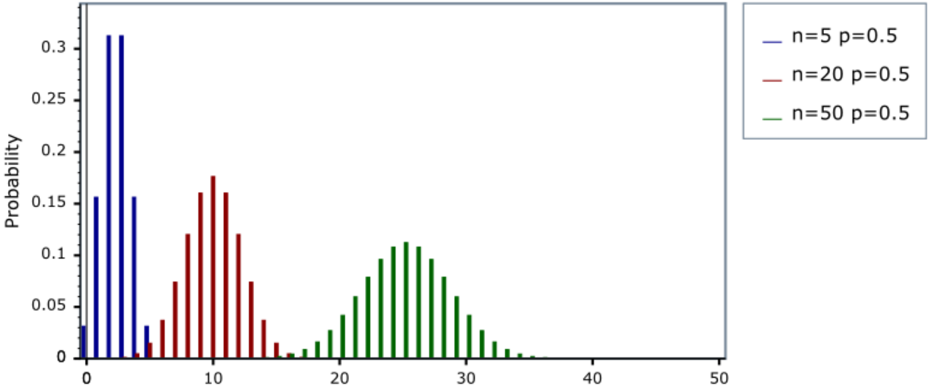
In theory, in 19 cases out of 20, overall results based on such samples will differ by no more than three percentage points in either direction from what would have been obtained by seeking to interview all American adults. For smaller subgroups, the margin of sampling error is larger. Shifts in results between polls over time also have a larger sampling error.

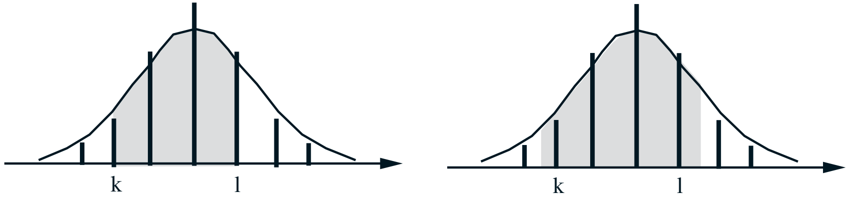
In addition to sampling error, the practical difficulties of conducting any survey of public opinion may introduce other sources of error into the poll. Variation in the wording and order of questions, for example, may lead to somewhat different results.

Complete questions and results are available at nytimes.com/polls.

CLT: Binomial Distribution

- Fix p, where $0 < p < 1$
- X_i: Bernoulli(p)
- $S_n = X_1 + \cdots + X_n$: Binomial(n, p)
  - mean np, variance $np(1-p)$
- CDF of $\dfrac{S_n - np}{\sqrt{np(1-p)}} \longrightarrow$ standard normal





De Moivre – Laplace Approximation to the Binomial

If S_n is a binomial random variable with parameters n and p, n is large, and k, ℓ are nonnegative integers, then

$$\mathbf{P}(k \le S_n \le \ell) \approx \Phi\left(\frac{\ell + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

- $\mathbf{P}(S_n \le 21) = \mathbf{P}(S_n < 22)$, because S_n is integer
- Compromise: consider $\mathbf{P}(S_n \le 21.5)$

CLT: Binomial Distribution

- When the 1/2 correction is used, CLT can also approximate the binomial p.m.f. (not just the binomial CDF)

$$\mathbf{P}(S_n = 19) = \mathbf{P}(18.5 \leq S_n \leq 19.5)$$

$$n = 36, p = 0.5$$

$$18.5 \leq S_n \leq 19.5 \quad \iff$$

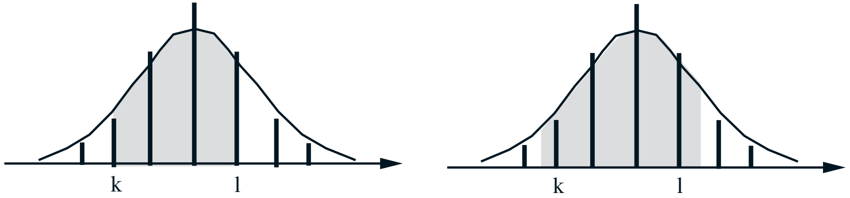
$$\frac{18.5 - 18}{3} \leq \frac{S_n - 18}{3} \leq \frac{19.5 - 18}{3} \quad \iff$$

$$0.17 \leq Z_n \leq 0.5$$

$$\begin{aligned} \mathbf{P}(S_n = 19) &\approx \mathbf{P}(0.17 \leq Z \leq 0.5) \\ &= \mathbf{P}(Z \leq 0.5) - \mathbf{P}(Z \leq 0.17) \\ &= 0.6915 - 0.5675 \\ &= 0.124 \end{aligned}$$

- Exact answer:

$$\binom{36}{19}\left(\frac{1}{2}\right)^{36} = 0.1251$$



De Moivre – Laplace Approximation to the Binomial

If S_n is a binomial random variable with parameters n and p, n is large, and k, ℓ are nonnegative integers, then

$$\mathbf{P}(k \leq S_n \leq \ell) \approx \Phi\left(\frac{\ell + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

- $\mathbf{P}(S_n \leq 21) = \mathbf{P}(S_n < 22)$, because S_n is integer

- Compromise: consider $\mathbf{P}(S_n \leq 21.5)$

CS145: Lecture 10 Outline

- Laws of Large Numbers
- Central Limit Theorem
- Finite Sample Bounds -*Advanced topic not covered in homeworks or exams!*

Large Deviation Bound – The Basic Idea

Advanced topic not covered in homeworks or exams!

Theorem

[Markov Inequality] *For any non-negative random variable, and for all* $a > 0$,

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

Theorem

[Chebyshev's Inequality] *For* **any** *random variable* X, *and any* $a > 0$,

$$Pr(|X - E[X]| \geq a) \leq \frac{Var[X]}{a^2}.$$

$$\text{For any } t > 0,\ Pr(X \geq a) \leq \frac{E[e^{tx}]}{e^{ta}}$$

Example – Chernoff Bound

Advanced topic not covered in homeworks or exams!

Theorem: Let $X_1, ..., X_n$ be independent random variables with

$$Pr(X_i = 1) = Pr(X_i = -1) = \frac{1}{2}.$$

Let $X = \sum_1^n X_i$. For any $a > 0$,

$$Pr(X \geq a) \leq e^{-\frac{a^2}{2n}}.$$

de Moivre – Laplace approximation: For any k, such that $|k - np| \leq a$

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{a^2}{2np(1-p)}}$$

Proof:

For any $t > 0$, $E[e^{tX_i}] = \frac{1}{2}e^t + \frac{1}{2}e^{-t}$.
$e^t = 1 + t + \frac{t^2}{2!} + \cdots + \frac{t^i}{i!} + \dots$
and
$e^{-t} = 1 - t + \frac{t^2}{2!} + \cdots + (-1)^i \frac{t^i}{i!} + \dots$

$$E[e^{tX_i}] \quad = \quad \frac{1}{2}e^t + \frac{1}{2}e^{-t} = \sum_{i \geq 0} \frac{t^{2i}}{(2i)!} \leq \sum_{i \geq 0} \frac{(\frac{t^2}{2})^i}{i!} = e^{t^2/2}$$

$$E[e^{tX}] = \prod_{i=1}^{n} E[e^{tX_i}] \leq e^{nt^2/2},$$

$$Pr(X \geq a) = Pr(e^{tX} > e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}} \leq e^{t^2 n/2 - ta}.$$

Setting $t = a/n$ yields $Pr(X \geq a) \leq e^{-\frac{a^2}{2n}}$.
By symmetry $Pr(|X| > a) \leq 2e^{-\frac{a^2}{2n}}$.

Asymptotic vs Bounded Sample

Advanced topic not covered in homeworks or exams!

Theorem (DeMoivre-Laplace-Liapounoff)

Let $x_1, ..., x_n$ be n independent, identically distributed random variables with mean μ and variance σ^2. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, then

$$P(a \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq b) \rightarrow \Phi(b) - \Phi(a)$$

as $n \rightarrow \infty$,

Theorem (Hoeffding's Inequality)

Let $X_1, \ldots, X_n$ be independent random variables such that for all $1 \leq i \leq n$, $E[X_i] = \mu$ and $Pr(a \leq X_i \leq b) = 1$. Then

$$Pr(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \geq \epsilon) \leq 2e^{-2n\epsilon^2/(b-a)^2}$$

Comparison of Bounded Sample Bounds

Advanced topic not covered in homeworks or exams!

Let $X_1, \ldots, X_n$**, independent, with** $Pr(X=-1) = Pr(X=1) = 1/2$**.**
$E[X_i] = 0$**.** $Var[X_i] = \sigma[X_i] = 1$

$Y = \sum_{i=1}^{n} X_i$, $E[Y] = 0$, $Var[Y] = n$

Chebyshev's Inequality: $P(|Y| \geq a) \leq \frac{Var[Y]}{a^2} = \frac{n}{a^2}$

For $a = \sqrt{n \log n}$, $P(|Y| \geq a) \leq \frac{1}{\log n}$

Chernoff Bound: $P(|Y| \geq a) \leq 2e^{-\frac{a^2}{2n}}$

For $a = \sqrt{n \log n}$, $P(|Y| \geq a) \leq \frac{2}{\sqrt{n}}$

Hoeffding's Bound: $P(|Y| \geq a) \leq 2e^{-\frac{2a^2}{4}}$

For $a = \sqrt{n \log n}$, $P(|Y| \geq a) \leq \frac{2}{\sqrt{n}}$