Unit 8 Limit Theorems and Classical Statistics

Inequalities, Convergence, and the Weak Law of Large Numbers

Weak Law of Large Numbers
Taking the average of more numbers converges to the actual mean
(Taking different sample means)

The definition of convergence here is a bit different as we are talking about a random variable converging to a number

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 $\mathbf{P}(X \geq a) \leq \dfrac{\mathbf{E}[X]}{a}$.

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

$$\geq \int_a^\infty a\, f_X(x)\,dx = a\,P(X \geq a)$$

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

$$Y = \begin{cases} 0, & \text{if } X < a \\ a, & \text{if } X \geq a \end{cases} \qquad a\,P(X \geq a) = E[Y] \leq E[X]$$

Assumes random variable is nonnegative

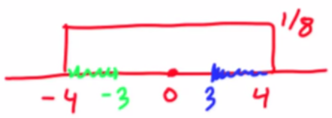
The Markov inequality

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($\lambda = 1$): $\mathbf{P}(X \geq a) \leq \dfrac{1}{a}$

- **Example:** X is Uniform$[-4, 4]$: $\mathbf{P}(X \geq 3) \leq$ $P(|X| \geq 3) \leq \dfrac{E[|X|]}{3} = \dfrac{2}{3}$

$= \dfrac{1}{2} P(|X| \geq 3) \leq \dfrac{1}{3}$



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

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

c is positive here

The Chebyshev inequality

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

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

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

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

Probability of being with k standard deviations is true no matter the distribution
In most cases, Chebyshev inequality is more telling than the Markov inequality because it also uses information about the variance

Exercise: Markov inequality

0/1 point (graded)

Let Z be a nonnegative random variable that satisfies $\mathbf{E}[Z^4] = 4$. Apply the Markov inequality to the random variable Z^4 to find the tightest possible (given the available information) upper bound on $\mathbf{P}(Z \geq 2)$.

$\mathbf{P}(Z \geq 2) \leq$ 1/8 ✖ **Answer:** 0.25

Solution:

We have

$$\mathbf{P}(Z \geq 2) = \mathbf{P}(Z^4 \geq 16) \leq \frac{\mathbf{E}[Z^4]}{16} = \frac{4}{16} = \frac{1}{4}.$$

Exercise: Chebyshev inequality

0/1 point (graded)

Let Z be normal with zero mean and variance equal to 4. For this case, the Chebyshev inequality yields:

$\mathbf{P}(|Z| \geq 4) \leq$ 1 ✖ **Answer:** 0.25

Solution:

We have

$$\mathbf{P}(|Z| \geq 4) \leq \frac{\mathsf{Var}(Z)}{4^2} = \frac{4}{4^2} = \frac{1}{4}.$$

Exercise: Chebyshev versus Markov

2/2 points (graded)

Let X be a random variable with zero mean and finite variance. The Markov inequality applied to $|X|$ yields

$$\mathbf{P}\big(\, |X| \geq a\big) \leq \frac{\mathbf{E}\big[\,|X|\,\big]}{a},$$

whereas the Chebyshev inequality yields

$$\mathbf{P}\big(\, |X| \geq a\big) \leq \frac{\mathbf{E}[X^2]}{a^2}.$$

a) Is it true that the Chebyshev inequality is stronger (i.e., the upper bound is smaller) than the Markov inequality, when a is very large?

Yes ✔ **Answer:** Yes

b) Is it true that the Chebyshev inequality is always stronger (i.e., the upper bound is smaller) than the Markov inequality?

No ✔ **Answer:** No

Solution:

a) Yes, because for very large a, the term $1/a^2$ will be much smaller than $1/a$.

b) No. For example, suppose that $a = 1$. It is certainly possible to have $\mathbf{E}[X^2] > \mathbf{E}\big[\,|X|\,\big]$, in which case the Markov inequality provides a stronger bound.

The Weak Law of Large Numbers (WLLN)

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

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

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

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

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

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

Sample mean is a random variable as it is a sum of random variables
It is the simplest way of trying to estimate the true mean

Interpreting the WLLN

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

WLLN: For $\epsilon > 0$, $\mathbf{P}\big(|M_n - \mu| \geq \epsilon\big) = \mathbf{P}\left(\left|\dfrac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0$, 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 $p = \mathbf{P}(A)$ $\qquad E[X_i] = p$
  - X_i: indicator of event A $\qquad X_i = 1$, if A occurs; 0, o.w.
  - the sample mean M_n is the **empirical frequency** of event A

Far off refers to a distance of epsilon

Exercise: Sample mean bounds

1/2 points (graded)

By the argument in the last video, if the X_i are i.i.d. with mean μ and variance σ^2, and if $M_n = (X_1 + \cdots + X_n)/n$, then we have an inequality of the form

$$\mathbf{P}\left(|M_n - \mu| \geq \epsilon\right) \leq \frac{a\sigma^2}{n},$$

for a suitable value of a.

a) If $\epsilon = 0.1$, then the value of a is: 10 ✖ **Answer:** 100

b) If we change $\epsilon = 0.1$ to $\epsilon = 0.1/k$, for $k \geq 1$ (i.e., if we are interested in k times higher accuracy), how should we change n so that the value of the upper bound does not change from the value calculated in part (a)?

n should

- stay the same
- increase by a factor of k
- increase by a factor of k^2
- decrease by a factor of k
- none of the above

✔

Solution:

a) Chebyshev's inequality yields

$$\mathbf{P}\left(|M_n - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2},$$

so that $a = 1/\epsilon^2 = 1/0.1^2 = 100.$

b) In order to keep the same upper bound, the term $n\epsilon^2$ in the denominator needs to stay constant. If we reduce ϵ by a factor of k, then ϵ^2 gets reduced by a factor of k^2. Thus, n will have to be increased by a factor of k^2.

The pollster's problem

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

  uniformly, independently

- $M_n = (X_1 + \cdots + X_n)/n$: fraction of "yes" in our sample
- Would like "small error," e.g.: $|M_n - p| < 0.01$
- Try $n = 10{,}000$
- $\mathbf{P}\big(|M_{10{,}000} - p| \geq 0.01\big) \leq \frac{\sigma^2}{n\varepsilon^2} = \frac{p(1-p)}{10^4 \cdot 10^{-4}} \leq \frac{1}{4}$ ← want $\leq 5\%$

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

Can't guarantee a error of no more than 1% but can say that the probability of getting a error above this is x%
Here we are saying 1/4 (25%), but that is really high
so we can work backwards and calculate how many people we need to sample to get the probability to less than 5%
But there are more accurate ways of doing this....

Exercise: Polling

2/4 points (graded)

We saw that if we want to have a probability of at least 95% that the poll results are within 1 percentage point of the truth, Chebyshev's inequality recommends a sample size of $n = 50,000$. This is very large compared to what is done in practice. Newspaper polls use smaller sample sizes for various reasons. For each of the following, decide whether it is a valid reason.

In the real world,

a) the accuracy requirements are looser.

Yes ✔ **Answer:** Yes

b) the Chebyshev bound is too conservative.

No ✖ **Answer:** Yes

c) the people sampled are all different, so their answers are not identically distributed.

No ✔ **Answer:** No

d) the people sampled do not have independent opinions.

Yes ✖ **Answer:** No

Solution:

a) Requiring the accuracy to be within one percentage point is too strict for most real world situations.

b) The Chebyshev bound is conservative as stated in the video.

c,d) No matter how opinions get formed, as long as we choose who to ask at random, independently and uniformly, the opinions reported will be i.i.d. random variables, so that the last two considerations do not apply.

Convergence "in probability"

WLLN: For any $\epsilon > 0$, $\mathbf{P}\big(|M_n - \mu| \geq \epsilon\big) \to 0$, as $n \to \infty$

- Would like to say that "M_n converges to μ"

$M_n \xrightarrow[n \to \infty]{i.p.} \mu$

- Need to define the word "converges"
- Sequence of random variables Y_n; not necessarily independent

Definition: A sequence Y_n **converges in probability** to a number a if:

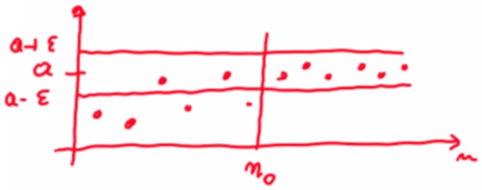
for any $\epsilon > 0$, $\lim_{n\to\infty} \mathbf{P}\big(|Y_n - a| \geq \epsilon\big) = 0$

Understanding convergence "in probability"

- Ordinary convergence
  - Sequence a_n;   number a

    $a_n \to a$

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

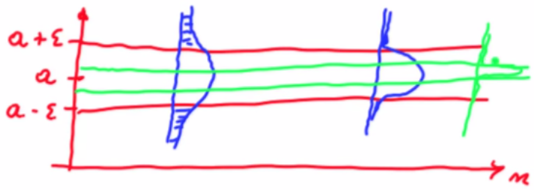


- For every $\epsilon > 0$, there exists n_0, such that for every $n \geq n_0$, we have $|a_n - a| \leq \epsilon$

- Convergence in probability
  - Sequence Y_n;   number a

    $Y_n \to a$

- for any $\epsilon > 0$, $\mathbf{P}\big(|Y_n - a| \geq \epsilon\big) \to 0$



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

The narrower the band of epsilon the longer it will take to get most of the probability in that band

Some properties

- Suppose that $X_n \to a$, $Y_n \to b$, in probability
- If g is continuous, then $g(X_n) \to g(a)$ $X_n^2 \to a^2$
- $X_n + Y_n \to a + b$
- **But:** $\mathbf{E}[X_n]$ need not converge to a

Convergence of random variables does not imply convergence of expected values

Convergence in probability examples

$Y_n \xrightarrow[n\to\infty]{i.p.} 0.$

$\varepsilon > 0 \quad P(|Y_n - 0| \geq \varepsilon) = 1/n \xrightarrow[n\to\infty]{} 0$

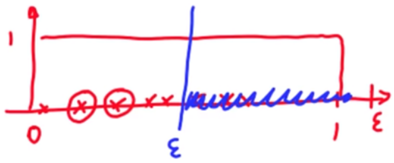
$\mathbf{E}[Y_n] = n^2 \cdot \frac{1}{n} = n \xrightarrow[n\to\infty]{} \infty$

- convergence in probability does **not** imply convergence of expectations

The expected value is very sensitive to the tail but the probability isn't affected when the probability mass of the tail is small

Convergence in probability examples

- X_i: i.i.d., uniform on $[0,1]$
- $Y_n = \min\{X_1, \ldots, X_n\}$



$Y_{n+1} \leq Y_n$

$\varepsilon > 0$

$\varepsilon > 1$

$\varepsilon \leq 1$

$$\begin{aligned}
\mathbf{P}\big(|Y_n - 0| \geq \epsilon\big) &= P(Y_n \geq \varepsilon). \\
&= P(X_1 \geq \varepsilon, \ldots, X_n \geq \varepsilon) \\
&= P(X_1 \geq \varepsilon) \cdots P(X_n \geq \varepsilon) \\
&= (1-\varepsilon)^n \xrightarrow[n\to\infty]{} 0
\end{aligned}$$

$Y_n \xrightarrow[n\to\infty]{i.p.} 0$

Exercise: Convergence in probability

0/3 points (graded)

a) Suppose that X_n is an exponential random variable with parameter $\lambda = n$. Does the sequence $\{X_n\}$ converge in probability?

No ✖ **Answer:** Yes

b) Suppose that X_n is an exponential random variable with parameter $\lambda = 1/n$. Does the sequence $\{X_n\}$ converge in probability?

Yes ✖ **Answer:** No

c) Suppose that the random variables in the sequence $\{X_n\}$ are independent, and that the sequence converges to some number a, in probability. Let $\{Y_n\}$ be another sequence of random variables that are dependent, but where each Y_n has the same distribution (CDF) as X_n. Is it necessarily true that the sequence $\{Y_n\}$ converges to a in probability?

No ✖ **Answer:** Yes

Solution:

a) In the first case, for any $\epsilon > 0$, we have $\mathbf{P}(X_n \geq \epsilon) = e^{-n\epsilon}$, which converges to zero. Therefore, we have convergence in probability.

b) In the second case, for any $\epsilon > 0$, we have $\mathbf{P}(X_n \geq \epsilon) = e^{-\epsilon/n}$, which converges to one. Therefore, we do not have convergence in probability.

c) Dependence will not make a difference because the definition of convergence in probability involves probabilities of the form $\mathbf{P}\big(|Y_n - a| \geq \epsilon\big)$. These probabilities are completely determined by the marginal distributions of the random variables Y_n, and these marginal distributions are the same as for the sequence X_n.

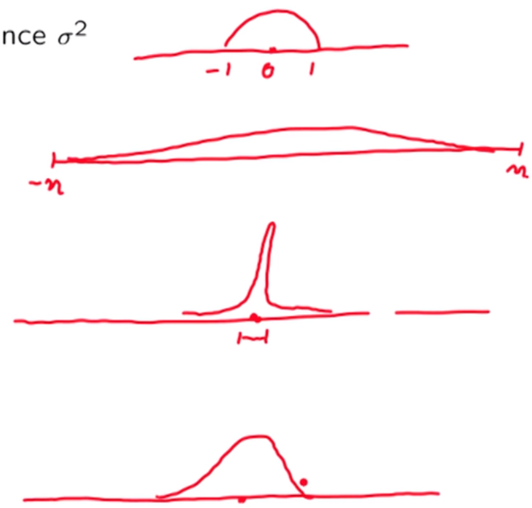
Related topics

- Better bounds/approximations on tail probabilities $\leq \sim \frac{1}{n}$
  - Markov and Chebyshev inequalities
  - Chernoff bound $P(|M_n - \mu| \geq a) \leq e^{-n \underbrace{h(a)}_{>0}}$
  - Central limit theorem "$M_n \sim N(\mu, \sigma^2/n)$"
- Different types of convergence
  - Convergence in probability
  - Convergence "with probability 1" $P(\{w : Y_n(w) \xrightarrow[n\to\infty]{} Y(w)\}) = 1$
  - Strong law of large numbers $M_n \xrightarrow[n\to\infty]{wp1} \mu$
  - Convergence of a sequence of distributions (CDFs) to a limiting CDF

Strong law of large numbers means sample mean will converge to actual mean with probability 1

The Central Limit Theorem

Different scalings of the sum of i.i.d. random variables



- $X_1, \ldots, X_n$ i.i.d., finite mean μ and variance σ^2
- $S_n = X_1 + \cdots + X_n$

  variance: $n\sigma^2$
- $M_n = \frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n}$

  variance: $\frac{\sigma^2}{n} \rightsquigarrow 0$
- $\frac{S_n}{\sqrt{n}} = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$

  variance: $\sigma^2 = \frac{n\sigma^2}{n}$

By dividing by sqrt(n) instead of n then the variance remains constant as we have more n
Instead of concentrating around one value and becoming 0

The Central Limit Theorem (CLT)

- $X_1, \ldots, X_n$ i.i.d., finite mean μ and variance σ^2
- $S_n = X_1 + \cdots + X_n$ variance: $n\sigma^2$
- $\frac{S_n}{\sqrt{n}} = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$ variance: σ^2

$Z_n = \frac{S_n - n\mu}{\sqrt{n}\,\sigma}$ $\mathbf{E}[Z_n] = 0$ $\text{var}(Z_n) = 1$

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

Central Limit Theorem: For every z: $\lim_{n\to\infty} \mathbf{P}(Z_n \le z) = \mathbf{P}(Z \le z)$

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

Exercise: CLT

2/2 points (graded)

Let X_n be i.i.d. random variables with mean zero and variance σ^2. Let $S_n = X_1 + \cdots + X_n$. Let Φ stand for the standard normal CDF. According to the central limit theorem, and as $n \to \infty$, $\mathbf{P}\left(S_n \leq 2\sigma\sqrt{n}\right)$ converges to $\Phi\left(a\right)$, where:

$a =$ 2 ✔ **Answer:** 2

Furthermore,

$\mathbf{P}\left(S_n \leq 0\right)$ converges to: 0.5 ✔ **Answer:** 0.5

(Here, enter the numerical value of the probability.)

Solution:

We have

$$\lim_{n\to\infty} \mathbf{P}\left(S_n \leq 2\sigma\sqrt{n}\right) = \lim_{n\to\infty} \mathbf{P}\left(\frac{S_n - 0}{\sigma\sqrt{n}} \leq 2\right) = \Phi\left(2\right).$$

Similarly,

$$\lim_{n\to\infty} \mathbf{P}\left(S_n \leq 0\right) = \lim_{n\to\infty} \mathbf{P}\left(\frac{S_n - 0}{\sigma\sqrt{n}} \leq 0\right) = \Phi\left(0\right) = \frac{1}{2}.$$

Usefulness of the CLT

$S_n = X_1 + \cdots + X_n \qquad Z_n = \dfrac{S_n - n\mu}{\sqrt{n}\,\sigma} \qquad Z \sim N(0,1)$

Central Limit Theorem: For every z: $\lim_{n\to\infty} \mathbf{P}(Z_n \leq z) = \mathbf{P}(Z \leq z)$

- universal and easy to apply; only means, variances matter
- fairly accurate computational shortcut
- justification of normal models

What exactly does the CLT say? — Theory

$S_n = X_1 + \cdots + X_n$ $\qquad Z_n = \dfrac{S_n - n\mu}{\sqrt{n}\,\sigma}$ $\qquad Z \sim N(0,1)$

Central Limit Theorem: For every z: $\lim_{n\to\infty} \mathbf{P}(Z_n \le z) = \mathbf{P}(Z \le z)$

- CDF of Z_n converges to normal CDF
- results for convergence of PDFs or PMFs (with more assumptions)
- results without assuming that the X_i are identically distributed
- results under "weak dependence"
- proof: uses "transforms": $\mathbf{E}[e^{sZ_n}] \to \mathbf{E}[e^{sZ}]$, for all s

What exactly does the CLT say? — Practice

$S_n = X_1 + \cdots + X_n$ $\qquad Z_n = \dfrac{S_n - n\mu}{\sqrt{n}\,\sigma}$ $\qquad Z \sim N(0,1)$

Central Limit Theorem: For every z: $\lim_{n\to\infty} \mathbf{P}(Z_n \le z) = \mathbf{P}(Z \le z)$

- The **practice** of normal approximations:
  - treat Z_n as if it were normal $\qquad S_n = \sqrt{n}\,\sigma Z_n + n\mu$
  - hence treat S_n as if normal: $\mathcal{N}(n\mu, n\sigma^2)$
- Can we use the CLT when n is "moderate"? $\qquad n = 30\ ?$
  - usually, yes
  - symmetry and unimodality help

unimodal - single peak

Exercise: CLT applicability

1/4 points (graded)

Consider the class average in an exam in a few different settings. In all cases, assume that we have a large class consisting of equally well-prepared students. Think about the assumptions behind the central limit theorem, and choose the most appropriate response under the given description of the different settings.

1. Consider the class average in an exam of a fixed difficulty.

- (selected) The class average is approximately normal
- The class average is not approximately normal because the student scores are strongly dependent
- The class average is not approximately normal because the student scores are not identically distributed

✔

2. Consider the class average in an exam that is equally likely to be very easy or very hard.

- (selected) The class average is approximately normal
- The class average is not approximately normal because the student scores are strongly dependent ✔
- The class average is not approximately normal because the student scores are not identically distributed

✖

3. Consider the class average if the class is split into two equal-size sections. One section gets an easy exam and the other section gets a hard exam.

- The class average is approximately normal ✔
- (selected) The class average is not approximately normal because the student scores are strongly dependent
- The class average is not approximately normal because the student scores are not identically distributed

✖

4. Consider the class average if every student is (randomly and independently) given either an easy or a hard exam.

- The class average is approximately normal ✔
- The class average is not approximately normal because the student scores are strongly dependent
- (selected) The class average is not approximately normal because the student scores are not identically distributed

✖

Solution:

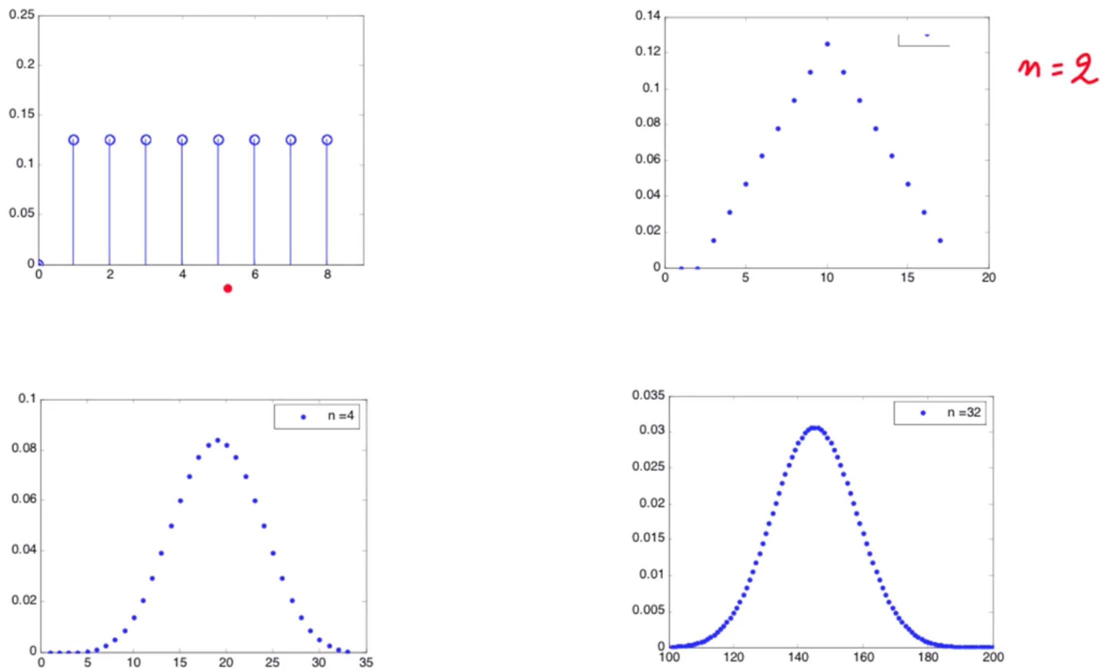
1. Since students are equally well-prepared and the difficulty level is fixed, the only randomness in a student's score comes from luck or accidental mistakes of that student. It is then plausible to assume that each student's score will be an independent random variable drawn from the same distribution, and the CLT applies.

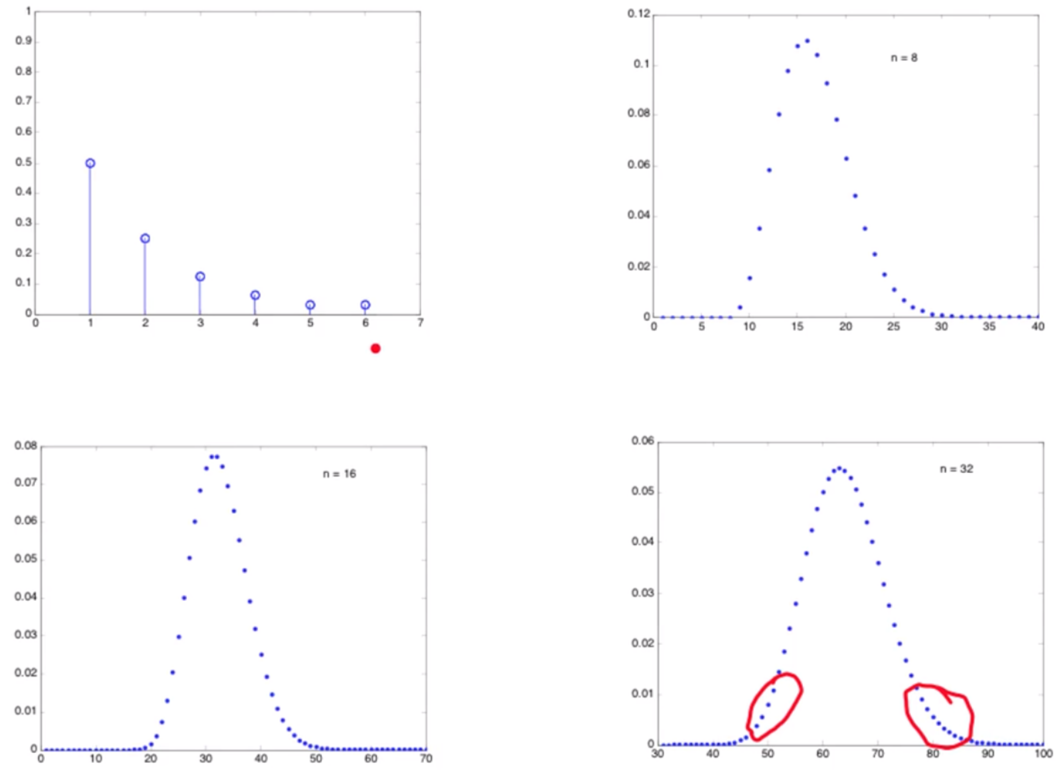
2. Here, the score of each student depends strongly on the difficulty level of the exam, which is random but common for all students. This creates a strong dependence between the student scores, and the CLT does not apply.

3. This is more subtle. The scores of the different students are not identically distributed. However, let Y_i be the score of the ith student from the first section and let Z_i be the score of the ith student in the second section. The class average is the average of the random variables $(Y_i + Z_i)/2$. Under our assumptions, these latter random variables can be modeled as i.i.d., and the CLT applies.

4. Unlike part (2), here the student scores are i.i.d., and the CLT applies.

Illustration of the CLT





The asymmetry of the original distribution means we need more values of n in order to get to a normal approximation

CLT examples

Example 1

- $\mathbf{P}(S_n \le a) \approx b$ given two parameters, find the third
- Package weights X_i, i.i.d. exponential, $\lambda = 1/2$;
- Load container with $n = 100$ packages

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\,\sigma}$$

$$\mu = \sigma = 2$$

$\mathbf{P}(S_n \ge 210)$

$= P\left(\frac{S_n - 200}{20} \geq \frac{210 - 200}{20}\right)$

$= P(Z_n \geq 0.5) \approx P(Z \geq 0.5)$

$= 1 - P(Z < 0.5) = 1 - \Phi(0.5)$

$= 1 - .6915 = 0.3085$

	.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

For an exponential the mean and variance are the inverse of lambda
Z is standard normal which Zn approximates to

Example 2

- $\mathbf{P}(S_n \le a) \approx b$ given two parameters, find the third
- Package weights X_i, i.i.d. exponential, $\lambda = 1/2$;
- Let $n = 100$. Choose the "capacity" a, so that $\mathbf{P}(S_n \ge a) \approx 0.05$.

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\,\sigma}$$

$$\mu = \sigma = 2$$

	.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

$$0.05 \approx P\left(\frac{S_n - 200}{20} \ge \frac{a-200}{20}\right)$$

$$\approx 1 - \underbrace{\Phi\left(\frac{a-200}{20}\right)}_{0.95}$$

$$\frac{a-200}{20} = 1.645 \qquad a = 232.9$$

Example 3

- $\mathbf{P}(S_n \le a) \approx b$ given two parameters, find the third
- Package weights X_i, i.i.d. exponential, $\lambda = 1/2$;
- How large can n be, so that $\mathbf{P}(S_n \ge 210) \approx 0.05$?

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\,\sigma}$$

$$\mu = \sigma = 2$$

	.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

$$P\left(\frac{S_n - 2n}{2\sqrt{n}} \ge \frac{210-2n}{2\sqrt{n}}\right)$$

$$\approx 1 - \underbrace{\Phi\left(\frac{210-2n}{2\sqrt{n}}\right)}_{0.95} \approx 0.05$$

$$\frac{210-2n}{2\sqrt{n}} = 1.645 \qquad \boxed{n = 89}$$

Example 4

- $\mathbf{P}(S_n \le a) \approx b$ given two parameters, find the third
- Package weights X_i, i.i.d. exponential, $\lambda = 1/2$;

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\,\sigma}$$

$$\mu = \sigma = 2$$

- Load container until weight exceeds 210
  N: number of packages loaded
- $\mathbf{P}(N > 100)$

$$= \mathbf{P}\left(\sum_{i=1}^{100} X_i \le 210\right)$$

$$\approx \Phi\left(\frac{210-200}{20}\right) = \Phi(0.5)$$

$$= 0.6915$$

	.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

N is not the sample mean so we translate it into something else = the sum of the weights of the packages
The value of that is the mean x 100 = 200

8. Exercise: CLT practice

Bookmark this page

Exercise: CLT practice

4/6 points (graded)

The random variables X_i are i.i.d. with mean 2 and standard deviation equal to 3. Assume that the X_i are nonnegative. Let $S_n = X_1 + \cdots + X_n$.

Use the CLT to find good approximations to the following quantities. You may want to refer to the normal table. In parts (a) and (b), give answers with 4 decimal digits.

Normal Table

Show

a) $\mathbf{P}(S_{100} \le 245) \approx$ 0.9332 ✔ **Answer:** 0.9332

b) We let N (a random variable) be the first value of n for which S_n exceeds 119.

$\mathbf{P}(N > 49) \approx$ 0.242 ✖ **Answer:** 0.8413

c) What is the largest possible value of n for which we have $\mathbf{P}(S_n \le 128) \approx 0.5$?

$n =$ 64 ✔ **Answer:** 64

Solution:

We will use Z_n to refer to the standardized random variable $(S_n - 2n) / (3\sqrt{n})$.

a) We have

$$\mathbf{P}(S_{100} \leq 245) = \mathbf{P}\left(\frac{S_{100} - 2 \cdot 100}{3 \cdot \sqrt{100}} \leq \frac{245 - 2 \cdot 100}{3 \cdot \sqrt{100}}\right) = \mathbf{P}(Z_n \leq 1.5) \approx 0.9332.$$

b) The event $N > 49$ is the same as the event $S_{49} \leq 119$. Its probability is

$$\mathbf{P}(S_{49} \leq 119) = \mathbf{P}\left(\frac{S_{49} - 2 \cdot 49}{3 \cdot \sqrt{49}} \leq \frac{119 - 2 \cdot 49}{3 \cdot \sqrt{49}}\right) = \mathbf{P}(Z_n \leq 1) \approx 0.8413.$$

c) We want n such that

$$0.5 \approx \mathbf{P}(S_n \leq 128) = \mathbf{P}\left(\frac{S_n - 2n}{3\sqrt{n}} \leq \frac{128 - 2n}{3\sqrt{n}}\right) = \Phi\left(\frac{128 - 2n}{3\sqrt{n}}\right).$$

But since $0.5 = \Phi(0)$, we must have $(128 - 2n) / (3\sqrt{n}) = 0$, so that $n = 128/2 = 64$.

A faster way to see the answer is to note that since the normal is symmetric around its mean, the relation $\mathbf{P}(S_n \leq 128) \approx 0.50$ tells us that 128 should be equal to the mean, $2n$, of S_n.

Normal approximation to the binomial

- X_i: independent, Bernoulli(p); $0 < p < 1$
- $S_n = X_1 + \cdots + X_n$: Binomial(n, p)
  - mean np, variance $np(1-p)$
- CDF of $\dfrac{\overbrace{S_n - np}^{Z_n}}{\sqrt{np(1-p)}} \longrightarrow$ standard normal
- $n = 36$, $p = 0.5$; find $\mathbf{P}(S_n \leq 21)$

$np = 18$ $\quad \sqrt{np(1-p)} = 3$

$$\mathbf{P}\left(\frac{S_n - 18}{3} \leq \frac{21 - 18}{3}\right)$$

$$= \mathbf{P}(Z_n \leq 1) \approx \Phi(1) = .8413$$

$$\sum_{k=0}^{21} \binom{36}{k} \left(\frac{1}{2}\right)^{36} = 0.8785$$

	.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

This approximation is off by about 4%

The 1/2 correction for integer random variables

- $0.8413 \approx \mathbf{P}(S_n \leq 21) = \mathbf{P}(S_n < 22)$, because S_n is integer

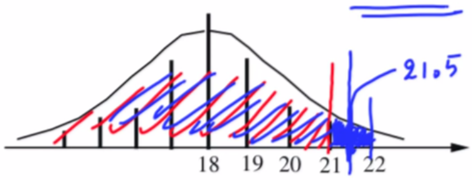
$$= P\left(\frac{S_n - 18}{3} < \frac{22-18}{3}\right)$$

$$= P(Z_n < 1.33) \approx \Phi(1.33) = 0.9082$$

true value 0.8785

$$\mathbf{P}(S_n \leq 21.5) = P\left(Z_n \leq \frac{21.5-18}{3}\right)$$

$$\approx \Phi(1.17) = .8790$$



	.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

The next approximate overestimates
We can see the blue section, the true value falls in the middle of this

De Moivre–Laplace CLT to the binomial

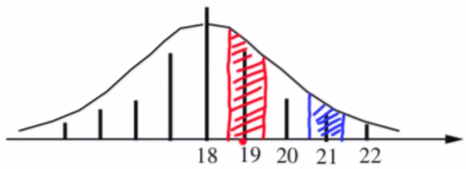
$$\mathbf{P}(S_n = 19) = P(18.5 \leq S_n \leq 19.5)$$

$$= P\left(\frac{18.5-18}{3} \leq Z_n \leq \frac{19.5-18}{3}\right)$$

$$= P(0.17 \leq Z_n \leq 0.5)$$

$$\approx \Phi(0.5) - \Phi(0.17)$$

$$= 0.6915 - 0.5675 = 0.124$$



- Exact answer:

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

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

10. Exercise: CLT for the binomial

Bookmark this page

Exercise: CLT for the binomial

3/3 points (graded)

Let X be binomial with parameters $n = 49$ and $p = 1/10$.

The mean of X is: 4.9 ✔ **Answer:** 4.9

The standard deviation of X is: 2.1 ✔ **Answer:** 2.1

The CLT, together with the $1/2$-correction, suggests that

$\mathbf{P}(X = 6) \approx$ 0.1661 ✔ **Answer:** 0.1623

You may want to refer to the normal table.

Note: In this case, the CLT may not provide a great approximation. The range of values that X is likely to take is quite narrow, so that its PMF consists of only a few entries of substantial size. But, regardless, we can still calculate what the CLT suggests.

Solution:

We have $\mathbf{E}[X] = np = 4.9$, and

$$\text{Var}(X) = np(1-p) = 49 \cdot \frac{1}{10} \cdot \frac{9}{10} = \frac{49 \cdot 9}{10^2},$$

so that the standard deviation of X is $21/10 = 2.1$.

The standardized version of X is $(X - 4.9)/2.1$. Thus,

$$\begin{aligned} \mathbf{P}(X = 6) &= \mathbf{P}(5.5 < X < 6.5) = \mathbf{P}\left(\frac{5.5 - 4.9}{2.1} \le \frac{X - 4.9}{2.1} \le \frac{6.5 - 4.9}{2.1}\right) \\ &\approx \Phi(0.76) - \Phi(0.29) \approx 0.7764 - 0.6141 = 0.1623. \end{aligned}$$

For comparison, the answer calculated by using the binomial PMF directly is

$$\mathbf{P}(X = 6) = \binom{49}{6}(0.1)^6(0.9)^{49-6} \approx 0.1507.$$

The pollster's problem revisited

- p: fraction of population that will vote "yes" in a referendum
- ith (randomly selected) person polled: $X_i = \begin{cases} 1, & \text{if yes,} \\ 0, & \text{if no.} \end{cases}$ $\quad E[X_i] = p = \mu$ $\quad \sigma = \sqrt{p(1-p)}$
- $M_n = (X_1 + \cdots + X_n)/n$: fraction of "yes" in our sample
- Would like "small error," e.g.: $|M_n - p| < 0.01$

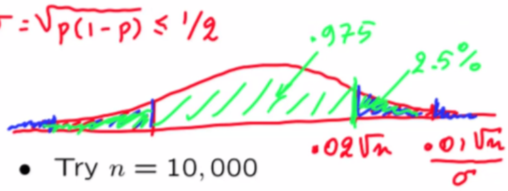
$$\mathbf{P}\big(|M_n - p| \geq .01\big) = P\left(|Z_n| \geq \frac{.01\sqrt{n}}{\sigma}\right) \approx P\left(|\underset{N(0,1)}{Z}| \geq \frac{.01\sqrt{n}}{\sigma}\right)$$

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\,\sigma} \qquad \left|\frac{S_n - np}{n}\right| \geq .01 \qquad \left|\overbrace{\frac{S_n - np}{\sqrt{n}\,\sigma}}^{Z_n}\right| \geq \frac{.01\sqrt{n}}{\sigma}$$

The pollster's problem revisited

$$\mathbf{P}\big(|M_n - p| \geq .01\big) \approx \mathbf{P}\left(|Z| \geq \frac{.01\sqrt{n}}{\sigma}\right) \leq \mathbf{P}\big(|Z| \geq .02\sqrt{n}\big) = 2\big(1 - \Phi(\underbrace{.02\sqrt{n}}_{2})\big) = 0.05$$

$$\sigma = \sqrt{p(1-p)} \leq 1/2$$



- Try $n = 10{,}000$

$$\text{prob} \leq 2(1 - \Phi(2)) =$$
$$= 2(1 - .9772) = 0.046$$

- Specs: $\mathbf{P}\big(|M_n - p| \geq .01\big) \leq .05$

$$\Phi(.02\sqrt{n}) = 0.975$$

$$.02\sqrt{n} = 1.96 \Rightarrow n = 9604$$

	.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

0.046 is the probability percentage that the error will be greater than 1 percent in green, working backwards and calculating how many people are needed to sample in order to get error greater than 1% with probability 5%

Classical Statistics

An unknown quantity is assumed as a constant not a random variable like in bayesian statistics

Classical statistics

- Inference using the Bayes rule:
  unknown Θ and observation X are both random variables
  - Find $p_{\Theta|X}$

$P_\Theta \quad P_{X|\Theta}$

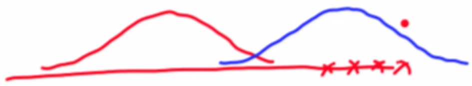
- Classical statistics: unknown constant θ

$$\theta \longrightarrow \boxed{p_X(x;\theta)} \xrightarrow[x]{X} \boxed{\text{Estimator}} \xrightarrow{\widehat{\Theta}} \qquad \widehat{\Theta} \leftarrow \text{r.v.} \qquad \hat{\theta} \leftarrow \text{estimate}$$

$g(x)$

$P_{X|\Theta}(x|\theta)$

  - also for vectors X and θ: $p_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta_1,\ldots,\theta_m)$
  - $p_X(x;\theta)$ are NOT conditional probabilities; θ is NOT random
  - mathematically: many models, one for each possible value of θ



x is a normal distribution which depends on theta in some way
semi-colon indicates that it is not a conditional probability but that theta is included

Problem types in classical statistics

- Classical statistics: unknown constant θ

$$\theta \longrightarrow \boxed{p_X(x;\theta)} \xrightarrow{X} \boxed{\text{Estimator}} \xrightarrow{\widehat{\Theta}}$$

- Hypothesis testing: $H_0 : \theta = 1/2$ versus $H_1 : \theta = 3/4$
- Composite hypotheses: $H_0 : \theta = 1/2$ versus $H_1 : \theta \neq 1/2$
- Estimation: design an **estimator** $\widehat{\Theta}$, to "keep estimation **error** $\widehat{\Theta} - \theta$ small"

Art!

Estimating a mean

- $X_1, \ldots, X_n$: i.i.d., mean θ, variance σ^2

$$\widehat{\Theta}_n = \text{sample mean} = M_n = \frac{X_1 + \cdots + X_n}{n}$$

$\widehat{\Theta}_n$: estimator (a random variable)

Properties and terminology:

$\hat{\Theta} = g(X)$

$E[\hat{\Theta}] = \sum_x g(x)\, p_X(x;\theta)$

- $\mathbf{E}[\widehat{\Theta}_n] = \theta$ (unbiased)
  for all θ
- WLLN: $\widehat{\Theta}_n \xrightarrow{\text{i.p.}} \theta$ (consistency)
  for all θ
- mean squared error (MSE): $\mathbf{E}\left[(\widehat{\Theta}_n - \theta)^2\right] = \text{var}(\hat{\Theta}_n) = \frac{\sigma^2}{n}$.

- An estimator is **consistent** if, as the sample size increases, the estimates (produced by the estimator) "converge" to the true value of the parameter being estimated. To be slightly more precise - consistency means that, as the sample size increases, the sampling distribution of the estimator becomes increasingly concentrated at the true parameter value.
- An estimator is **unbiased** if, on average, it hits the true parameter value. That is, the mean of the sampling distribution of the estimator is equal to the true parameter value.
- The two are not equivalent: **Unbiasedness** is a statement about the expected value of the sampling distribution of the estimator. **Consistency** is a statement about "where the sampling distribution of the estimator is going" as the sample size increases.

We estimate the unknown mean θ of a random variable X (where X has a finite and positive variance) by forming the sample mean $M_n = (X_1 + \cdots + X_n)/n$ of n i.i.d. samples X_i and then forming the estimator

$$\widehat{\Theta} = M_n + \frac{1}{n}.$$

Is this estimator unbiased?

No ✔ **Answer:** No

Is this estimator consistent?

Yes ✔ **Answer:** Yes

Consider now a different estimator, $\widehat{\Theta}_n = X_1$, which ignores all but the first measurement.

Is this estimator unbiased?

No ✖ **Answer:** Yes

Is this estimator consistent?

No ✔ **Answer:** No

Solution:

We have $\mathbf{E}[\widehat{\Theta}_n] = \theta + (1/n) \neq \theta$, so it is not unbiased. On the other hand, M_n converges (in probability) to θ, and $1/n$ converges to zero. So, their sum, $\widehat{\Theta}_n = M_n + (1/n)$ also converges (in probability) to θ, and the estimator is consistent.

The second estimator is unbiased, because $\mathbf{E}[\widehat{\Theta}_n] = \mathbf{E}[X_1] = \theta$. But it is not consistent. Its value stays the same (equal to X_1) for all n and therefore cannot converge to θ, unless X_1 is guaranteed to be equal to θ. But this is impossible since X has positive variance.

Problems

Problem 1. Convergence in probability

8/8 points (graded)

For each of the following sequences, determine whether it converges in probability to a constant. If it does, enter the value of the limit. If it does not, enter the number “999".

1. Let $X_1, X_2, \ldots$ be independent continuous random variables, each uniformly distributed between -1 and 1.

   - Let $U_i = \frac{X_1 + X_2 + \cdots + X_i}{i}$, $i = 1, 2, \ldots$. What value does the sequence U_i converge to in probability? (If it does not converge, enter the number “999". Similarly in all below.)

     0

   - Let $\Sigma_i = X_1 + X_2 + \cdots + X_i$, $i = 1, 2, \ldots$. What value does the sequence Σ_i converge to in probability?

     999

   - Let $I_i = 1$ if $X_i \geq 1/2$, and $I_i = 0$, otherwise. Define,

     $$S_i = \frac{I_1 + I_2 + \cdots + I_i}{i}.$$

     What value does the sequence S_i converge to, in probability?

     0.25

   - Let $W_i = \max\{X_1, \ldots, X_i\}$, $i = 1, 2, \ldots$. What value does the sequence W_i converge to in probability?

     1

   - Let $V_i = X_1 \cdot X_2 \cdots X_i$, $i = 1, 2, \ldots$. What value does the sequence V_i converge to in probability?

     0

2. Let $X_1, X_2, \ldots$, be independent identically distributed random variables with $\mathbf{E}[X_i] = 2$ and $\mathsf{Var}(X_i) = 9$, and let $Y_i = X_i/2^i$.

   - What value does the sequence Y_i converge to in probability?

     0

   - Let $A_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$. What value does the sequence A_n converge to in probability?

     0

   - Let $Z_i = \frac{1}{3}X_i + \frac{2}{3}X_{i+1}$ for $i = 1, 2, \ldots$, and let $M_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$ for $n = 1, 2, \ldots$. What value does the sequence M_n converge to in probability?

     2