Law of Large Numbers, Central Limit Theorem

Math 10A

November 14, 2017

November 15–18

Ribet in Providence on AMS business.

No SLC office hour tomorrow.

Thursday's class conducted by Teddy Zhu.

November 21

Class on hypothesis testing and p-values

December 1

- 8AM breakfast—send email to sign up;
- pop-in lunch at high noon (just show up).

Variance of a sum

If X_1 and X_2 are random variables, then

$$\begin{aligned}\mathrm{Var}[X_1 + X_2] &= E[(X_1 + X_2)^2] - (E[X_1 + X_2])^2 \\ &= (E[X_1^2] - E[X_1]^2) + (E[X_2^2] - E[X_2]^2) \\ &\quad + 2(E[X_1 X_2] - E[X_1]E[X_2]) \\ &= \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + 2(E[X_1 X_2] - E[X_1]E[X_2])\end{aligned}$$

If X_1 and X_2 are independent, the term $E[X_1 X_2] - E[X_1]E[X_2]$ is 0 because the expected value of a product of independent variables is the product of the expected values.

Hence the variance of a sum of two *independent* random variables is the sum of the variances of the random variables:

$$\mathrm{Var}[X_1 + X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2].$$

If the two random variables are not independent, this formula is very unlikely to hold. For example, suppose $X_2 = X_1$. Then

$$\mathrm{Var}[X_1 + X_2] = \mathrm{Var}[2 \cdot X_1] = 4\,\mathrm{Var}[X_1].$$

If the variance of X_1 is non-zero, $4\,\mathrm{Var}[X_1]$ will be different from $\mathrm{Var}[X_1] + \mathrm{Var}[X_1] = 2 \cdot \mathrm{Var}[X_1]$.

Many variables

There is a (somewhat technical) definition of what it means for a bunch of random variables X_1, X_2,..., X_n to be independent (i.e., mutually independent). If they are independent, then

$$\mathrm{Var}[X_1 + X_2 + \cdots + X_n] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n].$$

The variance of a sum of n independent variables is the sum of the variances of the variables.

Imagine again a (possibly biased) coin that comes up heads with probability p and tails with probability $q = 1 - p$.

There's a natural random variable X in the picture: $X = 0$ if we flip a coin once and get `T`; $X = 1$ if we get `H`.

We've seen that this random variable has mean p and variance pq.

iid's

Let's flip the coin n times and let $X_1, X_2, \ldots, X_n$ be the random variables associated with the first, second,…, nth flips.

The random variables X_1, X_2, etc. are clones of X. We say that they are *identically distributed*.

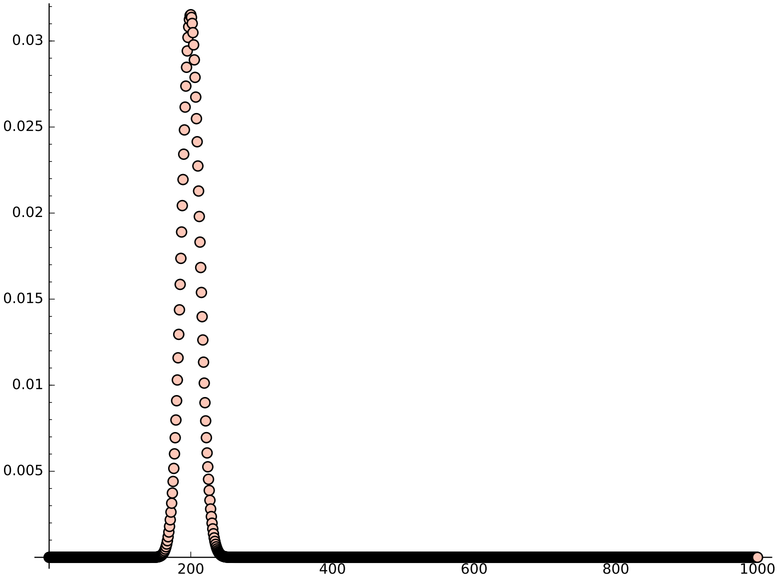
They are *independent* of each other: a given flip has no entanglement with the other flips.

The variables X_1, X_2,… are independent and identically distributed ("iid").
The random variable $X_1 + X_2 + \cdots + X_n$ counts the number of heads obtained when flipping a coin n times.

The random variable $X_1 + X_2 + \cdots + X_n$ counts the number of heads obtained when flipping a coin n times.

Its expected values is $p + p + \cdots + p = np$. If $\mathtt{H}$ comes up 1/5 of the time and we flip the coin 1000 times, we expect $1000 \times 1/5 = 200$ heads. This makes a lot of sense to us.

The variable $X_1 + X_2 + \cdots + X_n$ has variance $pq + pq + \cdots + pq = npq$ because the X_i are independent of each other. This makes sense to me because I'm teaching the course, but I don't have a pithy explanation to post to social media.



The distribution of $X_1 + X_2 + \cdots + X_{1000}$, $p = 0.2$.

If c is a constant, then

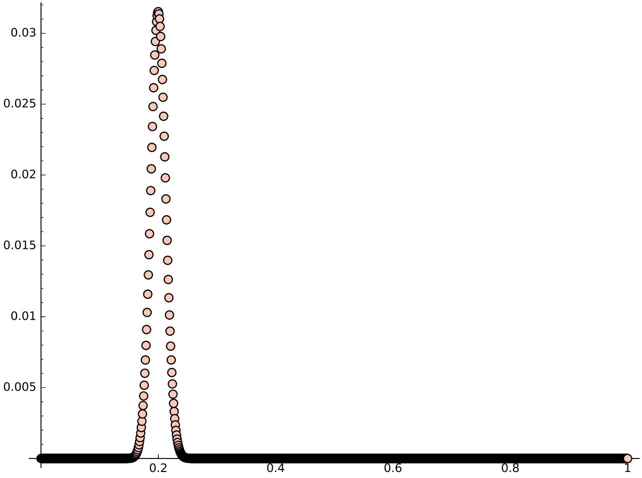
$$E[c(X_1 + X_2 + \cdots + X_n)] = c \cdot np,$$
$$\mathrm{Var}[c(X_1 + X_2 + \cdots + X_n)] = c^2 npq.$$

The reason for the second equation is that $\mathrm{Var}[X]$ involves squares: $E[X^2]$, $E[X]^2$.

Take $c = \dfrac{1}{n}$; then

$$E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] = p,$$
$$\mathrm{Var}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] = \frac{pq}{n}.$$

This graph is exactly like the previous graph, except that the x-axis has been squashed by a factor of 1000.



The distribution of $\dfrac{X_1 + X_2 + \cdots + X_{1000}}{1000}$, $p = 0.2$.

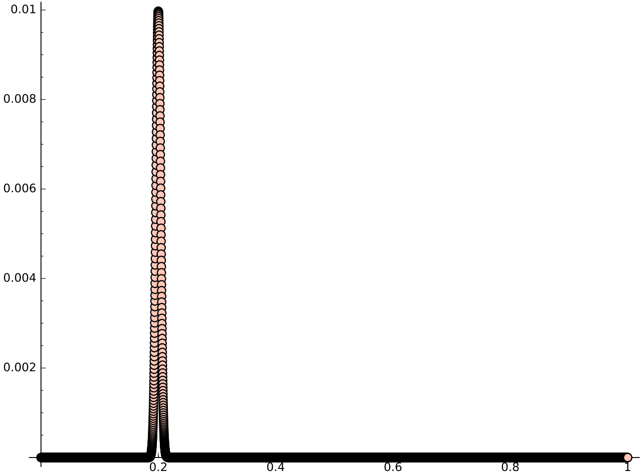
An "abstraction" with nothing really new in it: Let X be a random variable with mean μ and standard deviation σ. Let $X_1, X_2, \cdots, X_n$ be identically distributed independent clones of X, and let $\overline{X}$ be the average of the X_i:

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Then $\overline{X}$ has mean μ, variance $\dfrac{\sigma^2}{n}$ and standard deviation $\dfrac{\sigma}{\sqrt{n}}$.

We can say that $\overline{X}$ is "like" X in that it has the same mean but "less random" than X because its standard deviation is $\dfrac{\sigma}{\sqrt{n}}$ instead of σ.

What happens when n gets big?



The distribution of $\dfrac{X_1 + X_2 + \cdots + X_{10000}}{10000}$, $p = 0.2$.

What you see is what you get: the random variable $\overline{X}$ looks as if all of the probability is concentrated at the single value μ.

This is the *Law of Large Numbers:*

$$\text{As } n \to \infty, \text{ the average } \overline{X} = \frac{X_1 + \cdots + X_n}{n} \text{ tends to } \mu.$$

Remember: this is not just a good idea—it's the law.

To understand what's going on, remember that the standard deviation of $\overline{X}$ is $\dfrac{\sigma}{\sqrt{n}}$. As $n \to \infty$, the deviation of $\overline{X}$ approaches 0, so it's natural to think of $\overline{X}$ as a constant.

Not on the exam

If you want to read more about the Law, see the Wikipedia discussion, including the References and External links.

Break

December 1

- You know that you want to sign up for the 8AM breakfast. #reallygoodspecialbreakfast
- How could you possibly miss the pop-in lunch at high noon? #lastdayofclasses

Let X be a random variable with mean μ and standard deviation σ. Let $X_1, X_2, \cdots, X_n$ be identically distributed independent clones of X, and let $\overline{X}$ be the average of the X_i:

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Then $\overline{X}$ has mean μ, variance $\frac{\sigma^2}{n}$ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

Now for something different. First, subtract μ from $\overline{X}$: $\overline{X} - \mu$ has mean 0. It still has variance $\dfrac{\sigma^2}{n}$ and standard deviation $\dfrac{\sigma}{\sqrt{n}}$. Relative to the graph, all we've done is to shift the graph by $-\mu$ units so that its center is at 0 instead of at μ.

Next, consider

$$(\overline{X} - \mu) \cdot \frac{\sqrt{n}}{\sigma}.$$

This new random variable has mean 0 and standard deviation 1...

...just like the gold standard normal variable, the one with PDF equal to $\dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Now for something different. First, subtract μ from $\overline{X}$: $\overline{X} - \mu$ has mean 0. It still has variance $\frac{\sigma^2}{n}$ and standard deviation $\frac{\sigma}{\sqrt{n}}$. Relative to the graph, all we've done is to shift the graph by $-\mu$ units so that its center is at 0 instead of at μ.

Next, consider

$$(\overline{X} - \mu) \cdot \frac{\sqrt{n}}{\sigma}.$$

This new random variable has mean 0 and standard deviation 1…

…just like the gold standard normal variable, the one with PDF equal to $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

$$(\overline{X} - \mu) \cdot \frac{\sqrt{n}}{\sigma} \quad \stackrel{?}{\longleftrightarrow} \quad \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Possibly clarifying comments

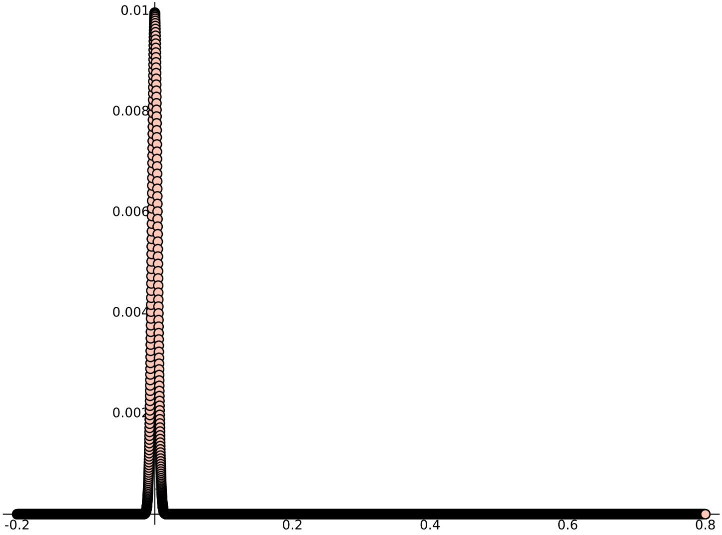
- The random variable $\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ depends on n. We might have called it $\overline{X}_n$ to stress the dependence. Let's do that from when appropriate.
- We could rewrite $\overline{X}\,\dfrac{\sqrt{n}}{\sigma}$ as $\dfrac{X_1 + X_2 + \cdots + X_n}{\sigma\sqrt{n}}$ and get an alternative expression for $(\overline{X} - \mu) \cdot \dfrac{\sqrt{n}}{\sigma}$.
- Data 8 instructors report that students in traditional statistics courses can never remember whether to multiply or divide by $\sqrt{n}$!

Biased coin example (again)

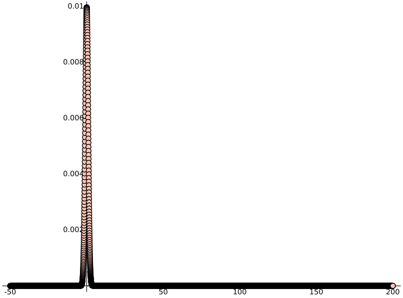
We continue discussing the biased coin that appeared before. To make things concrete, assume $p = \frac{1}{5}$, $q = \frac{4}{5}$; the coin comes up $\mathtt{H}$ with probability 0.2. For a single coin flip with $X = 0$ or 1, the mean is 0.2, the variance is $pq = \frac{4}{25}$ and the standard deviation is $\sigma = \frac{2}{5}$.

Let's flip the coin 10000 times; i.e., we take $n = 10^4$. The random variable $\overline{X}$ has values ranging from $0/10^4$ to $10^4/10^4$. The extreme values occur with very low probabilities $\left(\frac{4}{5}\right)^{10^4}$ and $\left(\frac{1}{5}\right)^{10^4}$, respectively.

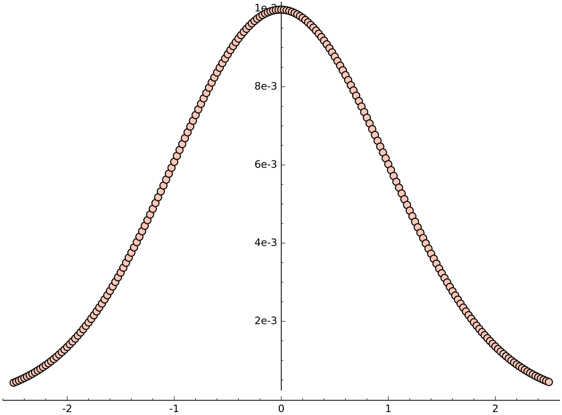
We saw the distribution of $\overline{X}$ before the break. Here's the probability distribution for $\overline{X} - \mu$:



To get something with standard deviation 1, we need to multiply by $\frac{\sqrt{n}}{\sigma} = 100 \times \frac{5}{2} = 250$. The values of $(\overline{X} - \mu) \cdot \frac{\sqrt{n}}{\sigma}$ will range from -50 to 200:



This graph looks exactly like the previous one. But it's fake news to say that they're the same. The most recent graph is obtained from the previous one by stretching out the *x*-axis by a factor of 250. The "technology" that produced the graph automatically squashes the graph horizontally so that it can fit on the screen. Sad!



This graph zeros in on the probabilities associated with the values of $(\overline{X} - \mu) \cdot \dfrac{\sqrt{n}}{\sigma}$ between ± 2.5. The picture looks a lot like a normal curve that was ordered up from Central Casting.

"Central" is the word.

Central Limit Theorem

For real numbers a and b with $a \leq b$:

$$P\left(a \leq \frac{(\overline{X}_n - \mu)\sqrt{n}}{\sigma} \leq b\right) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\,dx$$

as $n \to \infty$.

For further info, see the discussion of the Central Limit Theorem in the `10A_Prob_Stat` notes on `bCourses`.

The theorem is often paraphrased by the statement that the variables $\dfrac{(\overline{X}_n - \mu)\sqrt{n}}{\sigma}$ are becoming more and more like a standard normal variable.

For example, we might think

$$P\left(0 \leq \frac{(\overline{X}_n - \mu)\sqrt{n}}{\sigma} \leq 1\right) \approx \frac{1}{\sqrt{2\pi}} \int_0^1 e^{-x^2/2}\, dx \approx 0.3413$$

In the biased coin example, this means that the probability of getting between 2000 and 2040 heads in 10000 tosses is roughly 0.3413. To four decimal places, this probability is 0.3483. The probability of getting between 8000 and 8080 heads in 40000 tosses is roughly 0.3448.

Area under the Normal Curve from 0 to X

X	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41308	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950
3.3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965
3.4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976
3.5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983
3.6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989
3.7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992
3.8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997
4.0	0.49997	0.49997	0.49997	0.49997	0.49997	0.49997	0.49998	0.49998	0.49998	0.49998

With a table like this, we could look up the value 0.3413(4).

Then and now

Old-style problem: “A biased coin. . . . Estimate the probability of getting between 8000 and 8080 heads in 40000 tosses of the coin.” Answer: Guided by the Central Limit Theorem, we estimate the probability to be 0.3413.

2017-style exercise: “Use technology to compute a decimal value for the probability of getting between 8000 and 8080 heads in 40000 tosses of the coin.” Answer: 0.3448.

We can compute directly many quantities that our ancestors needed to estimate using tables.