MTH 7241 Fall 2022: Prof. C. King

Notes 3: Sequences and limit theorems

IID sequence

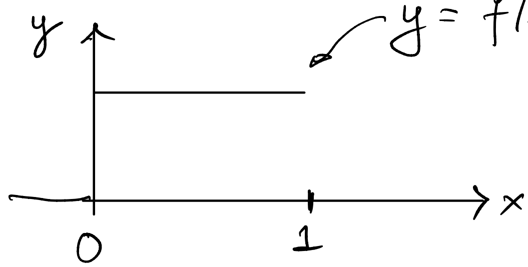
The random variables $Y_1, Y_2, \ldots$ are *independent and identically distributed* or **IID** if they are all independent and have the same pdf. Independence means that

$$\mathbb{P}(Y_{i_1} = k_{i_1}, \ldots, Y_{i_n} = k_{i_n}) = \mathbb{P}(Y_{i_1} = k_{i_1}) \ldots \mathbb{P}(Y_{i_n} = k_{i_n})$$

for all subsets of variables and all values $k_{i_1}, \ldots, k_{i_n}$. In particular they all have the same mean μ and variance σ^2. IID variables arise in sampling problems, where successive independent measurements are made on a random system. For example, Y_n could be the value of the nth roll of a die, or the nth toss of a coin.

Example. $Y_1 \sim$ uniform on $[0,1]$.

$$y = f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{else} \end{cases}$$



$Y_2 \sim U[0,1]$, independent of Y_1.

Mean: $\mathbb{E}[Y_1] = \mathbb{E}[Y_2] = \frac{1}{2} = \mu$

$VAR[Y_1] = VAR[Y_2] = \frac{1}{12} = \sigma^2$

Y_1, Y_2 are IID $\Rightarrow$ same pdf
independent.

$$\mathbb{P}\left(Y_1 \le \tfrac{1}{2}, Y_2 \le \tfrac{3}{4}\right) = \mathbb{P}\left(Y_1 \le \tfrac{1}{2}\right) \mathbb{P}\left(Y_2 \le \tfrac{3}{4}\right)$$
$$= \left(\tfrac{1}{2}\right)\left(\tfrac{3}{4}\right).$$

More generally, have

$$\left.\begin{array}{l} Y_1 \sim U[0,1] \\ Y_2 \sim U[0,1] \\ \vdots \\ Y_n \sim U[0,1] \end{array}\right\} \begin{array}{l} \text{all independent} \\ \Rightarrow \text{IID.} \end{array}$$

Random binary string

Let $Y_1, \ldots$ be IID Bernoulli r.v.'s, with

$$\mathbb{P}(Y = 1) = p, \quad \mathbb{P}(Y = 0) = 1 - p$$

The sequence $(Y_1, Y_2, \ldots, Y_n)$ is a random binary n-string, or random bit-string. The sum $\sum_i Y_i$ is the number of 1's in the string, and its pdf is binomial with parameters (n, p).

H T T H H T H T T H ....

$\downarrow$

1 0 0 1 1 0 1 0 0 1 ...

$\downarrow$

$Y_1 \quad Y_2 \quad Y_3 \quad Y_4 \quad \ldots$

$\mathbb{P}(Y_1 = 1) = p$

$\mathbb{P}(Y_1 = 0) = 1 - p$

Y_1	0	1
prob	$1-p$	p

$E[Y_1] = p$

$VAR[Y_1] = p(1-p)$.

$Y_1, Y_2, Y_3, \ldots, Y_n, \ldots$ are IID.

$$\mathbb{P}(1\,0\,0\,1\,1) = \mathbb{P}(Y_1 = 1, Y_2 = 0, Y_3 = 0, Y_4 = 1, Y_5 = 1)$$

$$= \mathbb{P}(Y_1 = 1)\,\mathbb{P}(Y_2 = 0) \cdots \mathbb{P}(Y_5 = 1)$$

$$= p(1-p)(1-p)\,p\,p$$

$$= p^3(1-p)^2.$$

$$\mathbb{P}(01101) = p^3(1-p)^2$$

← unchanged under permutation

$$= p^{\#1\text{'s}}(1-p)^{\#0\text{'s}}$$

$\mathbb{P}$(string w/ 3 1's & 2 0's)

$$= 10\, p^3(1-p)^2$$

$$= \binom{5}{3} p^3 (1-p)^2$$

← binomial coefficient

X = # 1's that occur in a sequence of length n

$$(Y_1, Y_2, Y_3, \dots Y_n)$$

$$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

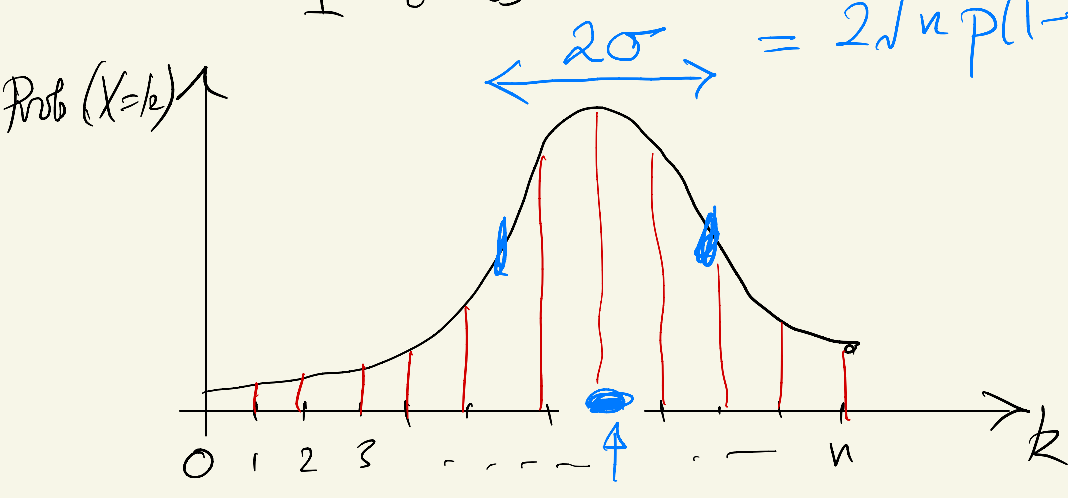
$$k = 0, 1, 2, 3 \sim n$$

$$\mathbb{E}[X] = np = \mu$$

$$\mathrm{VAR}[X] = np(1-p) = \sigma^2$$

$$\underbrace{\bigcirc \ \bigcirc \ \bigcirc \ \bigcirc \ \cdots\cdots \ \bigcirc}_{1 \text{ occurs } k \text{ times}}$$

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



$$\mathbb{E}[X] = np = \mu$$

Law of Large Numbers

Let $Y_1, Y_2, \dots$ be any collection of IID r.v.'s (independent and identically distributed random variables) with common mean μ and variance σ^2. The sample mean of the first n variables is

$$\overline{Y}_n = \frac{1}{n}(Y_1 + \cdots + Y_n)$$

The LNN (Law of Large Numbers) says that $\overline{Y}_n$ converges to the true mean μ as $n \to \infty$.

$$\mu = \mathbb{E}[Y_k]$$

$$\sigma^2 = VAR[Y_k].$$

$$\overline{Y}_n = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$$ sample mean.

random variable

Law of Large Numbers:

$$\overline{Y}_n \xrightarrow{\text{converges to}} \mu \quad \text{as } n \to \infty$$

sequence of random variables

constant.

Why is this true?

Mean value of $\bar{Y}_n$:

$$\mathbb{E}[\bar{Y}_n] = \mathbb{E}\left[\frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n)\right]$$

$$= \frac{1}{n}\mathbb{E}[Y_1 + Y_2 + \cdots + Y_n]$$

$$= \frac{1}{n}\left(\mathbb{E}[Y_1] + \mathbb{E}[Y_2] + \cdots + \mathbb{E}[Y_n]\right)$$

IID

$$= \frac{1}{n}(\mu + \mu + \mu + \cdots + \mu)$$

$$= \frac{n\mu}{n} = \mu.$$

$$VAR[\bar{Y}_n] = VAR\left[\frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n)\right]$$

$$= \frac{1}{n^2} VAR[Y_1 + Y_2 + \cdots + Y_n]$$

$$= \frac{1}{n^2}\left(VAR[Y_1] + VAR[Y_2] + \cdots + VAR[Y_n]\right)$$

$$= \frac{1}{n^2}\left(\sigma^2 + \sigma^2 + \cdots + \sigma^2\right)$$

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

$$= \frac{\sigma^2}{n}.$$

So $\bar{Y}_n$ has mean μ, variance $\frac{\sigma^2}{n}$

$\Rightarrow$ as $n \to \infty$, it converges to the constant value μ.

Note that $\overline{Y}_n = \frac{1}{n}(Y_1 + \cdots + Y_n)$ is the long-run average value of a sequence of measurements. For each possible value y_k define

$$n(k) = \#\{i : Y_i = y_k\}$$

so then

$$Y_1 + \cdots + Y_n = \sum_k y_k\, n(k)$$

The LLN is equivalent to the statement that the relative frequency of occurrence of each value converges to its probability:

$$\frac{n(k)}{n} \to \mathbb{P}(Y = y_k) \quad \text{as } n \to \infty$$

and therefore

$$\frac{1}{n}(Y_1 + \cdots + Y_n) = \sum_k y_k \frac{n(k)}{n} \to \sum_k y_k\, \mathbb{P}(Y = y_k) = \mathbb{E}[Y]$$

Real issue is what 'convergence' means. Given any $\epsilon > 0$, $\delta > 0$, there is $N < \infty$ such that for all $n \geq N$, $\overline{Y}_n$ will with probability at least $1 - \delta$ lie inside the interval $\mu \pm \epsilon$. Or more succinctly, for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}(Y_1 + \cdots + Y_n) - \mu\right| > \epsilon\right) = 0$$

Suppose pdf of Y is

Y	y_1	y_2	y_3
prob	p_1	p_2	p_3
frequency	n_1	n_2	n_3

$p_1 + p_2 + p_3 = 1$

$n_1 + n_2 + n_3 = n$

measure Y n times, n_1 = number of times we get y_1

n_2, n_3 are same

$$\overline{Y} = \frac{1}{n}(n_1 y_1 + n_2 y_2 + n_3 y_3)$$

$$\bar{Y} = \frac{n_1}{n} y_1 + \frac{n_2}{n} y_2 + \frac{n_3}{n} y_3$$

$\frac{n_1}{n}$ = fraction of times we measure the value y_1.

As $n \to \infty$:

$$\frac{n_1}{n} \longrightarrow P_1$$

$$\frac{n_2}{n} \longrightarrow P_2$$

$$\frac{n_3}{n} \longrightarrow P_3$$

So $\bar{Y} \longrightarrow P_1 y_1 + P_2 y_2 + P_3 y_3$

$$= \mathbb{E}[Y] = \mu$$

Markov's Inequality

Many useful results in probability are proved using inequalities. The relevant one for the LLN is called Markov's inequality and is easily stated. For any random variable X and for any numbers $a > 0$ and $k > 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{1}{a^k}\,\mathbb{E}[|X|^k]$$

The proof is easy:

some function

$$\begin{aligned}
\mathbb{E}[|X|^k] &= \sum_i |x_i|^k P(X = x_i) \\
&\geq \sum_{i:|x_i|\geq a} |x_i|^k P(X = x_i) \\
&\geq a^k \sum_{i:|x_i|\geq a} P(X = x_i) \\
&= a^k P(|X| \geq a)
\end{aligned}$$

An important special case of Markov's inequality is called *Chebyshev's inequality*: take $X = Y - \mathbb{E}Y$ and $k = 2$ to get

$$P(|Y - \mathbb{E}Y| \geq a) \leq \frac{1}{a^2}\,\mathrm{Var}(Y)$$

The LLN follows easily from this. Take

$$\bar{Y}_n \; \cancel{X} = \frac{1}{n}(Y_1 + \cdots + Y_n)$$

then $\mathbb{E}[X] = \mu$ and $\mathrm{VAR}[X] = \sigma^2/n$, hence

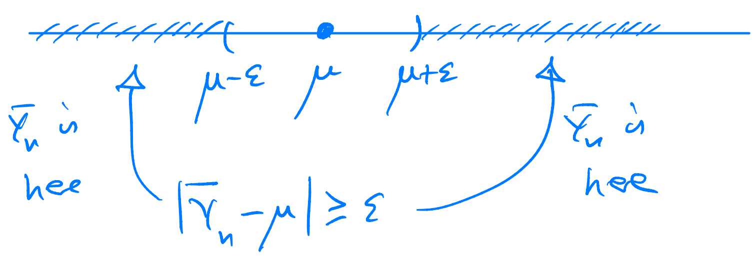
$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\,\epsilon^2} \to 0$$

ε = epsilon

$$\mathbb{E}[\bar{Y}_n] = \mu, \quad VAR[\bar{Y}_n] = \frac{\sigma^2}{n}$$

$$\mathbb{P}\left(|\bar{Y}_n - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}$$

$\frac{\sigma^2}{n\varepsilon^2} \to 0$ as $n \to \infty$



As $n \to \infty$, we must find $\bar{Y}_n$ lies in the interval $[\mu-\varepsilon, \mu+\varepsilon]$.

This is true for every $\varepsilon > 0$,

hence $\bar{Y}_n \to \mu$.

Example 1 *Toss an unbiased coin 1000 times. The running average $\overline{Y}_n$ converges to 0.5. Use Markov's inequality to lower bound*

$$P(0.4 < \overline{Y}_n < 0.6) \quad \textit{for } n = 1000$$

$$Y_k = \begin{cases} 1 & \text{if } k^{th} \text{ toss is Heads} \\ 0 & \text{if } k^{th} \text{ toss is Tails.} \end{cases}$$

$$\overline{Y}_n = \frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n) = \text{fraction of times we get Heads.}$$

Unbiased coin $\Rightarrow \mu = \frac{1}{2}$

Y_k	0	1
prob	$\frac{1}{2}$	$\frac{1}{2}$

$\sigma^2 = \frac{1}{4}, \quad \sigma = \frac{1}{2}.$

Markov's inequality: $(n = 1000)$

$$\mathbb{P}\left(\left|\overline{Y}_n - \frac{1}{2}\right| > 0.1\right) \le \frac{\sigma^2}{n(0.1)^2}$$

$$= \frac{1}{40}.$$

$$\Rightarrow \quad \mathbb{P}(0.4 \le \overline{Y}_n \le 0.6) \ge 1 - \frac{1}{40}$$

$$= \frac{39}{40}.$$

Central Limit Theorem

Theorem 1 *Let $Y_1, Y_2, \ldots$ be IID with finite mean $\mathbb{E}Y_i = \mu$ and finite variance* $\mathrm{VAR}[Y_i] = \sigma^2$. *Define*

$$Z_n = \frac{Y_1 + \cdots + Y_n - n\mu}{\sigma\sqrt{n}}$$

Then for all $a < b$, as $n \to \infty$,

$$P(a < Z_n \le b) \to P(a < Z \le b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx$$

$$\bar{Y}_n = \frac{1}{n}(Y_1 + Y_2 + \cdots)$$

$$\mathbb{E}[\bar{Y}_n] = \mu$$

$$\mathrm{VAR}[\bar{Y}_n] = \frac{\sigma^2}{n}$$

The integrand on the right side is the pdf of the standard normal. So another way to state the CLT is

$$Z_n \to Z \sim N(0,1) \quad \text{(convergence in distribution)}$$

$$Z_n = \frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}}$$

$$\Downarrow$$

$$\mathbb{E}[Z_n] = 0$$

$$\mathrm{VAR}[Z_n] = 1$$

Even more informally, we can say that for n large,

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i = \mu + \frac{\sigma}{\sqrt{n}} Z + \cdots$$

or even

$$\sum_{i=1}^n Y_i = n\,\mu + \sigma\,\sqrt{n}\, Z + \cdots$$

where $\cdots$ goes to zero faster than the leading order terms as $n \to \infty$. This is the most useful way to to think about the meaning of the CLT, and will guide you in its application.

Example. Previous example

unbiased coin, toss $n = 1000$ times.

$\mu = \frac{1}{2}$

$\sigma = \frac{1}{2}$.

$$\underbrace{\mathbb{P}(0.4 \le \bar{Y}_n \le 0.6)}_{\text{too hard to compute exactly}}$$

$\Rightarrow$ use CLT to approximate.

$$\bar{Y}_n \simeq \frac{1}{2} + \frac{\frac{1}{2}}{\sqrt{1000}} Z$$

$$\mathbb{P}(0.4 \le \bar{Y}_{1000} \le 0.6) = \mathbb{P}\left(-0.1 \le \frac{1}{20\sqrt{10}} Z \le 0.1\right)$$

$$= \mathbb{P}(-2\sqrt{10} \le Z \le 2\sqrt{10})$$

$$= \mathbb{P}(-6.2 \leq Z \leq 6.2)$$

$$= 1 - \mathbb{P}(Z > 6.2)$$

$$- \mathbb{P}(Z < -6.2)$$

$$= 1 - 2(9 \times 10^{-10}).$$

Example 2 *As noted, the standard normal* $Z \sim N(0,1)$ *has mean* 0 *and variance* 1, *and pdf*

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

The general normal $X \sim N(\mu, \sigma^2)$ *has mean* μ *and variance* σ^2, *and its pdf is*

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The pdf $f_X(x)$ *is used to compute expected values by*

$$\mathbb{E}[X^k] = \int_{-\infty}^{\infty} x^k\, f_X(x)\, dx$$

We note that every normal can be written as a translated and rescaled version of the standard normal: if $X \sim N(\mu, \sigma^2)$ *then*

$$X = \mu + \sigma\, Z$$

meaning that $(X - \mu)/\sigma$ *is a standard normal. This is very useful for calculations, for example: supposing* $X \sim N(2, 9)$, *what is the probability for* X *to be greater than* 8*? We have*

$$X = 2 + 3Z \Rightarrow \mathbb{P}(X > 8) = \mathbb{P}(Z > 2) = 0.023$$

(the number 0.023 *is obtained from z-tables).*

Tables of the Normal Distribution

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

Far Right Tail Probabilities

Z	P{Z to oo}	Z	P{Z to oo}	Z	P{Z to oo}	Z	P{Z to oo}
2.0	0.02275	3.0	0.001350	4.0	0.00003167	5.0	2.867 E-7
2.1	0.01786	3.1	0.0009676	4.1	0.00002066	5.5	1.899 E-8
2.2	0.01390	3.2	0.0006871	4.2	0.00001335	6.0	9.866 E-10
2.3	0.01072	3.3	0.0004834	4.3	0.00000854	6.5	4.016 E-11
2.4	0.00820	3.4	0.0003369	4.4	0.000005413	7.0	1.280 E-12
2.5	0.00621	3.5	0.0002326	4.5	0.000003398	7.5	3.191 E-14
2.6	0.004661	3.6	0.0001591	4.6	0.000002112	8.0	6.221 E-16
2.7	0.003467	3.7	0.0001078	4.7	0.000001300	8.5	9.480 E-18
2.8	0.002555	3.8	0.00007235	4.8	7.933 E-7	9.0	1.129 E-19
2.9	0.001866	3.9	0.00004810	4.9	4.792 E-7	9.5	1.049 E-21

Example 3 *A fair die is rolled 1000 times. Use the CLT to estimate the probability that the number of 6's is greater than 180. Define for all $1 \le i \le 1000$*

$$Y_i = \begin{cases} 1 & \text{if the } i\text{th roll is a } 6 \\ 0 & \text{if the } i\text{th roll is not } 6 \end{cases}$$

$p = \frac{1}{6}$

Let $X = Y_1 + Y_2 + \cdots + Y_{1000}$, then X is equal to the number of 6's rolled. Now $\mathbb{E}[Y] = 1/6$ and $\mathrm{VAR}[Y] = 5/36$, so the CLT gives

$p(1-p) = \frac{5}{36}$

$$X \simeq \frac{1000}{6} + \sqrt{1000}\,\sqrt{5/36}\,Z$$

$$\boxed{\sum Y_i = n\mu + \sigma\sqrt{n}\, Z}$$

CLT.

Therefore

$$\mathbb{P}(X > 180) \simeq \mathbb{P}\left(Z > \frac{180 - 1000/6}{\sqrt{5000/36}}\right) = \mathbb{P}(Z > 1.13) = 0.13$$

$1 - \mathbb{P}(Z \le 1.13)$
$= 1 - 0.8708$

$$\mathbb{P}\left(\frac{1000}{6} + \sqrt{1000}\sqrt{\frac{5}{36}}\, Z > 180\right) = \mathbb{P}\left(\sqrt{1000}\sqrt{\frac{5}{36}}\, Z > 180 - \frac{1000}{6}\right)$$

$$= \mathbb{P}\left(Z > \frac{180 - \frac{1000}{6}}{\sqrt{1000}\sqrt{\frac{5}{36}}}\right)$$

What is the accuracy of the CLT? The Berry-Esseen Theorem gives the general bound

$$\left| P(Z_n < a) - P(Z < a) \right| \leq \frac{C\rho}{\sigma^3 \sqrt{n}}$$ terrible bound! $\frac{1}{\sqrt{n}}$

where C is a constant (not more than 0.48), and $\rho = \mathbb{E}|Y_i - \mu|^3$.

In most practical applications

the rate of convergence is exponential. e^{-kn}

Central Limit Theorem: sketch of proof

The main idea of the proof is to consider the moment generating function of Z_n, that is

$$M_n(t) = \mathbb{E}[e^{tZ_n}], \quad t \in \mathbb{R}$$

and show that for every number t it converges to the moment generating function of the standard normal, which is

$$M(t) = \mathbb{E}[e^{tZ}] = e^{t^2/2}$$

(this result for the standard normal follows from a simple calculation using Gaussian integrals). Once this convergence has been shown for every t, it follows that all the moments of Z_n converge to the corresponding moments of Z, and this gives the result. The proof of convergence follows a calculation which we sketch below. First, for each $i = 1, \ldots, n$ we define

$$X_i = \frac{Y_i - \mu}{\sigma}$$

It follows that $\mathbb{E}[X_i] = 0$ and $\mathrm{VAR}[X_i] = 1$, and

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

Also, since the $\{X_i\}$ are independent we can compute

$$M_n(t) = \mathbb{E}[e^{tn^{-1/2} \sum_i X_i}] = \left(\mathbb{E}[e^{tn^{-1/2} X}]\right)^n$$

Now we just observe that for large n, using the Taylor series for the exponential

$$E[e^{tn^{-1/2}X}] = 1 + E[tn^{-1/2}X] + \frac{1}{2}\,\mathbb{E}[t^2 n^{-1} X^2] + \cdots = 1 + 0 + \frac{t^2}{2n} + \cdots$$

The remainder terms go to zero faster than n^{-1} as $n \to \infty$, so we can ignore these to compute

$$M_n(t) = \left(1 + \frac{t^2}{2n}\right)^n \to e^{t^2/2} = M(t)$$

Simple Random Walk

The simple random walk starts at zero, and at each step moves one unit either forwards (up) or backwards (down). We write X_n for the position of the random walk after n steps. So the possible values of X_n are

$$Ran(X) = \{-n, -n+2, \ldots, n-2, n\}$$

The pdf of X_n is the list of probabilities

$$\mathbb{P}(X = n - 2k), \quad k = 0, 1, \ldots, n$$

These can be calculated using the binomial distribution as

$$\mathbb{P}(X = n - 2k) = \mathbb{P}(k \text{ down}, n - k \text{ up}) = \binom{n}{k} 2^{-n}$$

R L L R R L R - - -

$\downarrow$

1 0 0 1 1 0 1 - - -

$\downarrow$

H T T H H T H - - -

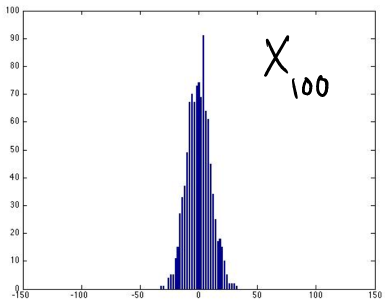
$$\mathbb{P}(X_n = n - 2k) = \mathbb{P}(n-k \text{ steps right, and } k \text{ steps left})$$

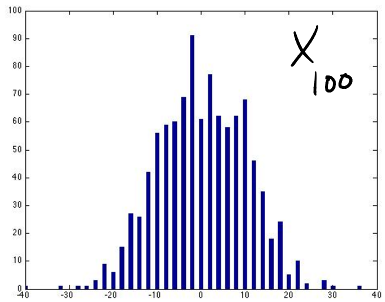
$$= \text{Binomial probability.}$$

$$= \binom{n}{k} \left(\tfrac{1}{2}\right)^k \left(\tfrac{1}{2}\right)^{n-k}$$

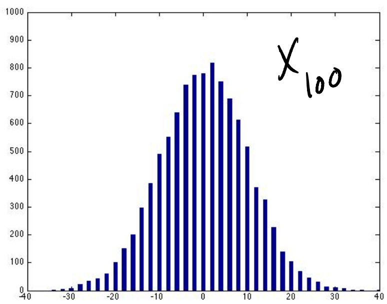
$$= \binom{n}{k} \left(\tfrac{1}{2}\right)^n$$

But first let's do an experiment. Repeat the random walk many times and compute the long-run fraction of occurrences of each possible value of X_n. This is called the empirical pdf of X_n. Here $n = 100$ steps and sample 1000 times.





Note that the re-scaled pdf has a definite shape. Here is a longer run, where $n = 100$ and we sample 10,000 times.



Bell curve

$$\mathbb{P}(X_{100} = 100-2k) = \binom{100}{k} 2^{-100}$$

this is actually a normal pdf b/c CLT.

The pdf looks like a normal curve. We can explain this observation using the CLT. The value $X_n = n - 2k$ is achieved by k down steps and $n - k$ up steps. So we can write

$$X_n = S_1 + S_2 + \cdots + S_n$$

where

$$S_k = \begin{cases} 1 & \text{if } k^{th} \text{ step goes up} \\ -1 & \text{if } k^{th} \text{ step goes down} \end{cases}$$

Each step S_k is an independent random variable with a very simple pmf:

$$\mathbb{P}(S_k = 1) = \mathbb{P}(S_k = -1) = \frac{1}{2}$$

Easily calculate that

$$\mathbb{E}[S_k] = 0, \quad \text{VAR}[S_k] = 1$$

Therefore since $X_n = S_1 + \cdots + S_n$ we get

$$\mathbb{E}[X_n] = 0, \quad \text{VAR}[X_n] = n$$

Note that the S_k are IID, so we can apply the CLT. In this case

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i = n^{-1/2} X_n$$

and therefore the CLT says

$$Z_n \to Z, \quad X_n \simeq n^{1/2}\, Z$$

So for large n, the position X_n is a rescaled standard normal. This explains the graphs shown above.