

Recall from last lecture:

X_1, X_2, X_3, \dots be a sequence of independent identically distributed (iid) random variables, with finite mean $\mu = E(X_i)$ and variance $\sigma^2 = \text{Var}(X_i)$.

$$\bar{X}_n := \frac{X_1 + X_2 + \dots + X_n}{n} \quad \text{sample average}$$

Weak Law of Large Numbers: For all $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0. \quad \text{i.e., } \bar{X}_n \xrightarrow{P} \mu$$

\bar{X}_n converges (in probability) to $\mu \in \mathbb{R}$.

Central Limit Theorem (Baby version):

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq a\right) = P(Z \leq a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-z^2/2} dz = \Phi(a).$$

i.e. $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim \text{Normal}(0,1)$

$Z \sim \text{Normal}(0,1)$
Standard Normal

$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ converges (in distribution) to a standard normal random variable.

Note: This gives information on "how" $(\bar{X}_n - \mu) \rightarrow 0$, as stated above (LLN).

Exercise: Suppose 10 fair dice are rolled. Use the CLT to approximate the probability that the sum of the results is between 30 and 40.

Table 5.1 Area $\Phi(x)$ Under the Standard Normal Curve to the Left of X .

X	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Let X_1, X_2, \dots, X_{10} be the results from each of the 10 dice. (iid)

$$E(X_i) = \frac{1+2+\dots+6}{6} = \frac{7}{2} = \mu$$

$$\text{Var}(X_i) = \dots = \frac{35}{12} = \sigma^2$$

$$\bar{X}_{10} = \frac{X_1 + \dots + X_{10}}{10}$$

$$X = 10 \bar{X}_{10} = X_1 + \dots + X_{10} \quad \text{integer valued.}$$

$$P(29.5 \leq X \leq 40.5) = P(2.95 \leq \bar{X}_{10} \leq 4.05)$$

$$= P\left(\frac{2.95 - \mu}{\sigma/\sqrt{10}} \leq \frac{\bar{X}_{10} - \mu}{\sigma/\sqrt{10}} \leq \frac{4.05 - \mu}{\sigma/\sqrt{10}}\right)$$

$$= P\left(-\frac{0.55}{\sqrt{\frac{35}{12}}} \leq \frac{\bar{X}_{10} - \mu}{\sigma/\sqrt{10}} \leq \frac{0.55}{\sqrt{\frac{35}{12}}}\right)$$

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq a\right) = P(Z \leq a)$$

$$P\left(\frac{\bar{X}_{10} - \mu}{\sigma/\sqrt{10}} \leq a\right) \approx P(Z \leq a) \quad \text{approx.}$$

-1.0184 1.0184

$$\approx P(-1.0184 \leq Z \leq 1.0184)$$

$$\Phi(a) = P(Z \leq a) \quad \Rightarrow \quad \Phi(1.0184) - \Phi(-1.0184)$$

$$\Phi(-a) = 1 - \Phi(a)$$

Using this approx., we treat $P(|\bar{X}_n - d| \leq 0.5) = 95/100$ as follows:

$$\begin{aligned} \frac{95}{100} &= P\left(\frac{|\bar{X}_n - d|}{\sigma/\sqrt{n}} \leq \frac{0.5}{\sigma/\sqrt{n}}\right) = P\left(\frac{|\bar{X}_n - d|}{\sigma/\sqrt{n}} \leq \frac{\sqrt{n}}{4}\right) \stackrel{\text{CLT}}{\approx} P(|Z| \leq \frac{\sqrt{n}}{4}) \\ &= P\left(Z \leq \frac{\sqrt{n}}{4}\right) - P\left(Z \leq -\frac{\sqrt{n}}{4}\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \end{aligned}$$

Need:

$$2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 = 0.95$$

$$\Phi\left(\frac{\sqrt{n}}{4}\right) = \frac{1.95}{2} = 0.975.$$

$1.96 \approx$

From table, $\frac{\sqrt{n}}{4} \approx 1.96$, which means $n \approx 4^2 \cdot 1.96^2 \approx 61.47$

So need at least 62 measurements to achieve the desired 95% accuracy. (Assuming CLT gives good enough approx. when $n=62$.)

Remark:

To avoid using any approx, one can instead

use Chebyshev:

$$E(\bar{X}_n) \stackrel{\text{iid}}{=} E(X_i) = \mu = d, \quad \text{Var}(\bar{X}_n) \stackrel{\text{iid}}{=} \frac{\text{Var}(X_i)}{n} = \frac{4}{n}$$

$$P(|\bar{X}_n - d| > 0.5) \leq \frac{4}{n(0.5)^2} = \frac{16}{n} = 0.05$$

then would need $n = \frac{16}{0.05} = \underline{\underline{320}}$ measurements.

Ex: Suppose the number of students that enroll in Calculus is a Poisson random variable with average $\mu = 100$. If more than 120 enroll, the university will split the class in 2 sections, while if less than 120 enroll, there will be a single section. What is the probability there will be 2 sections?

Recall: If $X \sim \text{Poisson}(\lambda)$, then:

$$E(X) = \lambda, \text{Var}(X) = \lambda$$

$$p(i) = p(X=i) = e^{-\lambda} \frac{\lambda^i}{i!}$$

$$P(X \leq n) = e^{-\lambda} \sum_{i=0}^n \frac{\lambda^i}{i!}$$

$$\lambda = 100$$

$$P(X \geq 120) = \sum_{i=120}^{+\infty} e^{-100} \frac{100^i}{i!}$$

$$= e^{-100} \sum_{i=120}^{+\infty} \frac{100^i}{i!}$$

Using a Computer

$$\approx 0.0282$$

$$\underline{\underline{2.82\%}}$$

From Lecture 19:

$$X_1 \sim \text{Poisson}(\lambda_1), X_2 \sim \text{Poisson}(\lambda_2)$$

$$\Rightarrow X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$$

We can use CLT to find an approximation;

$$X \sim \text{Poisson}(100)$$

$$X = X_1 + X_2 + \dots + X_{100}$$

$$X_i \sim \text{Poisson}(1), X_i \text{ iid}$$

integer valued

$$P(X \geq 120) \approx P(X \geq 119.5)$$

$$= P(X_1 + \dots + X_{100} \geq 119.5)$$

Table 5.1 Area $\Phi(x)$ Under the Standard Normal Curve to the Left of X .

X	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

$$\bar{X}_{100} = \frac{X_1 + \dots + X_{100}}{100}$$

$$E(\bar{X}_{100}) = E(X_i) = 1$$

$$\text{Var}(\bar{X}_{100}) = \frac{1}{100} \text{Var}(X_i) = 1$$

$$P\left(\frac{X_1 + \dots + X_{100}}{100} \geq \frac{119.5}{100}\right) = P(\bar{X}_{100} \geq 1.195)$$

$$= P\left(\frac{\bar{X}_{100} - \mu}{\sigma/\sqrt{100}} \geq \frac{1.195 - 1}{1/\sqrt{100}}\right)$$

$$= P\left(\frac{\bar{X}_{100} - \mu}{\sigma/\sqrt{100}} \geq 1.95\right)$$

CLT

$$\approx P(Z \geq 1.95) = 1 - \Phi(1.95)$$

$$\approx 1 - 0.9744 = 0.0256$$

2.56%

Central Limit Theorem (Grown-up version).

may have different distributions (need not be i.i.d.)

Let X_1, X_2, \dots be a sequence of independent random variables, with $\mu_i = E(X_i)$ and $\sigma_i^2 = \text{Var}(X_i)$. If X_i are uniformly bounded i.e. there exists $M \in \mathbb{R}$ s.t. for all i , $P(|X_i| < M) = 1$, and

$\sum_{i=1}^{+\infty} \sigma_i^2 = +\infty$, then for all $a \in \mathbb{R}$,

$Z \sim \text{Normal}(0,1)$
Standard Normal.

$$\lim_{n \rightarrow \infty} P\left(\frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \leq a\right) = P(Z \leq a) = \Phi(a).$$

i.e., letting $S_n^2 := \sum_{i=1}^n \sigma_i^2$, $\frac{1}{S_n} \sum_{i=1}^n (X_i - \mu_i) \xrightarrow{d} Z \sim \text{Normal}(0,1)$

We will not give a proof, but let us see how to recover the Baby version of CLT from it:

$E(X_i) = \mu_i = \mu.$ \leftarrow uniform boundedness $< +\infty$

has limit $+\infty$ as $n \rightarrow \infty$ if $\sigma > 0$.

$\text{Var}(X_i) = \sigma_i^2 = \sigma^2. \Rightarrow S_n^2 = \sum_{i=1}^n \sigma_i^2 = n \cdot \sigma^2 \Rightarrow S_n = \sqrt{n} \cdot \sigma.$

$$\underbrace{\frac{1}{\sqrt{n} \cdot \sigma} \sum_{i=1}^n (X_i - \mu)}_{\downarrow d} = \frac{1}{\sqrt{n} \cdot \sigma} \left(\left(\sum_{i=1}^n X_i \right) - n \cdot \mu \right) = \frac{1}{\sqrt{n} \cdot \sigma} (n \bar{X}_n - n \mu)$$

$$= \frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu) = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}$$

Z

so this reduces to what we had seen before,

Baby CLT: $\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} Z.$