

# Central Limit Theorem

## The Central Limit Theorem (CLT)

- WLLN:  $\frac{X_1 + \cdots + X_n}{n} \rightarrow E[X]$

- CLT:  $X_1 + \cdots + X_n \approx \text{normal}$

- precise statement
- universality, usefulness
- many examples
- refinement for discrete r.v.s
- application to polling

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

- $X_1, \dots, X_n$  i.i.d., finite mean  $\mu$  and variance  $\sigma^2$

- $S_n = X_1 + \dots + X_n$

variance:  $n\sigma^2$

- $M_n = \frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}$

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

- $\frac{S_n}{\sqrt{n}} = \frac{X_1 + \dots + X_n}{\sqrt{n}}$

variance:  $\sigma^2$

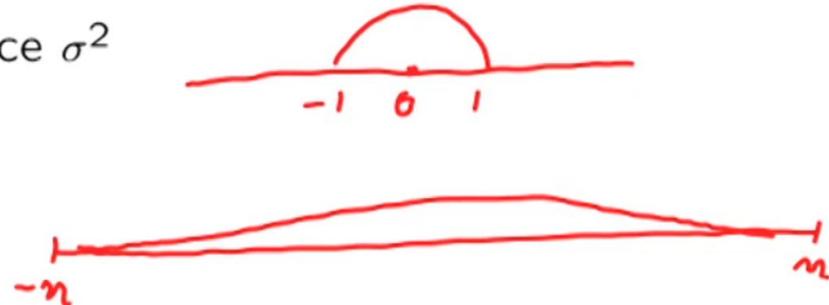
## Different scalings of the sum of i.i.d. random variables

- $X_1, \dots, X_n$  i.i.d., finite mean  $\mu$  and variance  $\sigma^2$



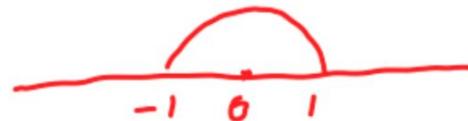
## Different scalings of the sum of i.i.d. random variables

- $X_1, \dots, X_n$  i.i.d., finite mean  $\mu$  and variance  $\sigma^2$
- $S_n = X_1 + \dots + X_n$   
variance:  $n\sigma^2$



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

- $X_1, \dots, X_n$  i.i.d., finite mean  $\mu$  and variance  $\sigma^2$



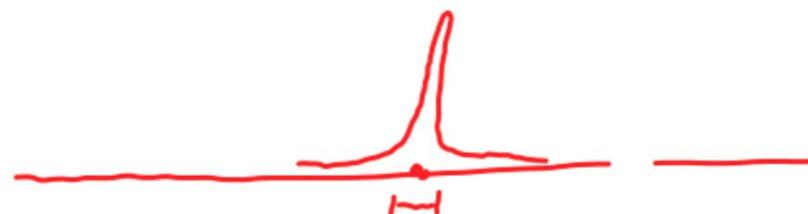
- $S_n = X_1 + \dots + X_n$

variance:  $n\sigma^2$



- $M_n = \frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}$

variance:  $\frac{\sigma^2}{n} \rightsquigarrow 0$



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

- $X_1, \dots, X_n$  i.i.d., finite mean  $\mu$  and variance  $\sigma^2$



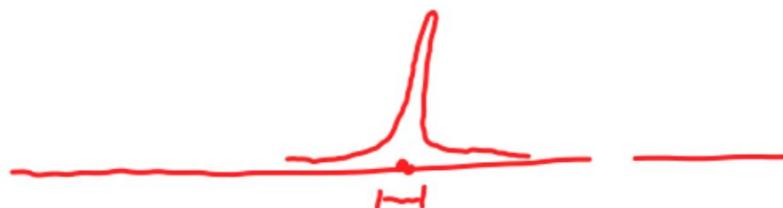
- $S_n = X_1 + \dots + X_n$

variance:  $n\sigma^2$



- $M_n = \frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}$

variance:  $\frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$



- $\frac{S_n}{\sqrt{n}} = \frac{X_1 + \dots + X_n}{\sqrt{n}}$

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



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

- $X_1, \dots, X_n$  i.i.d., finite mean  $\mu$  and variance  $\sigma^2$



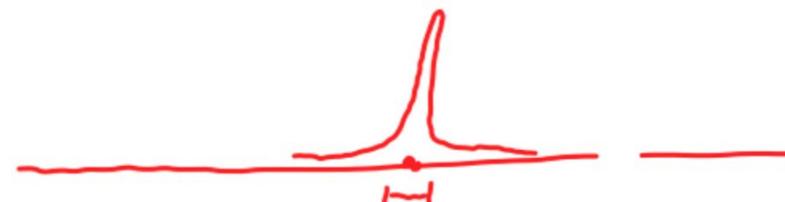
- $S_n = X_1 + \dots + X_n$

variance:  $n\sigma^2$



- $M_n = \frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}$

variance:  $\frac{\sigma^2}{n} \rightsquigarrow 0$



- $\frac{S_n}{\sqrt{n}} = \frac{X_1 + \dots + X_n}{\sqrt{n}}$

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



So the width of this distribution stays constant. Does it approach a certain shape?



## The Central Limit Theorem (CLT)

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

## The Central Limit Theorem (CLT)

- $X_1, \dots, X_n$  i.i.d., finite mean  $\mu$  and variance  $\sigma^2$

- $S_n = X_1 + \dots + X_n$  variance:  $n\sigma^2$

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

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

## The Central Limit Theorem (CLT)

- $X_1, \dots, X_n$  i.i.d., finite mean  $\mu$  and variance  $\sigma^2$

- $S_n = X_1 + \dots + X_n$  variance:  $n\sigma^2$

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

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

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

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

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

## Usefulness of the CLT

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

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

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

## What exactly does the CLT say? — Theory

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

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

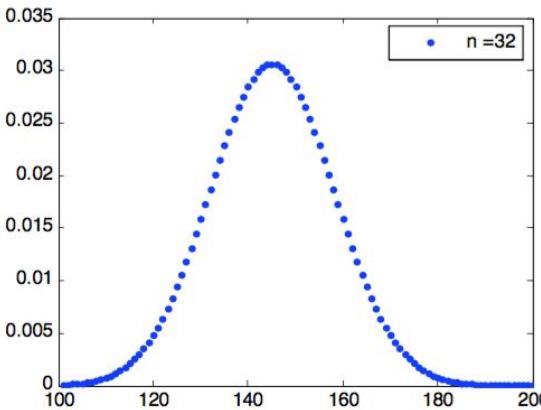
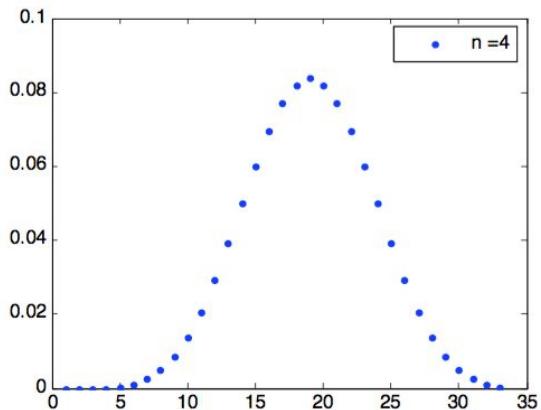
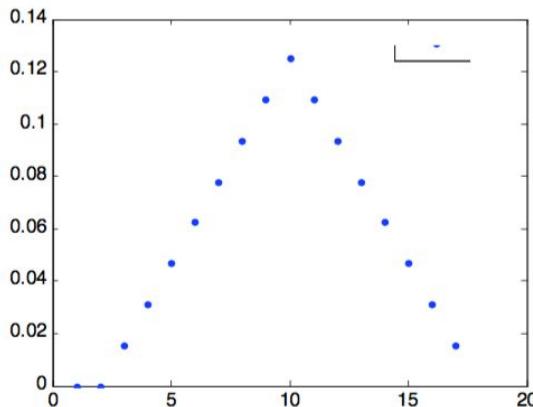
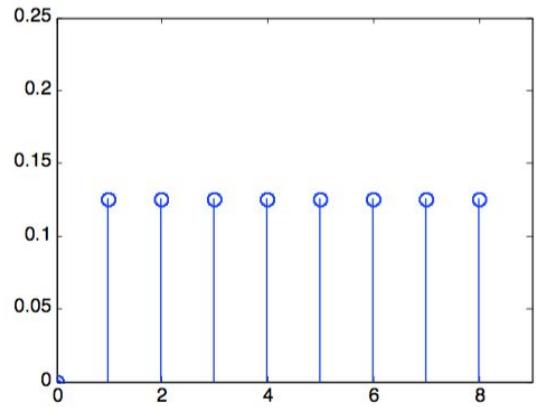
- CDF of  $Z_n$  converges to normal CDF
- results for convergence of PDFs or PMFs (with more assumptions)
- results without assuming that the  $X_i$  are identically distributed
- results under “weak dependence”
- proof: uses “transforms”:  $E[e^{sZ_n}] \rightarrow E[e^{sZ}]$ , for all  $s$

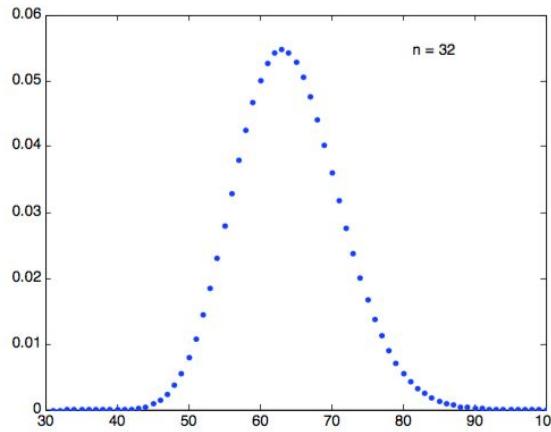
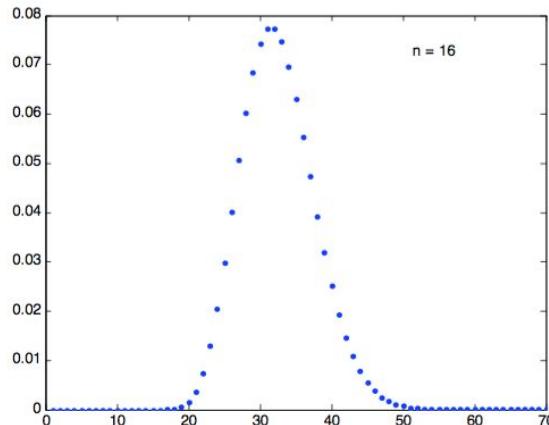
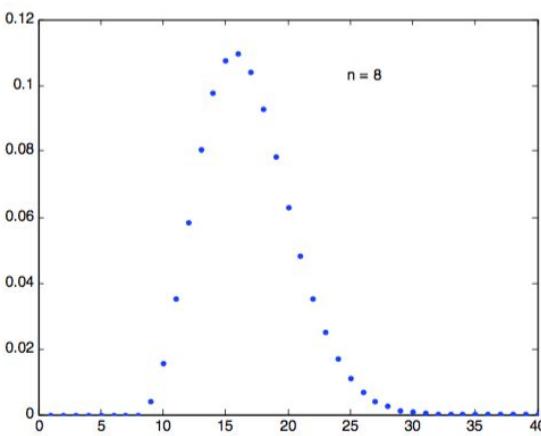
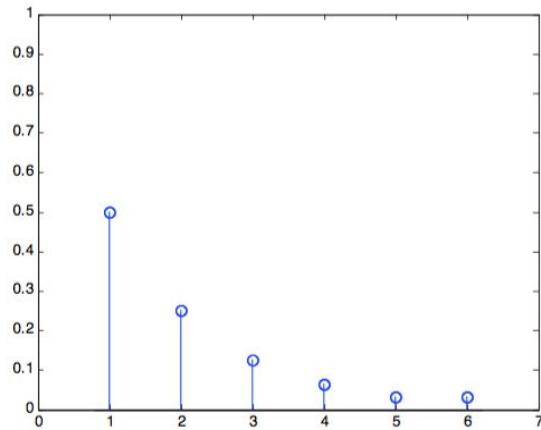
## What exactly does the CLT say? — Practice

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

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

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





## Example 1

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

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

$$\mu = \sigma = 2$$

$$P(S_n \geq 210)$$

	.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

## Example 1

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

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

$$\mu = \sigma = 2$$

$$P(S_n \geq 210)$$

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

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

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

$$= 1 - 0.6915 = 0.3085$$

	.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	0.6915	.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

## Example 2

- $P(S_n \leq a) \approx b$  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  $P(S_n \geq a) \approx 0.05$ .

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

$$\mu = \sigma = 2$$

	.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

## Example 2

- $P(S_n \leq a) \approx b$  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  $P(S_n \geq a) \approx 0.05$ .

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

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

0.95

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

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

$$\mu = \sigma = 2$$

	.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

### Example 3

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

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

$$\mu = \sigma = 2$$

	.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

### Example 3

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

$$P\left(\frac{S_n - \mu n}{\sigma \sqrt{n}} \geq \frac{210 - \mu n}{\sigma \sqrt{n}}\right) \approx 1 - \Phi\left(\frac{210 - \mu n}{\sigma \sqrt{n}}\right) \approx 0.05$$

0.95

$$\frac{210 - \mu n}{\sigma \sqrt{n}} = 1.645$$

$n = 89$

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

$$\mu = \sigma = 2$$

	.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

## Example 4

- $P(S_n \leq a) \approx b$  given two parameters, find the third
- Package weights  $X_i$ , i.i.d. exponential,  $\lambda = 1/2$ ;
- Load container until weight exceeds 210  
 $N$ : number of packages loaded
- $P(N > 100)$

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

$$\mu = \sigma = 2$$

	.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

## Example 4

- $P(S_n \leq a) \approx b$  given two parameters, find the third
- Package weights  $X_i$ , i.i.d. exponential,  $\lambda = 1/2$ ;
- Load container until weight exceeds 210  
 $N$ : number of packages loaded
- $P(N > 100)$

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

$$\approx \Phi\left(\frac{210 - 200}{20}\right) = \Phi(0.5) = 0.6915$$

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

$$\mu = \sigma = 2$$

	.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

## Normal approximation to the binomial

- $X_i$ : independent, Bernoulli( $p$ );  $0 < p < 1$
- $S_n = X_1 + \dots + X_n$ : Binomial( $n, p$ )
  - mean  $np$ , variance  $np(1 - p)$
- $n = 36, p = 0.5$ ; find  $P(S_n \leq 21)$

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

- CDF of  $\frac{S_n - np}{\sqrt{np(1 - p)}}$  → standard normal

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

	.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

## Normal approximation to the binomial

- $X_i$ : independent, Bernoulli( $p$ );  $0 < p < 1$
- $S_n = X_1 + \dots + X_n$ : Binomial( $n, p$ )
  - mean  $np$ , variance  $np(1 - p)$
- $n = 36, p = 0.5$ ; find  $P(S_n \leq 21)$

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

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

$$= P(Z_n \leq 1) \approx \Phi(1) = 0.8413$$

$Z_n$

- CDF of  $\frac{S_n - np}{\sqrt{np(1 - p)}}$  → standard normal

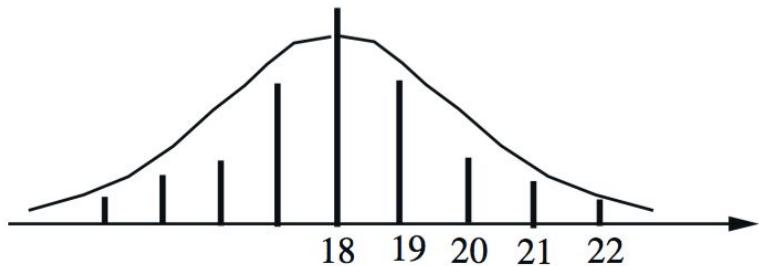
$$\sum_{k=0}^{21} \binom{36}{k} \left(\frac{1}{2}\right)^{36} = 0.8785$$

	.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

## The $1/2$ correction for integer random variables

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

$$P(S_n \leq 21.5)$$



true value 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

## The $1/2$ correction for integer random variables

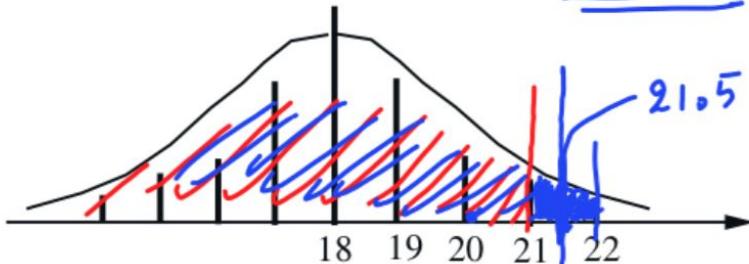
- $0.8413 \approx \underline{P(S_n \leq 21)} = \underline{P(S_n < 22)}$ , because  $S_n$  is integer

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

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

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

$$\approx \underline{\Phi(1.17)} = 0.8790$$

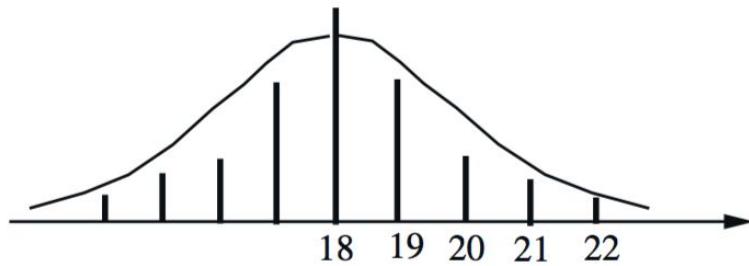


true value  $\underline{\underline{0.8785}}$

	.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

## De Moivre–Laplace CLT to the binomial

$$\mathbf{P}(S_n = 19)$$



- 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)

## The pollster's problem revisited

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

$$\mathbf{P}(|M_n - p| \geq .01)$$

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

## The pollster's problem revisited

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

$$E[X_i] = p = \mu$$
$$\sigma = \sqrt{p(1-p)}$$

$$P(|M_n - p| \geq 0.01) = P\left(|Z_n| \geq \frac{0.01\sqrt{n}}{\sigma}\right) \approx P\left(|Z| \geq \frac{0.01\sqrt{n}}{\sigma}\right)$$

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

$$\left| \frac{S_n - np}{n} \right| \geq 0.01$$

$$\left| \frac{S_n - np}{\sqrt{n}\sigma} \right| \geq \frac{0.01\sqrt{n}}{\sigma}$$

## The pollster's problem revisited

$$\mathbf{P}\left(|M_n - p| \geq .01\right) \approx \mathbf{P}\left(|Z| \geq \frac{.01\sqrt{n}}{\sigma}\right) \quad \mathbf{P}\left(|Z| \geq .02\sqrt{n}\right)$$

- Try  $n = 10,000$

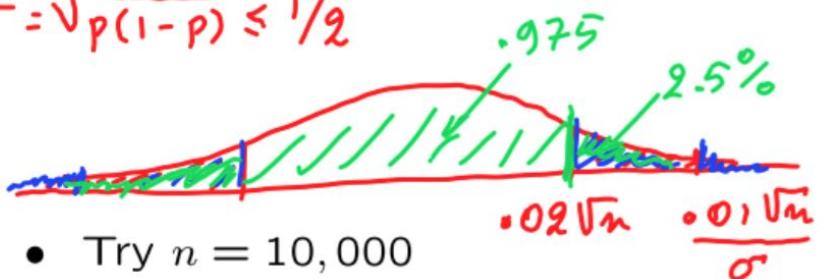
- Specs:  $\mathbf{P}\left(|M_n - p| \geq .01\right) \leq .05$

	.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

## The pollster's problem revisited

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

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



- Try  $n = 10,000$

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

$$= 2(1 - 0.9772) = 0.046$$

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

$$\Phi(0.02\sqrt{n}) = 0.975$$

$$0.02\sqrt{n} = 1.96 \Rightarrow n = 9604$$

	.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