Polling + Auctions

Anna Karlin

Agenda

idealized

- An application of The Central Limit Theorem – Polling
- A glimpse of auction theory

Magic mushrooms

Yesterday, Oregonians ~~are~~ were voting on whether to legalize the therapeutic use of “magic mushrooms”.

If you take a "heroic" dose, supposedly, “the ego dissolves and the user feels inseparable from the rest of the universe.”

Poll to determine fraction of population that will be voting in favor.

- call up a random sample
- report empirical fraction

is this good estimate? how choose n?

This poll is accurate to within 2%, 19 times out of 20.

$\frac{1}{20}$ chance estimate garbage

$\frac{19}{20}$ estimate is within 2% of the answer.

CLT → Polling on magic mushrooms

Population size N, p true fraction voting in favor.

We don't know p

for $i = 1$ to n $\left(\frac{1}{N}\right)$

pick uniformly random person to call

$$X_i = \begin{cases} 1 & \text{voting in favor} \\ 0 & \text{o.w.} \end{cases}$$

$X_1 \quad X_2 \quad \ldots \quad X_n$

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

$n = 500$

Estimate $\Rightarrow \frac{271}{500} = 0.542$

$$\sum_{i=1}^{n} X_i = 271$$

What kind of r.v. is X_i?

	type of r.v.	$E(X_i)$	$Var(X_i)$
a)	Bernoulli	p	$p(1-p)$
b)	Bernoulli	p	p^2
c)	Geometric	p	$p(1-p)$
d)	Poisson	p	$p(1-p)$

$$X_i = \begin{cases} 1 & \text{with prob } p \\ 0 & 1-p. \end{cases}$$

$$\boxed{\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

empirical mean.

	$E(\bar{X})$	$Var(\bar{X})$
a)	np	$np(1-p)$
b)	p	$p(1-p)$
c)	p	$\frac{p(1-p)}{n}$
d)	$\frac{p}{n}$	$\frac{p(1-p)}{n}$

$X_1, X_2, \ldots, X_n$ are independent.

i.i.d. $Ber(p)$

$$Var\left(\sum_{i=1}^{n} X_i\right) = \underline{np(1-p)}$$

$$Var(aY) = a^2 Var(Y)$$

The Central Limit Theorem

Consider i.i.d. (independent, identically distributed) random vars
$X_1, X_2, X_3, \ldots X_n$
Where X_i has $\mu = E[X_i]$ and $\sigma^2 = Var[X_i]$

standardize.

As $n \to \infty$,

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \longrightarrow N(0,1)$$

Restated: As $n \to \infty$,

$$\bar{X} = \boxed{\frac{1}{n}\sum_{i=1}^{n} X_i} \longrightarrow N\left(\mu, \frac{\sigma^2}{n}\right)$$

$\mu = p$
$\sigma^2 = p(1-p)$

$$\bar{X} \longrightarrow \boxed{N\left(p, \frac{p(1-p)}{n}\right)}$$

By CLT

In the limit as $n \to \infty$ $\bar{X}$ is

a) $N(0,1)$

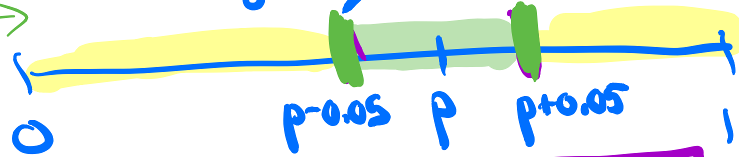
b) $\boxed{N\left(p, \frac{p(1-p)}{n}\right)}$

c) $N(p, np(1-p))$

d) I don't know.

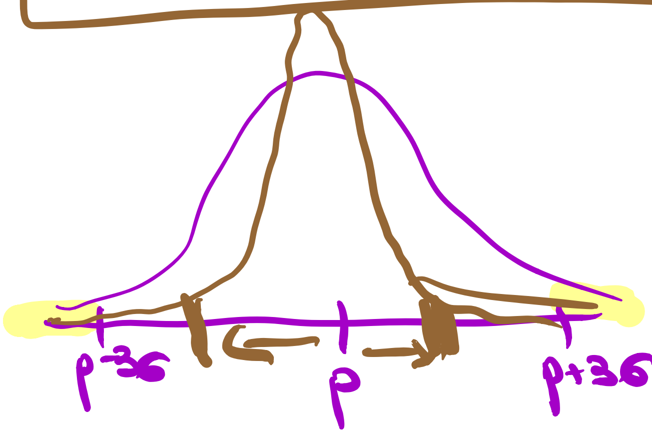
$$(*) \quad \boxed{Pr(|\bar{X} - p| > 0.05) < 0.02}$$

$Pr(\text{green}) > 0.98$



for what n does (*) hold?

98% of time my estimate is within 5% of correct answer.



$\boxed{\frac{p(1-p)}{n}}$

normal distn's,

$\mu \pm 3\sigma$ 0.999

$$Pr(\underbrace{p-0.05 \le \bar{X} \le p+0.05})$$

μ σ^2

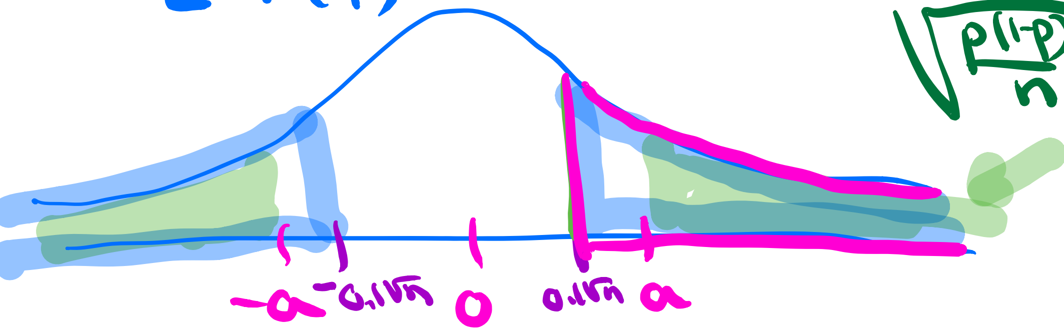
$$\Pr\left(|\bar{X} - p| > 0.05\right)$$

$$= \Pr\left(\left|\underbrace{\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}}}_{Z \sim N(0,1)}\right| > \underbrace{\boxed{\frac{0.05}{\sqrt{\frac{p(1-p)}{n}}}}}_{a}\right)$$

$$\bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

$$\bar{X} - p \sim N\left(0, \frac{p(1-p)}{n}\right)$$

$$\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$



$$\boxed{\Pr(|Z| > a)}$$

$$\le \Pr(|Z| > \underbrace{0.1\sqrt{n}})$$

green area ≤ blue area.

$$a = \frac{0.05\sqrt{n}}{\sqrt{p(1-p)}} \ge \frac{0.05\sqrt{n}}{\frac{1}{2}} = \boxed{0.1\sqrt{n}}$$

$$\underbrace{\sqrt{p(1-p)}}_{\text{maximized at } p = \frac{1}{2}} \le \frac{1}{2}$$

find n s.t.

$$\Pr(|Z| > 0.1\sqrt{n}) \leq 0.02$$

$$= \Pr(Z > 0.1\sqrt{n}) + \Pr(Z < -0.1\sqrt{n})$$

$$= 2\Pr(Z > 0.1\sqrt{n})$$

$$= 2(1 - \Pr(Z \leq 0.1\sqrt{n}))$$

find n s.t.

$$= \boxed{2(1 - \Phi(0.1\sqrt{n})) \leq 0.02}$$

$$1 - \Phi(0.1\sqrt{n}) \leq 0.01$$

$$0.99 \leq \Phi(0.1\sqrt{n})$$

✓

$$0.1\sqrt{n} \geq 2.33$$

$$n \geq \left(\frac{2.33}{0.1}\right)^2 = 543$$

take $n = 543$

$\bar{X}$ is within 0.05 of true p 98% of time

The Standard Normal CDF

$\Phi(2.33)=0.99$

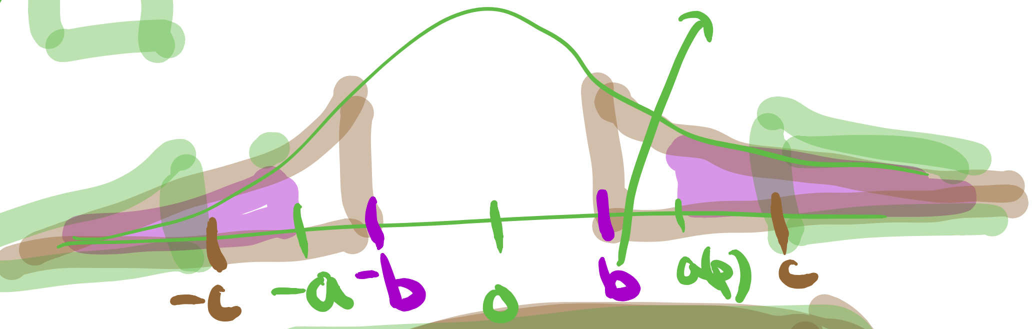
Φ Table: $\mathbb{P}(Z \le z)$ when $Z \sim \mathcal{N}(0,1)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999

$\Pr(|Z|>a)\ \boxed{\le}$

$$\frac{0.05\sqrt{n}}{\sqrt{p(1-p)}}$$

do not know p.



$\Pr(|Z|>a)$

$\Pr(|Z|>b)$

$\Pr(|Z|>c)$

$X_1 \sim Ber(p)$

$\Downarrow$

$X_1 + X_2$

$X_2 \sim Ber(p)$

Pr	X_1	X_2	X_1+X_2
$(1-p)^2$	0	0	0
$(1-p)p$	0	1	1
$p(1-p)$	1	0	1
p^2	1	1	2

$$Pr(X_1 + X_2 = 1)$$

$$= Pr(X_1 = 0, X_2 = 1)$$

$$+ Pr(X_1 = 1, X_2 = 0)$$