Box models

Expected Value + Standard Error

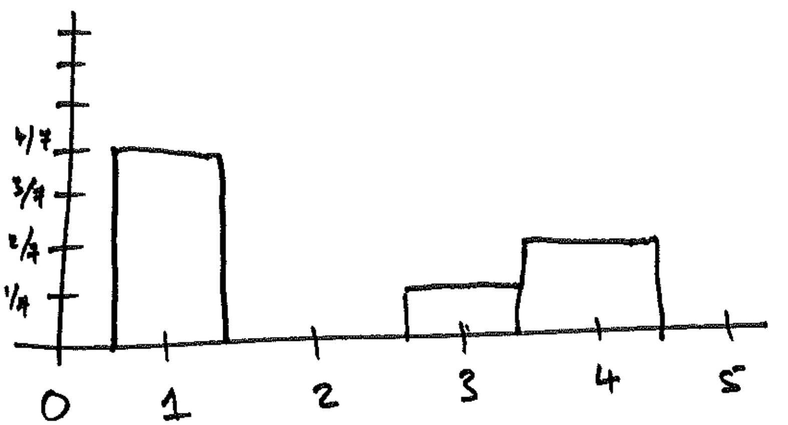
Normal approximation

→ sum of draws.

Probability Histograms.

$\boxed{1}\ \boxed{1}\ \boxed{1}\ \boxed{1}\ \boxed{3}\ \boxed{4}\ \boxed{4}$

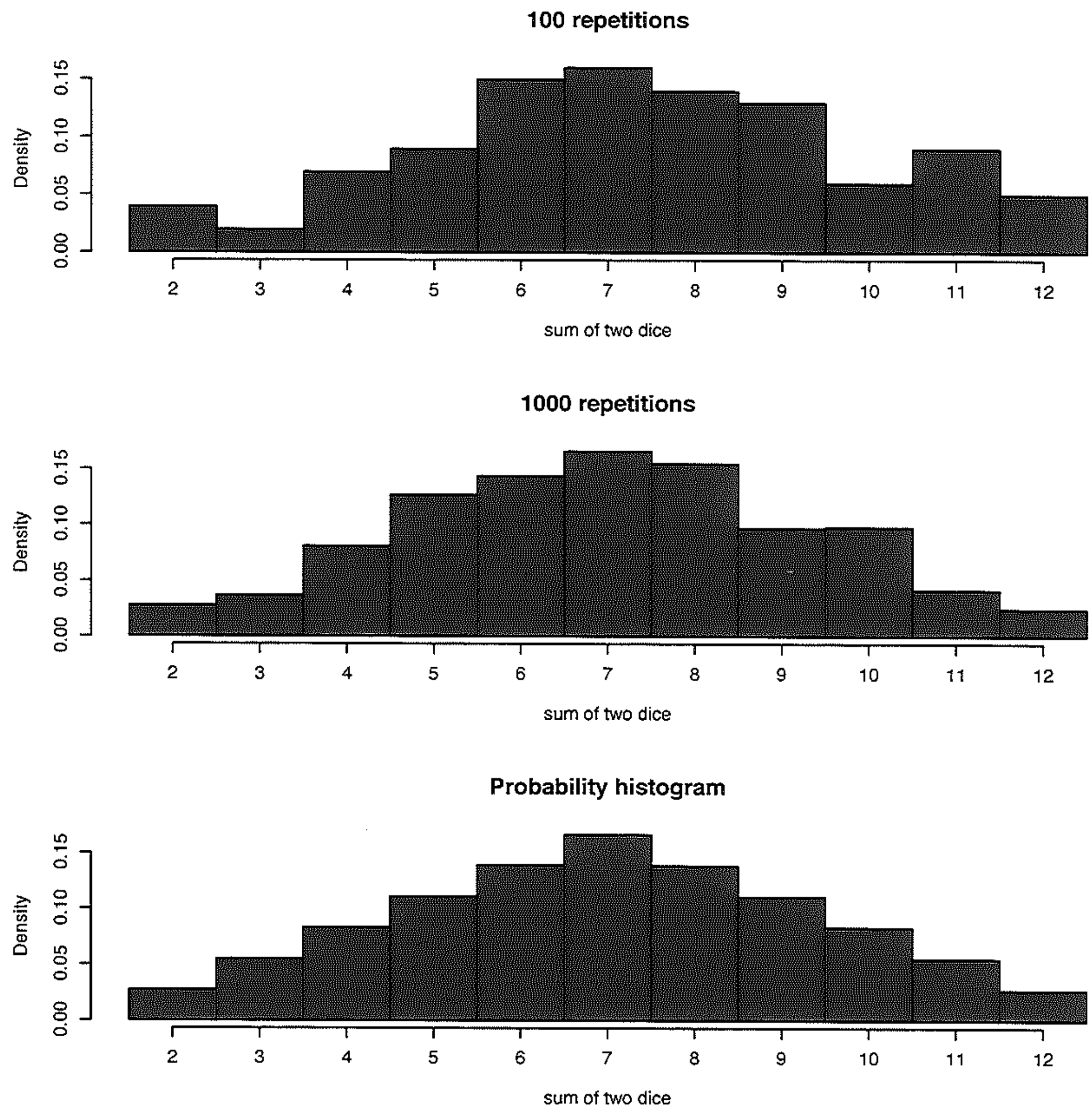
value	prob.
1	4/7
3	1/7
4	2/7



area of bar represents probability

total area = 1.

Represents chance not frequency of data.



100 repetitions
Density
0.00
0.05
0.10
0.15
2
3
4
5
6
7
8
9
10
11
12
sum of two dice
1000 repetitions
Density
0.00
0.05
0.10
0.15
2
3
4
5
6
7
8
9
10
11
12
sum of two dice
Probability histogram
Density
0.00
0.05
0.10
0.15
2
3
4
5
6
7
8
9
10
11
12
sum of two dice

If you repeat an experiment over and over, the experimental (empirical) histogram converges to the probability histogram.

Example : rolling two dice.

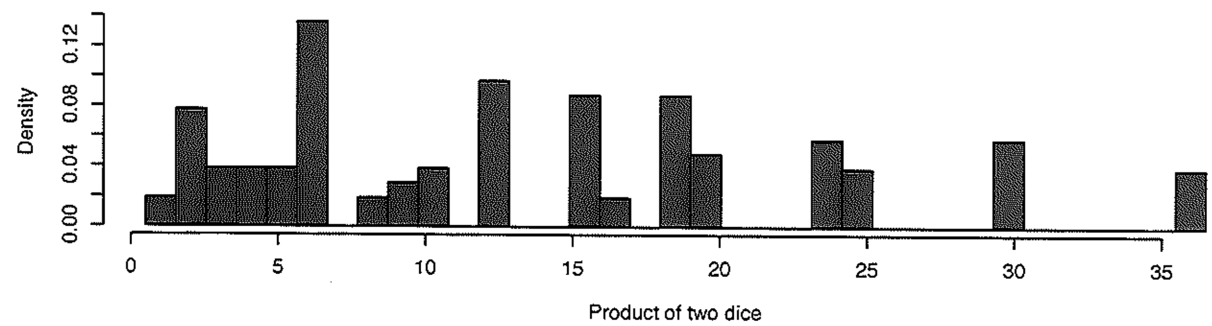
	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

sum	prob.
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

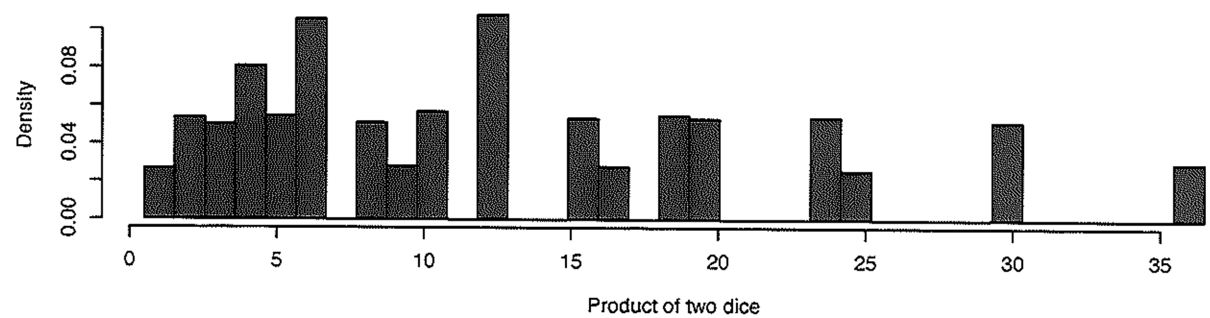
The relative size of the fluctuations is $\frac{\sqrt{N}}{N} = \frac{1}{\sqrt{N}}$

which gets smaller as N gets larger.

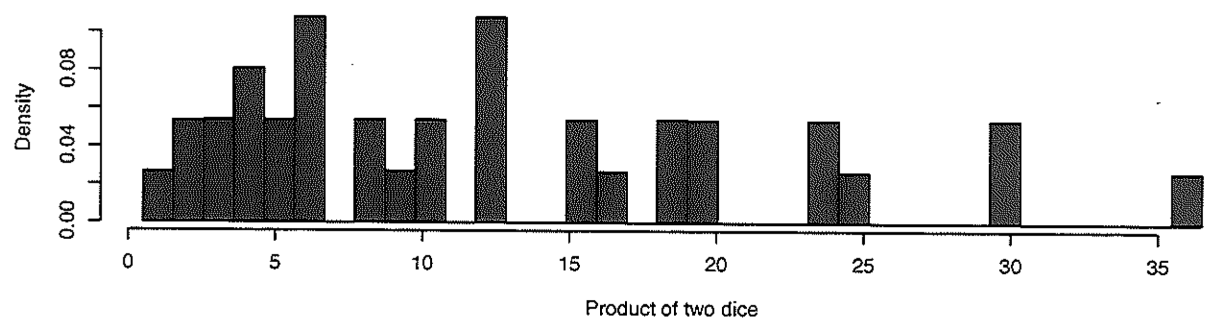
100 rolls



10000 rolls



Probability histogram



Product of the numbers on two dice.

	1	2	3	4	5	6
1	1	2	3	4	5	6.
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

product	1	2	3	4	5	6	7	8	9	10
	1/36	2/36	2/36	3/36	2/36	4/36	-	2/36	1/36	2/10

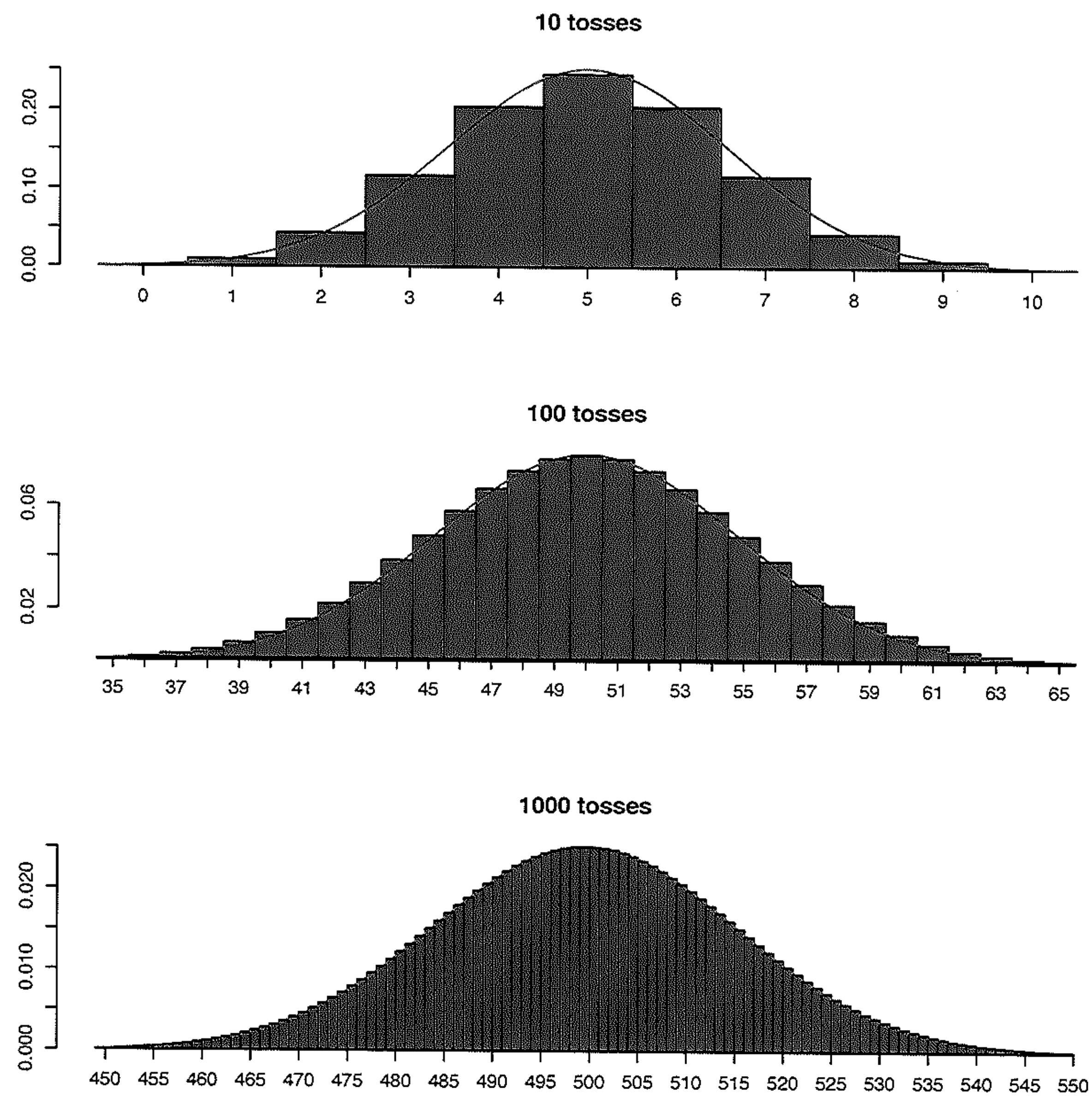
	11	12	13	14	15	16	17	18	19	20
	-	4/36	-	-	2/36	1/36	-	2/36	-	2/36

	21	22	23	24	25	26	27	28	29	30
				2/36	1/36					2/36

	31	32	33	34	35	36
						1/36

Green -
probability
histograms

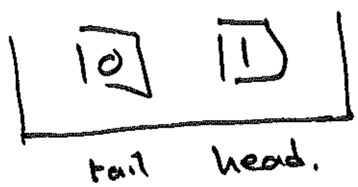
Red curve -
Normal curve



The histogram for the product is much less regular
- gaps.

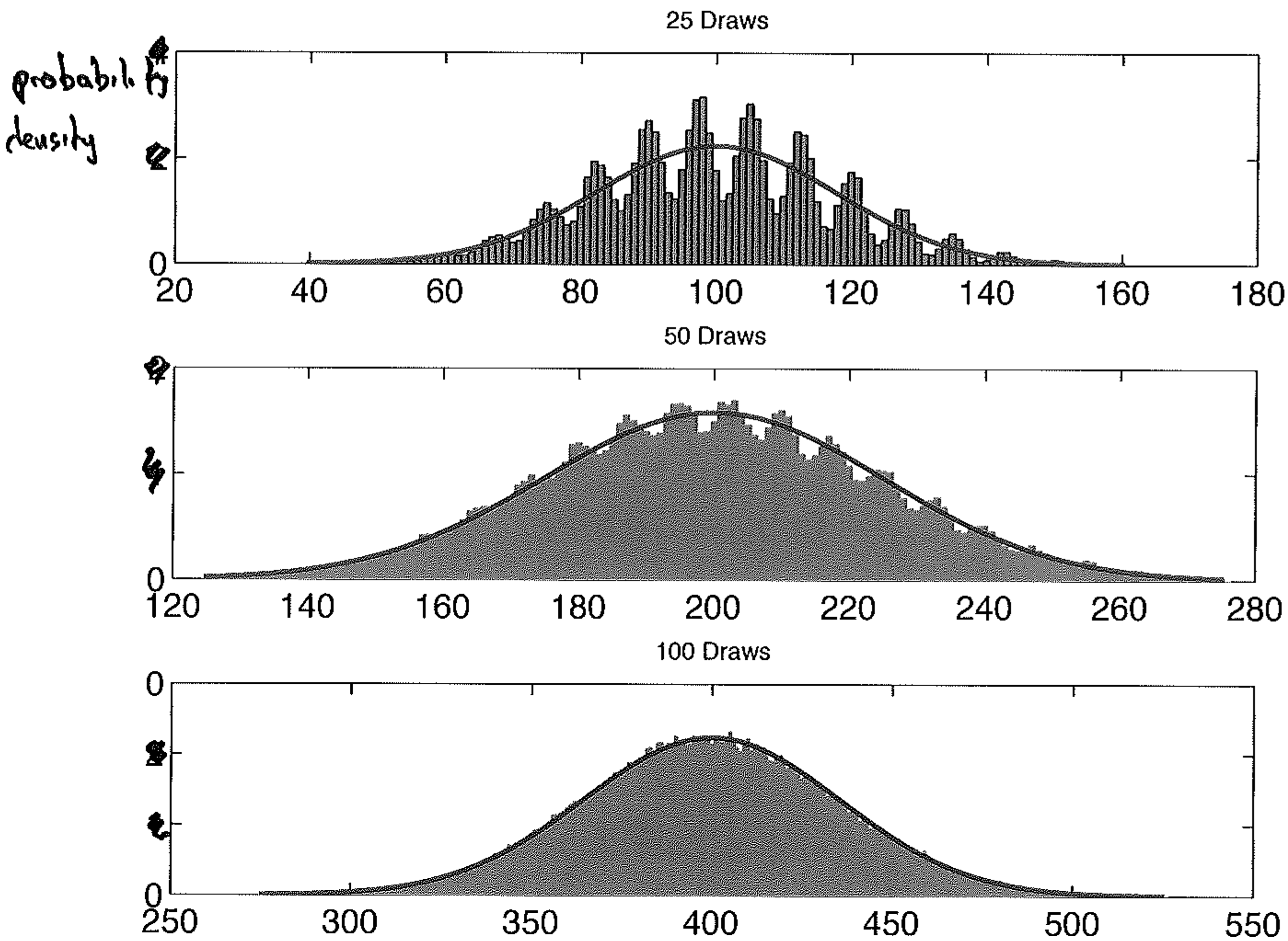
Empirical histogram still converges to the probability histogram

What's so special about the sum?



sum of draws is # heads in N tosses of a coin.

The probability histogram for the number of heads converges to a regular curve - the same normal curve we've seen before.

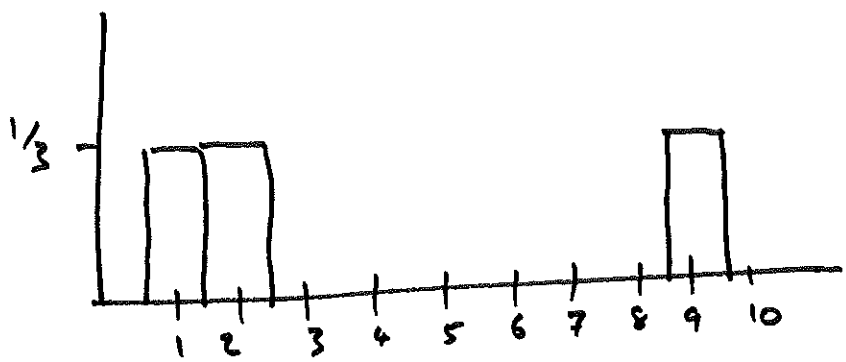

25 Draws
20 40 60 80 100 120 140 160 180
50 Draws
120 140 160 180 200 220 240 260 280
100 Draws
250 300 350 400 450 500 550
probability density

How general is this behaviour?

~~Does the sum of the~~

Does the distribution for the sum of the draws always converge to the normal curve?

Does it depend on what's in the box?



probability histogram

This looks nothing like a normal curve.

<u>However</u> - the distribution of the sum of draws from this box still converges to the normal curve, as the number of draws gets larger.

How ~~large~~ large the number of draws needs to be depends on the content of the box.

Central Limit Theorem.

In general, the probability histogram of a sum of draws can be approximated by the normal curve.

CLT: when drawing at random, with replacement from a box model, the probability histogram for the sum will follow the normal curve in the limit (large # of draws), even if the contents of the box do not follow the normal curve.

[the histogram must be put into standard units, and the # of draws must be "large"]

Example.

$\boxed{1}\ \boxed{3}\ \boxed{5}\ \boxed{7}$

400 draws

what's the chance that

the sum > 1500

Distribution of the Sum of draws

follows the normal curve.

Average of box $\dfrac{1+3+5+7}{4} = 4$

SD of box $\sqrt{\dfrac{(1-4)^2 + (3-4)^2 + (5-4)^2 + (7-4)}{4}}$ $\sqrt{\dfrac{9+1+1+9}{4}}$

$= \sqrt{5}$

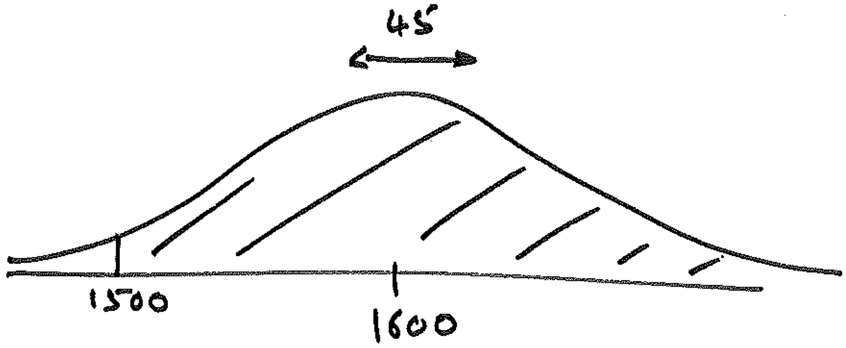
Expected value: # draws × ave of box

$400 \times 4 = 1600$

$S.E. = \sqrt{\#\text{draws}} \times SD_{box}$

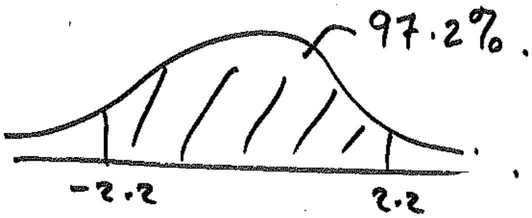
$= \sqrt{400} \times \sqrt{5}$

$\approx 45.$



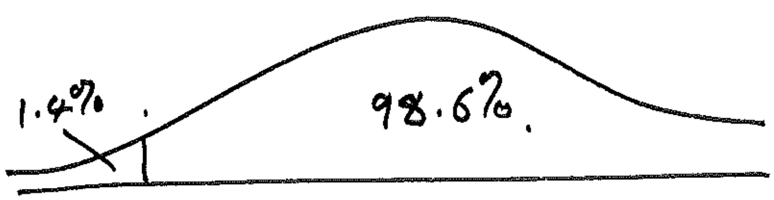
In standard units

$$1500 \longrightarrow \frac{1500 - 1600}{45} = -2.2.$$



2.8% in the two tails taken together

1.4% in each tail.



$$P(\text{sum of 400 draws} > 1500) \doteq \underline{\underline{98.6\%}}.$$