Unit 7: Chance Variation

The (real) Law of Averages

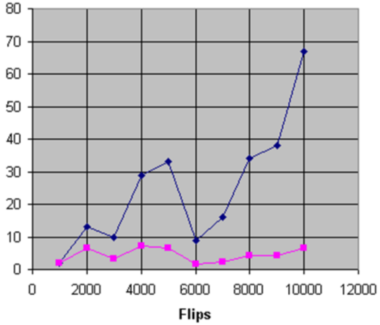
If we repeat an experiment more and more times, the fraction of times an event occurs will get closer to the probability of that event, but the difference

$$\text{(number of times event occurs)} - \text{(probability)(number of trials)}$$

is likely to go up.

Kerrich's coin-toss

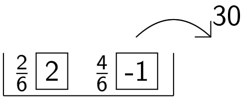
#tosses	#heads - (#tosses/2)	%x10
1000	2	2
2000	13	6.5
3000	10	3.33
4000	29	7.25
5000	33	6.67
6000	9	1.5
7000	16	2.29
8000	34	4.25
9000	38	4.22
10000	67	6.7



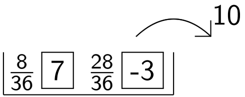
Repeat with a spreadsheet

Modeling games with boxes of tickets

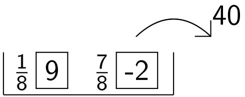
1. Roll one die. If you get a 5 or a 6, you win \$2; otherwise you lose \$1. You roll 30 times.



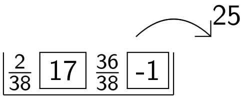
2. Roll two dice. If you get a 7 or 11, you win \$7; otherwise you lose \$3. You roll 10 times.



3. Toss 3 coins. If you get 3 heads, you win \$9; otherwise you lose \$2. You toss 40 times.



4. Roulette, playing "splits" (2 numbers of 38 are winners). It pays 17 to 1, and you are betting \$1 each time. You play 25 times.



EV and SE — Two of each!

Suppose we know the box (model) of a game, and we make many draws from the box with replacement (play the game many times).

What should we expect to get for a sum of draws (Expected Value of sum) and for the average of the draws (EV of avg)?

And how much variability should we expect in the sum (Standard Error of sum) and in the average (SE of avg)?

Why the EV and SE are really an AV and SD

Given a box model, we could make a "superbox" of all the possible sets of n draws from the original box.

Then we could replace each ticket in the superbox by a ticket with its sum (so that some sums are probably more frequent than others).

The EV of the sum is the AV of this new superbox.

And the SE of the sum is the SD of this new superbox.

The EV and SE of the average draws are just the EV and SE, respectively, of the sum divided by n.

A simple, sorta general example

Say the original box has two-fifths a's and three-fifths b's, where a and b are some numbers: $.4\ \boxed{a}\quad .6\ \boxed{b}$

Three draws is like one draw from the superbox

$(.4)(.4)(.4)\ \boxed{a, a, a}\quad (.4)(.4)(.6)\ \boxed{a, a, b}\quad (.4)(.6)(.4)\ \boxed{a, b, a}$

$\bullet\ \bullet\ \bullet\quad (.6)(.6)(.6)\ \boxed{b, b, b}$

Taking the sum of the three draws is like one draw from the new superbox

$(.4)(.4)(.4)\ \boxed{a + a + a}\quad (.4)(.4)(.6)\ \boxed{a + a + b}$

$\bullet\ \bullet\ \bullet\quad (.6)(.6)(.6)\ \boxed{b + b + b}$

The example II

Because we don't care what order we add, there are only four kinds of tickets in the second superbox:

$\binom{3}{0}(.4)^3$ $\boxed{a+a+a}$	$\binom{3}{1}(.4)^2(.6)$ $\boxed{a+a+b}$
$\binom{3}{2}(.4)(.6)^2$ $\boxed{a+b+b}$	$\binom{3}{3}(.6)^3$ $\boxed{b+b+b}$

So the average of the second superbox, i.e., the EV of the sum of three draws from the original box, is

$$\begin{aligned} EV \text{ of sum} &= \binom{3}{0}(.4)^3(3a) + \binom{3}{1}(.4)^2(.6)(2a+b) \\ &\quad + \binom{3}{2}(.4)(.6)^2(a+2b) + \binom{3}{3}(.6)^3(3b) \end{aligned}$$

The example III

The coefficient of a in this mess is

$$\begin{aligned} &3(.4)\left[\binom{3}{0}(.4)^2 + \binom{3}{1}(.4)(.6)\frac{2}{3} + \binom{3}{2}(.6)^2\frac{1}{3}\right] \\ &= 3(.4)\left[\binom{2}{0}(.4)^2 + \binom{2}{1}(.4)(.6) + \binom{2}{2}(.6)^2\right] \\ &= 3(.4)[.4 + .6]^2 = 3(.4) \end{aligned}$$

Similarly, the coefficient of b is 3(.6), so

$$EV \text{ of sum} = 3(.4a + .6b) = 3(AV \text{ of original box})$$

The example IV

Similarly, the SD of the second superbox, i.e., the SE of the sum of three draws from the original box, is

$$SE \text{ of sum} = \sqrt{3}(SD \text{ of original box})$$

The example IV

Similarly, the SD of the second superbox, i.e., the SE of the sum of three draws from the original box, is

$$SE \text{ of sum} = \sqrt{3}(SD \text{ of original box})$$

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	One die				Three dice			Sums			Avgs		
2	1	Avg	3.5		1	1	1	3	Avg	10.5	1	Avg	3.5
3	2	SD	1.71		2	1	1	4	SD	2.96	1.33	SD	0.99
4	3				3	1	1	5			1.67		
5	4				4	1	1	6			2		
6	5				5	1	1	7			2.33		
7	6				6	1	1	8			2.67		
8					1	2	1	4			1.33		
9					2	2	1	5			1.67		
10					3	2	1	6			2		
11					4	2	1	7			2.33		
12					5	2	1	8			2.67		
13					6	2	1	9			3		
14					1	3	1	5			1.67		
15					2	3	1	6			2		
16					3	3	1	7			2.33		
17					4	3	1	8			2.67		
18					5	3	1	9			3		
19					6	3	1	10			3.33		
20					1	4	1	6			2		
21					2	4	1	7			2.33		
22					3	4	1	8			2.67		
23					4	4	1	9			3		
24					5	4	1	10			3.33		
25					6	4	1	11			3.67		
26					1	5	1	7			2.33		
27					2	5	1	8			2.67		
28					3	5	1	9			3		
29					4	5	1	10			3.33		
30					5	5	1	11			3.67		
31					6	5	1	12			4		
32					1	6	1	8			2.67		
33					2	6	1	9			3		
34					3	6	1	10			3.33		

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	The box				Four draws				Counts			Pcts		
2	1	Avg	0.33		1	1	1	1	4	Avg	1.33	1	Avg	0.33
3	0	SD	0.47		0	1	1	1	3	SD	0.94	0.75	SD	0.24
4	0				0	1	1	1	3			0.75		
5					1	0	1	1	3			0.75		
6					0	0	1	1	2			0.5		
7					0	0	1	1	2			0.5		
8					1	0	1	1	3			0.75		
9					0	0	1	1	2			0.5		
10					0	0	1	1	2			0.5		
11					1	1	0	1	3			0.75		
12					0	1	0	1	2			0.5		
13					0	1	0	1	2			0.5		
14					1	0	0	1	2			0.5		
15					0	0	0	1	1			0.25		
16					0	0	0	1	1			0.25		
17					1	0	0	1	2			0.5		
18					0	0	0	1	1			0.25		
19					0	0	0	1	1			0.25		
20					1	1	0	1	3			0.75		
21					0	1	0	1	2			0.5		
22					0	1	0	1	2			0.5		
23					1	0	0	1	2			0.5		
24					0	0	0	1	1			0.25		
25					0	0	0	1	1			0.25		
26					1	0	0	1	2			0.5		
27					0	0	0	1	1			0.25		
28					0	0	0	1	1			0.25		
29					1	1	1	0	3			0.75		
30					0	1	1	0	2			0.5		
31					0	1	1	0	2			0.5		
32					1	0	1	0	2			0.5		
33					0	0	1	0	1			0.25		
34					0	0	1	0	1			0.25		

Shortcut formula for SD of a 2-value box

$$\text{SD} = (\text{larger} - \text{smaller}) \cdot \sqrt{(\text{fraction with larger})(\text{fraction with smaller})}$$

Shortcut formula for SD of a 2-value box

$$\text{SD} = (\text{larger} - \text{smaller}) \cdot \sqrt{(\text{fraction with larger})(\text{fraction with smaller})}$$

Shortcut formula for SD of a 2-value box

$$\text{SD} = (\text{larger} - \text{smaller}) \cdot \sqrt{(\text{fraction with larger})(\text{fraction with smaller})}$$

Ex: Roll two dice. If you get a 7 or 11, you win \$7; otherwise you lose \$3.

$$\text{Avg} = (8/36)7 + (28/36)(-3) = -7/9$$

$$\text{SD} = (7-(-3))\sqrt{(8/36)(28/36)} = (10/9)\sqrt{14} \approx 4.16$$

$\frac{8}{36}$ [7] $\frac{28}{36}$ [-3]

In general:

$$n = \text{number of draws (plays)}$$

$$\text{EV for sum} = (\text{AV of box}) \cdot n \qquad SE \text{ for sum} = (SD \text{ of box}) \cdot \sqrt{n}$$

$$\text{EV for avg} = \text{AV of box} \qquad SE \text{ for avg} = (SD \text{ of box})/\sqrt{n}$$

SE for sum gets larger as n gets larger, but only as $\sqrt{n}$.

So SE for avg actually gets smaller as n gets larger.

I don't know why the authors decided to put the sum and avg formulas into different chapters.

Also, they always compute the SE for avg as $(SE \text{ for sum})/n$ — but it is the same as the one above. Honest!

Example of EV and SE

Roll one die: If you get a 5 or 6, you win \$2; otherwise you lose \$1. You play 30 times.

$$\text{AV of game (box)} = \frac{2}{6}(2) + \frac{4}{6}(-1) = 0$$

$$\text{SD of game (box)} = (2 - (-1))/\sqrt{\frac{2}{6} \cdot \frac{4}{6}} = \sqrt{2}$$

$$\text{EV of sum of 30 plays} = 30(0) = 0$$

$$\text{EV of avg of 30 plays} = 0$$

$$\text{SE of sum of 30 plays} = \sqrt{2}\sqrt{30} = 2\sqrt{15}$$

$$\text{SE of avg of 30 plays} = \sqrt{2}/\sqrt{30} = 1/\sqrt{15}$$

Another example

Roll 2 dice. If you get a 7, you win \$4; if 11, win \$8; otherwise, lose \$2. Play 60 times.

$$
\begin{aligned}
\text{AV} &= \frac{6}{36}(4) + \frac{2}{36}(8) + \frac{28}{36}(-2) = -\frac{4}{9} \\
\text{SD} &= \sqrt{\frac{6}{36}(4-(-\frac{4}{9}))^2 + \frac{2}{36}(8-(-\frac{4}{9}))^2 + \frac{28}{36}(-2-(-\frac{4}{9}))^2} \\
&= \frac{2}{9}\sqrt{185} \approx 3.0
\end{aligned}
$$

Not the shortcut formula; it's a 3-value box.

$$
\begin{aligned}
\text{EV of sum} &= -\frac{4}{9}(60) \approx -26.7 \\
\text{SE of sum} &= \frac{2}{9}\sqrt{185}\sqrt{60} \approx 23.4
\end{aligned}
$$

How EV's and SE's change with n

Roll two dice. If you get a 7 or 11, you win \$7; otherwise you lose \$3.

$$\text{AV} = \frac{8}{36}(7) + \frac{28}{36}(-3) = -\frac{7}{9}$$

$$\text{SD} = (7 - (-3))\sqrt{\frac{8}{36} \cdot \frac{28}{36}} \approx 4.2$$

n	EV of sum	EV of avg	SE of sum	SE of avg
90	-70	$-7/9$	≈ 39	$\approx .44$
900	-700	$-7/9$	≈ 124	$\approx .14$
9000	-7000	$-7/9$	≈ 394	$\approx .044$
90000	-70000	$-7/9$	≈ 1240	$\approx .014$

Normal approx with EV and SE in place of AV and SD (I)

Again: Roll two dice. If you get a 7 or 11, you win \$7; otherwise you lose \$3. Play 90 times. What is the probability of winning at least \$5?

Again, EV of sum $= -70$, SE of sum ≈ 39.

So 5 in std units is $z = (5 - (-70))/39 \approx 1.9$.

$P(\text{win} \geq 5) = P(z \geq 1.9) = \frac{100-94}{2}\% = 3\%$

Normal table (Area between $-z$ and z)

z	Area(%)	z	Area(%)	z	Area(%)	z	Area(%)
0.0	0.0	1.15	74.99	2.3	97.86	3.45	99.944
0.05	3.99	1.2	76.99	2.35	98.12	3.5	99.953
0.1	7.97	1.25	78.87	2.4	98.36	3.55	99.961
0.15	11.92	1.3	80.64	2.45	98.57	3.6	99.968
0.2	15.85	1.35	82.3	2.5	98.76	3.65	99.974
0.25	19.74	1.4	83.85	2.55	98.92	3.7	99.978
0.3	23.58	1.45	85.29	2.6	99.07	3.75	99.982
0.35	27.37	1.5	86.64	2.65	99.2	3.8	99.986
0.4	31.08	1.55	87.89	2.7	99.31	3.85	99.988
0.45	34.73	1.6	89.04	2.75	99.4	3.9	99.99
0.5	38.29	1.65	90.11	2.8	99.49	3.95	99.992
0.55	41.77	1.7	91.09	2.85	99.56	4	99.9937
0.6	45.15	1.75	91.99	2.9	99.63	4.05	99.9949
0.65	48.43	1.8	92.81	2.95	99.68	4.1	99.9959
0.7	51.61	1.85	93.57	3	99.73	4.15	99.9967
0.75	54.67	1.9	94.26	3.05	99.771	4.2	99.9973
0.8	57.63	1.95	94.88	3.1	99.806	4.25	99.9979
0.85	60.47	2	95.45	3.15	99.837	4.3	99.9983
0.9	63.19	2.05	95.96	3.2	99.863	4.35	99.9986
0.95	65.79	2.1	96.43	3.25	99.885	4.4	99.9989
1	68.27	2.15	96.84	3.3	99.903	4.45	99.9991
1.05	70.63	2.2	97.22	3.35	99.919		
1.1	72.87	2.25	97.56	3.4	99.933		

Normal approx with EV and SE in place of AV and SD (II)

Roulette, playing splits (pay 17 to 1), play 100 times, betting same amount. Probability of breaking even (win ≥ 0)?

$$\text{EV of sum} = \left[\frac{2}{38}(17) + \frac{36}{38}(-1)\right](100) = -\frac{100}{19} \approx -5.3$$

$$\text{SE of sum} = (17 - (-1))\sqrt{\frac{2}{38} \cdot \frac{36}{38}}\sqrt{100} \approx 40.2$$

0 in std units is $z = (0 - (-5.3))/40.2 \approx .13$

$P(\text{win} \geq 0) = P(z \geq .13) = \frac{100-12}{2}\% = 44\%$ (not too bad!)

For counts or fractions of success, . . .

"success" means whatever you are counting happened

the box has values 1 (success) and 0 (failure),

the formulas for EV and SE of sum give the EV and SE of the success count,

and the formulas for EV and SE of avg give the EV and SE of fraction or % of successes

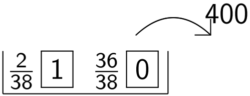
Example: How many wins? What %?

If a gambler plays splits in roulette 400 times, how many times should he expect to win, give or take how many?

$$\text{EV} = \left[\frac{2}{38}(1) + \frac{36}{38}(0)\right](400) \approx 21$$

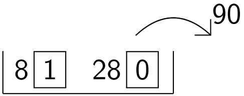
$$\text{SE} = (1-0)\sqrt{\frac{2}{38}\cdot\frac{36}{38}}\sqrt{400} \approx 4.5$$

What fraction (or %) of wins, give or take? (Divide each by 400.)



Normal approx with EV and SE in place of AV and SD (III)

Rolling two dice, probability of 7 or 11 is $\frac{8}{36}$. In 90 rolls, what is the probability of the fraction of 7's and 11's being between $\frac{7}{36}$ and $\frac{9}{36}$?



EV of fraction $= \frac{8}{36}$.

SE of fraction $= \sqrt{\frac{8}{36} \cdot \frac{28}{36}} / \sqrt{90} \approx .0438$

$\frac{7}{36}$, $\frac{9}{36}$ in std units: $z = \pm(\frac{1}{36})/.0438 \approx \pm.63$

$P(\frac{7}{36} \leq \text{fraction} \leq \frac{9}{36}) = P(-.63 \leq z \leq .63) \approx 48\%$

SE's of count and % $\implies$ Law of Averages

If we repeat an experiment more and more times,
i.e., as n gets larger,

the fraction of times an event occurs will be closer to the probability of that event,
i.e., SE for % gets smaller,

but the difference (number of times event occurs) $-$ (probability)(number of trials) is likely to go up,
i.e., SE for count gets larger.

Ex: Law of Averages

Roll two dice. Count how often you get a 7 or 11.

$$\text{AV} = \frac{8}{36}(1) + \frac{28}{36}(0) = 2/9$$

$$\text{SD} = (1-0)\sqrt{\frac{2}{9}\cdot\frac{7}{9}} \approx .42$$

n	EV of sum	EV of avg	SE of sum	SE of avg
90	20	2/9	≈ 3.9	$\approx .044$
900	200	2/9	≈ 12.5	$\approx .014$
9000	2000	2/9	≈ 39.4	$\approx .004$
90000	20000	2/9	≈ 124.7	$\approx .0014$

The usefulness of the SE

With which game is one more likely to make $\geq \$5$ in 100 rounds?

(a) Flip a coin: Win \$1 if H, lose \$1 if T.

(b) Flip 2 coins: Win \$(#H - #T).

(a) EV $= 0$, SE $= (1-(-1))\sqrt{(.5)(.5)}\sqrt{100} = \$10$

(b) EV $= 0$, SE $= \sqrt{\frac{(-2)^2+2(0)^2+(2)^2}{4}}\sqrt{100} = \$10\sqrt{2}$

So in (a), \$5 is $z = \frac{5-0}{10} = .5$; in (b) \$5 is $z = \frac{5-0}{10\sqrt{2}} \approx .35$. So winning $\geq \$5$ is more likely in (b).