

Slumpen och hur vi uppfattar den – med och utan tärning

Ingemar Holgersson är universitetslektor och lärarutbildare i matematik vid Högskolan Kristianstad.

En av de märkligaste upptäckterna när det gäller slumpen är att om man upprepar ett försök tillräckligt många gånger, så närmar sig sannolikhetsfördelningen samma fördelning, den s.k. normalfördelningen. Singlar man slant måste man upprepa försöket minst 20 gånger för att komma nära. Kastar man tärning behövs det färre. Men hur få? För att illustrera detta fenomen presenteras ett nytt schema för att bestämma antalet kombinationer som ger samma poäng vid olika antal kast. Det är en generalisering av Pascals triangel. I schemat börjar man med 6 ettor istället för två, och summerar över sex intilliggande celler istället för två, för att generera nästa kolumn. I tabellen nedan visas detta schema för resultatet av upp till 6 kast med en tärning.

Tabell 1. Ett schema för att bestämma antalet kombinationer vid upprepade kast med en tärning.

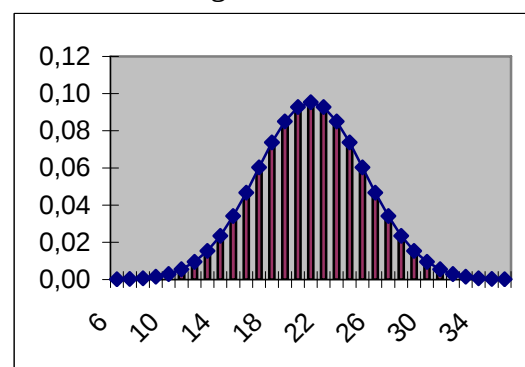
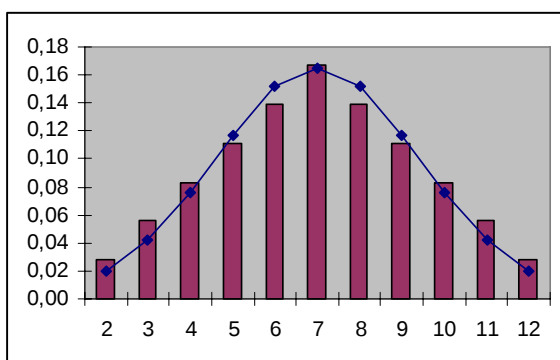
	A	B	C	D	E	F
1						
2						
3						
4						1
5						6
6					1	21
7					5	56
8				1	15	126
9				4	35	252
10			1	10	70	456
11			3	20	126	756
12		1	6	35	205	1161
13		2	10	56	305	1666
14	1	3	15	80	420	2247
15	1	4	21	104	540	2856
16	1	5	25	125	651	3431
17	1	6	27	140	735	3906
18	1	5	27	146	780	4221
19	1	4	25	140	780	4332
20		3	21	125	735	4221
21		2	15	104	651	3906
22		1	10	80	540	3431
23			6	56	420	2856
24			3	35	305	2247
25			1	20	205	1666
26				10	126	1161
27				4	70	756
28				1	35	456
29					15	252
30					5	126
31					1	56
32						21
33						6

34						
35						
36						
37	6	36	216	1296	7776	46656

1

Här är 6:an i cell B17 summan av talen i cellerna A14 till A19, och 5:an i cellen nedanför dvs B18, är summan av talen i A15 till A20, osv. Samma princip gäller för hela schemat så att talen i kolumn C är resultatet av motsvarande summor i kolumn B. Talen i kolumn D är summor av sex tal i kolumn C osv. I rad 37 ges summorna för varje kolumn, och de visar sig vara växande potenser av 6.

I figurerna nedan visas sannolikhetsfördelningarna för två kast resp sex kast med en tärning. Motsvarande normalfördelning, dvs med samma väntevärde och varians, ges också. Redan vid sex kast är alltså fördelningen väldigt nära en normalfördelning.



Det nya schemat kan lätt generaliseras till vilken likformig diskret fördelning som helst. För en tiosidig tärning t ex börjar man med tio ettor och summerar över tio celler etc. Men det kan faktiskt också generaliseras till att gälla för alla andra diskreta fördelningar också. Som ett exempel studerar jag kast med två tärningar där (den absoluta) skillnaden registreras.

När man kastar två tärningar blir skillnaden mellan ögonen inte likformigt fördelad. Antalet möjligheter att få skillnaden noll på är 6. Antalet möjligheter att få skillnaden ett är 10, att få två är 8, att få tre är 6, att få fyra är 4 och att få fem kan ske på 2 sätt. Eftersom alla dessa möjligheter är delbara med två kan de ges vikterna 3, 5, 4, 3, 2 och 1. I tabell 2 ges ett schema där man tar hänsyn till dessa vikter.

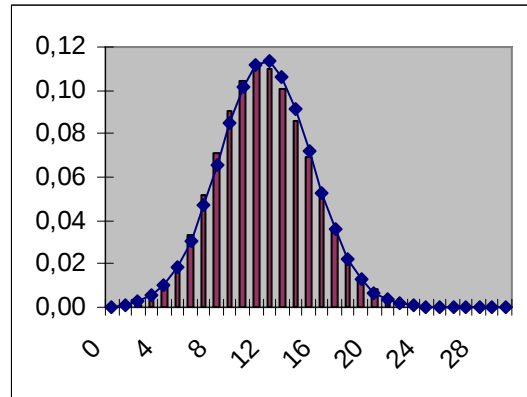
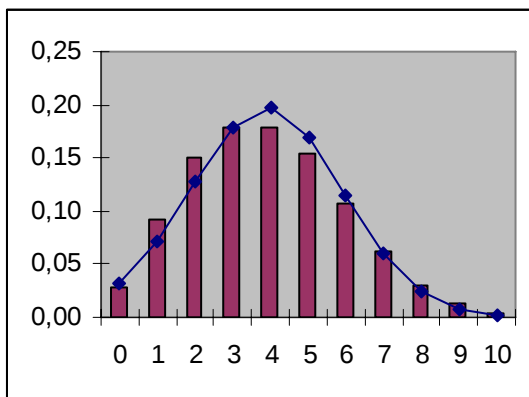
I tabell 2 ges talen i kolumn B genom att låta vikterna i A2 till A7 "passera genom" vikterna för ett kast i A13 till A18. Så är 9 in B11 resultatet av $3 \cdot 3$, vikten i B12 kommer från $3 \cdot 5 + 5 \cdot 3$ osv. På samma sätt ges vikterna i kolumn C genom att låta vikterna i A2 till A7 "passera genom" kolumn B, så att till exempel vikterna i C12 är lika med $3 \cdot 58 + 5 \cdot 49 + 4 \cdot 30 + 3 \cdot 9$. I rad 33 ges summan för ett, två, tre osv experiment. De är nu växande potenser av 18. Å andra sedan är talen i översta cellen på varje kolumn växande potenser av 3.

I figurerna under tabellen ges sannolikhetsfördelningarna för två respektive sex upprepningar av experimentet och motsvarande, dvs med samma väntevärde och varians, normalfördelningar. Eftersom grundfördelningen är sned är även övriga fördelningar sneda. Men eftersom normalfördelningen är symmetrisk blir avvikelser större i detta fall än för kast med en tärning.

Med det nya schemat har vi fått ett väldigt flexibelt instrument för att studera upprepade försök med vilken diskret fördelning som helst. Detta kan ske genom att ändra vikterna och summeringen beroende på hur många vikterna är.

Tabell 2. Ett schema för att bestämma vikterna för sannolikhetsfördelningarna vid kast med två tärningar där skillnaden registreras.

	A	B	C	D	E
1	Vikter				
2	1				
3	2				
4	3				
5	4				243
6	5				2025
7	3			81	8370
8				540	23265
9			27	1782	49605
10			135	3984	87260
11		9	333	6925	132110
12		30	566	10064	176435
13	3	49	768	12678	211100
14	5	58	888	14088	228785
15	4	58	889	13974	226394
16	3	50	774	12492	205630
17	2	35	594	10118	171985
18	1	20	405	7432	132715
19		10	243	4947	94560
20		4	126	2980	62163
21		1	56	1616	37620
22			21	780	20885
23			6	330	10580
24			1	120	4850
25				36	1987
26				8	715
27				1	220
28					55
29					10
30					1
31					
32					
33	18	324	5832	104976	1889568



I andra delen av föredraget diskuteras ett par exempel där vår intuitiva uppfattning av slumpen spelar oss en del spratt.

Det första exemplet är Benfords lag. Den säger att ifall ett datamaterial har en någorlunda slumpmässig bakgrund så blir fördelningen av den första siffran i data inte en jämn fördelning. Istället börjar data med en etta i ca 30 % av fallen, och i fallande andelar ned till ca 4,5 % för en nia. Fördelningen visar sig nämligen vara logaritmisk och ges av $\log(2)$, $\log(3/2)$ etc ned till $\log(10/9)$ för att ha en nia först. För att förstå varför det blir så här kan man t ex tänka sig att man anger världens längsta floder i antalet km eller i antalet miles. Det borde väl inte påverka fördelningen av första siffran. Men om det är så, dvs skalinvariant så kan man visa att det medför att fördelningen i så fall måste vara logaritmisk.

I föredraget ges några exempel på data som illustrerar Benfords lag. Intressant nog följer även en rent matematisk följd som Fibonacci-följden Benfords fördelning.

Benfords lag utnyttjas för att avslöja ifall ett större datamaterial, som t ex lämnas in till skattemyndigheterna, är fabricerat eller äkta. Om vi fabricerar fördelar vi begynnelsesiffrorna betydligt mer jämnt än vad Benfords lag anger.

Ett annat exempel på när vår intuitiva uppfattning spelar oss ett spratt är följande: Låt dina elever få kasta tärning och notera om resultatet blir udda eller jämnt. Men innan de börjar så låt dem förutsäga hur en sådan serie om 50 kast kommer att se ut. Låt dem sedan utföra experimentet och därefter jämföra resultaten. Den skillnad som står ut är att gissningarna skiftar mellan jämnt och udda betydligt oftare än vad slumpen gör. Det är få av eleverna som tror att det ska bli t ex udda mer än fem gånger i rad. I verkligheten är det inte helt ovanligt att det kan bli sju, åtta eller nio gånger i rad.