Calculating Probability

Transcript from Grimmett and Welsh

Exercises The most elementary problems in probability theory are those which involve experiments such as the shuffling of cards or the throwing of dice, and these usually give rise to situations in which every possible outcome is equally likely to occur. This is the case of Example 17 above. Such problems usually reduce to the problem of counting the number of ways in which some event may occur, and the following exercises are of this type.

11. Show that if a coin is tossed n times, then there are exactly

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

sequences of possible outcomes in which exactly k heads are obtained. If the coin is fair (so that heads and tails are equally likely on each toss), show that the probability of getting at least k heads is

$$\frac{1}{2^n}\sum_{r=k}^{n}\binom{n}{r}$$

12. We distribute r distinguishable balls into n cells at random, multiple occupancy being permitted. Show that

(i) there are n^r possible arrangements,

(ii) there are $\binom{r}{k}(n-1)^{r-k}$ arrangements in which the first cell contains exactly k

balls,

(iii) the probability that the first cell contains exactly k balls is

$$\binom{r}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{r-k}$$

The beginning of the following derivation comes from "Probability. An Introduction" Grimmett, G Welch, D. Oxford University Press 2003. page 10 exercise 12.

Exercise 12 part iii:

Show that the probability that the first cell contains exactly k balls is

$$\binom{r}{k}\left(\frac{1}{n}\right)^{k}\left(1-\frac{1}{n}\right)^{r-k}$$

<u>Answer</u>

Probability = number of possible arrangement/total number of possible arrangement

= part(ii)/part(i)

$$=\binom{r}{k}\frac{(n-1)^{r-k}}{n^r}$$

$$=\binom{r}{k}\left[n\left(1-\frac{1}{n}\right)\right]^{r-k}\frac{1}{n^r}$$

$$=\binom{r}{k}\frac{n^{r-k}}{n^r}\left(1-\frac{1}{n}\right)^{r-k}$$

$$=\binom{r}{k}\left(\frac{1}{n}\right)^{k}\left(1-\frac{1}{n}\right)^{r-k}$$

So in the following we can say that balls are occurrences of a distinct sequence and cells are loop positions:

P(1st cell contains k where there are r balls and n cells) $=\binom{r}{k}\frac{(n-1)^{r-k}}{n^r}$

Here number of cells n=3

$$\therefore P=\binom{r}{k}\frac{2^{r-k}}{3^r}$$

Problem: There are n balls and 3 cells numbered [a,b,c].

P(there are a balls in cell) $=\binom{n}{a}\frac{2^{n-a}}{3^n}$,n=a+b+c

Now we have n-a balls and 2 cells

P(There are b balls in cell2)= $\binom{n-a}{b}\frac{1^{n-a-b}}{2^{n-a}}=\binom{n-a}{b}\frac{1}{2^{n-a}}$

Now we have n-a-b balls= c balls and 1 cell, so there is a probability of 1 that those c balls will go in that one cell. The probability of getting the ordered triplet [a+b+c] therefore, is the product of these three probabilities

$$\therefore P = \binom{n}{a}\frac{2^{n-9}}{3^n} \times \binom{n-a}{b}\frac{1}{2^{n-a}} \times 1$$

$$\Rightarrow P = \binom{n}{a}\binom{n-a}{b}\frac{1}{3^n}$$

$$\Rightarrow P = \frac{(n!)}{a!(n-a)!} \times \frac{(n-a)!}{b!(n-a-b)!} \times \frac{1}{3^n}$$

$$\therefore P = \frac{(a+b+c)!}{a!b!c!3^{a+b+c}}$$

$$N.B. \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

$$n - a - b = c$$

or $P = \dfrac{n!}{3^n\, a!b!c!}$ where n=a+b+c

Calculating the Probability in log space

$$P = \frac{(a+b+c)!}{a!b!c!3^{a+b+c}}$$

$$\log P = \log\left[\frac{(a+b+c)!}{a!b!c!3^{a+b+c}}\right] = \log[(a+b+c)!] - \log(a!) - \log(b!) - \log(c!) - \log[3^{a+b+c}]$$

$$\Rightarrow \log P = \log[(a+b+c)!] - \log(a!) - \log(b!) - \log(c!) - (a+b+c)\log 3$$

Aside:

$$n! = n \cdot n-1 \cdot n-2 \cdot ..... \cdot 1$$

$$\log n! = \log(n \cdot n-1 \cdot n-2 \cdot ... \cdot 1) = \log n + \log(n-1) + \log(n-2) + ... + \log 1$$

$$\Rightarrow \log(n!) = \sum_{k=1}^{n} \log k$$

So

$$\log P = \sum_{n=1}^{a+b+c} \log n - \sum_{n=1}^{a} \log n - \sum_{n=1}^{b} \log n - \sum_{n=1}^{c} \log n - (a+b+c)\log 3$$

$$\Rightarrow Score = -\log P = (a+b+c)\log 3 + \sum_{n=1}^{a} \log n + \sum_{n=1}^{b} \log n + \sum_{n=1}^{c} \log n - \sum_{n=1}^{a+b+c} \log n$$

Consensus sequence for quadruplex type CCTGTT

G	19	1043	2361	2361	2253				2361			250	1	1	1935	2311	2308	2039	1335	786	296	223	842	1295	1193	1677	1861	1820	1430	534	441	233	380	635	312	281	792	1904	2176	2117	1577	287	136	63	33	36	32	23	10	7
C					12	2361	2361					3			13		4	27	65	81	28	22	43	56	72	45	39	55	106	99	25		6	92	1079	113	47	41	7	3	2	12	7	1	5					
T					29			2361		2361	2361	6			89	1	13	57	107	265	389	191	556	249	674	191	58	90	105	161	76	3	27	61	197	1011	62	143	17	23		6	6	4	1	4				
A					67							14			120	3	15	127	215	206	31	23	71	179	95	180	218	178	220	39	74	12	32	87	189	507	1415	225	75	21	31	11	11	18	9	2	4			
-	2342	1318										2088	2360	2360	204	46	21	111	639	1023	1617	1902	849	582	327	268	185	218	500	1528	1745	2113	1916	1486	584	449	45	48	86	197	751	2045	2201	2275	2313	2319	2325	2338	2351	2354
	-	-	G	G	G	C	C	T	G	T	T	-	-	-	G	G	G	G	G	-	-	-	-	G	G	G	G	G	G	-	-	-	-	-	C	T	A	G	G	G	G	-	-	-	-	-	-	-	-	-

Consensus sequence for quadruplex type CCTGTCA

G	443	1956	1956	1956				1956				1956	1956	1950	1508	985	149	372	433	331	1103	1432	1530	1673	1567	1002	640	163	253	399	1266	1847	1822	1439	613	68	37	17	6	4	2
C					1956	1956				1956					37	63	16	35	21	48	87	32	22	21	36	142	168	562	609	28	7	9		1	1	1					
T							1956		1956						78	170	855	207	49	561	307	121	53	51	61	63	83	224	252	799	21	27	6	1			1				
A											1956				167	209	21	66	36	54	80	153	112	110	119	120	124	179	631	712	642	28	9	18	12	4	3	1			
-	1513													6	166	529	915	1276	1417	962	379	218	239	101	173	629	941	828	211	18	20	45	119	497	1330	1883	1915	1938	1950	1952	1954
	-	G	G	G	C	C	T	G	T	C	A	G	G	G	G	G	-	-	-	-	G	G	G	G	G	G	-	-	C	T	G	G	G	G	-	-	-	-	-	-	-