Understanding the "Antoniak equation"

The equation I call the "Antoniak equation" (due to Antoniak, 1974) computes the probability that N samples from a Dirichlet process with concentration parameter α will draw M distinct atoms, $1 \leq M \leq N$. Most contemporary formulations have it as:

$$P(M=m \mid \alpha, N) = s(N,m)\,\alpha^m \frac{\Gamma(\alpha)}{\Gamma(\alpha+N)},$$

where $s(N, m)$ is the absolute value of the Stirling number of the first kind for N, m. By definition, $s(0, 0) = 1$, $s(0, m) = 0$ for $m > 0$, $s(N, m) = 0$ for $m > N$, and for remaining values:

$$s(N+1,m) = s(N,m-1) + N \cdot s(N,m).$$

A triangular representation of absolute values of Stirling numbers of the first kind follows:

N \ *m*	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0	1									
1	0	1								
2	0	1	1							
3	0	2	3	1						
4	0	6	11	6	1					
5	0	24	50	35	10	1				
6	0	120	274	225	85	15	1			
7	0	720	1764	1624	735	175	21	1		
8	0	5040	13068	13132	6769	1960	322	28	1	
9	0	40320	109584	118124	67284	22449	4536	546	36	1

Recall that conditioned on the prior N samples from the Dirichlet process, the probability of drawing a new atom is α / (α + N), while the probability of drawing an existing atom is N / (α + N). It is instructive now to enumerate the possible outcomes of drawing N samples from a Dirichlet process in terms of whether each sample draws a new atom. Let the binary string W indicate a particular sampling outcome such that $W(i) = 1$ if and only if sample i drew a new atom. The ratios mentioned just above allow us to compute the probability of any W, as in:

$$P(W=1,1,0,0,1) = \left(\frac{\alpha}{\alpha}\right)\left(\frac{\alpha}{\alpha+1}\right)\left(\frac{2}{\alpha+2}\right)\left(\frac{3}{\alpha+3}\right)\left(\frac{\alpha}{\alpha+4}\right). \quad (1)$$

After some algebra, this further factorization exists:

$$P(W=1,1,0,0,1) = \alpha^m \frac{\Gamma(\alpha)}{\Gamma(\alpha+N)}\ 1 \cdot 1 \cdot 2 \cdot 3 \cdot 1. \quad (2)$$

T. Stepleton (tss@ri.cmu.edu) • *18 April 2008*

Note that the product of the denominators in (1) will be the same for all *W* strings of length *N*. In (2), this product is factored out and expressed as the ratio of gamma functions. The equivalence may be more apparent when considering integer α values and replacing the gamma functions with factorials. Meanwhile, the α^m comes from factoring out the α values from terms corresponding to *W* values of 1. By elimination if nothing else, the remaining product **(which we'll call *Q* for short)** must be accounted for by the Stirling number of the first kind *s*(*N*, *m*). The diagram on the facing page, where each row shows all *W* strings of a particular length, shows how this works.

Each rounded rectangle corresponds to a different *W* string of length *m*. As binary strings, there are a power of two (specifically 2^{N-1}) of them in each row. Below the *W* strings in the rectangles are the corresponding *Q* products. Underlined 1s there were α values before we factored those out to get (2). Note that the first *Q* factor is always $\underline{1}$, since the first sample from a Dirichlet process will always draw a new atom.

Now note that the sum of *Q* products for each *N*, *m* (called SQ(*N*, *m*) from now on) is always *s*(*N*, *m*) in the diagram. Using the diagram as a base case, we can now sketch an inductive proof of the fact that SQ(*N*, *m*) = *s*(*N*, *m*) by observing that we can construct the *W* string for *N*, *m* in two ways: appending a 1 to *W* for *N*–1, *m*–1, or appending a 0 to *W* for *N*–1, *m*–1. Corresponding *Q* products are multiplied by $\underline{1}$ in the first case and *N*–1 in the second, *N*–1 being the number of samples already drawn. After summing the new *Q* products and a factorization step, we have:

$$\mathrm{SQ}(N,m) = \mathrm{SQ}(N-1,m-1) + (N-1)\mathrm{SQ}(N-1,m),$$

which, by the inductive hypothesis, is

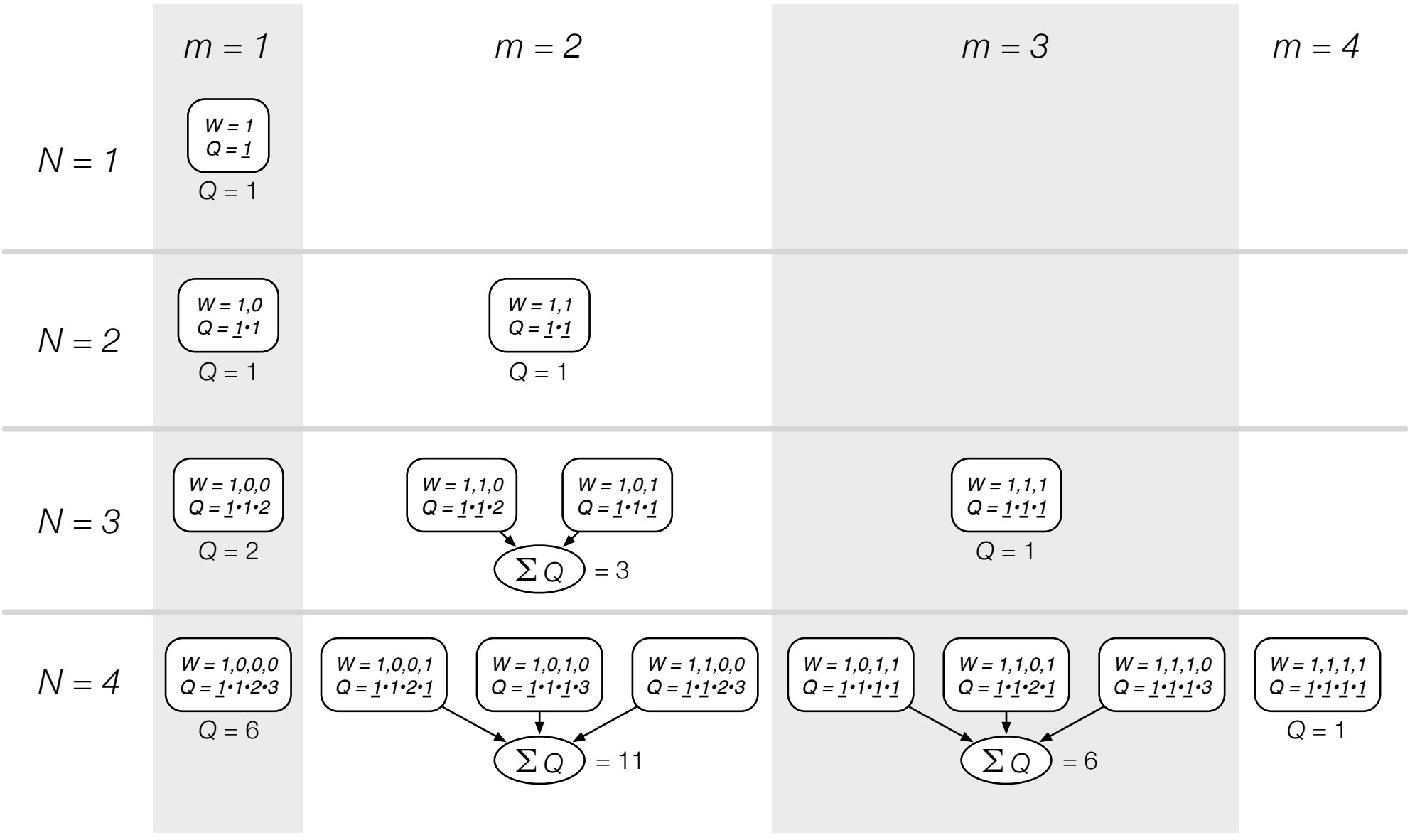
$$\mathrm{SQ}(N,m) = s(N-1,m-1) + (N-1)s(N-1,m),$$

which, by definition, is *s*(*N*, *m*).

References

Antoniak, C. (1974), “Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems,” *Annals of Statistics*, 2(6), pp. 1152-1174

Wikipedia contributors. (18 April 2008), “Stirling numbers of the first kind,” Wikipedia, the Free Encyclopedia,
http://en.wikipedia.org/w/index.php?title=Stirling_numbers_of_the_first_kind&oldid=205190063


m = 1
m = 2
m = 3
m = 4
N = 1
W = 1
Q = 1
Q = 1
N = 2
W = 1,0
Q = 1•1
Q = 1
W = 1,1
Q = 1•1
Q = 1
N = 3
W = 1,0,0
Q = 1•1•2
Q = 2
W = 1,1,0
Q = 1•1•2
W = 1,0,1
Q = 1•1•1
ΣQ
= 3
W = 1,1,1
Q = 1•1•1
Q = 1
N = 4
W = 1,0,0,0
Q = 1•1•2•3
Q = 6
W = 1,0,0,1
Q = 1•1•2•1
W = 1,0,1,0
Q = 1•1•1•3
W = 1,1,0,0
Q = 1•1•2•3
ΣQ
= 11
W = 1,0,1,1
Q = 1•1•1•1
W = 1,1,0,1
Q = 1•1•2•1
W = 1,1,1,0
Q = 1•1•1•3
ΣQ
= 6
W = 1,1,1,1
Q = 1•1•1•1
Q = 1