Discrete Second-Order Probability Distributions that Factor into Marginals

David Sundgren

University of Gävle

July 28, 2011

University of Gävle

Independence, sort of...

Let the first-order probability values be fractions $k_i/N, i = 1, \ldots, n, \sum_{i=1}^{n} k_i = N$. For simplicity we let N be understood and the second-order probability distribution be over the integers k_i.

Since $\sum_{i=1}^{n} k_i = N$, the variables cannot be independent, $p(k_1, \ldots, k_n) \neq p_1(k_1) \cdot \ldots \cdot p_n(k_n)$. Or, equivalently, $p_i(k_i|X) \neq p_i(k_i)$ when X is a set of other variables k_i.

But there could be a function f such that

$$p_i(k_i|X) = p_i(k_i) \frac{f(k_i + \sum_{k_j \in X} k_j, |X| + 1)}{f(\sum_{k_j \in X} k_j, |X|)} .$$

But then

$$p(k_1, \ldots, k_n) \propto p_1(k_1) \cdot \ldots \cdot p_n(k_n) .$$

Factoring into Marginals

We want

$$p(k_1, k_2, \ldots, k_n) = \frac{1}{K}\prod_{i=1}^{n} p_i(k_i)\,,$$

so for every $i, 1, \ldots, n$,

$$p_i(k_i) = \sum_{\Sigma_{j\neq i} k_j = N - k_i} \frac{1}{K}\prod_{i=1}^{n} p_i(k_i) = \frac{1}{K} p_i(k_i) \mathop{\ast}_{j\neq i} p_j\,(N - k_i)$$

where $*_{j\neq i}$ is the $n-1$-fold repeated convolution $p_1 * p_2 * \cdots * p_{i-1} * p_{i+1} * \cdots * p_n$ and $K = *_{i=1}^{n} p_i(N)$.
Thus

$$\mathop{\ast}_{j\neq i} p_j\,(N - k_i) = K\,, i = 1, \ldots, n\,.$$

The z-transform has the property that
$\mathcal{Z}\{p_1(k) * p_2(k)\} = \mathcal{Z}\{p_1(k)\}\mathcal{Z}\{p_2(k)\}$.

Time domain	Z domain
$p_2 * \ldots * p_n(N - k_1) = K$	$\mathcal{Z}\{p_2\} \ldots \mathcal{Z}\{p_n\} = \mathcal{Z}\{KH(c_1 - k_1)\}$
$p_1 * \ldots * p_n(N - k_2) = K$	$\mathcal{Z}\{p_1\} \ldots \mathcal{Z}\{p_n\} = \mathcal{Z}\{KH(c_2 - k_2)\}$
$\vdots$	$\vdots$
$p_1 * \ldots * p_{n-1}(N - k_n) = K$	$\mathcal{Z}\{p_1\} \ldots \mathcal{Z}\{p_{n-1}\} = \mathcal{Z}\{KH(c_n - k_n)\}$

So apart from the shifts from the Heaviside function $H(c_i - k_i)$ all marginals p_i are equal.

The Shift Property

If $p_i(k_i)$ can be written as a shifted function $q_i(k_i - a_i)$,

$$\mathcal{Z}\{p_i(k_i)\} = \mathcal{Z}\{q_i(k_i - a_i)\} = \mathcal{Z}\{q_i(k_i)\}z^{-a_i}$$

and

$$\prod_{j \neq i} \mathcal{Z}\{p_i\} = \frac{\prod_{j \neq i} \mathcal{Z}\{q_j\}}{z^{\sum_{j \neq i} a_j}} = \frac{Kz}{z-1} \frac{1}{z^{\sum_{j \neq i} a_j}}$$

Thus

$$\mathcal{Z}\{q_i\} = \left(\frac{Kz}{z-1}\right)^{\frac{1}{n-1}} = \frac{K^{\frac{1}{n-1}} \Gamma\left(k_i + \frac{1}{n-1}\right)}{k_i! \Gamma\left(\frac{1}{n-1}\right)}$$

The 2nd-order Distributions that Factor into Marginals

Finally, the joint distribution is

$$p(k_1,\dots,k_n) = \frac{(N-\sum_{i=1}^n a_i)!\prod_{i=1}^n \frac{\Gamma(k_i-a_i+\frac{1}{n-1})}{(k_i-a_i)!}}{(n-1)\Gamma\left(\frac{1}{n-1}\right)^{n-1}\Gamma\left(N+1-\sum_{i=1}^n a_i+\frac{1}{n-1}\right)}$$

and
the marginal distributions are

$$p_i(k_i) = \frac{(N-\sum_{j=1}^n a_j)!\Gamma\left(k_i-a_i+\frac{1}{n-1}\right)}{(n-1)\Gamma\left(N+1-\sum_{j=1}^n a_j+\frac{1}{n-1}\right)(k_i-a_i)!},$$

$i = 1,\dots,n.$

The multivariate Pólya distribution is obtained by drawing the underlying probabilities p_i from a Dirichlet distribution and integrating out $\mathbf{p} = (p_1, \ldots, p_n)$ from the multinomial distribution. In the same way, if we compound the Dirichlet distribution with parameters $1/(n-1)$ with the shifted multinomial distribution

$$\frac{(N - \sum_{i=1}^{n} a_i)! \prod_{i=1}^{n} p_i^{k_i - a_i}}{\prod_{i=1}^{n} (k_i - a_i)!}$$

we have

$$\int_{\mathbf{p}} \frac{1}{(n-1)^n \Gamma(n/(n-1))^{n-1} \prod_{i=1}^{n} p_i^{\frac{n-2}{n-1}}} \frac{(N - \sum_{i=1}^{n} a_i)! \prod_{i=1}^{n} p_i^{k_i - a_i}}{\prod_{i=1}^{n} (k_i - a_i)!} \, d\mathbf{p} =$$

$$\frac{(N - \sum_{i=1}^{n} a_i)!}{(n-1)\Gamma(1/(n-1))^{n-1}\Gamma(N+1+1/(n-1))} \prod_{i=1}^{n} \frac{\Gamma(k_i - a_i + 1/(n-1))}{(k_i - a_i)!}$$

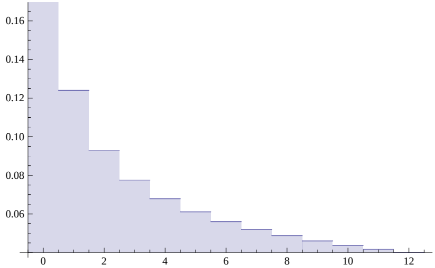
Let there be N balls with n different colours in an urn. After picking up a_i balls of type i I know that at least a_i of the N balls have this colour. If both the prior and the posterior factor into marginals the likelihood is a compound weighted hypergeometric distribution

$$\prod_{i=1}^{n} \frac{\binom{N-\sum_{j=1}^{n} a_j}{k_i - a_i}}{\binom{N}{k_i}} p_i^{-a_i},$$

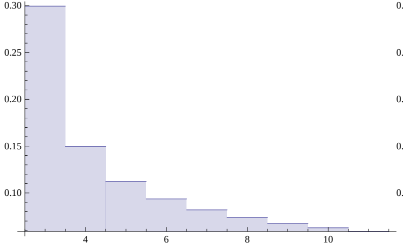
where $\mathbf{p}$ is drawn from the Dirichlet distribution with parameters $k_i + \frac{1}{n-1}$.

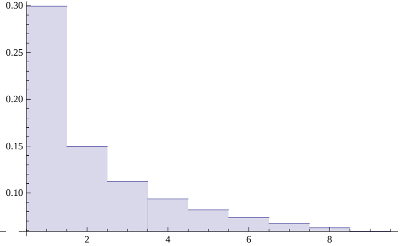
Example of Updating

Let $N = 12$ and $n = 3$. At first the plate is empty, that is $a_1 = a_2 = a_3 = 0$.
Then four balls are picked from the urn, three of which are of type 1 and one of type 2: $a_1 = 3, a_2 = 1, a_3 = 0$.



We see the posterior marginal distributions of k_1 and k_2 below, observe that $p_1(k) = p_2(k-2)$.





Given $k_1 = 6, k_2 = 4, k_3 = 2, \sum_{i=1}^{3} a_i = 4$ the likelihood function would have marginal probabilities such that

$$\Pr(a_1 = 0) = 0.157, \Pr(a_1 = 1) = 0.172, \Pr(a_1 = 2) = 0.191,$$
$$\Pr(a_1 = 3) = 0.218, \Pr(a_1 = 4) = 0.262$$

For a_2 we have

$$\Pr(a_2 = 0) = 0.111, \Pr(a_2 = 1) = 0.127, \Pr(a_2 = 2) = 0.152,$$
$$\Pr(a_2 = 3) = 0.202, \Pr(a_2 = 4) = 0.406$$

and