Ben Sauerwine
Progress Report March 3 2006

Consider my previous approximation for N >> P:

$$\frac{1}{N} Log\left[PathCount(\alpha, N, D, P)\right] \approx$$

$$\frac{P}{N}\log\binom{N}{\alpha N} + \frac{1}{N}\sum_{p=2}^{P}\log\left(1+\sum_{i=2}^{p}(-1)^{i-1}\frac{\binom{N}{\alpha N+(i-1)D}\binom{N}{\alpha N-(i-1)D}}{\binom{N}{\alpha N}\binom{N}{\alpha N}}\right)$$

Last week, I determined that an acceptable approximation for this approximation is:

$$\frac{1}{N} Log\left[PathCount(\alpha, N, D=1, P)\right] \approx$$

$$\frac{P}{N}\left[Log(\Gamma(N)) + Log\left({}_3\tilde{F}_2(1,1-(1-\alpha)N,1-\alpha N;(1-\alpha)N,\alpha N,-1)\right)\right]$$

which showed good convergence when N > P (see GVLongterm7), in the case where D = 1. It's not clear how letting D > 1 would affect this, except that it would require a higher order F function. To see this, recall that I can write

$$\Gamma(Di+\alpha N) = D^{Di+\alpha N-\frac{1}{2}}(2\pi)^{\frac{1-D}{2}}\prod_{k=0}^{D-1}\Gamma\left(i+\frac{\alpha N+k}{D}\right)$$, where D must be a positive integer.

Taking an expression from last week, I wanted a generating function such that

$$\left[\left(\frac{\partial}{\partial x}\right)^{i} F\right]\Bigg|_{x} = \frac{(-1)^i}{\Gamma(\alpha N+Di)\Gamma(\alpha N-Di)\Gamma((1-\alpha)N-Di)\Gamma((1-\alpha)N+Di)}$$

Attempting to rewrite this in terms of Gamma functions that generate the F series of functions using the identity above gives for two of the gamma functions (henceforth, let the $\hat{=}$ symbol represent a term equal to this desired generating function.

$$\hat{=}\frac{1}{\left(D^{2\alpha N-1}\right)(2\pi)^{1-D}}\frac{(-1)^i}{\left(D^{2D}\right)^i}\frac{1}{\Gamma(\alpha N-Di)\Gamma((1-\alpha)N-Di)\prod_{k=0}^{D-1}\Gamma\left(i+\frac{\alpha N+k}{D}\right)\Gamma\left(i+\frac{(1-\alpha)N+k}{D}\right)}$$

So far this looks like it might well be collapsible into something very nice, if I can just find a way to deal with the –Di terms. I recall that last week I dealt with them using:

$$\Gamma(z-n) = \frac{(-1)^n\Gamma(z)}{(1-z)_n} \quad n \in Integers$$

$$\frac{1}{\Gamma(\alpha N-Di)} = \frac{(-1)^{Di}(1-\alpha N)_{Di}}{\Gamma(\alpha N)} = \frac{(-1)^{Di}\Gamma(1-\alpha N+Di)}{\Gamma(\alpha N)\Gamma(1-\alpha N)}$$

$$= \frac{(2\pi)^{\frac{1-D}{2}}D^{\alpha N-\frac{1}{2}}}{\Gamma(\alpha N)\Gamma(1-\alpha N)}\left((-1)^D D^D\right)^i\prod_{k=0}^{D-1}\Gamma\left(i+\frac{(1-\alpha N)+k}{D}\right)$$

Now I can rewrite my generating function terms as

$$\hat{=}\frac{1}{\left(D^{2\alpha N-1}\right)(2\pi)^{1-D}}\frac{(-1)^i}{\left(D^{2D}\right)^i}\frac{1}{\Gamma(\alpha N-Di)\Gamma((1-\alpha)N-Di)\prod_{k=0}^{D-1}\Gamma\left(i+\frac{\alpha N+k}{D}\right)\Gamma\left(i+\frac{(1-\alpha)N+k}{D}\right)}$$

$$\frac{1}{\Gamma(\alpha N-Di)}=\frac{(2\pi)^{\frac{1-D}{2}}D^{\alpha N-\frac{1}{2}}}{\Gamma(\alpha N)\Gamma(1-\alpha N)}\left((-1)^D D^D\right)^i\prod_{k=0}^{D-1}\Gamma\left(i+\frac{(1-\alpha N)+k}{D}\right)$$

$$\frac{1}{\Gamma((1-\alpha)N-Di)}=\frac{(2\pi)^{\frac{1-D}{2}}D^{(1-\alpha)N-\frac{1}{2}}}{\Gamma((1-\alpha)N)\Gamma(1-(1-\alpha)N)}\left((-1)^D D^D\right)^i\prod_{k=0}^{D-1}\Gamma\left(i+\frac{(1-(1-\alpha)N)+k}{D}\right)$$

$$\hat{=}\left(\frac{(-1)^i}{\Gamma(\alpha N)\Gamma(1-\alpha N)\Gamma((1-\alpha)N)\Gamma(1-(1-\alpha)N)}\frac{\prod_{k=0}^{D-1}\Gamma\left(i+\frac{(1-\alpha N)+k}{D}\right)\Gamma\left(i+\frac{(1-(1-\alpha)N)+k}{D}\right)}{\prod_{k=0}^{D-1}\Gamma\left(i+\frac{\alpha N+k}{D}\right)\Gamma\left(i+\frac{(1-\alpha)N+k}{D}\right)}\right)$$

$$\hat{=}\left(\frac{(-1)^i\prod_{k=0}^{D-1}\Gamma\left(\frac{(1-\alpha N)+k}{D}\right)\Gamma\left(\frac{(1-(1-\alpha)N)+k}{D}\right)}{\Gamma(\alpha N)\Gamma(1-\alpha N)\Gamma((1-\alpha)N)\Gamma(1-(1-\alpha)N)}\times\frac{\prod_{k=0}^{D-1}\left(\frac{(1-\alpha N)+k}{D}\right)_i\left(\frac{(1-(1-\alpha)N)+k}{D}\right)_i}{\prod_{k=0}^{D-1}\Gamma\left(i+\frac{\alpha N+k}{D}\right)\Gamma\left(i+\frac{(1-\alpha)N+k}{D}\right)}\right)$$

Now collapsing this into the PFQ function using the same prescription as last week, I have

$$\hat{=}\left(\begin{array}{l}\frac{\prod_{k=0}^{D-1}\Gamma\left(\frac{(1-\alpha N)+k}{D}\right)\Gamma\left(\frac{(1-(1-\alpha)N)+k}{D}\right)}{\Gamma(\alpha N)\Gamma(1-\alpha N)\Gamma((1-\alpha)N)\Gamma(1-(1-\alpha)N)}\times\\ {}_{D+1}\tilde{F}_D\left(\{1\}\cup\bigcup_{k=0}^{D-1}\left\{\frac{(1-\alpha N)+k}{D},\frac{(1-(1-\alpha)N)+k}{D}\right\};\bigcup_{k=0}^{D-1}\left\{\frac{\alpha N+k}{D},\frac{(1-\alpha)N+k}{D}\right\};-1\right)\end{array}\right)$$

$$\hat{=}\frac{1}{\Gamma(\alpha N)\Gamma(1-\alpha N)}\mathrm{X}(D,\alpha,N)$$

which is plausible since it collapses to my expected result for D = 1. Now, through the new path-independence of this approximation I can collapse the overall result into: $\frac{1}{N}Log[PathCount(\alpha,N,D,P)]\approx\frac{P}{N}Log[\Gamma(N)\mathrm{X}(D,\alpha,N)]$, with the X function defined above as a function of the generating function chosen to eliminate a factor pulled out on the last progress report.

This week, I'd like to make explicit my assumptions that went into making this, and based on that try to find an approximation that works well at P = N or P < N.

<u>Suppose P>>N</u>

Recall that in the February 21 progress report, in my assumption while creating the determinant I assumed that no terms would ever truncate. In other words, every endpoint was reachable from every starting point. With P < N, only the terms corresponding to the edges will truncate, and so I will ignore this effect. Following the same logic from that progress report, I take the approximate determinant to be:

$$PathCount(\alpha, N, D, P) \approx \binom{N}{\alpha N}^{-P} \prod_{p=1}^{P} \left(\sum_{i=1}^{\min\left(\left\lfloor \frac{\alpha N}{D} \right\rfloor, \left\lfloor \frac{(1-\alpha)N}{D} \right\rfloor\right)} (-1)^{i-1} \binom{N}{\alpha N + (i-1)D} \binom{N}{\alpha N - (i-1)D} \right)$$

This should be a far easier expression to deal with than the N > P case! Let me simplify it a bit and then test it. Define the truncation distance:

$\mathrm{T}(\alpha, N, D) = \min\left(\left\lfloor \frac{\alpha N}{D} \right\rfloor, \left\lfloor \frac{(1-\alpha)N}{D} \right\rfloor\right)$. Now

$$PathCount(\alpha, N, D, P) \approx \binom{N}{\alpha N}^{-P} \left(\sum_{i=1}^{T(\alpha,N,D)} (-1)^{i-1} \binom{N}{\alpha N + (i-1)D} \binom{N}{\alpha N - (i-1)D} \right)^{P}$$

Now using results from last week, I have

$$\frac{1}{N} Log\left[PathCount(\alpha, N, D, P)\right]$$

$$= \frac{1}{N} \sum_{p=1}^{P} \log\left(\sum_{i=1}^{T(\alpha,N,D)} (-1)^{i-1} \binom{N}{\alpha N}^{-1} \binom{N}{\alpha N + (i-1)D} \binom{N}{\alpha N - (i-1)D} \right)$$

$$= \frac{P}{N} \log\left(\sum_{i=1}^{T(\alpha,N,D)} (-1)^{i-1} \binom{N}{\alpha N}^{-1} \binom{N}{\alpha N + (i-1)D} \binom{N}{\alpha N - (i-1)D} \right)$$

This is almost identical to the expression above, with one difference:

$$\Upsilon(D, \alpha, N) = \left(\begin{array}{l} \dfrac{\prod_{k=0}^{D-1} \Gamma\left(\frac{(1-\alpha N)+k}{D} \right) \Gamma\left(\frac{(1-(1-\alpha)N)+k}{D} \right)}{\Gamma((1-\alpha)N)\Gamma(1-(1-\alpha)N)} \times \\ {}_D\tilde{F}_D\left(\{1\} \cup \bigcup_{k=0}^{D-1} \left\{ \frac{(1-\alpha N)+k}{D}, \frac{(1-(1-\alpha)N)+k}{D} \right\}; \bigcup_{k=0}^{D-1} \left\{ \frac{\alpha N + k}{D}, \frac{(1-\alpha)N + k}{D} \right\}; -1 \right) \end{array} \right)$$

$$= \frac{P}{N} \log\left(\Gamma(N) \left[\Upsilon(D, \alpha, N) \right]_{First \quad T(\alpha,N,D) \quad Terms} \right)$$

It is important to note that in this case, the hypergeometric function will diverge at infinity, so it is critical that I only take the first legal terms. This is without a doubt easier to evaluate in the first form given, and so I will check it in that way.

Looking at GVLongterm8.nb, one sees that the percent error of cases with P>>N converges to a constant! The reason for this should be clear: recall that my approximation for the determinant experienced a slight error on each term, so it starts to play the role of a “slope”, or contributed entropy per path. In fact, the exact slope in the long should be related to the logarithm of the determinant of the matrix defined in the Feb

21 progress report under “Explicit Counting”, where the dimension of the matrix is determined by the maximum reachable endpoints: e.g., one needs only consider the relationship with the paths near enough that this path may interact, and then take the resulting determinant to the power of the number of paths.

Suppose $P = kN$

Let me consider again my approximate form for the determinant,

$$PathCount(\alpha, N, D, P) \approx \binom{N}{\alpha N}^{-P} \prod_{p=1}^{P} \left(\sum_{i=1}^{p} (-1)^{i-1} \binom{N}{\alpha N + (i-1)D} \binom{N}{\alpha N - (i-1)D} \right).$$

In this case, the P>>N form certainly cannot work very well since the number of terms in the determinant is still going to infinity. The N>>P form certainly can’t work well because there are a lot of truncations not being accounted for.

Let me try probabilistically weighting each contribution in the product. Namely, I consider the likelihood that the endpoint associated with this term will be reachable from an arbitrary path. Thus each probability will approach 1 as N becomes very large. Thus, the maximum number of endpoints encompassed by any particular cone from a starting point will be $\frac{N}{D} = \frac{kP}{D}$ (I think. Is there an alpha-dependence here? I don’t think so.). Let $k = 1, D = 1, \alpha = 0.5$ so that I may get a feel for the problem.

Now, paths at the center will be able to reach all endpoints. Paths at the border will be able to reach only half, and will decrease linearly going outwards. Then I can write that the probability of an endpoint of this extremity being reachable is $\frac{1}{P}$ for paths of the most extreme displacement, 1 for the least extreme, and increases linearly going outward.

$$PathCount(\alpha = 0.5, N = P(odd), D = 1, P(odd))$$
$$\approx \binom{N}{\alpha N}^{-P} \prod_{p=1}^{P} \left(\sum_{i=1}^{\frac{p-1}{2}} (-1)^{i-1} \frac{P-2i}{P} \binom{N}{\alpha N + (i-1)D} \binom{N}{\alpha N - (i-1)D} \right)$$

Let’s try this approximation in GVLongterm8, before any further simplification.