Fundamental Value Information

The market simulator uses a “fundamental” to give a common value to traded securities. In general, this fundamental could be a realization of any stochastic process, but we tend to use a mean reverting version that makes Gaussian jumps. In the following sections are some equations that describe useful properties of the fundamental process.

Mean Revering Gaussian Fundamental

The mean reverting Gaussian fundamental is a combination of two stochastic processes. The first decides when a “jump” happens. This is an independent Bernoulli draw with success rate ϕ at every time step. The second is a is a mean reverting Gaussian jump that happens on every Bernoulli success. To sample the mean reverting Gaussian after a jump, the old value is averaged with the mean μ, by proportion κ, where $\kappa = 0$ implies no mean reversion, and $\kappa = 1$ implies every jump is an independent draw from the mean. After adjustment a zero mean Gaussian is drawn with variance σ^2. Because actually sampling from these distributions at every time step would be prohibitively expensive ($O(n)$), we sample from the fundamental lazily whenever it is requested.

If we want to sample forward in time, the number of jumps that happen between t and $t + \delta$ is distributed by a Binomial with parameters δ and ϕ (the jump probability).

$$\text{Jumps after } \delta \sim \text{Binomial}(\delta, \phi)$$

If we want to sample the fundamental at time t between time $t - \delta$ and $t - \gamma$, where m jumps occurred in the $\delta + \gamma$ time frame, the number of jumps that happened before t is distributed by a Hypergeometric with population size $\delta + \gamma$, number of successes m, and δ draws.

$$\text{Jumps before } t \text{ between points} \sim \text{Hypergeometric}(\delta + \gamma, m, \delta)$$

We can formally write the mean reverting jump distribution of the fundamental in terms of f_j, where f_j represents the fundamental after j steps.

$$f_{j+1} \sim \mathcal{N}(\kappa\mu + (1 - \kappa)f_j, \sigma^2)$$

For brevity, it is simpler to use the compliment of the mean reversion instead of κ. We define $\lambda \equiv 1 - \kappa$. If we want to sample the fundamental forward in time after δ jumps this formula can be applied recursively to yield

$$f_{j+\gamma} \sim \mathcal{N}\left((1 - \lambda^{\gamma})\mu + \lambda^{\gamma} f_j, \frac{1 - \lambda^{2\gamma}}{1 - \lambda^2}\sigma^2\right)$$

Things get more complicated if we want to sample the fundamental between to times. First, we'll calculate the likelihood of observing the fundamental in the past given a future observation. In this case we recursively calculate this the same way we did the forward case and end up with

$$f_{j-\delta} \sim \mathcal{N}\left((1 - \lambda^{-\delta})\mu + \lambda^{-\delta} f_j, \frac{1 - \lambda^{-2\delta}}{\lambda^2 - 1}\sigma^2\right)$$

Next we find the joint distribution over f_j conditioned on $f_{j-\delta}$ and $f_{j+\gamma}$. We can use the fact that the product of two Gaussian PDFs (not random variables) is a new Gaussian with the following parameters

$$\mathcal{N}(\mu_1, \sigma_1^2)\mathcal{N}(\mu_2 . \sigma_2^2) = \mathcal{N}\left(\frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)$$

Using the previous three equations, we can combine them all into the posterior of the fundamental given that δ jumps happened before it, and γ jumps happened after it

$$\begin{aligned}
\mu_j &= \frac{(\lambda^{\delta} - 1)(\lambda^{\gamma} - 1)(\lambda^{\delta+\gamma} - 1)}{\lambda^{2\delta+2\gamma} - 1}\mu + \frac{\lambda^{\delta}(\lambda^{2\gamma} - 1)}{\lambda^{2\delta+2\gamma} - 1} f_{j-\delta} + \frac{\lambda^{\gamma}(\lambda^{2\delta} - 1)}{\lambda^{2\delta+2\gamma} - 1} f_{j+\gamma} \\
\sigma_j^2 &= \frac{(\lambda^{2\delta} - 1)(\lambda^{2\gamma} - 1)}{(\lambda^2 - 1)(\lambda^{2\delta+2\gamma} - 1)}\sigma^2 \\
f_j &\sim \mathcal{N}(\mu_j, \sigma_j^2)
\end{aligned}$$

Appendix

The Hypergeometric distribution is a somewhat expensive distribution to sample from. For repeated sampling from the same distribution, the standard inverse CMF method can be used, which only takes $O(\log n)$ time after an initial $O(n)$ computation. For the Hypergeometric distribution, the PMF of successive samples has a simple recurrence relation

$$p(X = k + 1) = \frac{(K - k)(n - k)}{(k + 1)(N - K - n + k + 1)} p(X = k)$$

where N is the population size, K is the number of successes in the population, and n is the sample size.

However, to use this relation, we need to have an initial value for $p(X = 0)$. This is generally expensive to compute accurately, so for our purposes we use Stirling's Approximation to speed up computation.

$$
\begin{aligned}
p(X = 0) &= \frac{\binom{N-K}{n}}{\binom{N}{n}} \\
&= \frac{(N-K)!(N-n)!}{(N-K-n)!N!} \\
&\approx \frac{(N-K)^{N-K+\frac{1}{2}}(N-n)^{N-n+\frac{1}{2}}}{(N-K-n)^{N-K-n+\frac{1}{2}}N^{N+\frac{1}{2}}} \\
\log(p(X = 0)) &\approx (N-K+\frac{1}{2})\log(N-K) + (N-n+\frac{1}{2})\log(N-n) \\
&\quad - (N-K-n+\frac{1}{2})\log(N-K-n) - (N+\frac{1}{2})\log N
\end{aligned}
$$