*Tristan Fletcher*

# Fisher Kernel for High Frequency Market Microstructure Prediction

### Model Specification

If we denote the volume at each depth $i$ of the order book ($i \in [1 \ldots D]$ assuming $D/2$ levels on the bid side and $D/2$ on the ask) at time $t$ as $V_t^i$, then we can use $\Delta V_i$ to represent the rate of change of volume at a depth over a given time interval $\tau$:

$$\Delta V_i = \frac{\left|V_{t+\tau}^i - V_t^i\right|}{\tau} \tag{1}$$

We can model $\Delta V_i$ using a Poisson process, i.e. $\Delta V_i \sim Poi(\lambda_i)$:

$$P(\Delta V_i = x) = \frac{e^{-\lambda_i \tau}(\lambda_i \tau)^x}{x!} \tag{2}$$

Using $f(x_i|\lambda_i)$ to represent the log likelihood of a rate $x_i$ observed at depth $i$ for a model parameterized by $\lambda_i$, the log likelihood can be expressed:

$$\begin{aligned} L(\lambda_i) &= \log f(x_i|\lambda_i) \\ &= \log\left(\frac{e^{-\lambda_i \tau}(\lambda_i \tau)^{x_i}}{x_i!}\right) \\ &= -\lambda_i + x_i \log(\lambda_i) - \log(x_i!) \end{aligned} \tag{3}$$

### Fisher Score for Poisson Process

It would useful to use observed rates to make predictions about future characteristics of the order book, for example whether the actual mid price would go up or down relative to its current value at some point in the near future. A plausible method for incorporating the information contained in the rates is to construct a Fisher Kernel based on the Poisson processes described above. The resultant mapping of the rates observed at each depth into a feature space can then be used in a classifier such as a Support or Relevance Vector Machine.

Using the model specified in (2) for each order depth, a Fisher Score vector can be calculated by taking the derivative of the log likelihood of the model with respect to each of the model's parameters.

Differentiating (3) wrt $\lambda_i$:

$$\frac{dL(\lambda_i)}{d\lambda_i} = -1 + \frac{x_i}{\lambda_i} \tag{4}$$

The parameter $\lambda_i$ needs to be estimated at each depth $i$. The maximum likelihood estimate of $\lambda_i$ can be calculated by first adjusting (3) to take into account a set of $N$ observations:

$$
\begin{aligned}
L(\lambda_i) &= \log \prod_{j=1}^{N} f(x_i^j|\lambda_i) \\
&= \sum_{j=1}^{N} \log \left( \frac{e^{-\lambda_i \tau}(\lambda_i \tau)^{x_i^j}}{x_i^j!} \right) \\
&= -N\lambda_i + \left( \sum_{j=1}^{N} x_i^j \right) \log(\lambda_i) - \sum_{j=1}^{N} \log(x_i^j!) \qquad (5)
\end{aligned}
$$

Differentiating (5) wrt $\lambda_i$ and setting this to zero yields our ML estimate of $\lambda_i$:

$$
\begin{aligned}
\frac{dL(\lambda_i)}{d\lambda_i} &= -N + \left( \sum_{j=1}^{N} x_i^j \right) \frac{1}{\lambda_i} = 0 \\
\Rightarrow \hat{\lambda}_i &= \frac{1}{N} \sum_{j=1}^{N} x_i^j \qquad (6)
\end{aligned}
$$

So $\hat{\lambda}_i$ is estimated for each depth for a sequence of $N$ observations and then used to calculate the Fisher Score of a new set of observations.

**Fisher Score for Gaussian Distributed Poisson Rates**

The previous model can be extended to take into account that the set of rates $\boldsymbol{\lambda}$ are not constant and also that there may be relationships between the $\lambda_i$ at different levels. One way of doing this is to assume a Gaussian distribution on $\boldsymbol{\lambda}$, i.e. $\boldsymbol{\lambda} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. A Fisher Score for the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can then be derived.

The Log Likelihood of a vector of observations $\mathbf{x}$ for a multivariate Gaussian is:

$$
L = -\frac{1}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \qquad (7)
$$

Differentiating (7) wrt $\boldsymbol{\mu}$ we get:

$$
\frac{dL}{d\boldsymbol{\mu}} = \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \qquad (8)
$$

and differentiating it wrt $\boldsymbol{\Sigma}$:

$$
\begin{aligned}
\frac{dL}{d\boldsymbol{\Sigma}} &= -\frac{1}{2} \frac{d}{d\boldsymbol{\Sigma}} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \frac{d}{d\boldsymbol{\Sigma}} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \\
&= -\frac{1}{2} \boldsymbol{\Sigma}^{-1} + \frac{1}{2} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} \qquad (9)
\end{aligned}
$$

Once again we have to use estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ from previous observations. The ML estimates of these parameters after $N$ observations are as follows:

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_j$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N}\sum_{j=1}^{N}(\mathbf{x}_j - \hat{\boldsymbol{\mu}})(\mathbf{x}_j - \hat{\boldsymbol{\mu}})^T \tag{10}$$

The Fisher Score is then constructed using these ML estimates for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (8) and (9).