

1 Introduction

This note describes the MCMC algorithm used to fit the overflow model. The bulk of this material may end up in a technical appendix or as a section of the paper. Section 2 briefly describes the model and section 3 describes the MCMC strategy for computing the posterior distribution.

2 The Overflow Model

The overflow model assumes we observe several pieces of data, 1) a noisy measurement of the arrival time of block k , t_k , measured in minutes since the zero'th block, 2) the number of transactions in block k , $M_k - M_{k-1}$ so that M_k is the number of transactions on the entire block chain up to and including block k , and 3) the amount of data the transactions require to be stored in the block, D_k , measured in MBs, all observed for $k = 1, 2, \dots, n_b$.

2.1 Measurement error for block arrival times

Let $m_k = \text{median}(t_{k-1:11})$ where $t_{k-1:11} = (t_{k-11}, t_{k-10}, \dots, t_{k-1})'$. Then by design $t_k > m_k$ for all k . Let τ_k denote the actual arrival time of the k 'th block. Then for $k = 1, 2, \dots, n_b$ t_k has a truncated normal distribution:

$$\log t_k | t_{1:(k-1)}, \tau_{1:n_b}, \sigma^2 \sim N(\log \tau_k, \sigma^2) 1(t_k > m_k).$$

2.2 Block interarrival times

By the design of the bitcoin protocol, blocks are a Poisson process with rate $1/10$. Interarrival times, defined as $\delta_k = \tau_k - \tau_{k-1}$ with $\tau_0 = 0$, are exponentially distributed with rate parameter $1/10$ so that $E[\delta_k] = 10$. In other words for $k = 1, 2, \dots, n_b$

$$\delta_k \stackrel{iid}{\sim} \text{Exp}(1/10).$$

2.3 Transaction arrival process

We will assume that transactions arrive according to a non-homogenous Poisson process. Let $N(\tau)$ denote the number of transaction in the entire chain up to time τ , and let $\lambda(\tau)$ denote some intensity function. Then any collection of non-overlapping increments $\{N(\tau_i + \delta_i) - N(\tau_i)\}$ are independent with distribution

$$N(\tau_i + \delta_i) - N(\tau_i) \sim \text{Poi}(\Lambda(\tau_i + \delta_i) - \Lambda(\tau_i))$$

where

$$\Lambda(\tau) = \int_0^\tau \lambda(u) du.$$

Suppose the intensity function is a gamma process with shape function $\phi\Psi(\tau)$ and rate parameter ϕ . In other words

$$\Lambda(\tau_i + \delta_i) - \Lambda(\tau_i) \stackrel{iid}{\sim} G(\phi[\Psi(\tau_i + \delta_i) - \Psi(\tau_i)], \phi)$$

with mean $\Psi(\tau_i + \delta_i) - \Psi(\tau_i)$. Then for block $k = 1, 2, \dots, n_b$ define $\lambda_k = \Lambda(\tau_k) - \Lambda(\tau_{k-1})$, $\eta_k = N(\tau_k) - N(\tau_{k-1})$, and $\psi_k = \Psi(\tau_k) - \Psi(\tau_{k-1})$. Then assume

$$\begin{aligned} \eta_k | \lambda_{1:n_b} &\stackrel{iid}{\sim} \text{Poi}(\lambda_k) \\ \lambda_k | \delta_{1:n_b} &\stackrel{iid}{\sim} G(\phi\psi_k, \phi). \end{aligned}$$

In order for the gamma process to be well defined we require $\phi > 0$ and $\Psi(\tau)$ to be nondecreasing and right continuous with $\Psi(0) = 0$. An easy way to achieve this is to define $\Psi(\tau) = \int_0^\tau \psi(t)dt$ where $\psi(t)$ is continuous and nonnegative. Then $\psi(t)$ is the intensity of the Gamma process, but is also intimately related to the intensity of the Poisson process. The mean of $N(\tau + \delta) - N(\tau)$ is

$$\begin{aligned} E[N(\tau + \delta) - N(\tau)] &= E\{E[N(\tau + \delta) - N(\tau)|\Lambda(\tau + \delta) - \Lambda(\tau)]\} = E[\Lambda(\tau + \delta) - \Lambda(\tau)] \\ &= \Psi(\tau + \delta) - \Psi(\tau) = \int_\tau^{\tau+\delta} \psi(t)dt. \end{aligned}$$

In this way, $\psi(t)$ can be seen as the expected intensity function of the $N(\tau)$ process. The Gamma-Poisson structure also allows for a more flexible model on $N(\tau)$ by allowing for overdispersion. In a Poisson model the mean and variance are the same, but in this case they are proportional:

$$\begin{aligned} V[N(\tau + \delta) - N(\tau)] &= E\{V[N(\tau + \delta) - N(\tau)|\Lambda(\tau + \delta) - \Lambda(\tau)]\} + V\{E[N(\tau + \delta) - N(\tau)|\Lambda(\tau + \delta) - \Lambda(\tau)]\} \\ &= E[\Lambda(\tau + \delta) - \Lambda(\tau)] + Var[\Lambda(\tau + \delta) - \Lambda(\tau)] \\ &= [\Psi(\tau + \delta) - \Psi(\tau)](1 + 1/\phi) = \frac{\phi + 1}{\phi} \int_\tau^{\tau+\delta} \psi(t)dt. \end{aligned}$$

This allows for overdispersion but not underdispersion since $(\phi + 1)/\phi \in [1, \infty)$ for $\phi \geq 0$.

To complete this portion of the model we need to specify $\psi(t)$. A flexible structure that captures many modeling choices is $\psi(t) = \exp(\mathbf{z}(t)'\boldsymbol{\beta})$ where $\mathbf{z}(t)' = (z_1(t), \dots, z_p(t))$ is a vector of time-varying covariates continuously differentiable in time and $\boldsymbol{\beta}' = (\beta_1, \dots, \beta_p)$ is a vector of regression coefficients, for example $\psi(t) = \exp(\beta_1 + t\beta_2)$ in order to capture a simple trend. In that case we have

$$\Psi(\tau + \delta) - \Psi(\tau) = \int_\tau^{\tau+\delta} e^{\mathbf{z}(t)'\boldsymbol{\beta}} dt = \int_\tau^{\tau+\delta} e^{\beta_1 + t\beta_2} dt = \frac{e^{\beta_1 + \tau\beta_2}}{\beta_2} (e^{\delta\beta_2} - 1)$$

and thus

$$\psi_k = \frac{e^{\beta_1 + \tau_{k-1}\beta_2}}{\beta_2} (e^{\delta_k\beta_2} - 1).$$

2.4 Transaction data

Let $i = 1, 2, \dots, n_{trans}$ index transactions and d_i denote the amount of data for transaction i measured in MBs. Then we assume that for $i = 1, 2, \dots, n_{trans}$

$$d_i \stackrel{iid}{\sim} G(\alpha_d, \beta_d)$$

with mean α_d/β_d . Along with the Poisson process for $N(\tau)$, this implies that $D(\tau)$, the amount of transaction data that has arrived by time τ , is a compound non-homogenous Poisson process conditional on $\Lambda(\tau)$.

2.5 Overflow and observations of M_k and D_k

Each block has an upper limit for the amount of transaction data it can hold defined by the block miner. Denote this upper limit by \bar{D}_k — because different miners use different limits and the same miner uses different limits at different times, each block has its own upper limit. Then the number of transactions in blocks $1, 2, \dots, k$ combined, M_k , could potentially be less than the number of transactions that have arrive by the time the k 'th block arrives, N_k . Specifically:

$$M_k = \max \left\{ m \in \{M_{k-1}, M_{k-1} + 1, \dots, N_k\} : \sum_{i=M_{k-1}+1}^m d_i \leq \bar{D}_k \right\}$$

so that

$$D_k = \sum_{i=M_{k-1}+1}^{M_k} d_i$$

with the convention that empty summations are defined as zero, i.e. $\sum_{i=5}^4 d_i \equiv 0$. Finally, the upper limits are modeled as a categorical distribution over a set of prespecified possible limits, $\Delta_1, \dots, \Delta_{n_{limits}}$ with probability $p_{ik} = P(\bar{D}_k = \Delta_i)$, $\sum_i p_{ik} = 1$ and $p_{ik} \geq 0$. In other words

$$\bar{D}_k \stackrel{iid}{\sim} Cat(\Delta_{1:n_{limits}}, p_{1:n_{limits},k}).$$

3 Markov Chain Monte Carlo

In order to simulate from the posterior we construct a Gibbs sampler using data augmentation to expand the state space and Metropolis steps where appropriate. We use data augmentation in two ways. First, in order to more effectively deal with the truncated distribution in the measurement error model for block arrival times we introduce $\tilde{\mathbf{t}}_k = (\tilde{t}_{1k}, \dots, \tilde{t}_{\tilde{n}_k k}, t_k)$ as the original t_k along with \tilde{n}_k draws from the untruncated lognormal distribution for t_k that occurred before t_k was drawn from the restricted support. Second, the models for $\eta_{1:n_b}$ and $d_{1:n_{trans}}$ induce a model for $(M_{1:n_b}, D_{1:n_{trans}})$ since the latter are a deterministic function of the former, but this model is not easy work with. Instead, we explicitly use $\eta_{1:n_b}$ and $d_{1:n_{trans}}$ in the Gibbs sampler as additional data augmentation.

3.1 Data augmentation for truncated distributions

For block $k = 1, 2, n_b$ the measurement error model on block arrival times is

$$\log t_k \sim N(\log \tau_k, \sigma^2) 1(t_k > m_k)$$

where $m_k = \text{median}(t_{k-1:11})$ with density

$$p(\log \mathbf{t}_{1:n_b} | \boldsymbol{\tau}_{1:n_b}, \sigma^2) \propto \sigma^{-n_b} \frac{\exp \left[-\frac{1}{2\sigma^2} \sum_{k=1}^{n_b} (\log t_k - \log \tau_k)^2 \right]}{\prod_{k=1}^{n_b} \left[1 - \Phi \left(\frac{\log m_k - \log \tau_k}{\sigma} \right) \right]}$$

where $\Phi(\cdot)$ is the standard normal cdf. Drawing from the full conditional distributions of σ^2 and τ_k can be challenging since both parameters enter into the normalizing constant of $p(\log \mathbf{t}_{1:n_b} | \boldsymbol{\tau}_{1:n_b}, \sigma^2)$. Consider an underlying process which gives rise to the truncated distribution — instead of simply drawing t_k from the truncated lognormal distribution, this process draws from the untruncated lognormal distribution until it obtains a draw from the restricted region of the parameter space, i.e. until it successfully obtains a t_k such that $t_k > m_k$. Let \tilde{n}_k be the number of failures before the observed success t_k . Then \tilde{n}_k has a negative binomial distribution, specifically $\tilde{n}_k \sim NB(1, p_k)$ with density

$$p(\tilde{n}_k | \tau_k, \sigma^2) = p_k^{\tilde{n}_k} (1 - p_k)$$

where p_k is the probability of failure, i.e. $p_k = \Phi([\log m_k - \log \tau_k]/\sigma)$. Conditional on \tilde{n}_k the density of the \tilde{n}_k failures, $\tilde{t}_{1k}, \dots, \tilde{t}_{\tilde{n}_k k}$, is again truncated lognormal, but this time truncated to the opposite end of the parameter space. That is for $i = 1, \dots, \tilde{n}_k$

$$\log \tilde{t}_{ik} \stackrel{iid}{\sim} N(\log \tau_k, \sigma^2) 1(t_k < m_k).$$

So the data augmentation step of the Gibbs sampler consists of the following substeps: for $k = 1, 2, \dots, n_b$,

1. Draw $\tilde{n}_k \sim NB(1, p_k)$ where $p_k = \Phi([\log m_k - \log \tau_k]/\sigma)$.

2. If $\tilde{n}_k > 0$, for $i = 1, \dots, \tilde{n}_k$ draw $\tilde{t}_{ik} \sim N(\log \tau_k, \sigma^2)1(t_k < m_k)$ and form $\tilde{\mathbf{t}}_k = (\tilde{t}_{1k}, \dots, \tilde{t}_{\tilde{n}_k k}, t_k)'$.

For simplicity in the other Gibbs steps define $\tilde{t}_{\tilde{n}_k+1,k} \equiv t_k$.

It is tempting to imbue \tilde{n}_k and \tilde{t}_{ik} with some interpretation relative to the model, e.g. as is [CITE GELMAN PAPER]. For example, \tilde{n}_k may represent the number of blocks that were rejected by the Bitcoin protocol for having a timestamp earlier than the median of the previous eleven blocks. While tempting this interpretation is not strictly speak correct. When a block is rejected by the protocol a new block does not appear instantaneously since it takes time for another miner to discover the block. So the arrival times and thus measurement error distributions of the rejected block and the accepted block are not the same. While the interpretation is not correct, it is good to be inspired by computational tricks to create better model [CITE SAME GELMAN PAPER] and perhaps something similar to this negative binomial structure can be used to add rejected blocks as a component of the model.

3.2 measurement error variance

The full conditional distribution of the measurement error variance, σ^2 , depends only on $\tilde{\mathbf{t}}_{1:n_b}$ and $\tau_{1:n_b}$. It is

$$p(\sigma^2 | \dots) \propto \sigma^{-(n_b + \sum_{k=1}^{n_b} \tilde{n}_k)} \exp \left[-\frac{1}{2\sigma^2} \sum_{k=1}^{n_b} \sum_{i=1}^{\tilde{n}_k} (\log \tilde{t}_{ik} - \log \tau_k)^2 \right] p(\sigma^2)$$

where $p(\sigma^2)$ is the density of the prior distribution on σ^2 . We will suppose that σ has a half- t prior, $\sigma \sim t_\nu(S)1(\sigma > 0)$ with degrees of freedom and scale parameter S . This prior is equivalent to the marginal prior on σ when $\sigma^2 | \omega \sim IG(\nu/2, \omega)$ and $\omega \sim G(1/2, 1/(\nu S^2))$. Then letting $a_\sigma = (n_b + \sum_{k=1}^{n_b} \tilde{n}_k)/2$ and $b_\sigma = \sum_{k=1}^{n_b} \sum_{i=1}^{\tilde{n}_k} (\log \tilde{t}_{ik} - \log \tau_k)^2$ the full conditional distribution of (σ^2, ω) is

$$p(\sigma^2, \omega | \dots) \propto (\sigma^2)^{-(a_\sigma + \nu)/2-1} \left[-\frac{b_\sigma}{\sigma^2} \right] \exp \left[-\frac{\omega}{\sigma^2} \right] \omega^{(\nu+1)/2-1} \exp \left[-\frac{\omega}{\nu S^2} \right].$$

So we draw σ^2 and ω in two separate Gibbs steps:

1. Draw $\omega \sim G\left(\frac{\nu+1}{2}, \frac{1}{\sigma^2} + \frac{1}{\nu S^2}\right)$ and form

$$a_\sigma = n_b + \sum_{k=1}^{n_b} \tilde{n}_k \quad \text{and} \quad b_\sigma = \sum_{k=1}^{n_b} \sum_{i=1}^{\tilde{n}_k} (\log \tilde{t}_{ik} - \log \tau_k)^2 / 2;$$

2. Draw $\sigma^2 \sim IG([a_\sigma + \nu]/2, b_\sigma + \omega)$.

3.3 block arrival times

The full conditional distribution of the block arrival times is more complicated:

$$p(\tau_{1:n_b} | \dots) \propto \exp \left[-\frac{1}{2\sigma^2} \sum_{k=1}^{n_b} \sum_{i=1}^{\tilde{n}_k+1} (\log \tilde{t}_{ik} - \log \tau_k)^2 - \frac{1}{10} \tau_{n_b} \right] \frac{\phi^{\phi[\Psi(\tau_k) - \Psi(\tau_{k-1})]}}{\Gamma(\phi[\Psi(\tau_k) - \Psi(\tau_{k-1})])} \lambda_k^{\phi[\Psi(\tau_k) - \Psi(\tau_{k-1})] - 1}$$

where $\tau_0 \equiv 0$ and $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the gamma function. Even $\Psi(\tau)$ is some nice, known form, this density is nonstandard because the τ_k 's enter the Gamma function. As a result, we have several options for this step:

1. Draw each τ_k conditional on all the others using random walk Metropolis steps for $k = 1, \dots, n_b$.
2. Approximate the density of τ_k sufficiently well and use independent Metropolis steps for $k = 1, \dots, n_b$.

3. Approximate the density of $\tau_{1:n_b}$ sufficiently well and use an independent Metropolis step.

Options 2 and 3 could both be replaced with a rejection sampling step under some circumstances, but a Metropolis step is likely cheaper computationally [CITE THE LIU PAPER].

Ideas:

1. ARMS – ARS except with a Metropolis correction for non-log concave densities (look it up).
2. RW Metropolis and tune the proposal variances.

3.4 transaction process dispersion

The transaction process dispersion parameter is ϕ — $V[N(\tau)] = E[N(\tau)](1 + 1/\phi)$. It has the full conditional density

$$p(\phi|\dots) \propto \prod_{k=1}^{n_b} \frac{\phi^{\phi\psi_k} \lambda_k^{\phi\psi_k-1} e^{\lambda_k\phi}}{\Gamma(\phi\psi_k)} p(\phi).$$

This is another complicated density, but ARS should work decently to sample from it. If not, various Metropolis options are available. Another option: transform so that ϕ is absorbed into the shape function. Then ϕ has the full conditional

$$p(\phi|\dots) \propto \phi^{\sum_{k=1}^{n_b} \psi_k} e^{\phi \sum_{k=1}^{n_b} \lambda_k} p(\phi).$$

With $\phi \sim G(a_\phi, b_\phi)$ in the prior, this becomes

1. Draw $\phi \sim G(a_\phi + \sum_{k=1}^{n_b} \psi_k, b_\phi + \sum_{k=1}^{n_b} \lambda_k)$.

3.5