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 4 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 measured in minutes since the zero'th block and denoted by t_k, 2) the number of transactions in block k denoted by $M_k - M_{k-1}$ 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 transaction data stored in block k measured in MBs and denoted by D_k, all observed for $k = 1, 2, \ldots, n_b$.

2.1 Measurement error for block arrival times

Let $m_k = \text{median}(t_{k-1:11})$ where $t_{k-1:11} = (t_{k-11}, t_{k-10}, \ldots, t_{k-1})'$. 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 $k = 1, 2, \ldots, n_b$, t_k has a truncated normal distribution:

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

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 $k = 1, 2 \ldots, n_b$

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

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 transactions that have arrive by time τ, and let $\lambda(\tau)$ denote some intensity function such that $\lambda(\tau) > 0$ for all times τ. Then any collection of non-overlapping increments $\{N(\tau_i + \delta_i) - N(\tau_i)\}$ where $\delta_i > 0$ for all i are independent with distribution

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

where

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

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

$$\Lambda(\tau_i + \delta_i) - \Lambda(\tau_i) \overset{ind}{\sim} G(\Psi(\tau_i + \delta_i) - \Psi(\tau_i), \phi)$$

with mean $[\Psi(\tau_i + \delta_i) - \Psi(\tau_i)]/\phi$. Then for block $k = 1, 2, \ldots, n_b$ 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

$$\eta_k | \lambda_{1:n_b} \overset{ind}{\sim} Poi(\lambda_k)$$
$$\lambda_k | \delta_{1:n_b} \overset{ind}{\sim} G(\psi_k, \phi).$$

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. While $\psi(t)$ can be thought of as the intensity of the Gamma process it is also intimately related to the intensity of the Poission process. The mean of $N(\tau+\delta) - N(\tau)$ is

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

In this way, $\psi(t)/\phi$ 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:

$$\begin{aligned} V[N(\tau+\delta) - N(\tau)] &= E\{V[N(\tau+\delta) - N(\tau)|\Lambda(\tau+\delta) - \Lambda(\tau)]\} + V\{E[N(\tau+\delta) - N(\tau)|\Lambda(\tau+\delta) - \Lambda(\tau)]\} \\ &= E[\Lambda(\tau+\delta) - \Lambda(\tau)] + V[\Lambda(\tau+\delta) - \Lambda(\tau)] \\ &= [\Psi(\tau+\delta) - \Psi(\tau)](1/\phi + 1/\phi^2) = \frac{\phi+1}{\phi} E[N(\tau+\delta) - N(\tau)]. \end{aligned}$$

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 $\psi(t) = \exp(\boldsymbol{z}(t)'\boldsymbol{\beta})$ where $\boldsymbol{z}(t)' = (z_1(t), \ldots, z_p(t))$ is a vector of time-varying covariates continuously differentiable in time and $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_p)$ 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

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

and thus

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

2.4 Transaction data

Let $i = 1, 2, \ldots, n_d \equiv N_{n_b}$ index transactions and d_i denote the amount of data for transaction i measured in MBs. Then we assume that for $i = 1, 2, \ldots, n_d$

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

with mean α_d/β_d. Along with the Poisson process for $N(\tau)$ this implies that the amount of transaction data that has arrived by time τ, $D(\tau)$, 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 $1, 2, \ldots, k$ combined, M_k, could potentially be less than the number of transactions that have arrived by the time the k'th block arrives, N_k. Specifically:

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

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. $\sum_{i=5}^{4} d_i \equiv 0$.

3 Priors

The model has the following unknown parameters: σ^2, ϕ, β, α_d, β_d, and $\bar{D}_k$ for $k = 1, 2, \ldots, n_b$. We will suppose that these parameters are mutually independent in the prior, i.e.

$$p(\sigma^2, \phi, \beta, \alpha_d, \beta_d, \bar{D}_1, \ldots, \bar{D}_{n_b}) = p(\sigma^2)p(\phi)p(\beta)p(\alpha_d)p(\beta_d)p(\bar{D}_1)\ldots p(\bar{D}_{n_b}).$$

3.1 Prior on σ^2

The expected size of the measurement error on t_k is controlled by σ^2 — the larger σ^2, the farther away we expect t_k to be from τ_k. We will use a half-t prior on σ which can be written as the following prior on σ^2:

$$\sigma^2|\omega \sim IG(v/2, \omega); \qquad \omega \sim G(1/2, s^2/2).$$

This yields a $t_v(0, s^2)1(\sigma > 0)$ prior on σ where s^2 is a scale parameter and v is a degrees of freedom parameter. We use $v = 1$ so that the prior on σ has a half-Cauchy distribution with median s. To gain some intuition in order to set an appropriate value for s, suppose that the measurement error distribution was not truncated. Then we would expect 68% of observed block arrival times to be within σ of the actual arrival time. A reasonable a priori guess for σ is then one minute, so we set $s = 1$ so that the median value of σ in its prior is one. This prior is only weakly informative — the right tail is gently sloped so that the data will quickly overwhelm the prior as observations pile up [CITE GELMAN VARIANCE PAPER].

3.2 Prior on ϕ

The conditionally conjugate prior for ϕ is a gamma distribution, so we will suppose that $\phi \sim G(a_\phi, b_\phi)$. To choose a_ϕ and b_ϕ recall that ϕ is a dispersion parameter:

$$V[N(\tau)|\phi, \Psi(\tau)] = \frac{\phi + 1}{\phi} E[N(\tau)|\phi, \Psi(\tau)] = \frac{\phi + 1}{\phi} \frac{\Psi(\tau)}{\phi}.$$

Setting $\phi = 1$ means that the variance of $N(\tau)$ is twice its mean. Absent strong prior information about ϕ, we will center our prior on 2 with a high level of uncertainty. The mean of the gamma distribution is a_ϕ/b_ϕ with variance a_ϕ/b_ϕ^2, so we will set $b_\phi = 1/100$ to reflect a high degree of uncertainty, and $a_\phi = 1/50$ so that $a_\phi/b_\phi = 2$.

3.3 Prior on $\boldsymbol{\beta}$

As we will see in the next section, there is no convenient conditionally conjugate form for $\boldsymbol{\beta}$, so there is no computational trade-off associated with choosing a prior which more accurately reflects prior knowledge. We will consider $\boldsymbol{\beta}$ much like a regression parameter and set

$$\boldsymbol{\beta} \sim N(\boldsymbol{b}_\beta, \boldsymbol{S}_\beta)$$

where $\boldsymbol{b}_\beta$ is a mean vector and $\boldsymbol{S}_\beta$ is a covariance matrix. We will use a weakly informative prior with $\boldsymbol{b}_\beta = \boldsymbol{0}$ and $\boldsymbol{S}_\beta = s_\beta^2 \boldsymbol{I}$ where $s_\beta = 100$. [MAYBE SOMETHING WITH FATTER TAILS HERE? MAYBE NOTE THAT NONINFORMATIVE PRIOR IS NOT A GOOD IDEA SINCE WE WANT TO GUARANTEE POSTERIOR PROPRIETY — SAME FOR s_β TOO LARGE (VIVEK'S PAPER)]

3.4 Priors on α_d and β_d

The conditionally conjugate prior for $beta_d$ is a gamma while α_d has no convenient conditionally conjugate form. We will assume that $\beta_d \sim G(a_\beta, b_\beta)$ and $\alpha_d \sim G(a_\alpha, b_\alpha)$. In order to choose a_α, b_α, a_β, and b_β, consider that $E[d_i|\alpha_d, \beta_d] = \alpha_d/\beta_d$. But

$$E\left[\frac{\alpha_d}{\beta_d}\right] = E[\alpha_d]E\left[\frac{1}{\beta_d}\right] = \frac{a_\alpha}{b_\alpha}\frac{b_\beta}{a_\beta - 1}.$$

Similarly

$$V\left[\frac{\alpha_d}{\beta_d}\right] = V[\alpha_d] + V\left[\frac{1}{\beta_d}\right] = \frac{a_\alpha}{b_\alpha^2} + \frac{b_\beta^2}{(a_\beta - 1)^2(a_\beta - 2)}.$$

We will choose these parameters so that $E[\alpha_d/\beta_d] \approx 0.01$ MB and $V[\alpha_d/\beta_d]$ is large so that the prior is weakly informative. We will set $a_\beta = 2.1$ to ensure the variance calculated above exists. Then $a_\alpha = 0.1$, $b_\alpha = 1$, and $b_\beta = 1$ yields

$$E\left[\frac{\alpha_d}{\beta_d}\right] = \frac{1}{21} \approx 0.09; \qquad V\left[\frac{\alpha_d}{\beta_d}\right] \approx 10.$$

3.5 Priors on the $\bar{D}_k$'s

We will put a discrete prior on each of the $\bar{D}_k$'s. Theoretically a miner can put any upper limit she wants on a mined bitcoin, but in practice most miners use one of several possible values. We will use the set of values $\mathcal{D} = \{0.25, 0.35, 0.5, 0.75, 0.9, 0.95, 1\}$ [DOUBLE CHECK THESE — IN PARTICULAR, ARE THERE OTHER COMMON VALUES? WHAT ABOUT THE ELGIUS MINING POOL'S CAPS?]. For each $\bar{D}_k$ we will use a discrete uniform prior on $\mathcal{D}$ so that $P(\bar{D}_k = j) = 1/7$ for all $j \in \mathcal{D}$.

[OPTIONS FOR MORE INTERESTING THINGS HERE: DIFFERENT PRIOR FOR EACH k SO THAT LATER BLOCKS HAVE A HIGHER PROBABILITY OF HAVING A HIGHER CAP. THIS USES THE DATA TWICE, SO A BETTER OPTION MAY JUST BE MODELING THE PROBABILITIES ON $\mathcal{D}$ AS EVOLVING OVER TIME BASED ON SOME HIGHER LEVEL PARAMETERS]

4 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 $\tilde{\boldsymbol{t}}_k = (\tilde{t}_{1k}, \ldots, \tilde{t}_{\tilde{n}_k k}, t_k)$ as the original t_k along with $\tilde{n}_k$ draws from the untruncated normal distribution for t_k that occurred before t_k was drawn from the restricted support. Second, the models for $\eta_{1:n_b}$ and $d_{1:n_d}$ induce a model for $(M_{1:n_b}, D_{1:n_d})$ since the latter are a detrministic function of the former, but this model is not easy work with. Instead, we explicitly use $\eta_{1:n_b}$ and $d_{1:n_d}$ in the Gibbs sampler as additional data augmentation.

4.1 Data augmentation for truncated distributions — $\tilde{n}_k$ and $\tilde{t}_k$

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

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

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

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

where $\Phi(.)$ 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 $p(\boldsymbol{t}_{\mathbf{1:n_b}}|\boldsymbol{\tau}_{1:n_b}, \sigma^2)$. Consider an underlying process which gives rise to the truncated distribution — instead of simply drawing t_k from the truncated normal distribution, this process draws from the untruncated normal distribution until it obtains a draw from the restricted region of the parameter space, i.e. until it successfully obtains a $t_k > m_k$. 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 $\tilde{n}_k \sim NB(1, \rho_k)$ with density

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

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

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

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

1. Draw $\tilde{n}_k \sim NB(1, \rho_k)$ where $\rho_k = \Phi([m_k - \tau_k]/\sigma)$.
2. If $\tilde{n}_k > 0$, for $i = 1, \ldots, \tilde{n}_k$ draw $\tilde{t}_{ik} \sim N(\tau_k, \sigma^2)1(t_k < m_k)$ and form $\tilde{\boldsymbol{t}}_k = (\tilde{t}_{1k}, \ldots, \tilde{t}_{\tilde{n}_k k}, t_k)'$.

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 in [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 elevent blocks. While tempting this interpretation is not strictly speaking 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, though we do not explore this possibility here.

4.2 Measurement error variance — σ^2

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

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

where $p(\sigma^2)$ is the density of the prior distribution on σ^2. With the half-t prior on σ discussed above and letting $a_\sigma = (n_b + \sum_{k=1}^{n_b} \tilde{n}_k)/2$ and $b_\sigma = \sum_{k=1}^{n_b} \sum_{i=1}^{\tilde{n}_k} (\tilde{t}_{ik} - \tau_k)^2$ the full conditional distribution of (σ^2, ω) is

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

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

1. Draw $\omega \sim G\left(\frac{v+1}{2}, \frac{1}{\sigma^2} + \frac{1}{vs^2}\right)$ and form[1]

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

2. Draw $\sigma^2 \sim IG\left(a_\sigma, b_\sigma/2 + \omega\right)$.

[1] The inconsistency in notation between the definition of a_σ and b_σ is a consequence of matching notation to the R code supplied for implementing this MCMC algorithm.

4.3 Block arrival times — τ_k

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

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

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

1. Draw each τ_k condtional on all the others using random walk Metropolis steps for $k = 1, \ldots, n_b$.
2. Approximate the density of τ_k sufficiently well and use independent Metropolis steps for $k = 1, \ldots, n_b$.
3. Approximate the density of $\tau_{1:n_b}$ sufficiently well and use an independent Metropolis step.
4. Use adaptive rejection sampling (ARS) to draw from the density of τ_k.
5. Use ARMS — ARS except with a Metropolis correction for densities which are not log concave (look it up).

4.4 Transaction process dispersion — ϕ

The full conditional density of ϕ is

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

With $\phi \sim G(a_\phi, b_\phi)$ in the prior, this is a Gamma density so the step is simple:

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

4.5 Regression parameters — $\boldsymbol{\beta}$

The full conditional posterior of $\boldsymbol{\beta}$ is closely connected to the functional form of the covariates in time. No matter the functional form, however, the density is complex and a Metropolis step is likely warranted. Recall that since we defined $\psi(t) = e^{\boldsymbol{z}(t)'\boldsymbol{\beta}}$ we have $\psi_k(\boldsymbol{\beta}) = \int_{\tau_{k-1}}^{\tau^k} e^{\boldsymbol{z}(t)'\boldsymbol{\beta}}dt$ — in this subsection we write $\psi_k(\boldsymbol{\beta})$ to explicitly acknowledge that ψ_k is a function of $\boldsymbol{\beta}$. Then the full conditional of $\boldsymbol{\beta}$ is

$$p(\boldsymbol{\beta}|\ldots) \propto p(\boldsymbol{\beta})\prod_{k=1}^{n_b}\frac{\phi^{\psi_k(\boldsymbol{\beta})}\lambda_k^{\psi_k(\boldsymbol{\beta})-1}}{\Gamma[\psi_k(\boldsymbol{\beta})]}$$

where in the prior $\boldsymbol{\beta} \sim p(\boldsymbol{\beta})$ independent of the other parameters. This density is complex and nonstandard since $\psi_k(\boldsymbol{\beta})$ enters the gamma function. A Metropolis step is likely warranted here as well.

4.6 Transaction arrival times and sizes in MB — $N(\tau_k)$ and D_k

First consider transaction sizes, $\boldsymbol{d} = (d_1, d_2, \ldots, d_{n_d})'$. Let $\boldsymbol{d}_k$ denote the sizes of the $M_k - M_{k-1}$ transactions that were stored in block k so that $\boldsymbol{d} = (\boldsymbol{d}_1', \ldots, \boldsymbol{d}_{n_b}')'$. The full conditional distribution of the $\boldsymbol{d}_k$'s is

$$p(\boldsymbol{d}_{1:n_b}|...) \propto \prod_{k=1}^{n_b}\prod_{i=1}^{M_k-M_{k-1}} d_i^{\alpha_d-1}e^{-d_i\beta_d}1\left(\sum_{i=1}^{M_k-M_{k-1}} d_{ik} = D_k\right)1\left(d_{1k} > \bar{D}_{k-1} - D_{k-1} \text{ or } M_{k-1} = N_{k-1}\right).$$

The second indicator function is required because if $M_{k-1} < N_{k-1}$ then $N_{k-1} - M_{k-1}$ transactionss must have overflowed into block k. So the amount of data from the first transaction of block k, d_{1k}, must have been greater than the difference between the amount of data in block $k-1$ and the data cap for block $k-1$. Similarly if $M_{k-1} = N_{k-1}$ then we know there was no overflow into block k, so we have no reason to think that d_{1k} has a restricted support. Let $\boldsymbol{o}_k = \boldsymbol{d}_k/D_k$. Then the $\boldsymbol{o}_k$'s have potentially truncated independent Dirichlet distributions. That is

$$p(\boldsymbol{o}_{1:n_b}|...) \propto \prod_{k=1}^{n_b} \prod_{i=1}^{M_k - M_{k-1}} o_{ik}^{\alpha_d - 1} 1\left(d_{1k} > [\bar{D}_{k-1} - D_{k-1}]/D_k \text{ or } M_{k-1} = N_{k-1}\right).$$

So the $\boldsymbol{d}$ step to draw from $p(\boldsymbol{d}|...)$ is as follows: Let $\mathbf{1}_n$ denote an n-vector of ones. For $k = 1, 2, \ldots, n_b$

1. If $N_k = M_k$ draw $\boldsymbol{o}_k \sim Dir(\alpha_d \mathbf{1}_{M_k - M_{k-1}})$.
   Otherwise draw $\boldsymbol{o}_k \sim Dir(\alpha_d \mathbf{1}_{M_k - M_{k-1}}) 1(d_{1k} > [\bar{D}_{k-1} - D_{k-1}]/D_k)$.
2. Set $\boldsymbol{d}_k = \boldsymbol{o}_k D_k$.

To draw from the truncated Dirichlet distribution above, first note that by the properties of the Dirhichlet the marginal distribution of o_{1k} is $Beta(\alpha_d, \alpha_d(M_k - M_{k-1} - 1))$ and conditional on o_{1k}, $\boldsymbol{o}_{-1,k}/(1 - o_{1k}) \sim Dir(\alpha_d \mathbf{1}_{M_k - M_{k-1} - 1})$. Then o_{1k} can be drawn from the truncated Beta distribution using the inverse cdf method, resulting in the following two step algorithm to draw from the desired truncated Dirichlet distribution:

1. Draw $u \sim U(0,1)$ and set $o_{1k} = F^{-1}(u * [1 - F((\bar{D}_{k-1} - D_{k-1})/D_k)])$ where F is the cdf of the $Beta(\alpha_d, \alpha_d(M_k - M_{k-1} - 1))$ distribution.
2. Draw $\tilde{\boldsymbol{o}}_{2:(M_k - M_{k-1})} \sim Dir(\alpha_d \mathbf{1}_{M_k - M_{k-1} - 1})$ and set $o_{ik} = \tilde{o}_i$ for $i = 2, 3, \ldots, M_k - M_{k-1}$.

To draw the N_k's first note that their support is restricted by the value of the M_k's. Specifically for $k = 1, 2, \ldots, n_b$ and $N_0 = 0$ we have $M_k \leq N_k \leq N_{k+1}$. Second, when

$$\sum_{i=1}^{M_k - M_{k-1} + 1} d_{M_{k-1} + i} \leq \bar{D}_k$$

we know that $N_k = M_k$ because otherwise transaction $M_k + 1$ would be in block k. If the summation is greater than D_k then it is still possible that $N_k = M_k$ because transaction $d_{M_k + 1}$ could still have arrived later than block k. In that case, the density of N_k given $N_{0:(k-1)}$ and all other parameters and processes is given by

$$p(N_k | N_{0:(k-1)}, \ldots) \propto \frac{\lambda_k^{N_k - N_{k-1}} e^{-\lambda_k}}{(N_k - N_{k-1})!} 1(N_k \geq M_k) 1(N_k \geq N_{k-1}).$$

In other words, $\eta_k = N_k - N_{k-1}$ has a truncated Poisson distribution where the support is restricted to $\{M_k - N_{k-1}, M_k - N_{k-1} + 1, \ldots\}$. We can easily simulate from this distribution using the inverse cdf method. So we can draw from $p(\boldsymbol{N}_{1:n_b}| \ldots)$ as follows: For $k = 1, 2, \ldots, n_b$

1. If $\sum_{i=1}^{M_k - M_{k-1} + 1} d_{M_k + i} \leq \bar{D}_k$ set $N_k = M_k$.
   Otherwise, draw $\eta_k \sim Poi(\lambda_k) 1(\eta_k \geq M_k - N_{k-1})$ and set $N_k = N_{k-1} + \eta_k$.

4.7 Block size caps — $\bar{D}_k$

Since the caps are independent discrete uniform in the prior, they are independent in their full conditional and for each block k, there are two possibilities:

1. When $N_k > M_k$, $D_k \leq \bar{D}_k < D_k + d_{M_k+1}$.
2. When $N_k = M_k$, $D_k \leq \bar{D}_k$.

In the first case this should completely determine the value of $\bar{D}_k$ since transaction sizes are much smaller than the differences between possible caps. In both cases we draw $\bar{D}_k$ from a discrete uniform distribution over the set of possible caps given the relevant constraint above — sometimes that distribution puts all of the mass on a single value.

4.8 Transaction data parameters — α_d and β_d

The full conditional distribution of (α_d, β_d) is

$$p(\alpha_d, \beta_d|...) \propto p(\alpha_d)p(\beta_d)\prod_{i=1}^{n_d}\frac{\beta_d^{\alpha_d}}{\Gamma(\alpha_d)}d_i^{\alpha_d-1}e^{-\beta_d d_i}.$$

When $\beta_d \sim G(a_\beta, b_\beta)$ in the prior, its full conditional becomes

$$p(\beta_d|...) \propto \beta_d^{\alpha_d n_d + a_\beta - 1}e^{-\beta_d\left(\sum_{i=1}^{n_d} d_i + b_\beta\right)},$$

which is another gamma distribution. The full conditional of α_d is complex no matter what its prior is since α_d enters the gamma function, and has the form

$$p(\alpha_d|...) \propto p(\alpha_d)\frac{\left(\beta_d^{n_d}\prod_{i=1}^{n_d} d_i\right)^{\alpha_d}}{\Gamma(\alpha_d)^{n_d}}$$

which in the case of a gamma prior on α_d is

$$p(\alpha_d|...) \propto \alpha_d^{a_\alpha - 1}\frac{\left(e^{-b_\alpha}\beta_d^{n_d}\prod_{i=1}^{n_d} d_i\right)^{\alpha_d}}{\Gamma(\alpha_d)^{n_d}}.$$

This density can be drawn from using rejection sampling or an independence Metropolis step, e.g. ARS or ARMS.