

# Banerjee Text

## Outline and Most Important Results

The most relevant portion of this text for my purposes is the discussion on Bayesian model fitting and inference for shot noise processes. To build up to this result I will discuss the metropolis Hastings algorithms, the shot noise process, and how they tie together. I will also include a list of references I hope to read in preparation for our implementation.

## The Metropolitan-Hastings Algorithm

The Metropolitan-Hastings Algorithm is a rejection algorithm that requires only a function proportional to the distribution to be sampled. This is in contrast to the, more intuitive, Gibbs Sampler which requires we know the full conditional distributions,  $p(\theta_i|\theta_{i \neq j}, y)$ . However, we often do not these full conditional distributions. Rather, we know them up to a proportionality constant. (This part I don't full understand)

The main purpose of the algorithm is for generation from full conditionals. Suppose we want to generate from a joint posterior distribution  $p(\theta|y) \propto h(\theta) \equiv f(y|\theta)p(\theta)$ . First, we specify a candidate density  $q(\theta^*|\theta^{(t-1)})$  that is a valid density function for every possible value of the conditioning variable. Given a starting value  $\theta^{(0)}$  at iteration  $t = 0$ , the algorithm proceeds:

*Metropolis-Hastings Algorithm:* For  $(t \in 1 : T)$ , repeat:

1. Draw  $\theta^*$  from  $q(\cdot|\theta^{(t-1)})$
2. Compute the ratio  $r = \frac{h(\theta^*)q(\theta^{(t-1)}|\theta^*)}{h(\theta^{(t-1)})q(\theta^*|\theta^{(t-1)})}$
3. If  $r \geq 1$ , set  $\theta^{(t)} = \theta^*$ . Otherwise, set  $\theta^{(t)} = \theta^{(t-1)}$  with probability  $1 - r$ .

Then, under mild conditions, a draw  $\theta^{(t)}$  converges in distribution to a draw from the true posterior density  $p(\theta|y)$  as  $t \rightarrow \infty$ .

Apparently, it is common in practice to set  $q(\theta^*|\theta^{(t-1)}) = q(\theta^*)$ , i.e. the proposal density ignores the current value of the variable. This algorithm is sometimes referred to as a **Hastings independence chain**. Though easy to implement, this algorithm will converge slowly unless the chosen  $q$  is rather close to the true posterior.

## Shot Noise Processes

Similar to a Neyman-Scott process, this shot noise process is defined in two stages. For a given region,  $D$ , create a point pattern  $S$  over  $D$  from a Poisson process. Then, assign a random mass to each sampled location. The process realization at  $s$  is  $Y(s) = \sum_{s_i \in S} h(s - s_i; m(s_i))$ . Forms for  $h$  include  $h(s - s_i; m(s_i)) = f(s - s_i)m(s_i)$  where  $f$  is a density over  $D$  and  $m(s_i)$  a positive random variable. The  $m$ 's might be IID, a regression on some covariate, or a process realization over  $D$ . So,  $m(s_i)$  represents the contribution to  $Y(s)$  from the point at  $s_i$  and  $Y(s)$  accumulates the "shots" arising from the realization,  $S$ .

Generally, we can compute  $E_S Y(S)$  as a single integral over  $S$ , given  $m(\cdot)$  or as a double integral over the randomness in  $m$  using Campbell's theorem (I need to look up what this is).

Shot noise processes are introduced as models for intensities. That is,  $Y(s)$  is viewed as  $\lambda(s)$ , creating a random intensity.

## Example

Suppose we write  $\lambda(s) = e^{X^T(s)\beta}$

$\lambda_0(s)$  where the exponential term is a deterministic specification and  $\lambda_0(s)$  is a mean 1 shot noise process so that  $\lambda(s)$  is centered around the deterministic component. Suppose we adopt the form  $\lambda_0(s) = \sum_{s_i \in S} f(s - s_i)m(s_i)$  with  $S$  drawn from an homogeneous Poisson process( $\lambda$ ) and  $m(s_i)$  a constant,  $m$ . From Campbells theorem, that I still need to learn (though the intuition from the result is super clear), we have  $\mathbb{E}[\lambda_0(s)] = m\lambda = 1$ , so  $m = 1/\lambda$ .

The likelihood arises as follows

$$\begin{aligned} L(\beta, \lambda_0(s), s \in D; S_{obs}) &= e^{-\lambda(D)} \prod_i \lambda(s_i) \\ &= e^{-\int_D e^{X^T(s)\beta} \lambda_0(s) ds} \prod_i e^{X^T(s_i)\beta} \lambda_0(s_i) \end{aligned}$$

This is something I hope to model and look more into. But, the text says that the likelihood is conditional on  $S$ , a realization from an HPP( $\lambda$ ), i.e.,  $\lambda_0(s)$  is a function of  $S$ . So we first draw  $S$  given  $\lambda$ , and then we draw  $S_{obs}$  given  $\beta$  and  $S$ . Actually, this makes sense.

## Fitting a Modified Metropolis-Hastings Probability

This portion is important if we want to simulate, but I am having a difficult time understanding it currently. I will show it to you when we meet next, to try and work it out. I will update this research log if I start to better understand it. This is also something I would really like to model. I think it is pretty close to what we want. It doesn't seem like much has been done in the way of analyzing clustering like we are doing...

### Additional Resources for this portion

Berthelson and Moller(2003, 2004, 2006, 2008) (I don't know how to do the  $\phi$  letter in Moller's name. I do know it's there, though)