Non-parametric Bayes clustering による重力波検出器の非定常大型雑音の特徴づけ

間 野 修 平，端 山 和 大，KAGRA Collaborations

統計数理研究所

日本物理学会年会第 70 回年次大会

2015 年 3 月 22 日

@ 早稲田大学

Clustering of Glitches

Classification of non-stationary large noises (broadband noise, known as glitches, triggers) is useful for diagnosis of interferometer.

An un-supervised learning problem → clustering.

Recognized by Mukherjee (2006) Class. Quant. Grav.:

- Input from kleine Welle: duration, central scale, SNR, and 15 highest wavelet coefficients → each glitch is a 18-dimensional vector.
- Dissimilarity matrix (probably) by Euclidean distance.
- Visualization by multidimensional scaling.
- Hierarchical clustering (UPGMA?).

Clustering of Glitches (cont.)

Improvement by Mukherjee, et al. (2012) Phys. Rev. D:

- For the waveform, distances based on the length of the longest common sub-sequence (local alignment by dynamic programming).
- Non-hierarchical clustering (k-means). The number of clusters, $\boldsymbol{K}$, is chosen such that the minimizer of intra/inter distances.
- Within a cluster, hierarchical clustering by central frequency, amplitude, SNR, and quality score.

Our proposal:

- Clustering by power spectral density (PSD), which is a functional data.
- Reformulate the problem in terms of Bayesian non-parametric clustering (density estimate, Lo, 1984), which captures probabilistic nature of the problem.

Bayesian Paradigm: the beta-binomial model

Suppose s successes in $Bernoulli(p)$ trial of the length n. The binomial likelihood $Binom(n, \theta)$ with the beta prior $Beta(\theta, \theta)$ yields posterior $Beta(s+\theta, n-s+\theta)$, because

$$\pi(p|s) \propto \mathbb{P}(s|p)\pi_\theta(p) \propto p^s(1-p)^{n-s}p^{\theta-1}(1-p)^{\theta-1}.$$

The beta prior is conjugate. A Bayesian estimate of the parameter is

$$\hat{p}_{Bayes} = \mathbb{E}(p|s) = \frac{n}{n+2\theta}\hat{p}_{MLE} + \frac{2\theta}{n+2\theta}\frac{1}{2}, \qquad \hat{p}_{MLE} = \frac{s}{n}.$$

Thus θ, the hyperparameter, is a strength of the prior belief. The marginal likelihood is

$$\mathbb{P}_\theta(s) = \int_0^1 \mathbb{P}(s|p)\pi_\theta(p)dp = \binom{n}{s}\frac{(\theta)_s(\theta)_{n-s}}{(2\theta)_n}.$$

Frequently used empirical estimate of the hyperparameter is the maximizer of the marginal likelihood.

Bayesian Clustering

Suppose sampling from a 2-components (cluster) mixture model. The likelihood is

$$\pi(x_1, ..., x_n|p) = \prod_{i=1}^{n} \{p\pi(x_i|c_i = 1) + (1 - p)\pi(x_i|c_i = 2)\}$$

and set the beta prior for p. In Bayesian clustering, our target is the joint posterior of memberships $\mathbb{P}(c_1, ..., c_n|x_1, ..., x_n)$. For the Gibbs sampler, we need $\mathbb{P}(c_i|c_{-i}, x_1, ..., x_n)$. By Bayes' rule,

$$\mathbb{P}(c_i|c_{-i}, x_1, ..., x_n) \propto \pi(x_i|c_1, ..., c_n, x_{-i})\mathbb{P}(c_i|c_{-i}, x_{-i}),$$

where $\pi(x_i|c, x_{-i}) = \pi(x_i|c_i)$ and $\mathbb{P}(c_i|c_{-i}, x_{-i}) = \mathbb{P}(c_i|c_{-i})$. Moreover, noting exchangeability in sampling, we have

$$\mathbb{P}(c_i = 1||\{j : c_{j\neq i} = 1\}| = s) = \frac{\mathbb{P}_n(s+1)}{\mathbb{P}_{n-1}(s)}\frac{s+1}{n} = \frac{\theta + s}{2\theta + n - 1}.$$

Bayesian Non-parametric Clustering

Extension to k-components is straightforward, the Dirichlet-multinomial model, where p is sampled from the k-dimensional Dirichlet distribution. But we have interest in the number of clusters.

Sampling from an “infinite-dimensional” Dirichlet distribution is possible by replacing the Dirichlet distribution by the Dirichlet process (random measure). This is non-parametric, since we do not use a fixed prior. The Dirichlet process is a conjugate prior.

Definition (Ferguson, 1973)

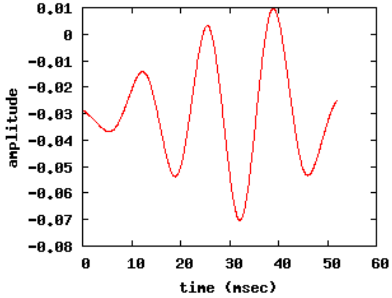
Let μ be a finite measure on $(\mathcal{X}, \mathcal{B})$. A random measure D on $\mathcal{X}$ is called a Dirichlet process if for every finite measureable partition $\{B_1, ..., B_k\}$ of $\mathcal{X}$, $(D(B_1), ..., D(B_k)) \sim Dirichlet(\mu(B_1), ..., \mu(B_k))$.

Bayesian Non-parametric Clustering (cont.)

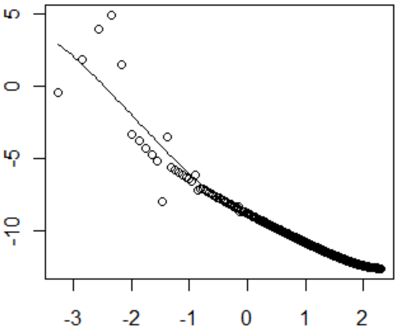
- For the Dirichlet process, $\mathbb{P}(c_i|c_{-i})$, which is needed in the Gibbs sampler, is given by the Chinese restaurant process (CRP):
  - Create a new cluster with probability $\frac{\theta}{\theta + n - 1}$;
  - Member of the i-th cluster with probability $\frac{n_i}{\theta + n - 1}$.
- We set $\pi(x_i|c_i) \sim MVN(\mu_i, \sigma)$, where μ_i are sampled by the Metropolis-Hastings algorithm.
- Empirical estimate of θ are chosen as the maximizer of the marginal likelihood with the MCMC estimate

$$\pi_\theta(x) = \sum_c \pi(x|c)\mathbb{P}_\theta(c) \hat{=} \frac{1}{M}\sum_{i=1}^{M} \pi(x|c^{(i)}), \qquad c^{(i)} \sim CRP(\theta).$$

Preprocessing

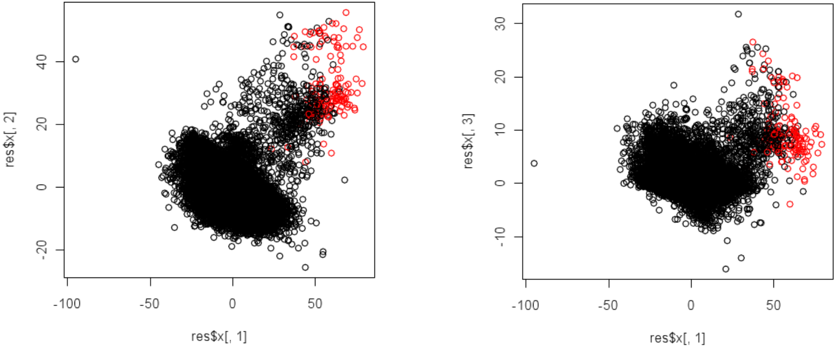


26,139 records by TAMA300



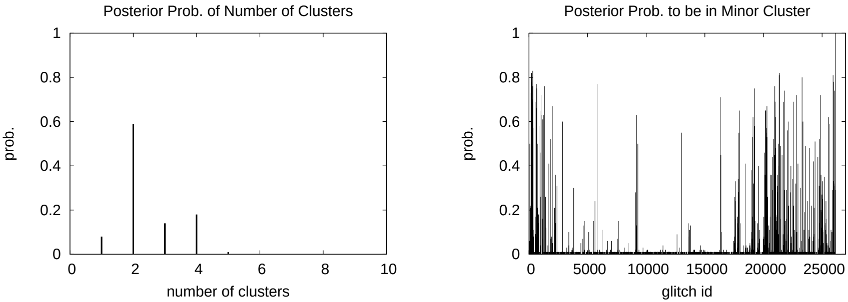
PSD → B-spline regression coefficients (10 basis) → PCA (94.8% variance explained by only the 2 PCs).

Visualization



Glitches are projected onto PCs. This is a snapshot. A clustering is a realization of a random event, in contrast to the k-means. Empirical estimate $\hat{\theta}$ =0.007.

Number of Clusters and Memberships



The number of clusters and memberships are random variables. The posterior probability of membership is given for each cluster for each glitch.

Summary

- Classification of broadband noises are useful for diagnosis of interferometer.
- Regarding PSD of glitches as a functional data a non-parametric Bayesian clustering was proposed.
- The method provides posterior distributions of the number of clusters and memberships for each cluster for each glitch.
- The posterior probability gives magnitude of confidence of the assignment. In fact, our result shows that irregular glitches accumulate in some specific period and then the lock will be lost.

Acknowledgment

We would like to thank TAMA collaboration for providing us the dataset.