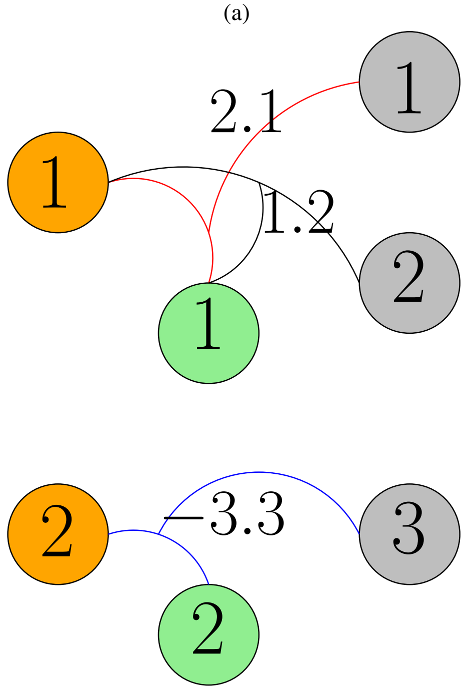
Supplementary Materials

Mode 1	Mode 2	Mode 3	Value
1	1	1	2.1
1	1	2	1.2
2	2	3	-3.3

(a)



(b)

Figure 1: Hypergraph representation of a sparse $2 \times 2 \times 3$ tensor. Nodes in different codes represent different modes. Each (hyper-)edge represents an existent entry, where the edge weight is the entry value.

1 Sparse Tensor Models

1.1 Completely Random Measures and Gamma Processes

A completely random measure (Kingman, 1967, 1992; Lijoi et al., 2010) μ on a $\mathbf{R}_+^d$ is a random variable that takes values in the space of measures on $\mathbf{R}_+^d$ such that for any collection of disjoint subsets $A_1, \ldots, A_n \subset \mathbf{R}^d$, the random variables $\mu(A_1), \ldots, \mu(A_n)$ are independent. This independence condition has the implication that CRMs are discrete measures. That is,

$$\mu = \sum_{i=1}^{\infty} w_i \delta_{\boldsymbol{\theta}_i}. \tag{1}$$

The theory of CRMs is intimately connected to Poisson Point Processes (PPP). We can characterize CRMs by the mean measure of a PPP. If $(w_i, \boldsymbol{\theta}_i) \in (\mathbf{R}_+, \mathbf{R}_+^d)$ has the distribution of a Poisson Point Process with intensity (mean) measure $\nu(dw d\boldsymbol{\theta})$, then the resulting discrete measure is a CRM. If we assume that the weights are independent of the locations in the CRM, the measure ν can be decomposed as $\nu(dw d\boldsymbol{\theta}) = \rho(w)\mu_0(d\boldsymbol{\theta})$.

A Gamma process (Hougaard, 1986; Brix, 1999) with the base measure μ_0, denoted by $\Gamma\text{P}(\mu_0)$, is the CRM that arises when

$$\nu(dw d\theta) = w^{-1} e^{-w} dw \mu_0(d\boldsymbol{\theta}).$$

Since

$$\int w^{-1} e^{-w} dw = \infty$$

for any measurable subset $\Theta \subset \mathbf{R}^d$ with $\mu_0(\Theta) > 0$, the ΓP will have an infinite number of atoms (locations). This is why in our sparse tensor process where we set $\mu_0 = \lambda_\alpha$, the Lebesgue measure with support restricted to $[0, \alpha]^d$, we still generate an infinite number of nodes in each mode (see (2) in the main paper). However when the PPP with the product of ΓPs as the mean measure is sampled to generate tensor entries, only a finite number of those nodes in each mode become active, because the the number of entries is finite (with probability one); see Sec. 3.1 of the main paper for more details.

Now suppose $g \sim \Gamma\text{P}(\mu_0)$, then it can be shown $g(\Theta)$ follows a Gamma distribution with the shape parameter $\mu_0(\theta)$ for any measureable $\Theta \subset \mathbf{R}_+^d$. This implies that if μ_0 is a finite measure, then $g(\mathbf{R}_+^d)$ is finite almost surely and $g/g(\mathbf{R}_+^d)$ is a well defined probability measure. Furthermore,

$$g/g(\mathbf{R}_+^d) \sim \text{DP}(\mu_0(\mathbf{R}_+^d), \mu_0/\mu_0(\mathbf{R}_+^d))$$

where DP is a Dirichlet process with the strength $\mu_0(\mathbf{R}_+^d)$ and base probability measure $\mu_0/\mu_0(\mathbf{R}_+^d)$.

1.2 Sparsity

Now we will prove Lemma 3.1 and Corollary 3.1.1. Our sparse tensor process is summarized as

$$\begin{aligned} W_k^\alpha &\sim \Gamma\text{P}(\lambda_\alpha)(1 \leq k \leq K), \\ T &\sim \text{PPP}(W_1^\alpha \times \cdots \times W_K^\alpha). \end{aligned} \tag{2}$$

We will first list a few lemmas the will be important to finish the proof.

Lemma 1.1 (Campbell's Theorem (Kingman, 1992))**.** *Let* Π *be a Poisson Process on* S *with mean measure* ν *and suppose* $f : S \to \mathbf{R}$ *is a measureable function, then*

$$\mathbb{E}\left[\sum_{x\in\Pi} f(x)\right] = \int_S f(x)\nu(dx).$$

Lemma 1.2 ((Caron and Fox, 2014) Lemma 17)**.** *Let* μ *be a random almost surely positive measure on* $\mathbf{R}^+$ *and let*

$$N|\mu \sim \textit{PoissonPoint}(\mu).$$

Define $\hat{N}_t = N[0,t]$ *and* $\hat{\mu}_t = \mu([0,t])$ *then*

$$\hat{N}_t|\mu \sim \textit{Poisson}(\hat{\mu}_t).$$

Furthermore if $\hat{\mu}_t \to \infty$ *and* $\lim_{t\to\infty} \frac{\hat{\mu}_{t+1}}{\hat{\mu}_t} = 1$*, then*

$$\frac{\hat{N}_t}{\hat{\mu}_t} \to 1 \;\; a.s.$$

Lemma 1.3 (Poisson Superposition Theorem (Cinlar and Agnew, 1968))**.** *Suppose* Π_1 *and* Π_2 *are Poisson point process on* S *with mean measure* μ_1 *and* μ_2 *respectively. Then* $\Pi_1 + \Pi_2$ *is a Poisson point process on* S *with mean measure* $\mu = \mu_1 + \mu_2$

Lemma 1.4 (Marking Theorem (Kingman, 1993))**.** *Let* Π *be a Poisson process on* S *with mean measure* μ*. Suppose for each* $X \in \Pi$ *we associate a mark* $m_X \in M$ *from a distribution* $p_x(\dot{})$*, that may depend on* X *but not other points. Then the cartesian product* $\{(X, m_X)|X \in \Pi\}$ *is a Poisson process on* $S \times M$ *with mean measure* $\mu(dx)p_x(dm)$*.*

1.2.1 Proof of Lemma 3.1 and Corollary 3.1.1

We will prove Lemma 3.1 in two steps. For simplicity we will assume λ_α is the Lebesgue measure on $[0,\alpha]$ and λ is the Lebesgue measure on $[0,\infty]$.The extension to the Lebesgue measure on $[0,\alpha]^d$ is straightforward.

It follows from the properties of the ΓP that if $W^\infty \sim \Gamma\text{P}(\lambda)$ and if $W^\alpha \sim \Gamma\text{P}(\lambda_\alpha)$ then the distribution of the measure W^∞ restricted to $[0,\alpha]$ is identical to W^α. Thus instead of generating a new CRM for W^α each time with α increased, we assume the same CRM, W^∞ is restricted to the growing set $[0,\alpha]$.

Let M_k^α be the number of active nodes in mode k and let N^α be the number of entries. Let

$$A^\alpha_{k,\theta_i^k} = [0,\alpha] \times \cdots \times \{\theta_i^k\} \times \cdots \times [0,\alpha].$$

Then we have

$$M_k^\alpha = \#\{\theta_i^k \in [0,\alpha] | T(A^\alpha_{k,\theta_i^k}) > 0\}.$$

In the first step, we will show $\lim_{\alpha\to\infty} \frac{\alpha}{M_k^\alpha} = 0 \;\; a.s.$ for all $k \in \{1,\dots,K\}$. Then in the second step , we will show that $\limsup_{\alpha\to\infty} N^\alpha/\alpha^K < \infty \;\; a.s.$ Together this implies

$$\lim_{\alpha\to\infty} \frac{N^\alpha}{\prod_{k=1}^K M_k^\alpha} = 0 \;\; a.s$$

because

$$\frac{N^\alpha}{\prod_{k=1}^K M_k^\alpha} = \frac{N^\alpha}{\alpha^K}\prod_{k=1}^K \frac{\alpha}{M_k^\alpha}.$$

Step 1. First note that $T(A^\alpha_{k,\theta_i^k})|\{W_k^\infty\}_{k=1}^K$ has a Poisson distribution so

$$\Pr(T(A^\alpha_{k,\theta_i^k}) > 0|\{W_k^\infty\}_{k=1}^K)) = 1 - \exp\left(-W_k^\infty(\{\theta_i^k\}) \times \prod_{j\neq k} W_j^\infty([0,\alpha])\right).$$

Additionally, the set of points $\{T(A^\alpha_{k,\theta_i^k}) > 0\}_i$ can be interpreted as random binary marks on the Gamma process W_k^∞ when conditioned on $\{W_j^\infty\}_{j\neq k}$. Hence, according to the Poisson marking theorem (Lemma 1.4), the marked Gamma process $\{(\theta_i^k, T(A^\alpha_{k,\theta_i^k}) > 0)\}$ conditioned on $\{W_j^\infty\}_{j\neq k}$ is generated by a Poisson point process on $\mathbb{R}_+ \times \mathbb{R}_+ \times \{0,1\}$. Thus $M_k^\alpha|\{W_i^\infty([0,\alpha])\}_{i\neq k}$ is a Poisson random variable. We compute the expectation of M_k^α given the ΓPs of the other modes to characterize the distribution of $M_k^\alpha|\{W_i^\infty\}_{i\neq k}$. Using the law of total expectation, we have

$$\begin{aligned}
\mathbb{E}[M_k^\alpha|\{W_j^\infty\}_{j\neq k}] &= \mathbb{E}\left[\sum_{\theta_i\in[0,\alpha]} \mathbb{1}(T(A^\alpha_{k,\theta_i^k}) > 0)\middle|\{W_j^\infty\}_{j\neq k}\right] \\
&= \mathbb{E}\left[\sum_{\theta_i\in[0,\alpha]} \mathbb{E}[\mathbb{1}(T(A^\alpha_{k,\theta_i^k}) > 0)|\{W_k^\infty\}_{k=1}^K]\middle|\{W_j^\infty\}_{j\neq k}\right] \\
&= \mathbb{E}\left[\sum_{\theta_i\in[0,\alpha]} 1 - \exp\left(-W_k^\infty(\{\theta_i^k\} \times \prod_{j\neq k} W_j^\infty([0,\alpha])\right)\middle|\{W_j^\infty\}_{j\neq k}\right].
\end{aligned}$$

For the expectation, because (θ_i^k, w_i^k) is a Poisson process due to the construction of the CRM, we can apply Lemma 1.1. Together this gives

$$\begin{aligned}
&\mathbb{E}[M_k^\alpha|\{W_j^\infty([0,\alpha])\}_{j\neq k}] \\
&= \int_0^\infty\int_0^\infty \left(1 - \exp\left(-w \times \prod_{j\neq k} W_j^\infty([0,\alpha])\right)\right) w^{-1}e^{-w} dw d\lambda_\alpha \\
&= \alpha\int_0^\infty \left(1 - \exp\left(-w \times \prod_{i\neq k} W_i^\infty([0,\alpha])\right)\right) w^{-1}e^{-w} dw.
\end{aligned}$$

Let

$$\psi(t) = \int_0^\infty (1 - \exp(-wt))w^{-1}e^{-w}dw,$$

then our work shows

$$M_k^\alpha|\{W_j^\infty\}_{j\neq k} \sim \text{Poisson}\left(\alpha\cdot\psi\left(\prod_{j\neq k} W_j^\infty([0,\alpha])\right)\right).$$

As $W_j^\infty([0,\alpha])$ is Gamma distributed with shape parameter α, $\lim_{\alpha\to\infty} W_j^\alpha([0,\alpha]) = \infty\, a.s.$ We also have $\lim_{t\to\infty}\psi(t)=\infty$. This follows immediately from the monotone convergence theorem as $\int_0^\infty w^{-1}e^{-w}dw=\infty$. Together this implies

$$\lim_{\alpha\to\infty}\frac{\alpha\psi\left(\prod_{j\neq k} W_j^\infty([0,\alpha])\right)}{\alpha} = \infty\, a.s. \tag{3}$$

Applying Lemma 1.2 the Poisson process with mean measure, τ where $\tau([a,b]) = b\psi\left(\prod_{i\neq k} W_i^\infty([0,b])\right) - a\psi\left(\prod_{i\neq k} W_i^\infty([0,a])\right)$ then implies

$$Pr\left(\lim_{\alpha\to\infty}\frac{M_k^\alpha}{\alpha\cdot\psi(\prod_{i\neq k} W_i^\alpha([0,\alpha]))} = 1\middle|\{W_i^\infty\}_{i\neq k}\right) = 1.$$

Taking the expectation on both sides of the above expression implies

$$\lim_{\alpha\to\infty}\frac{M_k^\alpha}{\alpha\cdot\psi(\prod_{i\neq k} W_i^\alpha([0,\alpha]))} = 1\, a.s.$$

Combining the above with with equation (3) completes the first step and implies

$$\lim_{\alpha\to\infty}\frac{\alpha}{M_k^\alpha} = 0\ \ a.s.$$

Step 2. As it is possible for the point process to sample more than one point at a single location, the number of points generated from the point process may not equal to the number of (distinct) tensor entries. Let D^α be the actual number of points sampled. Note $N^\alpha < D^\alpha$.

Now consider $j\in\mathbb{N}$ and $D^j = T([0,j]^K)$. We have

$$D^j|\{W_1^\infty,\dots W_K^\infty\}\sim Poisson\left(\prod_{k=1}^K W_k^\infty([0,j])\right).$$

By the independence of the CRM on disjoint sets, it follows immediately by the strong law of large numbers

$$\lim_{j\to\infty}\frac{W_k^\infty([0,j])}{j} = \frac{\sum_{i=1}^j W_k^\infty((i-1,i])}{j} = E[W_k^\infty([0,1])] = 1\, a.s.$$

as $W_k^\infty((i-1,i])$ are $i.i.d$ Gamma random variables. This implies

$$\lim_{j\to\infty}\frac{\prod_{k=1}^K W_k^\infty([0,j])}{j^K} = 1\, a.s. \tag{4}$$

But applying Lemma 1.2 implies

$$Pr\left(\lim_{j\to\infty}\frac{D^j}{\prod_{k=1}^K W_k^j([0,j])} = 1\middle|\{W_i^\infty\}_{i=1}^K\right) = 1$$

Taking the expectation of both sides of the above expression and combining with equation (4) implies

$$\lim_{j\to\infty}\frac{D^j}{j^K} = 1\,a.s.$$

The above only holds for natural numbers. To extend to real numbers note for any α, there exists, $j \in \mathbb{N}$ such that $j \leq \alpha \leq j+1$. Thus

$$\frac{j^K}{(j+1)^K}\frac{D^j}{j^K} \leq \frac{D^\alpha}{\alpha^k} \leq \frac{(j+1)^K}{j^K}\frac{D^{j+1}}{(j+1)^K},$$

so taking $\alpha \to \infty$ proves

$$\lim_{\alpha\to\infty}\frac{D^\alpha}{\alpha^K} = 1.$$

Recalling $N^\alpha \leq D^\alpha$ completes the proof.

Proof of Corollary 3.1.1 By the Lemma 1.3 (Poisson superposition theorem)

$$T \sim \text{PPP}(\sum_{r=1}^{R} W_{1,r}^\alpha \times \cdots \times W_{K,r}^\alpha)$$

can be constructed as

$$T = \sum_{r=1}^{R}\text{PPP}(W_{1,r}^\alpha \times \cdots \times W_{K,r}^\alpha).$$

Now lemma 3.1 applies to each of the individual Poisson processes which implies the result.

References

Brix, A. (1999). Generalized gamma measures and shot-noise cox processes. *Advances in Applied Probability*, pages 929–953.

Caron, F. and Fox, E. B. (2014). Sparse graphs using exchangeable random measures. *arXiv preprint arXiv:1401.1137*.

Cinlar, E. and Agnew, R. (1968). On the superposition of point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 576–581.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396.

Kingman, J. (1967). Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78.

Kingman, J. (1992). *Poisson Processes*, volume 3. Clarendon Press.

Kingman, J. F. C. (1993). *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York. Oxford Science Publications.

Lijoi, A., Prünster, I., et al. (2010). Models beyond the Dirichlet process. *Bayesian nonparametrics*, 28(80):342.