Tensors in Algebraic Statistics

Luis Sierra Muntané
luis.sierra@mail.utoronto.ca

Department of Statistical Sciences
University of Toronto

Ottawa Mathematics and Statistics Conference 2025

Joint work with Marta Casanellas and Piotr Zwiernik.

Measures as tensors

Measures as tensors

Tensors are multilinear maps that can be identified with m-way arrays. A *pure tensor* $e^1_{i_1} \otimes \cdots \otimes e^m_{i_m}$ acts as

$$\begin{aligned} \mathbb{R}^{r_1} \times \cdots \times \mathbb{R}^{r_m} &\longrightarrow \mathbb{R} \\ (e^1_{j_1}, \ldots, e^m_{j_m}) &\mapsto \begin{cases} 1 & \text{if } (j_1, \ldots, j_m) = (i_1, \ldots, i_m), \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

Measures as tensors

Tensors are multilinear maps that can be identified with m-way arrays. A *pure tensor* $e^1_{i_1} \otimes \cdots \otimes e^m_{i_m}$ acts as

$$\begin{aligned} \mathbb{R}^{r_1} \times \cdots \times \mathbb{R}^{r_m} &\longrightarrow \mathbb{R} \\ (e^1_{j_1}, \ldots, e^m_{j_m}) &\mapsto \begin{cases} 1 & \text{if } (j_1, \ldots, j_m) = (i_1, \ldots, i_m), \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

For discrete r.v.s $X_i \in \mathcal{X}_i = \{0, \ldots, r_i - 1\}$, $i = 1, \ldots, m$
The p.m.f. $p = (p(x))_{x \in \mathcal{X}}$ of X can be then identified with a point

$$p(x) = \mathbb{P}(X_1 = x_1, \ldots, X_m = x_m) \in \mathbb{R}^{\mathcal{X}} := \bigotimes_{i=1}^{m} \mathbb{R}^{\mathcal{X}_i},$$

Measures as tensors

Tensors are multilinear maps that can be identified with m-way arrays. A *pure tensor* $e^1_{i_1} \otimes \cdots \otimes e^m_{i_m}$ acts as

$$\begin{aligned} \mathbb{R}^{r_1} \times \cdots \times \mathbb{R}^{r_m} &\longrightarrow \mathbb{R} \\ (e^1_{j_1}, \ldots, e^m_{j_m}) &\mapsto \begin{cases} 1 & \text{if } (j_1, \ldots, j_m) = (i_1, \ldots, i_m), \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

For discrete r.v.s $X_i \in \mathcal{X}_i = \{0, \ldots, r_i - 1\}$, $i = 1, \ldots, m$
The p.m.f. $p = (p(x))_{x \in \mathcal{X}}$ of X can be then identified with a point

$$p(x) = \mathbb{P}(X_1 = x_1, \ldots, X_m = x_m) \in \mathbb{R}^{\mathcal{X}} := \bigotimes_{i=1}^{m} \mathbb{R}^{\mathcal{X}_i},$$

whose projectivized space can be identified with the $m-1$ dimensional simplex

$$\Delta^{\mathcal{X}} \quad := \quad \{p \in \mathbb{R}^{\mathcal{X}} : \; p_x \geq 0, \sum_{x \in \mathcal{X}} p_x = 1\}.$$

Independence model

Independence model

The discrete random vector $X = (X_1, \ldots, X_m) \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ satisfies the *full independence model* if $p \in \Delta^{\mathcal{X}}$ can be written for all $x = (i_1, \ldots, i_m) \in \mathcal{X}$ as a (pure) tensor product

Independence model

The discrete random vector $X = (X_1, \ldots, X_m) \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ satisfies the *full independence model* if $p \in \Delta^{\mathcal{X}}$ can be written for all $x = (i_1, \ldots, i_m) \in \mathcal{X}$ as a (pure) tensor product

$$p_{i_1 i_2 \cdots i_m} = p^1_{i_1} p^2_{i_2} \cdots p^m_{i_m}$$

Independence model

The discrete random vector $X = (X_1, \dots, X_m) \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ satisfies the *full independence model* if $p \in \Delta^{\mathcal{X}}$ can be written for all $x = (i_1, \dots, i_m) \in \mathcal{X}$ as a (pure) tensor product

$$p_{i_1 i_2 \cdots i_m} = p^1_{i_1} p^2_{i_2} \cdots p^m_{i_m} = p^1 \otimes \cdots \otimes p^m.$$

Independence model

The discrete random vector $X = (X_1, \ldots, X_m) \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ satisfies the *full independence model* if $p \in \Delta^{\mathcal{X}}$ can be written for all $x = (i_1, \ldots, i_m) \in \mathcal{X}$ as a (pure) tensor product

$$p_{i_1 i_2 \cdots i_m} = p^1_{i_1} p^2_{i_2} \cdots p^m_{i_m} = p^1 \otimes \cdots \otimes p^m.$$

$$\operatorname{Seg}\left(\mathbb{P}_{\mathbb{R}}^{\mathcal{X}_1} \times \cdots \times \mathbb{P}_{\mathbb{R}}^{\mathcal{X}_m}\right) \subset \mathbb{P}_{\mathbb{R}}^{\mathcal{X}} = \operatorname{Proj}\left(\mathbb{R}^{\mathcal{X}_1} \otimes \cdots \otimes \mathbb{R}^{\mathcal{X}_m}\right),$$

which correspond to semialgebraic constraints.

Independence model

The discrete random vector $X = (X_1, \ldots, X_m) \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ satisfies the *full independence model* if $p \in \Delta^{\mathcal{X}}$ can be written for all $x = (i_1, \ldots, i_m) \in \mathcal{X}$ as a (pure) tensor product

$$p_{i_1 i_2 \cdots i_m} = p^1_{i_1} p^2_{i_2} \cdots p^m_{i_m} = p^1 \otimes \cdots \otimes p^m.$$

$$\mathrm{Seg}\left(\mathbb{P}_{\mathbb{R}}^{\mathcal{X}_1} \times \cdots \times \mathbb{P}_{\mathbb{R}}^{\mathcal{X}_m}\right) \subset \mathbb{P}_{\mathbb{R}}^{\mathcal{X}} = \mathrm{Proj}\left(\mathbb{R}^{\mathcal{X}_1} \otimes \cdots \otimes \mathbb{R}^{\mathcal{X}_m}\right),$$

which correspond to semialgebraic constraints. As an example with two variables X_1, X_2:

Independence model

The discrete random vector $X = (X_1, \ldots, X_m) \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ satisfies the *full independence model* if $p \in \Delta^{\mathcal{X}}$ can be written for all $x = (i_1, \ldots, i_m) \in \mathcal{X}$ as a (pure) tensor product

$$p_{i_1 i_2 \cdots i_m} = p^1_{i_1} p^2_{i_2} \cdots p^m_{i_m} = p^1 \otimes \cdots \otimes p^m.$$

$$\mathrm{Seg}\left(\mathbb{P}^{\mathcal{X}_1}_{\mathbb{R}} \times \cdots \times \mathbb{P}^{\mathcal{X}_m}_{\mathbb{R}}\right) \subset \mathbb{P}^{\mathcal{X}}_{\mathbb{R}} = \mathrm{Proj}\left(\mathbb{R}^{\mathcal{X}_1} \otimes \cdots \otimes \mathbb{R}^{\mathcal{X}_m}\right),$$

which correspond to semialgebraic constraints. As an example with two variables X_1, X_2:

$$p = p^1 \otimes p^2 = \begin{pmatrix} p^1_0 \\ \vdots \\ p^1_{r_1-1} \end{pmatrix} \begin{pmatrix} p^2_0 & \cdots & p^2_{r_2-1} \end{pmatrix}$$

Moment Tensors

For $u = (u_1, \ldots, u_m) \in \mathbb{N}^m$ and a vector $x = (x_1, \ldots, x_m) \in \mathcal{X}$ define the monomial

Moment Tensors

For $u = (u_1, \ldots, u_m) \in \mathbb{N}^m$ and a vector $x = (x_1, \ldots, x_m) \in \mathcal{X}$ define the monomial

$$\mu_u = \mathbb{E}X^u = \sum_{x \in \mathcal{X}} p_x x^u, \quad x^u = x_1^{u_1} \cdots x_m^{u_m},$$

which is a *moment of order* k if $k = u_1 + \cdots + u_m$. Setting $u = (0, \ldots, 0)$ we recover the p.m.f. condition $\mu_{0 \ldots 0} = 1$.

Moment Tensors

For $u = (u_1, \dots, u_m) \in \mathbb{N}^m$ and a vector $x = (x_1, \dots, x_m) \in \mathcal{X}$ define the monomial

$$\mu_u = \mathbb{E}X^u = \sum_{x \in \mathcal{X}} p_x x^u, \quad x^u = x_1^{u_1} \cdots x_m^{u_m},$$

which is a *moment of order* k if $k = u_1 + \cdots + u_m$. Setting $u = (0, \dots, 0)$ we recover the p.m.f. condition $\mu_{0\dots0} = 1$.

In the binary case $(X_1, X_2) \in \{0,1\}^2$. We have $\mu_{00} = p_{00} + p_{01} + p_{10} + p_{11} = 1$, $\mu_{10} = p_{10} + p_{11}$, $\mu_{01} = p_{01} + p_{11}$, $\mu_{11} = p_{11}$.

Moment Tensors

For $u = (u_1, \ldots, u_m) \in \mathbb{N}^m$ and a vector $x = (x_1, \ldots, x_m) \in \mathcal{X}$ define the monomial

$$\mu_u = \mathbb{E}X^u = \sum_{x \in \mathcal{X}} p_x x^u, \quad x^u = x_1^{u_1} \cdots x_m^{u_m},$$

which is a *moment of order* k if $k = u_1 + \cdots + u_m$. Setting $u = (0, \ldots, 0)$ we recover the p.m.f. condition $\mu_{0\ldots0} = 1$.

In the binary case $(X_1, X_2) \in \{0,1\}^2$. We have
$\mu_{00} = p_{00} + p_{01} + p_{10} + p_{11} = 1$, $\mu_{10} = p_{10} + p_{11}$, $\mu_{01} = p_{01} + p_{11}$, $\mu_{11} = p_{11}$.

The corresponding cumulants are given by the transformation $\kappa_{00} = 0$, $\kappa_{10} = \mu_{10}$, $\kappa_{01} = \mu_{01}$, and $\kappa_{11} = \mu_{11} - \mu_{10}\mu_{01}$.

Moment Tensors

For $u = (u_1, \ldots, u_m) \in \mathbb{N}^m$ and a vector $x = (x_1, \ldots, x_m) \in \mathcal{X}$ define the monomial

$$\mu_u = \mathbb{E}X^u = \sum_{x \in \mathcal{X}} p_x x^u, \quad x^u = x_1^{u_1} \cdots x_m^{u_m},$$

which is a *moment of order* k if $k = u_1 + \cdots + u_m$. Setting $u = (0, \ldots, 0)$ we recover the p.m.f. condition $\mu_{0\ldots 0} = 1$.

In the binary case $(X_1, X_2) \in \{0,1\}^2$. We have
$\mu_{00} = p_{00} + p_{01} + p_{10} + p_{11} = 1$, $\mu_{10} = p_{10} + p_{11}$, $\mu_{01} = p_{01} + p_{11}$, $\mu_{11} = p_{11}$.

The corresponding cumulants are given by the transformation $\kappa_{00} = 0$, $\kappa_{10} = \mu_{10}$, $\kappa_{01} = \mu_{01}$, and $\kappa_{11} = \mu_{11} - \mu_{10}\mu_{01}$.

$\implies$ Independence model is given by $\kappa_{11} = 0$. For the study of binary cumulant varieties see (Sturmfels and Zwiernik, 2013).

Remark (Absolutely continuous measures)

The absolutely continuous case is also possible, where the moment tensors are *symmetric* (McCullagh, 2018):

$$\mu_{i_1 \dots i_r}(X) = \mathbb{E} X_{i_1} \cdots X_{i_r} = \frac{\partial^r}{\partial s_{i_1} \cdots \partial s_{i_r}} M_X(\mathbf{s}) \Big|_{\mathbf{s}=0} \,,$$

$$\kappa_{i_1 \dots i_r}(X) \;=\; \mathrm{cum}(X_{i_1}, \dots, X_{i_r}) \;=\; \frac{\partial^r}{\partial s_{i_1} \cdots \partial s_{i_r}} \log M_X(\mathbf{s}) \Big|_{\mathbf{s}=0} \,.$$

Remark (Absolutely continuous measures)

The absolutely continuous case is also possible, where the moment tensors are *symmetric* (McCullagh, 2018):

$$\mu_{i_1 \cdots i_r}(X) = \mathbb{E} X_{i_1} \cdots X_{i_r} = \frac{\partial^r}{\partial s_{i_1} \cdots \partial s_{i_r}} M_X(\mathbf{s}) \Big|_{\mathbf{s}=0},$$

$$\kappa_{i_1 \cdots i_r}(X) = \operatorname{cum}(X_{i_1}, \ldots, X_{i_r}) = \frac{\partial^r}{\partial s_{i_1} \cdots \partial s_{i_r}} \log M_X(\mathbf{s}) \Big|_{\mathbf{s}=0}.$$

$$X \text{ Gaussian} \iff \kappa_X(\mathbf{s}) = \mu^\top \mathbf{s} + \frac{1}{2} \mathbf{s}^\top \Sigma \mathbf{s}.$$

Remark (Absolutely continuous measures)

The absolutely continuous case is also possible, where the moment tensors are *symmetric* (McCullagh, 2018):

$$\mu_{i_1\cdots i_r}(X) = \mathbb{E}X_{i_1}\cdots X_{i_r} = \frac{\partial^r}{\partial s_{i_1}\cdots\partial s_{i_r}} M_X(\mathbf{s})\Big|_{\mathbf{s}=0},$$

$$\kappa_{i_1\cdots i_r}(X) \;=\; \operatorname{cum}(X_{i_1},\ldots,X_{i_r}) \;=\; \frac{\partial^r}{\partial s_{i_1}\cdots\partial s_{i_r}} \log M_X(\mathbf{s})\Big|_{\mathbf{s}=0}.$$

$$X \text{ Gaussian} \iff \kappa_X(\mathbf{s}) = \mu^\top\mathbf{s} + \frac{1}{2}\mathbf{s}^\top\Sigma\mathbf{s}.$$

Useful for multivariate Gaussian models $\mathcal{N}_m(\mu,\Sigma)$, where conditional independences correspond to linear restrictions $\left(\Sigma^{-1}\right)_{ij} = 0$ (Uhler, 2017).

Tensor rank

An $n_1 \times n_2$ matrix has rank $\leq k$ if

$$M = u^1 \otimes v^1 + \cdots + u^k \otimes v^k, \quad \text{for } u^i \in \mathbb{R}^{n_1},\, v^i \in \mathbb{R}^{n_2}.$$

Tensor rank

An $n_1 \times n_2$ matrix has rank $\leq k$ if

$$M = u^1 \otimes v^1 + \cdots + u^k \otimes v^k, \quad \text{for } u^i \in \mathbb{R}^{n_1},\, v^i \in \mathbb{R}^{n_2}.$$

An m-way tensor has rank $\leq k$ if

$$T = \sum_{j=1}^{k} u^{1,j} \otimes u^{2,j} \otimes \cdots \otimes u^{m,j}.$$

Tensor rank

An $n_1 \times n_2$ matrix has rank $\leq k$ if

$$M = u^1 \otimes v^1 + \cdots + u^k \otimes v^k, \quad \text{for } u^i \in \mathbb{R}^{n_1},\, v^i \in \mathbb{R}^{n_2}.$$

An m-way tensor has rank $\leq k$ if

$$T = \sum_{j=1}^{k} u^{1,j} \otimes u^{2,j} \otimes \cdots \otimes u^{m,j}.$$

We say the rank is *nonnegative* if $u^{i,j} \in \mathbb{R}^{r_i}_{\geq 0}$; $\mathrm{rank}(T) \leq \mathrm{rank}_+(T)$.

Tensor rank

An $n_1 \times n_2$ matrix has rank $\leq k$ if

$$M = u^1 \otimes v^1 + \cdots + u^k \otimes v^k, \quad \text{for } u^i \in \mathbb{R}^{n_1},\, v^i \in \mathbb{R}^{n_2}.$$

An m-way tensor has rank $\leq k$ if

$$T = \sum_{j=1}^{k} u^{1,j} \otimes u^{2,j} \otimes \cdots \otimes u^{m,j}.$$

We say the rank is *nonnegative* if $u^{i,j} \in \mathbb{R}^{r_i}_{\geq 0}$; $\mathrm{rank}(T) \leq \mathrm{rank}_+(T)$.

Finding such a decomposition is `NP`-hard! (Hillar and Lim, 2013)

Naive Bayes

Consider the m-variate the dataset $\{(X_1, \ldots, X_m)\}$ with $X_i \in \mathcal{X}_i = [r_i]$, where each point belongs to the class $H \in \mathcal{C} = [k]$.

Naive Bayes

Consider the m-variate the dataset $\{(X_1, \ldots, X_m)\}$ with $X_i \in \mathcal{X}_i = [r_i]$, where each point belongs to the class $H \in \mathcal{C} = [k]$. Making the assumption

$$X_i \perp\!\!\!\perp X_j \,|\, H, \quad \forall i, j \in [m]$$

Naive Bayes

Consider the m-variate the dataset $\{(X_1, \ldots, X_m)\}$ with $X_i \in \mathcal{X}_i = [r_i]$, where each point belongs to the class $H \in \mathcal{C} = [k]$. Making the assumption

$$X_i \perp\!\!\!\perp X_j \mid H, \quad \forall i, j \in [m]$$

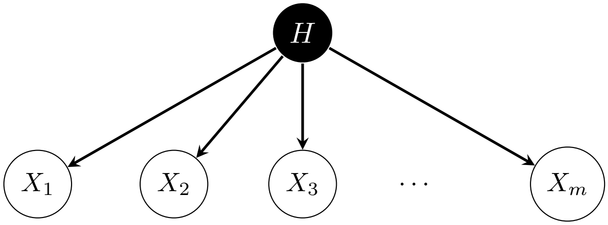
$$\mathbb{P}(H = h \mid X = x) \propto p(h) \prod_{i=1}^{m} \mathbb{P}(X_i = x_i \mid H = h)$$

Naive Bayes

Consider the m-variate the dataset $\{(X_1, \dots, X_m)\}$ with $X_i \in \mathcal{X}_i = [r_i]$, where each point belongs to the class $H \in \mathcal{C} = [k]$. Making the assumption

$$X_i \perp\!\!\!\perp X_j \,|\, H, \quad \forall i, j \in [m]$$

$$\mathbb{P}(H = h \,|\, X = x) \propto p(h) \prod_{i=1}^{m} \mathbb{P}(X_i = x_i \,|\, H = h)$$



Naive Bayes II

Since the X_i are discrete, we can encode the conditionals as

$$u^{i,h} = (\mathbb{P}(X_i = 0 \,|\, h), \dots, \mathbb{P}(X_i = r_i - 1 \,|\, h)) \in \mathbb{R}^{\mathcal{X}_i},$$

Naive Bayes II

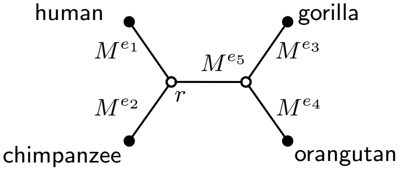
Since the X_i are discrete, we can encode the conditionals as

$$u^{i,h} = (\mathbb{P}(X_i = 0 \,|\, h), \ldots, \mathbb{P}(X_i = r_i - 1 \,|\, h)) \in \mathbb{R}^{\mathcal{X}_i},$$

and then the marginal over the observables is naturally given as a tensor

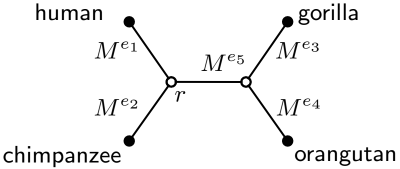
$$\mathbb{P}(X = x) = \sum_{h \in \mathcal{C}} p(h) \cdot u^{1,h} \otimes \cdots \otimes u^{m,h},$$

whose nonnegative rank is therefore no larger than k.



Consider species as vertices of a directed tree $\mathcal{T} = (V, E)$, $m = |L|$ leaves, $\deg(v) > 2$ for $u \in V \setminus L$, with nucleotides $X_u = \{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}$.

Phylogenetic Trees



Consider species as vertices of a directed tree $\mathcal{T} = (V, E)$, $m = |L|$ leaves, $\deg(v) > 2$ for $u \in V \setminus L$, with nucleotides $X_u = \{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}$.

$$\begin{array}{c}\mathtt{ACGTGACTGATCAGCTGACT} \\ \Big\downarrow M^e \\ \mathtt{ACGTTACTGATCAGCTGACT}\end{array}$$

$M^e_{x,y} = \mathbb{P}(X_v = y \,|\, X_u = x)$ are transition matrices for $e = \{u \to v\}$.

Phylogenetic Trees II

$$\mathbb{P}\left(\{x_v\}_{v\in V}\right) = \mathbb{P}(x_r) \prod_{e=\{u\to w\}\in E} M^e_{x_w,x_u},$$

since all interior nodes are latent, and marginalizing over them realizes a tensor

$$p_{\mathcal{T}} \in \mathbb{R}^{\mathcal{X}} = \mathbb{R}^4 \otimes \overset{m}{\cdots} \otimes \mathbb{R}^4,$$

Phylogenetic Trees II

$$\mathbb{P}\left(\{x_v\}_{v\in V}\right) = \mathbb{P}(x_r) \prod_{e=\{u\to w\}\in E} M^e_{x_w,x_u},$$

since all interior nodes are latent, and marginalizing over them realizes a tensor

$$p_{\mathcal{T}} \in \mathbb{R}^{\mathcal{X}} = \mathbb{R}^4 \otimes \overset{m}{\cdots} \otimes \mathbb{R}^4,$$

so that the leaf nucleotides follow the distribution

$$p_{i_1\ldots i_n} = \sum_{\{x_v\}_{v\in V}\,|\,x_j=\{\mathtt{A},\mathtt{C},\mathtt{G},\mathtt{T}\},j\in[m]} \mathbb{P}(\{x_v\}_{v\in V}).$$

Phylogenetic Trees II

$$\mathbb{P}\left(\{x_v\}_{v\in V}\right) = \mathbb{P}(x_r) \prod_{e=\{u\to w\}\in E} M^e_{x_w,x_u},$$

since all interior nodes are latent, and marginalizing over them realizes a tensor

$$p_{\mathcal{T}} \in \mathbb{R}^{\mathcal{X}} = \mathbb{R}^4 \otimes \overset{m}{\cdots} \otimes \mathbb{R}^4,$$

so that the leaf nucleotides follow the distribution

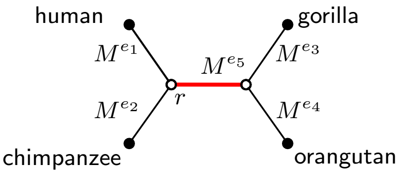
$$p_{i_1\dots i_n} = \sum_{\{x_v\}_{v\in V}\,|\,x_j=\{\mathtt{A},\mathtt{C},\mathtt{G},\mathtt{T}\},j\in[m]} \mathbb{P}(\{x_v\}_{v\in V}).$$

Now consider a split of the leaves $L = A \sqcup B$

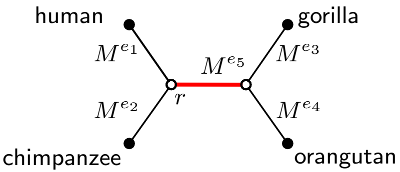
$$\begin{array}{rcl} \mathbb{R}^{\mathcal{X}} \cong \mathbb{R}^{\mathcal{X}_A} \otimes \mathbb{R}^{\mathcal{X}_B} & \longrightarrow & \mathcal{M}_{|\mathcal{X}_A|\times|\mathcal{X}_B|}(\mathbb{R}) \\ p = (p_{x_1\dots x_n}) & \mapsto & \mathrm{flatt}_{A|B}(p), \end{array}$$

where the matrix $\mathrm{flatt}_{A|B}(p)$ has entries $\mathbb{P}(X_A = x_A, X_B = x_B)$.

Phylogenetic Trees II



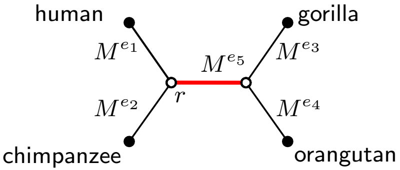
Phylogenetic Trees II



Denoting $L = \{\text{human, chimpanzee, gorilla, orangutan}\}$ by $\{1, 2, 3, 4\}$, the flattened tensor is given by

$$\text{flatt}_{12|34}(p) = \begin{pmatrix} p_{0000} & p_{0001} & p_{0002} & \cdots & p_{0033} \\ p_{0100} & p_{0101} & p_{0102} & \cdots & p_{0133} \\ p_{0200} & p_{0201} & p_{0202} & \cdots & p_{0233} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{3300} & p_{3301} & p_{3302} & \cdots & p_{3333} \end{pmatrix}.$$

Phylogenetic Trees II

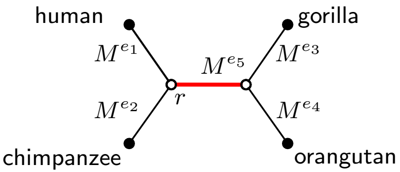


Denoting $L = \{\text{human, chimpanzee, gorilla, orangutan}\}$ by $\{1, 2, 3, 4\}$, the flattened tensor is given by

$$\mathrm{flatt}_{12|34}(p) = \begin{pmatrix} p_{0000} & p_{0001} & p_{0002} & \cdots & p_{0033} \\ p_{0100} & p_{0101} & p_{0102} & \cdots & p_{0133} \\ p_{0200} & p_{0201} & p_{0202} & \cdots & p_{0233} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{3300} & p_{3301} & p_{3302} & \cdots & p_{3333} \end{pmatrix}.$$

$\implies$ This is a Naive Bayes model with $m = 2$,

Phylogenetic Trees II



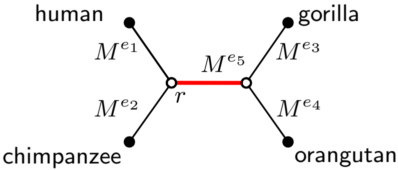
Denoting $L = \{\text{human, chimpanzee, gorilla, orangutan}\}$ by $\{1, 2, 3, 4\}$, the flattened tensor is given by

$$\text{flatt}_{12|34}(p) = \begin{pmatrix} p_{0000} & p_{0001} & p_{0002} & \cdots & p_{0033} \\ p_{0100} & p_{0101} & p_{0102} & \cdots & p_{0133} \\ p_{0200} & p_{0201} & p_{0202} & \cdots & p_{0233} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{3300} & p_{3301} & p_{3302} & \cdots & p_{3333} \end{pmatrix}.$$

$\implies$ This is a Naive Bayes model with $m = 2$,

$\implies$ $p_{\mathcal{T}}$ has (nonnegative) rank at most 4,

Phylogenetic Trees II



Denoting $L = \{\text{human, chimpanzee, gorilla, orangutan}\}$ by $\{1, 2, 3, 4\}$, the flattened tensor is given by

$$\mathrm{flatt}_{12|34}(p) = \begin{pmatrix} p_{0000} & p_{0001} & p_{0002} & \cdots & p_{0033} \\ p_{0100} & p_{0101} & p_{0102} & \cdots & p_{0133} \\ p_{0200} & p_{0201} & p_{0202} & \cdots & p_{0233} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{3300} & p_{3301} & p_{3302} & \cdots & p_{3333} \end{pmatrix}.$$

$\Longrightarrow$ This is a Naive Bayes model with $m = 2$,

$\Longrightarrow$ $p_{\mathcal{T}}$ has (nonnegative) rank at most 4,

$\Longrightarrow$ $\mathrm{flatt}_{A|B}(p_{\mathcal{T}})$ has matrix rank at most 4.

Phylogenetic Trees III

Theorem (Allman and Rhodes (2008))

For $\mathcal{T}$ a phylogenetic tree and $A|B$ a split of its leaves L, let $p_{\mathcal{T}}$ be the tensor distribution obtained from a Markov process on $\mathcal{T}$. Then, assuming some genericity conditions

1. *If $A|B$ is an edge split, $\mathrm{flatt}_{A|B}(p_{\mathcal{T}})$ has rank less than or equal to 4.*
2. *If $A|B$ is not an edge split, the rank of $\mathrm{flatt}_{A|B}(p_{\mathcal{T}})$ is larger than 4*

Identifiability of tensor models (Allman et al., 2009), applications to genomic reconstruction (Fernández-Sánchez and Casanellas, 2016).

Mixture Models

From a family of distributions $\mathcal{P}$ we can construct the mixture model

$$\mathrm{Mixt}^k\left(\mathcal{P}\right) = \left\{ \sum_{i=1}^{k} \pi_i p^i \;:\; \pi \in \Delta^{k-1},\, p^i \in \mathcal{P} \right\}.$$

Upon taking the Zariski closure, mixtures models in statistics correspond to join varieties and secant varieties in algebraic geometry.

Examples: Gaussian mixtures, topic models, Latent Dirichlet Allocation (LDA) etc.

Model fitting: EM algorithm Dempster et al. (1977).

Challenges and Remarks

- The Best Low-Rank Approximation Problem is ill-posed.
- Most Tensor problems are NP-hard Hillar and Lim (2013).
- Average case polynomial-time heuristics being developed.
- `algstat` R package for algebraic statistics.
- Alpha tensor by Google DeepMind for tensor decompositions.

References

E. S. Allman and J. A. Rhodes. Phylogenetic ideals and varieties for the general Markov model. *Advances in Applied Mathematics*, 40(2):127–148, feb 2008. ISSN 01968858. doi: 10.1016/j.aam.2006.10.002.

E. S. Allman, C. Matias, and J. A. Rhodes. Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.*, 37(6A):3099–3132, 2009. ISSN 0090-5364. doi: 10.1214/09-AOS689. URL https://doi.org/10.1214/09-AOS689.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1): 1–38, 1977. ISSN 00359246.

M. Drton, B. Sturmfels, and S. Sullivant. *Lectures on Algebraic Statistics*, volume 39 of *Oberwolfach Seminars*. Springer, 2009. doi: 10.1007/978-3-7643-8905-5.

J. Fernández-Sánchez and M. Casanellas. Invariant versus classical approach when evolution is heterogeneous across sites and lineages. *Sys Bio*, 65:280–291, 2016.

C. J. Hillar and L.-H. Lim. Most tensor problems are np-hard. *J. ACM*, 60(6), 11 2013. ISSN 0004-5411. doi: 10.1145/2512329. URL https://doi.org/10.1145/2512329.

P. McCullagh. *Tensor methods in statistics: Monographs on statistics and applied probability*. Chapman and Hall/CRC, 2018.

B. Sturmfels and P. Zwiernik. Binary cumulant varieties. *Annals of Combinatorics*, 17(1): 229–250, Mar 2013. ISSN 0219-3094. doi: 10.1007/s00026-012-0174-1. URL https://doi.org/10.1007/s00026-012-0174-1.

C. Uhler. Gaussian graphical models: An algebraic and geometric perspective, 2017. URL https://arxiv.org/abs/1707.04345.