Models via sufficient statistics and partial exchangeability

Reading circle on Bayesian theory

24.10.2006

Jaakko Riihimäki

Outline

- Models via sufficient statistics
  - Summary statistics
  - Predictive and parametric sufficiency
  - Exponential family
- Models via partial exchangeability (extended data structures)
  - Several sequences of observables
  - Structured layouts
  - Covariates
  - Hierarchical models
- Exercise

Models via sufficient statistics

Summary statistics

- **Definition 4.6. (Statistic)**

  Given random quantities $x_1, \ldots, x_m$, with specified sets of possible values $X_1, \ldots, X_m$, respectively, a random vector $\mathbf{t}_m : X_1 \times \cdots \times X_m \to \mathcal{R}^{k(m)}$ $(k(m) \leq m)$ is called a $k(m)$-dimensional statistic.

- Familiar examples: sample size, mean, median, total, sum of squares...

- Data reduction if $k(m) < m$, and often a fixed dimension $k(m) = k$

Models via sufficient statistics

Predictive sufficiency

- **Definition 4.7. (Predictive sufficiency)**

  Given a sequence of random quantities $x_1, x_2, \ldots$, with probability measure P, where x_i takes values in $X_i, i = 1, 2, \ldots$ the sequence of statistics $\mathbf{t}_1, \mathbf{t}_2, \ldots$, with $\mathbf{t}_j$ defined on $X_1 \times \cdots \times X_j$, is predictive sufficient for $x_1, x_2, \ldots$ if, for all $m \geq 1$,$r \geq 1$ and $\{i_1, \ldots, i_r\} \cap \{1, \ldots, m\} = \emptyset$,

  $$p(x_{i_1}, \ldots, x_{i_r} | x_1, \ldots, x_m) = p(x_{i_1}, \ldots, x_{i_r} | \mathbf{t}_m),$$

  where $p(\cdot|\cdot)$ is the conditional density induced by P.

- In other words, future observations and past observations are conditionally independent given $\mathbf{t}_m$

Models via sufficient statistics

Parametric sufficiency

- **Definition 4.8. (Parametric sufficiency)**

  If $x_1, x_2, \ldots$ is an infinitely exchangeable sequence of random quantities, where x_i takes values in $X_i = X, i = 1, 2, \ldots$ the sequence of statistics $\mathbf{t}_1, \mathbf{t}_2, \ldots$ with $\mathbf{t}_j$ defined on $X_1 \times \cdots \times X_j$, is parametric sufficient for $x_1, x_2, \ldots$ if, for any $n \geq 1$,

  $$dQ(\boldsymbol{\theta}|x_1, \ldots, x_n) = dQ(\boldsymbol{\theta}|\mathbf{t}_n),$$

  for any $dQ(\theta)$ defining an exchangeable predictive probability model via the representation

  $$p(x_1, \ldots, x_n) = \int \prod_{i=1}^{n} p(x_i|\boldsymbol{\theta}) dQ(\boldsymbol{\theta}).$$

Models via sufficient statistics

Sufficient summaries

- There is an equivalence between predictive and parametric sufficiencies (Proposition 4.9.)

- Parametric sufficiency is also equivalent to the following conditions:

  – Neyman factorisation criterion (Proposition 4.10.)

$$p(x_1, \dots, x_m | \boldsymbol{\theta}) = h_m(\boldsymbol{t}_m, \boldsymbol{\theta}) g(x_1, \dots, x_m)$$

  – Conditional independence (Proposition 4.11.)

$$p(x_1, \dots, x_m | \boldsymbol{\theta}, \boldsymbol{t}_m) \text{ is independent of } \boldsymbol{\theta}$$

- Minimal sufficient statistic (Definition 4.9.)

Models via sufficient statistics

Sufficiency and exponential family

- Representations relating to sufficient statistics of fixed dimension
- **Definition 4.10. (One-parameter exponential family)**

  A probability density $p(x|\theta)$ belongs to the one-parameter exponential family if it is of the form

  $$p(x|\theta) = Ef(x|f, g, h, \phi, \theta, c) = f(x)g(\theta)\exp\{c\phi(\theta)h(x)\}, x \in X$$

  where, given f, h ϕ and c, $[g(\theta)]^{-1} = \int_X f(x)\exp\{c\phi(\theta)h(x)\}dx < \infty$.

- The family is either regular (X doesn't depend on θ) or otherwise non-regular.

Models via sufficient statistics

Sufficiency and exponential family

- Sufficient statistics for the one-parameter exponential family (Proposition 4.12.)

  If $x_1, x_2, \ldots, x_n \in X$, is an exchangeable sequence such that, given regular $Ef(\cdot|\cdot)$,

  $$p(x_1, \ldots, x_n) = \int_{\Theta} \prod_{i=1}^{n} Ef(x_i|f, g, h, \phi, \theta, c) dQ(\theta)$$

  for some $dQ(\theta)$, then $\boldsymbol{t}_n = [n, h(x_1) + \ldots + h(x_n)]$, for $n = 1, 2, \ldots,$ is a sequence of sufficient statistics.

- One-parameter cases can be generalised to k-parameter exponential family (Definition 4.11. & Proposition 4.13.)

Models via sufficient statistics

Canonical exponential family

- The description of exponential family can be changed into canonical form that is convenient for some cases (Definition 4.12.)

  The probability density $p(\boldsymbol{y}|\boldsymbol{\psi}) = Cef(\boldsymbol{y}|a, b, \boldsymbol{\psi}) = a(\boldsymbol{y})\exp\{\boldsymbol{y}^t\boldsymbol{\psi} - b(\boldsymbol{\psi})\}$, $\boldsymbol{y} \in Y$, derived from $Ef_k(\cdot|\cdot)$ via the transformations $\boldsymbol{y} = (y_1, \ldots, y_k)$, $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_k)$, $y_i = h_i(x)$, $\psi_i = c_i\phi_i(\boldsymbol{\theta})$, $i = 1, \ldots, k$, is the canonical form of representation of the exponential family.

- First two moments can be derived from $b(\boldsymbol{\psi})$ (Proposition 4.14.)

- Sufficient statistic in the canonical exponential family can be expressed as a sum of $\boldsymbol{y}_i$'s (Proposition 4.15.)

Models via sufficient statistics

- Until now, the exchangeable belief distributions are constructed by assuming a mixing over finite-parameter exponential family forms
- Now, we consider whether there are structural assumptions about an exchangeable sequence which imply that the mixing must be over exponential family forms
- Previously, we started from exchangeability and invariance assumptions
- Now, we start from assumptions about conditional distributions, motivated by sufficiency ideas

Models via sufficient statistics

- **Proposition 4.16. (Representation theorem under sufficiency)**

  If $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots$ is any exchangeable sequence such that, for all $n \geq 2$ and $k < n$,

$$p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k | \boldsymbol{y}_1 + \ldots + \boldsymbol{y}_n = \boldsymbol{s}) = \prod_{i=1}^{k} a(\boldsymbol{y}_i) a^{(n-k)}(\boldsymbol{s} - \boldsymbol{s}_k) / a^{(n)}(\boldsymbol{s}),$$

  where $\boldsymbol{s}_k = \boldsymbol{y}_1 + \ldots + \boldsymbol{y}_k$ and $a(\cdot)$ defines $Cef(\boldsymbol{y}|a, b, \boldsymbol{\psi})$, then

$$p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = \int \prod_{i=1}^{n} Cef(\boldsymbol{y}_i | a, b, \boldsymbol{\psi}) dQ(\boldsymbol{\psi}),$$

  for some $dQ(\boldsymbol{\psi})$.

- Information measures and exponential family
  - Discrepancy of an approximation

Models via Partial exchangeability

Models for extended data structures

- Several sequences of observables (i)
- Structured layouts (ii)
- Covariates (iii)
- Hierarchical models (iv)

Models via Partial exchangeability

Several sequences of observables

- Model is extended to several sequences
- Unrestricted exchangeability for 0-1 sequences (Definition 4.13)
- Representation theorem for several sequences of 0-1 random quantities (Proposition 4.18, generalisation of the Proposition 4.1)
- Further on, this representation theorem can be generalised for sequences with predictive sufficient statistics (Definition 4.14.)

Models via Partial exchangeability

Structured layouts

- Furthermore, model is extended to structured layouts
- For instance, situations where random quantities are of the form x_{ijk}, where subscripts indicate replicates, different contexts and various treatments
- Complete exchangeability often unacceptable
  $\longrightarrow$ restricted set of permutations of the subscripts (partial exchangeability)

Models via Partial exchangeability

Covariates

- Situations where sequences of observables $x_{i1}, x_{i2}, \ldots, i = 1, \ldots, m$ are functionally dependent (in some sense) on the observed values, $\boldsymbol{z}_i, i = 1, \ldots, m$, of a related sequence of (random) quantities
- Some typical forms in examples 4.12-4.14, where it is proceeded as if:
  - Random quantities are conditionally independent, given the values of covariates, and given the unknown parameter $\boldsymbol{\phi}$
  - The latter are assigned a prior distribution $dQ(\boldsymbol{\phi})$
- Parameters are replaced by more complex functional forms involving the covariates

Models via Partial exchangeability

Hierarchical models

- Judgements about relationships among various sequences lead to structured forms of prior specification $dQ(\theta_i, \ldots, \theta_m)$
- For instance, the following hierarchical structure can be used

$$p(y_1(n_1), \ldots, y_m(n_m)|\theta_1, \ldots, \theta_m) = \prod_{i=1}^{m} p(y_i(n_i)|\theta_i)$$

$$g(\theta_1, \ldots, \theta_m|\boldsymbol{\phi}) = \prod_{i=1}^{m} g(\theta_i|\boldsymbol{\phi})$$

$$\Pi(\boldsymbol{\phi})$$

- Hierarchical modelling provides a powerful and flexible approach to the representations of beliefs about observables in extended data structures

Exercise

- When studying the cost-effectiveness of a treatment period, a questioning is performed before and after the treatment. In the questionary some explanatory variables concerning patient characteristics are inquired, besides some other questions reflecting the patient's quality of life. Then these questionaries are combined with corresponding cost information to assess the cost-effectiveness through these measured values. What can you say about the exchangeability of the following situations?

  – You have all the questionaries and information about the costs from each of the patients.
  – In addition to previous, you also know in which hospital the patient was treated.
  – In addition to previous, you also know whether the patient has filled in the questionary by his/herself or has he/she used an assistant.

  What kind of model would you use in each of these cases?