Bayesian Modelling and Computation

Probability model with parameter θ has density

$$f_{X,Z|\Theta}(x, z \mid \theta) \quad \text{w.r.t.} \quad \mu_X(dx)\,\mu_Z(dz)$$

x: observable; z: latent variable "nuisance"

Define $f_{X|\Theta}(x|\theta) = \int f_{X,Z|\Theta}(x,z|\theta)\,\mu_Z(dz)$

Prior dist $f_\Theta(\theta)\,\mu_\Theta(d\theta)$ represent "prior belief"

Define The posterior distribution is a Regular Conditional Probability Dist

$$\mu_{\Theta|X}(d\theta|x) = \frac{f_{X|\Theta}\, f_\Theta(\theta)}{\int f_{X|\Theta}(x|\theta) f_\Theta(\theta)\,\mu_\Theta(d\theta)}\,\mu_\Theta(d\theta)$$

i.e. $\forall x$, $\mu_{\Theta|X}(d\theta|x)$ is a probability and $\forall$ measurable A $x \to \mu_{\Theta|X}(\theta\in A|x)$ is Borel measurable.

Remark There exists whenever $(\mathbb{X}, \mathcal{X})$ is Borel-isomorphic. or $\mathbb{X}$ is Polish (separable, complete metric) and $\mathcal{X}$ is Borel σ-algebra.

What is Bayesian Computation?

① Compute Bayes estimate under some loss function L

$$\hat\theta_{Bayes} = \arg\min_t \int L(t,\theta)\,\mu_{\Theta|X}(d\theta|x)$$

When $\theta \in \mathbb{R}$ if $L(t,\theta) = (t-\theta)^2$ then $\hat\theta_{Bayes} = \mathbb{E}_{\Theta|X}(\theta)$

$L(t,\theta) = |t-\theta|$ then $\hat\theta_{Bayes} = \text{median}(f_{\Theta|X})$

② Credible interval: an integrable range of the posterior, typically centred around mean.

③ Prediction: Observation is $X_{1:n} = (X_1,\dots,X_n)$ likelihood given is

$$f(X_{1:n}|t) = \prod_{i=1}^n f(X_i|\theta) \quad \text{a prediction rule is}$$

$$f(X_{n+1:n+m}|X_{1:n}) = \int \prod_{i=n+1}^{n+m} f(x_i|\theta)\,\mu(d\theta|X_{1:n})$$

④ Model Comparison is done through Bayes Factors. Given 2 models M_1, M_2, likelihoods f_1, f_2, priors μ_1, μ_2. A B.F. is the ratio of the posterior probability of each model

$$\frac{\pi(M_1)\int f_1(x|\theta)\,\mu_1(d\theta)}{\pi(M_2)\int f_2(x|\theta)\,\mu_2(d\theta)}$$ — "evidence" or "marginal likelihood" under M_1

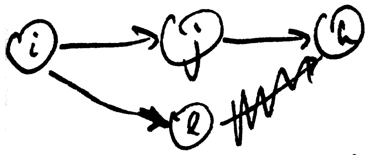
$\pi(M_i)$ is prior belief in Model i

Bayesian computation = Numerical integration on quadrature

Graphical Models — a repn of conditional independence for a set of r.v.s $X_V = \{X_v : v \in V\}$

Def A Bayes network is a directed acyclic graph (DAG) which satisfies the following property: if $p(v)$ is the set of parents of $v \in V$ (nodes w with an edge $w \to v$), then

$$f(x_V) = \prod_{v \in V} f(x_v \mid x_{p(v)})$$



Def A Markov blanket of $v \in V$ is a set of nodes S s.t. given X_S, X_v is independent of $X_{V \setminus S \setminus \{v\}}$

or $X_v \perp X_{V \setminus S \setminus \{v\}} \mid X_S$

In a Bayes network, the following is a Markov blanket for $v \in V$

$$p(v) \cup \text{children}(v) \cup \bigcup_{c \in \text{children}(v)} p(c)$$

Def A Markov Random Field is an (undirected) graph $G = (V, E)$ satisfying the global Markov property: for any partition $V = V_1 \cup V_2 \cup W$ s.t. W separates V_1, V_2 (no edges between V_1 and V_2)

$$X_{V_1} \perp X_{V_2} \mid X_W$$

Def A Gibbs Random Field is a graph $G = (V, E)$ together with a set of "potentials" ψ_c indexed by the cliques in G, such that

$$f(x_V) = \underbrace{\frac{1}{Z_{G,\psi}}}_{\text{partition fn}} \prod_{c \text{ a clique of } G} \psi_c(x_c)$$

Def A factor graph is a bipartite graph $G = (\underbrace{V}_{\text{variable nodes}} \cup \underbrace{F}_{\text{factor nodes}}, E)$ with potentials $\{\psi_a : a \in F\}$ has

$$f(x_V) = \frac{1}{Z_{G,\psi}} \prod_{a \in F} \psi_a(x_{\partial a})$$

∂a ↳ neighbours of a in G