INFÉRENCE VARIATIONNELLE

1 Warm-up

Soit (Z, X) un couple de variables aléatoires sur $\mathbb{R}^d \times \mathbb{R}^m$. On note $(z, x) \mapsto p(z, x)$ la densité jointe de la loi de (X, Z) par rapport à la mesure de Lebesgue. On considère également une famille $\mathcal{Q}$ de densités par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. On introduit alors la ELBO (Evidence Lower Bound) de la façon suivante: pour tout $q \in \mathcal{Q}$,

$$\mathcal{L}(q) = \mathbb{E}_q \left[\log \frac{p(Z, X)}{q(Z)} \middle| X \right] = \int \log \frac{p(z, X)}{q(z)} q(z) \mathrm{d}z \,.$$

1. Montrer que $\mathrm{KL}(q \| p(\cdot|X)) = \log p(X) - \mathcal{L}(q)$.
2. En déduire que $\mathcal{L}(q) \leq \log p(X)$.
3. Supposons que q soit de la forme $q : (z_1, \ldots, z_d) \mapsto \prod_{j=1}^d q_j(z_j)$ où les $\{q_j\}_{1 \leq j \leq d}$ sont des densités sur $\mathbb{R}$. Fixons $1 \leq j_0 \leq d$ et tous les q_j, $j \neq j_0$. Montrez que

$$q_{j_0}^* = \mathrm{Argmax}_{q_{j_0}} \mathcal{L}(q)$$

est la densité proportionnelle à $z_j \mapsto \exp\{\mathbb{E}_{-j_0}[\log p(z_{j_0}, Z_{-j_0}, X)]\}$, où $Z_{-j_0} = (Z_j)_{j \neq j_0}$ et $\mathbb{E}_{-j_0}$ est l'espérance lorsque la densité de Z_{-j_0} est $\prod_{j=1, j \neq j_0}^d q_j$.

2 Inférence variationnelle : modèle gaussien

Soient α_0 et β_0 deux reéls strictement positifs et μ_0 un réel. On considère les variables aléatoires suivantes : $\sigma^2 \sim \mathcal{IG}(\alpha_0, \beta_0)$, $\mu \sim \mathcal{N}(\mu_0, \sigma^2)$ et $X = (X_i)_{1 \leq i \leq n} \sim \otimes_{i=1}^n \mathcal{N}(\mu, \sigma^2)$ où $\mathcal{IG}$ est la loi inverse gamma.

1. Écrire la densité jointe des variables $Z = (\mu, \sigma^2)$ et X.
2. On considère une famille variationnelle où $q : (\mu, \sigma^2) \mapsto q_\mu(\mu) q_{\sigma^2}(\sigma^2)$. Écrire la ELBO associée.
3. Écire la mise à jour de q_μ dans une étape de l'algorithme CAVI.
4. Écire la mise à jour de q_{σ^2} dans une étape de l'algorithme CAVI.

3 Inférence variationelle pour les modèles exponentiels

On considère un couple de variables aléatoires $(Z, X) \in \mathbb{R}^d \times \mathbb{R}^m$. On note $(z, x) \mapsto p(z, x)$ la densité jointe de ce couple par rapport à la mesure de Lebesgue. Nous souhaitons utiliser dans cet exercice une approche variationnelle pour estimer la loi a posteriori $p(z|x)$. Pour cela on se donne une famille de densités sur $\mathbb{R}^d$:

$$\mathcal{Q} = \left\{ (z_1, \ldots, z_d) \mapsto \prod_{j=1}^d q_j(z_j) \,;\, q_j \text{ est une densité sur } \mathbb{R} \right\} .$$

1. Rappeler l'algorithme CAVI (Coordinate Ascent Variational Inference) pour estimer itérativement q^*.

2. Supposons que le modèle soit tel que pour tout $j \in \mathbb{R}$,

$$p(z_j|z_{-j}, x) = h(z_j)\exp(\eta(z_{-j})^T s(z_j) - a(z_{-j}))\,,$$

où $z_{-j} = (z_u)_{1\leqslant u\leqslant d, u\neq j}$ et où η, s et a sont des fonctions connues (la dépendance en x de ces fonctions est omise par simplicité). Montrer que si les densités $(q_u)_{1\leqslant u\leqslant d, u\neq j}$ sont fixées alors la mise à jour de l'algorithme CAVI de la j-ème densité est (à une constante multiplicative près),

$$q_j^*(z_j) \mapsto h(z_j)\exp\left\{\mathbb{E}_{-j}[\eta(Z_{-j})^T s(Z_j)]\right\}\,,$$

où $\mathbb{E}_{-j}$ est l'espérance sous la loi de densité $\prod_{u=1,u\neq j}^d q_u(z_u)$.

3. La convergence de l'algorithme CAVI dépend-elle de l'initialisation des densité $(q_u)_{1\leqslant u\leqslant d}$?