

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 $q \in \mathcal{Q}$,

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

1. Montrer que $\text{KL}(q \| p(\cdot | X)) = \log p(X) - \mathcal{L}(q)$.

Nous remarquons que

$$\begin{aligned} \log p(X) &= \mathbb{E}_q [\log p(X)] = \int \log p(X) q(z) dz = \int \log \frac{p(z, X)}{p(z | X)} q(z) dz \\ &= \int \log \frac{p(z, X) q(z)}{p(z | X) q(z)} q(z) dz \\ &= \text{KL}(q \| p(\cdot | X)) + \mathcal{L}(q). \end{aligned}$$

2. En déduire que $\mathcal{L}(q) \leq \log p(X)$.

Il suffit de remarquer, grâce à l'inégalité de Jensen, qu'une divergence de Kullback-Leibler est toujours positive. Dans notre cas :

$$-\text{KL}(q \| p(\cdot | X)) = \int \log \frac{p(z | X)}{q(z)} q(z) dz \leq \log \int \frac{p(z | X)}{q(z)} q(z) dz \leq 0.$$

3. Supposons que q soit de la forme $q : (z_1, \dots, z_d) \mapsto \prod_{j=1}^d q_j(z_j)$ 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

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

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 $\prod_{j=1, j \neq j_0}^d q_j$.

Par définition,

$$\mathcal{L}(q) = \int \log \frac{p(z, X)}{q(z)} q(z) dz = \mathbb{E}_{q_{j_0}} [\mathbb{E}_{-j_0} [\log p(Z_{j_0}, Z_{-j_0}, X)]] - \sum_{j=1}^d \mathbb{E}_{q_j} [\log q_j(Z_j)]$$

La densité $q_{j_0}^$ est donc solution de*

$$\text{Argmax}_{q_{j_0}} \{ \mathbb{E}_{q_{j_0}} [\mathbb{E}_{-j_0} [\log p(Z_{j_0}, Z_{-j_0}, X)]] - \mathbb{E}_{q_{j_0}} [\log q_{j_0}(Z_{j_0})] \}$$

Définissons alors la densité $\tilde{q}_{j_0} : z \mapsto c_{j_0} \exp\{\mathbb{E}_{-j_0}[\log p(z_{j_0}, Z_{-j_0}, X)]\}$ où c_{j_0} est la constante de normalisation permettant d'obtenir une densité. On obtient alors,

$$\text{Argmax}_{q_{j_0}} \{ \mathbb{E}_{q_{j_0}} [\log \tilde{q}_{j_0}(Z_{j_0})] - \mathbb{E}_{q_{j_0}} [\log q_{j_0}(Z_{j_0})] \} = \text{Argmin}_{q_{j_0}} \text{KL}(q_{j_0} \| \tilde{q}_{j_0}),$$

ce qui permet de conclure que $q_{j_0}^ = \tilde{q}_{j_0}$.*

2 Inférence variationnelle : modèle gaussien

Soient α_0 et β_0 deux réels 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 .

Pour tout x, μ, σ^2 , la logdensité jointe de $z = (\mu, \sigma^2)$ et x est donnée par :

$$\begin{aligned} \log p(z, x) &= \log p(\sigma^2) + \log p(\mu | \sigma^2) + \log p(x | \mu, \sigma^2) \\ &= -\frac{1}{2} \log(2\pi\sigma^2) - \frac{(\mu - \mu_0)^2}{2\sigma^2} \\ &\quad + \alpha_0 \log \beta_0 - \log \Gamma(\alpha_0) - (\alpha_0 + 1) \log(\sigma^2) - \frac{\beta_0}{\sigma^2} \\ &\quad - \frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

2. On considère une famille variationnelle où $q : (\mu, \sigma^2) \mapsto q_\mu(\mu)q_{\sigma^2}(\sigma^2)$. Écrire la ELBO associée.

La ELBO s'écrit

$$\begin{aligned} \mathcal{L}(q) &= \mathbb{E}_q \left[\log \frac{p(Z, X)}{q(Z)} \right] \\ &= \mathbb{E}_q [\log p(\sigma^2) + \log p(\mu | \sigma^2) + \log p(x | \mu, \sigma^2)] - \mathbb{E}_{q_\mu} [q_\mu(\mu)] - \mathbb{E}_{q_{\sigma^2}} [q_{\sigma^2}(\sigma^2)]. \end{aligned}$$

3. Écrire la mise à jour de q_μ dans une étape de l'algorithme CAVI.

On sait que la mise à jour s'écrit, à une constante additive près,

$$\log q_\mu^*(\mu) = \mathbb{E}_{q_{\sigma^2}} \left[-\frac{(\mu - \mu_0)^2}{2\sigma^2} - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right].$$

Ainsi, à une constante additive près,

$$\log q_\mu^*(\mu) = -\mathbb{E}_{q_{\sigma^2}} \left[\frac{n+1}{2\sigma^2} \left(\mu - \frac{\mu_0 + n\bar{x}_n}{n+1} \right)^2 \right].$$

où $\bar{x}_n = \sum_{i=1}^n x_i / n$. Ainsi,

$$\log q_\mu^*(\mu) = -\frac{1}{2}(n+1)\mathbb{E}_{q_{\sigma^2}}[1/\sigma^2] \left(\mu - \frac{\mu_0 + n\bar{x}_n}{n+1} \right)^2.$$

On en déduit que q_μ^* est la densité de la loi gaussienne de moyenne $(\mu_0 + n\bar{x}_n)/(n+1)$ et dont l'inverse de la variance est $(n+1)\mathbb{E}_{q_{\sigma^2}}[1/\sigma^2]$ (qui est calculable lorsque q_{σ^2} est une loi inverse gamma).

4. Écrire la mise à jour de q_{σ^2} dans une étape de l'algorithme CAVI.

On sait que la mise à jour s'écrit, à une constante additive près,

$$\begin{aligned} \log q_{\sigma^2}^*(\sigma^2) &= -\frac{1}{2} \log(2\pi\sigma^2) - (\alpha_0 + 1) \log(\sigma^2) - \frac{n}{2} \log(2\pi\sigma^2) \\ &\quad + \mathbb{E}_{q_\mu} \left[-\frac{(\mu - \mu_0)^2}{2\sigma^2} - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right]. \end{aligned}$$

Ainsi, à une constante additive près,

$$\log q_{\sigma^2}^*(\sigma^2) = -\left(\frac{1}{2} + \alpha_0 + 1 + \frac{n}{2}\right) \log \sigma^2 - \frac{\mathbb{E}_{q_\mu}[(\mu - \mu_0)^2 + \sum_{i=1}^n (x_i - \mu)^2]}{2\sigma^2}.$$

On reconnaît une loi inverse gamma de paramètres

$$\alpha = 1/2 + \alpha_0 + 1 + n/2$$

$$\beta = \mathbb{E}_{q_\mu} \left[(\mu - \mu_0)^2 + \sum_{i=1}^n (x_i - \mu)^2 \right] / 2.$$

3 Inférence variationnelle 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 :

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

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

Fixons $1 \leq j_0 \leq d$. On sait que pour toutes densités $(q_j)_{1 \leq j \leq d}$, en notant \mathcal{L} la ELBO,

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

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 $\prod_{j=1, j \neq j_0}^d q_j$. L'algorithme CAVI fonctionne donc de la façon suivante.

- (a) Initialiser toutes les densités $(q_j)_{1 \leq j \leq d}$ aux valeurs $(q_j^{(0)})_{1 \leq j \leq d}$.
- (b) Jusqu'à convergence, répéter pour $p \geq 0$:
 - i. Choisir aléatoirement $1 \leq j_0 \leq d$.
 - ii. Pour tout $j \neq j_0$, $q_j^{(p+1)} = q_j^{(p)}$.
 - iii. Poser $q_{j_0}^{(p+1)} = \text{Argmax}_{q_{j_0}} \mathcal{L}(q)$ où $q : (z_1, \dots, z_d) \mapsto q_{j_0}(z_{j_0}) \prod_{j=1, j \neq j_0}^d q_j^{(p+1)}(z_j)$.

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 $z_{-j} = (z_u)_{1 \leq u \leq d, u \neq j}$ 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 $(q_u)_{1 \leq u \leq d, u \neq j}$ sont fixées alors la mise à jour de l'algorithme CAVI de la j -ème densité est (à une constante multiplicative près),

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

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

Il suffit d'écrire, pour $1 \leq j \leq d$, la fonction $z_j \mapsto \exp\{\mathbb{E}_{-j}[\log p(z_j, Z_{-j}, X)]\}$ en utilisant la forme exponentielle de l'énoncé et de supprimer les termes multiplicatifs ne dépendants pas de z_j .

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

Oui. La convergence de l'algorithme dépend de l'initialisation de l'algorithme. C'est bien sûr un point très important en pratique (il en est de même pour l'algorithme Expectation Maximization).