

Champs moyen et méthodes variationnelles

Alain Franc, UMR BioGeCo

Version 1.1
24 septembre 2007

Table des matières

1	Calcul heuristique	3
2	Notations	4
3	Distance de Kullback-Leibler	5
4	Approximation par une loi plus simple	6
5	Application	8
6	Interprétation	9
7	Illustration numérique	9

Source : Le calcul de cette note (approximation, section 4), est un résumé adapté de [2], pp. 422–427.

Contenu

Cette note est organisée ainsi :

1. Présente des calculs heuristiques tels que pratiqués avec beaucoup d'intuition juste par les physiciens
2. Présente des notations pour formaliser la notion de champ de Gibbs sur graphe, qui est le cadre adopté ici, et permet de comprendre le calcul de la section 1
3. Introduit la distance de Kullback-Leibler, qui est une mesure parmi d'autres de distances entre lois ; comme elle n'est pas symétrique, elle est parfois appelée *divergence* de K.L.
4. Est le cœur de cette note, et introduit la notion de champ moyen comme une meilleure approximation d'une loi donnée par une loi plus simple, au sens de K.L., et fait le pont avec les approches variationnelles
5. Est une application de cette approche au modèle d'Ising (l'extension au modèle de Potts sur un graphe est immédiate),
6. Etablit le lien avec l'approche heuristique des physiciens.
7. Conclut par une illustration numérique, que l'on espère parlante.

1 Calcul heuristique

► On se donne un matériau magnétique, symbolisé par n spins alignés sur un cercle, qui peuvent être orientés chacun vers le haut ($s = +1$) ou vers le bas $s = -1$). On note s_i le spin en i . Un exemple est donné dans la figure 1. La probabilité qu'un spin soit orienté vers le haut ou vers le bas est fonction de l'orientation des spins voisins. Comme cette dépendance est invariante par translation sur le graphe, on notera de façon compacte

$$p(\uparrow) = \mathbb{P}\{s_i = 1 \mid s_{i-1}, s_{i+1}\} \quad (1.1)$$

► On définit alors une énergie associée à un spin s_i

$$E_i = -\frac{1}{2} J s_i (s_{i-1} + s_{i+1}) \quad (1.2)$$

avec

$$E_{\uparrow} = -\frac{1}{2} J (s_{i-1} + s_{i+1}) \quad (1.3)$$

En notant $\beta = 1/kT$, la loi sur les spins est donnée par

$$p(\uparrow) = \frac{1}{Z} \exp -\beta E_{\uparrow} \quad (1.4)$$

► On ne peut calculer cette loi sans connaître la configuration exacte de l'ensemble des spins, et le calcul de Z requiert la somme de 2^n termes. Aussi, des approximations sont utiles. On pose pour cela

$$x = \mathbb{P}(s = +1) \quad (1.5)$$

comme la probabilité qu'un spin soit orienté vers le haut, supposée invariante par translation. Ne connaissant pas l'état des spins en $i - 1$ et $i + 1$, on fait l'hypothèse qu'ils sont égaux à leur espérance, soit $(+1) \times x + (-1) \times (1 - x) = 2x - 1$. D'où l'équation

$$x = \mathbb{P}(s = +1) = \frac{1}{Z'} \exp \beta J (2x - 1) \quad (1.6)$$

► De même

$$1 - x = \mathbb{P}(s = -1) = \frac{1}{Z'} \exp -\beta J (2x - 1) \quad (1.7)$$

On en déduit facilement

$$Z' = \exp \beta J (2x - 1) + \exp -\beta J (2x - 1) \quad (1.8)$$

et

$$\begin{aligned} \mathbb{P}(s = -1) - \mathbb{P}(s = +1) &= 2x - 1 \\ &= \frac{\exp \beta J (2x - 1)}{\exp \beta J (2x - 1) + \exp -\beta J (2x - 1)} - \dots \\ &\dots - \frac{\exp -\beta J (2x - 1)}{\exp \beta J (2x - 1) + \exp -\beta J (2x - 1)} \\ &= \frac{\exp \beta J (2x - 1) - \exp -\beta J (2x - 1)}{\exp \beta J (2x - 1) + \exp -\beta J (2x - 1)} \\ &= \tanh \beta J (2x - 1) \end{aligned} \quad (1.9)$$

En posant

$$u = 2x - 1 \quad (1.10)$$

on obtient une équation implicite en u

$$u = \tanh \beta J u \quad (1.11)$$

dont la solution est appelée *approximation champ moyen* du modèle¹.

2 Notations

► On se donne un graphe $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, où les sites $i \in \mathcal{V}$ sont numérotés de 1 à n , et les liens $(i, j) \in \mathcal{E} \subset \mathcal{V} \times \mathcal{V}$

► On se donne une fonction d'état sur \mathcal{G}

$$\begin{aligned} s : \mathcal{V} &\longrightarrow \Lambda = \{-1, +1\} \\ i &\longrightarrow s_i \end{aligned} \quad (2.1)$$

¹Souvent abrégée en M.F. pour *mean field*.

- On note Ω l'ensemble des états possibles

$$\Omega = \{(s_1, \dots, s_n) : s_i \in \Lambda\} \quad (2.2)$$

soit

$$\Omega \equiv \Lambda^n \quad (2.3)$$

- On définit une loi de probabilité P sur Ω par

$$P = (p_\omega)_\omega, \quad \omega \in \Omega, \quad p_\omega \geq 0, \quad \sum_\omega p_\omega = 1 \quad (2.4)$$

- On définit de façon (provisoirement) imparfaite² une énergie sur Ω par

$$\begin{aligned} E : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longrightarrow E_\omega \end{aligned} \quad (2.5)$$

Un exemple souvent utilisé dans la suite est

$$E_\omega = -\frac{1}{2} \sum_{i,j} J_{i,j} s_i s_j - \sum_i h_i s_i \quad (2.6)$$

3 Distance de Kullback-Leibler

- On appelle distance de Kullback-Leibler entre deux probabilités Q et P , et on note $D_{KL}(Q||P)$ la quantité

$$D_{KL}(Q||P) = \sum_\omega q_\omega \text{Log} \frac{q_\omega}{p_\omega} \quad (3.1)$$

On remarque que, pour que cette définition ait un sens, il est nécessaire que

$$\forall \omega, \quad p_\omega, q_\omega > 0 \quad (3.2)$$

²Une définition rigoureuse devrait faire intervenir la notion de clique. Elle est donnée dans [1].

- Dans la suite, on se restreindra sans limite de validité aux lois de la forme

$$p_\omega = \frac{1}{Z(\beta)} \exp -\beta E_\omega, \quad Z(\beta) = \sum_\omega \exp -\beta E_\omega \quad (3.3)$$

- Dans ce contexte, on peut écrire

$$\begin{aligned} D_{KL}(Q\|P) &= \sum_\omega q_\omega \operatorname{Log} \frac{q_\omega}{p_\omega} \\ &= \sum_\omega q_\omega (\operatorname{Log} q_\omega - \operatorname{Log} p_\omega) \\ &= - \sum_\omega q_\omega (-\operatorname{Log} q_\omega + \operatorname{Log} Z + \beta E_\omega) \\ &= \beta \langle E \rangle_Q - H_Q + \operatorname{Log} Z \end{aligned} \quad (3.4)$$

où, si $A : \Omega \longrightarrow \mathbb{R}$,

$$\langle A \rangle_Q = \sum_\omega q_\omega A_\omega \quad (3.5)$$

4 Approximation par une loi plus simple

- Pour une loi de la forme $p_\omega = \frac{1}{Z} \exp -\beta E_\omega$, la fonction $Z(\beta)$ contient beaucoup d'information. Mais son calcul est inextricable. Aussi, une stratégie est d'approcher la loi P par une loi Q où le calcul de Z_Q est possible, au mieux au sens de la distance de Kullback-Leibler.

- Aussi, étant donnée une loi définie par $p_\omega = \frac{1}{Z(\beta)} \exp -\beta E_\omega$, on cherche une loi $Q(\alpha)$ d'une famille définie par les paramètres α telle que $D_{KL}(Q_\alpha\|P)$ soit minimale. Comme $D_{KL}(Q\|P) = \beta \langle E \rangle_Q - H_Q + \operatorname{Log} Z$, cela revient à rechercher le paramètre α tel que

$$\beta \tilde{F}(\alpha) = \beta \langle E \rangle_Q - H_Q \quad (4.1)$$

soit minimum (Z ne dépend pas de α).

► Nous allons poursuivre le calcul dans le cas où

$$\begin{cases} E_\omega = -\frac{1}{2} \sum_{i,j} J_{i,j} s_i s_j - \sum_i h_i s_i \\ q_\omega = \frac{1}{Z_Q} \exp \sum_i \alpha_i s_i \end{cases} \quad (4.2)$$

► Ces lois ont été choisies comme 'toy model' en physique des statistique (voir la section 1). Dans le vocabulaire de la physique statistique, un nœud est un *spin*, qui peut être vers le haut ou vers le bas. Les coefficients de couplage J_{ij} se lisent comme des intensités de couplage parallèles ou anti-parrallèles de spins couplés. Les coefficients h_i se lisent comme un champ extérieur de même intensité qui s'applique au spin i .

► Pour la loi Q , les variables s_i sont indépendantes

$$Q = \otimes_i q_i \quad (4.3)$$

Aussi

$$H_Q(\alpha) = \sum_i H_i(\alpha_i) \quad (4.4)$$

où $H_i(\alpha_i)$ est l'entropie de la loi définie par $p_i(s_i = +1) = e^\alpha / (e^{-\alpha} + e^\alpha)$.

► Comme les variables s_i sont indépendantes, on peut écrire

$$\begin{aligned} \langle E \rangle_Q &= \left\langle -\frac{1}{2} \sum_{i,j} J_{i,j} s_i s_j - \sum_i h_i s_i \right\rangle \\ &= -\frac{1}{2} \sum_{i,j} J_{i,j} \langle s_i s_j \rangle - \sum_i h_i \langle s_i \rangle \\ &= -\frac{1}{2} \sum_{i,j} J_{i,j} \langle s_i \rangle \langle s_j \rangle - \sum_i h_i \langle s_i \rangle \end{aligned} \quad (4.5)$$

- On calcule aisément que (en omettant les indices)

$$\begin{aligned}\langle s \rangle &= e^\alpha / (e^{-\alpha} + e^\alpha) - e^{-\alpha} / (e^{-\alpha} + e^\alpha) \\ &= \frac{e^\alpha - e^{-\alpha}}{e^{-\alpha} + e^\alpha} \\ &= \tanh \alpha\end{aligned}\tag{4.6}$$

- Le minimum de $\beta \tilde{F}(\alpha)$ est obtenu pour les paramètres $\alpha = (\alpha_1, \dots, \alpha_n)$ tels que

$$\forall i, \quad \frac{\partial \beta \tilde{F}(\alpha)}{\partial \alpha_i} = 0\tag{4.7}$$

ce qui, tout calcul fait³ produit

$$\alpha_i = \beta \left(h_i + \sum_j J_{ij} \langle s_j \rangle \right), \quad s_j = \tanh \alpha_j\tag{4.8}$$

- Ce système est un système non linéaire de n équations à n inconnues. Une manière courante de le résoudre est d'itérer cycliquement sur chaque variable, et converger vers un point fixe⁴, selon

$$\alpha_i^{t+1} = \beta \left(h_i + \sum_j J_{ij} \tanh \alpha_j^t \right)\tag{4.9}$$

5 Application

- Le modèle d'Ising est un modèle sur un graphe où

$$J_{i,j} = J \quad \Leftrightarrow \quad (i, j) \in \mathcal{E}\tag{5.1}$$

³Le calcul complet est assez long, et est omis en première écriture.

⁴Les questions techniques liées à la convergence de cet algorithme ne sont pas abordées ici.

et $J_{i,j} = 0$ sinon. En ce cas, si le graphe est de degré homogène z , l'équation devient

$$\alpha = \beta (h + z J \tanh \alpha) \quad (5.2)$$

► Si le graphe n'est pas de degré homogène, cette équation devient

$$\alpha_i = \beta \left(h_i + J \sum_{j \in \mathcal{N}(i)} \tanh \alpha_j \right) \quad (5.3)$$

6 Interprétation

► On peut lire l'équation

$$\alpha_i = \beta \left(h_i + \sum_j J_{ij} \langle s_j \rangle \right) \quad (6.1)$$

comme l'application en i d'un champ d'intensité

$$\tilde{h}_i = h_i + \sum_j J_{ij} \langle s_j \rangle \quad (6.2)$$

comme si chaque spin était soumis à un champ généré par les états moyens $\langle s_j \rangle$ des spins auxquels il est couplé. C'est la raison pour laquelle cette approximation est appelée approximation du *champ moyen*.

► Historiquement, elle a été proposée en 1908 par le physicien Weiss, avant que le formalisme des lois de probabilité ne soit rigoureusement établi, par une démarche heuristique : ne sachant pas quel est l'état précis d'un spin, on pose l'hypothèse nulle que son état est l'état moyen attendu sous les conditions rencontrées de couplage (voir section 1). Il reste à prouver que l'on obtient ainsi directement le système fermé (4.8).

7 Illustration numérique

► Afin de rendre plus concrets ces calculs, on peut les réaliser numériquement pour $n = 6$. Le nombre d'état est $|\Omega| = 2^6 = 64$. Pour chaque $\alpha \in [-2, +2]$,

avec un certain pas, on calcule la loi q_α donnée par

$$q_\omega = \prod_{i=1}^6 \frac{e^{s_i \alpha}}{e^\alpha + e^{-\alpha}}, \quad \omega = (s_1, \dots, s_6), \quad s_i \in \{-1, +1\} \quad (7.1)$$

Pour (β, J) fixés, on calcule la loi p_ω . Il est alors possible de calculer

$$\delta_\alpha = D_{KL}(q_\alpha \| p) \quad (7.2)$$

Les figures 2 à 4 indiquent le graphe de δ_α pour différentes valeurs de β et $J = 1$. Elles sont à mettre en relation avec les figures 5 à 7 qui indiquent la distribution du spin moyen sous p , où on observe bien une transition vers la bimodalité pour $\beta = \beta_c$. Comme $n = 6$, donc fini, il n'y a aucune discontinuité, et la transition est douce. Observons que même pour $n = 6$, ce qui est relativement petit, la transition est nette. Que dire pour $n = 100$, qui est un nombre faible en physique statistique ...

► Remarquons que la transition de phase β_c n'est pas la valeur de β pour laquelle la distance de K.L. avec le champ moyen devient bimodale. Cette dernière transition se réalise pour $\beta = 0.5$, qui est la valeur critique pour le champ moyen (la loi q , pas la loi p). La différence entre les deux est plus visible pour n plus grand, montrée ici pour $n = 10$ (figures 8 et 9).

Références

- [1] P. Brémaud. *Markov Chains, Gibbs fields, Monte Carlo simulation, and Queues*, volume 31 of *Texts in Applied Mathematics*. Springer Verlag, Berlin, 1999.
- [2] D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge, UK, 2003.

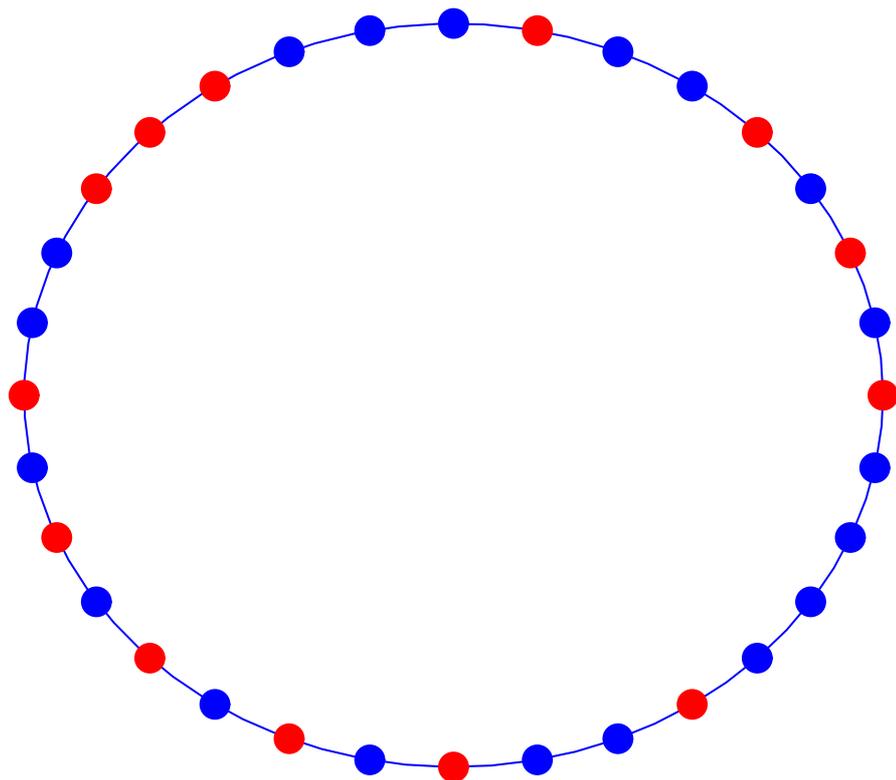


FIG. 1 – Exemple de spins vers le haut(rouges) ou vers le bas (bleus) sur un graphe circulaire. $n = 32$.

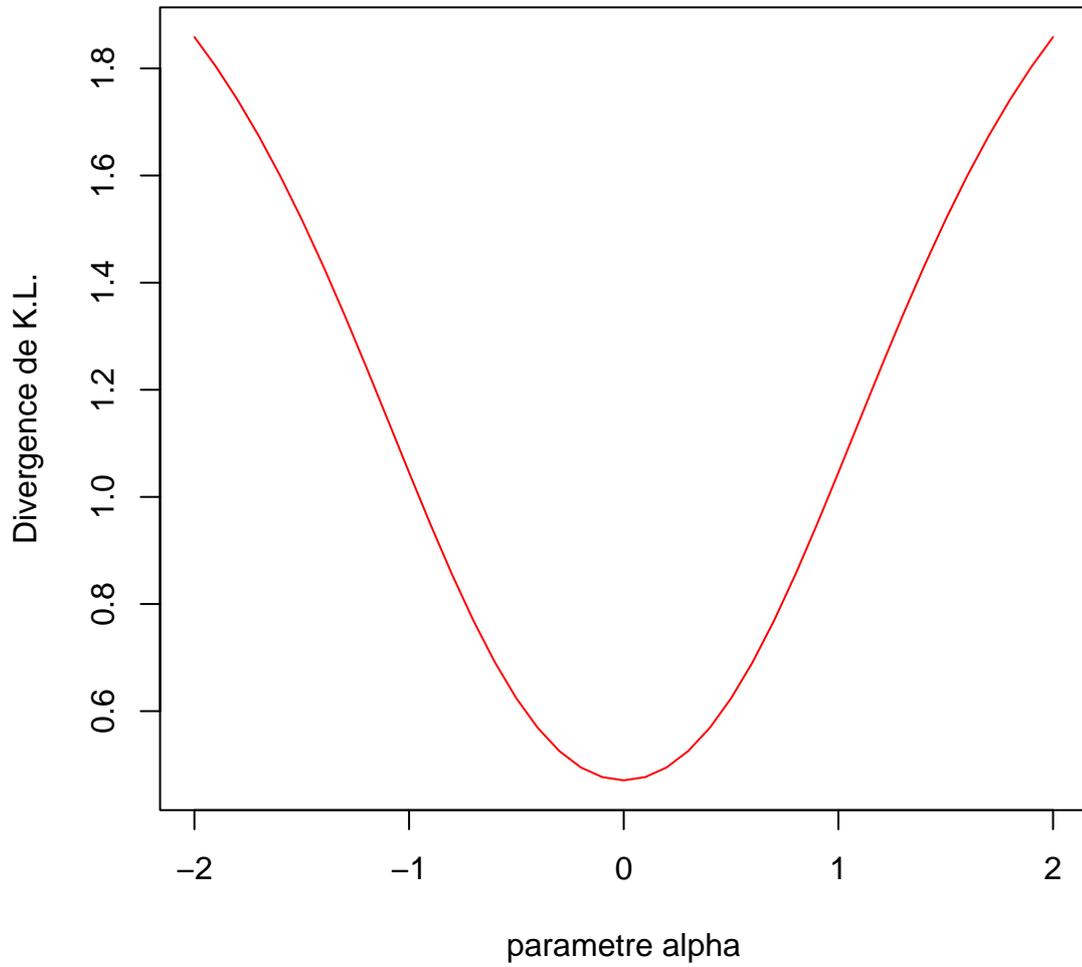


FIG. 2 – Distance de Kullback-Leibler (en y) en fonction de α (en x) pour $J = 1$ et $\beta = 0.4$. Comme $\beta < \beta_c$, la meilleure approximation est donnée par la loi $p(\uparrow) = p(\downarrow) = \frac{1}{2}$ et $\alpha^* = 0$.

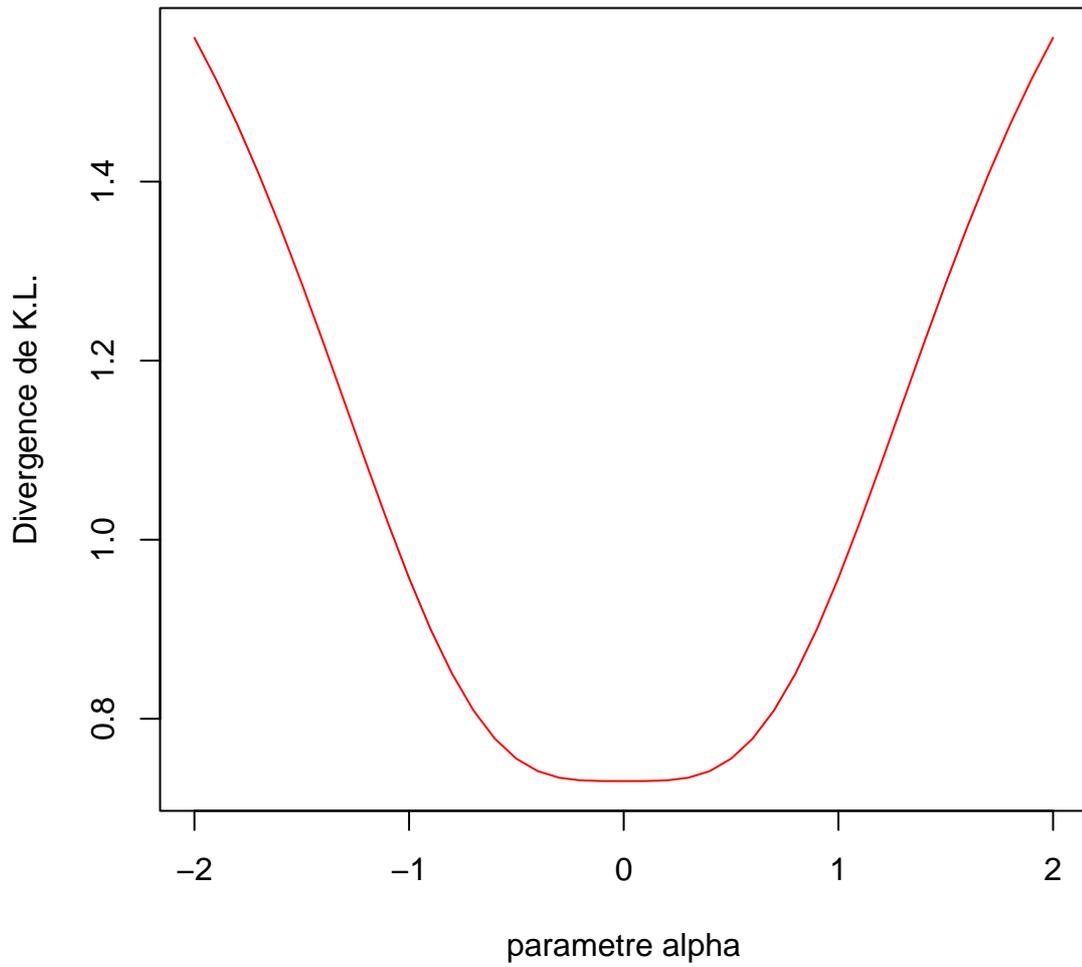


FIG. 3 – Idem, $\beta = 0.5$. Au voisinage du point critique, le minimum s'aplatit (l'estimation de α devient très difficile).

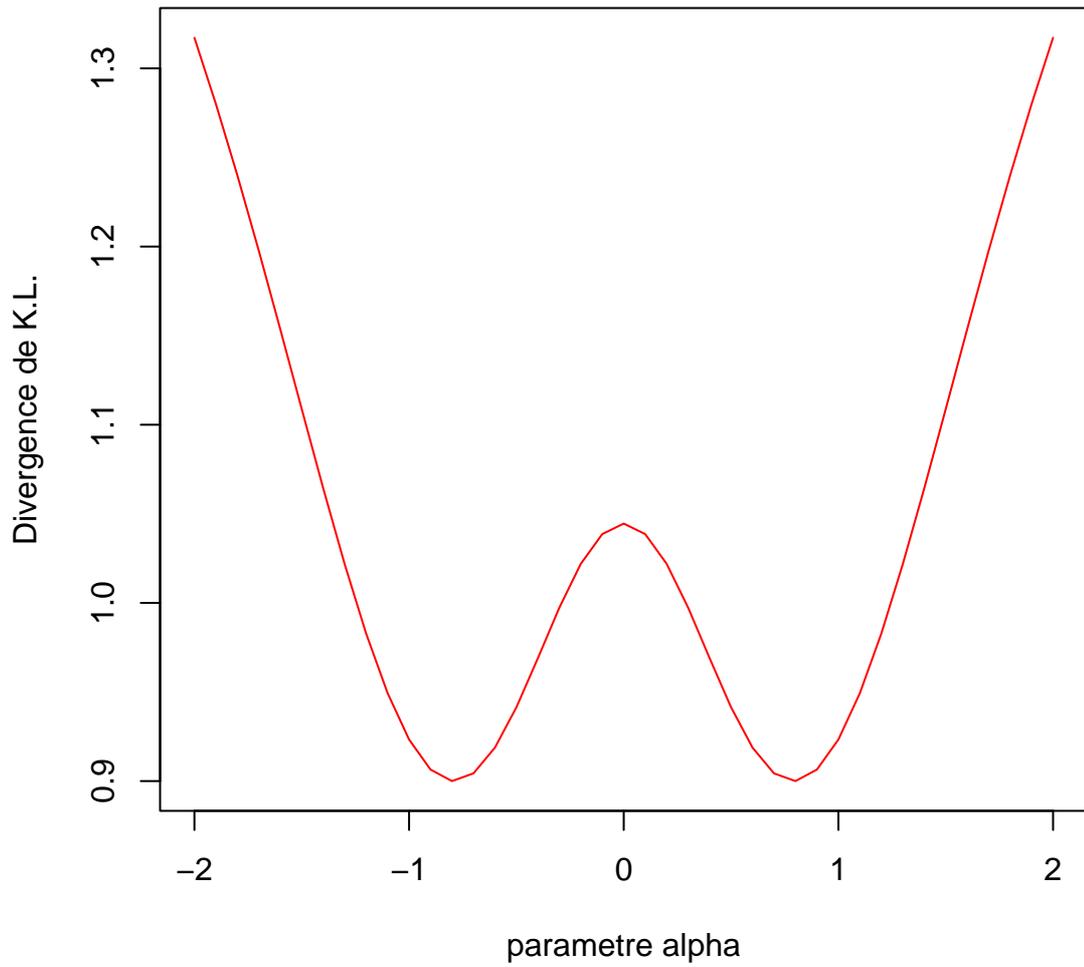


FIG. 4 – Idem, $\beta = 0.6$. Pour $\beta > \beta_c$, la distance de Kullback-Leibler possède deux minimum (ici symétriques par rapport à $\alpha = 0$).

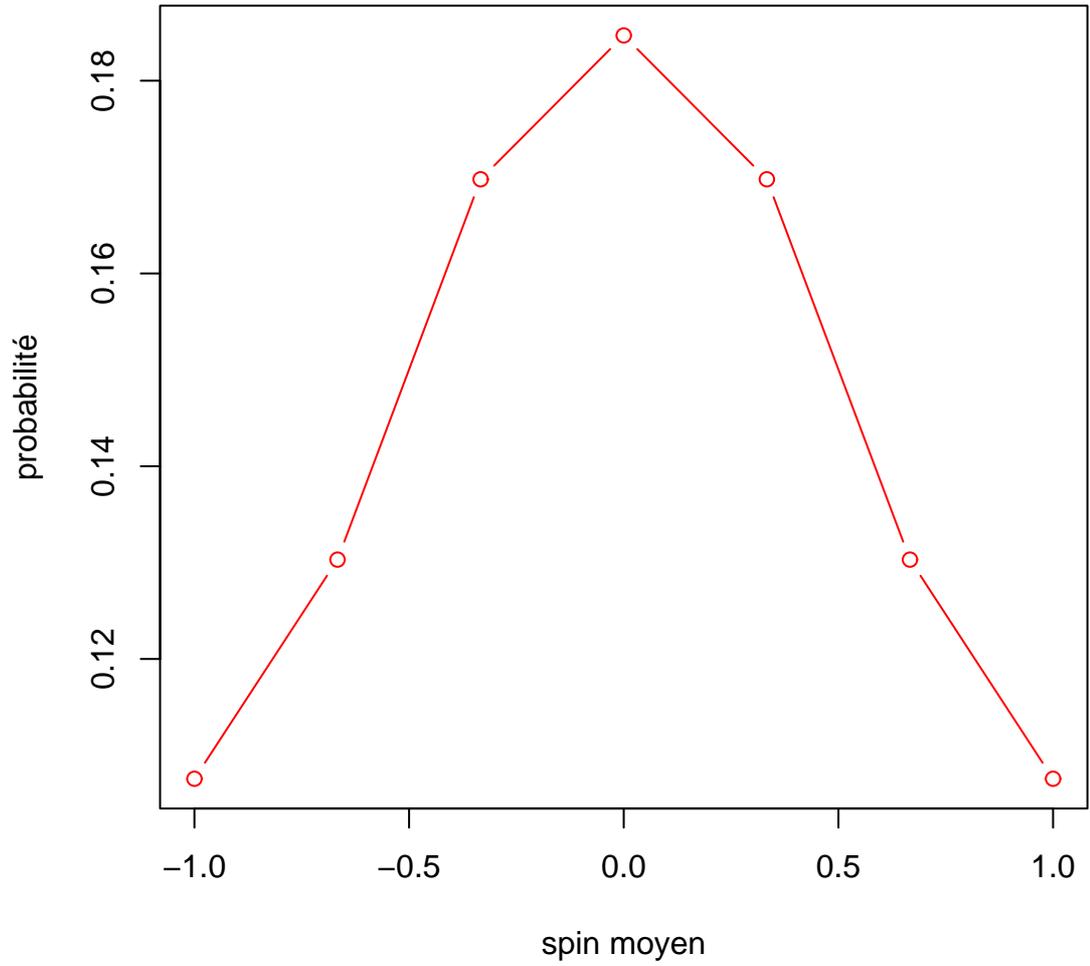


FIG. 5 – En abscisse, les valeurs possibles du spin moyen $\bar{s}_\omega = \frac{1}{n} \sum_i s_i^{(\omega)}$ et en ordonnée la probabilité d’observer ce spin moyen sous la loi p . $n = 6$, $J = 1$, $\beta = 0.4$. Pour $\beta < \beta_c$, cette probabilité a un mode pour $\bar{s} = 0$.

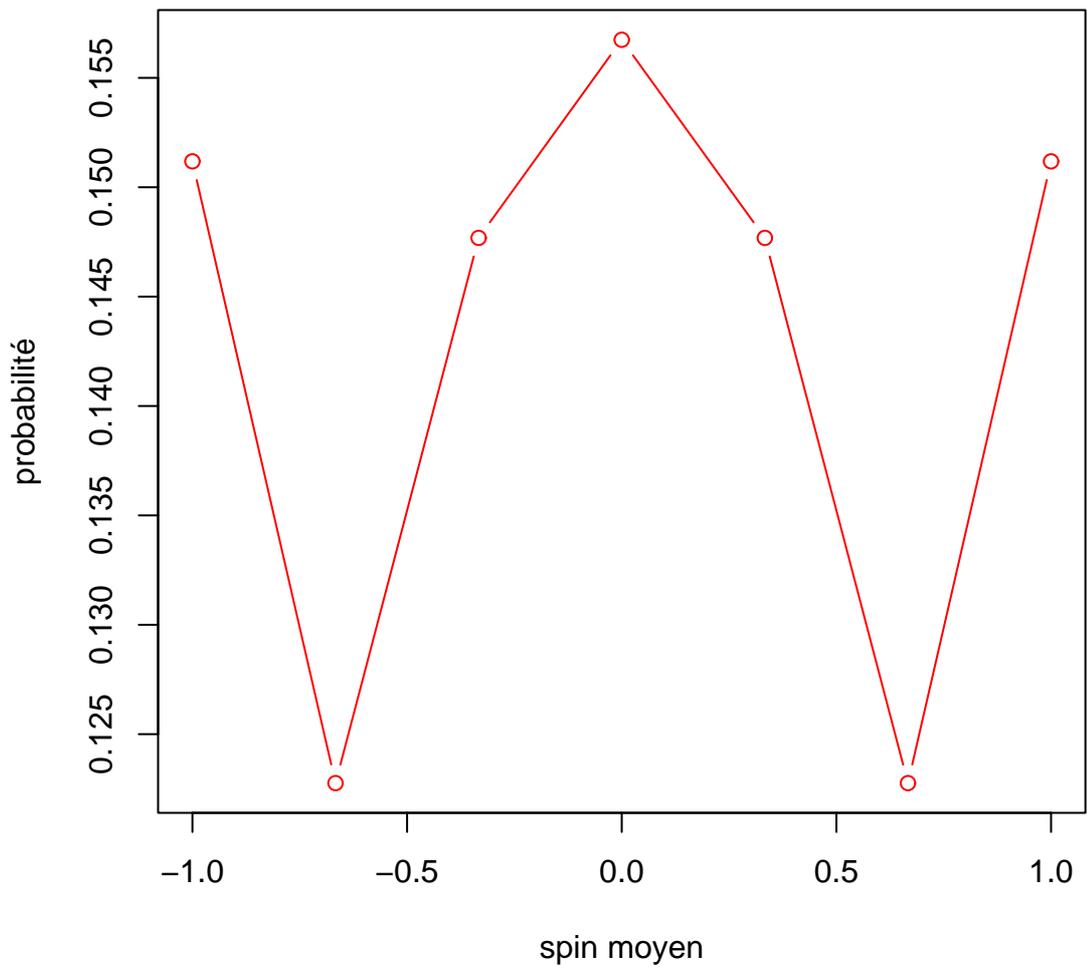


FIG. 6 – Idem, pour $\beta = 0.5$. Au voisinage du point critique, tous les spins moyens sont à peu près équiprobables ... (attention aux valeurs en ordonnée !)

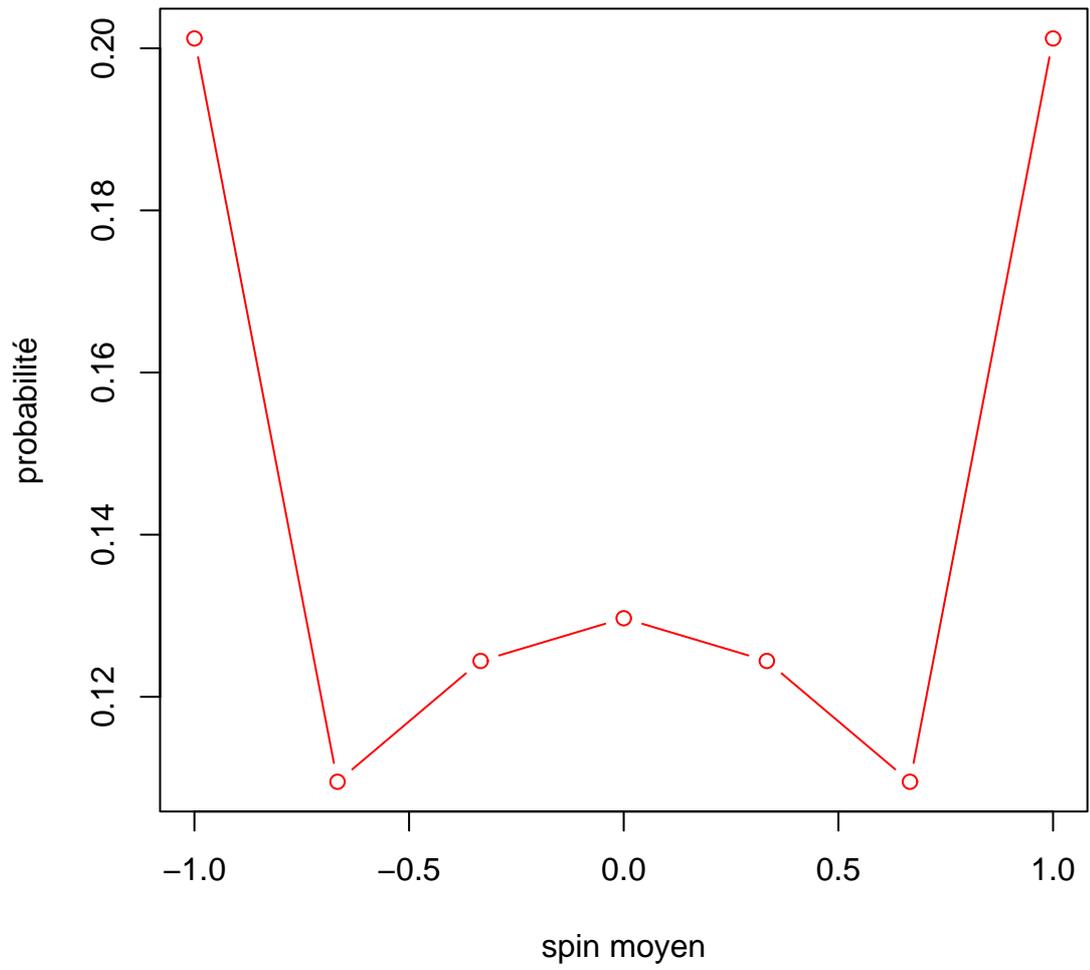


FIG. 7 – Idem, pour $\beta = 0.6$. Pour $\beta > \beta_c$, la distribution est bimodale, les modes étant $\bar{s} = \pm 1$.

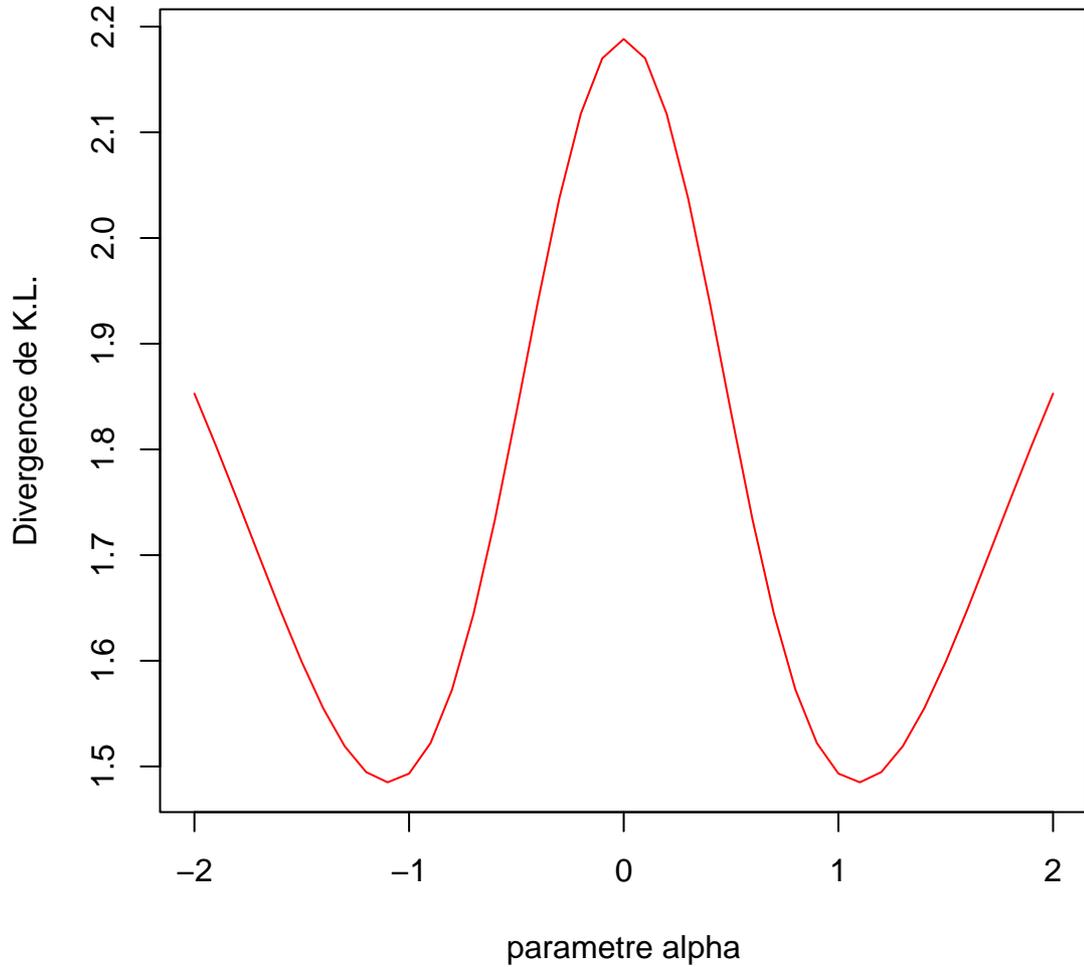


FIG. 8 – Distance de K.L. en fonction de α pour la valeur β_c de la loi p , ici trouvée numériquement ($\beta = .68$, voir figure 9). On remarque que cette distribution possède franchement deux minima. Cet écart est du à l'écart entre la valeur critique de β pour le champ moyen et pour le système spatialement explicite. $n = 10$.

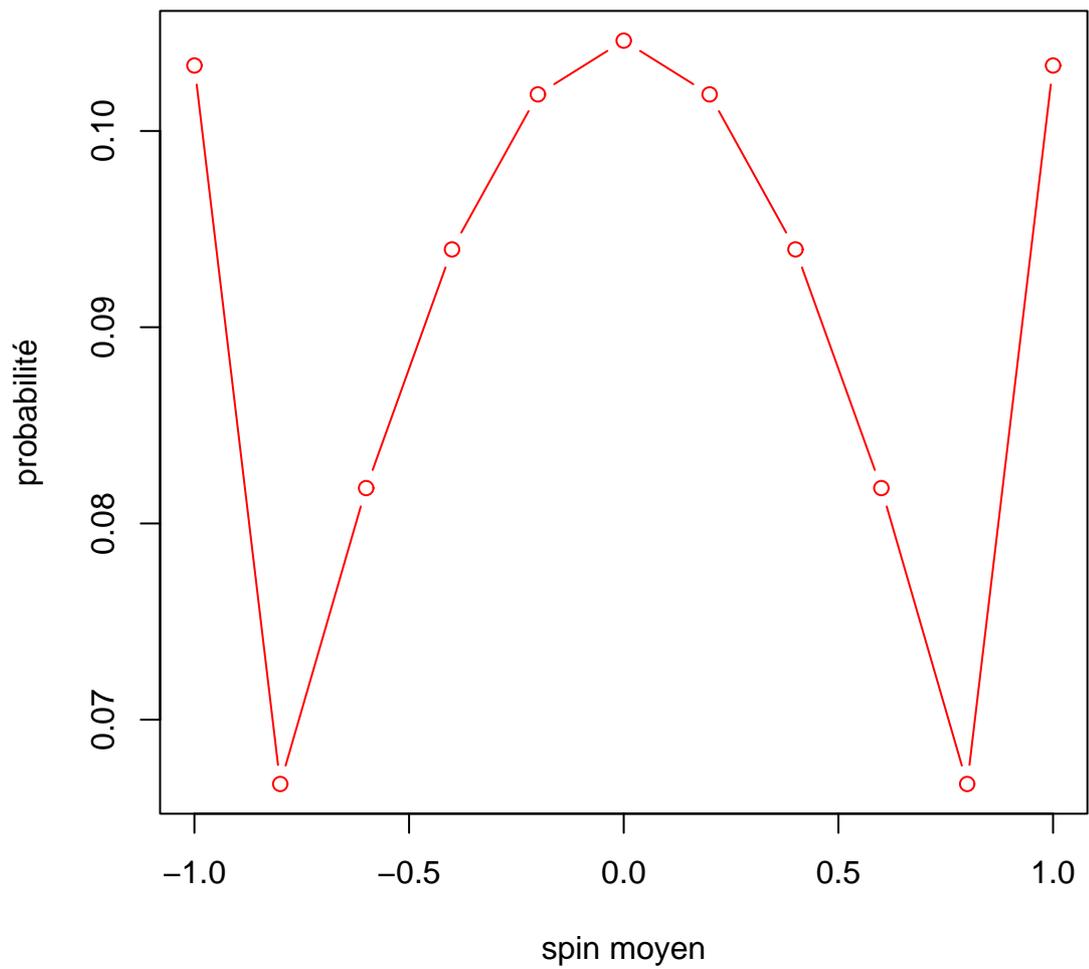


FIG. 9 – Approximation numérique de la transition unimodal / bimodal pour le spin moyen, pour $n = 10$. $\beta = 0.68$