Konvergenshastigheter hos Markovkedjor
Laboration 2: Isingmodellen

Olle Häggström

Betrakta Isingmodellen med parameter β i fallet då grafen $G = (V, E)$ utgörs av ett kvadratiskt gitter av storlek $(k \times k)$. I Gibbsmåttet $\pi = \pi_{G,\beta}$ har varje konfiguration $\xi \in \{-1, 1\}^V$ sannolikhet

$$\pi_{G,\beta}(\xi) = \frac{1}{Z_{G,\beta}} \exp\left(\beta \sum_{\langle x,y \rangle \in E} \xi(x)\xi(y)\right)$$

där $Z_{G,\beta}$ är en normaliseringskonstant.

I föreläsningarna studerade vi en MCMC-algoritm för $\pi_{G,\beta}$, baserad på den $\{-1, 1\}^V$-värda Markovkedjan $(X_0, X_1, \ldots)$ med följande övergångsmekanism: Vid varje tidpunkt n väljs en nod $x \in V$ på måfå, och ett slumptal U likformigt på $[0, 1]$. Noden x uppdateras enligt

$$X_{n+1}(x) = \begin{cases} +1 & \text{om } U < \frac{1}{1+\exp(2\beta(b^-(X_n,x)-b^+(X_n,x)))} \\ -1 & \text{annars,} \end{cases}$$

där $b^-(X_n, x)$ och $b^+(X_n, x)$ betecknar antalet minus- respektive plus-grannar till x i X_n.

Antag att kedjan startar enligt någon viss startfördelning $\mu^{(0)}$. Vissa uppskattningar av $d_{TV}(\mu^{(n)}, \pi)$ gjordes i föreläsningarna, men i praktiken är det värdefullt att komplettera dessa med numeriska resultat. Till skillnad mot i lab 1 kan vi inte direkt beräkna $d_{TV}(\mu^{(n)}, \pi)$ (pga att tillståndsrummet är så stort: det har kardinalitet 2^{k^2}), så vi tillgriper istället simuleringar. Speciellt intressant är att skatta *värsta fallet*, dvs maximum av $d_{TV}(\mu^{(n)}, \pi)$ över alla startfördelningar $\mu^{(0)}$. Låt d_n beteckna detta maximum.

Vi kommer att ha glädje av att koppla två eller flera exemplar av kedjan från olika starttillstånd, och då använder vi samma koppling som i föreläsningarna: vid varje given tidpunkt väljer samtliga kedjor samma nod x att uppdatera, och samma slumptal U.

Följande lemma är bra att ha. För två konfigurationer $\xi, \xi' \in \{-1, 1\}^V$ skriver vi $\xi \preceq \xi'$ om $\xi(x) \leq \xi'(x)$ i alla noder $x \in V$.

Lemma 1: *Koppla två exemplar $(X_0, X_1, \ldots)$ och $(Y_0, Y_1, \ldots)$ enligt ovan, med starttillstånd $X_0 = \xi$ och $Y_0 = \xi'$ valda på sådant sätt att $\xi \preceq \xi'$. Denna dominering bevaras då i tiden, dvs $X_n \preceq Y_n$ för alla n.*

Labbens tre huvuduppgifter är nu följande:

(a) Bevisa Lemma 1.

(b) Gör en undre skattning av d_n genom att (simuleringsvägen) skatta hur mycket $\mu^{(n)}$-sannolikheten för någon händelse E (med känd π-sannolikhet) avviker från $\pi(E)$. Om k är udda kan vi ta

$$E = \{\text{mindre än hälften av noderna antar värdet } +1\},$$

som ju av symmetriskäl har π-sannolikhet $\frac{1}{2}$. (Hur kan vi lämpligen modifiera E då antal noder är jämnt?) För just detta E är värsta fallet att starta i konfigurationen med $+1$ i alla noder. Detta följer av Lemma 1 (varför?).

(c) För en övre skattning av d_n, antag att vi kan starta två kedjor $(X_0, X_1, \ldots)$ och $(Y_0, Y_1, \ldots)$ med startvärden valda enligt π respektive värsta startfördelningen $\mu^{(0)}$. På vanligt sätt har vi då

$$d_n = d_{TV}(\mu^{(n)}, \pi) \leq \mathbf{P}(X_n \neq Y_n)\,.$$

Men tyvärr kan vi inte genomföra detta då vi inte vet i praktiken hur man startar enligt π (om vi visste det vore ju MCMC-simulering onödigt), och, vad värre är, vi vet inte ens vad $\mu^{(0)}$ är för en fördelning.

Hopplöst? Nej! Låt oss köra ytterligare två Markovkedjor, betecknade $(Z_0, Z_1, \ldots)$ och $(W_0, W_1, \ldots)$, med start i Z_0 ="enbart -1" och W_0 ="enbart $+1$". Pga Lemma 1 har vi då för varje n att

$$Z_n \preceq X_n \preceq W_n$$

och

$$Z_n \preceq Y_n \preceq W_n\,.$$

Om nu Z_n och W_n råkar sammanfalla, så måste därför även X_n och Y_n anta deras gemensamma värde[1]. Alltså har vi

$$d_n = d_{TV}(\mu^{(n)}, \pi) \leq \mathbf{P}(X_n \neq Y_n) \leq \mathbf{P}(Z_n \neq W_n)\,,$$

och vi kan alltså få en övre skattning av d_n genom att enbart köra kedjorna $(Z_0, Z_1, \ldots)$ och $(W_0, W_1, \ldots)$.

Prova skattningsmetoderna i (b) och (c) för givna k och β för att se hur d_n avtar i n. Ligger övre och undre skattningar tillräckligt nära varandra för att ge en någotsånär god idé om hur länge MCMC-algoritmen behöver köras? Prova sedan att variera k och β. Hurpass väl stämmer resultaten med de resultat om konvergenshastigheter jag talade om i mina föreläsningar? Går det speciellt att se en avgörande skillnad mellan fallen då β är under respektive över det kritiska värdet $\beta_c = \frac{1}{2}\ln(1+\sqrt{2})$?

[1] För den som händelsevis känner till Propp–Wilsonalgoritmen för Isingmodellen bör detta klinga bekant. Men det är inte riktigt samma sak, eller hur?