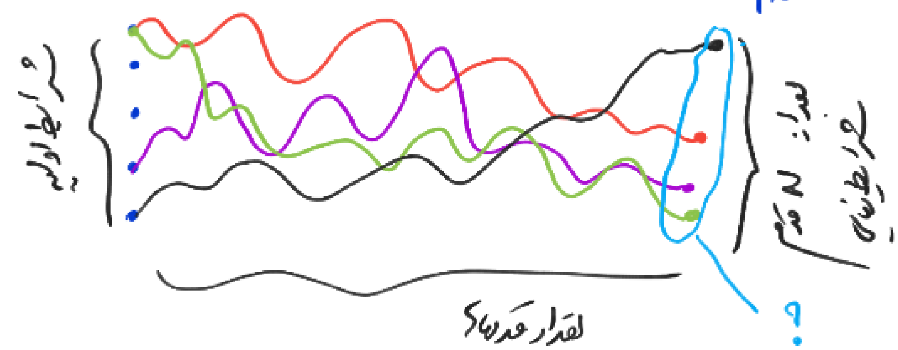
بسم الله الرحمن الرحیم

۹۹/۲/۲۳

ادامه بحث "MCMC"

آزمون همگرایی "Convergence Test"

در بستر جستجو در فضای پارامتری در چارچوب MC ، چون از بستر اعداد تصادفی استفاده می‌کنیم لذا تضمینی وجود ندارد که در حالت کلی در هر دفعه اجرای برنامه (جستجو در فضای پارامتری) به مقدار بهینه پارامترهای آزاد $\{\Theta\}$ برسیم.



$T^N(\text{initial state}) = \text{final state}$ (Unique) } آیا این حالت پیش بیاید

$T^N\{\theta_{initial}\} \to \{\theta_{Best}\}$ / اصطلاحاً می‌گوییم که برنامه همگرا شده است.

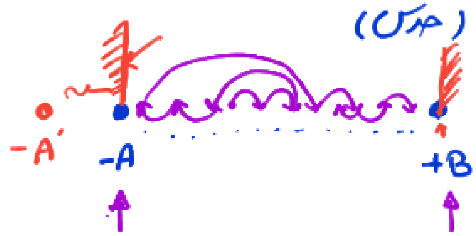
پس وضعیت نهایی قابل اعتماد است؟

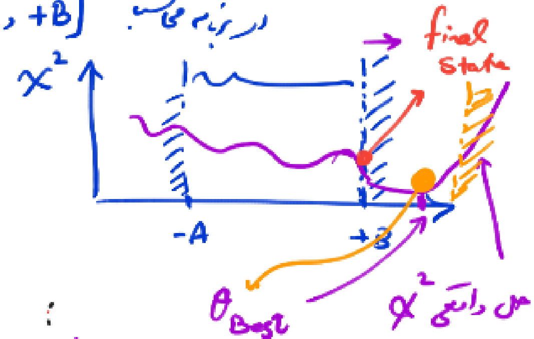
بنابراین الزاماً باید معیارهای همگرایی را بررسی کنیم.

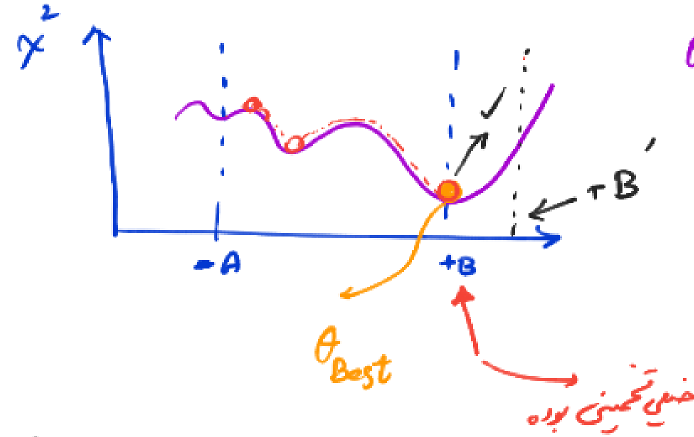
* ① مقادیر حالت نهایی الزاماً باید از مرزهای مشخص شده برای $\{\Theta\}$ در برنامه دور باشد.

فرض کنیم که $\theta \in (-\infty, \infty)$

در برنامه محاسبه $\theta \in [-A, +B]$ (حدس)

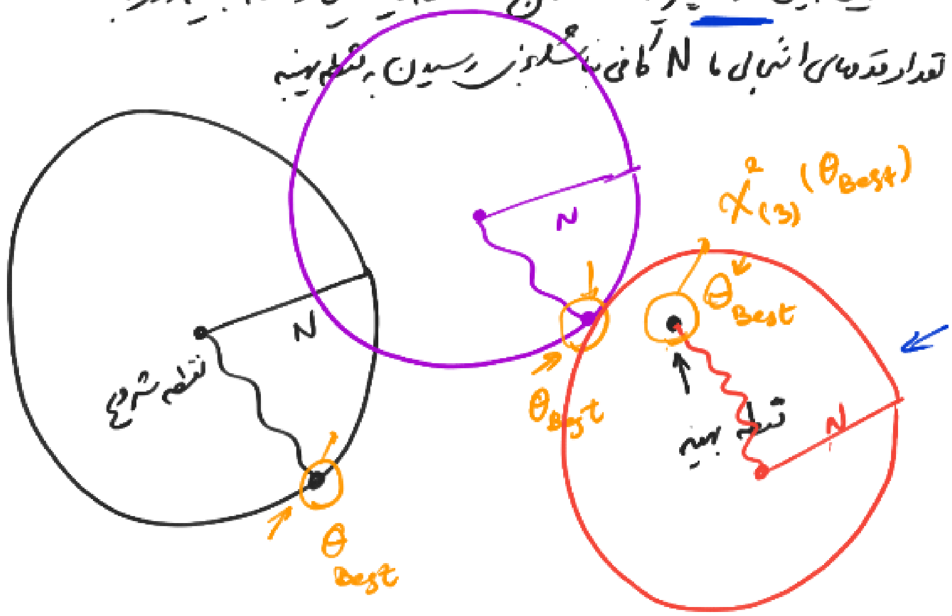






② حالت نهایی بستگی به تکرارپذیر باشد. ممکن است نقطه اولیه خیلی از نقطه بهینه دور باشد

تعداد قدم‌های انتخابی ما N کافی نباشد برای رسیدن به نقطه بهینه

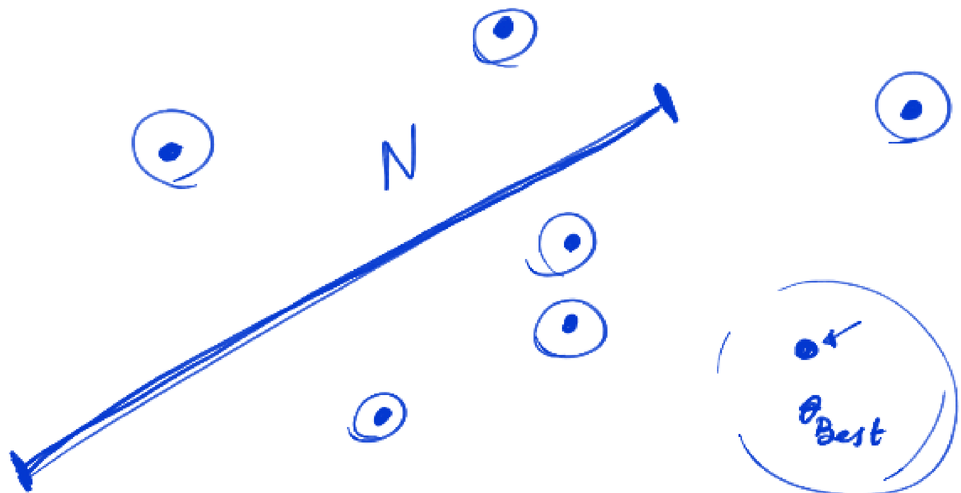


$$\underbrace{\chi^2_{(3)}(\theta = \theta^{(3)}_{Best})}_{\downarrow} < \chi^2_{(2)}(\theta = \theta^{(2)}_{Best}) < \chi^2_{(1)}(\theta = \theta^{(1)}_{Best})$$

$\theta^{(3)}_{Best}$ جواب واقعی

چون انتخاب نقطه اولیه و حرکت‌ها مبتنی بر اعداد تصادفی (best) لذا نمی‌توانیم با اطمینان بگوییم که

$\theta_{Best}^{(3)}$ همان مقدار واقعی است که دنبالش می‌گردیم.



$$\boxed{R = \frac{\left(\frac{N-1}{N}\right)W + \left(1 + \frac{1}{M}\right)B}{W}}$$

۳

$R < 1.0$ شرط همگرایی

* W : واریانس مربوط به θ ها در یک اجرا (در یک زنجیره)
* B : واریانس مربوط کل آزمایش‌ها

M : تعداد اجرای برنامه، تعداد کل آزمایش‌ها

$2N$: تعداد کل نمونه برداری‌ها در هر اجرا (در هر زنجیره). N تای دوم را مورد استفاده قرار می‌دهیم به دلیل اینکه نقطه شروع نقشی در نتایج نداشته باشد.

$2N = 1000000 \longrightarrow N = 500000$

④

⑤ نسبت به کمیت‌هایی که نشان دهنده‌ی Degree of freedom هستند

Degree of freedom (DoF)

$$\chi^2(\{\theta\}) \longrightarrow \chi^2_\nu = \frac{\chi^2}{N-M} \sim 1$$

⑤⁺

N ← تعداد نقاط اندازه‌گیری

M ← تعداد پارامترهای آزاد

⑥ حول θ_{Best} → بازه اطمینان کامل دارد مقدار دقیق ← به صورت تعیین جستجو کنیم

مقدار دقیق θ_{Best} را به دست آوریم

یک الگوریتم جستجو در فضای پارامتری مبتنی بر الگوریتم MCMC

- $\{D\}: \{D_i\}, \; i=1, \dots N_D$ تعداد اندازه‌گیری
- $\{\theta\}: \{\theta_j\}, \; j=1 \dots N_m$ تعداد پارامترهای آزاد

MCMC Algorithm

Import Data

Select $\{\theta\}_{old}$ ← نقطه شروع در فضای پارامترها (Initial state)

کاملاً بر اساس اطلاعات اولیه از مدل

Compute $\chi^2_{old}(\{\theta\}_{old})$

loop on MCMC تعداد قدم‌ها را مشخص می‌کند

Ⓐ $\boxed{\{\theta\}_{old} \xrightarrow[a_{ij}]{\text{جهش می‌کنیم}} \{\theta\}_{New}}$

Compute $\chi^2_{New}(\{\theta\}_{New})$

$$\Delta\chi^2 = \chi^2_{New}(\{\theta\}_{New}) - \chi^2_{old}(\{\theta\}_{old})$$

یادآوری $\dfrac{P_j}{P_i} = \dfrac{P_{New}}{P_{old}} = e^{-\frac{\Delta\chi^2}{2}} = e^{-\frac{\chi^2_{New}}{2}} \Big/ e^{-\chi^2_{old}/2}$ یادآوری

یادآوری کنیم فرض کردم چون اطلاعات اولیه
نداشتیم، تابع درست‌نمایی کار می‌کنم به جای
تابع احتمال پسینی posterior

$$P = \mathcal{L} = e^{-\chi^2/2}$$

(B) $\boxed{\text{Check Acceptance Rate } AR \equiv \min\left\{1, \frac{P_{New}}{P_{old}}\right\}}$ >1

Write $\{\theta\}_{old}$, $e^{-\frac{\chi^2_{old}}{2}}$

End loop

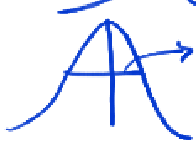
Subroutine (A) (A) زیرروال

$$\{\Delta\theta\} \rightarrow N(0, \{\sigma_\theta\})$$

$$\{\theta_{New}\} = \{\theta_{old}\} + \{\Delta\theta\} \rightarrow [\theta_1 \quad \theta_2]$$

Box Muller

یک عدد تصادفی است که دارای تابع توزیع گوسی با واریانس $\{\sigma_\theta\}$



$2\sigma_\theta$

آهسته پیش می‌رود در σ_θ بزرگ
اگر σ_θ کوچک هم باشد دیر همگرا می‌شود

Subroutine (B)

ξ = Call Random Number — بین عدد تصادفی (0,1)!

If $\xi \leqslant \dfrac{P_{New}}{P_{old}}$ then — احتمال — تابع توزیع کت

$\quad \{\theta\}_{old} = \{\theta\}_{New}$

$\quad \chi^2_{old} = \chi^2_{New}$

End if

$AR = \min\left\{1, \dfrac{P_{New}}{P_{old}}\right\}$

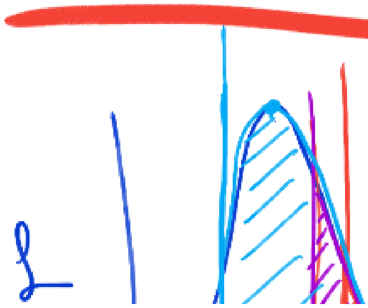
MCMC - Disadvantages — نواقص MCMC

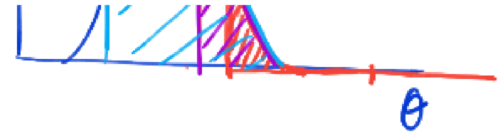
① $\{\theta\}_{New}$, $\{\theta\}_{old}$ are highly correlated

$\langle \theta(s)\, \theta(s+\tau) \rangle$

یعنی دیر همگرا می‌شود.

② Delay in Convergence for highly Non-Linear χ^2





Hamiltonian MC (HMC)

به دلیل محدودیت‌های ذاتی MCMC، روش‌های دیگری پیشنهاد شده‌اند از جمله HMC

ایده اصلی این است که از $\{\theta\}_{old} \longrightarrow \{\theta\}_{new}$ به جای اینکه به صورت یک

ولگشت تصادفی حرکت کنیم، حرکت توسط معادلات دینامیک هامیلتونی هدایت شود. پس به طور کلی داریم:

$$\mathcal{H}(q, p)$$

$$\mathcal{H}(\{\theta\}, \{\xi\}) = U(\{\theta\}) + K(\{\xi\})$$

نقش x

نقش سرعت

پارامترهای آزاد مدل

متغیرهای جدیدی هستند که نقش اندازه حرکت را بازی می‌کنند

$$\mathcal{H}(\{\theta\}, \{\xi\}) = \frac{\chi^2(\{\theta\})}{2} + K(\{\xi\})$$

$$\{\xi\} \longrightarrow \{\xi\} = \{\dot{\theta}\}$$

$$\{\dot{\xi}\} = -\frac{\partial \mathcal{H}}{\partial \{\theta\}}$$

Recall

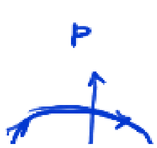
$$\frac{dp}{dt} = -\frac{\partial \mathcal{H}}{\partial q}$$

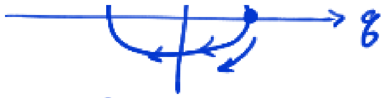
$$\frac{dq}{dt} = \frac{\partial \mathcal{H}}{\partial p}$$

نوسانگر هماهنگ

$$\left(\mathcal{H} = \frac{p^2}{2} + \frac{1}{2} q^2\right)$$

$$\dot{q} = \frac{dq}{dt} = p$$





$$\dot p = -q \rightarrow \ddot q = -q \rightarrow \boxed{\ddot q + q = 0}$$

نوسانگر هماهنگ

$$q(t) = \cos(t), \quad p(t) = -\sin(t)$$

$$\mathcal{L} = p(\{D\}|\{\theta\},\{\xi\}) = e^{-\mathcal{H}(\{\theta\},\{\xi\})}$$

Likelihood

$$P_{\text{کل}} = e^{-\frac{\chi^2(\{\theta\})}{2} - \overbrace{K(\{\xi\})}^{\xi^2/2}}$$

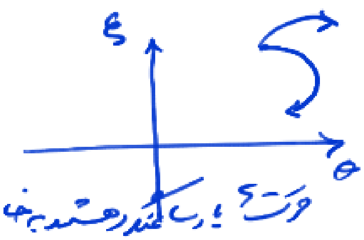
$$P = e^{-\frac{\chi^2(\{\theta\})}{2}} \underbrace{N_{\{\xi\}}(0,1)}_{e^{-\frac{\xi^2}{2}}}$$

در این حالت تابع درست‌نمایی دارای دو بخش است. یک بخش مربوط به پتانسیل که مستقیماً توسط مقدار $\{\theta\}$ مشخص می‌شود

یک بخش دیگر مربوط به انرژی جنبشی که توسط مقدار $\{\xi\}$ مشخص می‌شود

تغییرات $\delta\theta$ دقیقاً از θ_{old} : θ_{New} و ... $\rightarrow \xi$

می‌کنیم در غیر این صورت داده شود



حرکت ξ را رسم کرده‌ایم به خاطر معادلات دینامیکی.