Bayes Pharma 2019 Lyon

https://www.bayes-pharma.org

A prezentációk letölthetők a honlapról.

STAN, rstan

https://mc-stan.org

Michael Clark: Bayesian Basics

Michael Clark: Become a Bayesian with R & Stan

Hamilton Monte Carlo (HMC) módszer:

Betancourt MA: Conceptual Introduction to Hamiltonian Monte Carlo

Alapkönyv (Bayes OK, de szinte semmi kódolás nincs benne):

Andrew **Gelman**, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. **Rubin** (2013) Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Press.

Markov lánc Monte Carlo eljárások

Cél: a **posterior eloszlásból véletlen minta generálása** a likelihood függvény és az a priori eloszlás ismeretében.

Tudjuk, hogy a posterior eloszlás a likelihood függvény és az a priori eloszlás szorzatával arányos

$$\boldsymbol{P}(\boldsymbol{\theta}|\boldsymbol{Y}) \propto \boldsymbol{L}(\boldsymbol{Y}|\boldsymbol{\theta}) \cdot \boldsymbol{P}(\boldsymbol{\theta})$$

Y az adatok vektora, θ pedig a modell paramétereinek a vektora

Hogyan lehet úgy véletlen számokat generálni, hogy az eloszlást csak egy arányossági tényező erejéig ismerjük?

Kiindulunk egy **kezdő paramétervektorból**. Valamilyen szabály szerint **kisorsolunk ebből egy következőt**. Ezt az eljárást sokszor megismételjük.

Az így kapott minta elemei nem függetlenek!

Markov lánc: állapotok olyan véletlenszerű $X_0, X_1, \dots, X_t, \dots$ sorozata, ahol az **új állapotba történő átmenetek valószínűségei csak az aktuális állapottól függenek**, a korábbiaktól nem.

$$\boldsymbol{P(X_{t+1} = a_{t+1} | X_t = a_t, X_{t-1} = a_{t-1}, \dots, X_0 = a_0) =}$$

$$\boldsymbol{P(X_{t+1} = a_{t+1} | X_t = a_t)}$$

Ha az átmenetvalószínűség nem függ t-től, akkor a Markov lánc **homogén**, különben **inhomogén**.

A múlt és a jövő állapotai csak a jelenlegi állapoton keresztül függenek össze.

Példa: $X_{t+1} = \alpha \cdot X_t + \varepsilon_{t+1},\ X_t \perp\!\!\!\perp \varepsilon_{t+1},\ \varepsilon_{t+1} \sim N(0, \sigma^2)$
Gauss-Markov folyamat.

Ez Markov lánc, ha $|\alpha| < 1$. .

Stacionér eloszlás (equilibrium): az állapotok olyan valószínűségeloszlása, ami **nem változik t-ről t+1-re**.

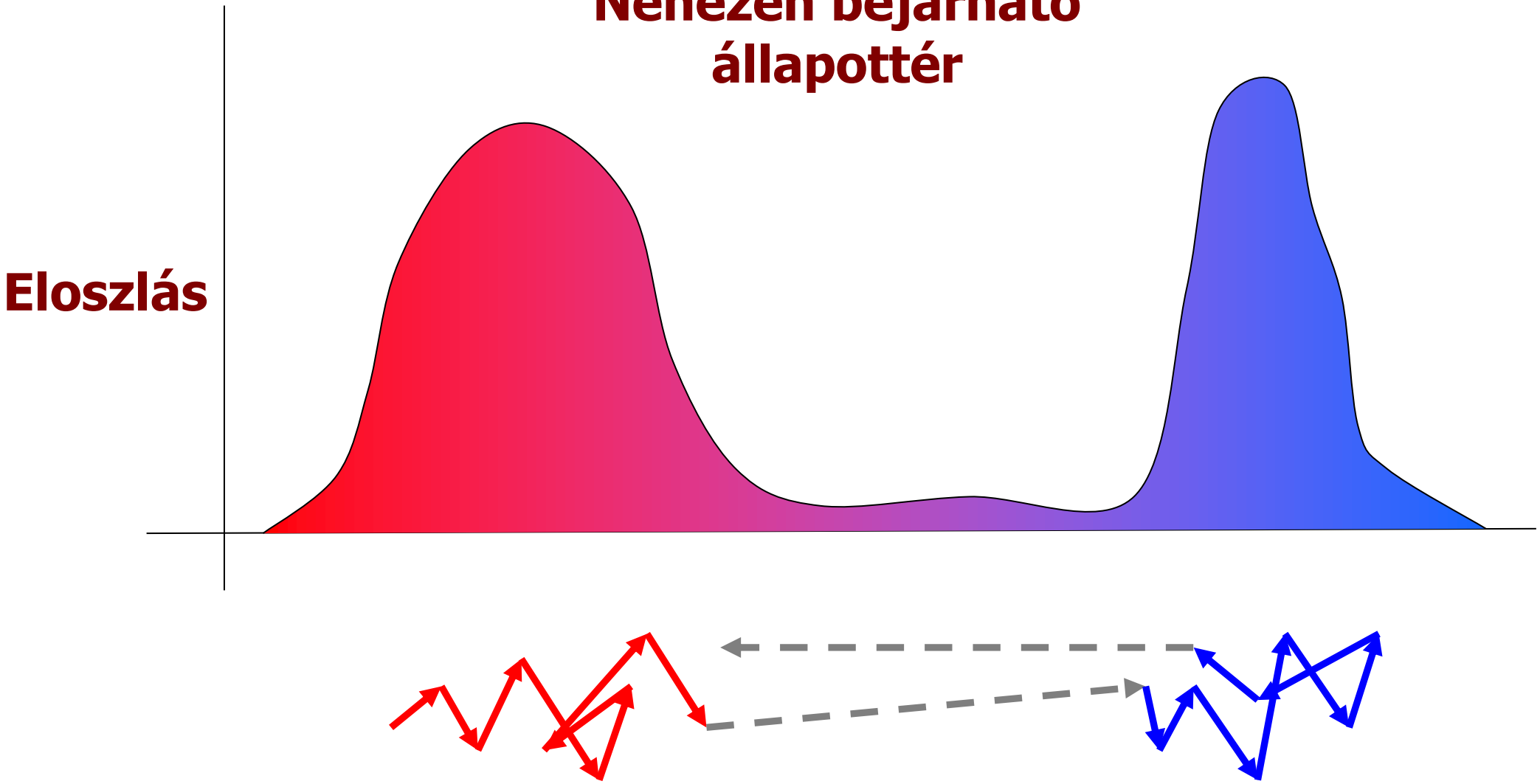
Nem mindig létezik stacionér eloszlás.

Sokszor viszont minden kezdeti állapotból **előbb-utóbb stacionér eloszlású lesz a lánc** egymást követő állapotaiból álló minta.

Jellemzően olyankor, ha az állapotváltozások nem mutatnak szisztematikus irányt, pl. növekedést vagy csökkenést.

Nem létezik stacionér eloszlás, ha az **állapottér** egymástól elválasztott, **nem átjárható** komponensekből áll.

Nehezen bejárható állapottér



Metropolis algoritmus

Olyan Markov láncot szeretnénk generálni, aminek stacionér eloszlása megegyezik a posterior eloszlással.

Rögzítünk egy $\boldsymbol{q}(\mu|\boldsymbol{\theta})$ átmenetvalószínűség-függvényt. Kiindulunk egy kezdő $\boldsymbol{\theta_0}$ paramétervektorból. Először generálunk $\boldsymbol{\theta_0}$-ból egy μ „javaslat”-ot. Ezt azonban csak akkor fogadjuk el, ha „köze van” a posterior eloszláshoz. Ez pontosabban azt jelenti, hogy kiszámítjuk a

$$\frac{P(\mu|Y)}{P(\theta_0|Y)} = \frac{L(Y|\mu)\cdot P(\mu)}{L(Y|\theta_0)\cdot P(\theta_0)}$$

hányadost és **ha ez legalább 1, akkor biztosan elfogadjuk μ-t, különben pedig csak a hányados valószínűségével kisorsolva fogadjuk el**. Elfogadáskor $\theta_1 = \mu$, elutasításkor $\theta_1 = \theta_0$. Ezt ismételgetve kapjuk a $\theta_0, \theta_1, \theta_2, \theta_3, \ldots$ mintát.

A kapott sorozat Markov lánc, hiszen az új paramétervektorra való átmenet valószínűségeit az aktuális paramétervektor meghatározza.

Tétel: Ha a $\boldsymbol{q}(\mu|\boldsymbol{\theta})$ **átmenetvalószínűség-függvény szimmetrikus**, azaz fennáll, hogy $\boldsymbol{q}(\mu|\boldsymbol{\theta}) = \boldsymbol{q}(\boldsymbol{\theta}|\mu)$, akkor a $\theta_0, \theta_1, \theta_2, \theta_3, \ldots$ paramétersorozatnak **csak a posterior eloszlás *lehet* a stacionér eloszlása**.

Arra azonban nincs elméleti garancia, hogy a sorozathoz létezik stacionér eloszlás!

Metropolis-Hastings algoritmus

Ha $\boldsymbol{q}(\mu|\boldsymbol{\theta})$ **nem szimmetrikus**, akkor θ aktuális paraméter mellett a μ javaslat felől a

$$\frac{P(\mu|Y)\cdot q(\theta|\mu)}{P(\theta|Y)\cdot q(\mu|\theta)}$$

hányados alapján döntünk. Ha ez legalább 1, akkor biztosan elfogadjuk, különben a tört értékének megfelelő valószínűséggel fogadjuk el. **A Metropolis-Hastings algoritmus a Metropolis algoritmus általánosítása**, hiszen ha $\boldsymbol{q}(\mu|\boldsymbol{\theta})$ szimmetrikus, akkor a törtet az előző alakúra lehet egyszerűsíteni.

Miért működik?

A Metropolis-algoritmusra ellenőrizzük, hogy **ha létezik stacionér eloszlás, akkor ez megegyezik a posterior eloszlással**. Tegyük fel, hogy $\boldsymbol{g(\theta)}$ stacionér eloszlás, megmutatjuk, hogy akkor $\boldsymbol{g(\theta)}$ csak a $\boldsymbol{p(\theta) = P(\theta|Y)}$ posterior eloszlás lehet. A $\boldsymbol{P(\theta_t, \theta_{t+1})}$ együttes valószínűséget a feltételes valószínűség definíciója szerint kétféleképpen felírva kapjuk, hogy

$$\boldsymbol{P(\theta_t, \theta_{t+1}) = P(\theta_t|\theta_{t+1}) \cdot g(\theta_{t+1}) = P(\theta_{t+1}|\theta_t) \cdot g(\theta_t).}$$

Ebből következik, hogy

$$\boldsymbol{\frac{P(\theta_{t+1}|\,\theta_t)}{P(\theta_t|\theta_{t+1})} = \frac{g(\theta_{t+1})}{g(\theta_t)}.}$$

A $\boldsymbol{P}(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t)$ átmenetvalószínűség a $\boldsymbol{q}(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t)$ javaslat-valószínűség és az elfogadás valószínűségének a szorzata.

Tegyük fel, hogy $\boldsymbol{p}(\boldsymbol{\theta}_{t+1}) < p(\boldsymbol{\theta}_t)$, ekkor az elfogadás valószínűsége $\boldsymbol{p}(\boldsymbol{\theta}_{t+1})/\boldsymbol{p}(\boldsymbol{\theta}_t)$, így az átmenetvalószínűség $\boldsymbol{P}(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t) = \boldsymbol{q}(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t) \cdot \boldsymbol{p}(\boldsymbol{\theta}_{t+1})/\boldsymbol{p}(\boldsymbol{\theta}_t)$. A másik irányban az átmenetvalószínűség $\boldsymbol{P}(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}) = \boldsymbol{q}(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}) \cdot \mathbf{1}$. Kihasználva, hogy $\boldsymbol{q}(\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t) = \boldsymbol{q}(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1})$, kapjuk, hogy

$$\frac{\boldsymbol{P}(\boldsymbol{\theta}_{t+1}|\,\boldsymbol{\theta}_t)}{\boldsymbol{P}(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1})} = \frac{\boldsymbol{p}(\boldsymbol{\theta}_{t+1})}{\boldsymbol{p}(\boldsymbol{\theta}_t)}.$$

Ez csak úgy lehetséges, ha $\boldsymbol{g}(\boldsymbol{\theta})$ és $\boldsymbol{p}(\boldsymbol{\theta})$ megegyeznek. A $\boldsymbol{p}(\boldsymbol{\theta}_{t+1}) > p(\boldsymbol{\theta}_t)$ feltétezés mellett hasonló számolással ugyanezt kapjuk.

Hamilton Monte Carlo módszer

Ez a rész a **Betancourt MA**: Conceptual Introduction to Hamiltonian Monte Carlo dolgozat kivonata.

A **Markov lánc Monte Carlo (MCMC) módszerek célja** valamilyen véges dimenziós Q paramétertéren értelmezett $\pi(q)$ valószínűségi sűrűségfüggvénynek (a q paraméter megfigyelt adatoktól függő posterior eloszlásának) a megismerése. Ehhez több különböző, előre definiált $f: Q \to \mathbf{R}$ valós értékű függvénynek az

$$E_\pi(f) = \int_Q f(q)\,\pi(q)dq$$

várható értékét (f-nek a teljes Q paramétertéren vett átlagát) közelítjük az f függvénynek a $q_1, q_2, \dots, q_N, \dots$ Markov láncon megfigyelt

$$\hat{f}_N = \frac{1}{N}\sum_{n=0}^{N} f(q_n)$$

átlagával.

Definíció: (stacionárius folyamat **ergodikus**). Ha a $q_1, q_2, \dots, q_N, \dots$ Markov lánc már stacionáriusnak tekinthető (tagjainak eloszlása már közelítőleg megegyezik a posterior eloszlással), akkor az ergodikusság azt jelenti, hogy a

$$\hat{f}_N = \frac{1}{N} \sum_{n=0}^{N} f(q_n)$$

trajektória-átlag olyan $\hat{f}_\infty$ állandóhoz tart, amely nem függ a $q_1, q_2, \dots, q_N, \dots$ trajektóriától. Ebben az esetben megmutatható, hogy **a trajektória-átlag megegyezik a teljes paramétertéren vett átlaggal**, azaz

$$E_\pi(f) = \hat{f}_\infty.$$

Amennyiben a Q **paramétertér dimenziója nagy** (pl. random hatásos modellek esetében), akkor az MCMC trajektória a paramétertérnek geometriailag csak **kis részét képes bejárni** véges (gyakorlatban elfogadható számú) iterációval. A dimenzió növekedésével ui. a tér egyre nagyobb része lesz „külső, perifériküs", nehezen feltárható:

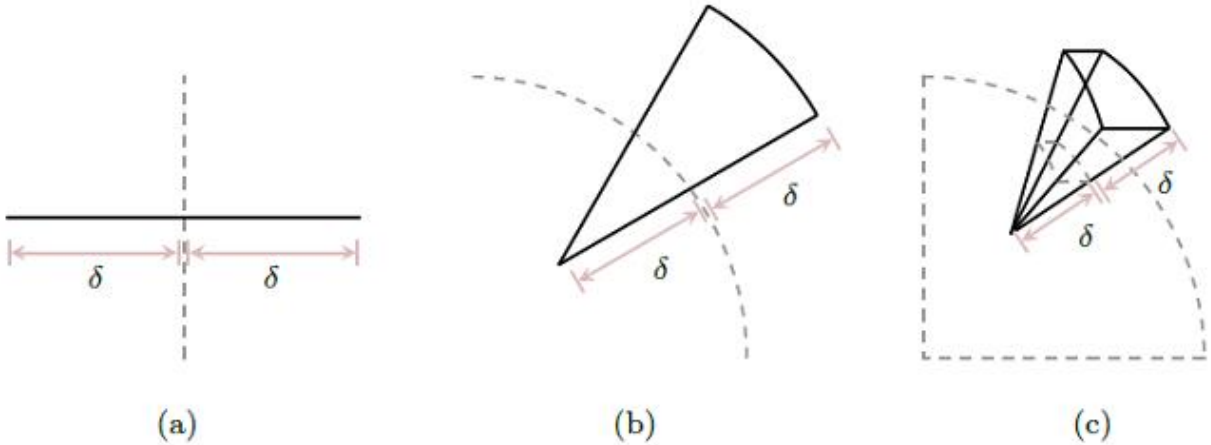


FIG 2. *The dominance of volume away from any point in parameter space can also be seen from a spherical perspective, where we consider the volume contained radial distance δ both interior to and exterior to a D-dimensional spherical shell, shown here with dashed lines. (a) In one dimension the spherical shell is a line and volumes interior and exterior are equivalent. (b) In two dimensions the spherical shell becomes circle and there is more volume immediately outside the shell than immediately inside. (c) The exterior volume grows even larger relative to the interior volume in three dimensions, where the spherical shell is now a the surface of a sphere. In fact, with increasing dimension the exterior volume grows exponentially large relative to the interior volume, and very quickly the volume around the mode is dwarfed by the volume away from the mode.*

Szerencsére a Q téren értelmeztünk egy $\pi(q)$ valószínűségi sűrűségfüggvényt, így (patologikus, a gyakorlatban nem lényeges esetektől eltekintve) a paramétertér geometriailag távoli tartományai szükségképpen kis valószínűségűek. Akármilyen nagy dimenziós is a tér, egyértelműen létezik, **behatárolható egy korlátos részhalmaza, aminek 1-hez közeli a valószínűsége**. Ezt a térrészt **tipikus halmaz**nak nevezzük.

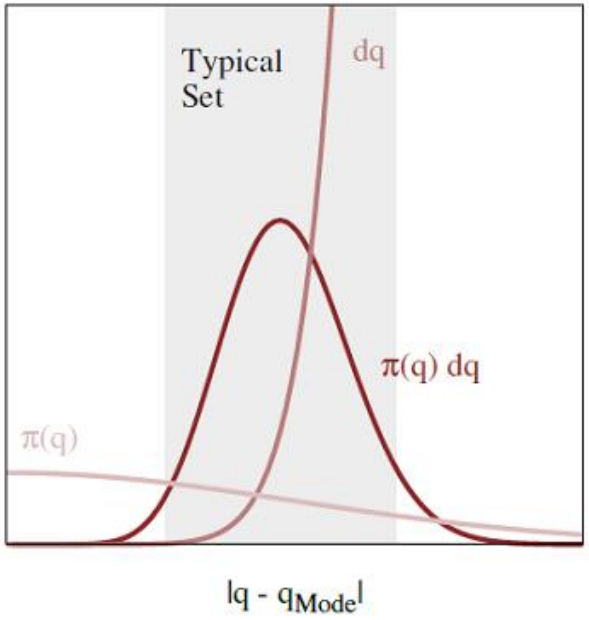


FIG 3. *In high dimensions a probability density, $\pi(q)$, will concentrate around its mode, but the volume over which we integrate that density, $\mathrm{d}q$, is much larger away from the mode. Contributions to any expectation are determined by the product of density and volume, $\pi(q)\,\mathrm{d}q$, which then concentrates in a nearly-singular neighborhood called the typical set (grey).*

A tipikus halmaz bár korlátos, jellemzően sokkal bővebb, mint a $\pi(q)$ sűrűségfüggvény maximumhelyének (a móduszának) a lokális környezete. A frekventista likelihood-elméletben elegendő a maximum likelihood becslést (az eloszlás móduszának becslését) meghatározni, **a Bayes-i MCMC módszerek a tipikus halmaz megismerésére irányulnak**.

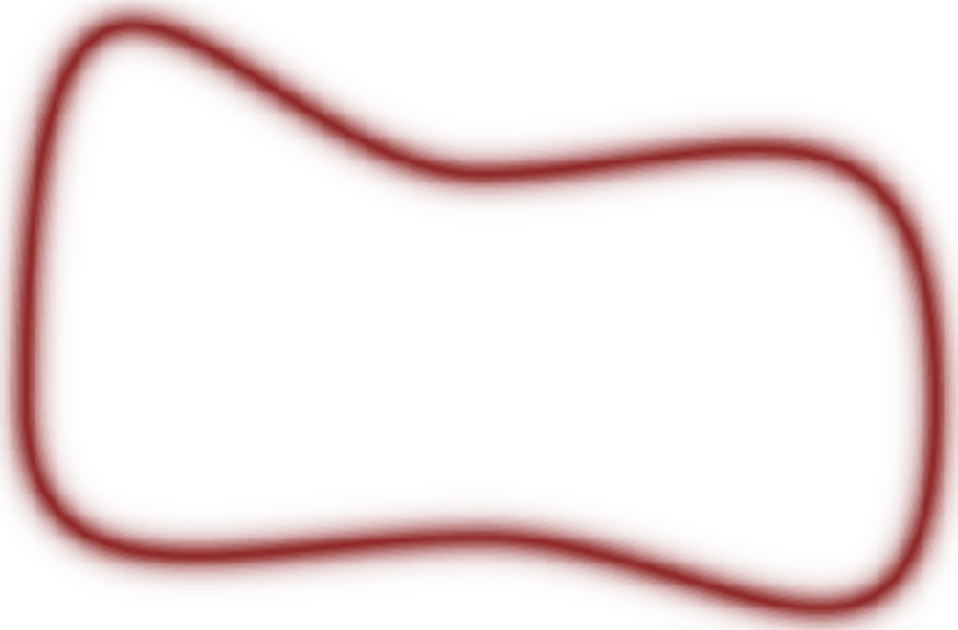


FIG 4. *In high-dimensional parameter spaces probability mass, $\pi(q)\,\mathrm{d}q$, and hence the dominant contributions to expectations, concentrates in a neighborhood called the* typical set. *In order to accurately estimate expectations we have to be able to identify where the typical set lies in parameter space so that we can focus our computational resources where they are most effective.*

A tipikus halmaz **megismerése három fázisra bontható**: (a) az MCMC iterációk **elérik**, (b) **körülhatárolják**, (c) részletesen és pontosan **feltárják**.

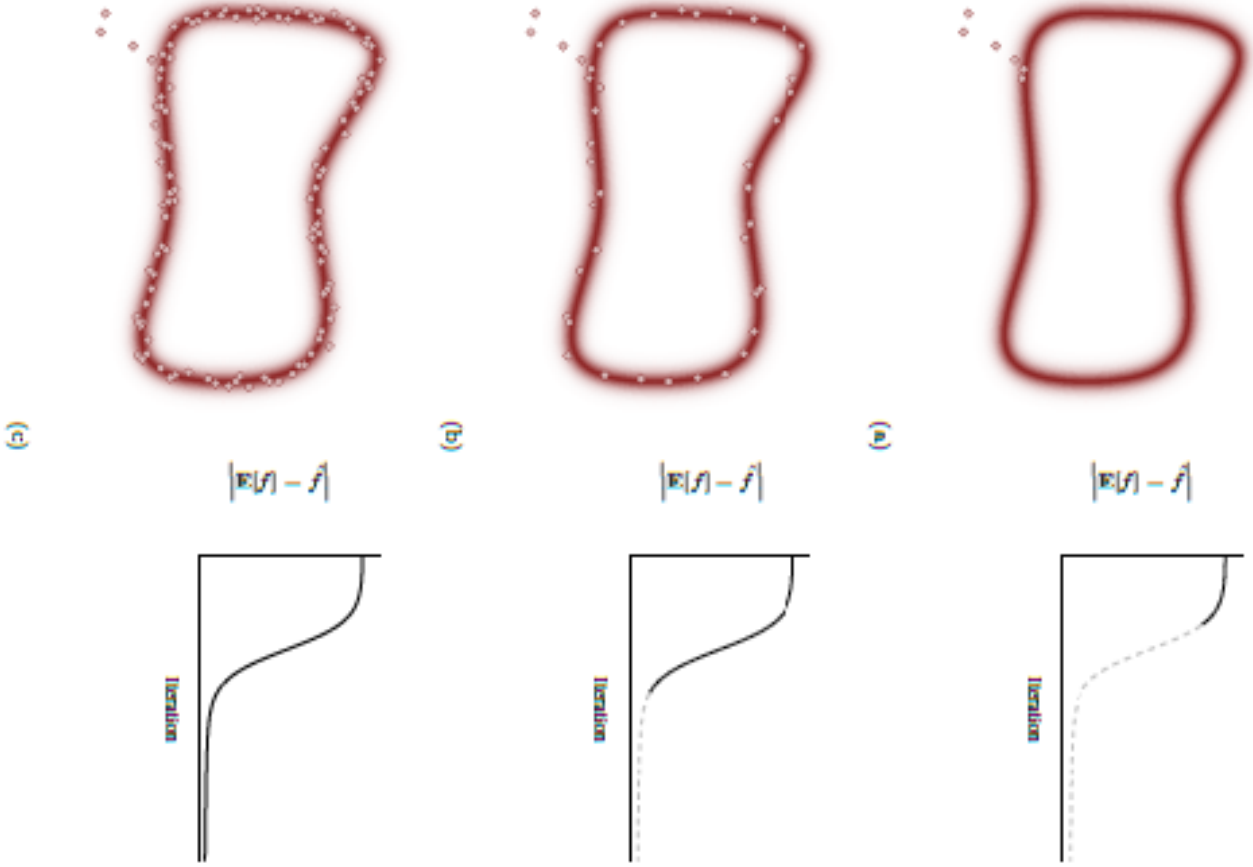
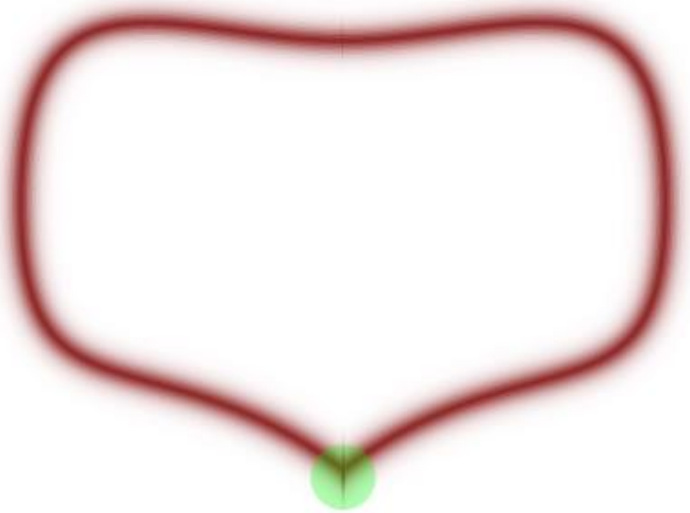


FIG 7. *Under ideal circumstances, a Markov chain explores the target distribution in three phases. (a) First the Markov chain converges to the typical set and estimators suffer from initial but ultimately transient biases. (b) Once the Markov chain finds the typical set and makes the first sojourn through it, this initial bias rapidly vanishes and the estimators become much more accurate. (c) As the Markov chain continues it explores more details of the typical set, gradually reducing the precision of the Markov chain Monte Carlo estimators towards zero.*

Ideális esetben az első két fázis a bemelegítő (warmup) iterációk során megtörténik, az iterációk eloszlása eléri a posterior eloszlást. A harmadik fázisban az iterációk konvergálnak a posterior eloszláshoz, teljesül az **ergodikusság** (ami a nagy számok törvényével analóg), továbbá az iterációkkal kiszámolt $\hat{f}_N = \frac{1}{N}\sum_{n=0}^{N} f(q_n)$ **átlagok közelítőleg normális eloszlásúvá válnak** (ami a központi határeloszlástétel megfelelője).

Patologikus esetek legtöbbször akkor adódnak, ha a $\pi(q)$ sűrűségfüggvény (ami arányosságtól eltekintve a likelihood és a prior szorzata) a paramétertér valamely kis környezetében lokálisan nagyon változékony (**nagy a $\pi(q)$ görbülete**). Ilyenkor **az iterációk a patologikus környezetet elérve beragadhatnak**, csapdába kerülnek, amiből csak sok iterációs lépés után szabadulnak ki.



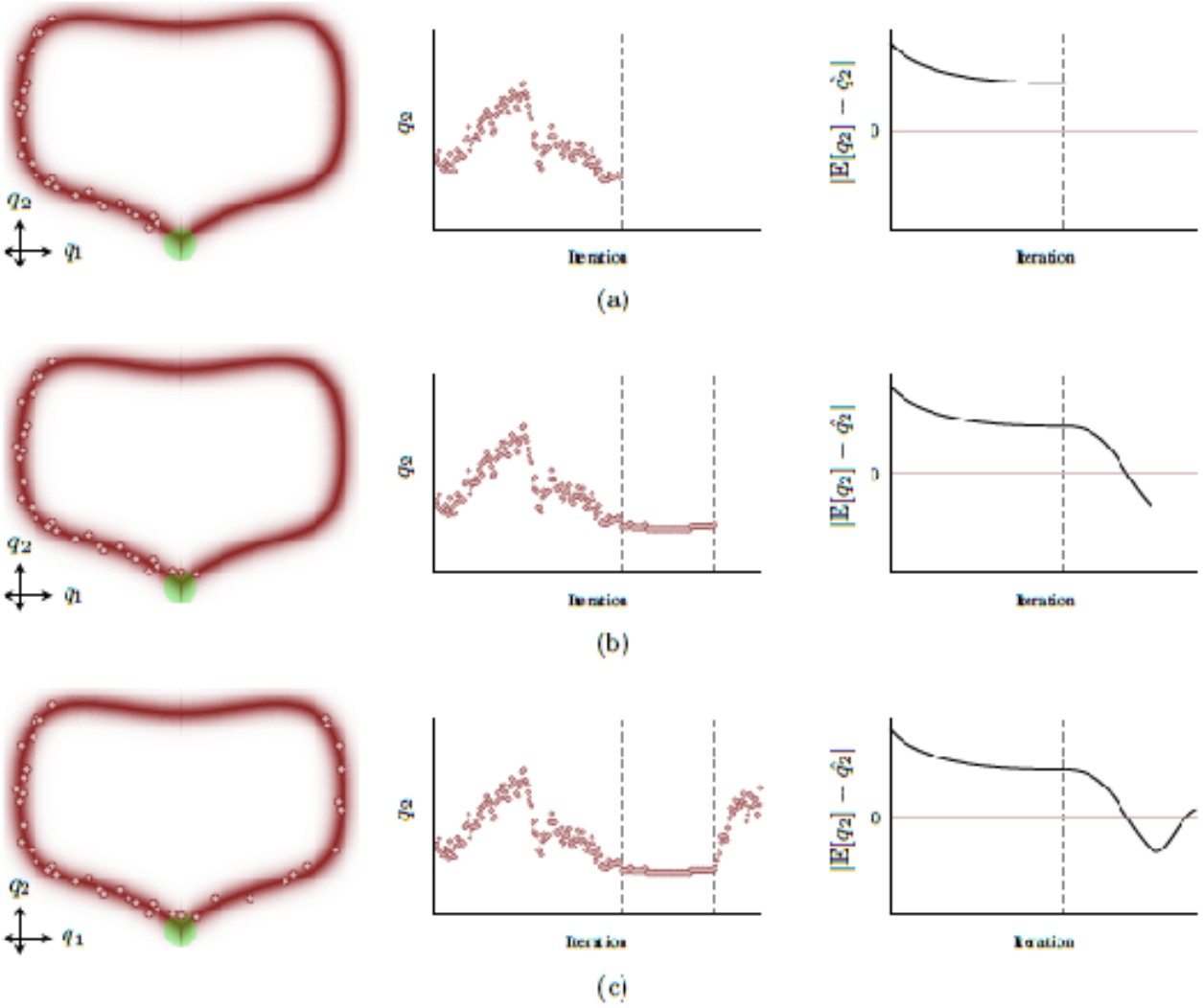


FIG 9. *In practice, pathological regions of the typical set usually cause Markov chains to get "stuck". (a) Initially the Markov chain explores well-behaved regions of the typical set, avoiding the pathological neighborhood entirely and biasing Markov chain Monte Carlo estimators. (b) If the Markov chain is run long enough then it will get stuck near the boundary of the pathological region, slowly correcting the Markov chain Monte Carlo estimators. (c) Eventually the Markov chain escapes to explore the rest of the typical set. This process repeats, causing the resulting estimators to oscillate around the true expectations and inducing strong biases unless the chain is improbably stopped at exactly the right time.*

Példa (random hatás **centrált paraméterezés**e):

```
parameters {
  real mu_beta;
  real<lower=0> sigma;
  vector[K] beta;
model {
  mu_beta ~ normal(0, 10);
  sigma ~ cauchy(0, 5);
  beta ~ normal(mu_beta, sigma); // Vektorizált (mind a K)
```

Amikor `sigma` éppen kicsi, akkor `beta` iteráltjai csak kicsit változhatnak. Ebben a (`sigma, beta`) környezetben a likelihood extrém nagy görbületű.

Ugyanez **nemcentrált paraméterezés**sel (`sigma` és `beta_raw` nem kapcsolódik össze):

```
  vector[K] beta_raw;
  real<lower=0> sigma;
transformed parameters {
  vector[K] beta;
  beta = mu_beta + sigma * beta_raw;
model {
  sigma ~ cauchy(0, 5);
  beta_raw ~ std_normal();
```

Az MCMC-iterációk konstruálásakor elsőre talán úgy gondolnánk, hogy a paraméterteret a $\pi(q)$ valószínűségi sűrűségfüggvény (illetve a likelihood x prior) gradiense ($\pi(q)$ leggyorsabb változásának iránya) mentén érdemes feltérképezni. **Ezzel a módszerrel azonban a móduszhoz konvergálnak az iterációk** (mint a maximum likelihood-nál):

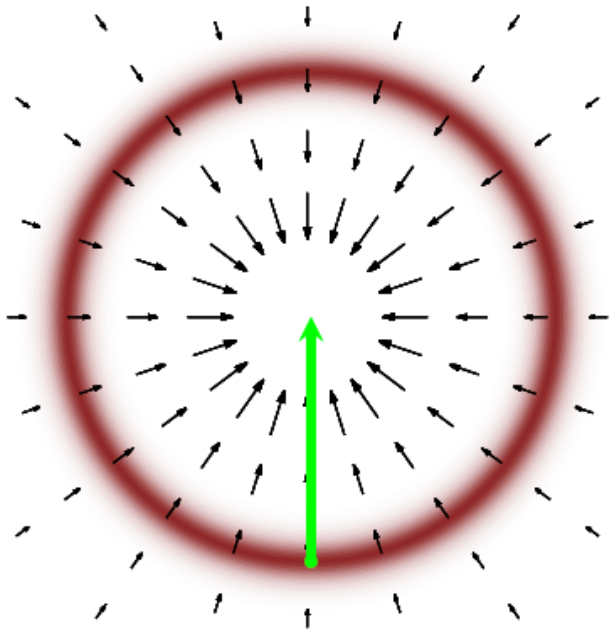


FIG 13. *The gradient of the target probability density function encodes information about the geometry of the typical set, but not enough to guide us through the typical set by itself. Following along the gradient instead pulls us away from the typical set and towards the mode of the target density. In order to generate motion through the typical set we need to introduce additional structure that carefully twists the gradient into alignment with the typical set.*

Annak érdekében, hogy a tipikus halmazban maradjunk, a gradiensekre nagyjából ortogonális pályán kell tartani az iterációkat. Mintha a módusz, mint bolygó körül egy műhold keringene. **A műholdat a sebessége (impulzusa) tartja stabil röppályán.**

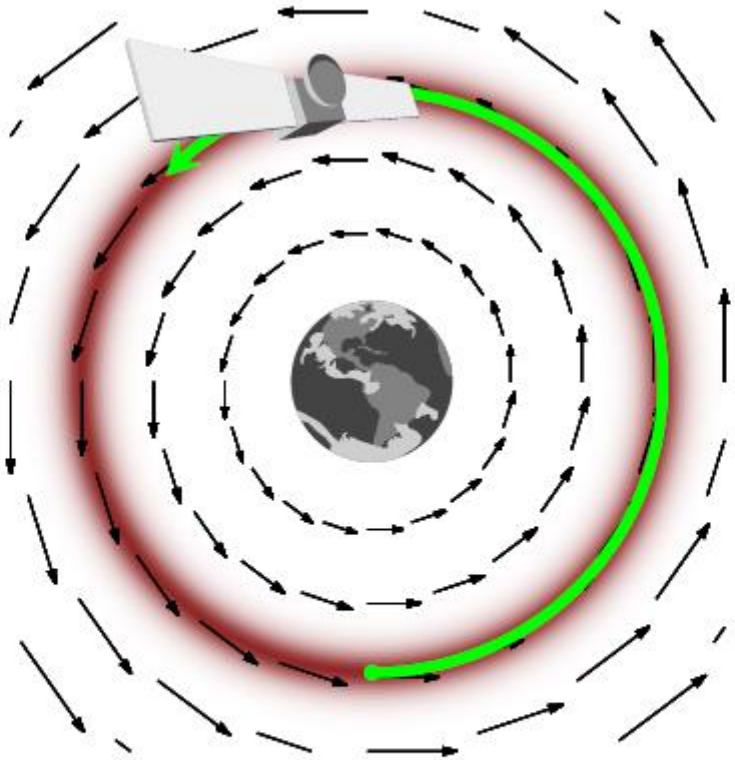


FIG 17. *When we introduce exactly the right amount of momentum to the physical system, the equations describing the evolution of the satellite define a vector field aligned with the orbit. The subsequent evolution of the system will then trace out orbital trajectories.*

A **Hamilton Monte Carlo módszer** alapötlete, hogy a $\pi(q)$ eloszláshoz és a q paraméterhez értelmezünk egy p **impulzus**t, és **a Markov láncot** a Q paramétertér helyett a (q,p) párokból álló **fázistéren értelmezzük**. A fázistérre kiterjesztjük a $\pi(q)$ céleloszlást, így nyerjük a

$$\pi(q,p) = \pi(p|q) \cdot \pi(q)$$

kanonikus eloszlást.

Ha a Q paramétertér D dimenziós (azaz D számú paraméter van a modellben), akkor minden dimenzióhoz (paraméterhez) értelmezünk egy-egy impulzust. Így **a fázistér $2D$-dimenziós lesz**.

A kanonikus eloszlásnak a Q téren a marginális eloszlása a $\pi(q)$ céleloszlás, így **a fázistéren iterált trajektóriát Q-ra vetítve feltárjuk a paraméterteret**:

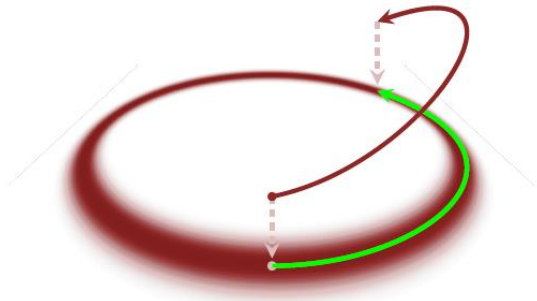


FIG 19. *By constructing a probability distribution on phase space that marginalizes to the target distribution, we ensure that the typical set on phase space projects to the typical set of the target distribution. In particular, if we can construct trajectories that efficiently explore the joint distribution (black) they will project to trajectories that efficiently explore the target distribution (green).*

A fizikai analógiát folytatva, a műhold akkor marad stabil röppályán, ha energiája állandó. Konzervatív fizikai rendszerekben (ahol a helyzeti és mozgási energia összege állandó) értelmezik az ún. Hamilton-függvényt, ami a kétféle energia összege. A Hamilton-függvény megadása kijelöli azokat a trajektóriákat, amelyek mentén az energia állandó marad.

A fázistér $\boldsymbol{\pi(q,p)}$ **kanonikus eloszlásához rendeljük hozzá** a

$$H(q,p) = -\log \pi(q,p)$$

Hamilton-függvényt. Ez természetes módon felbomlik

$$H(q,p) = -\log \pi(p|q) - \log \pi(q)$$

alakúra, ahol $K(p,q) = -\log \pi(p|q)$ a **kinetikus energia**, $V(q) = -\log \pi(q)$ a **potenciális (helyzeti) energia**.

A fázistéren a $\big(q(t), p(t)\big)$ trajektóriákat a

$$\frac{dq}{dt} = +\frac{\partial H}{\partial p} = \frac{\partial K}{\partial p}$$
$$\frac{dp}{dt} = -\frac{\partial H}{\partial q} = -\frac{\partial K}{\partial q} - \frac{\partial V}{\partial q}$$

Hamilton-egyenletek definiálják.

Tétel: **A trajektóriák mentén a H Hamilton-függvény értéke (az energia) állandó**.

Bizonyítás: A Hamilton-függvény idő szerinti derivátja

$$\frac{dH}{dt} = \frac{\partial H}{\partial q} \cdot \dot{q} + \frac{\partial H}{\partial p} \cdot \dot{p} = \frac{\partial H}{\partial q} \cdot \frac{\partial H}{\partial p} - \frac{\partial H}{\partial p} \cdot \frac{\partial H}{\partial q} = 0,$$

ezért $H(t)$ nem függ az eltelt t időtől.

A **Hamilton Markov állapotváltás** a fázistérben három lépésben történik.

1. Annak érdekében, hogy a paramétertérből a fázistérbe jussunk, a q paraméterhez **kisorsoljuk a p impulzust**: $p \sim \pi(p|q)$.
2. A fázistérben valamennyi ideig kiintegrájuk a Hamilton-egyenleteket (azaz **meghatározzuk a $(q(t), p(t))$ folytonos időskálás görbét**).
3. A görbe végét **(a végső állapotot) visszavetítjük** a paramétertérbe

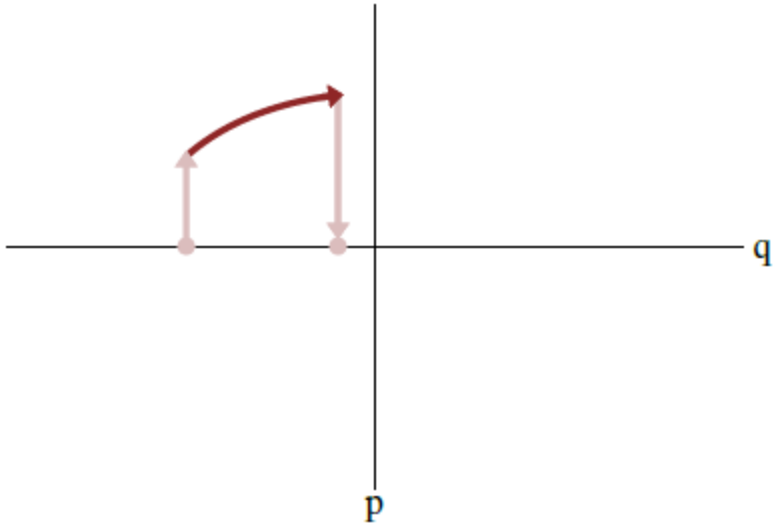


FIG 20. *Every Hamiltonian Markov transition is comprised of a random lift from the target parameter space onto phase space (light red), a deterministic Hamiltonian trajectory through phase space (dark red), and a projection back down to the target parameter space (light red).*

A fázistér természetes geometriája

A Hamilton-trajektóriák mentén állandó az energia, ezért **minden trajektória teljes egészében hozzátartozik valamelyik**

$$H^{-1}(E) = \{q, p | H(q, p) = E\}$$

energiaszint-halmazhoz. A legtöbb gyakorlatban fontos alkalmazásban **az energiaszint-halmazok „sima" és korlátos, koncentrikus $2D - 1$-dimenziós felületekre bontják a fázisteret**:

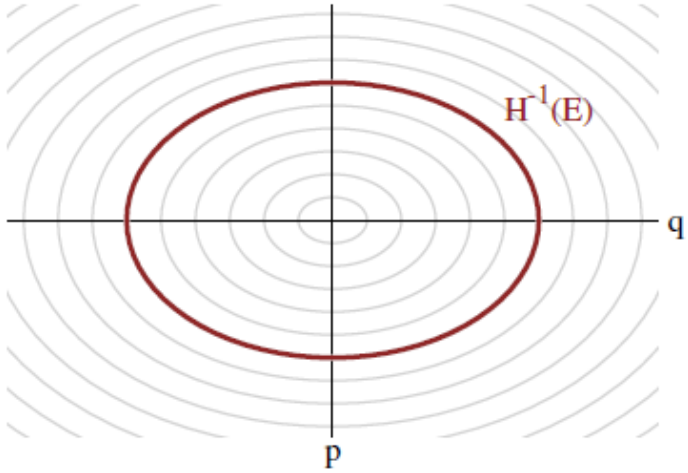


FIG 21. *Phase space naturally decomposes into level sets of the Hamiltonian, $H^{-1}(E)$. Instead of specifying a point in phase space with its position and momentum, we can specify it with an energy, E, and its position on the corresponding level set, $\theta_E \in H^{-1}(E)$.*

Állítás: A fázistér kanonikus eloszlása eleget tesz a

$$\pi(q,p) = \pi(\theta_E|E) \cdot \pi(E)$$

mikrokanonikus felbontásnak. Az energiaszint-halmazokon értelmezett $\pi(\theta_E|E)$ feltételes eloszlások neve **mikrokanonikus eloszlás**.

A Hamilton-Markov átmenet úgy is elmondható, hogy a **fázistérbe lépés** (1) után **kisorsolunk egy energiaszintet és ezen maradva a Hamilton-egyenleteket kiintegráljuk** (elrepülünk a fázistér másik részébe) (2), majd az **érkezési állapotot visszavetítjük** a paramétertérbe.

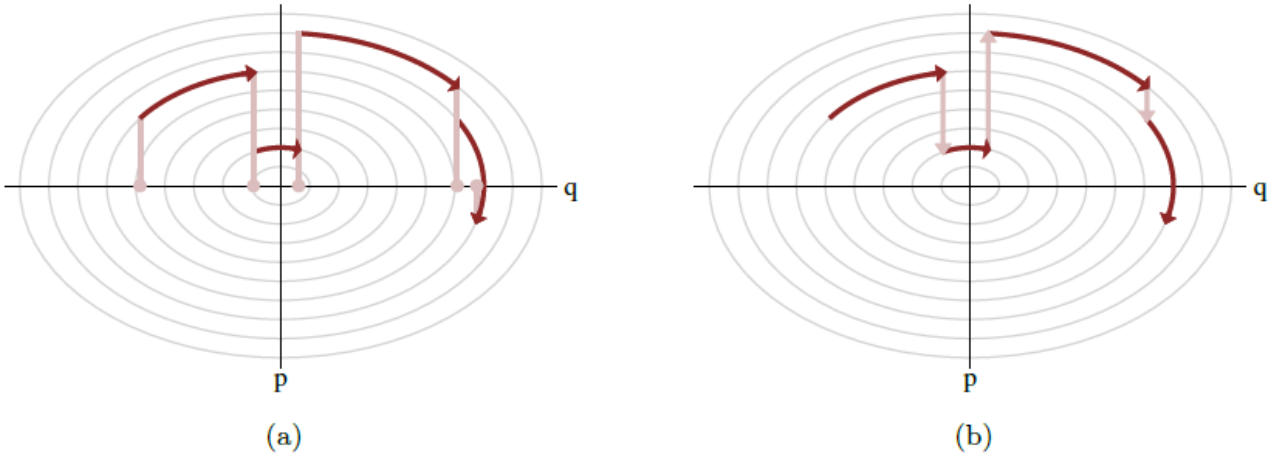


FIG 22. *(a) Each Hamiltonian Markov transition lifts the initial state onto a random level set of the Hamiltonian, which can then be explored with a Hamiltonian trajectory before projecting back down to the target parameter space. (b) If we consider the projection and random lift steps as a single momentum resampling step, then the Hamiltonian Markov chain alternates between deterministic trajectories along these level sets (dark red) and a random walk across the level sets (light red).*

Az iterációkat az ide-oda vetítgetés helyett vizsgálhatjuk magán a fázistéren is. Ekkor egy-egy lépés az **energia véletlen sorsolásából** és az energiaszinten maradva **egy görbén való determinisztikus elmozdulásból** (utazásból) áll.

Az iterációk lassúak, ha a $\pi(E|q)$ feltételes energiaeloszlás sokkal keskenyebb, mint a $\pi(E)$ marginális energiaeloszlás. **Ideális esetben** $\pi(E|q) \approx \pi(E)$ vagyis **az energiaeloszlás független a q aktuális paramétertől**.

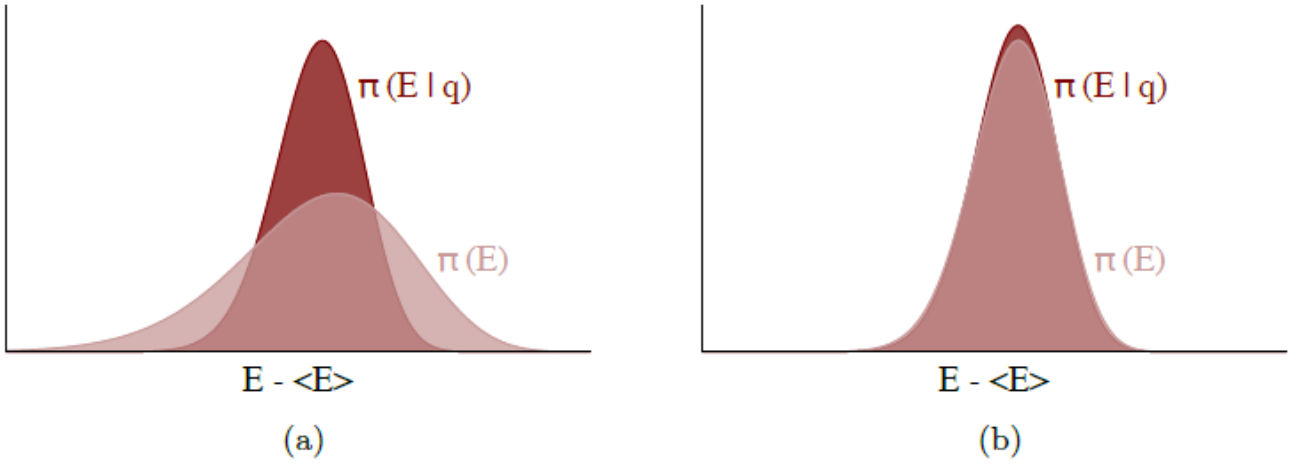


FIG 23. *The momentum resampling in a Hamiltonian Markov transition randomly changes the energy, inducing a random walk between level sets. (a) When the energy transition distribution, $\pi(E \mid q)$ is narrow relative to the marginal energy distribution, $\pi(E)$, this random walk will explore the marginal energy distribution only very slowly, requiring many expensive transitions to survey all of the relevant energies. (b) On the other hand, when the two distributions are well-matched the random walk will explore the marginal energy distribution extremely efficiently.*

A $\pi(E|q)$ paramétertől függő feltételes energiaeloszlás és a $\pi(E)$ teljes fázistéren vett energiaeloszlás eltérését mérhetjük a **hiányzó információ Bayes-i részarányával** (Bayesian fraction of missing information):

$$E-BFMI=\frac{E_{\pi}[Var(E|q)]}{Var(E)}\approx E-\widehat{BFMI}=\frac{\sum_{n=1}^{N}(E_n-E_{n-1})^2}{\sum_{n=0}^{N}(E_n-\bar{E})^2},$$

ahol $E_0, E_1, \dots, E_N, \dots$ a Markov-lánc energiaszintjeinek a sorozata. **Ha ez 0.3 alá kerül, akkor a tapasztalatok szerint a Markov lánc túl lassan járja be a fázisteret.**

A kinetikus energia optimalizásása

Ha a paramétertéren valamilyen skálán értelmezzük az euklideszi távolságfogalmat, akkor ezt egy alkalmas M skálázó mátrixszal át lehet vinni az impulzusokra is, ha a paramétert és az impulzusát skálázó mátrixok egymás inverzei (l. részletesen a Betancourt-dolgozatban). Ebben az esetben $\pi(p|q) = N(p|0, M)$, azaz 0 várható értékű (többdim.) normális, és az **Euklidesz-Gauss kinetikus energia**

$$K(q,p) = \frac{1}{2} p^T \cdot M^{-1} \cdot p + \frac{1}{2} \cdot \log|M| + const.$$

Ekkor „jó" energiamegmaradáshoz jutunk, azaz „szép" energiaszint-halmazokhoz. Általánosabb esetben a p impulzus kovarianciája függhet a q paramétertől, $\pi(p|q) = N\big(p|0, M(q)\big)$. Ekkor kapjuk a **Riemann-Gauss kinetikus energiát**,

$$K(q,p) = \frac{1}{2} p^T \cdot M^{-1}(q) \cdot p + \frac{1}{2} \cdot \log|M(q)| + const.$$

Az ennek megfelelő HMC neve **Riemann-Gauss Monte Carlo** módszer (Girolami & Calderhead 2011).

A Hamilton Monte Carlo módszer implementálása

Az energiaszint kisorsolása után a trajektóriát egy differenciálegyenlet-rendszer numerikus megoldása szolgáltatja. Ez azonban **sok esetben pontatlan, különösen, ha a céleloszlás görbülete nagy**. Az ún. szimplektikus integrátorok (melyek közül az rstan a **leapfrog** – békaugrás – integrátor-t alkalmazza) megtartják a Hamilton-függvényt, így átlagosan nem térhetnek el messze a pontos trajektóriától.

A posterior eloszlást pontosan elérhetjük, ha **a numerikusan integrált trajektóriát csak egy javaslatnak tekintjük, melynek elfogadásáról a Metropolis-Hastings algoritmussal döntünk**. Az R-ben az elfogadási valószínűséget az iterációk első (adaptációs, bemelegítő) szakaszában állítja be a stan, de mi is befolyásolhatjuk a control paraméter beállításával.

A kiintegrált trajektóriákból előre adott (az adaptációs szakaszban megfelelően beállított) **egyforma (ekvidisztans) lépésközzel értékeket választunk. Ezek lesznek a Hamilton Markov lánc (HMC) egymást követő elemei.**

További optimalizációk

A leapfrog integrálási szakaszokat kötegelve hajtja végre az algoritmus: ha már van egy n leapfrog-ból álló trajektóriánk, akkor ehhez újabb n lépéses leapfrog-szakaszt ragasztunk, előbb kisorsolva, hogy a trajektória melyik végét hosszabbítsuk meg, azaz hogy időben előre, vagy időben visszafelé haladjunk. Ezt csináljuk egymás után treedepth alkalommal. Ezek szerint **a treedepth és a leapfrog paraméterek között nagyjából a**

$$\boldsymbol{leapfrog = 2^{treedepth}}$$

összefüggés áll fenn.

No U-turn sampler (NUTS)

Ez **a jelenleg legjobbnak tartott HMC algoritmus**. Szabályt ad arra, hogy meddig hosszabbítsuk a trajektóriákat. Cél, hogy egymástól minél távolabb kerüljenek a fázistérben az új iterációk, de ne túl távol, azaz **a trajektória a két végén éppen kanyarodjon visszafelé**. Így egy-egy trajektória kb. az energiaszint felét járja be. A szerzők (Hoffman & Gelman 2014) a visszakanyarodást úgy definiálják, hogy **a trajektória a kezdetét és végét összekötő szakasszal zárjon be hegyesszöget**:

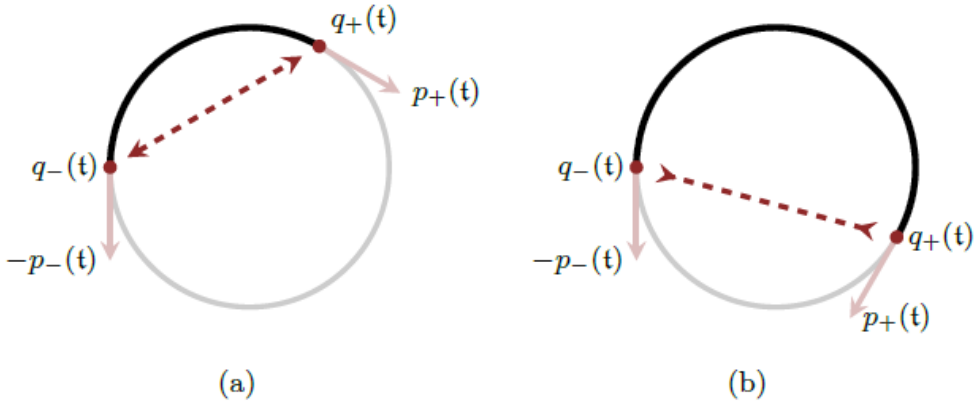


FIG 42. *(a) When the No-U-Turn termination criterion is not satisfied, expanding the trajectory in either direction typically extends the trajectory further across the energy level set (grey) towards unexplored neighborhoods. (b) When the No-U-Turn termination criterion is satisfied, however, further expansion typically contracts the boundaries of the trajectory towards each other and neighborhoods that have already been explored. Stopping when the No-U-Turn termination criterion is just satisfied typically ensures that each trajectory efficiently explores about half of the energy level set. Although this can be too aggressive, it works well empirically and is far superior to any scheme with only a static integration time.*

Képlettel a No-U-Turn megállítási kritérium (**Euklidesz-Gauss** Monte Carlo esetén):

$$p_+(t)^T \cdot \big(q_+(t) - q_-(t)\big) < 0$$

és ugyanakkor

$$p_-(t)^T \cdot \big(q_-(t) - q_+(t)\big) < 0.$$

(**Riemann-Gauss** Monte Carlo esetén annyiban módosul a feltétel, hogy a $q_+(t) - q_-(t)$ vektort az impulzusok trajektória menti integráljával helyettesítjük.)

Az alábbi link a különféle MCMC algoritmusok működését demonstrálja.

https://chi-feng.github.io/mcmc-demo/app.html#RandomWalkMH,banana

Jó kísérletezést!