

Statisztikai tanulás az idegrendszerben, 2019.

Algoritmikus következtetés valószínűségi modellekben

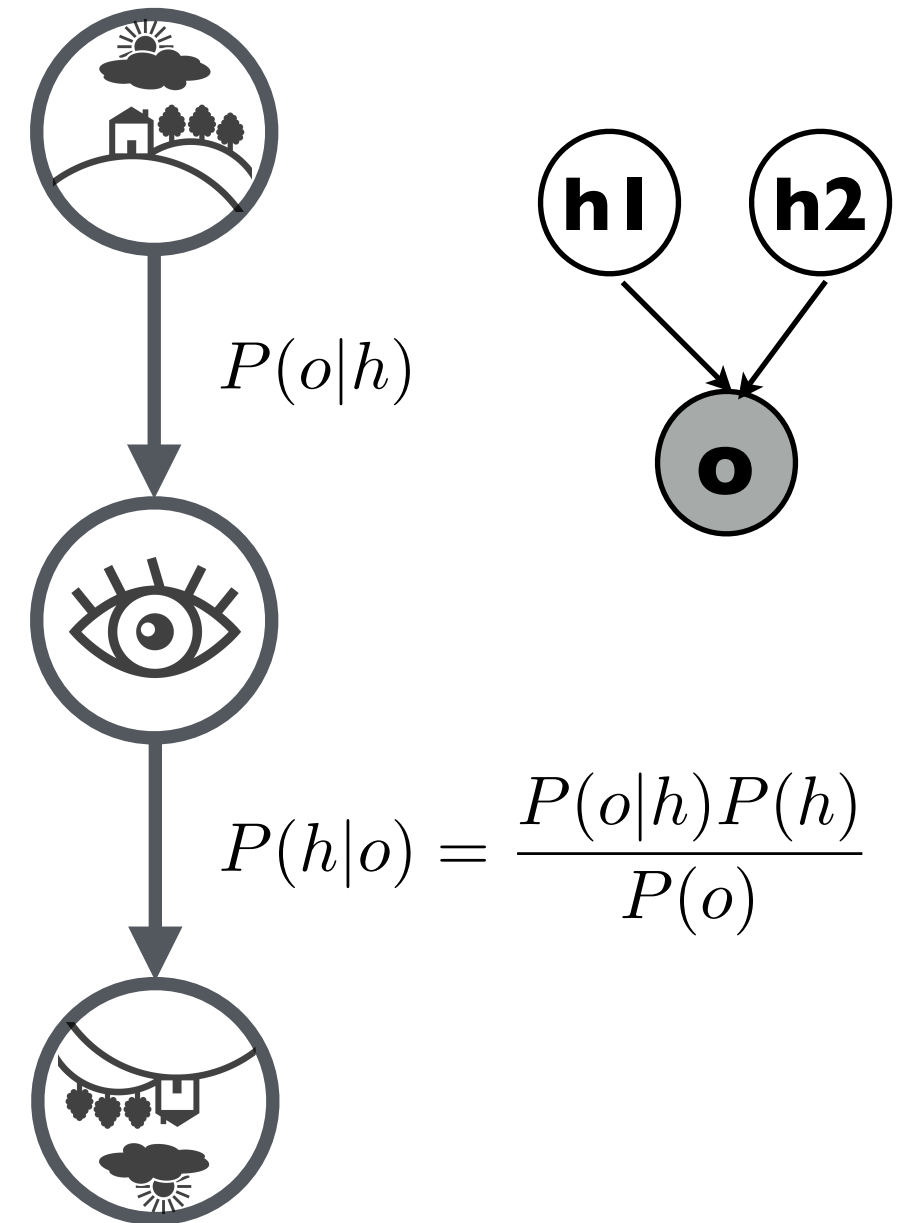
Bányai Mihály

banyai.mihaly@wigner.mta.hu

<http://golab.wigner.mta.hu/people/mihaly-banyai/>

Megfigyelt és rejtett változók

- bizonyos változókról rendelkezünk adatokkal
 - jelölés: sötét kör
- a többi csak a jelenség struktúrájáról alkotott feltételezéseinket reprezentálja
 - látens, rejtett változók: üres kör
- Mire lehetünk kíváncsiak
 - poszterior eloszlás a látensek fölött
 - marginális poszteriorok
 - a prediktív eloszlás: a megfigyelt változók marginális eloszlása

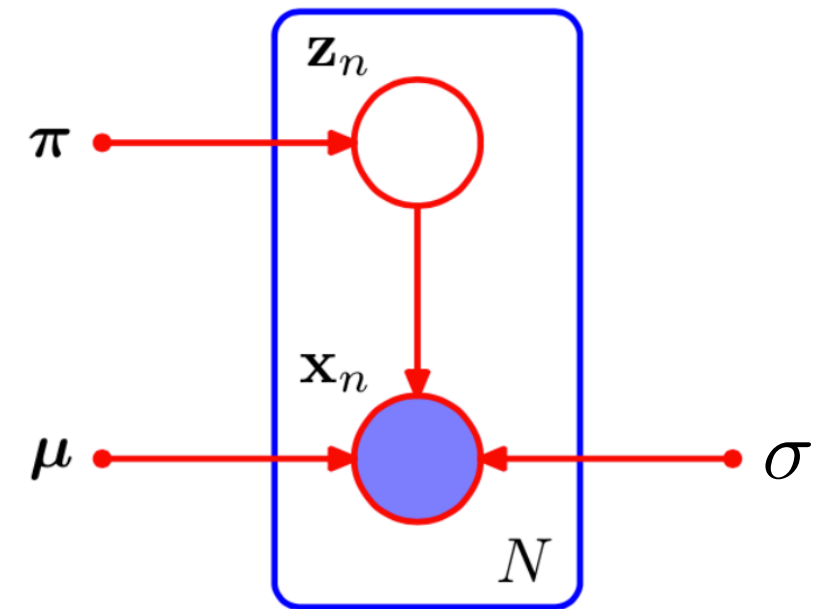
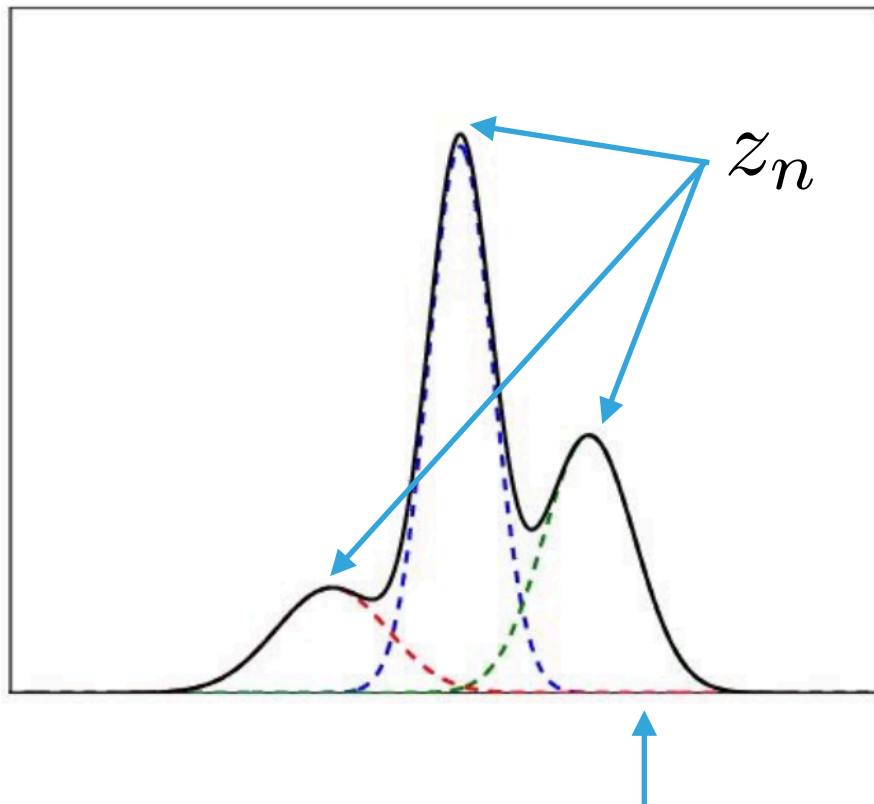


MIXTURE OF GAUSSIANS

- Valószínűségi modell:

$$p(z_n = i | \pi) = \pi_i$$

$$p(x_n | \mu, \sigma, z_n) = \mathcal{N}(x; \mu_{z_n}, \sigma_{z_n})$$

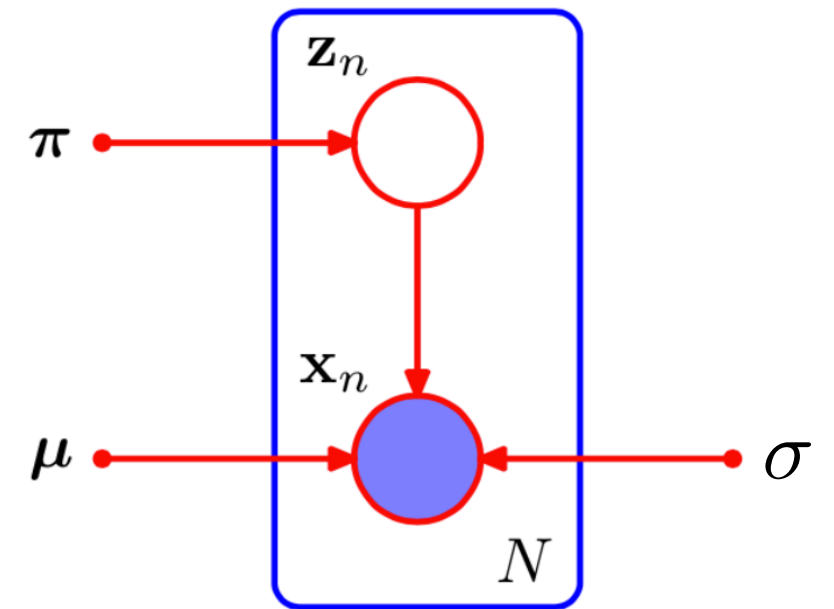


MIXTURE OF GAUSSIANS

- Valószínűségi modell:

$$p(z_n = i | \pi) = \pi_i$$

$$p(x_n | \mu, \sigma, z_n) = \mathcal{N}(x; \mu_{z_n}, \sigma_{z_n})$$



- Poszterior:

$$p(z, \pi, \mu, \sigma | x) = \frac{p(x | z, \mu, \sigma) p(z | \pi) p(\mu, \sigma, \pi)}{p(x)}$$

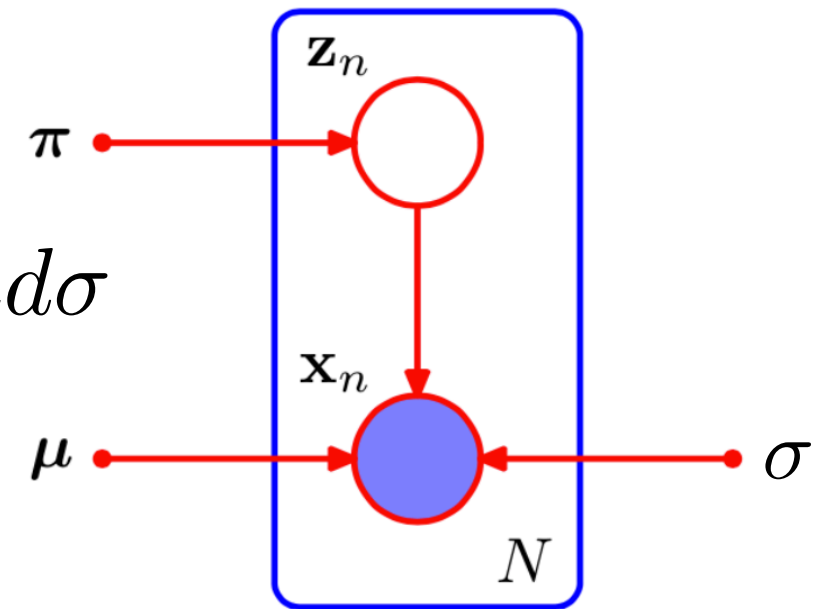
$$p(x | z, \mu, \sigma) = \prod_{n=1}^N p(x_n | z_n, \mu, \sigma)$$

$$= \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma_{z_n}^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma_{z_n}^2}\right)$$

MIXTURE OF GAUSSIANS

$$p(\pi, \mu, \sigma | x) = ? \sum p(z, \pi, \mu, \sigma | x)$$

$$p(x) = ? \int \int \int \sum_z p(x, \pi, \mu, \sigma, z) d\pi d\mu d\sigma$$



► Posztterior:

$$p(z, \pi, \mu, \sigma | x) = \frac{p(x | z, \mu, \sigma) p(z | \pi) p(\mu, \sigma, \pi)}{p(x)}$$

$$p(x | z, \mu, \sigma) = \prod_{n=1}^N p(x_n | z_n, \mu, \sigma)$$

$$= \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma_{z_n}^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma_{z_n}^2}\right)$$

érzékelés

**percepció
(észlelés)**

jelenlegi
megfigyelésből
kikövetkeztetett
információ

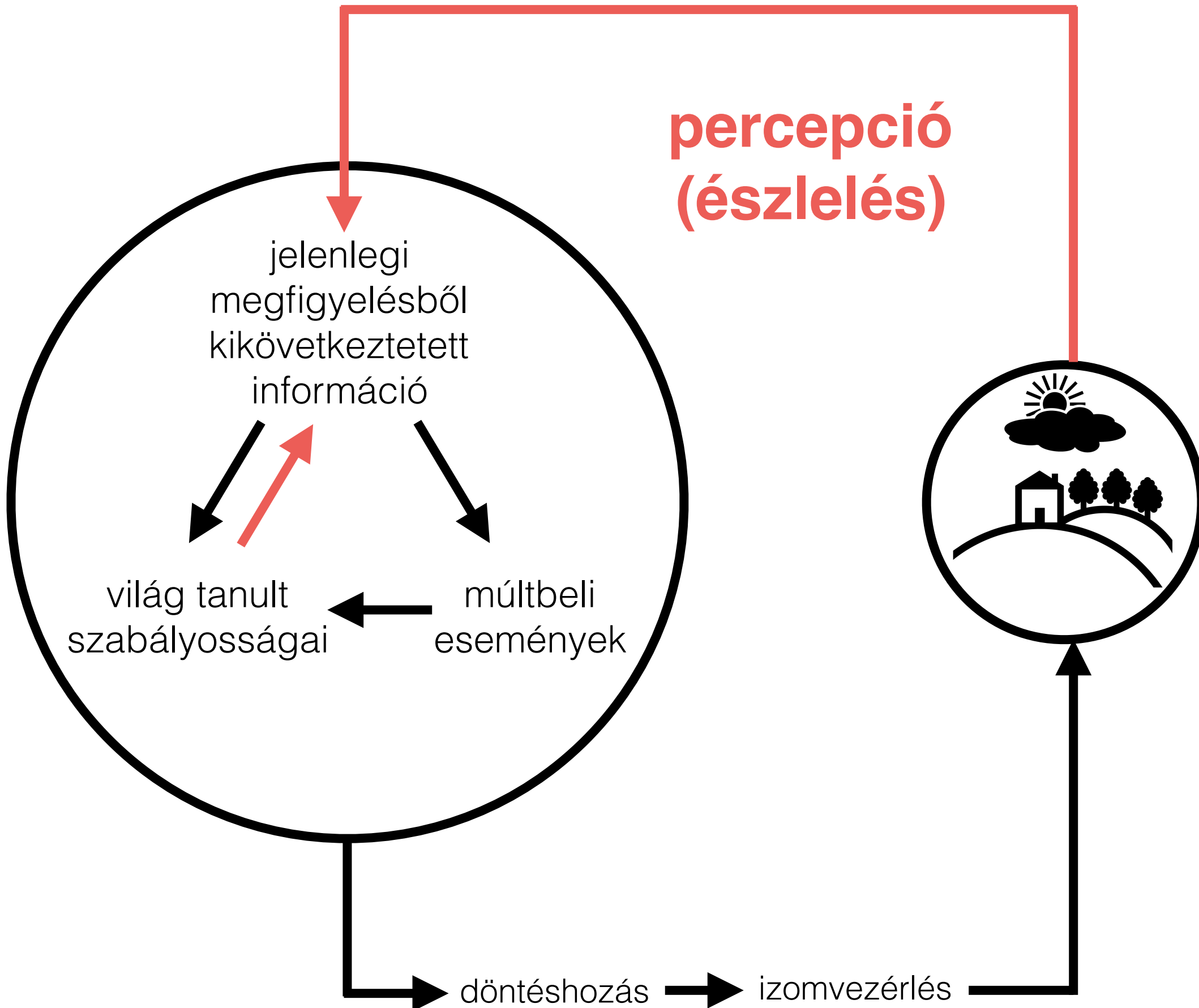
világ tanult
szabályosságai

múltbeli
események

környezet

döntéshozás → izomvezérlés

cselekvés



Poszterior közelítő becslése

- Sztochasztikus
 - mintákat veszünk belőle
 - véletlenszámgenerátorra van szükség
 - a minták számától függ a pontosság
- Determinisztikus
 - egyszerűbben kezelhető eloszlásokkal közelítjük
 - nem használunk véletlenszámokat
 - a közelítő eloszlás formájától függ a pontosság

MINTAVÉTELEZÉS

► Problémák:

- **1. Független mintákat szeretnénk venni egy eloszlásból (pl. a poszteriorból)**

$$x^{(r)} \sim P(x)$$

- **2. Egy tetszőleges függvény várhatóértékének kiszámítása**

$$\mathbb{E}(f(x)) = \langle f \rangle_{P(x)} = \int f(x) P(x) dx$$

MINTAVÉTELEZÉS

- ▶ Tegyük fel, hogy valaki egy (bonyolult) kockajátékra invitál.
- ▶ Megadja, hogy milyen esetben mennyi nyereményt fizet.
- ▶ Azt is megmondja, hogy mennyibe kerül egy játék.
- ▶ Azt szeretnénk eldönteni, hogy akarunk-e játszani.
- ▶ A játék várhatóértékét szeretnénk kiszámolni.
- ▶ **Ötlet:** játsszuk le a játékot (magunkban) sokszor és vegyük az eredmények átlagát.

MINTAVÉTELEZÉS

- ▶ Játsszuk le a játékot (magunkban) sokszor és vegyük az eredmények átlagát.

Nyeremény várhatóértéke

$$F = \mathbb{E}(f(x))$$

Nyeremény egy x kimenetelnél

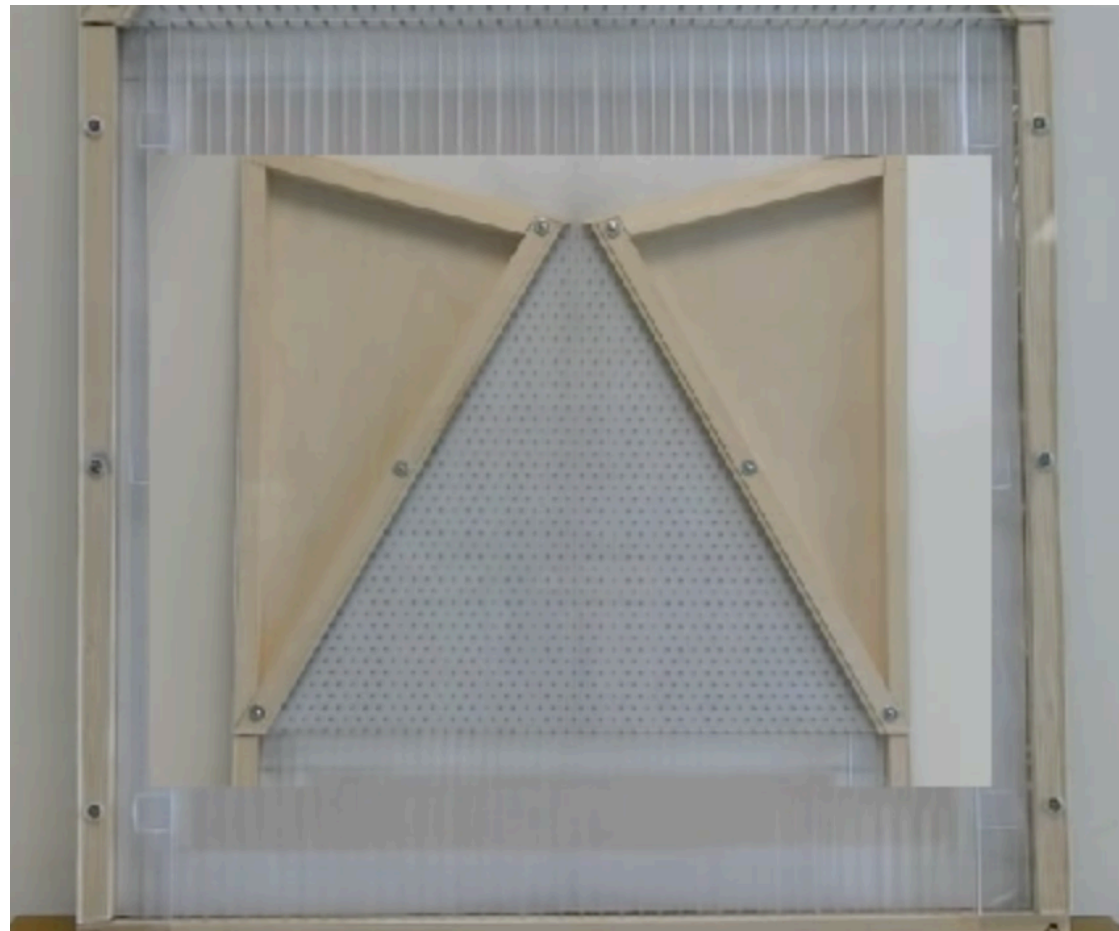
$$x_n \sim P(x)$$

Minták a játékból

$$\hat{F} = \sum_{n=1}^N \frac{1}{N} f(x_n)$$

Becslés a várhatóértékre

GALTON-DESZKA

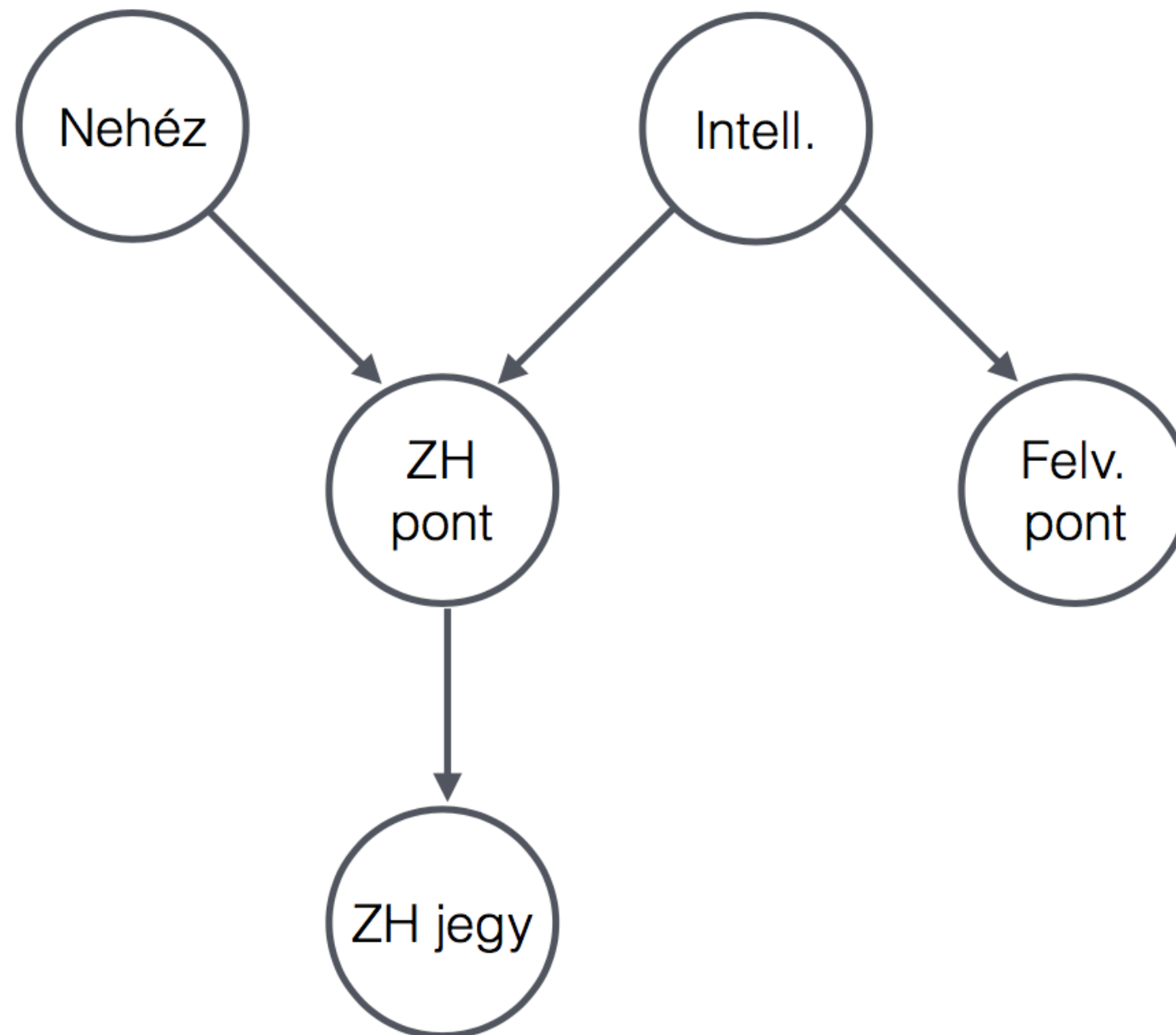


<https://www.youtube.com/watch?v=EMK5WPRjGgA>

MILYEN ELOSZLÁSOKBÓL TUDUNK KÖNNYEN MINTÁKAT VENNI?

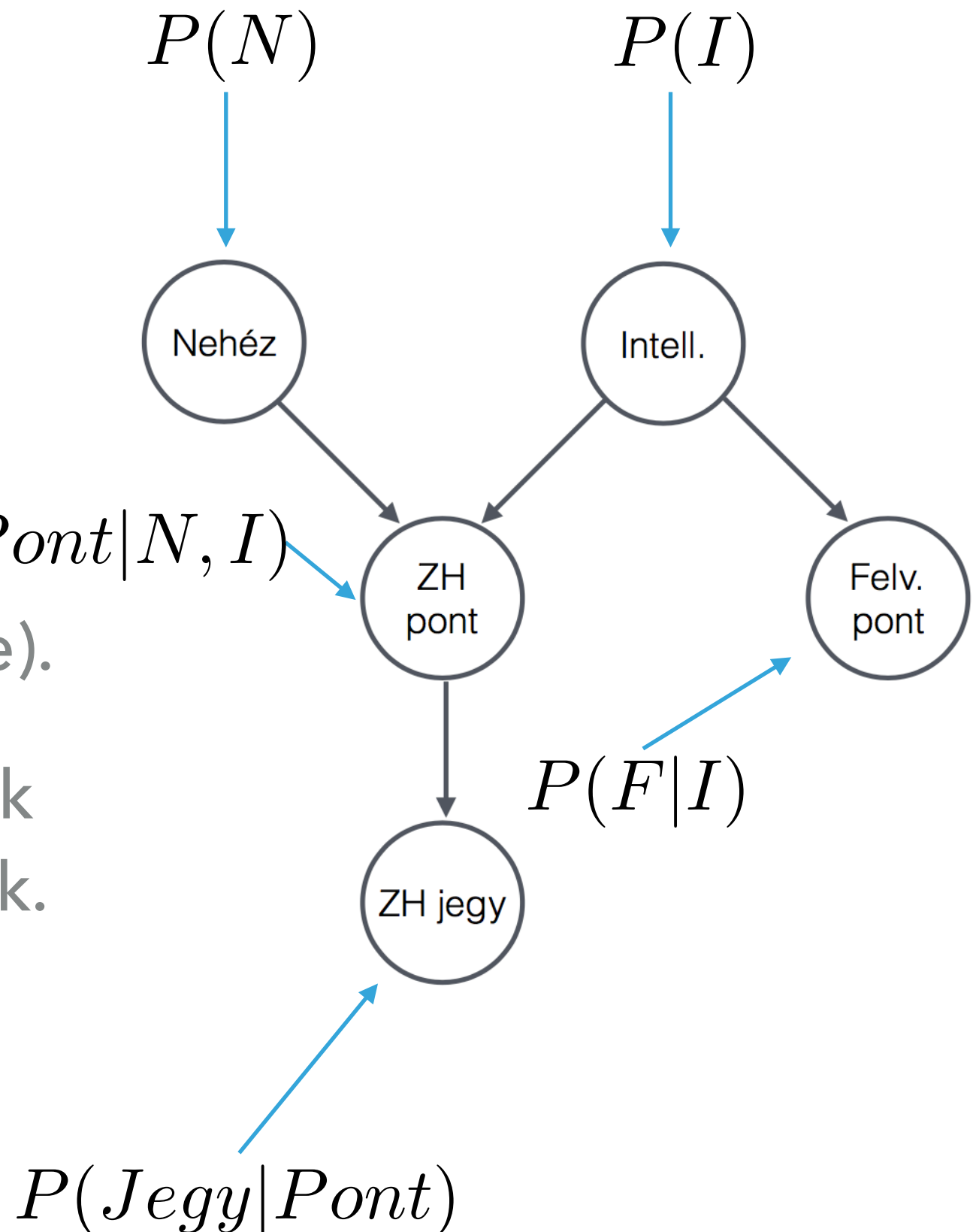
- ▶ Dobókocka (egyenletes eloszlás egy hat elemű halmazon)
- ▶ Egyenletes eloszlás a $[0,1]$ intervallumon (pseudo-random generátor)
- ▶ Pénzérme feldobás
- ▶ Tudjuk-e ezeket használni, hogy más eloszlásokból vegyünk mintákat?

ANCESTRAL SAMPLING



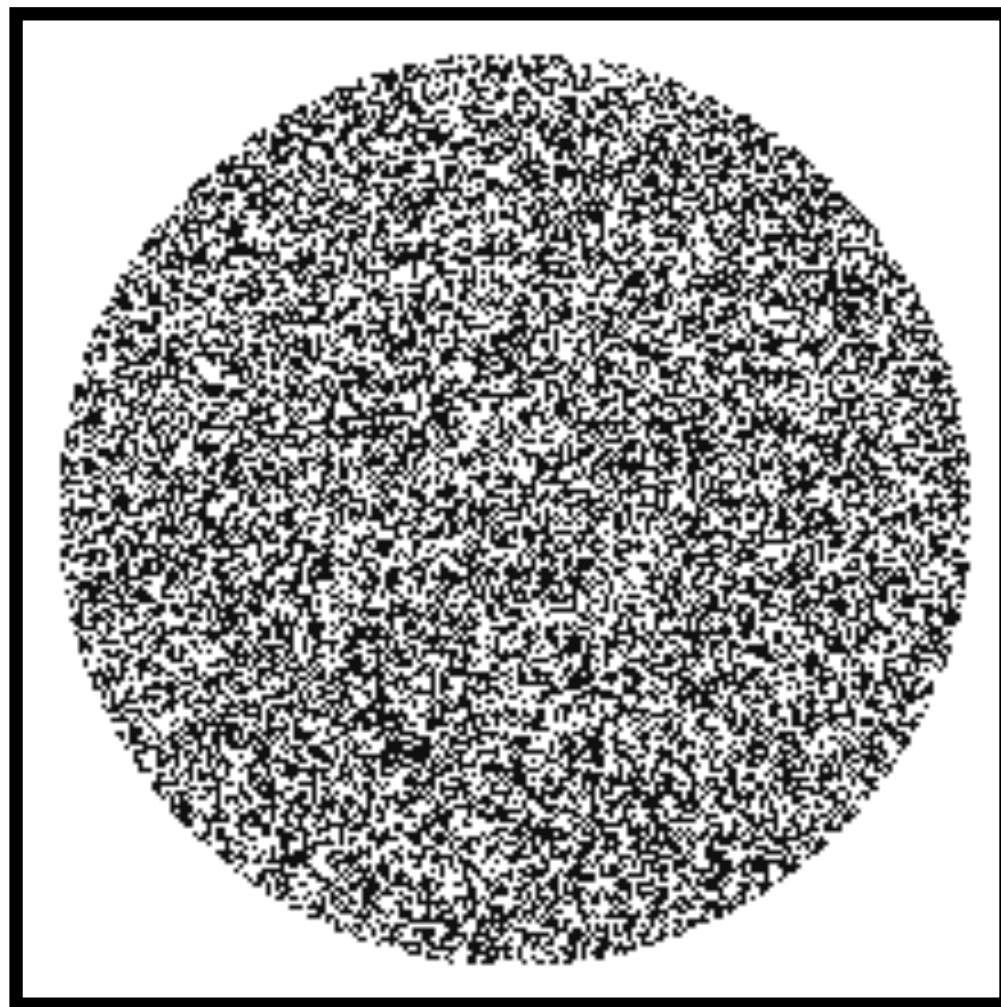
ANCESTRAL SAMPLING

- ▶ Szeretnénk egy mintát az együttes eloszlásból.
- ▶ 1. Először mintát veszünk a "legfelső változókból" $P(Pont|N, I)$ (amelyeknek nincsen szülője).
- ▶ 2. Majd azokból, amelyeknek a szüleit már mintavételeztük.
- ▶ 2. lépést ismételjük



REJECTION SAMPLING

- ▶ Hogyan mintavételezzünk pontokat egyenletesen egy körlapon?



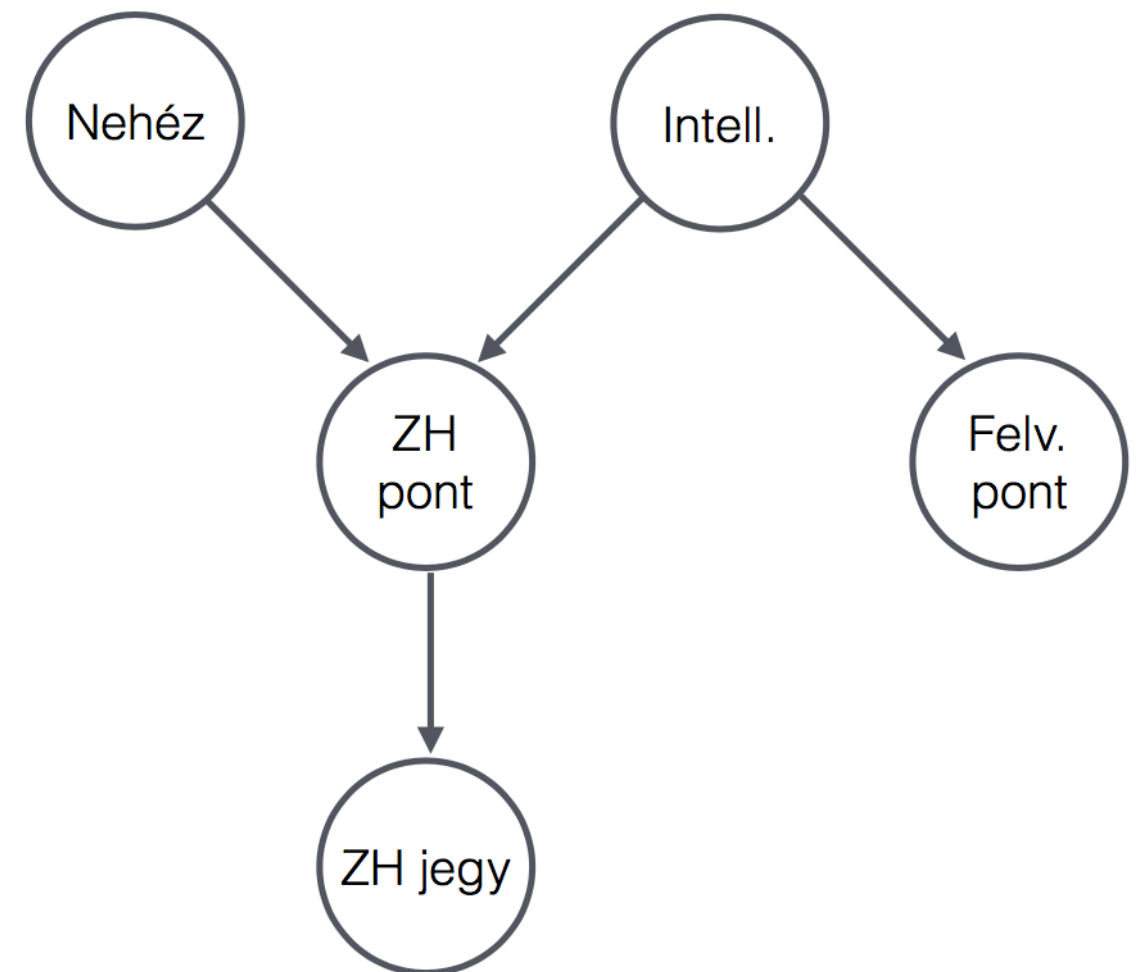
ANCESTRAL+REJECTION SAMPLING

- ▶ Ha tudunk mintákat venni $P(N, I, F, Pont, Jegy)$ eloszlásból, hogy tudunk inferenciát végezni, vagyis, kondicionálisokat számolni?

- ▶ pl: $P(I | Jegy = 5) = ?$

- ▶ Rejection sampling!

Azokat a mintákat nézzük, amelyekben a Jegy a megadott értékű



MINTAVÉTELEZÉS

► Problémák:

- 1. Független mintákat szeretnénk venni egy eloszlásból (pl. a poszteriorból)

$$x^{(r)} \sim P(x)$$

- Tegyük fel, hogy csak P valamely többszöröséhez férünk hozzá:

$$P^*(x) = Z \cdot P(x)$$

$$P(x|Data) \propto \boxed{P(Data|x)P(x)} = P^*(x)$$

$$P(Data) = \int \int \int \sum_z P(\mu, \sigma, \pi, z, Data) d\mu d\sigma d\pi$$

MINTAVÉTELEZÉS

- ▶ Problémák:

- ▶ 1. Független mintákat szeretnénk venni egy eloszlásból (pl. a poszteriorból)

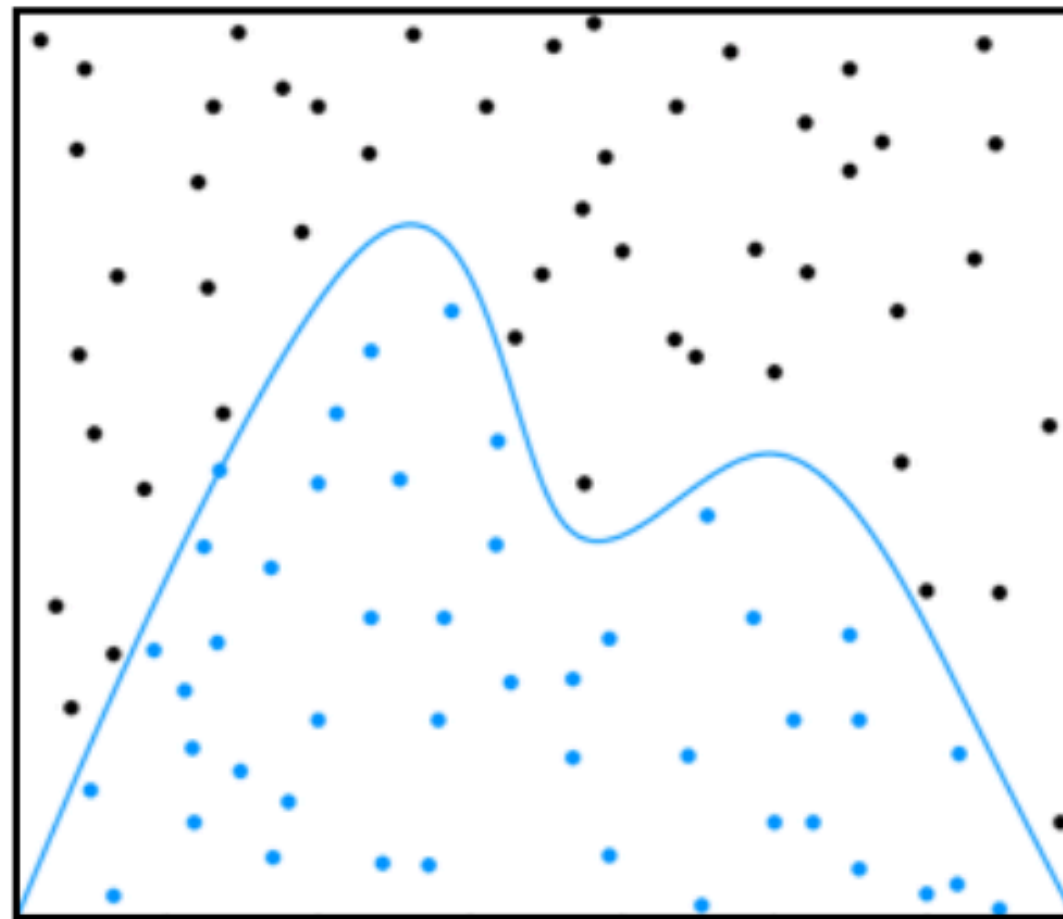
$$x^{(r)} \sim P(x)$$

- ▶ Tegyük fel, hogy csak P valamely többszöröséhez férünk hozzá:

$$P^*(x) = Z \cdot P(x)$$

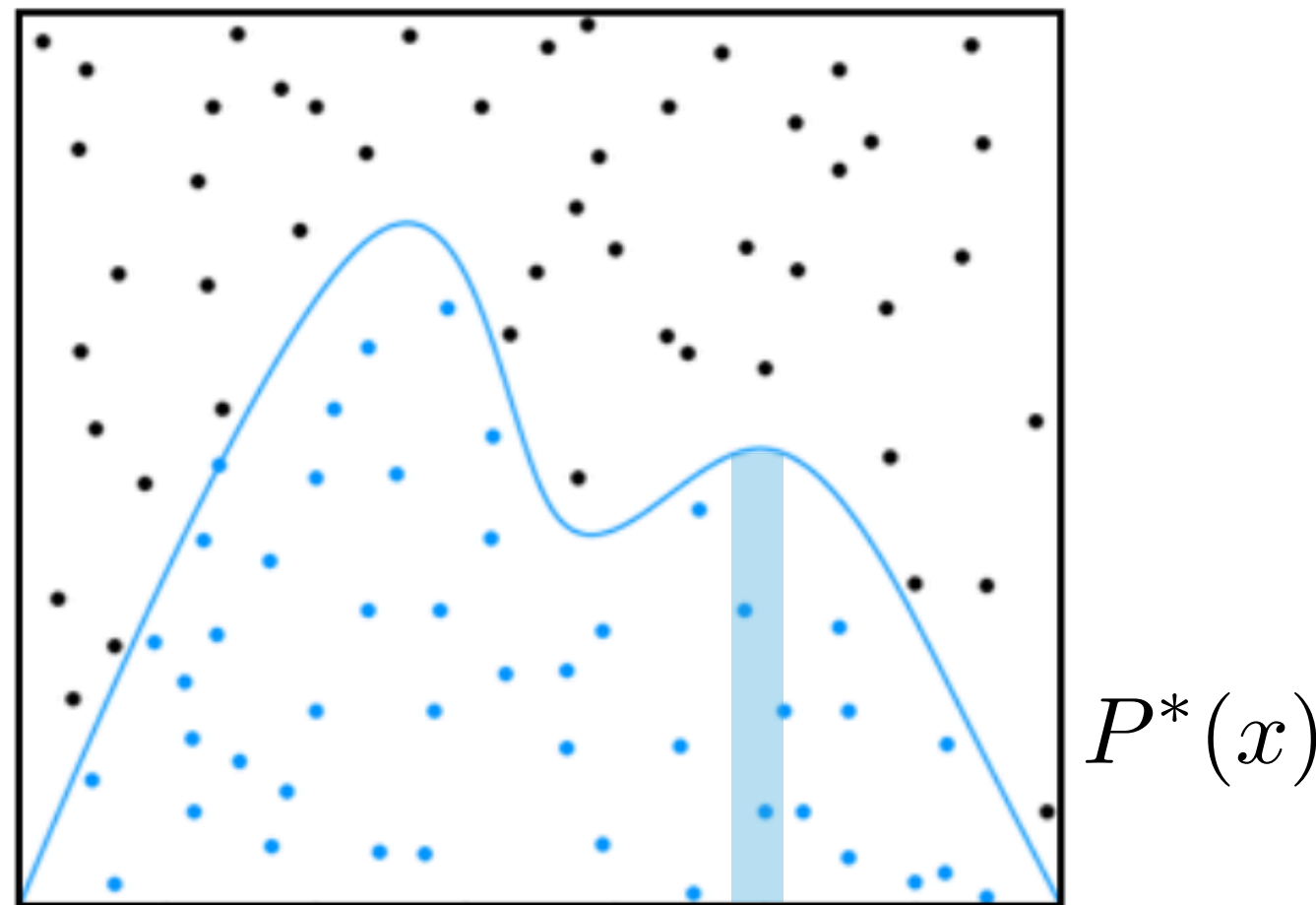
- ▶ Lényeg: ki tudjuk értékelni $P^*(x)$ -et tetszőleges x -ben.

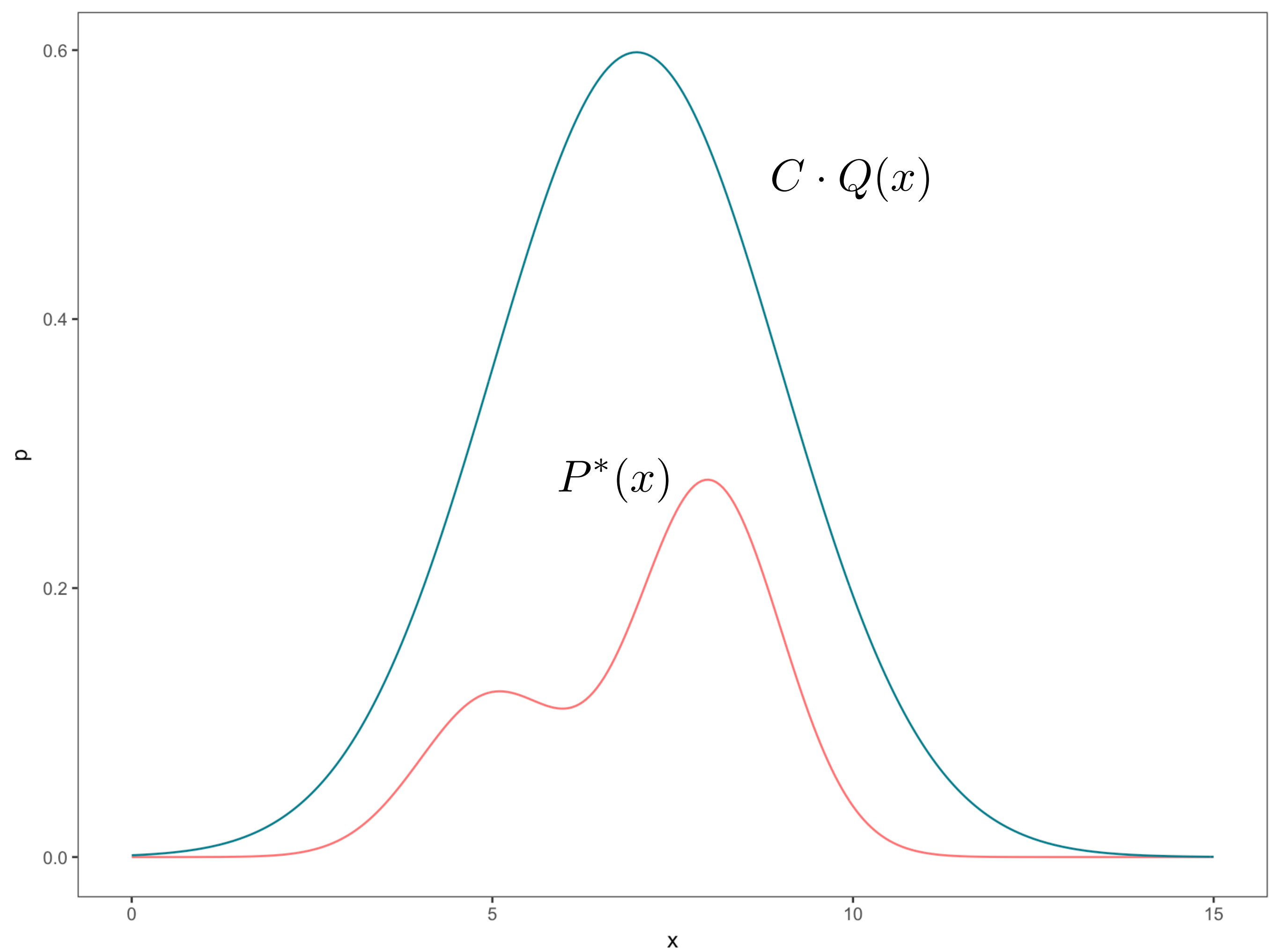
REJECTION SAMPLING

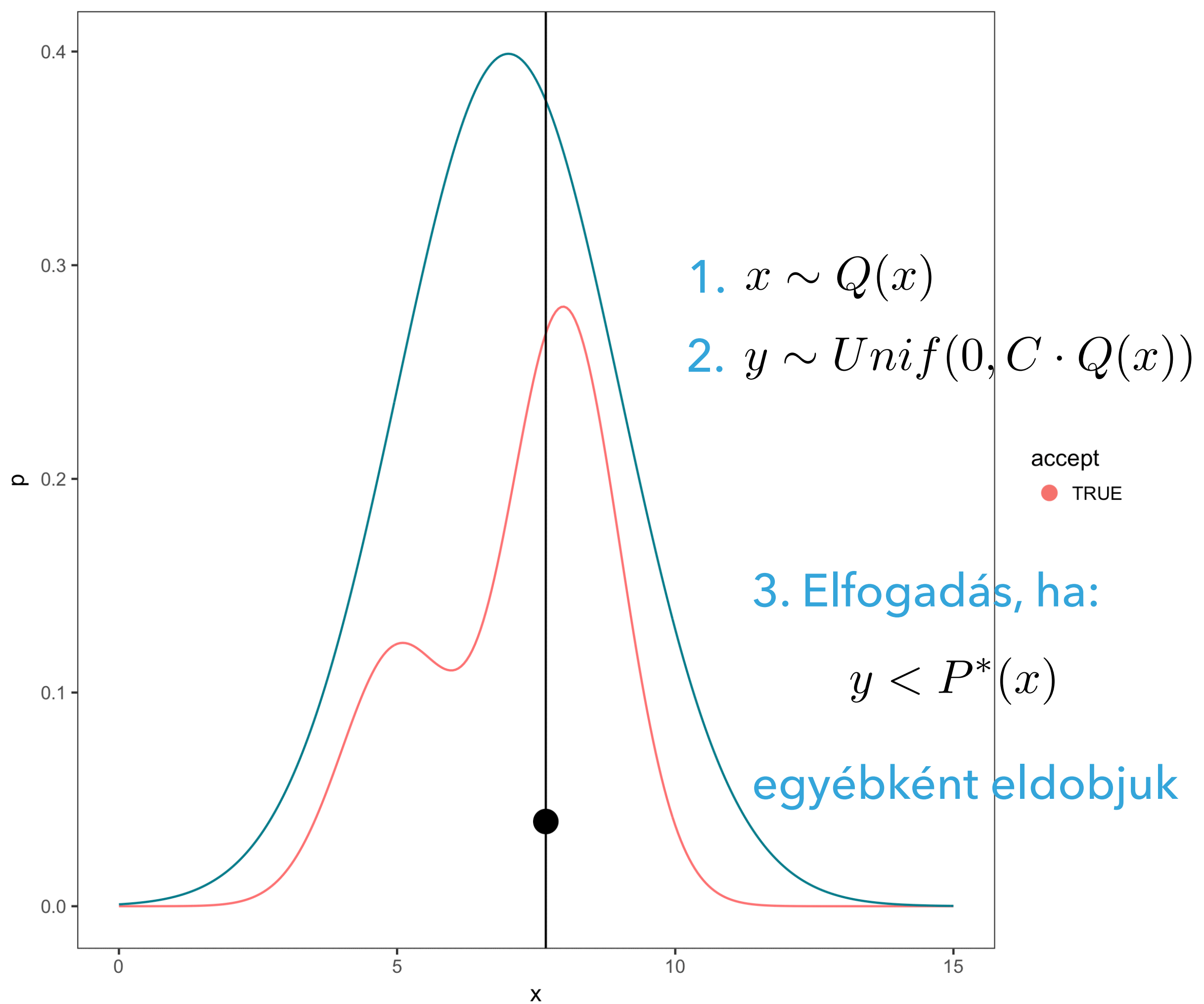


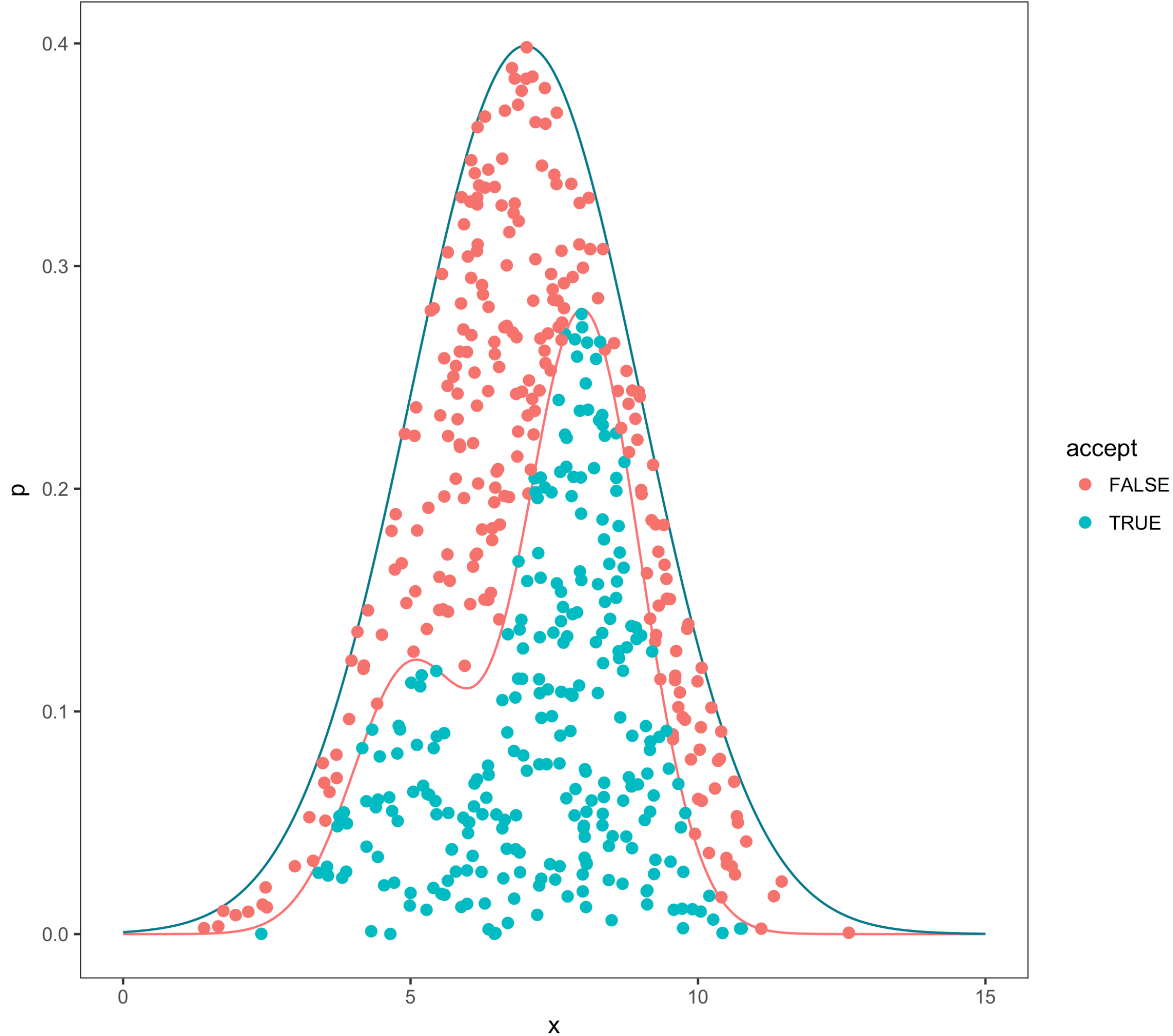
$$P^*(x)$$

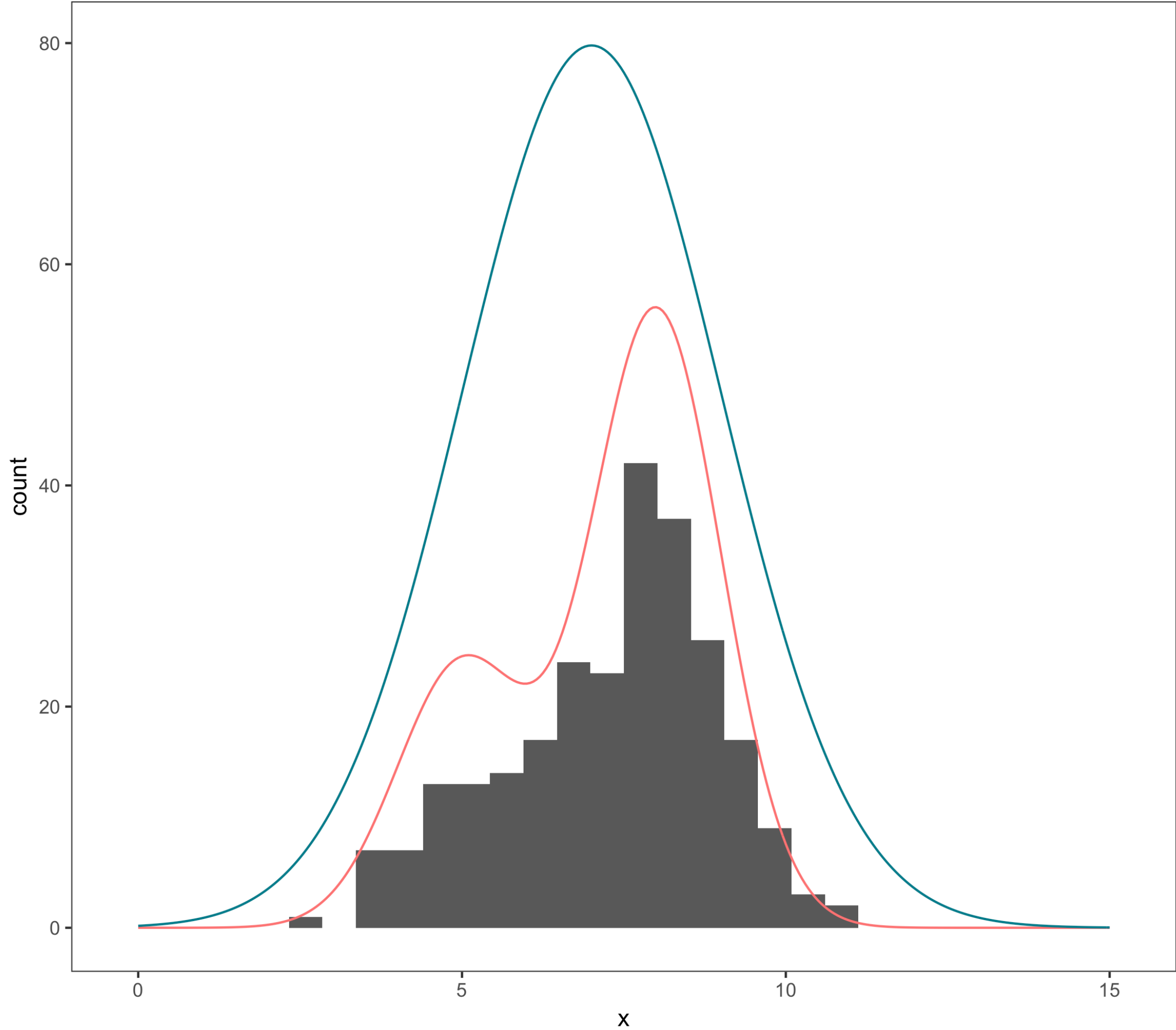
REJECTION SAMPLING











REJECTION SAMPLING PROBLÉMA

- ▶ Dimenzió növekedésével csökken az elfogadott minták aránya
- ▶ Nehéz az eloszlást felülről becsülni.

IMPORTANCE SAMPLING

- ▶ Egy függvény várható értékét szeretnénk kiszámolni

$$E_x(f(x)) = \int f(x)p(x)dx$$

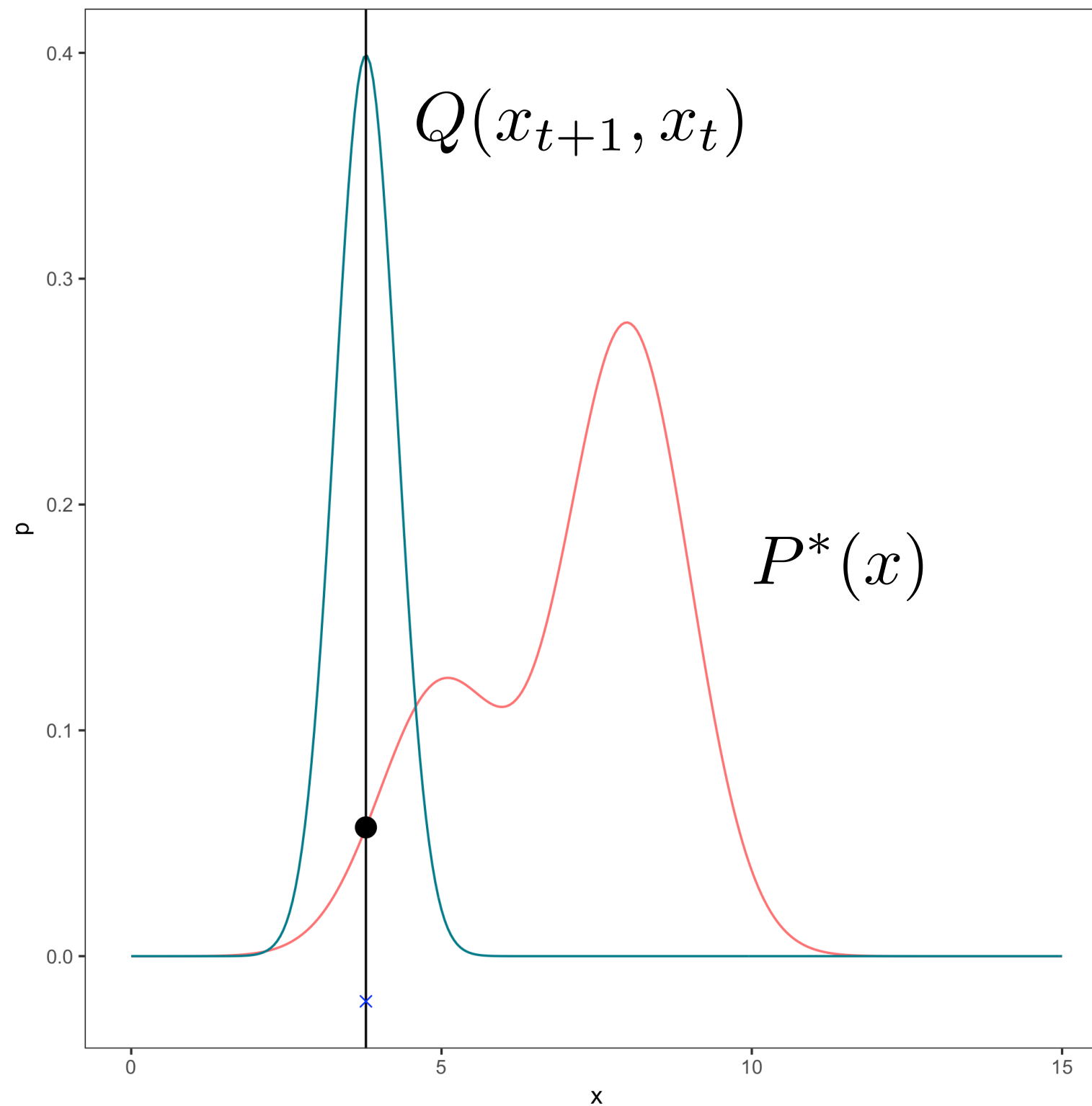
- ▶ Nem tudunk $p(x)$ eloszlásból mintavételezni, de van egy mintavételező eljárásunk, ami $q(x)$ -ből mintavételez.
- ▶ A várhatóértéket becsülhetjük úgy, hogy:

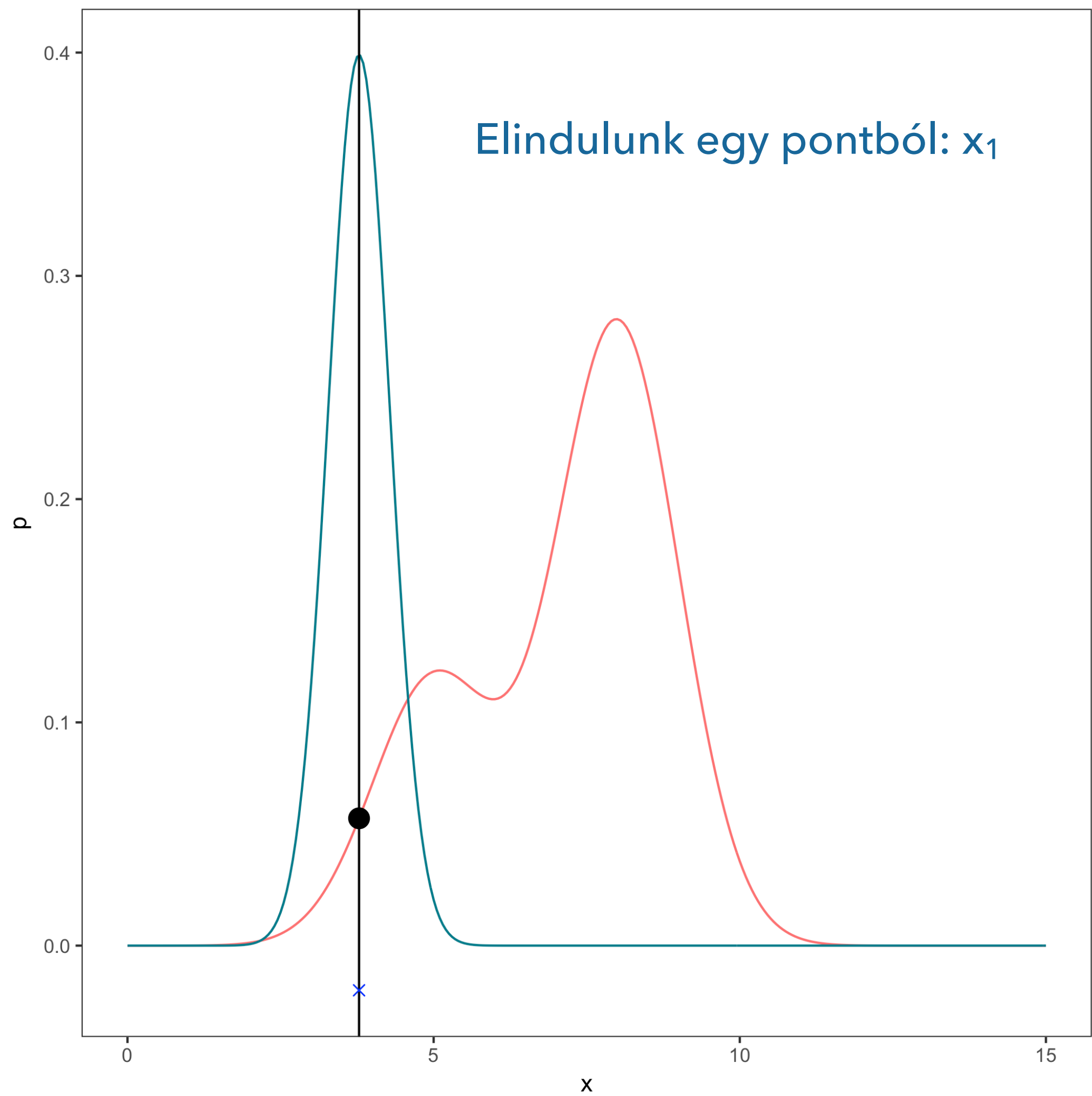
$$E_x(f(x)) \approx \sum_{i=1}^N \frac{1}{N} f(x_i) \frac{p(x_i)}{q(x_i)}$$

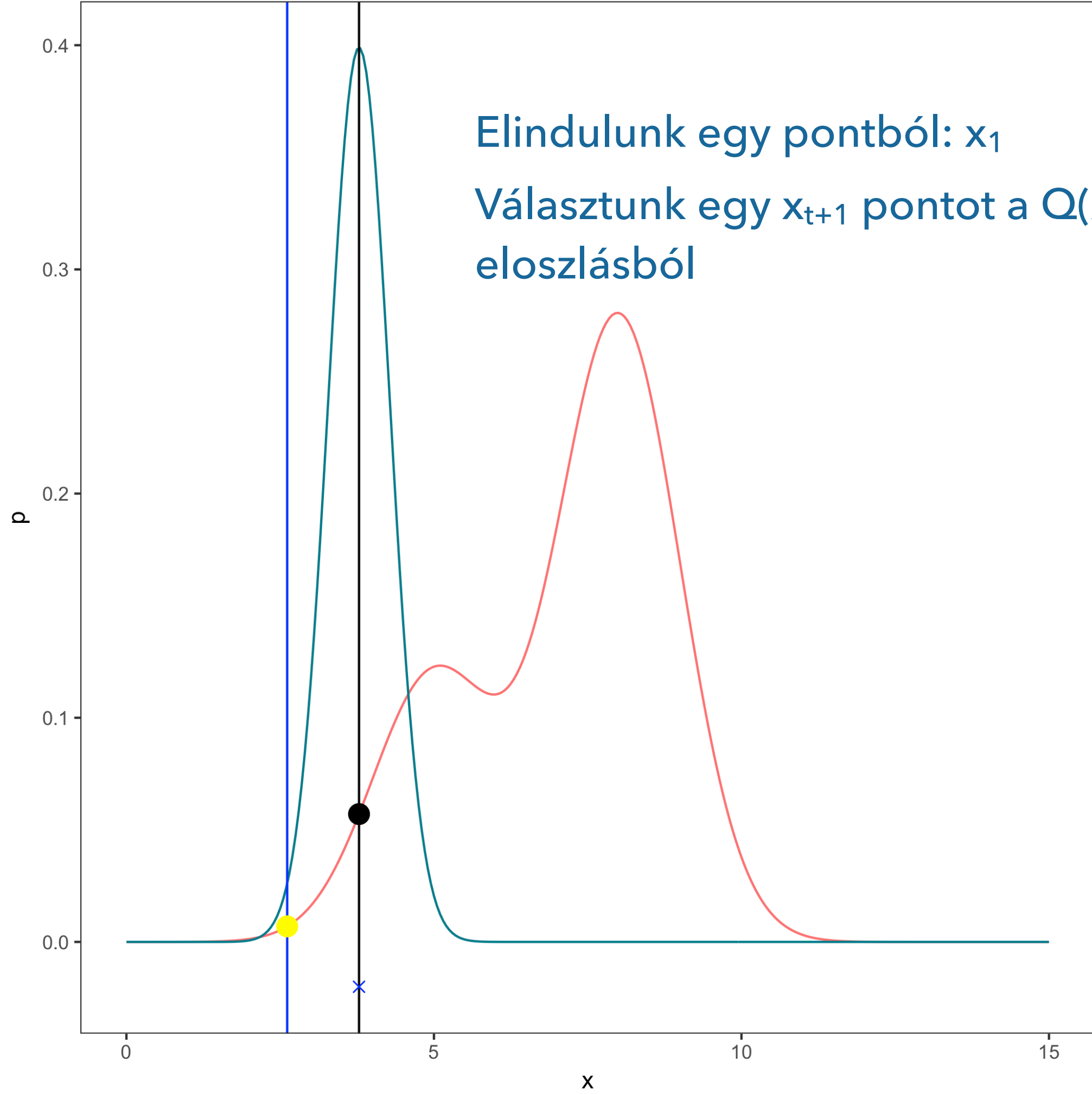
MARKOV CHAIN MONTE CARLO

- ▶ Ahhoz, hogy a dimenzionalitás problémáját kikerüljük, megpróbálhatjuk felhasználni az eddigi mintáinkban rejlő információt.
- ▶ Viszont elveszítjük a függetlenséget!

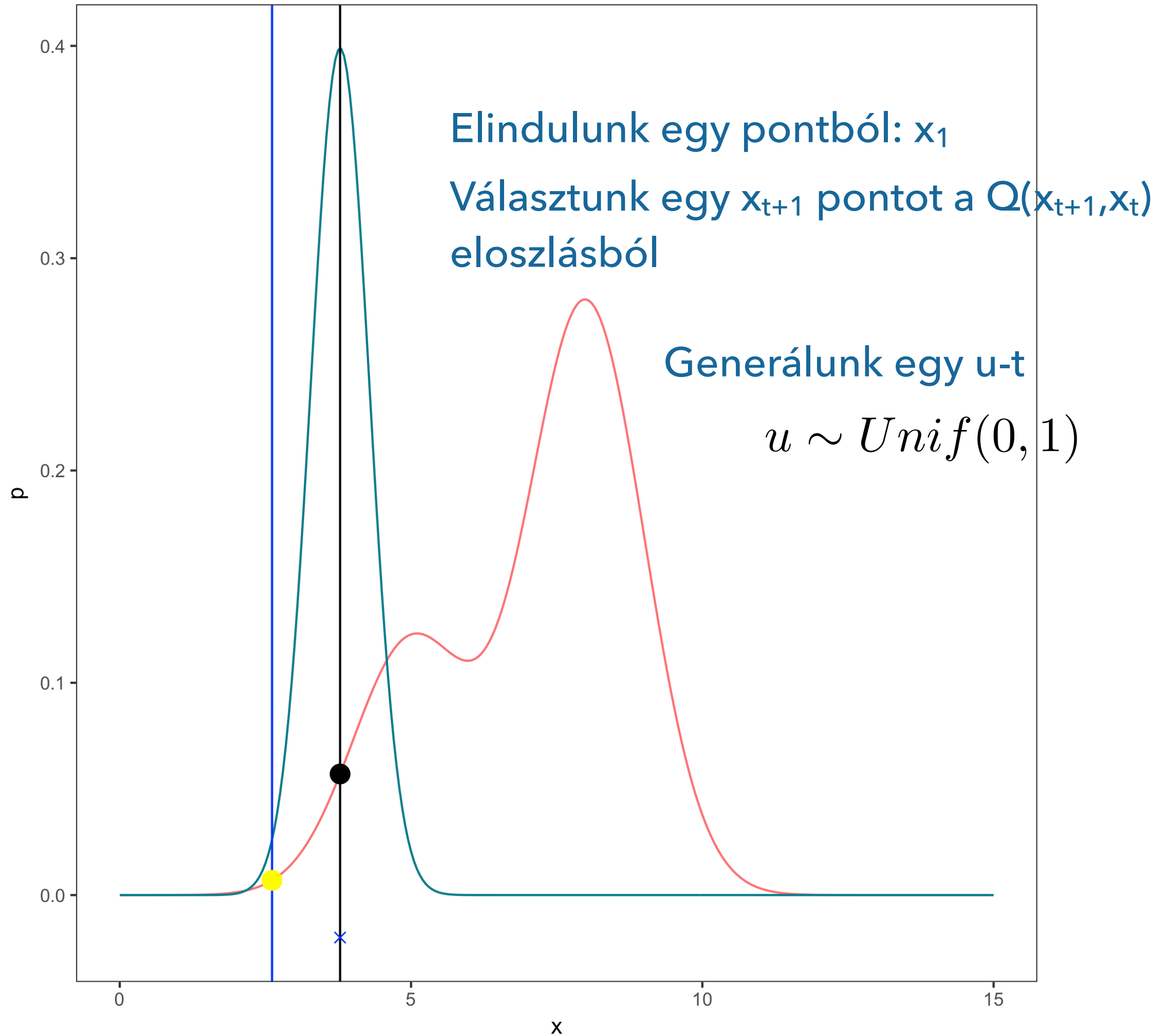
METROPOLIS-HASTINGS ALGORITMUS

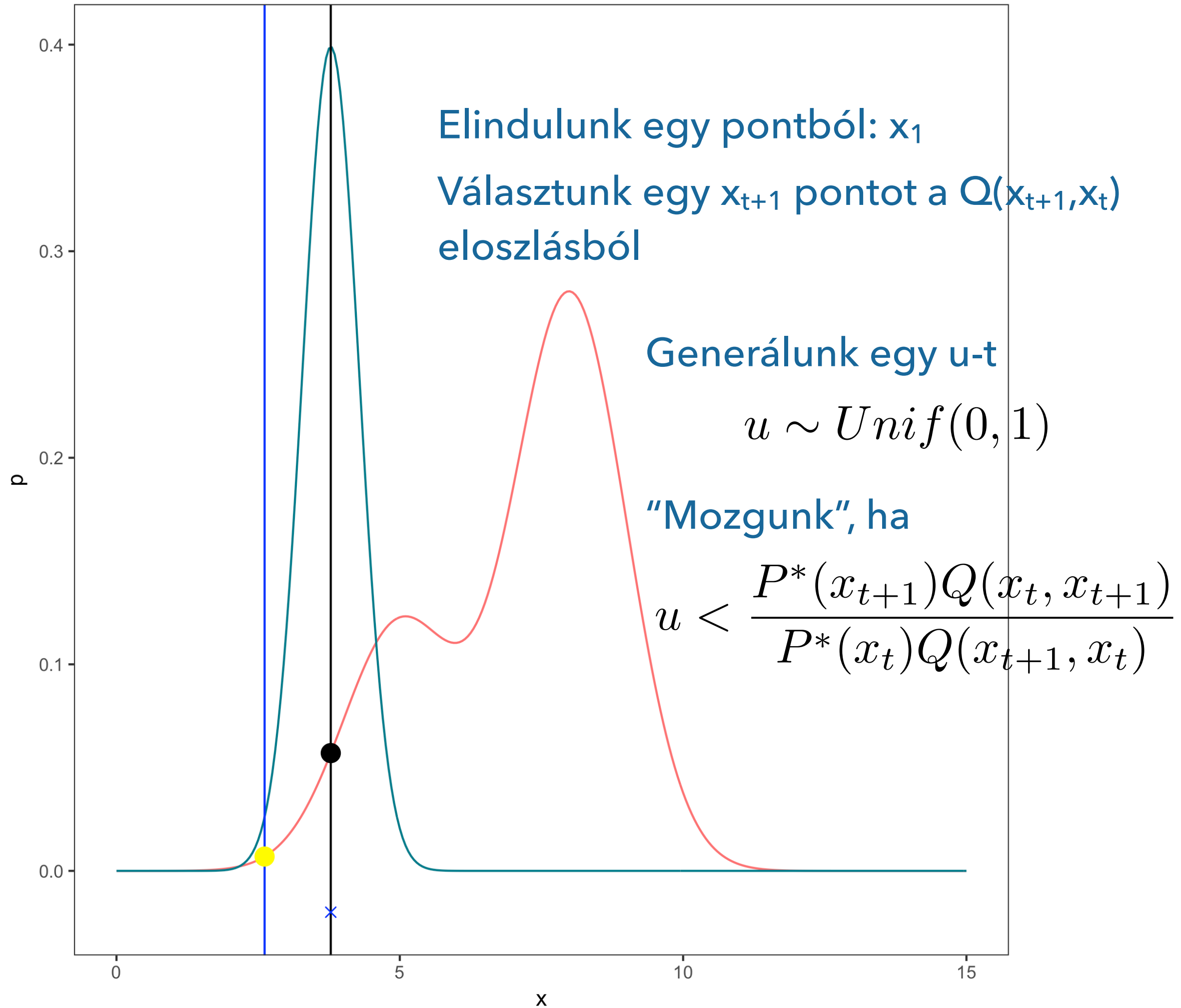


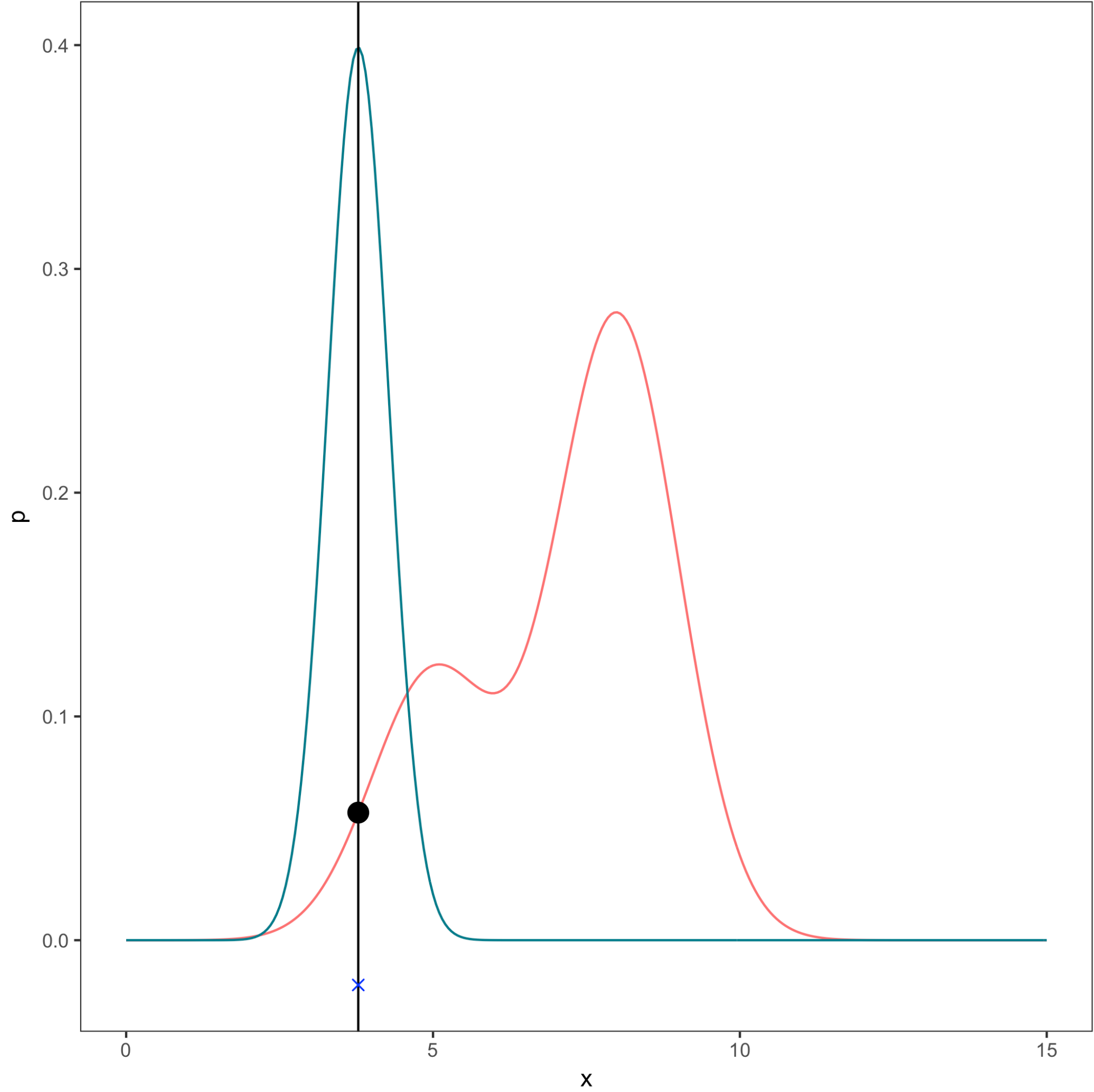


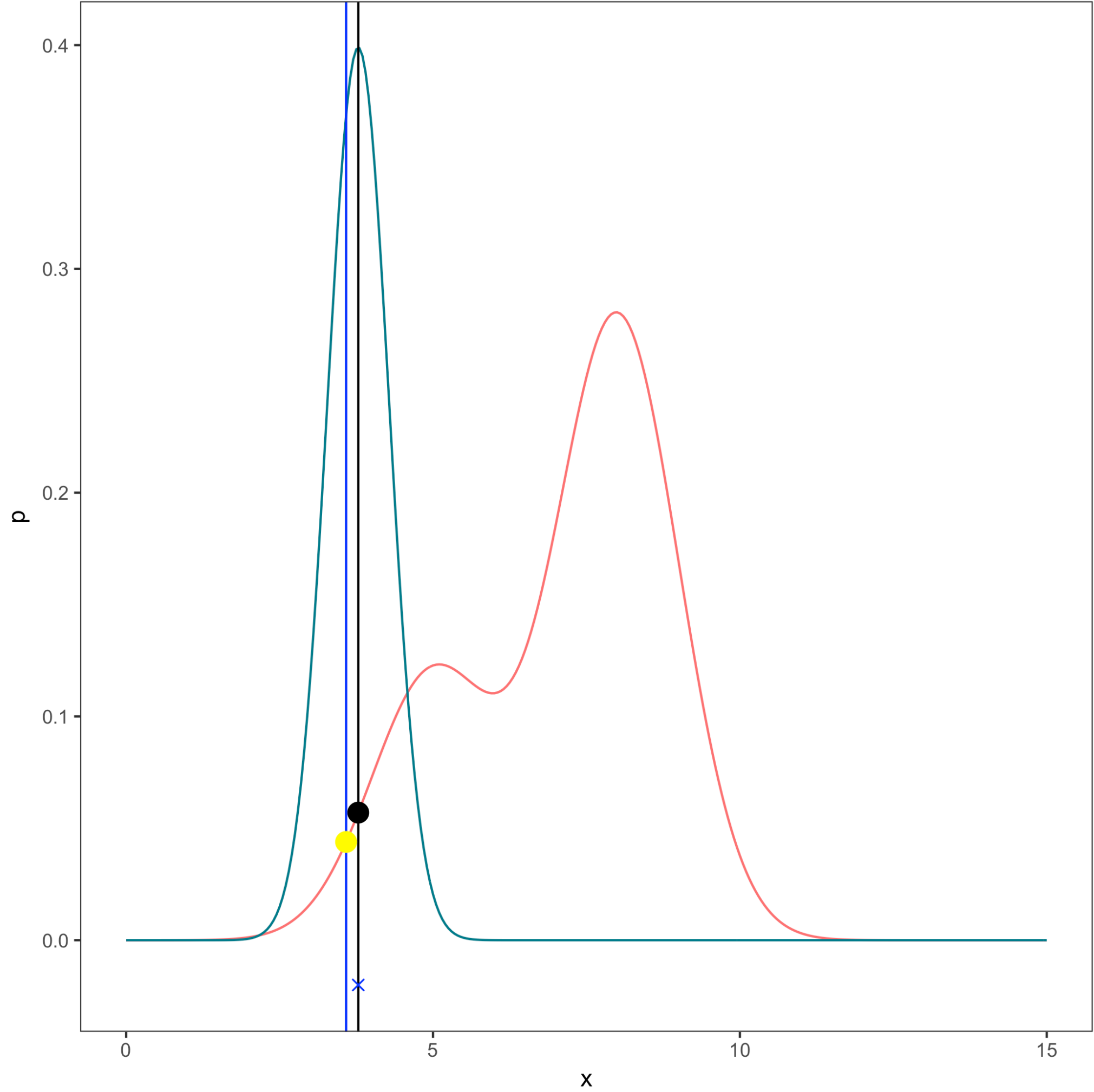


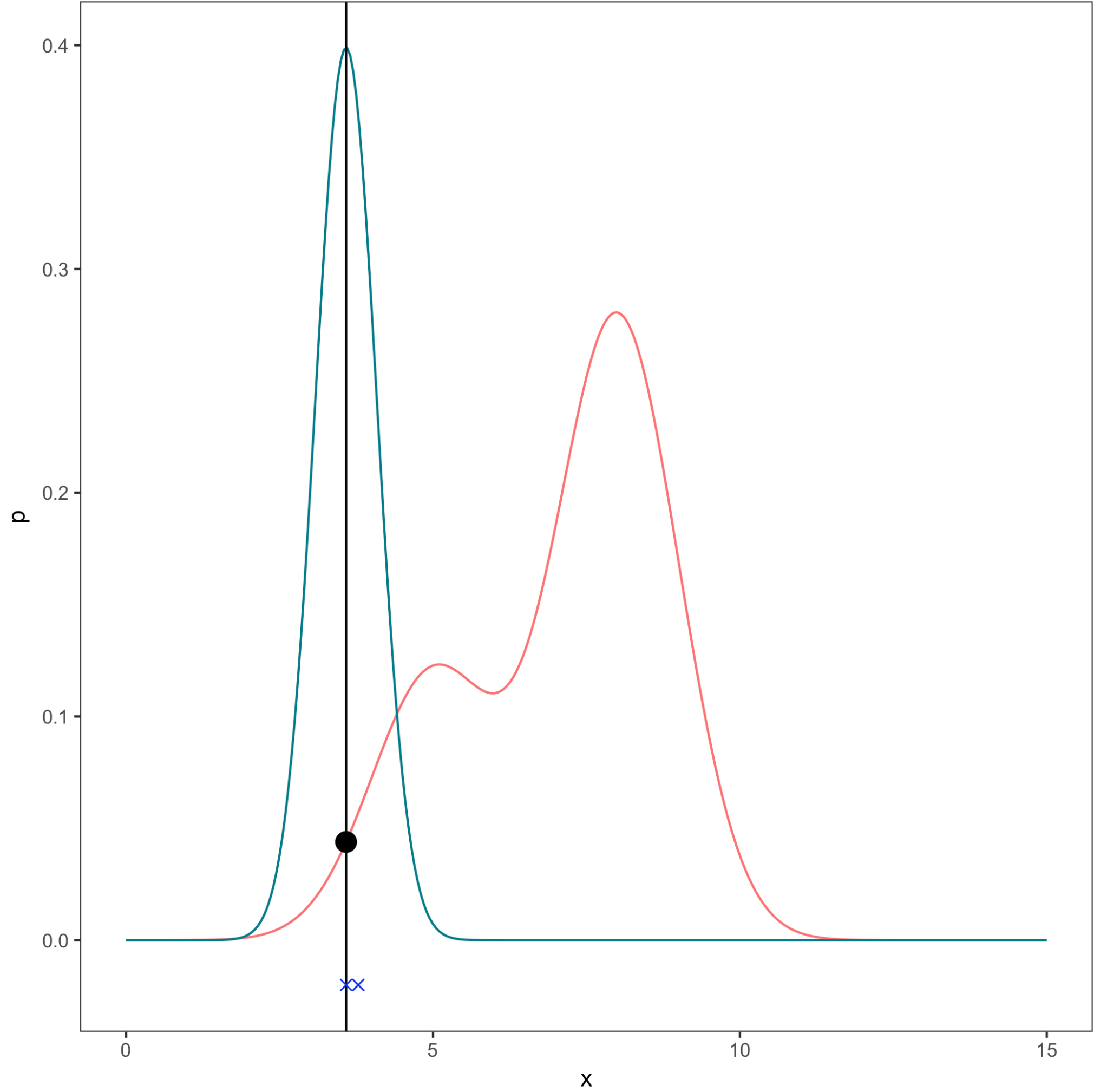
Elindulunk egy pontból: x_1
Választunk egy x_{t+1} pontot a $Q(x_{t+1}, x_t)$
eloszlásból

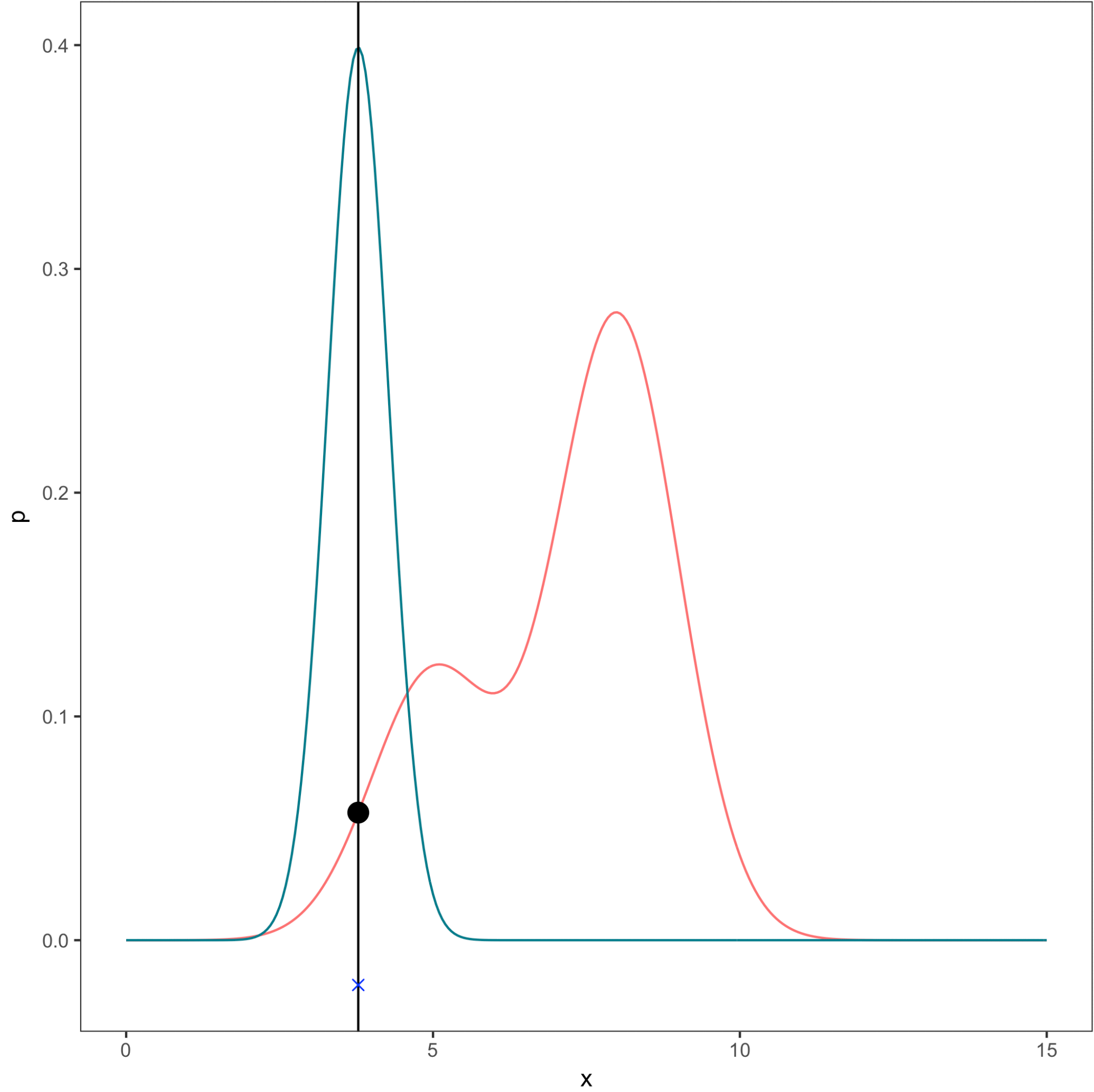


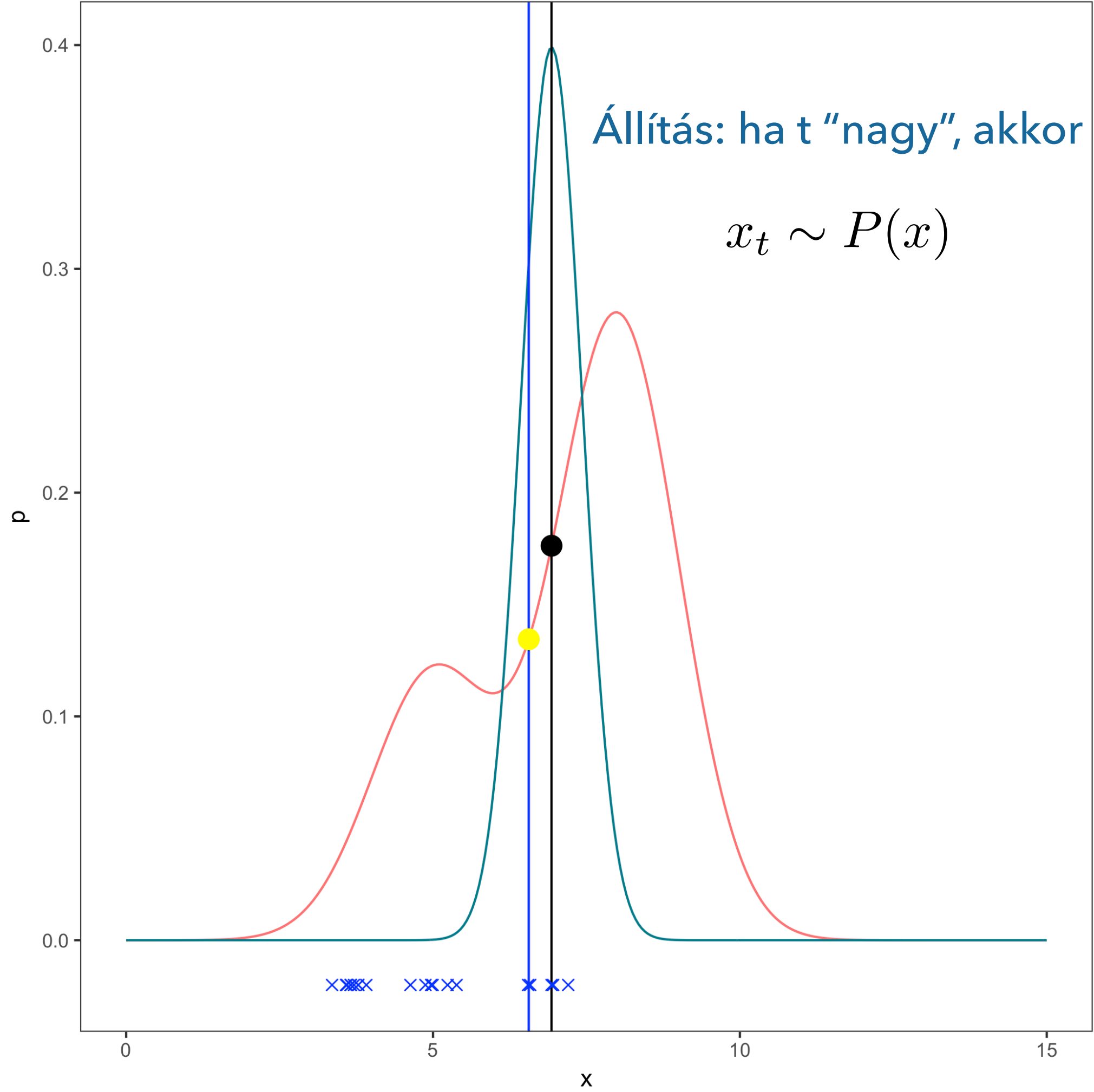


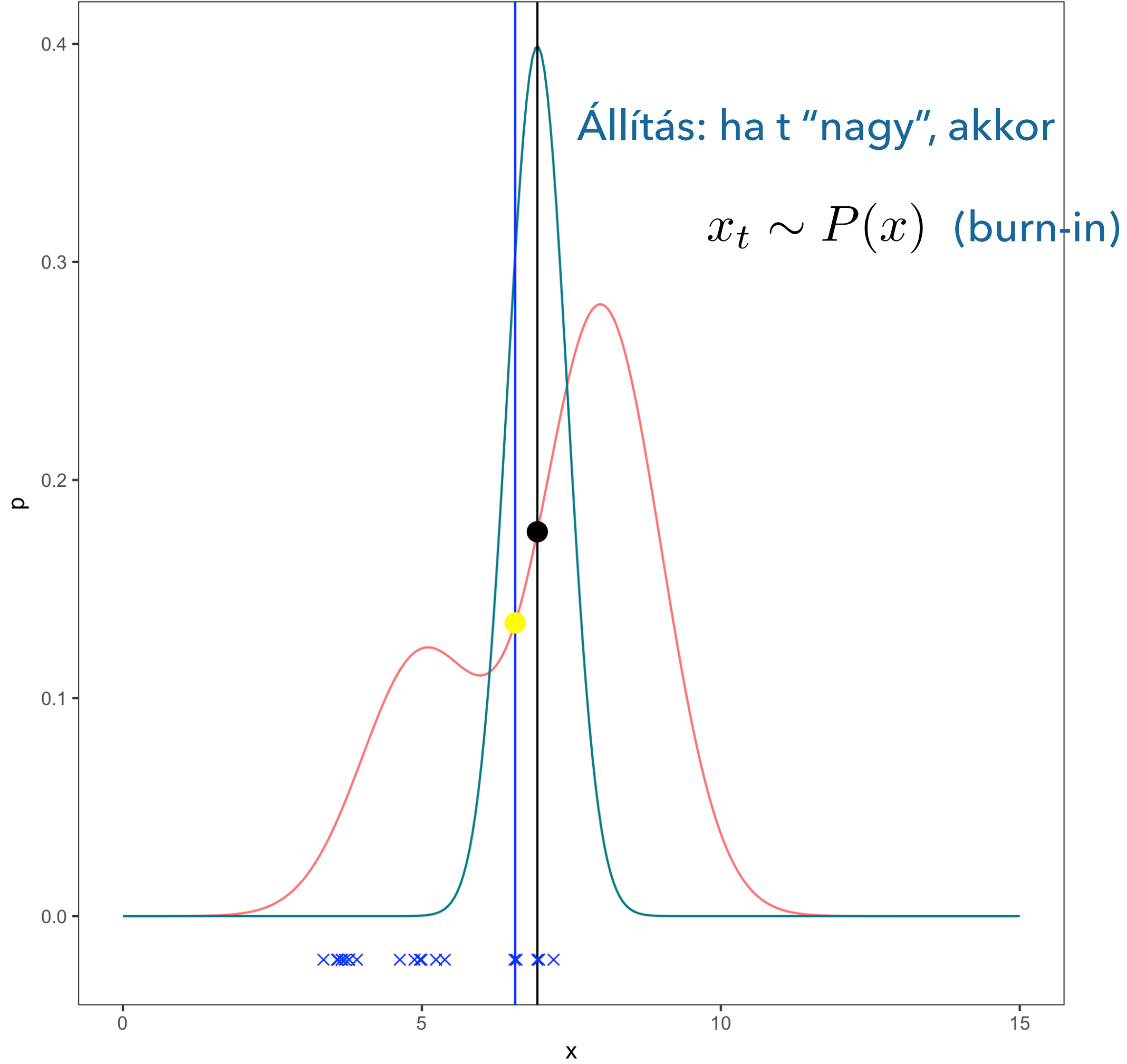


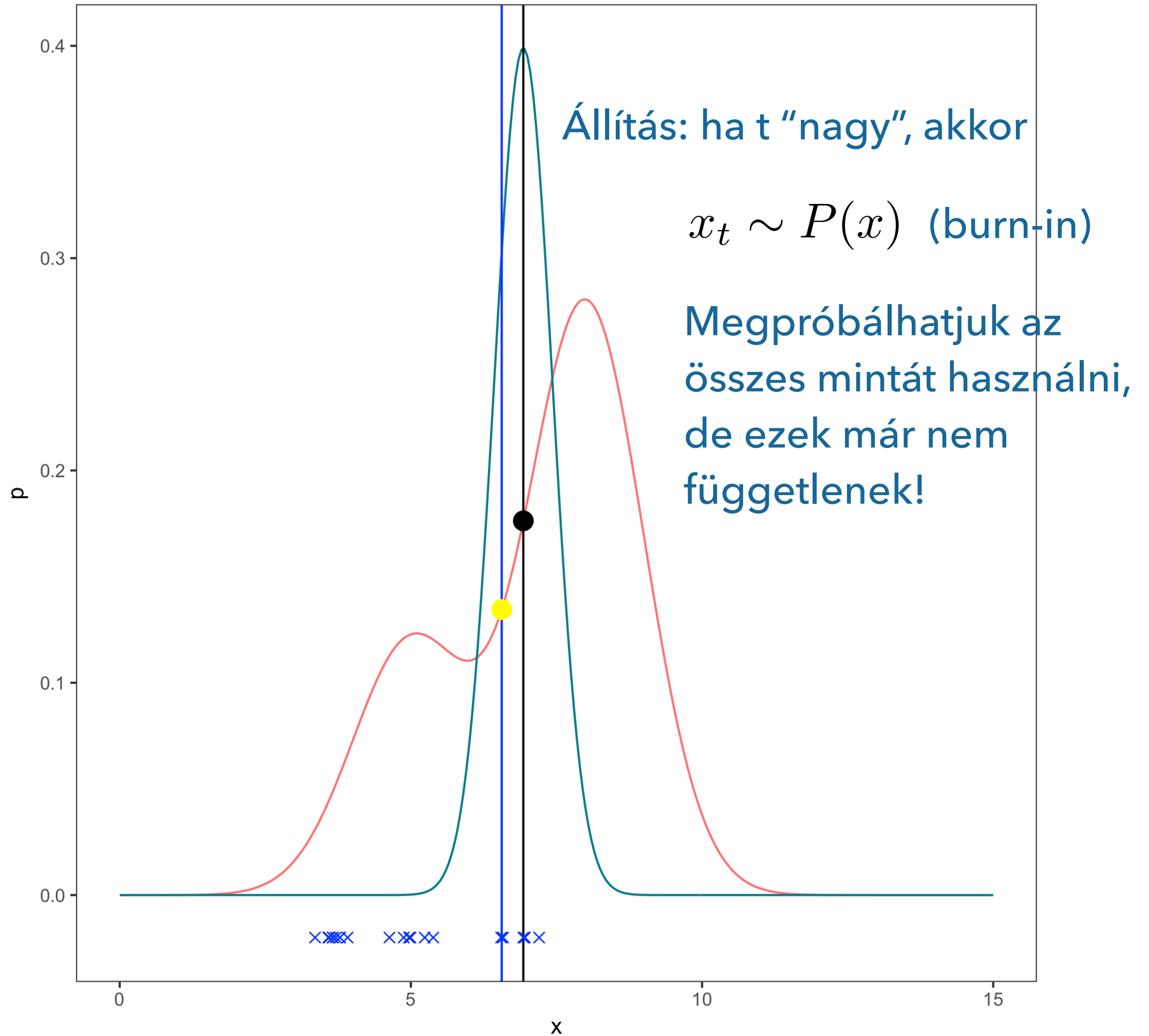


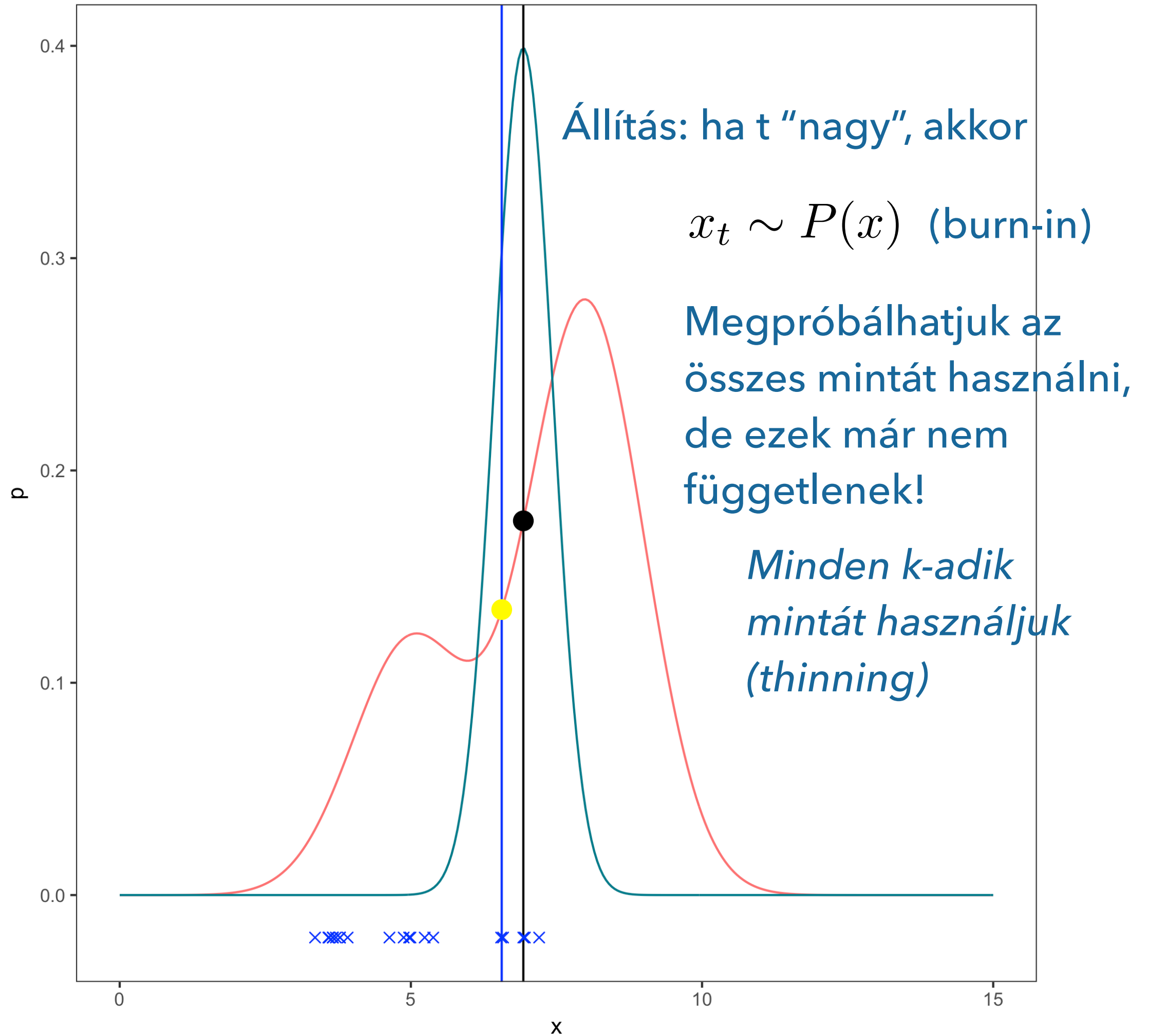












METROPOLIS HASTINGS VS. REJECTION SAMPLING

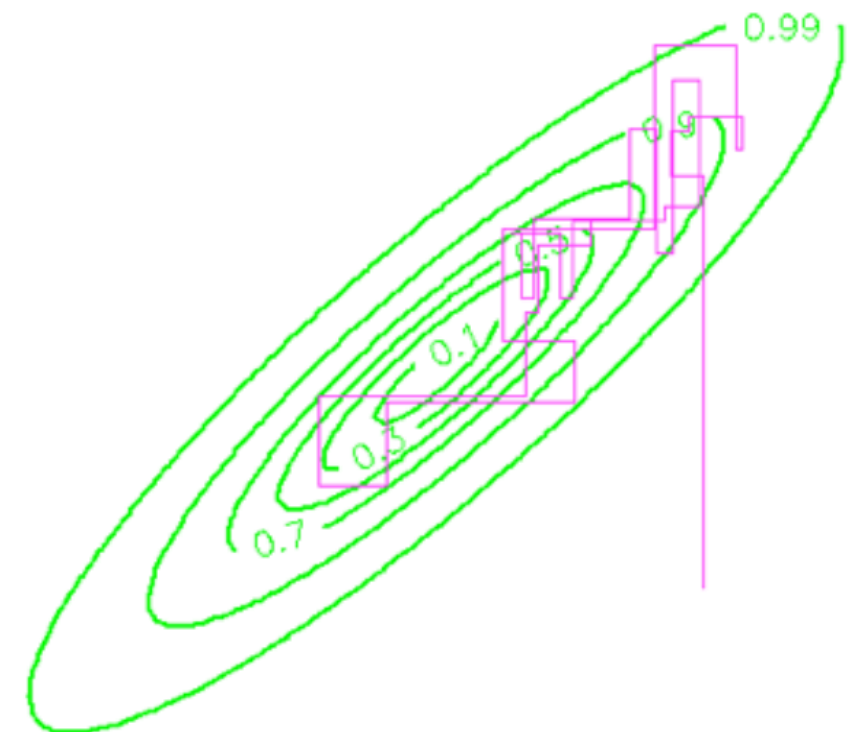
- ▶ MH: az algoritmikus lépések nem függenek a céleloszlás $P(x)$ alakjától. Emiatt általános inferencia-algoritmus fejleszthető belőle. (Általános, mert bármilyen modellre alkalmazható)
 - ▶ a rejection sampling esetében megfelelő $Q(x)$ mintavételező eloszlás kell ($CQ(x) > P^*(x)$)
- ▶ Kevésbé sújtja a “dimenzionalitás átka”

GIBBS SAMPLING

- ▶ Az egyes változókat sorban mintavételezzük, a többi változóra kondicionálva:

$$x_{t+1}^{(1)} \sim p(x^{(1)} | x_t^{(2)})$$

$$x_{t+1}^{(2)} \sim p(x^{(2)} | x_{t+1}^{(1)})$$



HAMILTONIAN MONTE CARLO (HMC)

- ▶ Markov-lánc
- ▶ Kihhasználja a mintavételezni kívánt eloszlás alakját.
 - ▶ Ehhez szükséges, hogy differenciálható legyen az eloszlásfüggvény $\rightarrow$ nincsenek diszkrét változók

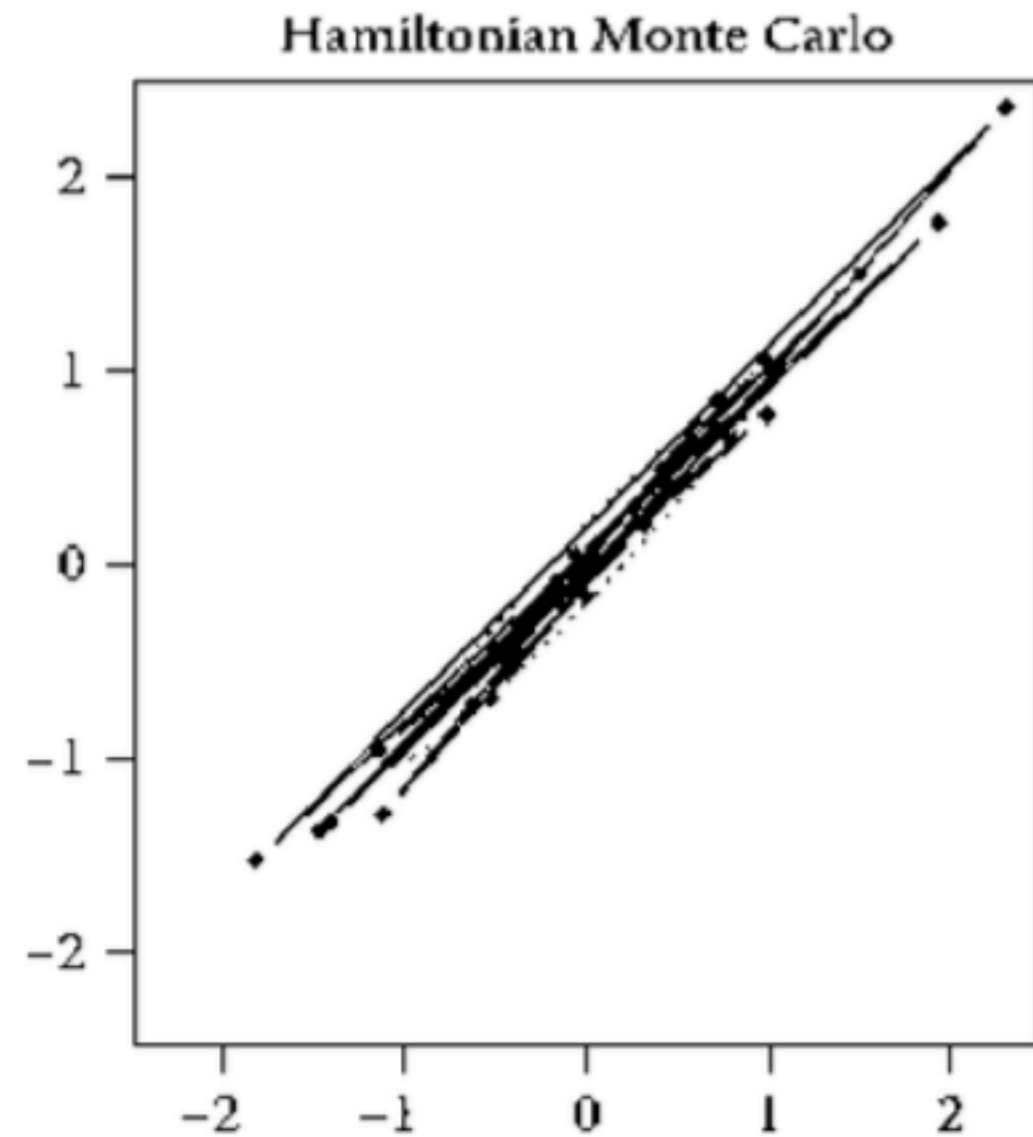
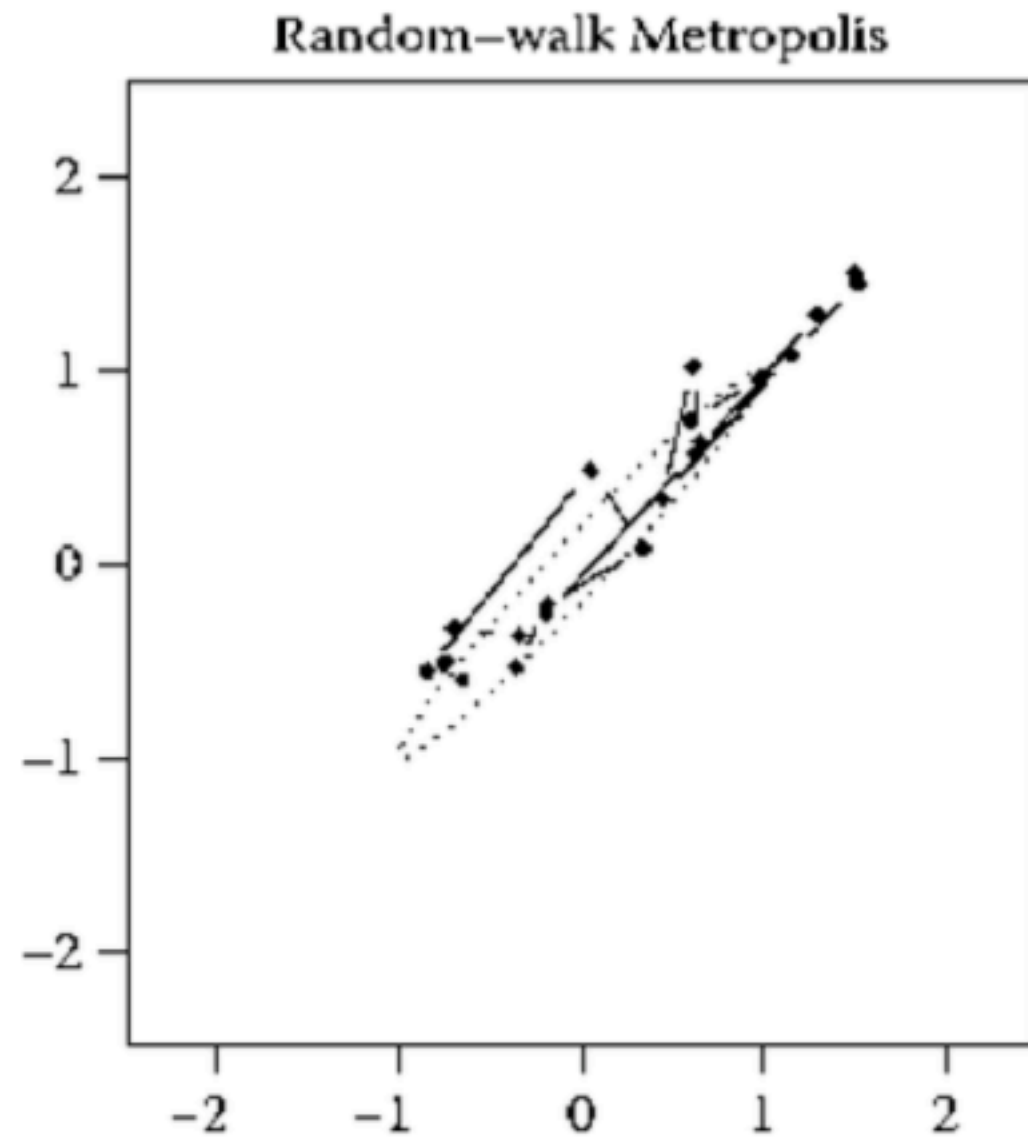
Hamiltonian MC

- bevezetünk egy segédváltozót, az impulzust
- a Metropolis-Hastings proposalt egy dinamikus rendszer állítja elő
- csak akkor rejectelünk, ha a diszkrét léptetés lemászott a trajektóriáról
- kell a grádiens
- gyorsabb keveredés

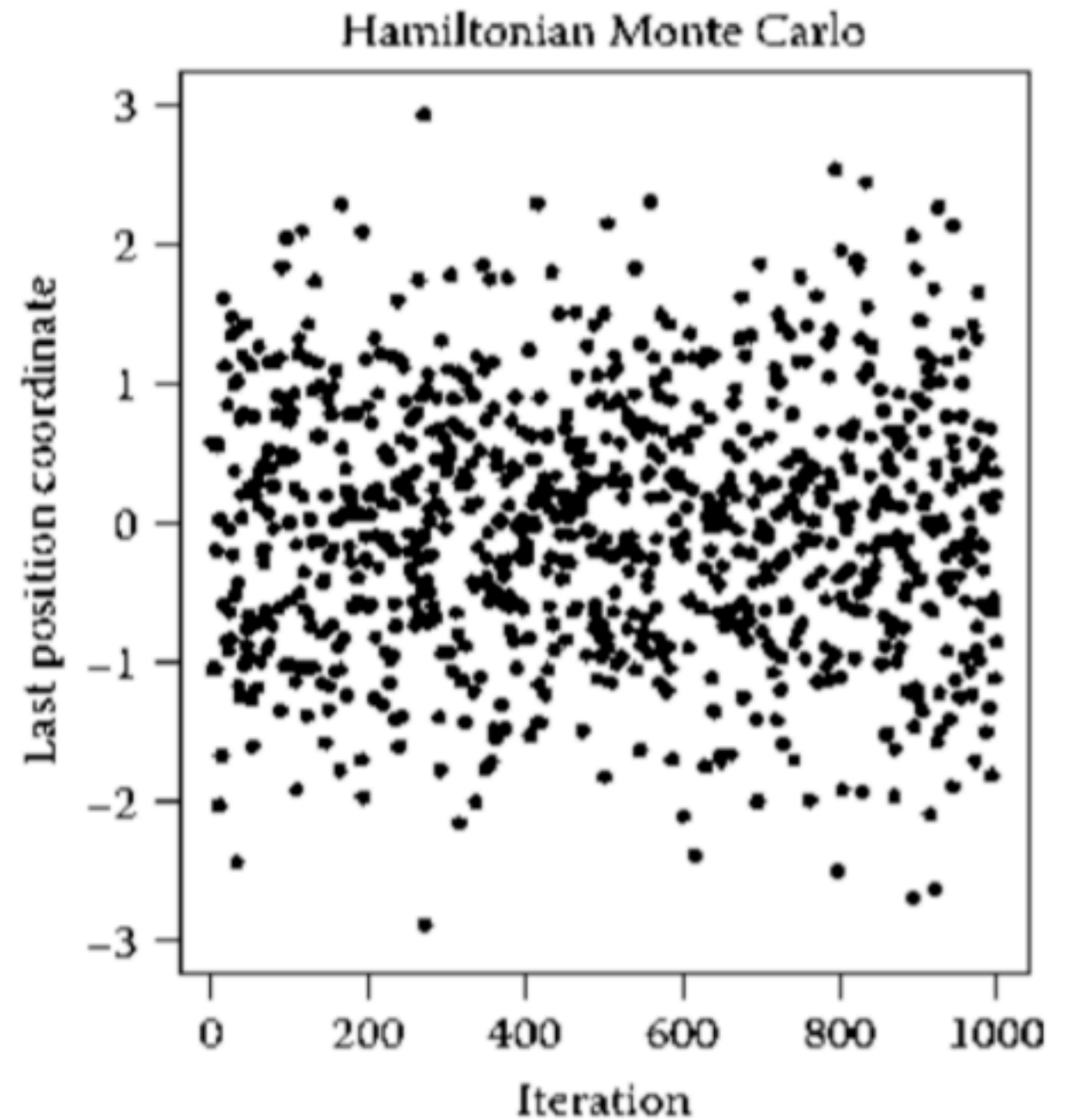
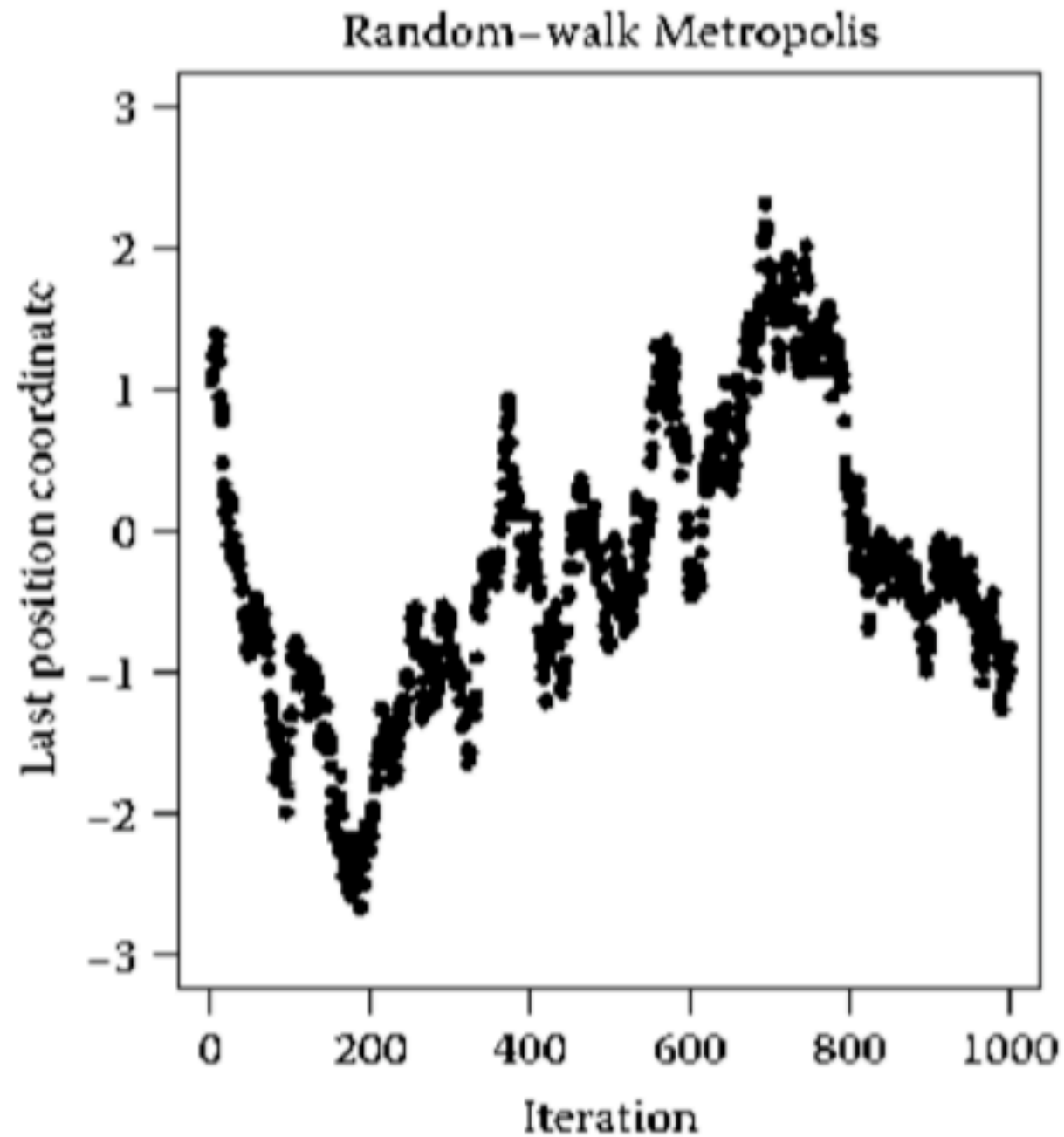
$$\dot{\mathbf{u}} = -\frac{1}{\tau} \frac{\partial \log P(\mathbf{u}, \mathbf{v}|\mathbf{x})}{\partial \mathbf{v}}$$
$$\dot{\mathbf{v}} = \frac{1}{\tau} \frac{\partial \log P(\mathbf{u}, \mathbf{v}|\mathbf{x})}{\partial \mathbf{u}},$$



HAMILTONIAN MONTE CARLO VS. METROPOLIS-HASTINGS

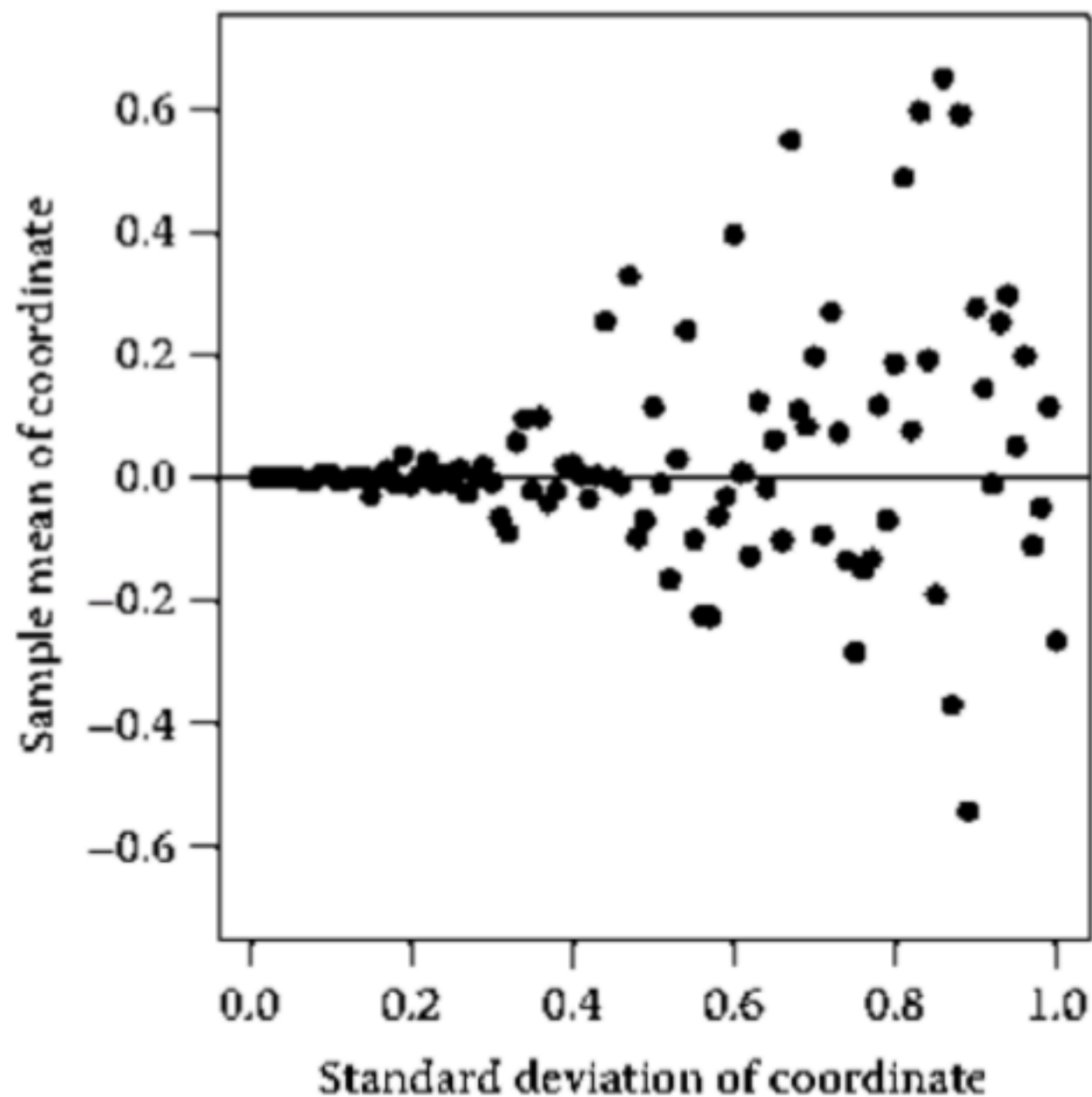


HAMILTONIAN MONTE CARLO VS. METROPOLIS-HASTINGS

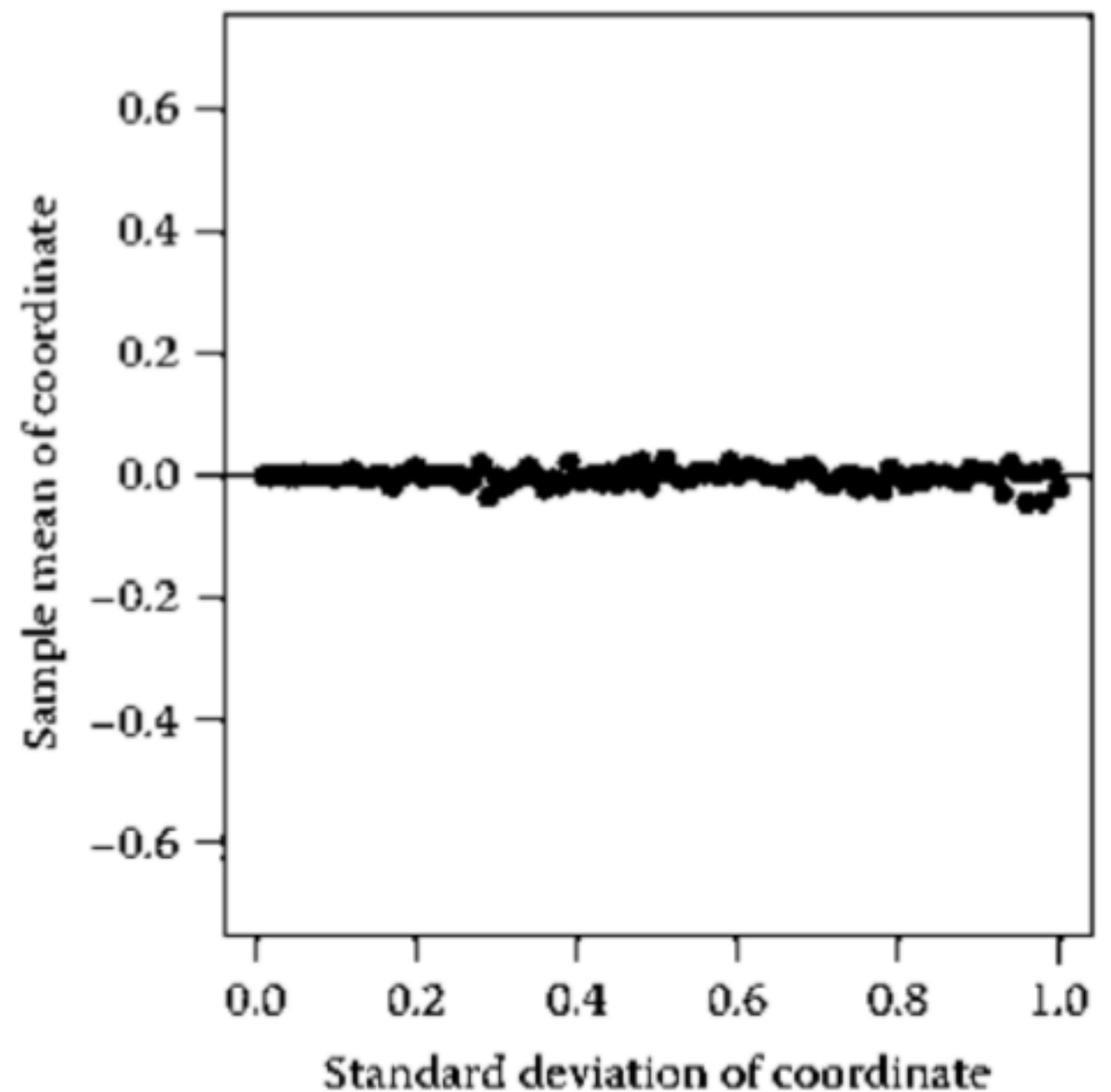


HAMILTONIAN MONTE CARLO VS. METROPOLIS-HASTINGS

Random-walk Metropolis



Hamiltonian Monte Carlo



Probabilisztikus programozási nyelvek

- Deklaratív nyelvek (mint pl. a PROLOG)
- A generatív modellt és a megfigyeléseket kell specifikálni
- Látenis változók poszterior eloszlására vonatkozó inferencia nyelvi elemként használható
- Ismertebb példák: BUGS, Church, Stan, Edward, Pyro

Stan

- Valószínűségi modelldefiníciós nyelv
- Hamiltonian MCMC
 - Autodiff
 - NUTS
- Interfaces: Python, R, Matlab, command line, etc.
- Compiles to C++

SEGÉDANYAGOK

- ▶ David Mackay video lectures 12. rész
https://www.youtube.com/watch?v=sN_0iGWcyLI
- ▶ Bishop: [Pattern recognition and machine learning](#),
11. fejezet
- ▶ Brooks, Gelman, Jones & Meng: Handbook of Markov
Chain Monte Carlo

Házi feladat

- Valósíts meg ancestral samplinget ebben a modellben egy jupyter notebookban
- Válassz numpy-ben vagy scipy-ben implementált eloszlásokat a változóknak
- Ábrázold a mintákat

