2022-06-23

MCMC παραδείγματα

$q(y|x) = ?$ transition kernel

σππ της επόμενης μέτρησης y δεδομένης της παρούσας μέτρησης x.

Εύκολη επιλογή: θεωρούμε κατανομή για το y ανεξάρτητη του x

Η κατανομή αυτή πρέπει να επιτρέπει πεδίο τιμών τουλάχιστον αυτό της f

π.χ. αν $f(x)$: $\boxed{\mathcal{U}(10,20)}$ $10 \le X \le 20$

Αν εγώ πάρω κατανομή για το y: $\mathcal{U}(5,15)$

ακατάλληλη: επιτρέπει τιμές του Y: $5 \le Y \le 15$

Αν παίρναμε για $Y \sim \mathcal{U}(8,30)$ οκ.

$8 \le Y \le 30$.

Αν τώρα δημιουργηθεί ένα y: π.χ. $y = 25$

αυτόματα δεν θα γίνει δεκτό γιατί

$$f(x) = \begin{cases} \frac{1}{10} & 10 \le x \le 20 \\ 0, & x \text{ αλλού} \end{cases}$$

Επομένως $f(y) = f(25) = 0$

κ' τότε ρ = πιθ. αποδοχής του y

$$= \min\left\{ \frac{\overset{=0}{f(y)}\, q(x|y)}{f(x)\, q(y|x)}, 1 \right\} = 0.$$

$\Rightarrow \rho = 0 \Rightarrow y = 25$ αποκλείεται να γίνει δεκτό.

Αν X που θέλουμε να προσομοιώσουμε έχει πεδίο τιμών $a \leq X \leq b$

τότε μπορούμε να πάρουμε $Y \sim \mathcal{U}(a, b)$

$$q(y) = \begin{cases} \dfrac{1}{b-a}, & a \leq y \leq b \\ 0, & \text{αλλού} \end{cases}$$

Ο αλγόριθμος απλοποιείται:

$$q(x|y) = q(x) = \frac{1}{b-a}$$

$$q(y|x) = q(y) = \frac{1}{b-a}$$

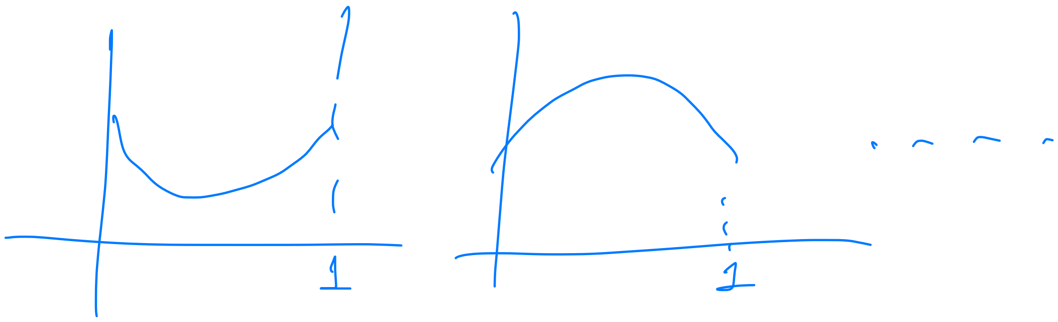
$$\rho(x,y) = \min\left\{ \frac{f(y)\,\frac{1}{b-a}}{f(x)\,\frac{1}{b-a}}, 1 \right\} = \min\left\{ \frac{f(y)}{f(x)}, 1 \right\}$$

Παράδειγμα 1

Γεννήτρια $X \sim Beta(a, b)$, $a = 5.2$, $b = 2.9$

$0 \leq X \leq 1$

$$f(x) = \begin{cases} C\, x^{a-1}(1-x)^{b-1}, & 0 \leq x \leq 1 \\ 0, & \text{αλλού} \end{cases}$$



Εναλλακτικά R function $\underbrace{d\,\underbrace{beta}_{\text{Beta distr.}}(\underbrace{a, b}_{\text{παράμετροι.}})}_{\text{συν. πυκνότητας (density)}}$

MCMC algorithm

Επειδή $X \in [0,1]$, θα χρησιμοποιήσουμε ως transition kernel

$Y \sim \mathcal{U}(0,1)$ (ανεξάρτητη της X)

$$q(y) = \begin{cases} 1, & 0 \leq y \leq 1 \\ 0, & \text{αλλού} \end{cases}$$

Αλγόριθμος

① Ξεκινάμε με αυθαίρετο $x_1 \in [0,1]$.

② Για $j = 2, \dots, N$

Δημιουργούμε $y \sim U(0,1)$

$$\rho(x,y) = \min\left\{ \frac{f(y)}{f(x_{j-1})}, 1 \right\}$$

όπου $f(x) = \text{dbeta}(x, a, b)$

Δημιουργούμε $u \sim U(0,1)$

Αν $u \le \rho \Rightarrow x_j = y$

Αν $u > \rho \Rightarrow x_j = x_{j-1}$

Θεωρητική ιδιότητα

Αν $N \to \infty \Rightarrow X_N \approx \text{Beta}(a,b)$

Επομένως για να δημιουργήσουμε δείγμα από n ανεξ ισόνομες $x_1, \dots, x_n \sim \text{Beta}$ θα πρέπει να τρέξουμε τον MCMC n φορές ανεξάρτητες μεταξύ τους κ' κάθε φορά να παίρνουμε την τελική τιμή X_N

Επίσης όμως ισχύει ότι αν

πάρουμε μια μεγάλη υλοποίηση του MCMC

$$(\underbrace{X_1, X_2, \ldots, X_N})$$ για μεγάλο N

κ' πάρουμε το τελευταίο κομμάτι της

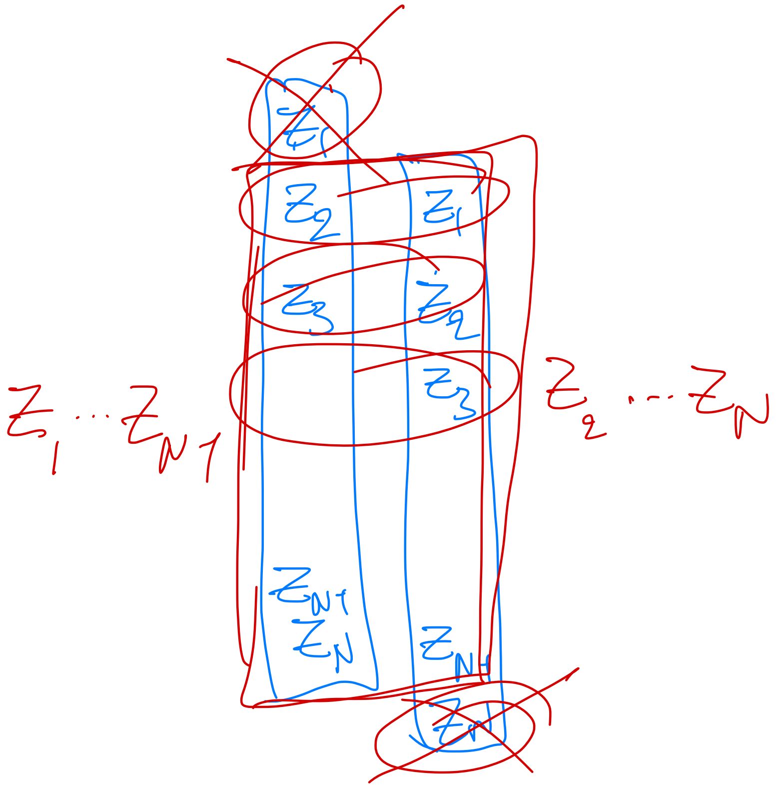
(π.χ. το δεύτερο μισό ή κ' παραπάνω)

αυτές οι τιμές επίσης $\sim$ Beta

όμως είναι συσχετισμένες

Έχουμε πάρει Z = 500 τιμές του X (τελευταίες)

$$Z = \begin{pmatrix} z_1 \\ \vdots \\ z_{499} \end{pmatrix} \qquad \tilde{Z} = \begin{pmatrix} z_2 \\ z_3 \\ \vdots \\ z_{500} \end{pmatrix}$$



Παράδειγμα 2

Έστω $X = (X_1, X_2)$ $\quad 0 \le X_1 \le 1$, $0 \le X_2 \le 1$

όπου η από κοινού συνάρτηση πυκνότητας

$$f(x_1, x_2) = \begin{cases} c\,\underline{(x_1 + x_2)}, & 0 \le x_1 \le 1,\ 0 \le x_2 \le 1 \\ 0, & \text{αλλού} \end{cases}$$

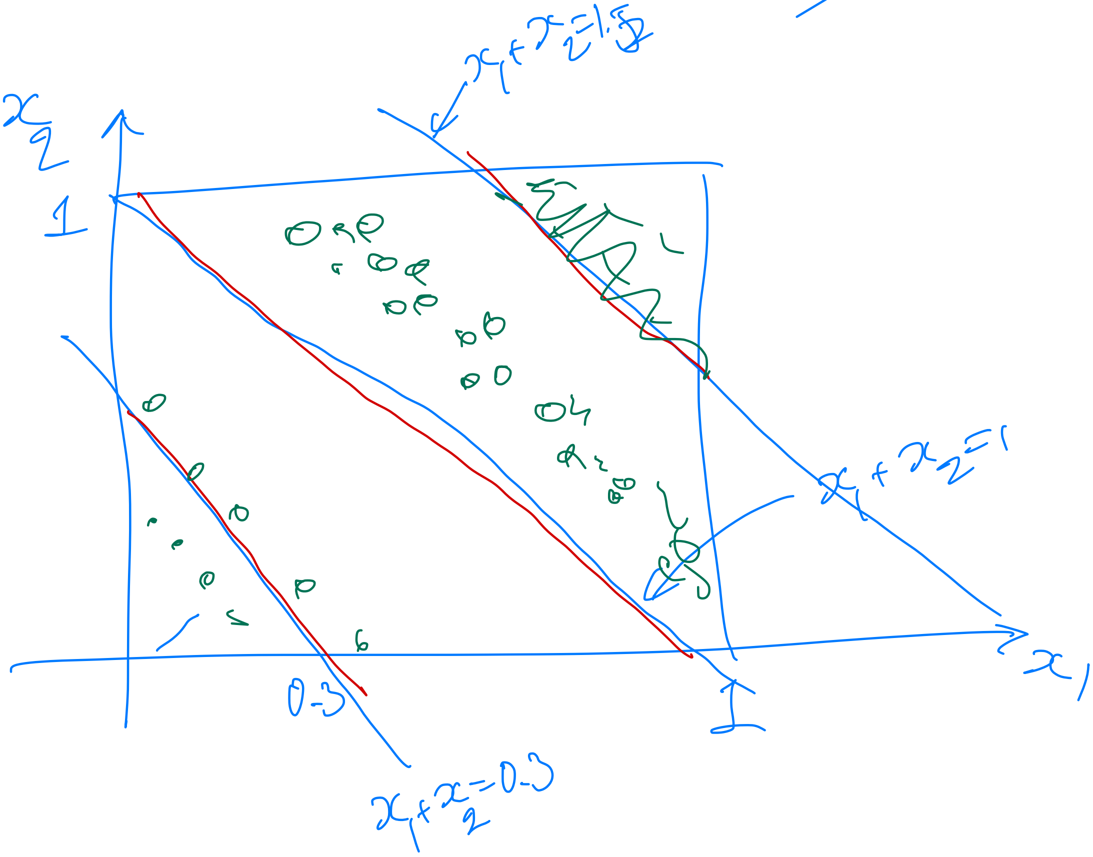
$c = ?$ $\qquad \displaystyle\int_0^1\!\!\int_0^1 f(x_1, x_2)\,dx_1\,dx_2 = 1 \cdots$

$\Rightarrow \cdots$ $c = ?$

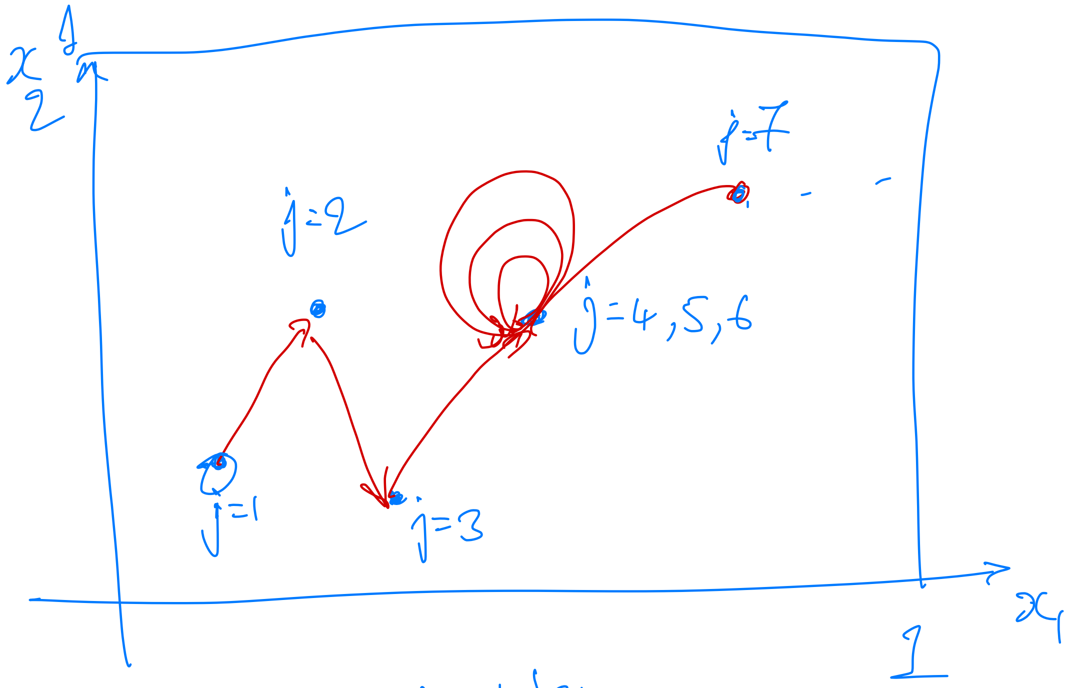
Τυχαία παρατήρηση $x=(x_1, x_2)$ ζεύγος

Τυχαίο δείγμα: (x_{11}, x_{12})
(x_{21}, x_{22})
$\vdots$
(x_{n1}, x_{n2})

ανεξάρτητα ζεύγη



Η χρονοσειρά που θα δημιουργηθεί θα αποτελείται από ζεύγη



υποψήφιο νέο βήμα

$$q(y) = q(y_1, y_2)$$

$Y = (Y_1, Y_2)$: παίρνουμε (ανεξάρτητα)

$Y_1 \sim \mathcal{U}(0,1)$ ⟵ $q_{Y_1}(y_1) = 1$

$Y_2 \sim \mathcal{U}(0,1)$ ⟵ $q_{Y_2}(y_2) = 1$

Y_1, Y_2 ανεξάρτητες

$$q(y_1, y_2) = \begin{cases} 1, & 0 \le y_1 \le 1, \; 0 \le y_2 \le 1 \\ 0, & \text{αλλού} \end{cases}$$

$(= q_{Y_1}(y_1)\, q_{Y_2}(y_2))$

Αλγόριθμος MCMC

1. Ξεκινάμε με αυθαίρετο $\underline{x}_1 = (x_{11}, x_{12})$

2. Για $j = 2, \dots, N$

Δημιουργούμε $Y_1 \sim U(0,1)$, $Y_2 \sim U(0,1)$ ανεξ.

$Y = (Y_1, Y_2)$

$$p = \min\left\{ \frac{f(y_1, y_2)}{f(x_{j-1,1}, x_{j-1,2})}, 1 \right\}$$

$$\left[f(x_1, x_2) = c(x_1 + x_2) \right]$$

$$= \min\left\{ \frac{y_1 + y_2}{x_{j-1,1} + x_{j-1,2}}, 1 \right\}$$

Δημιουργ. $u \sim U(0,1)$

Αν $u \le p \Rightarrow x_j = y = (y_1, y_2)$

$u > p \Rightarrow x_j = x_{j-1} = (x_{j-1,1}, x_{j-1,2})$

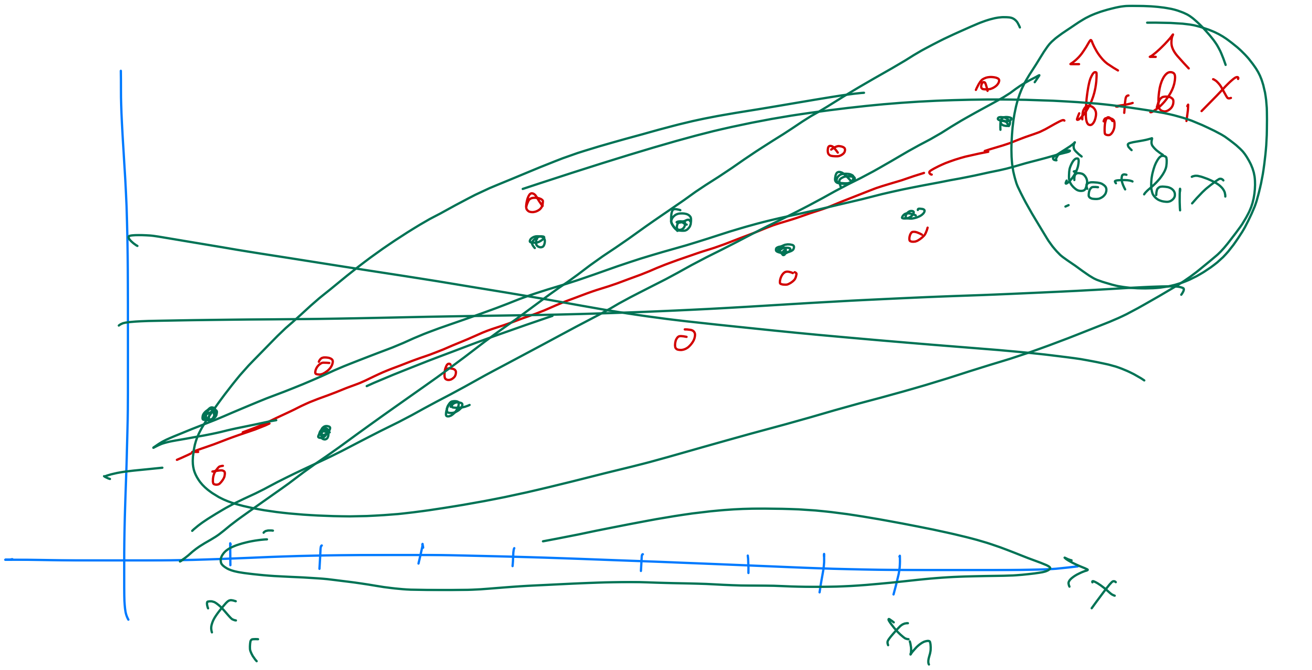
Παράδειγμα 3 Προσομοίωση Παλινδρόμησης

Έστω το μοντέλο $Y = b_0 + b_1 X + \varepsilon$

$\varepsilon \sim \mathcal{N}(0, \sigma^2)$

Έστω $(x_1, \dots, x_n)$ οι τιμές της ανεξ. μεταβλητής σε ένα δείγμα μεγέθους n.

Θέλουμε να "δούμε" (οπτικά) την μεταβλητότητα της γραμμής παλινδρόμησης όταν μεταβάλλονται τα y σε διαφορετικά δείγματα με τα ίδια x.



Προσομοίωση

Δίνεται $n=50$

" $x=(x_1,\dots,x_n)$

π.χ. $x=(\underbrace{10.1, 10.1, 10.2, \dots 15}_{50})$

Έστω ότι $b_0=5$, $b_1=2$ $E(Y)=5+2x$

$\sigma=10$ ~~αγνώστο~~ αγνώστο

Πως συμπεριφέρεται η εκτιμητική γραμμή παλινδρόμησης για διαφορετικά δείγματα με τα ίδια x κάθε φορά.

Δημιουργούμε N δείγματα μεγέθους $\underline{\underline{n}}$

$(N=1000)$ με το ίδιο x.

Για $j=1,\dots,N$

Εκτιμούμε $\hat{b}_{0,j}$, $\hat{b}_{1,j}$ για το δείγμα j

$$\hat{B} = \begin{array}{l} \begin{bmatrix} \hat{b}_{01} & \hat{b}_{11} \\ \hat{b}_{02} & \hat{b}_{12} \\ \vdots & \\ \hat{b}_{0N} , & \hat{b}_{1N} \end{bmatrix} \begin{array}{l} \leftarrow 1^{o} \text{ σενάριο} \\ \\ \\ \leftarrow \text{σενάριο } N. \end{array} \\ \qquad\qquad\quad N \times 2 \end{array}$$

Παράδειγμα 4

Σε ένα πληθυσμό η ηλικία X ενός ατόμου $X \sim \text{Gamma}(5, 0.1)$

Ένα άτομο ηλικίας $X = x$ έχει αρτηριακή πίεση $Y \sim N$, με μέση τιμή $9 + x/12$ κ' τυπική απόκλιση $0.04x$.

$$Y \mid X = x \sim N\left(\underline{9 + x/12},\ \underline{(0.04x)^2}\right)$$

Θέλουμε να εκτιμήσουμε $P(Y > 15) = \theta$

όπου Y : αρτ. πίεση τυχαίων ατόμων

(δηλ. % πληθ. πάσχει $Y > 15$)

Να εκτιμήσουμε το θ μέσω προσομοίωσης χρησιμοποιώντας 1000 σενάρια (1000 τυχαία άτομα).

Αλγόριθμος

Δημιουργούμε $N=1000$ υλοποιήσεις του X

(1000 άτομα τυχαίας ηλικίας $X \sim Gamma(5, 0.1)$)

μέσω της rgamma( )

Για

$$\begin{array}{cc} x & y \\ \left[\begin{array}{c} x_1 \\ \vdots \\ \vdots \\ x_N \end{array}\right. & \boxed{\begin{array}{c} y_1 \\ \vdots \\ \vdots \\ y_N \end{array}} \end{array}$$

← αρτ. πιέσεις των 1000 τυχαίων ατόμων

Για κάθε $j=1,\dots,N$

$$y_j \sim rnorm(1, \underbrace{9 + x_j/12}_{\mu}, \underbrace{0.04\, x_j}_{\sigma})$$

$$\hat{\theta} = \frac{(\text{αρ. } y_j \text{ όπου } y_j > 15)}{N} = \boxed{\text{\% στο δείγμα όπου } Y > 15}$$

Παράδειγμα 5

Μια ασφαλιστική εταιρεία έχει $n=500$ ασφαλισμένους καθένας από τους οποίους έχει 5% πιθανότητα να πάθει ατύχημα στη διάρκεια της επόμενης χρονιάς.

Για ένα ατύχημα η αποζημίωση $Y \sim Exp(\lambda)$

με $E(Y) = 600$ $\quad \left(\lambda = 1/600\right)$.

$E(Y) = \frac{1}{\lambda}$

Έστω S το συνολικό ποσό που θα πρέπει να πληρώσει η εταιρεία σε αποζημιώσεις

$$\theta = P(S > 35000)$$

Προσομοίωση

$N = 5000$ σενάρια

Σε κάθε σενάριο $j = 1, \ldots, N$

---

① Προσομοίωση X = αρ. ατυχημάτων

n ασφ. $= 500$

καθ. πιθ. $p = 5\%$ για ατύχημα $\Big\} \Rightarrow$

$\Rightarrow X \sim \text{Binomial}(n, p) \leftarrow$ (rbinom)

$\quad 500, \; 0.05$

Αν $X = k$ προσομοιώνουμε

τα k ποσά αποζημιώσεων $Y_1, \ldots, Y_k$

$Y_i \sim \text{Exp}(1/600), \; i = 1, \ldots, k$

S_j = συνολικό ποσό στο σενάριο j

$$= \sum_{i=1}^{k} Y_i$$

---

② $\hat{\theta} = \dfrac{\text{αρ. σεναρίων όπου } S > 35000}{N}$