

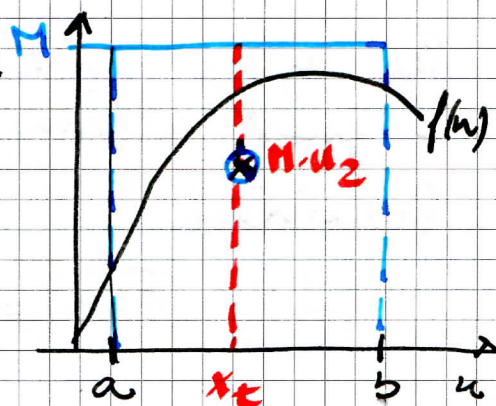
Campionamento di numeri casuali da distribuzioni di probabilità.

REJECTION METHOD

Supponiamo di non essere in grado di calcolare analiticamente la CDF di una distribuzione di probabilità e quindi di non poter usare il metodo dell'inversione.

Esiste un metodo alternativo introdotto da von Neumann nel 1951.

Supponiamo di dover campionare da una pdf $f(u)$ della quale conosciamo la forma analitica nell'intervallo $[a, b]$ e inoltre



$$\forall u \in [a, b] \quad f(u) < M$$

L'algoritmo prevede di

1) generare una variabile casuale uniforme in $[a, b]$:

$$u_1 \in \mathcal{U}(0, 1) \rightarrow u_t = a + (b - a)u_1$$

2) generiamo un altro numero $u_2 \in \mathcal{U}(0, 1)$ e ~~calcoliamo~~ la disuguaglianza
verificare

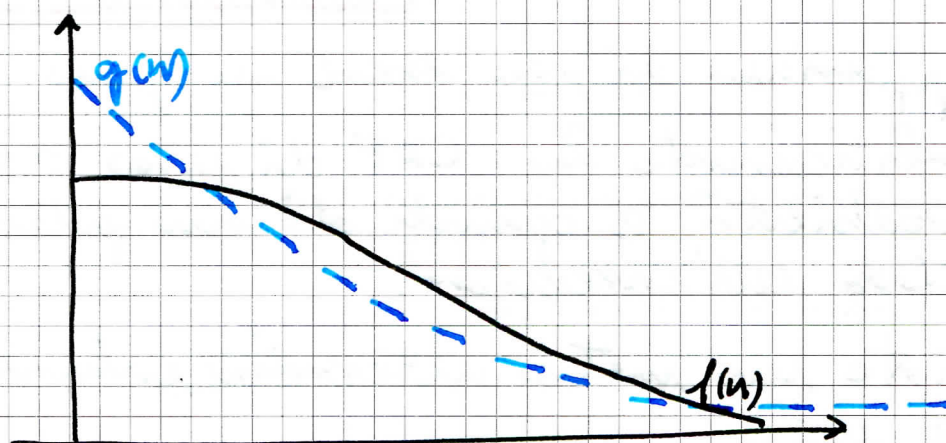
$$u_2 \cdot M \leq f(u_t)$$

in tal caso accettare u_t
altrimenti riprendiamo dal passo (1) dell'algoritmo.

L'efficienza del metodo è data dal rapporto delle aree

$$\epsilon = \frac{\int_a^b f(u) du}{M(b-a)}$$

Consideriamo ora il caso generale di due funzioni densità di probabilità molto simili $f(u)$ e $g(u)$



Per $g(u)$ conosciamo la forma analitica della $G(u)$ e quest'ultima può essere facilmente invertita

Riscriviamo

$$F(u) = \int_a^u du' f(u') = \int_a^u du' \frac{f(u')}{g(u')} \cdot g(u') = \int_a^u du' h(u') g(u')$$

se $f(u)$ e $g(u)$ sono molto simili, $h(u)$ non è lontana da 1.

In ogni caso, nell'intervallo $[a, b]$ $h(u)$ deve essere limitata

$$\forall u \in [a, b] \Rightarrow h(u) < h_{\max}$$

Algoritmo

- 1) sceglie $u_1 \in \mathcal{U}(0,1)$
- 2) calcola $u_t = G^{-1}(u_1)$
- 3) sceglie un nuovo $u_2 \in \mathcal{U}(0,1)$
- 4) se $u_2 < \frac{h(x_t)}{h_{\max}}$ accetta u_t
altrimenti ripete dallo step 1.

Vogliamo dimostrare che u_t si distribuisce
come $f(u_t) = h(u_t) \cdot g(u_t)$

Calcoliamo la probabilità

$$\Pr \left\{ u_2 < \frac{h(u_t)}{h_{\max}} \right\} = \int_0^{\frac{h(u_t)}{h_{\max}}} du_2 = \frac{h(u_t)}{h_{\max}}$$

$W(u_t, u_2)$ sia la probabilità congiunta
per le due variabili casuali x_t e u_2

Dal teorema di Bayes

$$\Pr \left\{ u_t \mid u_2 < \frac{h(u_t)}{h_{\max}} \right\} = \frac{W(u_t, u_2) \Pr \left\{ u_2 < \frac{h(u_t)}{h_{\max}} \right\}}{\int_a^b W(u_t, u_2) \cdot \Pr \left\{ u_2 < \frac{h(u_t)}{h_{\max}} \right\} dx_t}$$

dato che u_t e u_2 sono variabili casuali
indipendenti,

$$W(u_t, u_2) = g(u_t) \cdot \underbrace{u(u_2)}_{u_2 \in U(0,1)} = g(u_t)$$

sostituendo nell'eq. precedente

$$\begin{aligned} \Pr \left\{ u_t \mid u_2 < \frac{h(u_t)}{h_{\max}} \right\} &= \frac{g(u_t) \cdot h(u_t) \cdot \frac{1}{h_{\max}}}{\int_a^b du_t \cdot g(u_t) \cdot \frac{h(u_t)}{h_{\max}}} \\ &= \frac{g(u_t) \cdot h(u_t)}{\int_a^b f(u_t) du_t} = g(u_t) \cdot h(u_t) \\ &= f(u_t) \end{aligned}$$

Esempio

Campionare da una distribuzione di probabilità
gaussiana

$$f(u) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < u < +\infty$$

operare una trasformazione di variabile
passando ad una distrib. normalizzata

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

una volta campionato da z è possibile
ricavarne

$$u = \mu + \sigma z$$

Riscriviamo la $f(z)$ come

$$f(z) = \sqrt{\frac{2}{\pi}} e^{-\frac{z^2}{2}} \quad \text{con } u \geq 0$$

dato che la distribuzione è simmetrica attorno
allo zero, una volta campionato u è
possibile generare il x_{rep} (+o-) campionando
da una distribuzione uniforme

$$u \in U(0,1)$$

e $u < 0.5 \rightarrow x_{\text{rep}} +$
altrimenti $\rightarrow x_{\text{rep}} -$

Riscriviamo la pdf come

$$f(x) = \sqrt{\frac{2e}{\pi}} e^{-u} e^{-\frac{(x-1)^2}{2}}$$

$$\text{con } g(u) = e^{-u}$$

$$h(u) = \sqrt{\frac{2e}{\pi}} e^{-\frac{(x-1)^2}{2}}$$

algoritmo

1) generiamo $u_1 \in U(0,1)$ $V_1 = -\ln u_1$

2) la condizione di accettazione è

$$u_2 \in U(0,1)$$

$$u_2 < \frac{h(V_1)}{h_{\max}}$$

$$\text{con } h_{\max} = \sqrt{\frac{2e}{\pi}}$$

cioè

$$u_2 < e^{-\frac{(V_1-1)^2}{2}}$$

che è equivalente a

$$\ln u_2 < -\frac{(V_1-1)^2}{2}$$

$$-\ln u_2 > \frac{(V_1-1)^2}{2}$$

ma ~~non~~ ^{ricordando} $V_2 = -\ln u_2$ allora

$$V_2 > \frac{(V_1-1)^2}{2}$$

ALGORITMO NUMERICO

1) generare V_1 e V_2
da una distrib. esp

$$u_1, u_2 \in U(0,1)$$

$$V_1 = -\ln u_1$$

$$V_2 = -\ln u_2$$

2) se $V_2 > \frac{(V_1-1)^2}{2}$, proseguire

generare $u_3 \in U(0,1)$

$$u_3 < 0.5 \rightarrow Z = -V_1$$

$$\text{altrimenti} \rightarrow Z = V_1$$

3) se $V_2 < \frac{(V_1-1)^2}{2}$ ripetere dal step 1