HW 1

① Seja $Y \sim Exp(\lambda)$ e $a > 0$. Consideramos a variavel depois de restringir seu suporte. Ou seja, $X = Y | Y \geq a$ e.g. X tem mesma lei de Y com a condição de estar em $[a, \infty+)$. Calcule $F_X(x)$, a função acumulada de X, e $F_X^{-1}(u)$ a função quantil de X. Descreva um algoritmo para simular X de $U \sim U[0,1]$.

Temos que $f_Y(y) = \lambda e^{-\lambda y}$.

Daí a CDF de X é

$$F_X(x) = P(X \leq x) = P(Y \leq x | Y \geq a)$$

$$= \frac{P(a \leq Y \leq x)}{P(Y \geq a)}$$

- $P(a \leq Y \leq x) = \int_a^x \lambda e^{-\lambda y} dy$
- $P(Y \geq a) = \int_a^\infty \lambda e^{-\lambda y} dy = e^{-\lambda a}$

$$\int_a^x \lambda e^{-\lambda y} dy = \left[-e^{-\lambda y}\right]_a^x = e^{-\lambda a} - e^{-\lambda x}$$

Então

$$F_X(x) = \frac{e^{-\lambda a} - e^{-\lambda x}}{e^{-\lambda a}} = 1 - e^{-\lambda(x-a)}, \quad x \geq a$$

Para a função quantil $F_X^{-1}(u)$, deve satisfazer $F_X(F_X^{-1}(u)) = u$. Como $F_X(x) = 1 - e^{-\lambda(x-a)}$, resolvendo para x

$$u = 1 - e^{-\lambda(x-a)} \Rightarrow e^{-\lambda(x-a)} = 1 - u$$
$$\Rightarrow x = a - \frac{\ln(1-u)}{\lambda} := F_X^{-1}(u)$$

Para simular X apartir de uma Uniforme$[0,1]$, basta computar $X = F_X^{-1}(U) = a - \frac{\ln(1-U)}{\lambda}$, onde $U \sim Unif[0,1]$

② Sejam a e b dados com $a<b$. Mostre que podemos simular $X = Y | a \le Y \le b$ de $U \sim Unif[0,1]$, usando

$$X = F_Y^{-1}(F_Y(a)(1-U) + F_Y(b)U).$$

eq. mostre que X tem é dado pela formula acima, então $P(X \le x) = P(Y \le x | a \le Y \le b)$. Aplique a formula para o problema anterior.

$F_X(x) = P(X \le x)$ é 0 fora de $a \le x \le b$. Neste caso.

$$P(X \le x) = P(F_Y^{-1}(F_Y(a)(1-U) + F_Y(b)U) \le x)$$

$$= P(F_Y(a)(1-U) + F_Y(b)U \le F_Y(x))$$

$$= P\left(U \le \frac{F_Y(x) - F_Y(a)}{F_Y(b) - F_Y(a)}\right) = \frac{F_Y(x) - F_Y(a)}{F_Y(b) - F_Y(a)}.$$

Para a exponencial condicional, para $F_Y(y) = 1 - exp(-\lambda y)$, e $b = \infty$ então $F_Y(b) = 1$

$$F_Y(a)(1-U) + F_Y(b)U = (1 - exp(-\lambda a))(1-U) + U$$

$$\Rightarrow \quad F_Y^{-1}(v) = -\lambda^{-1} \log(1-v)$$

$$\Rightarrow \quad X = -\lambda^{-1} \log(1 - (1 - exp(-\lambda a))(1-U) + U)$$

$$\Rightarrow \quad = -\lambda^{-1} \log(exp(-\lambda a)(1-U))$$

$$= a - \lambda^{-1} \log(1-U).$$

③ Aqui é um algoritmo simples para simular $X = Y \mid Y > a$ para $Y \sim Exp(\lambda)$:

(a) Seja $Y \sim Exp(\lambda)$. Simula $Y = y$
(b) Se $Y > a$ para e retorna $X = y$, e caso contrário, vai para o passo (a).

Mostre que este é ~~s~~ um algoritmo de rejeição, escrevendo as densidades propostas e alvo como π e q, bem como o limite $M = \max_x \pi(x)/q(x)$. Calcule o número esperado de tentativas para a primeira aceitação. Por que a inversão é preferida para $a \gg 1/\lambda$?

As densidades proposta e alvo são

$$q(y) = \lambda \exp(-\lambda y), \quad y > 0$$
$$\pi(x) = \lambda \exp(-\lambda(x-a)) \mathbb{I}(x \geq a).$$

O limite $M = \max_x \pi(x)/q(x)$ é $M = \exp(\lambda a)$ e então a probabilidade de aceitação do algoritmo de rejeição é

$$\pi(y) / Mq(y) = \begin{cases} 1 & y \geq a \\ 0 & y < a \end{cases}$$

Como aceitamos quando $y \geq a$ e rejeitamos caso contrário. Este algoritmo corresponde ao descrito acima. O número de tentativas N é distribuído geometricamente com probabilidade p, onde p é a probabilidade de amostrar $Y \geq a$, que é $p = \exp(-\lambda a)$. O valor esperado é $EN = 1/p = \exp(\lambda a)$. Se $a\lambda \gg 1$, então EN é grande. A inversão fornece uma amostra com uma avaliação de função única.

④ Considere o seguinte algoritmo de rejeição "squeeze" para amostrar de uma distribuição $\pi(x) = \tilde{\pi}(x)/Z_\pi$ num espaço de estado $\mathcal{X}$ tal que

$$h(x) \leq \tilde{\pi}(x) \leq M\tilde{g}(x)$$

onde h é uma função, $M>0$ e $g(x) = \tilde{g}(x)/Z_g$ é a densidade de uma distribuição que podemos amostrar. O algoritmo segue como:

(a) Amostre independentemente $X \sim g$, $U \sim \text{Unif}[0,1]$.
(b) Aceita X se $U \leq h(X)/(M\tilde{g}(X))$.
(c) Se X não é aceito ~~volta ao~~ no passo (b), amostre independentemente $V \sim \text{Unif}[0,1]$ e aceite X se

$$V \leq \frac{\tilde{\pi}(X) - h(X)}{M\tilde{g}(X) - h(X)}.$$

Mostre que a probabilidade de aceitação um $X=x$ proposto em cada passo de (b) ou (c) é

$$\frac{\tilde{\pi}(x)}{M\tilde{g}(X)}.$$

Temos que $P(X=x \text{ ser aceito}) =$
$$P(X=x \text{ ser aceito em (b)}) + P(X=x \text{ ser aceito em (c)})$$

onde
$$P(X=x \text{ ser aceito em (b)}) = \int_0^1 \mathbb{I}\left(u \leq \frac{h(x)}{M\tilde{g}(x)}\right) du$$
$$= \frac{h(x)}{M\tilde{g}(x)}.$$

e

$$P(X=x \text{ ser aceito em } (c)) = \int_0^1\int_0^1 \mathbb{I}\left(u > \frac{h(x)}{M\tilde{q}(x)}\right)\mathbb{I}\left(v \leq \frac{\tilde{\pi}(x)-h(x)}{M\tilde{q}(x)-h(x)}\right) du\,dv$$

$$= \left(\frac{\tilde{\pi}(x)-h(x)}{M\tilde{q}(x)-h(x)}\right)\left(1-\frac{h(x)}{M\tilde{q}(x)}\right)$$

$$= \frac{\tilde{\pi}(x)-h(x)}{M\tilde{q}(x)-h(x)} - \frac{h(x)(\tilde{\pi}(x)-h(x))}{M\tilde{q}(x)(M\tilde{q}(x)-h(x))}$$

$$= \frac{M\tilde{q}(x)\tilde{\pi}(x) - M\tilde{q}(x)h(x) - h(x)\tilde{\pi}(x) + h^2(x)}{M\tilde{q}(x)(M\tilde{q}(x)-h(x))}$$

$$= \frac{\cancel{(M\tilde{q}(x)-h(x))}(\tilde{\pi}(x)-h(x))}{M\tilde{q}(x)\cancel{(M\tilde{q}(x)-h(x))}}$$

$$= \frac{\tilde{\pi}(x)-h(x)}{M\tilde{q}(x)}.$$

O que completa a prova

⑤ Deduza da questão anterior que a distribuição das amostras aceitas pelo algoritmo é π.

$$P(X \in A \mid X \text{ foi aceito}) = \frac{P(X \in A, A \text{ foi aceito})}{P(A \text{ foi aceito})}$$

$$= \frac{\int_A \frac{\tilde{\pi}(x)}{M\tilde{q}(x)} q(x)\,dx}{\int_X \frac{\tilde{\pi}(x)}{M\tilde{q}(x)} q(x)\,dx}$$

$$= \frac{\int_A \tilde{\pi}(x)\,dx}{\int_X \tilde{\pi}(x)\,dx}$$

Pois $Z_\pi = \int_X \tilde{\pi}(x)\,dx \rightarrow \; = \int_A \pi(x)\,dx$

⑥ Da questão ④ mostre que a probabilidade de que a etapa (c) tenha que ser executada é

$$1 - \frac{\int_{\mathcal{X}} h(x)\,dx}{MZ_q}$$

$$P(X \text{ ser rejeitado no passo (b)}) = \int_{\mathcal{X}} \int_0^1 \mathbb{I}\left(u \leq \frac{h(x)}{M\tilde{q}(x)}\right) q(x)\,du\,dx$$

$$= \int_{\mathcal{X}} \left(1 - \frac{h(x)}{M\tilde{q}(x)}\right) q(x)\,dx$$

$$= \int_{\mathcal{X}} q(x) - \frac{h(x)}{MZ_q}\,dx$$

$$= \underbrace{\int q(x)\,dx}_{1} - \frac{\int_{\mathcal{X}} h(x)\,dx}{MZ_q}$$

$$= 1 - \frac{\int_{\mathcal{X}} h(x)\,dx}{MZ_q}.$$

(7) Da questão (4). Seja $\tilde{\pi}(x) = \exp(-x^2/2)$ e $\tilde{g}(x) = \exp(-|x|)$. Usando o fato que

$$\tilde{\pi}(x) \geq 1 - \frac{x^2}{2}, \quad \forall x \in \mathbb{R},$$

como você poderia usar o algoritmo para amostrar $\pi(x)$. Qual é a probabilidade de não ter que avaliar $\tilde{\pi}(x)$? Por que seria benéfico usar este algoritmo em vez do procedimento padrão de amostragem por rejeição?

Pela sugestão

$$l(x) = \begin{cases} 1 - \frac{x^2}{2} & \text{se } |x| \leq \sqrt{2} \\ 0 & \text{c.c.} \end{cases}$$

Temos

$$\sup \frac{\tilde{\pi}(x)}{\tilde{g}(x)} = \sup_{x \in \mathcal{X}} \exp(-x^2/2 + |x|) = \sqrt{e} \stackrel{\text{def}}{=} M$$

Com probabilidade $\int_{\mathcal{X}} l(x)\,dx / (M Z_g) = \frac{2\sqrt{2}(1 - 2/6)}{\sqrt{e} \times 2} = \frac{2}{3}\sqrt{\frac{2}{e}} \approx 0{,}57$

não temos que avaliar $\tilde{\pi}(x)$

Como $l(x)$ tem ~~significante~~ computação significante mais barata do que $\exp(-x^2/2)$, podemos esperar ganhos.

(8) Considere o seguinte algoritmo conhecido como Marsaglia's polar method.

(a) Gera independentemente $U_1, U_2 \sim Unif[-1,1]$ até $Y = U_1^2 + U_2^2 \leq 1$.

(b) Defina

$$Z = \sqrt{-2\log_2(Y)}$$

e retorne

$$X_1 = Z\frac{U_1}{\sqrt{Y}}, \quad X_2 = Z\frac{U_2}{\sqrt{Y}}.$$

Defina $\vartheta = \arctan(U_2/U_2)$. Mostre que a distribuição conjunta de Y e ϑ tem densidade

$$f_{Y,\vartheta}(y,\theta) = \mathbb{I}_{[0,1]}(y)\frac{\mathbb{I}_{[0,2\pi]}(\theta)}{2\pi}.$$

Podemos interpretar o passo (a) como um algoritmo de rejeição para amostrar de uma distribuição uniforme na esfera unitária D ou seja

$$D := \{(U_1, U_2) \in \mathbb{R} : U_1^2 + U_2^2 \leq 1\}.$$

O mapeamento $(y,\theta) = T(u_1,u_2) = (u_1^2+u_2^2, \arctan(u_1/u_2))$ é bijetiva de D a $(0,1) \times (0,2\pi)$ com inversa $(u_1,u_2) = T^{-1}(y,\theta) = (\sqrt{y}\cos\theta, \sqrt{y}\,\mathrm{sen}\,\theta)$. O Jacobiano de T^{-1} é

$$\left|\det\begin{pmatrix}\frac{\partial u_1}{\partial y} & \frac{\partial u_1}{\partial \theta} \\ \frac{\partial u_2}{\partial y} & \frac{\partial u_2}{\partial \theta}\end{pmatrix}\right| = \left|\det\begin{pmatrix}\frac{\cos\theta}{2\sqrt{y}} & -\sqrt{y}\sin\theta \\ \frac{\sin\theta}{2\sqrt{y}} & \sqrt{y}\cos\theta\end{pmatrix}\right| = \frac{1}{2}$$

então

$$f_{Y,\vartheta}(r,\theta) = f_{U_1,U_2}(T^{-1}(y,\theta)) \cdot \frac{1}{2} = f_{U_2}(\sqrt{y}\cos\theta, \sqrt{y}\,\mathrm{sen}\,\theta) \cdot \frac{1}{2}$$

$$= \mathbb{I}_{[0,1]}(y) \cdot \frac{\mathbb{I}_{[0,2\pi]}(\theta)}{2\pi}.$$

(9) Da questão anterior mostre que X_1 e X_2 são independentes com distribuição normal padrão.

Pela questão anterior $Y \sim U[0,1]$ e $\vartheta \sim U[0,2\pi]$ independentes. Daí $Z^2 = -2\log Y \sim Exp(1/2)$ e $\vartheta \sim U[0,2\pi]$ são independentes. Portanto,

$$X_1 = Z\cos\vartheta, \; X_2 = Z\,\mathrm{sen}\,\vartheta$$

são independentes de normal padrão, pelo Box-Muller algoritmo.

10 Qual os potenciais benefícios de uma abordagem usando Box-Muller algoritmo?

Avaliações de funções ~~trit~~ trigonométricas tendem a ser custosos, o algoritmo ignora estas avaliações.