ESMA 5015: Accept Reject

Alejandro Ouslan

Spring 2025

Contents

1 notes for 2025-03-06

$$
w \sim gamma(\alpha, \beta) \quad \begin{aligned} E(w) &= \alpha\beta \\ var(w) &= \alpha\beta^2 \end{aligned}
$$

$$
\begin{aligned}
F_y(g) &= \frac{1}{2}e^{-y} + \frac{1}{2}ye^{-y} \\
E[y] &= \int_0^\infty y f_y(y)dy \\
&= \frac{1}{2}E[y_1] + \frac{1}{2}E[y_2] \\
y_1 &\sim gamma(1,1) \\
y_2 &\sim gamma(2,1) \\
&= \frac{1}{2}1 + \frac{1}{2}2 \\
&= \frac{3}{2}
\end{aligned}
$$

$$
\begin{aligned}
var(y) &= E[y^2] - E[y]^2 \\
E[y^2] &= \int_0^\infty y^2 f_y(y)dy \\
&= .5E[y_1^2] + .5E[y_2^2] \\
&= \frac{1}{2}2 + \frac{1}{2}6 \\
&= 4
\end{aligned}
$$

$$
\begin{aligned}
E[y_1^2] &= var(y_1) + E[y_1]^2 \\
&= 1 + 1^2 \\
&= 2
\end{aligned}
$$

$$
\begin{aligned}
E[y_2^2] &= var(y_2) + E[y_2]^2 \\
&= 2 + 2^2 \\
&= 6
\end{aligned}
$$

$$
\begin{aligned}
&= 4 - \frac{3}{2}^2 \\
&= 4 - \frac{9}{4} \\
&= \frac{7}{4}
\end{aligned}
$$

2 Consideren un algoritmo Accept-Reject para simular $X \sim N(0,1)$

1. usando $Y \sim cauchy(0,1)$

$$F_Y(y) = \frac{1}{\pi}\frac{1}{1+y^2} \quad y \in \mathbb{R}$$

2. usando $Y \sim double - exponential(0,1)$

$$f_Y(y) = \frac{1}{2}e^{-|y|} \quad y \in \mathbb{R}$$

2.1 theory

1. calculate $M = \sup_x \frac{f_X(x)}{f_Y(y)}$

$$
\begin{aligned}
&= \sup_x \frac{\frac{1}{\sqrt{2\pi}} e^{-x^2/2}}{\frac{1}{\pi}\frac{1}{1+x^2}} \\
&= \sup_x \frac{\pi}{\sqrt{2}} e^{-x^2/2}(1+x^2)
\end{aligned}
$$

sea $h(x) = e^{-x^2/2}(1+x^2)$

$$
\begin{aligned}
\frac{dh(x)}{dx} &= -xe^{-x^2/2}(1+x^2) + e^{-x^2/2}2x \\
&= \vdots \\
&= -xe^{-x^2/2}(x^2-1)
\end{aligned}
$$

$\frac{dh(x)}{dx} = 0 \Rightarrow x = 0, \pm 1$ puede demostra maxiomo occurra en $x = \pm 1 \implies M = \frac{w\sqrt{\pi}}{\sqrt{2}} e^{-1/2} = 1.52$

2.2 algorithm

1. simular $Y \sim cauchy(0,1)$ y $U \sim U(0,1)$ independientes.

2. si

$$
\begin{aligned}
U < \frac{f_X(Y)}{M f_Y(Y)} \\
&= \frac{1}{\sqrt{2\pi}} e^{-1/2} \frac{\sqrt{\pi}}{\sqrt{\pi}}(1+Y^2) \\
&= \frac{1}{2}(1+Y^2) e^{\frac{-Y^2+1}{2}}
\end{aligned}
$$

   aceptar $X = Y$

3. si no, regresar a paso 1.

2.3 steps

1.

$$
\begin{aligned}
M = \sup_x \frac{f_X(x)}{f_Y(y)} &= \sup_x \frac{\int \frac{1}{\sqrt{2\pi}} e^{-x^2/2}}{\frac{1}{2} e^{-|x|}} \\
&= \sup_x \frac{\sqrt{2}}{\sqrt{\pi}} e^{-x^2/2+|x|}
\end{aligned}
$$

$$
h(x) = \ln\left(e^{-x^2/2+|x|}\right)
$$

$$
\frac{dh(x)}{dx} = \frac{d}{dx}\left(-\frac{x^2}{2} + |x|\right) = \begin{cases} -x-1 & x < 0 \\ -x+1 & x > 0 \end{cases}
$$

$$
\frac{dh(x)}{dx} = 0 \Rightarrow x = \pm 1
$$

debe demostrar que el supremo ocure en $x = \pm 1$

2.4 algorithm

1. $M = \frac{\sqrt{2}}{\sqrt{\pi}} e^{-1/2+1} = 1.31$

2. simular $Y \sim double - exponential(0,1)$ y $U \sim U(0,1)$ independientes.

3. si

$$U < \frac{f_X(Y)}{Mf_Y(Y)}$$
$$= e^{\frac{-y^2}{2}+|y|-1/2}$$

entonces aceptar $X = Y$

4. si no, regresar a paso 2.

3 Generar una $X \sim Beta(\alpha, \beta)$

Donte $0 < x < 1$ y $\alpha, \beta > 0$

Utilizar $y \sim U(0, 1)$.

$$M = \sup_x \frac{f_X(x)}{f_Y(y)} = \frac{\frac{\gamma(\alpha+\beta)}{\gamma(\alpha)\gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}}{1}$$
$$h(x) = \ln\left(X^{\alpha-1}(1-x)^{\beta-1}\right)$$
$$\frac{dh(x)}{dx} = \frac{\alpha-1}{x} - \frac{\beta-1}{1-x}$$
$$\frac{dh(x)}{dx} = 0 \Rightarrow x = \frac{\alpha-1}{\alpha+\beta-2} \text{ y } \beta! = 1, \alpha! = 1$$

puede demostrar que en esta X OCURRE UN maxiomo

3.1 algorithm

1. Generar $y \sim U(0, 1)$ y $U \sim U(0, 1)$ independientes.

2. Si

$$u < \frac{f_X(y)}{Mf_Y(y)} = \frac{\frac{\gamma(\alpha+\beta)}{\gamma(\alpha)\gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}}{\frac{\gamma(\alpha+\beta)}{\gamma(\alpha)\gamma(\beta)} \frac{\alpha-1}{\alpha+\beta-2}^{\alpha-1}(1-\frac{\alpha-1}{\alpha+\beta-2})^{\beta-1}}$$

define $X = y$

3. si no, regresar a paso 1.

4 Ejemplo generar $X \sim Beta(\alpha = 2.7, \beta = 6.3)$

$M = 2.6667444$ que ocurre en $x = \frac{2.7-1}{2.7+6.3-2} = 0.2428$

4.1 algorithm

1. Generar $y \sim U(0, 1)$ y $U \sim U(0, 1)$ independientes.

2. Si

$$u < \frac{f_X(y)}{Mf_Y(y)} = \frac{\frac{\gamma(\alpha+\beta)}{\gamma(\alpha)\gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}}{2.666744}$$

Define $X = y$

3. si no, regresar a paso 1.

5 Assignment

para la cauchy y double double-exponential

1. Mathematics
2. Algorithm Implementar codigo y grafica con la distribuicion deseada

$$P(x \leq y) = \int_{-\infty}^{x} f(t)dt$$

6 Bayesian Inference

Accept y Metropolis Hasting (topico de la proxima clase) surgen naturlament en estadisticas Bayesiana. En el Analisis Bayesiano ademas de espexifixar el modelo de los datos observadpos $X = x_1, x_2, \ldots, x_n$ dado un vector de parametros desxonoxidos θ definido por $f(x|\theta)$, se define θ como una variable aleatoria que tiene una distribucion priori $\pi(\theta)$. El Conocimiento de θ se axtualiza con cm el conocimiento que se obtiene de los datos basadosk la inferencia conxerniente a θ. en la distribucion posterior definida por

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta} \text{ Teorema de Bayes}$$

$$\pi(\theta|x)\alpha f(x|\theta)\pi(\theta)$$

6.1 Ejemplo

$x_1, x_2, \ldots, x_n$ iid $Bernulli(\theta)$

6.1.1 Frecuentista

IC de $(1-\alpha)100\%$ para θ

$$\hat{\theta} \pm z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

6.1.2 Bayesiana

$$f(x_1, x_2, \ldots, x_n|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

$$\frac{\pi(\theta|x_1, x_2, \ldots, x_n) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\int_0 \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta}$$

En este caso la distribucion a priori es cojugada com $f(x|\theta)$ porque $\pi(\theta|x)$ pertenece a la misma familia de la distribucion pariori.

6.1.3 Calculation

$$\begin{aligned} E[\theta|x_1, x_2, \ldots, x_n] &= \frac{\sum x_i + \alpha}{\sum x_i\alpha + n - \sum x_i + \beta} \\ &= \frac{\sum x_i + \alpha}{n + \alpha + \beta} \end{aligned}$$

Si n es mucho mayor que $\alpha + \beta$ el promedio posterior se inclina hacia el promedio muestral

6.2 Assignment

Considere $X_1, X_2, \ldots, X_n; \sigma^2$ conocido y la distribucion a priori

$$\pi(\theta) = N(\mu_0, \tau^2)$$

Demuestre que

$$\pi(\theta|x_1, x_2, \ldots, x_n)\ N(\frac{n\hat{x} + \frac{\mu}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}})$$

7 Markor Chain Monte Calro (MCMC)

Definition 1. *Un metodo MCMC para simular de una distibucion F es cualquer metodoque produzxa una cadena de MARKOV erogodica X_N cuya distribucion estacionari es F*

Elejemplos dde estos metodos:

1. Metropolis-Hastings
2. Gibbs Sampling
3. Data Augmentation Algorithm

Tipicamnet ese descartan las variables iniciales de la cadena para "asegurarce" seleccionalr las que tiene la distribucion deseada.(que converga la distribucion)

7.1 Metropolis Hating

Vimos accept -reject . Y sea que es dificil calcular una distribucion cadidata que resulte en una M adecuada ($M > \infty$, of facil de calcular), Metropolis Hatings ofrece una alternativa para simular de una distrobucion F

7.2 Algorithmo

1. Sea $X \sim f_x(x)$target
2. Sea $y \sim f_y(y)$candidate
3. Genera $y \sim f_y(y)$. Define $Z_0 = y$ para $i = 1, 2, 3, \ldots$
4. Genera $U \sim U(0, 1)$
5. Define $P_i = \min\{\frac{f_x(y_i)f_y(z_{i-1})}{f_y(y_i)f_x(z_{i-1})}, 1\}$
6. $Z_i = \begin{cases} y_i \text{ si } u_i \leq P_i \\ z_{i-1} \text{ si } u_i > P_i \end{cases}$
7. ENtonces, $i \to \infty$ Z_i converge en distribucion a la distribucion $F_x(X)$

- Convergenia en distribucion. una secuencia $x_1, x_2, \ldots$ converge en distribucion a una variable alatoria X si $\lim_{n\to\infty} F_{xn}(x) = F_X(X)$ para todas los punetos desde F_X es continua
- Este algoritmo produce lo que se conoce como una cadena de markov que converge a $f_x(x)$ y ino una variable con distribucion $f_x(x)$ y no una variable alatoria con distribucion $f_x(x)$ como comolo hace accept-reject.
- Se destaca las variables alatoria iniciales de la cadena para "asegurar" que la variable alatoria escogida tiene distribucion $F_x(x)$
- La variable aleatora generada no simula idependientes $n\ to\infty$

- Con los metodos MCMC tenemos una cadena acodada y se cumple un resultado abnalogo a la ley de ldo numeros grandes por lo que
$$\frac{1}{n}\sum_{i=1}^{n} g(x_i) \to E[g(X)]$$
aunque X_i no son idependients.
- Pero hubiera necesidad de idepenencia par una aplicaciciopn particualr una alternatica es generar xadena paralela hasta la convergencia yh toma el ultimo elemento de cada uno.

7.2.1 Ejecicio de practica

Compara Accept-Reject con Metropolis Hasting en la estimaxion de $E[x^2]$ cuando $X \sim \Gamma(\alpha, 1)$ usando $Y \sim \Gamma(|\alpha|, \frac{\alpha}{\alpha\mathbb{R}^+})$

Definition 2. *Un asecuancia $x_0, X_1, \ldots, X_n$ de variable aleatoria es una cadenan de Markov si*
$$P(X_n \in A|x_0, x_1, \ldots, x_{n-i}) = P(X_n \in A|x_{n-1})$$