Módulo 5: Aprendizaje no supervisado

Métodos de Muestreo, 1ra Parte

Diplomatura Cs. de Datos - FaCENA-UNNE

Docentes: Magdalena Lucini, Luis Duarte, Griselda Bóbeda

Generación números aleatorios

Supongamos se quiere generar números aleatorios (o pseudoaleatorios). En python, por ejemplo, se puede usar la librería `random` donde, entre otras, se tiene las siguientes funciones:

- `randrange(a, b[, step])`: devuelve un número aleatorio entre a y b con un salto de tamaño *step*
- `randint(a, b)`: retorna un entero aleatorio z tal que $a \leq z \leq b$
- `uniform(a, b)`: devuelve un número real entre a y b.

Generación de distribuciones

En numerosas aplicaciones, métodos y técnicas usadas en machine learning, ciencias de datos y estadística, necesitaremos generar realizaciones de variables aleatorias que siguen alguna distribución probabilística.

Por ejemplo:

- Generar realizaciones (muestras) $x_i, i = 1, \ldots, n$ de una variable aleatoria X con distribución F_X para poder estimar el/los parámetros de esa distribución.
- Calcular (aproximar) la esperanza de una función $f(\mathrm{z})$ con respecto a una distribución de probabilidades $p(\mathrm{z})$ (discreta o continua). $E[f] = \int (f(\mathrm{z}) p(\mathrm{z}) d\mathrm{z}$, $\hat{f} = \frac{1}{L} \sum_{i=1}^{L} f(\mathrm{z}^i)$, $E[\hat{f}] = E[f]$
- Generar muestras de una probabilidad a posteriori $p(\mathrm{z}/\mathcal{D})$ para luego calcular (aproximar) probabilidades marginales, esperanzas, etc.

Métodos de muestreo numérico

- Objetivo: generar eficientemente muestras de una distribución , aún en altas dimensiones. Esto es, generar números aleatorios que se distribuyan según una determinada densidad $p_o(\mathrm{z})$, denominada densidad objetivo.
- Basados en la idea de aproximaciones Monte Carlo
- Cada uno de los algoritmos asume disponible una fuente de números aleatorios con densidad $t(\mathrm{z})$, denominada densidad propuesta y convierte muestras de $t(\mathrm{z})$ a muestras de $p_o(\mathrm{z})$.
- Generando la cantidad de muestras suficientes se puede alcanzar el nivel de precisión deseado.

Diferentes técnicas de muestreo

Métodos directos

- Se generan muestras de números aleatorios cuya densidad es la densidad propuesta, y se usa una transformación adecuada para obtener muestras de la densidad objetivo.
- En general, son métodos rápidos y las muestras son independientes (tan independientes como las generadas por la distribución propuesta).
- Dificilmente aplicables en la práctica pues no es simple encontrar la transformación adecuada.

Diferentes técnicas de muestreo

Métodos de aceptación -rechazo

- Se generan muestras de números aleatorios cuya densidad es la densidad propuesta (disponible), y el algoritmo acepta o rechaza esas muestras mediante un test.
- Las muestras son independientes (tan independientes como las generadas por la distribución propuesta).
- Costo computacional alto si la tasa de aceptación es baja (probabilidad baja de aceptar la muestra)

Diferentes técnicas de muestreo

Importance sampling

- Se asignan pesos a las muestras generadas por la densidad propuesta para obtener una aproximación de la probabilidad representada por la densidad objetivo.
- La aproximación a la densidad objetivo lograda con este método , junto a los principios básicos de MCMC puede ser usada como generador de números aleatorios.

MCMC: Markov Chain Monte Carlo

- Técnicas basadas en la construcción de una cadena de Markov que converge a la densidad objetivo.
- En general, son ,métodos "universales", ya que se pueden aplicar en casi cualquier caso.
- Las muestras no necesariamente serán independientes

Métodos Directos

Consideremos una variable aleatoria X con función de densidad objetivo $p_o(\mathrm{x})$, con función de distribución acumulada $F_X(\mathrm{x})$.

Método de Inversión - distribuciones "simples"

Si $U \sim \mathcal{U}[0,1]$, entonces la variable aleatoria $Z = F_X^{-1}(U)$ tiene densidad $p_o(\mathrm{x})$

Algorithm 1 Método de inversión

generar $u_i \sim \mathcal{U}(0,1)$
devolver $z_i = F_X^{-1}(u_i)$

Ejemplo: Simulación distribución exponencial

Si $X \sim \mathcal{E}(\lambda) \Rightarrow$

$F_X(x) = 1 - e^{-\lambda x}, \quad \mathrm{si\ x} \geq 0$

Como

$1 - e^{-\lambda x} = u \Leftrightarrow x = -\frac{\ln(1-u)}{\lambda}$

Figure: 1000 muestras $\mathcal{U}(0,1)$

generar $u_i \sim \mathcal{U}(0,1)$
devolver $x_i = -\frac{\ln(1-u_i)}{\lambda}$

Figure: 1000 muestras exponencial, $\lambda = 5$

No siempre es posible encontrar la expresión analítica de F_X^{-1}

Ejemplo: Simulación distribución exponencial

Si $X \sim \mathcal{E}(\lambda) \Rightarrow$

$F_X(x) = 1 - e^{-\lambda x}, \quad \text{si x} \geq 0$

Como

$1 - e^{-\lambda x} = u \Leftrightarrow x = -\frac{\ln(1-u)}{\lambda}$

generar $u_i \sim \mathcal{U}(0,1)$ devolver $x_i = -\frac{\ln(1-u_i)}{\lambda}$

Figure: 1000 muestras $\mathcal{U}(0,1)$

Figure: 1000 muestras exponencial, $\lambda = 5$

No siempre es posible encontrar la expresión analítica de F_X^{-1}

Ejemplo: Simulación distribución exponencial

Si $X \sim \mathcal{E}(\lambda) \Rightarrow$

$F_X(x) = 1 - e^{-\lambda x}, \quad \mathrm{si\ x} \geq 0$

Como

$1 - e^{-\lambda x} = u \Leftrightarrow x = -\frac{\ln(1-u)}{\lambda}$

generar $u_i \sim \mathcal{U}(0,1)$
devolver $x_i = -\frac{\ln(1-u_i)}{\lambda}$

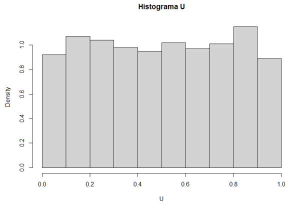


Figure: 1000 muestras $\mathcal{U}(0,1)$

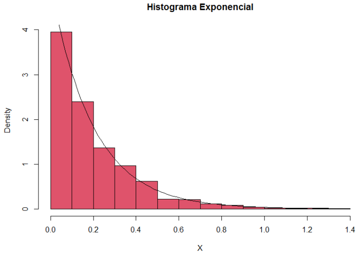


Figure: 1000 muestras exponencial, $\lambda = 5$

No siempre es posible encontrar la expresión analítica de F_X^{-1}

Ejemplo: Simulación distribución exponencial

Si $X \sim \mathcal{E}(\lambda) \Rightarrow$

$F_X(x) = 1 - e^{-\lambda x}, \quad \mathrm{si\ x} \geq 0$

Como

$1 - e^{-\lambda x} = u \Leftrightarrow x = -\frac{\ln(1-u)}{\lambda}$

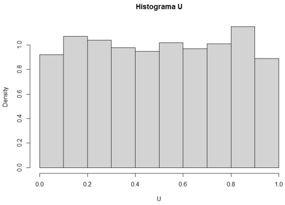


Figure: 1000 muestras $\mathcal{U}(0,1)$

---

generar $u_i \sim \mathcal{U}(0,1)$
devolver $x_i = -\frac{\ln(1-u_i)}{\lambda}$

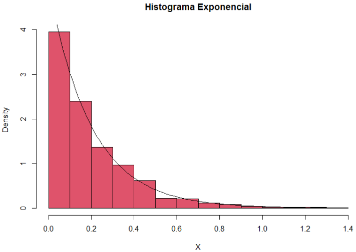


Figure: 1000 muestras exponencial, $\lambda = 5$

No siempre es posible encontrar la expresión analítica de F_X^{-1}

Método de Inversión

Nombre	Densidad	$F(x)$	$F^{-1}(U)$	Forma simplificada
$\exp(\lambda)$ $(\lambda > 0)$	$\lambda e^{-\lambda x}$, si $x \geq 0$	$1 - e^{-\lambda x}$	$-\frac{\ln(1-U)}{\lambda}$	$-\frac{\ln U}{\lambda}$
Cauchy	$\frac{1}{\pi\left(1+x^2\right)}$	$\frac{1}{2} + \frac{\arctan x}{\pi}$	$\tan\left(\pi\left(U - \frac{1}{2}\right)\right)$	$\tan \pi U$
Triangular en $(0, a)$	$\frac{2}{a}\left(1 - \frac{x}{a}\right)$, si $0 \leq x \leq a$	$\frac{2}{a}\left(x - \frac{x^2}{2a}\right)$	$a\left(1 - \sqrt{1-U}\right)$	$a\left(1 - \sqrt{U}\right)$
Pareto ($a, b > 0$)	$\frac{ab^a}{x^{a+1}}$, si $x \geq b$	$1 - \left(\frac{b}{x}\right)^a$	$\frac{b}{(1-U)^{1/a}}$	$\frac{b}{U^{1/a}}$
Weibull ($\lambda, \alpha > 0$)	$\alpha\lambda^\alpha x^{\alpha-1} e^{-(\lambda x)^\alpha}$, si $x \geq 0$	$1 - e^{-(\lambda x)^\alpha}$	$\frac{(-\ln(1-U))^{1/\alpha}}{\lambda}$	$\frac{(-\ln U)^{1/\alpha}}{\lambda}$

Muestreo de aceptación - rechazo (rejection sampling) Von Neuman, 1951

- Método universal: en principio puede usarse para generar muestras de cualquier tipo de densidad objetivo $p_o(x)$.
- Se considera una función de densidad de probabilidad (fdp) propuesta $t(x)$ fácil de muestrear.
- Se elige una constante C tal que la curva $Ct(x)$ está siempre por encima de la curva de $p_o(x)$, i.e,

$$Ct(x) \geq p_o(x), \; \forall x \in \mathcal{S}$$

  con $\mathcal{S}$ dominio de $p_o(x)$.
- Se genera una muestra $x' \sim t$, y se calcula $p_A(x') = \frac{p_o(x')}{Ct(x')}$ que es la probabilidad de aceptar esa muestra.
- Se genera $u' \sim \mathcal{U}(0, 1)$. Si $u' \leq p_A(x') \Rightarrow$ se acepta x', caso contrario se la descarta.

Método de aceptación-rechazo

Algorithm 2 Método de aceptación-rechazo

generar $u \sim \mathcal{U}(0,1)$
generar $x \sim t$
if $u \leq \frac{p_o(x)}{Ct(x)}$ **then**
  aceptar x como muestra
else
  rechazar x como muestra y volver al punto inicial
end if

Ejemplo método de aceptación-rechazo

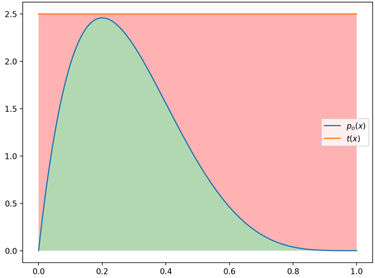
$p_o(x) = 30x(1-x)^4,\ 0 \leq x \leq 1$

$t(x) = 2.5\ 0 \leq x \leq 1$ mayoriza a $p_o(x)$
NO es fdp

$r(x) = t(x)/k\ :\ k = \int_{-\infty}^{\infty} t(x)dx$
$r(x) = 1,\ 0 \leq x \leq 1$

$F_r(x) = x,\ 0 \leq x \leq 1$

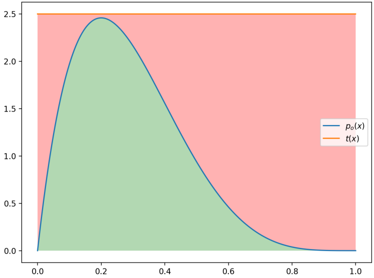
Si es fdp



Pasos

1. Generar x al "azar". ¿ Cómo?
2. P(aceptar x) $= \frac{p_o(x)}{t(x)}$
3. Generar $u_x \sim \mathcal{U}(0,1)$
4. Si $u_x \leq$ P(aceptar x) $\rightarrow$ devolver x
   Si $u_x >$ P(aceptar x) $\rightarrow$ ir a 1

Ejemplo método de aceptación-rechazo



$p_o(x) = 30x(1-x)^4$, $0 \leq x \leq 1$

$t(x) = 2.5$ $0 \leq x \leq 1$ mayoriza a $p_o(x)$
NO es fdp

$r(x) = t(x)/k \ : \ k = \int_{-\infty}^{\infty} t(x)dx$
$r(x) = 1, \ 0 \leq x \leq 1$

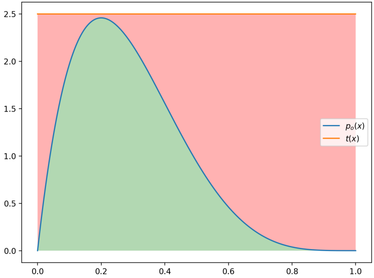
$F_r(x) = x, \ 0 \leq x \leq 1$

Si es fdp

Pasos

1. Generar x al "azar". ¿ Cómo?
2. P(aceptar x) $= \frac{p_o(x)}{t(x)}$
3. Generar $u_x \sim \mathcal{U}(0,1)$
4. Si $u_x \leq$ P(aceptar x) $\rightarrow$ devolver x
   Si $u_x >$ P(aceptar x) $\rightarrow$ ir a 1

Ejemplo método de aceptación-rechazo



$p_o(x) = 30x(1-x)^4,\ 0 \leq x \leq 1$

$r(x) = \frac{t(x)}{k}$
$F_r(x) = x,\ \ 0 \leq x \leq 1$

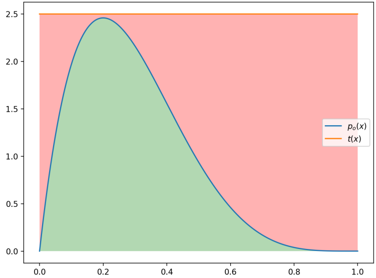
Transformada Inversa $x = F_r^{-1}(u_r)$,
$0 \leq u_r \leq 1$

$u_r \sim \mathcal{U}(0,1)$

Pasos

1. Generar $u_r \sim \mathcal{U}(0,1)$
   Generar $x \sim r(x)$ usando u_r
2. $\mathrm{P}(\text{aceptar } x) = \frac{p_o(x)}{t(x)}$
3. Generar $u_x \sim \mathcal{U}(0,1)$
4. Si $u_x \leq \mathrm{P}(\text{aceptar } x) \rightarrow$ devolver x
   Si $u_x > \mathrm{P}(\text{aceptar } x) \rightarrow$ ir a 1

Ejemplo método de aceptación-rechazo



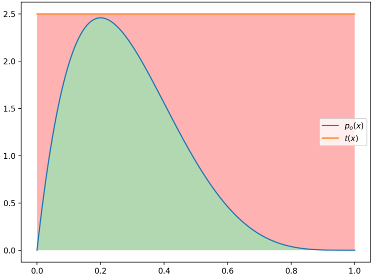
Pasos

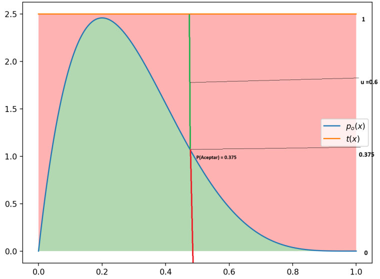
1. Generar $u_r \sim \mathcal{U}(0,1)$
   Generar $x \sim r(x)$ usando u_r
2. $\text{P}(\text{aceptar } x) = \frac{p_o(x)}{t(x)}$
3. Generar $u_x \sim \mathcal{U}(0,1)$
4. Si $u_x \leq \text{P}(\text{aceptar } x) \rightarrow$ devolver x
   Si $u_x > \text{P}(\text{aceptar } x) \rightarrow$ ir a 1

Pasos

1. $u_r = 0.5$
   $x = 0.5$
2. $\text{P}(\text{aceptar } x) = \frac{0.9375)}{2.5} = 0.375$
3. u_x=0.6
4. rechaza x y vuelve al punto 1

Ejemplo método de aceptación-rechazo





Pasos

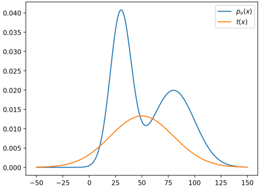
1. Generar $u_r \sim \mathcal{U}(0,1)$
   Generar $x \sim r(x)$ usando u_r
2. $\mathrm{P}(\text{aceptar } x) = \frac{p_o(x)}{t(x)}$
3. Generar $u_x \sim \mathcal{U}(0,1)$
4. Si $u_x \leq \mathrm{P}(\text{aceptar } x) \rightarrow$ devolver x
   Si $u_x > \mathrm{P}(\text{aceptar } x) \rightarrow$ ir a 1

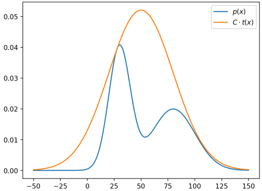
Pasos

1. $u_r = 0.5$
   $x = 0.5$
2. $\mathrm{P}(\text{aceptar } x) = \frac{0.9375)}{2.5} = 0.375$
3. u_x=0.6
4. rechaza x y vuelve al punto 1

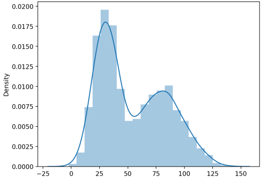
Método de aceptación -rechazo

- Puede demostrarse que la P(Aceptar X)$= 1/C$ y está relacionado con la elección de la función $t(x)$
- Lo ideal es elegir $t(x)$ lo más parecida posible a $p_o(x)$





C$=\max_x(\frac{p_o(x)}{t(x)})$



N = 5000 muestras

$$\hat{\mathbb{E}}_p[X] = \frac{1}{N}\sum_{i=1}^{N} x^i = 55.27$$

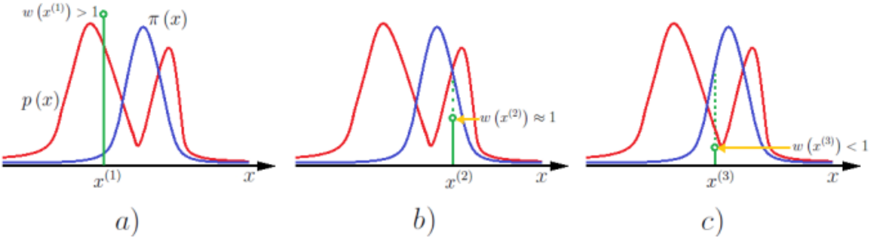
Muestreo por Importancia (importance sampling)

- Proveen de un marco estadístico para aproximar momentos (esperanzas, varianzas) bajo una distribución $p_o(x)$ cuando no es posible muestrearla directamente.
- NO es un método para generar muestras de la fdp objetivo p_o.
- En el muestreo por importancia se asocian *pesos* a las muestras producidas por la densidad propuesta $t(x)$
- Estos *pesos* representan la importancia estadística de la muestra
- Los pesos son $w(x') = \frac{p_o(x')}{t(x')} \propto \frac{p(x')}{t(x')}$, donde $p(x) \propto p_o(x)$
- La importancia es:
  - directamente proporcional a $p(x)$: mientras más alto el valor de $p_o(x')$, más mas de probabilidad estará concentrada alrededor de x' y por ende hay que darle más importancia a esa muestra.
  - inversamente proporcional a $t(x)$: si $t(x')$ es bajo, la probabilidad de proponer otra muestra x' es baja, por lo que hay que darle importancia a esa muestra.

Muestreo por importancia

Situaciones en muestreo por importancia

- $p(x')$ alto, $t(x')$ bajo: la importancia $w(x')$ será grande (Fig a).
- $p(x') \approx t(x')$la importancia $w(x') \approx 1$ (Fig b)
- $p(x')$ bajo, $t(x')$ alto: la importancia $w(x')$ será pequeña. (Fig c)
- En general, los pesos miden cuán distintas son las fdp objetivo y propuesta.



Muestreo por Importancia

$$
\begin{aligned}
\mathbb{E}_p[f(X)] &= \int_{-\infty}^{\infty} f(x) p_o(x)\ dx \\
&= \int_{-\infty}^{\infty} f(x) \frac{p_o(x)}{t(x)} t(x)\ dx \\
&= \int_{-\infty}^{\infty} \left(\frac{f(x) p_o(x)}{t(x)} \right) t(x)\ dx \\
&= \mathbb{E}_t \left[\frac{f(Y)\ p_o(Y)}{t(Y)} \right]
\end{aligned}
$$

$$
\begin{aligned}
\hat{\mathbb{E}}_p[f(X)] &= \hat{\mathbb{E}}_t \left[\frac{f(Y)\ p_o(Y)}{t(Y)} \right] \\
&= \frac{1}{N} \sum_{i=1}^{N} \frac{f(x^i) p_o(x^i)}{t(x^i)}
\end{aligned}
$$

Puede ser necesario normalizar la fdp objetivo

Muestreo por importancia

Método más general (no se conocen constantes de normalización)

Algorithm 3 Muestreo por importancia

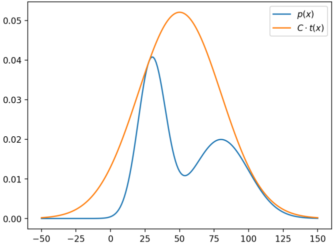
1. generar N muestras x^i, $i = 1, \ldots, N$ de la fdp propuesta $t(x)$
2. asociar pesos a cada uno de los x^i haciendo

$$w_i = \frac{p(x^i)}{t(x^i)}$$

, con $p(x) \propto p_o(x)$

3. normalizar los pesos $\tilde{w}_i = \frac{w_i}{\sum_{k=1}^{N} w_k}$, $i = 1, \ldots N$
4. Se aproxima a $p_o(x)$ usando $\tilde{h}_N(x) = \sum_{i=1}^{N} \tilde{w}_i \delta(x - x^i)$

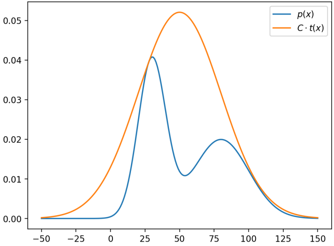
Cálculo de esperanzas usando Muestreo por Importancia



- $p_o(x)$ fdp $\mathcal{N}(30, 10) + \mathcal{N}(80, 20)$, $t(x)$ fdp $N(50, 30)$
- X con fdp $p_o(x)$
- N $=$5000
- generar 5000 muestras de $t(x)$
- calcular pesos
- $\hat{\mathbb{E}}[X] = \sum_{i=1}^{N} \tilde{w}_i x^i = 50$

La elección de la densidad propuesta afecta al desempeño de este método

Cálculo de esperanzas usando Muestreo por Importancia



- $p_o(x)$ fdp $\mathcal{N}(30, 10) + \mathcal{N}(80, 20)$, $t(x)$ fdp $N(50, 30)$
- X con fdp $p_o(x)$
- N $=$5000
- generar 5000 muestras de $t(x)$
- calcular pesos
- $\hat{\mathbb{E}}[X] = \sum_{i=1}^{N} \tilde{w}_i x^i = 50$

La elección de la densidad propuesta afecta al desempeño de este método

Sampling Importance Resampling (SIR)

- El éxito de los métodos de rechazo depende en parte de la elección de la constante C
- Para muchos pares de combinaciones de fdp objetivo y propuesta se dificulta encontrar una constante C que sea lo suficientemente grande para acotar la distribución objetivo y a la vez garantizar una alta tasa de aceptación.
- El método Sampling Importance Resampling (SIR) hace uso de la fdp propuesta t para generar muestras de la fdp objetivo p_o, pero evita el paso de determinar la constante C.
- Básicamente consiste de dos pasos: generación de muestras en el primer paso, mientras que en el segundo se calculan pesos como en el muestreo porimportancia y se genera (resampling) un segundo conjunto de muestras usando esos pesos.
- Las muestras resultantes se utilizan para aproximar p_o. La aproximación es buena para muestras grandes ($N \to \infty$)

Sampling Importance Resampling

Objetivo: generar muestras de la fdp objetivo p_o.

Algorithm 4 SIR

1. generar N muestras y^i, $i = 1, \ldots, N$ de la fdp propuesta t
2. calcular los pesos

$$w_i = \frac{p_o(y^i)}{t(y^i)}$$

y normalizarlos $\tilde{w}_i = \frac{w_i}{\sum_{k=1}^{N} w_k}$, $i = 1, \ldots N$

3. Para $i = 1, \ldots, N$ se elige un nuevo conjunto de muestras $(x^1, \ldots, x^N)$ (con reemplazo) usando los pesos como probabilidades:

$$x^i = \begin{cases} y^1 & \text{con probabilidad } \tilde{\mathrm{w}}_1 \\ \vdots & \vdots \\ y^N & \text{con probabilidad } \tilde{\mathrm{w}}_\mathrm{N} \end{cases}$$

Aplicación SIR: Ejercicio

Objetivo: Generar muestras de una variable aleatoria con fdp objetivo p_o definida como la suma de gaussianas $\mathcal{N}(30, 10) + \mathcal{N}(80, 20)$

Usar como fdp propuesta:

1. $t1$ fdp de una variable uniforme en $\mathcal{U}(0, 110)$
2. $t2$ fdp de $\mathcal{N}(50, 30)$.