

MAT-468: Sesión 1, Simulación Estocástica

Felipe Osorio

<http://fosorios.mat.utfsm.cl>






Departamento de Matemática, UTFSM



1. Álgebra lineal numérica.
2. Solución de ecuaciones no lineales y optimización.
3. Generación de variables aleatorias.
4. Métodos Monte Carlo.
5. Aplicaciones en Estadística.



Bibliografía básica

-  Gentle, J.E. (2007).
Matrix Algebra: Theory, Computation and Applications in Statistics.
Springer, New York.
-  McLachlan, G.J., Krishnan, T. (2008).
The EM Algorithm and Extensions.
Wiley, New York.
-  Monahan, J.F. (2011).
Numerical Methods of Statistics.
Cambridge University Press, Cambridge.
-  Robert, C.P., Casella, G. (2004).
Monte Carlo Statistical Methods.
Springer, New York.
-  Tanner, M.A. (1996).
Tools for Statistical Inference.
Springer, New York.



Observación

Una mala implementación puede hacer que un buen algoritmo sea **inútil**.

- ▶ Algoritmos pobres: El caso de Microsoft Excel
 - ▶ Knüsel (1998, 2002, 2005).
 - ▶ McCullogh y coautores (1999, 2002a, 2002b, 2005, 2008).
 - ▶ Pottel (2001).
- ▶ Alternativas como Gnumeric o R, disponen de algoritmos robustos.



Considere el siguiente ejemplo:

$$\sum_{i=1}^n (x_i - \bar{x}_n)^2 = \sum_{i=1}^n x_i^2 - n\bar{x}_n^2 = (n-1)s^2.$$

Esto da origen (al menos) a dos algoritmos

- ▶ Algoritmo de 1 paso.
- ▶ Algoritmo de 2 pasos.



Cálculo de la varianza muestral. Algoritmo de 2-pasos¹

Algoritmo 1: Varianza muestral usando un algoritmo de 2-pasos.

Entrada: Conjunto de n datos $\mathbf{x} = (x_1, \dots, x_n)^\top$.

Salida : Promedio y varianza muestrales, \bar{x} y s^2 .

```
1 begin
2    $M \leftarrow x_1$ 
3   for  $i = 2$  to  $n$  do
4     |  $M \leftarrow M + x_i$ 
5   end
6    $M \leftarrow M/n$ 
7    $T \leftarrow (x_1 - M)^2$ 
8   for  $i = 2$  to  $n$  do
9     |  $T \leftarrow T + (x_i - M)^2$ 
10  end
11   $\bar{x} \leftarrow M$ 
12   $s^2 \leftarrow \frac{1}{n-1}T$ 
13 end
```

¹Desventajas: Requiere dos ciclos for



Cálculo de la varianza muestral. Algoritmo de 1-paso²

Algoritmo 2: Varianza muestral usando un algoritmo de 1-paso.

Entrada: Conjunto de n datos $x = (x_1, \dots, x_n)^\top$.

Salida : Promedio y varianza muestrales, \bar{x} y s^2 .

```
1 begin
2    $M \leftarrow x_1$ 
3    $T \leftarrow x_1^2$ 
4   for  $i = 2$  to  $n$  do
5      $M \leftarrow M + x_i$ 
6      $T \leftarrow T + x_i^2$ 
7   end
8    $\bar{x} \leftarrow M/n$ 
9    $s^2 \leftarrow \frac{1}{n-1}T - \frac{n}{n-1}\bar{x}^2$ 
10 end
```

²**Desventajas:** Puede generar cancelamientos 'catastróficos'



Un algoritmo online (West, 1979)

Considere

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i.$$

De este modo,

$$\begin{aligned}\bar{x}_n &= \frac{1}{n} \left(\sum_{i=1}^{n-1} x_i + x_n \right) = \frac{1}{n} \left((n-1)\bar{x}_{n-1} + x_n \right) \\ &= \frac{1}{n} \left(n\bar{x}_{n-1} - \bar{x}_{n-1} + x_n \right) \\ &= \bar{x}_{n-1} + \frac{\delta_n}{n},\end{aligned}\tag{1}$$

con $\delta_n = x_n - \bar{x}_{n-1}$.

Ecuación en (1) corresponde a un **algoritmo recursivo** para el cálculo del promedio.



Un algoritmo online (West, 1979)

Análogamente, podemos considerar la suma de cuadrados,

$$T_n = \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

Podemos escribir,

$$\begin{aligned} T_n &= \sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2 + (x_n - \bar{x}_n)^2 \\ &= \sum_{i=1}^{n-1} \left[(x_i - \bar{x}_{n-1})^2 - 2\frac{\delta_n}{n}(x_i - \bar{x}_{n-1}) + \frac{\delta_n^2}{n^2} + \left(\delta_n - \frac{\delta_n}{n}\right)^2 \right]. \end{aligned}$$

Notando que $\sum_{i=1}^{n-1} (x_i - \bar{x}_{n-1}) = 0$, obtenemos:

$$T_n = \sum_{i=1}^{n-1} (x_i - \bar{x}_{n-1})^2 + \left(1 - \frac{1}{n}\right)\delta_n^2 = T_{n-1} + \left(1 - \frac{1}{n}\right)\delta_n^2. \quad (2)$$

Ecuaciones (1) y (2) llevan al siguiente algoritmo.



Cálculo de la varianza muestral. Algoritmo online (1-paso)³

Algoritmo 3: Promedio y varianza muestrales usando un algoritmo online.

Entrada: Conjunto de n datos $\mathbf{x} = (x_1, \dots, x_n)^\top$.

Salida : Promedio y varianza muestrales, \bar{x} y s^2 .

```
1 begin
2    $M \leftarrow x_1$ 
3    $T \leftarrow 0$ 
4   for  $i = 2$  to  $n$  do
5      $\delta \leftarrow (x_i - M)/i$ 
6      $M \leftarrow M + \delta$ 
7      $T \leftarrow T + i(i - 1)\delta^2$ 
8   end
9    $\bar{x} \leftarrow M$ 
10   $s^2 \leftarrow \frac{1}{n - 1}T$ 
11 end
```

³**Ventajas:** Es un algoritmo estable que sólo requiere de un paso.



Otra variante del algoritmo online (West, 1979)

Sea

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{1}{n-1} T_n,$$

y usando que $T_{n-1} = ((n-1) - 1)s_{n-1}^2$. Podemos escribir,

$$\begin{aligned} s_n^2 &= \frac{1}{n-1} \left[(n-1)s_{n-1}^2 - s_{n-1}^2 + \left(1 - \frac{1}{n}\right)\delta_n^2 \right] \\ &= s_{n-1}^2 + \frac{1}{n-1} \left[(n-1)\frac{\delta_n^2}{n} - s_{n-1}^2 \right]. \end{aligned} \quad (3)$$

Observación:

Ecuación (3) lleva a una ligera modificación del [Algoritmo 3](#) para el cálculo de \bar{x}_n y s_n^2 .



Inestabilidad en el cálculo de s^2

Chan y Lewis (1979) introdujeron el **número condición** para evaluar la sensibilidad de s^2 en una muestra $\mathbf{x} = (x_1, \dots, x_n)^\top$, como:

$$\kappa = \frac{\|\mathbf{x}\|_2}{\sqrt{n-1}s}.$$

Suponga que introducimos una perturbación

$$\mathbf{x}(\varepsilon) = \mathbf{x} + \varepsilon\mathbf{u}, \quad \mathbf{u} = (u_1, \dots, u_n)^\top,$$

tal que $\|\mathbf{u}\|_2 = \|\mathbf{x}\|_2$, y defina

$$s^2(\varepsilon) = \frac{1}{n-1} \sum_{i=1}^n (x_i(\varepsilon) - \bar{x}(\varepsilon))^2, \quad \bar{x}(\varepsilon) = \bar{x} + \varepsilon\bar{u}.$$

Es posible mostrar (Tarea) que:

$$\frac{|s(\varepsilon) - s|}{s} \leq \varepsilon\kappa + O(\varepsilon^2).$$



Inestabilidad en el cálculo de s^2

Chan y Lewis (1979) introdujeron el **número condición** para evaluar la sensibilidad de s^2 en una muestra $\mathbf{x} = (x_1, \dots, x_n)^\top$, como:

$$\kappa = \frac{\|\mathbf{x}\|_2}{\sqrt{n-1}s}.$$

Suponga que introducimos una perturbación

$$\mathbf{x}(\varepsilon) = \mathbf{x} + \varepsilon\mathbf{u}, \quad \mathbf{u} = (u_1, \dots, u_n)^\top,$$




tal que $\|\mathbf{u}\|_2 = \|\mathbf{x}\|_2$, y defina

$$s^2(\varepsilon) = \frac{1}{n-1} \sum_{i=1}^n (x_i(\varepsilon) - \bar{x}(\varepsilon))^2, \quad \bar{x}(\varepsilon) = \bar{x} + \varepsilon\bar{u}.$$

Es posible mostrar (**Tarea**) que:

$$\frac{|s(\varepsilon) - s|}{s} \leq \varepsilon\kappa + O(\varepsilon^2).$$



-  Golub, G.H., van Loan, C.F. (1996)
Matrix Computations.
The Johns Hopkins University Press, Baltimore.
-  Dongarra, J.J., Moler, C.B., Bunch, J.R., Stewart, G.W. (1979).
Linpak Users' Guide.
Society for Industrial and Applied Mathematics (SIAM), Philadelphia.
-  Lawson, C., Hanson, R., Kincaid, D., Krogh, F. (1979).
Basic linear algebra subprograms for Fortran usage.
ACM Trans. Math. Software 5, 308-371.



Operaciones $O(n)$: BLAS nivel 1

DSCAL: Escala un vector: $\mathbf{x} \leftarrow \alpha \mathbf{x}$.

DDOT: Producto punto: $\mu \leftarrow \mathbf{x}^\top \mathbf{y}$.

DNRM2: Norma Euclidiana: $\mu \leftarrow \|\mathbf{x}\|_2$.

DASUM: Suma valores absolutos: $\mu \leftarrow \sum_{i=1}^n |x_i|$.

DCOPY: Copia un vector: $\mathbf{y} \leftarrow \mathbf{x}$.

DSWAP: Intercambia vectores: $\mathbf{x} \leftrightarrow \mathbf{y}$.

DAXPY: Actualización 'AXPY': $\mathbf{y} \leftarrow \alpha \mathbf{x} + \mathbf{y}$.



Ciclos for “enrollados”

```
subroutine daxpy(n,a,x,y)
double precision x(*),y(*),a
integer n
```

```
!
! constant times a vector plus a vector.
!
```

```
if (n.le.0) return
if (a .eq. 0.0d0) return
```

```
!
```

```
do i = 1,n
  y(i) = y(i) + a * x(i)
end do
```

```
return
```



Ciclos for “desenrollados”

```
subroutine daxpy(n,a,x,y)
double precision x(*),y(*),a
integer n
```

```
!  
!  
! constant times a vector plus a vector.  
! uses unrolled loops for increments equal to one.  
! jack dongarra, linpack, 3/11/78.  
!
```

```
if (n.le.0) return  
if (a .eq. 0.0d0) return
```

```
!  
m = mod(n,4)  
mp1 = m + 1  
do i = mp1,n,4  
  y(i)      = dy(i)      + a * x(i)  
  y(i + 1) = y(i + 1) + a * x(i + 1)  
  y(i + 2) = y(i + 2) + a * x(i + 2)  
  y(i + 3) = y(i + 3) + a * x(i + 3)  
end do
```

```
return
```



Operaciones $O(n^2)$ y $O(n^3)$: BLAS niveles 2 y 3⁴

Nivel 2: Operaciones **matriz-vector**:

$$\text{DGEMV: } \mathbf{y} \leftarrow \alpha \mathbf{A}\mathbf{x} + \beta \mathbf{y}.$$

DTRSV: Resuelve $\mathbf{A}\mathbf{x} = \mathbf{b}$, o $\mathbf{A}^\top \mathbf{x} = \mathbf{b}$, \mathbf{A} triangular.

$$\text{DGER: } \mathbf{A} \leftarrow \mathbf{A} + \alpha \mathbf{x}\mathbf{y}^\top.$$

Nivel 3: Operaciones **matriz-matriz**:

$$\text{DGEMM: } \mathbf{C} \leftarrow \alpha \mathbf{A}\mathbf{B} + \beta \mathbf{C}.$$

$$\text{DSYRK: } \mathbf{C} \leftarrow \alpha \mathbf{A}\mathbf{A}^\top + \beta \mathbf{C}.$$

⁴Implementación utiliza **BLAS nivel 1**



Ejemplo: GAXPY

Considere una “versión generalizada” del axpy:

$$\mathbf{y} \leftarrow \mathbf{A}\mathbf{x} + \mathbf{y},$$

donde $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$ y $\mathbf{y} \in \mathbb{R}^m$.

De este modo, para realizar la multiplicación

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 7 \\ 8 \end{pmatrix} = 7 \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} + 8 \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 23 \\ 53 \\ 83 \end{pmatrix}.$$

Podemos invocar “repetidas veces” a la función: $\mathbf{y} \leftarrow \alpha\mathbf{x} + \mathbf{y}$ (axpy).



Definición 1 (Descomposición LDL)

Si $A \in \mathbb{R}^{n \times n}$ es matriz simétrica y no singular, entonces existe L matriz triangular inferior y $D = \text{diag}(d_1, \dots, d_n)$, tal que

$$A = LDL^T.$$

Definición 2 (Descomposición Cholesky)

Si $A \in \mathbb{R}^{n \times n}$ es simétrica y definida positiva, entonces existe una única matriz triangular inferior $G \in \mathbb{R}^{n \times n}$ (factor Cholesky) con elementos diagonales positivos, tal que

$$A = GG^T.$$



Definición 1 (Descomposición LDL)

Si $A \in \mathbb{R}^{n \times n}$ es matriz simétrica y no singular, entonces existe L matriz triangular inferior y $D = \text{diag}(d_1, \dots, d_n)$, tal que

$$A = LDL^T.$$

Definición 2 (Descomposición Cholesky)

Si $A \in \mathbb{R}^{n \times n}$ es simétrica y definida positiva, entonces existe una única matriz triangular inferior $G \in \mathbb{R}^{n \times n}$ (factor Cholesky) con elementos diagonales positivos, tal que

$$A = GG^T.$$



Descomposición Ortogonal-Triangular (QR)

Definición 3 (Descomposición QR)

Sea $A \in \mathbb{R}^{m \times n}$, entonces existe $Q \in \mathcal{O}_m$ y $R \in \mathbb{R}^{m \times n}$, tal que

$$A = QR,$$

donde

$$R = \begin{pmatrix} R_1 \\ 0 \end{pmatrix}$$

con $R_1 \in \mathbb{R}^{n \times n}$ matriz triangular superior, aquí suponemos que $m \geq n$.

Observación: Si $\text{rg}(A) = r$, entonces las primeras n columnas de Q forman una base ortonormal para $\mathcal{M}(A)$.

Note que, si $A = QR$ entonces

$$A^T A = R^T Q^T Q R = R^T R = R_1^T R_1,$$

y R_1 corresponde al factor Cholesky de $A^T A$.



Descomposición Ortogonal-Triangular (QR)

Definición 3 (Descomposición QR)

Sea $A \in \mathbb{R}^{m \times n}$, entonces existe $Q \in \mathcal{O}_m$ y $R \in \mathbb{R}^{m \times n}$, tal que

$$A = QR,$$

donde

$$R = \begin{pmatrix} R_1 \\ 0 \end{pmatrix}$$

con $R_1 \in \mathbb{R}^{n \times n}$ matriz triangular superior, aquí suponemos que $m \geq n$.

Observación: Si $\text{rg}(A) = r$, entonces las primeras n columnas de Q forman una base ortonormal para $\mathcal{M}(A)$.

Note que, si $A = QR$ entonces

$$A^T A = R^T Q^T Q R = R^T R = R_1^T R_1,$$

y R_1 corresponde al factor Cholesky de $A^T A$.



Definición 4 (Descomposición de Schur)

Sea $A \in \mathbb{C}^{n \times n}$. Entonces existe una matriz unitaria $U \in \mathbb{C}^{n \times n}$ ($U^H U = I$) y una matriz triangular M cuyos elementos diagonales son los valores propios de A , tal que

$$U^H A U = M.$$

Definición 5 (Descomposición espectral)

Sea $A \in \mathbb{C}^{n \times n}$ matriz Hermitiana ($A^H = A$). Entonces existe una matriz unitaria $U \in \mathbb{C}^{n \times n}$ tal que

$$U^H A U = \Lambda,$$

donde $\Lambda = \text{diag}(\lambda)$ es matriz diagonal cuyos elementos diagonales son los valores propios de A .

⁵ $Z \in \mathbb{C}^{m \times n}$, $Z = A + iB$, $Z^H = A^T - iB^T$



Definición 4 (Descomposición de Schur)

Sea $\mathbf{A} \in \mathbb{C}^{n \times n}$. Entonces existe una matriz unitaria $\mathbf{U} \in \mathbb{C}^{n \times n}$ ($\mathbf{U}^H \mathbf{U} = \mathbf{I}$) y una matriz triangular \mathbf{M} cuyos elementos diagonales son los valores propios de \mathbf{A} , tal que

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \mathbf{M}.$$

Definición 5 (Descomposición espectral)

Sea $\mathbf{A} \in \mathbb{C}^{n \times n}$ matriz Hermitiana ($\mathbf{A}^H = \mathbf{A}$). Entonces existe una matriz unitaria $\mathbf{U} \in \mathbb{C}^{n \times n}$ tal que

$$\mathbf{U}^H \mathbf{A} \mathbf{U} = \mathbf{\Lambda},$$

donde $\mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ es matriz diagonal cuyos elementos diagonales son los valores propios de \mathbf{A} .

⁵ $\mathbf{Z} \in \mathbb{C}^{m \times n}$, $\mathbf{Z} = \mathbf{A} + i\mathbf{B}$, $\mathbf{Z}^H = \mathbf{A}^\top - i\mathbf{B}^\top$



Definición 4 (Descomposición SVD)

Sea $\mathbf{A} \in \mathbb{R}^{m \times n}$ con $\text{rg}(\mathbf{A}) = r$, entonces existen matrices $\mathbf{U} \in \mathcal{O}_m$, $\mathbf{V} \in \mathcal{O}_n$, tal que

$$\mathbf{A} = \mathbf{U} \begin{pmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{V}^\top,$$

donde $\mathbf{D}_r = \text{diag}(\delta_1, \dots, \delta_r)$ con $\delta_1 \geq \delta_2 \geq \dots \geq \delta_r > 0$, que son llamados **valores singulares** de \mathbf{A} .



Aplicaciones: Evaluar la densidad normal multivariada

Tenemos que $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\boldsymbol{\Sigma} > 0$ tiene función densidad

$$f(\mathbf{y}) = |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}) \right\}, \quad \mathbf{y} \in \mathbb{R}^p.$$

De este modo, si **deseamos evaluar $f(\mathbf{y})$** debemos calcular $|\boldsymbol{\Sigma}|$ y $\boldsymbol{\Sigma}^{-1}$, lo que puede ser realizado muy eficientemente usando la **descomposición Cholesky**.

Note que,

$$|\boldsymbol{\Sigma}| = |\mathbf{G}\mathbf{G}^\top| = |\mathbf{G}||\mathbf{G}^\top| = |\mathbf{G}|^2 = \left(\prod_{i=1}^p g_{ii} \right)^2,$$

mientras que

$$\boldsymbol{\Sigma}^{-1} = (\mathbf{G}\mathbf{G}^\top)^{-1} = \mathbf{G}^{-\top} \mathbf{G}^{-1}.$$

De este modo, para evaluar $(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ basta resolver el **sistema triangular** $\mathbf{G}\mathbf{z} = (\mathbf{y} - \boldsymbol{\mu})$ y hacer

$$\mathbf{z}^\top \mathbf{z} = \{\mathbf{G}^{-1}(\mathbf{y} - \boldsymbol{\mu})\}^\top \{\mathbf{G}^{-1}(\mathbf{y} - \boldsymbol{\mu})\}.$$



Podemos obtener A^{-1} , mediante resolver el sistema lineal matricial

$$AX = I$$

El método preferido para resolver el sistema anterior es considerar la **descomposición QR**. Es decir,

$$(QR)X = I \quad \implies \quad RX = Q^T,$$

y luego, obtenemos $X (= A^{-1})$ resolviendo el sistema triangular $RX = Q^T$.

Finalmente, la **matriz inversa**⁶ puede ser escrita como:

$$X = R^{-1}Q^T.$$

⁶Algunas de estas operaciones se pueden realizar "In place".



Aplicaciones: Evaluar la exponencial de una matriz

Considere la **exponencial de una matriz**:

$$\exp(\mathbf{A}) = e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{1}{2}\mathbf{A}^2 + \frac{1}{3!}\mathbf{A}^3 + \dots = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!},$$

donde $\mathbf{A}^0 = \mathbf{I}$.

Si \mathbf{A} es matriz simétrica $n \times n$, entonces podemos escribir su **descomposición espectral**⁷ como $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\top}$ con $\mathbf{U} \in \mathcal{O}_n$ y $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$. De este modo

$$\begin{aligned}\exp(\mathbf{A}) &= \sum_{k=0}^{\infty} \frac{\{\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\top}\}^k}{k!} = \mathbf{I} + \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\top} + \frac{1}{2}\mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^{\top} + \dots \\ &= \mathbf{U} \left\{ \sum_{k=0}^{\infty} \frac{\mathbf{\Lambda}^k}{k!} \right\} \mathbf{U}^{\top} = \mathbf{U} e^{\mathbf{\Lambda}} \mathbf{U}^{\top},\end{aligned}$$

donde $\exp(\mathbf{\Lambda}) = \text{diag}(e^{\lambda_1}, \dots, e^{\lambda_n})$.

⁷Este es un caso particular de la **caracterización Jordan** (basado en JCF).

