EST-46114 Métodos Multivariados **Nombre:** ____________________________

Primavera 2018 **C.U.:** ____________

Tarea 1 **1/Feb/2018**

---

1. Suponga que un conjunto de n datos p dimensionales antes de ser observados pueden describirse con la variables aleatorias $X_1, \ldots, X_n$. Supongamos que las X_is condicionalmente independientes con distribución homogénea, $N(x|\mu, \Sigma)$; siendo μ un vector p dimensional y Σ una matriz de dimensión $p \times p$ simétrica positivo definida. La descripción del desconocimiento acera de la especificidad de esta distribución es complementada suponiendo que (μ, Σ) son aleatorias y que ambas siguen una distribución normal-WishartInversa $(\mu, \Sigma) \sim N - WiI(\mu, \Sigma | m_0, s_0, S_0, n_0)$, con m_0, s_0, S_0, n_0 hiperparámetros fijos. Esquemáticamente, nos referimos al siguiente modelo bayesiano,

$$
\begin{aligned}
X_i|\mu, \Sigma &\sim^{i.i.d.} N(x|\mu, \Sigma) &(1)\\
\mu|\sigma &\sim N(\mu|m_0, s_0\Sigma) &(2)\\
\Sigma &\sim WiI(\Sigma|n_0, S_0). &(3)
\end{aligned}
$$

   Con base en lo anterior, deriva la distribución condicional de (μ, Σ) dado un conjunto de datos observados $x_1, \ldots, x_n$.

2. Con base en lo anterior, deriva la expresión analítica de la descomposición espectral de Σ. Así, desarrolla una función (en `R`, `Python` o `Matlab`) que:

   - lea un conjunto de datos genéericos escalares $x_1, \ldots, x_n$ expreseados como aun matriz de $n \times p$
   - lea el conjunto de hiperparámetros m_0, s_0, S_0, n_0
   - lea un número arbirtrario entero M
   - genere un conjunto de tamaño M de datos simulados *i.i.d.* $\left(\mu^{(m)}, \Sigma^{(m)}\right)_{m=1}^{M}$ de la distribción final del inciso anterior
   - a cada $\Sigma^{(m)}$ le genere su descomposición espectral
   - almacede los datos simulados y los correspondientes a la descomposición espectral.

   Por invarianza, se tiene que los datos $\left(\Lambda^{(m)}, P^{(m)}\right)$ de la descomposición espectral corresponden a una muestra *i.i.d.* de la distribución condicional de (Λ, P) dados los datos, $x_1, \ldots, x_n$, cuya distribución existe pero no es necesaria de evidenciar analíticamente (es muy complejo!).

   Hint: Para las derivaciones analíticas apoyense en el libro Press (2005, cap.3).