Lema de Fisher-Cochran: *Sean $X_1, X_2, \ldots$ v.a.i.i.d. con distribución $N(\mu, \sigma)$. Sean*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad y \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Entonces

a) *$\bar{X}$ y S^2 son v.a. independientes;*

b) $\bar{X} \sim N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$

c) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

En la demostración del lema de Fisher-Cochran utilizaremos un corolario derivado del siguiente resultado (para demostrarlo basta utilizar la función característica de una gamma):

Lema (gamma): *Si $\xi_1, \ldots, \xi_n$ son v.a. independientes con $\xi_i \sim gamma(a, p_i)$, $i = 1, \ldots, n$, es decir, con densidad*

$$f_{\xi_i}(x) = \frac{a^{p_i}}{\Gamma(p_i)} x^{p_i - 1} e^{-ax}, \qquad x > 0,$$

entonces $\sum_{i=1}^{n} \xi_i \sim gamma(a, \sum_{i=1}^{n} p_i)$.

Como consecuencia de este lema sobre la distribución gamma y del hecho de que la distribución χ^2_p es una gamma(1/2,p/2), tenemos el siguiente resultado sobre la distribución χ^2:

Corolario (χ^2): *Si $\xi_1, \ldots, \xi_n$ son v.a. independientes con $X_i \sim \chi^2_{p_i}$, $i = 1, \ldots, n$, entonces $\sum_{i=1}^{n} \xi_i \sim \chi^2_{\sum_{i=1}^{n} p_i}$.*

Demostración del lema de Fisher-Cochran: Primero observemos que podemos suponer sin pérdida de generalidad que $\mu = 0$ y $\sigma = 1$. Si $\mu \neq 0$ ó $\sigma \neq 1$, basta expresar $\bar{X}$ y S^2 como

$$\bar{X} = \mu + \frac{\sigma}{n}\sum_{i=1}^{n} Z_i = \mu + \sigma\bar{Z} \qquad \text{y} \qquad S^2 = \frac{\sigma^2}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2,$$

siendo $Z_i = (X_i - \mu)/\sigma$ (o, equivalentemente, $X_i = \mu + \sigma Z_i$) y hacer la demostración para las Z_i, que son N(0,1).

Suponemos, pues, que $X \sim N(0,1)$ y queremos probar que $\bar{X} \sim N\left(0, \dfrac{1}{\sqrt{n}}\right)$ y que $(n-1)S^2 \sim \chi^2_{n-1}$.

a) Basta probar que $\bar{X}$ y S^2 son, respectivamente, funciones de vectores aleatorios independientes entre sí. Primero observemos que podemos escribir S^2 como función de sólo $n-1$ desviaciones a la media muestral:

$$S^2 = \frac{1}{n-1}\left((X_1 - \bar{X})^2 + \sum_{i=2}^{n}(X_i - \bar{X})^2\right) = \frac{1}{n-1}\left(\left[\sum_{i=2}^{n}(X_i - \bar{X})\right]^2 + \sum_{i=2}^{n}(X_i - \bar{X})^2\right) \tag{1}$$

En la última igualdad hemos aplicado que $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$ y, por tanto, $X_1 - \bar{X} = -\sum_{i=2}^{n}(X_i - \bar{X})$. La última expresión de (1) quiere decir que S^2 se puede escribir como función de sólo el vector aleatorio $(X_2 - \bar{X}, \ldots, X_n - \bar{X})$.

A continuación probaremos que $\bar{X}$ y $(X_2-\bar{X},\ldots,X_n-\bar{X})$ son "objetos" aleatorios independientes. Por ejemplo, basta calcular la función de densidad conjunta $f(y_1,y_2,\ldots,y_n)$ del vector aleatorio

$$(Y_1,Y_2,\ldots,Y_n)=(\bar{X},X_2-\bar{X},\ldots,X_n-\bar{X})$$

y ver que se puede expresar como producto de las funciones de densidad $f(y_1)$ de $Y_1=\bar{X}$ y $f(y_2,\ldots,y_n)$ de $(Y_2,\ldots,Y_n)=(X_2-\bar{X},\ldots,X_n-\bar{X})$.

Para hallar $f(y_1,y_2,\ldots,y_n)$ observemos que

$$\begin{aligned}
y_1 &= \bar{x}=\frac{1}{n}x_1+\ldots+\frac{1}{n}x_n\\
y_2 &= x_2-\bar{x}=-\frac{1}{n}x_1+\frac{n-1}{n}x_2-\frac{1}{n}x_3-\ldots-\frac{1}{n}x_n\\
&\vdots\\
y_n &= x_n-\bar{x}=-\frac{1}{n}x_1-\ldots-\frac{1}{n}x_{n-1}+\frac{n-1}{n}x_n
\end{aligned}$$

es decir,

$$\begin{pmatrix} y_1\\ \vdots\\ y_n\end{pmatrix}=\begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n}\\ -\frac{1}{n} & \frac{n-1}{n} & \cdots & -\frac{1}{n}\\ \vdots & \vdots & & \vdots\\ -\frac{1}{n} & -\frac{1}{n} & \cdots & \frac{n-1}{n}\end{pmatrix}\begin{pmatrix} x_1\\ \vdots\\ x_n\end{pmatrix}. \tag{2}$$

La función (2) es lineal con jacobiano

$$\left|\frac{\partial(y_1,y_2,\ldots,y_n)}{\partial(x_1,x_2,\ldots,x_n)}\right|=\frac{1}{n}.$$

Por otro lado,

$$\begin{aligned}
x_1 &= y_1-y_2-\ldots-y_n\\
x_2 &= y_2+y_1\\
&\vdots\\
x_n &= y_n+y_1.
\end{aligned}$$

Entonces

$$\begin{aligned}
f(y_1,y_2,\ldots,y_n) &= f(x_1,x_2,\ldots,x_n)\left|\frac{\partial(x_1,x_2,\ldots,x_n)}{\partial(y_1,y_2,\ldots,y_n)}\right|\\
&= \frac{n}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}\left[\left(y_1-\sum_{i=2}^n y_i\right)^2+\sum_{i=2}^n(y_i+y_1)^2\right]\right) \qquad (3)\\
&= \underbrace{\left(\frac{n}{2\pi}\right)^{1/2}e^{-\frac{n}{2}y_1^2}}_{\text{función sólo de } y_1}\underbrace{\left(\frac{n}{(2\pi)^{n-1}}\right)^{1/2}e^{-\frac{1}{2}\left(\sum_{i=2}^n y_i^2+\left(\sum_{i=2}^n y_i\right)^2\right)}}_{\text{función sólo de } y_2,\ldots,y_n}. \qquad (4)
\end{aligned}$$

Para pasar de (3) a (4) basta desarrollar el cuadrado que aparece en cada uno de los dos exponentes y simplificar términos coincidentes.

b) Sea $M_X(t) = \mathbb{E}(e^{tX})$ la función generatriz de momentos (f.g.m.) de la v.a. X. Como $X \sim N(0,1)$, tenemos que $M_X(t) = e^{t^2/2}$. En general, la f.g.m. de una $N(\mu, \sigma)$ es $M(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$. La f.g.m. de la media muestral es

$$\begin{aligned} M_{\bar{X}}(t) &= \mathbb{E}\left(e^{t\bar{X}}\right) = \mathbb{E}\left(e^{\frac{t}{n}\sum_{i=1}^n X_i}\right) \\ &= \prod_{i=1}^n \mathbb{E}\left(e^{\frac{t}{n}X_i}\right) = \left(M_X\left(\frac{t}{n}\right)\right)^n = \exp\left(\frac{t^2}{2n}\right), \end{aligned}$$

que corresponde a una $N(0, 1/\sqrt{n})$. Evidentemente la demostración de (b) se puede hacer de manera totalmente equivalente usando la función característica $\mathbb{E}(e^{itX})$ en lugar de la f.g.m.

c) Para la prueba de este apartado utilizaremos un argumento de inducción. Denotamos por $\bar{X}_k$ y S_k^2 la media y la varianza muestrales de las primeras k observaciones, $X_1, \ldots, X_k$, de la muestra, es decir,

$$\bar{X}_k = \frac{1}{k}\sum_{i=1}^k X_i \qquad \text{y} \qquad S_k^2 = \frac{1}{k-1}\sum_{i=1}^k (X_i - \bar{X}_k)^2.$$

Se puede comprobar que, para todo $k = 2, 3, 4, \ldots$, se cumple que

$$(k-1)S_k^2 = \sum_{i=1}^k (X_i - \bar{X}_k)^2 = (k-2)S_{k-1}^2 + \frac{k-1}{k}(X_k - \bar{X}_{k-1})^2. \tag{5}$$

Primero analizamos el caso $k = 2$. Por (5) tenemos que

$$S_2^2 = \frac{1}{2}(X_2 - X_1)^2 = \left(\frac{X_2 - X_1}{\sqrt{2}}\right)^2.$$

Como las X_i son N(0,1) independientes, se tiene que $(X_2 - X_1)/\sqrt{2} \sim$ N(0,1) y, por tanto, $S_2^2 \sim \chi_1^2$, es decir, el apartado (c) se cumple para $n = 2$.
Ahora suponemos que el apartado (c) se cumple para $n = k$, lo que significa que $(k-1)S_k^2 \sim \chi_{k-1}^2$. Vamos a probar que también se cumple para $n = k+1$. Por (5) tenemos que

$$kS_{k+1}^2 = (k-1)S_k^2 + \frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2.$$

Observemos que $X_{k+1} - \bar{X}_k \sim N(0, \sqrt{(k+1)/k}$, por ser diferencia de dos normales independientes. Entonces

$$\frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2 \sim (N(0,1))^2 \equiv \chi_1^2.$$

Por otro lado, el apartado (a) y la independencia de las X_i implican que S_k^2 y $\bar{X}_k$ son independientes, luego $(k-1)S_k^2$ y $\frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2$ son v.a. independientes con distribuciones χ_{k-1}^2 y χ_1^2 respectivamente. Aplicando el corolario sobre la χ^2, se obtiene que kS_{k+1}^2 sigue una χ_k^2.

$\square$