Distribución Normal Multivariada

Hamdi Raïssi

IES PUCV

hamdi.raissi@pucv.cl

Objetivos de esta parte

Objetivo 1 : Teóricos

a- Revisar los conocimientos de variables gaussianas univariadas.

b- Estudiar la generalización al caso multivariado.

c- Profundizar con ejercicios en clase.

Objetivo 2 : Prácticos

a- Tener una idea intuitiva de la distribución gaussiana multivariada con simulaciones.

b- Ilustrar el uso del conocimiento de los vectores gaussianos para la comprensión de herramientas y modelos en estadística

Bibliografía : Methods of Multivariate Analysis. A.C. Rencher, Wiley (2002).

Distribución normal en estadística tiene un rol central

En regresión lineal :

Supuesto normal del vector de los errores

$\Rightarrow$ Prueba de significatividad de parámetros

$\Rightarrow$ Detección de outliers

$\Rightarrow$ Intervalos de confianza de predicciones

$\Rightarrow$ etc....

Distribución normal en estadística tiene un rol central

Ley límite del TCL :

Muchas pruebas utilizan la ley normal

$\Rightarrow$ A veces queremos hacer prueba no solo sobre un parámetro, pero sobre varios en conjunto.

$\Rightarrow$ Se utiliza la distribución normal multivariada.

Hay mucho más ejemplos....

Función caracteristica de una variable aleatoria.

Definición

$\phi_X(t) = E\left[\exp(itX)\right].$

- Se escribe entonces de forma más explicita :

$$\phi_X(t) = \int_{\mathbb{R}} e^{itx} f_X(x) dx,$$

donde $f_X(x)$ es la densidad de X.

- Si X y Y son independientes, entonces $\phi_{(X+Y)}(t) = \phi_X(t)\phi_Y(t)$.

Repaso : Distribución normal univariada.

- Densidad de la ley normal estandar :

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

- Mostrar que es una densidad.
- Calcular la esperanza y la varianza de una variable que sigue la ley normal estandar.
- Mostrar que la función caracteristica de una variable aleatoria que sigue una ley normal estandar es dada por : $\exp(-t^2/2)$.

De manera general si $X \sim \mathcal{N}(m, \sigma^2)$, entonces su funcción caracteristica es dada por :

$$\phi_X(t) = \exp\left(itm - \frac{1}{2}\sigma t^2\right),$$

Función caracteristica de un vector aleatorio.

Un vector aleatorio $X = (X_1, \ldots, X_p)'$.

Definición

$\phi_X(t) = E\left[\exp(i < t, X >)\right]$, donde $< ., . >$ es el producto escalar (o sea $< t, X >= \sum_{j=1}^{p} t_j X_j$.

Si los X_j son independientes, entonces $\phi_X(t) = \phi_{X_1}(t) \times \cdots \times \phi_{X_p}(t)$.
Caso bivariado :

$$\phi_X(t) = \int\int_{\mathbb{R}^2} e^{i(t_1 x_1 + t_2 x_2)} f_{(X_1,X_2)}(x_1, x_2) dx_1 dx_2,$$

donde $f_{(X_1,X_2)}(x_1, x_2)$ es la densidad de X.

Distribución normal multivariada.

Definición

Un vector aleatorio $X = (X_1, \ldots, X_p)'$ es un vector aleatorio gaussiano si cualquiera combinación lineal $\sum_{j=1}^{p} a_j X_j$ sigue una ley normal.

En adelante anotamos $X \sim \mathcal{N}_p(m, \Sigma)$, donde :

- $m = E(X)$
- $\Sigma = V(X)$ (matriz de covarianza)

Ejemplo : $X_1, \ldots, X_p$ son estimadores de parametros (por ejemplo promedios). Si $X = (X_1, \ldots, X_p)'$ es un vector aleatorio gaussiano, entonces combinación lineal $\sum_{j=1}^{p} a_j X_j$ sigue una ley normal.

Matriz de covarianza.

$$
\begin{aligned}
\Sigma &= E\{(X - E(X))(X - E(X))'\} \\
&= \begin{pmatrix}
V(X_1) & \cdots & \cdots & \cdots \\
\vdots & V(X_2) & cov(X_i, X_j) & \vdots \\
\vdots & cov(X_j, X_i) & \ddots & \vdots \\
\vdots & \cdots & \cdots & V(X_d)
\end{pmatrix}
\end{aligned}
$$

Matriz de covarianza.

- Por definición, Σ es simetrica $(cov(X_i, X_j) = cov(X_j, X_i))$
- Vamos a suponer que Σ es definida positiva (entonces invertible).

$\Rightarrow$ No hay variables redundantes (si hay redundencia, la matriz de covarianza no sería invertible).

- Entonces existe una matriz P tal que $P\Sigma P' = I_p$ y $P'P = \Sigma^{-1}$.

Distribución normal multivariada.

- Eso implica que cada componente sigue la ley normal (elegir $a_j = 1$ y el resto 0).
- El revés no es verdad cuando los componentes no son independientes. Podemos tener X_1 sigue $\mathcal{N}(0,1)$, y $X_2 = (2\epsilon - 1)X_1$ sigue la ley normal, con ϵ sigue una ley $\mathcal{B}(0.5)$. Sin embargo

$$0 < P(X_1 + X_2 = 0) = 0.5.$$

  (Si $X_1 + X_2$ hubiese sido de ley normal entonces $P(X_1 + X_2 = 0) = 0$, con $Var(X_1 + X_2) > 0$.)
- La definición implica el caso de la ley Dirac (el caso donde $\Sigma = 0$) !

Simulación de variables normales multivariadas en R : Caso no normal, aunque que sus componentes son normales.

```
n<-3000
X1<-rnorm(n)
e<-2*rbinom(n, 1, 0.5)-1
X2<-e*X1
plot(X1, X2, main="No Gaussian!!")
```

Función caracteristica de una distribución normal multivariada.

Definición 2 (equivalente al anterior)

La función caracteristica de un vector aleatorio gaussiano $X = (X_1, \ldots, X_p)'$ es dada por :

$$\phi_X(t) = \exp\left(it'm - \frac{1}{2}t'\Sigma t\right),$$

$t \in \mathbb{R}^p$ y donde se recuerda que

- $E(X) = m$
- $V(X) = \Sigma$ (matriz de covarianza)

Un caso particular.

$X \sim \mathcal{N}(0, I_p)$:

- $X \sim \mathcal{N}(0, I_p) \Leftrightarrow$ los X_i son independientes gaussianos.
- $\Leftarrow$ los X_i son independientes entonces sus covarianzas son ceros.
- $\Rightarrow \phi_X(t) = \phi_{X_1}(t) \dots \phi_{X_p}(t)$ (aplicando la definición anterior con $\Sigma = I_p$).

A partir de $X \sim \mathcal{N}(0, I_p)$, podemos construir (y simular) cualquiera ley normal :

Propiedad

Sea Y un vector gaussiano, de esperanza m y de varianza Σ definida positiva. Entonces existe un vector gaussiano $X \sim \mathcal{N}(0, I_p)$, tal que $X = P(Y - m)$.

- Tenemos $E(Y - m) = 0$.
- Podemos definir $X = P(Y - m)$, y ver que este vector es gaussiano y tal que $E(X) = 0$ y $V\{X\} = P\Sigma P' = I_p$ (propiedad de las matrices definidas positivas asegura la existencia de P).
- Utilizar la linealidad de la esperanza en los dos puntos anteriores.

Distribución condicional.

Propiedad

Sea $Z = (X', Y')' \sim \mathcal{N}(m, \Sigma)$, donde X y Y son vectores gaussianos, Σ es definida positiva, y

$$m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Entonces

$$Y \mid X = x \sim \mathcal{N}(m_2 + \Sigma_{21}\Sigma_{11}^{-1}(x - m_1), \widetilde{\Sigma}_{22}),$$

donde $\widetilde{\Sigma}_{22} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.

Distribución condicional.

Observaciones :

- La mejor predicción de Y dado X es lineal.
- $Y \mid X = x$ no depende de x si $\Sigma_{21} = 0$. De hecho Y y X son independientes en este caso....
- La varianza de $Y \mid X = x$ no depende del valor x que tomada por X. Es el supuesto de homoscedasticidad (vamos a ver la heteroscedasticidad en el contexto de la regresión).
- En el caso dependiente, la varianza condicional es más chica que la varianza de Y :

$$\Sigma_{22} - \widetilde{\Sigma}_{22} = \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \geq 0.$$

La interpretación es que la información pertinente de X ayuda a comprender Y.

Densidad de una distribución normal multivariada.

Propiedad

La densidad de una variable normal multivariada es :

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-m)'\Sigma^{-1}(x-m)\right)$$

Correlación de Pearson

- Densidad de la ley normal bivariante $(X,Y)'$:

$$
\begin{aligned}
f(x,y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\exp\left(-\frac{1}{2(1-r^2)}\left(\frac{(x-E(X))^2}{\sigma_x^2}\right.\right.\\
&+ \left.\left.\frac{(y-E(Y))^2}{\sigma_y^2} - \frac{2r(x-E(X))(y-E(Y))}{\sigma_x\sigma_y}\right)\right)
\end{aligned}
$$

- σ_x desviación estándar de X y
- El coeficiente de correlación de Pearson r entre X y Y :

$$
r = \frac{E\left\{(X-E(X))(Y-E(Y))\right\}}{\sigma_x\sigma_y}
$$

- Si $r=0$ la densidad de $(X,Y)'$ es igual al producto de la densidad de X y Y $\Rightarrow$ Independancia.

Observación : r es entre -1 y 1.

Simulación de variables normales multivariadas en R : Caso bivariado.

```
library(mnormt)
n<-3000
mu <- c(0,0)
Sigma <- matrix(c(1,0,0,1), 2, 2)
mat<-rmnorm(n, mu, Sigma)
plot(mat[,1], mat[,2], main="Gaussian")
```

Simulación de variables normales multivariadas en R : Caso 3D.

```
library(mnormt)
n<-3000
mu <- c(0,0,0)
Sigma <- matrix(c(1,0,0,0,1,0,0,0,1), 3, 3)
mat<-rmnorm(n, mu, Sigma)
library(scatterplot3d)
scatterplot3d(mat[,1 :3])
```

TCL multivariado.

Generalización del TCL univariado.

Propiedad

Sean $X_1, X_2, ..., X_n$ vectores aleatorios iid con $E(X_1) = m$ y $\Sigma = Var(X_1) < \infty$. Anotamos $\overline{X} = 1/n \sum_{i=1}^{n} X_i$. Entonces

$$\sqrt{n}(\overline{X} - m) \Rightarrow \mathcal{N}(0, \Sigma),$$

cuando $n \to \infty$.

Justifica pruebas y otros herramientas multivariados cuando no estamos en el caso gaussiano. Cuidado que n debe ser "grande" para que se aplica este teórema.

Aplicación 1 : Construir tests estadísticos

El siguiente resultado es la consecuencia del teorema de Cochrane :

Propiedad

Sea un vector aleatorio $X = (X_1, \ldots, X_n)'$ tal que $X \sim \mathcal{N}(0, \sigma^2 I_n)$. Sea $M_n = 1/n \sum_{i=1}^n X_i$ y $V_n = 1/n \sum_{i=1}^n (X_i - M_n)^2$. Entonces tenemos :

$$\sqrt{n} M_n \sim \mathcal{N}(0,1) \quad \text{y} \quad nV_n \sim \chi^2_{n-1}.$$

Más encima M_n y V_n son independientes.

- De este resultado obtenemos el muy utilizado test de Student, utilizado en muchas situaciones en estadística $(S = \sqrt{n} M_n / \sqrt{\frac{n}{n-1} V_n} \sim T_{n-1})$.
- Ver ejercicio en la guía para ver la prueba de este resultado sorprendente a primer vista !

Aplicación 2 : Vectores gaussianos y regresión

Sea $(Y, X_1, \ldots, X_d)'$ un vector gaussiano.

- De la propriedad de distribución condicional que hemos visto, se deduce que $E(Y \mid X_1, \ldots, X_p)$ tiene una forma lineal.
- En practica, eso significa que si los datos siguen la ley normal, el mejor predictor es entregado por el modelo lineal.
- De forma clásica se utilisa el modelo de regresión lineal :

$$\begin{cases} Y = X\beta + \epsilon, \\ E(\epsilon) = 0, \\ Var(\epsilon) = \sigma^2 I_n, \quad \epsilon \quad \text{es un vector gaussiano} \end{cases}$$

donde n es el número de observaciones y $X = (X_1, \ldots, X_p)$ matriz $n \times p$ contiene los regresores.

Aplicación 3 : Heteroscedasticidad y autocorrelación

En algunos casos la hipótesis de no correlación de los errores es cuestionable :

- Los errores son autocorreladas cuando las variables son temporales o espaciales.

Consecuencias :

- La estimación por MCO no es necesariamente la mejor
- Las pruebas que hacemos bajo las hipótesis estandard ya no son válidos.

Aplicación 3 : Heteroscedasticidad y autocorrelación

Consideramos el modelo :

$$\begin{cases} Y = X\beta + \epsilon, \\ E(\epsilon) = 0, \\ Var(\epsilon) = \Sigma, \quad \epsilon \quad \text{es un vector gaussiano} \end{cases}$$

- Σ es una matriz simetrica definida positiva $\Rightarrow \quad \exists$ una matriz P $n \times n$ invertible tal que :

  $$P\Sigma P' = I_n.$$

  $(P'P = \Sigma^{-1})$.
- Recuerdamos que la hipótesis estandard es : $Var(\epsilon) = \sigma^2 I_n$

Aplicación 3 : Heteroscedasticidad y autocorrelación

- Los errores son autocorreladas : $Cov(\epsilon_i, \epsilon_j) = \Sigma_{ij}$, $i \neq j$.
- Las varianzas no son constantes : $Var(\epsilon_i) = \Sigma_{ii}$.

Este modelo es más general (engloba) el modelo estandard (es un caso particular).

Aplicación 3 : Heteroscedasticidad y autocorrelación

Eliminamos los efectos indeseables escribiendo :

$$PY = PX\beta + P\epsilon.$$

o sea definiendo $\widetilde{Y} = PY$, $\widetilde{X} = PX$ y $\widetilde{\epsilon} = P\epsilon$, tenemos

$$\begin{cases} \widetilde{Y} = \widetilde{X}\beta + \widetilde{\epsilon}, \\ E\left(\widetilde{\epsilon}\right) = PE\left(\epsilon\right) = 0, \\ Var\left(\widetilde{\epsilon}\right) = PVar\left(\epsilon\right)P' = I_n. \end{cases}$$

El estimador MCO de este modelo es :

$$\boxed{\widehat{\beta}_{MCG} = \left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'\widetilde{Y} = \left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}Y}$$

Es el estimador mínimos cuadrados generalizados.

Aplicación 3 : Heteroscedasticidad y autocorrelación

Propiedades :

- Bajo autocorrelación y heteroscedasticidad la varianza del estimador MCO $\hat{\beta}_{MCO} = (X'X)^{-1}X'Y$ es :

$$Var(\hat{\beta}_{MCO}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$$

- La varianza del estimador MCG es :

$$Var(\hat{\beta}_{MCG}) = (X'\Sigma^{-1}X)^{-1}$$

- Bajo autocorrelación y heteroscedasticidad el estimador $\hat{\beta}_{MCG}$ es BLUE.

Aplicación 3 : Heteroscedasticidad y autocorrelación

Se mejora la estimación del modelo lineal (ver código R).

Aplicación 3 : Heteroscedasticidad y autocorrelación

Conclusión :

- El estimador MCO es meno preciso que el estimador MCG.
- Los tests Student y Fisher estandard ya no son válidos.

Observaciones de puntos no vistos ahora :

- Para construir el estimador MCG es necesario disponer de un estimador de Σ obtenido en un etapa preliminar.
- Es importante de tener en cuenta la heteroscedasticidad y la autocorrelación de los datos si necesario ... y entonces de ser capaz de detectarlos !

Aplicación 4 : Medir distancias en estadística

Definición-Propiedad

Sea dos vectorios aleatorios gaussianos $X = (X_1, \ldots, X_n)'$ y $Y = (Y_1, \ldots, Y_n)'$ tal que $X - Y \sim \mathcal{N}(0, \Sigma)$, donde Σ es definida positiva. Definimos la distancia de Mahalanobis como

$$D = \sqrt{(X-Y)'\Sigma^{-1}(X-Y)}.$$

Tenemos $D^2 \sim \chi^2_n$.

El último punto puede servir a probar si la distancia de Mahalanobis es significativa.

Aplicación 4 : Medir distancias en estadística

- De esta propiedad podemos medir distancias entre individuos : sirve en "clustering".
- Es diferente de la distancia Euclidiana = $\sqrt{(X-Y)'(X-Y)}$.
- La diferencia es que la distancia de Mahalanobis toma en cuenta las varianzas (y las correlaciones) de las variables que sirven a medir distancias.

Aplicación 4 : Medir distancias en estadística

Ejemplo :

- Estamos haciendo clustering de individuos según tres caracteristicas X_1, X_2, X_3.
- Tenemos evidencia que el vector sigue una ley normal.
- Sin embargo la tercera variable es medida con una incertitumbre grande (varianza alta consecuencía de equipo de medición de mala calidad quizás).
- Las dos primera son muy correlacionadas (información redundante hasta cierto punto).

La distancia de Mahalanobis permite de resolver esos problemas...