Tema 6. Variables aleatorias continuas

Resumen del tema

6.1. Definición de variable aleatoria continua

- *Identificación de una variable aleatoria continua X*: es preciso conocer su **función de densidad**, $f(x)$, que debe verificar:
  - ⋆ $f(x) \geq 0$ para todo número real x.
  - ⋆ El área total bajo la curva $y = f(x)$ vale 1.
  - ⋆ La probabilidad de que la variable aleatoria X esté comprendida entre a y b, $P(a \leq X \leq b)$, viene determinada por el área bajo la curva $y = f(x)$ entre $x = a$ y $x = b$.
- Los valores concretos de la función de densidad no tienen ningún significado especial pues las probabilidades vienen determinadas por áreas bajo la curva determinada por la función de densidad y no por valores de la función de densidad. En todo caso, este hecho nos informa de que en las distribuciones continuas la probabilidad de que la variable aleatoria tome un valor concreto, $P(X = a)$, es cero, como corresponde al área de un rectángulo de base un punto y altura $f(a)$. Resumiendo, si X es una variable aleatoria continua, entonces:

$$P(X = a) = 0 \quad \text{para todo } a\,.$$

- La **representación gráfica de la función de densidad** de una variable aleatoria continua es equivalente al polígono de frecuencias relativas de una variable estadística continua cuando la amplitud de los intervalos es infinitesimal.
- La **media** y la **varianza** de una variable aleatoria continua se determinan mediante una operación matemática denominada *integral*.
- La **función de distribución** de una variable aleatoria continua X se define igual que para cualquier variable aleatoria; es decir:

$$F_X(t) = P(X \leq t) \quad \text{para todo } t\,.$$

- El valor de $F_X(t)$ coincide con el área bajo la curva $y = f(x)$ desde el valor más pequeño que puede tomar la variable hasta el valor t.
- Para algunas variables aleatorias continuas los resultados de la función de distribución se pueden determinar con cualquier paquete estadístico, como MINITAB o SPSS.
- Si X es una variable aleatoria continua, entonces:
  - ⋆ $P(X < a) = P(X \leq a) = F_X(a)$ para todo a.
  - ⋆ $P(X > a) = P(X \geq a) = 1 - F_X(a)$ para todo a.
  - ⋆ $P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = F_X(b) - F_X(a)$ para todo a y b.

6.2. La distribución Normal

6.2.1. Distribución Normal

Una variable aleatoria continua X tiene una distribución **Normal de parámetros μ y σ** si su función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \qquad \text{para todo } x\,,$$

donde μ es cualquier número, σ es cualquier número positivo y, en general, $\exp(t)$ significa e^t, siendo e la base de los logaritmos neperianos.

Son equivalentes las dos afirmaciones siguientes: "X tiene una distribución Normal de parámetros μ y σ" y "X es una variable aleatoria Normal de parámetros μ y σ".

La variable aleatoria Normal de parámetros μ y σ será denotada por:

$$\mathcal{N}(\mu, \sigma)\,.$$

Se cumplen las siguientes propiedades:

- La media, la mediana y la moda de una variable aleatoria $\mathcal{N}(\mu, \sigma)$ coinciden entre sí y tienen por valor al parámetro μ.
- La desviación típica de la distribución $\mathcal{N}(\mu, \sigma)$ es igual al parámetro σ.
- La curva que representa a la función de densidad de la distribución $\mathcal{N}(\mu, \sigma)$ es simétrica respecto de la recta vertical de ecuación $x = \mu$.
- El área comprendida entre el eje horizontal y la curva que representa a la función de densidad de la distribución $\mathcal{N}(\mu, \sigma)$ vale 1 (como ocurre con cualquier distribución continua).

6.2.2. Distribución Normal Estándar

A la variable aleatoria Normal de parámetros 0 y 1 se le llama variable aleatoria **Normal Estándar**, o *Normal Típica*, y se le denota por $\mathcal{N}(0, 1)$.

6.2.3. Uso de la tabla de la función de distribución

La tabla de la función de distribución de la variable aleatoria Normal Estándar, Z, da las probabilidades a la izquierda de números positivos; es decir, $P(Z \leq t)$, con $t > 0$. A partir de las propiedades de simetría y de que el área total bajo la curva de densidad es la unidad, pueden deducirse todos los casos: probabilidades a la izquierda o a la derecha de números positivos o negativos.

6.2.4. Uso de la tabla de los cuantiles

Además de tener tabulados los resultados de la función de distribución de la variable aleatoria Normal Estándar, también tenemos tabulados los valores inversos de la función de distribución; es decir, los cuantiles.

El **cuantil** (o percentil) al $100p\,\%$ de la variable aleatoria Normal Estándar se denota por Z_p y es el valor que verifica:

$$P(\mathcal{N}(0, 1) \leq Z_p) = p\,,$$

es decir, el área comprendida entre la curva de densidad de la distribución $\mathcal{N}(0,1)$ y el eje horizontal, a la izquierda de Z_p, es igual a p.

Otra interpretación es la siguiente: el valor Z_p deja por debajo el $100p\,\%$ de todos los resultados de una variable aleatoria Normal Estándar.

El resultado de Z_p se puede determinar con cualquier paquete estadístico, como MINITAB o SPSS (para cualquier valor de p) y con las tablas de los cuantiles de $\mathcal{N}(0,1)$ (para algunos valores de p).

6.2.5. Tipificación

Se conoce por **tipificación** a la transformación realizada con una variable aleatoria cuando se le resta su media y se divide por su desviación típica.

Si la variable aleatoria X es Normal de parámetros μ y σ, $X \equiv \mathcal{N}(\mu,\sigma)$, entonces la variable aleatoria que resulta cuando tipificamos:

$$Z = \frac{X-\mu}{\sigma}$$

es una Normal Estándar; es decir, $Z \equiv \mathcal{N}(0,1)$.

6.3. Otras distribuciones continuas importantes

6.3.1. Distribución chi-cuadrado de Pearson

Si $Z_1, Z_2, \ldots, Z_n$ son variables aleatorias independientes, todas ellas con distribución Normal Estándar, entonces la variable aleatoria $Z_1^2 + Z_2^2 + \cdots + Z_n^2$ sigue una distribución denominada **chi-cuadrado de Pearson con n grados de libertad**, que se denota por χ_n^2.

El cuantil al $100p\,\%$ de χ_n^2 se representa por $\chi_{n\,,p}^2$ y es el valor que verifica:

$$P(\chi_n^2 \le \chi_{n\,,p}^2) = p\,,$$

es decir, el área comprendida entre la curva de densidad de la distribución χ_n^2 y el eje horizontal, a la izquierda de $\chi_{n\,,p}^2$, es igual a p.

Otra interpretación es la siguiente: el valor $\chi_{n\,,p}^2$ deja por debajo el $100p\,\%$ de todos los resultados de una variable aleatoria chi-cuadrado de Pearson con n grados de libertad.

El resultado de $\chi_{n\,,p}^2$ se puede determinar con cualquier paquete estadístico, como MINITAB o SPSS (para cualquier valor de n y p) y con las tablas de los cuantiles de χ_n^2 (para algunos valores de n y p).

6.3.2. Distribución t de Student

Si Z sigue una distribución Normal Estándar y χ_n^2 es independiente de Z, entonces la variable aleatoria

$$\frac{Z}{\sqrt{\dfrac{\chi_n^2}{n}}}$$

sigue una distribución denominada t **de Student con n grados de libertad**, que se denota por t_n.

El cuantil al $100p\,\%$ de t_n se representa por $t_{n\,,p}$ y es el valor que verifica:

$$P(t_n \le t_{n\,,p}) = p\,,$$

es decir, el área comprendida entre la curva de densidad de la distribución t_n y el eje horizontal, a la izquierda de $t_{n\,,\,p}$, es igual a p.

Otra interpretación es la siguiente: el valor $t_{n\,,\,p}$ deja por debajo el $100p\,\%$ de todos los resultados de una variable aleatoria t de Student con n grados de libertad.

El resultado de $t_{n\,,\,p}$ se puede determinar con cualquier paquete estadístico, como MINITAB o SPSS (para cualquier valor de n y p) y con las tablas de los cuantiles de t_n (para algunos valores de n y p).

6.3.3. Distribución F de Snedecor

Si tenemos dos variables aleatorias chi-cuadrado independientes, χ^2_m y χ^2_n, entonces la variable aleatoria

$$\frac{\frac{\chi^2_m}{m}}{\frac{\chi^2_n}{n}}$$

sigue una distribución denominada F **de Snedecor con** m **grados de libertad en el numerador y** n **grados de libertad en el denominador**, que se denota por $F_{m\,,\,n}$.

El cuantil al $100p\,\%$ de $F_{m\,,\,n}$ se representa por $F_{m\,,\,n\,,\,p}$ y es el valor que verifica:

$$P(F_{m\,,\,n} \leq F_{m\,,\,n\,,\,p}) = p\,,$$

es decir, el área comprendida entre la curva de densidad de la distribución $F_{m\,,\,n}$ y el eje horizontal, a la izquierda de $F_{m\,,\,n\,,\,p}$, es igual a p.

Otra interpretación es la siguiente: el valor $F_{m\,,\,n\,,\,p}$ deja por debajo el $100p\,\%$ de todos los resultados de una variable aleatoria F de Snedecor con m grados de libertad en el numerador y n grados de libertad en el denominador.

El resultado de $F_{m\,,\,n\,,\,p}$ se puede determinar con cualquier paquete estadístico, como MINITAB o SPSS (para cualquier valor de m, n y p) y con las tablas de los cuantiles de $F_{m\,,\,n}$ (para algunos valores de m, n y p).

Ejemplos que se van a resolver en clase

Ejemplo 6.1. Si $Z \equiv \mathcal{N}(0,1)$ calcular las siguientes probabilidades:

a) $P(Z < 0'321)$.
b) $P(Z \geq 1'275)$.
c) $P(Z < -2'152)$.
d) $P(Z \geq -0'456)$.
e) $P(-1'434 \leq Z \leq 1'568)$.

Ejemplo 6.2. Si $Z \equiv \mathcal{N}(0,1)$ determinar los siguientes cuantiles e interpretar los resultados.

a) Mediana de Z.
b) Tercer cuartil de Z.
c) Primer cuartil de Z.

Ejemplo 6.3. En una determinada asignatura de un Grado en Información y Documentación se sabe que las calificaciones siguen una distribución Normal de media $5'5$ y desviación típica $1'5$. Si en un año académico hay 150 alumnos matriculados en esta asignatura, calcular el número de alumnos que obtendrán una calificación:

a) menor o igual que 3.
b) mayor o igual que 8.
c) comprendida entre 4 y 6.

Ejemplo 6.4. Determinar los siguientes cuantiles e interpretar los resultados.

a) Mediana de χ^2_{10}.
b) Tercer cuartil de χ^2_{30}.

Ejemplo 6.5. Determinar los siguientes cuantiles e interpretar los resultados.

a) Tercer cuartil de t_{25}.
b) Primer cuartil de t_{60}.

Ejemplo 6.6. Determinar los siguientes cuantiles e interpretar los resultados.

a) Cuantil al 95 % de $F_{20\,,\,10}$.
b) Cuantil al 10 % de $F_{20\,,\,10}$.

Problemas propuestos

Problema 6.1. Si Z es una variable Normal Estándar, determinar:

a) $P(Z \leq 2'21)$.
b) $P(Z < 3'47)$.
c) $P(Z \leq -1'75)$.
d) $P(Z > 2'46)$.
e) $P(Z \geq 3'24)$.
f) $P(Z > -3'08)$.
g) $P(1'12 \leq Z \leq 2'68)$.
h) $P(-0'85 < Z < 1'27)$.
i) $P(-2'97 < Z \leq -1'33)$.

Problema 6.2. Si X es una variable Normal con media $8'46$ y desviación típica $1'14$, hallar:

a) $P(X \leq 9'11)$.
b) $P(X < 12'33)$.
c) $P(X \leq 6'41)$.
d) $P(X > 10'52)$.
e) $P(X \geq 12'61)$.
f) $P(X > 4'01)$.
g) $P(6'11 \leq X \leq 11'91)$.
h) $P(7'53 < X < 10'33)$.
i) $P(5'05 \leq X < 6'83)$.

Problema 6.3. Hallar el valor de los siguientes cuantiles:

a) $Z_{0'58}$.
b) $Z_{0'42}$.
c) $Z_{0'999}$.
d) $Z_{0'001}$.

Problema 6.4. El cociente intelectual de 5.600 alumnos del Grado en Información y Documentación de diversas universidades sigue una distribución Normal de media 130 y desviación típica 6. Calcular cuántos de ellos tienen un cociente intelectual:

a) mayor que 140.
b) entre 125 y 135.
c) menor que 120.

Problema 6.5. Calcular el valor de los siguientes cuantiles:

a) $\chi^2_{6\,,\,0'01}$.
b) $\chi^2_{6\,,\,0'99}$.
c) $\chi^2_{72\,,\,0'975}$.

Problema 6.6. Sea X una variable aleatoria que sigue una distribución chi-cuadrado de Pearson con 15 grados de libertad. Determinar el valor de a que verifica la siguiente igualdad:

a) $P(X \leq a) = 0'05$.

b) $P(X > a) = 0'99$.

Problema 6.7. Calcular el valor de los siguientes cuantiles:

a) $t_{26\,,\,0'9}$.

b) $t_{26\,,\,0'1}$.

c) $t_{75\,,\,0'8}$.

Problema 6.8. Sea X una variable aleatoria que sigue una distribución t de Student con 20 grados de libertad. Determinar el valor de a que verifica la siguiente igualdad:

a) $P(X \leq a) = 0'99$.

b) $P(X \geq a) = 0'25$.

Problema 6.9. Calcular el valor de los siguientes cuantiles:

a) $F_{8\,,\,6\,,\,0'975}$.

b) $F_{25\,,\,50\,,\,0'01}$.

c) $F_{45\,,\,35\,,\,0'01}$.

Problema 6.10. Sea X una variable aleatoria que sigue una distribución F de Snedecor con 10 grados de libertad en el numerador y 8 grados de libertad en el denominador. Determinar el valor de a que verifica la siguiente igualdad:

a) $P(X < a) = 0'9$.

b) $P(X > a) = 0'05$.

Soluciones de los problemas propuestos

Solución del problema 6.1. **a)** $0'986447$, **b)** $0'9997398$, **c)** $0'040059$, **d)** $0'006947$, **e)** $0'0005976$, **f)** $0'998965$, **g)** $0'127676$, **h)** $0'700295$, **i)** $0'09027$.

Solución del problema 6.2. **a)** $0'715661$, **b)** $0'9996505$, **c)** $0'03593$, **d)** $0'035148$, **e)** $0'0001363$, **f)** $0'9999519$, **g)** $0'979078$, **h)** $0'743389$, **i)** $0'074964$.

Solución del problema 6.3. **a)** $0'20189$, **b)** $-0'20189$, **c)** $3'09023231$, **d)** $-3'09023231$.

Solución del problema 6.4. **a)** $0'04746 \cdot 5600 = 265'776 \simeq 266$ alumnos, **b)** $0'593462 \cdot 5600 = 3323'3872 \simeq 3323$ alumnos, **c)** $0'04746 \cdot 5600 = 265'776 \simeq 266$ alumnos.

Solución del problema 6.5. **a)** $0'87209$, **b)** $16'8119$, **c)** $97'356547$.

Solución del problema 6.6. **a)** $7'26094$, **b)** $5'22935$.

Solución del problema 6.7. **a)** $1'315$, **b)** $-1'315$, **c)** $0'844772$.

Solución del problema 6.8. **a)** $2'528$, **b)** $0'687$.

Solución del problema 6.9. **a)** $5'5996$, **b)** $0'416684$, **c)** $0'477478$.

Solución del problema 6.10. **a)** $2'538$, **b)** $3'3472$.