Estadística: estadística
Grado en Relaciones Laborales | Curso 2019-2020
Tema 6. Variables aleatorias

Alejandro Saavedra Nieves

Introducción

- En el Tema 1 hemos estudiado variables, entendiéndolas como mediciones que se efectúan sobre los individuos de una muestra. Así, la Estadística Descriptiva nos permitía analizar los distintos valores que tomaban las variables sobre una muestra ya observada. Se trataba, pues, de un estudio posterior a la realización del experimento aleatorio.
- En este tema trataremos las variables situándonos antes de la realización del experimento aleatorio. Por tanto, haremos uso de los conceptos del tema anterior (Probabilidad), mientras que algunos desarrollos serán análogos a los del tema de Estadística Descriptiva.

Variable aleatoria

Al realizar un experimento aleatorio generalmente estamos interesados en alguna función del resultado más que en el resultado en sí mismo. Por ejemplo, al arrojar un dado dos veces podríamos estar interesados sólo en la suma de los puntos obtenidos y no en el par de valores que dio origen a ese valor de la suma. De manera informal, esa cantidad de interés se denomina variable aleatoria.

- Variable porque toma distintos valores
- aleatoria porque el valor observado no puede ser predicho antes de la realización del experimento, aunque sí se sabe cuáles son sus posibles valores.

Dado que el valor de una variable aleatoria (v.a.) es determinado por el resultado de un experimento, podremos asignar probabilidades a los posibles valores o conjuntos de valores de la variable.

Variable aleatoria

Llamamos **variable aleatoria** a una aplicación del espacio muestral asociado a un experimento aleatorio en $\mathbb{R}$, que a cada resultado de dicho experimento le asigna un número real, obtenido por la medición de cierta característica.

$$\begin{array}{rccc} X: & \Omega & \longrightarrow & \mathbb{R} \\ & \omega & \longrightarrow & X(\omega) \end{array}$$

Denotamos la variable aleatoria por una letra mayúscula. El conjunto imagen de esa aplicación es el conjunto de valores que puede tomar la variable aleatoria, que serán denotados por letras minúsculas.

Variables aleatorias

De modo idéntico a lo dicho en el tema de Descriptiva, podemos clasificar las variables aleatorias en **discretas** y **continuas** en función del conjunto de valores que pueden tomar.

- Variables discretas: los valores se encuentran separados entre sí en conjuntos discretos, como $\mathbb{Z}$ o $\mathbb{N}$. Para dichas variables veremos:
    - Función de probabilidad o función de masa
    - Función de distribución
- Variables continuas: el conjunto de valores que puede tomar es un intervalo. Para dichas variables veremos:
    - Función de densidad
    - Función de distribución

Si X es una variable discreta, su distribución viene dada por los valores que puede tomar y las probabilidades de que aparezcan. Si $x_1 < x_2 < .. < x_n$ son los posibles valores de la variable X, las diferentes probabilidades de que ocurran estos sucesos,

$$\begin{aligned} p_1 &= P(X = x_1), \\ p_2 &= P(X = x_2), \\ &\vdots \\ p_n &= P(X = x_n). \end{aligned}$$

constituyen la distribución de X. Esta función se denomina **función de probabilidad o función de masa**. La función de probabilidad se puede representar análogamente al diagrama de barras.

Ejemplo: Los servicios médicos de una empresa establecen un período de entre 7 y 9 días de baja para un trabajador que ha sufrido una fuerte contusión. Además se estima que

- La probabilidad de que el período de baja sea de 7 días es 0.4.
- La probabilidad de que el período de baja sea de 8 días es 0.5.
- La probabilidad de que de que el período de baja sea de 9 día es 0.1.

Comprueba que se trata efectivamente de una distribución de probabilidad y a represéntala.

Variables aleatorias discretas. Función de distribución

La **función de distribución** de una variable aleatoria se define como:

$$\begin{array}{rccc} F: & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & x_0 & \longrightarrow & F(x_0) = P(X \leq x_0) \end{array}$$

Ejemplo: Los servicios médicos de una empresa establecen un período de entre 7 y 9 días de baja para un trabajador que ha sufrido una fuerte contusión. Además se estima que

- La probabilidad de que el período de baja sea de 7 días es 0.4.
- La probabilidad de que el período de baja sea de 8 días es 0.5.
- La probabilidad de que de que el período de baja sea de 9 día es 0.1.

Calcula y representa la función de distribución. Interpreta los resultados.

Variables aleatorias discretas. Función de distribución

Suponiendo que la variable X toma los valores $x_1 \leq x_2 \leq \ldots \leq x_n$, la función de distribución viene definida por:

$$\begin{aligned}
F(x_1) &= P(X \leq x_1) = P(X = x_1) \\
F(x_2) &= P(X \leq x_2) = P(X = x_1) + P(X = x_2) \\
&\vdots \\
F(x_n) &= P(X \leq x_n) = P(X = x_1) + \ldots + P(X = x_n) = 1
\end{aligned}$$

La función de distribución es siempre no decreciente y verifica que,

$$\begin{aligned}
F(-\infty) &= 0, \\
F(+\infty) &= 1.
\end{aligned}$$

- Los conceptos que permiten resumir una distribución de frecuencias utilizando valores numéricos pueden utilizarse también para describir la distribución de probabilidad de una variable aleatoria.

Media y varianza de variables aleatorias

Para distinguir entre las propiedades de los conjuntos de datos y las de las distribuciones de probabilidad, usaremos cierta terminología y ciertos símbolos que describimos a continuación.

- Las propiedades de los datos se llaman propiedades muestrales. Por ejemplo, hablamos en el Tema 1 de la media muestral $\bar{x}$ o de la desviación típica muestral S_X.
- Las propiedades de las distribuciones de probabilidad se llaman propiedades poblacionales.
  - μ denota la media poblacional.
  - σ denota la desviación típica poblacional.

Media y Varianza poblacional de una variable aleatoria discreta

- Consideremos el ejemplo del trabajador que ha sufrido una fuerte contusión. Estamos interesados en el número de días de baja del jugador.

x_i	p_i
7	0.4
8	0.5
9	0.1

- ¿Cómo definirías el número medio (o número esperado) de días que el jugador pasará de baja?

$$\mathbb{E}(X) = \mu = \sum_i x_i p_i = 7 \cdot 0.4 + 8 \cdot 0.5 + 9 \cdot 0.1 = 7.7$$

- ¿Cómo definirías la varianza de la variable X?

$$\text{Var}(X) = \sigma^2 = \sum_i (x_i - \mu)^2 p_i = (7-7.7)^2 \cdot 0.4 + (8-7.7)^2 \cdot 0.5 + (9-7.7)^2 \cdot 0.1 = 0.41$$

(Con respecto al Tema 1, p_i juega el papel de la frecuencia relativa)

Propiedades de la media y varianza de una variable aleatoria discreta

Propiedades

Sea X una variable aleatoria discreta con valores x_i. Entonces:

- $\mathbb{E}(a+bX) = a + b\mathbb{E}(X)$
- $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$
- $\text{Var}(a+bX) = b^2 Var(X)$

Ejemplo: Consideremos el ejemplo del trabajador que ha sufrido una fuerte contusión. Por cada lesión que sufre el trabajador el seguro le debe pagar 5000 euros, además de 1000 euros por cada día de baja. ¿Cuánto dinero espera recibir el trabajador del seguro?

Principales modelos de distribuciones discretas

- Estudiaremos distribuciones de variables aleatorias discretas que han adquirido una especial relevancia por ser adecuadas para modelizar una gran cantidad de situaciones.
- Caracterizaremos estas distribuciones mediante la función de masa y función de distribución.
- Calcularemos también los momentos (media y varianza) y destacaremos las propiedades de mayor utilidad.

Variable Bernoulli

En muchas ocasiones nos encontramos ante experimentos aleatorios con sólo dos posibles resultados: éxito y fracaso

Ejemplos: cara o cruz en el lanzamiento de una moneda, ganar o perder un partido, aprobar o suspender un examen, recuperarse o no recuperarse de una enfermedad...

Se pueden modelizar estas situaciones mediante la variable aleatoria

$$X = \begin{cases} 1 & \text{si éxito} \\ 0 & \text{si Fracaso} \end{cases}$$

La probabilidad de éxito es p y la probabilidad de fracaso es $q = 1 - p$. Un experimento de este tipo se llama **experimento de Bernoulli** $Be(p)$.

- Calcula la función de masa y la función de distribución de una $Be(p)$.
- Si $X \in Be(p)$, entonces:
    - $\mu = p$
    - $\sigma^2 = p(1 - p)$

Ejemplo: La probabilidad de que una bombilla no funcione es 0.6. Si tenemos un lote de tres bombillas, ¿cuál es la probabilidad de que exactamente una de ellas no funcione?

Cada bombilla es independiente de las demás y podemos considerarlo como un ensayo de Bernoulli, donde el éxito de funcionar ($p = 0.4$). Lo que hacemos es repetir el experimento 3 veces y queremos calcular la probabilidad de que el número de éxitos sea igual a 2 (es decir, 2 funcionan y 1 falla)

Variable Binomial

Empezando con una prueba de Bernoulli con probabilidad de éxito p, vamos a construir una nueva variable aleatoria al repetir n veces la prueba de Bernoulli. La variable aleatoria **binomial** X es el número de éxitos en n repeticiones de una prueba de Bernoulli con probabilidad de éxito p.
Debe cumplirse:

- Cada prueba individual puede ser un éxito o un fracaso
- La probabilidad de éxito, p, es la misma en cada prueba
- Las pruebas son independientes. El resultado de una prueba no tiene influencia sobre los resultados siguientes

Principales modelos de distribuciones discretas: Binomial

Variable Binomial

La variable aleatoria **binomial** X es el número de éxitos en n repeticiones de una prueba de Bernoulli con probabilidad de éxito p, es decir:

$$X = \text{Número de éxitos en las } n \text{ pruebas}$$

Denotaremos esta variable como $Bin(n, p)$.

- ¿Qué valores toma una $Bin(n, p)$?
- ¿Cuál es su función de masa?

Variable Binomial

La variable aleatoria **binomial** X es el número de éxitos en n repeticiones de una prueba de Bernoulli con probabilidad de éxito p, es decir:

$$X = \text{Número de éxitos en las } n \text{ pruebas}$$

La probabilidad de obtener k éxitos en n pruebas es

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

- La media y la varianza de una $Bin(n, p)$ son:
  - $\mu = n \cdot p$
  - $\sigma^2 = n \cdot p \cdot (1-p)$

El coeficiente binomial

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

es el número de subconjuntos diferentes de k elementos que se pueden definir a partir de un total de n elementos (**combinaciones** de n elementos tomados de k en k).

Coeficientes binomiales

El coeficiente binomial

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

representa el número de subconjuntos diferentes de k elementos que se pueden definir a partir de un total de n elementos (**combinaciones** de n elementos tomados de k en k).

Se denota por $n!$ el factorial de n: $n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$.

Por ejemplo, si para un partido de dobles de la Copa Davis tenemos a tres jugadores ({Robredo, Feliciano López, Verdasco}), el entrenador tendrá

$$\binom{3}{2} = \frac{3!}{2!1!} = \frac{3 \cdot 2 \cdot 1}{(2 \cdot 1)(1)} = 3$$

posibles formas de elegir a los jugadores del partido ({Robredo, Feliciano López}, {Robredo, Verdasco}, {Feliciano López, Verdasco}).

- En muchas circunstancias (llamadas a una centralita telefónica, número de accidentes laborales, ...) el número de individuos susceptibles de dar lugar a un éxito es muy grande.
- Para modelizar estas situaciones mediante una distribución binomial tendremos problemas al escoger el parámetro n (demasiado grande o incluso difícil de determinar) y al calcular la distribución de probabilidad (la fórmula resulta inviable).

Variable Poisson

Una variable aleatoria X tiene distribución de **Poisson** de parámetro λ, y lo denotamos $X \in Poisson(\lambda)$, si es discreta y

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad \text{si} \quad k \in \{0, 1, 2, 3, \ldots\}$$

La variable de Poisson mide el número de éxitos en un intervalo de tiempo fijado.

La media y la varianza de la Poisson de parámetro λ son:

- $\mu = \lambda$
- $\sigma^2 = \lambda$

- Utilizaremos la distribución de Poisson como aproximación de la distribución binomial cuando n sea grande y p pequeño, en base al límite que hemos visto.
- En concreto, utilizaremos una distribución de Poisson con $\lambda = n \cdot p$. Esto es, X es aproximada por una distribución $Poission(n \cdot p)$.
- Como criterio podremos aproximar cuando $n > 50$ y $p < 0.1$.

Ejemplo

La probabilidad de que una persona se desmaye en un concierto es $p = 0.005$. ¿Cuál es la probabilidad de que en un concierto al que asisten 3000 personas se desmayen 18?
La variable $X =$Número de personas que se desmayan en el concierto sigue una distribución $Bin(3000, 0.005)$. Queremos calcular

$$P(X = 18) = \binom{3000}{18} \cdot 0.005^{18} \cdot 0.995^{2982} = 0.07071.$$

Estos valores están fuera de las tablas de la binomial y son difíciles de calcular, por eso es preferible aproximar por una Poisson de parámetro $\lambda = np = 3000 \cdot 0.005 = 15$. Entonces:

$$P(X = 18) \approx P(Poisson(15) = 18) = e^{-15} \frac{15^{18}}{18!} = 0.07061.$$

Principales modelos de distribuciones discretas: Poisson

Aunque la distribución de Poisson se ha obtenido como forma límite de una distribución Binomial, tiene muchas aplicaciones sin conexión directa con las distribuciones binomiales. Por ejemplo, la distribución de Poisson puede servir como modelo del número de éxitos que ocurren durante un intervalo de tiempo o en una región específica.

Definimos el **proceso de Poisson** como un experimento aleatorio que consiste en contar el número de ocurrencias de determinado suceso en un intervalo de tiempo, verificando:

- El número medio de sucesos por unidad de tiempo es constante. A esa constante la llamamos **intensidad del proceso**.
- Los números de ocurrencias en subintervalos disjuntos son independientes.

Ejemplo

El número de nacimientos en un hospital constituye un proceso de Poisson con intensidad de 10 nacimientos por semana. ¿Cuál es la probabilidad de que se produzcan al menos tres nacimientos en una semana?

$$P(X \geq 3) = 1 - P(X < 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

$$= 1 - \left[e^{-10}\,\frac{10^0}{0!} + e^{-10}\,\frac{10^1}{1!} + e^{-10}\,\frac{10^2}{2!} \right]$$

¿Cuál es la probabilidad de que se produzcan 5 nacimientos un día?

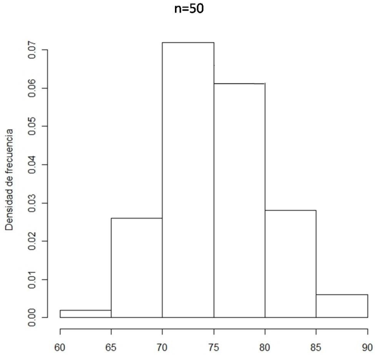
Variables aleatorias continuas

- Una variable aleatoria es **continua** cuando puede tomar cualquier valor en un intervalo.
  - el peso de una persona
  - el salario de la plantilla de una empresa
  - el tiempo de recuperación de una operación,...
- El estudio de las variables continuas es más sutil que el de las discretas. Recordemos que la construcción del histograma es más delicado que el del diagrama de barras ya que depende de la elección de las clases.

Ejemplo

En un estudio sobre atención a la tercera edad se desea evaluar la edad a la que las personas mayores deciden ingresar en un centro geriátrico.

- Se registra la edad a la que ingresaron los 50 residentes de un determinado centro gerontológico y se construye el histograma correspondiente.



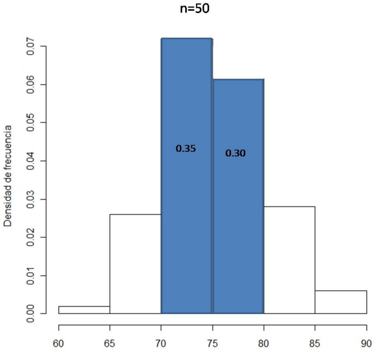
Variables aleatorias continuas

Ejemplo

En un estudio sobre atención a la tercera edad se desea evaluar la edad a la que las personas mayores deciden ingresar en un centro geriátrico.

- Se registra la edad a la que ingresaron los 50 residentes de un determinado centro gerontológico y se construye el histograma correspondiente.

Sea A el suceso "El residente ingresa con edad entre 70 y 80 años".

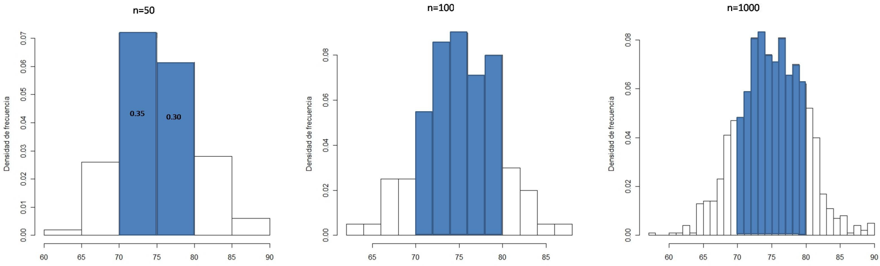


Ejemplo

En un estudio sobre atención a la tercera edad se desea evaluar la edad a la que las personas mayores deciden ingresar en un centro geriátrico.

- Se registra la edad a la que ingresaron los 50 residentes de un determinado centro gerontológico y se construye el histograma correspondiente.
- Se registra la edad a la que ingresaron los 100 residentes de un determinado centro gerontológico y se construye el histograma correspondiente.
- Se registra la edad a la que ingresaron los 1000 residentes de un determinado centro gerontológico y se construye el histograma correspondiente.

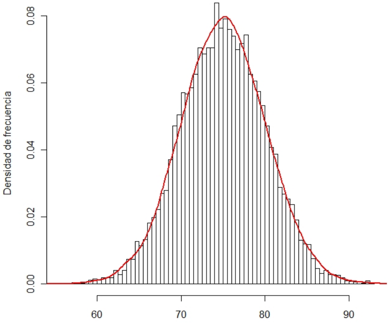
Sea A el suceso "El residente ingresa con edad entre 70 y 80 años".



Ejemplo

En un estudio sobre atención a la tercera edad se desea evaluar la edad a la que las personas mayores deciden ingresar en un centro geriátrico.

- Idealmente, se registra la edad de todos los residentes de centros gerontológicos y se construye el histograma correspondiente.



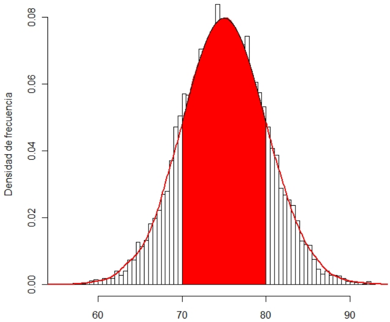
Variables aleatorias continuas

Ejemplo

En un estudio sobre atención a la tercera edad se desea evaluar la edad a la que las personas mayores deciden ingresar en un centro geriátrico.

- Idealmente, se registra la edad de todos los residentes de centros gerontológicos y se construye el histograma correspondiente.

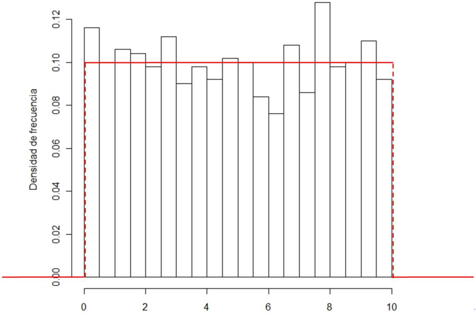
Sea A el suceso "El residente ingresa con edad entre 70 y 80 años".



- Tomando más observaciones de una variable continua y haciendo más finas las clases, el histograma tiende a estabilizarse en una curva suave que describe la distribución de la variable.
- Esta función, $f(x)$, se llama **función de densidad** de la variable X.
- La función de densidad constituye una idealización de los histogramas de frecuencia o un **modelo** del cual suponemos que proceden las observaciones.
- La función de densidad cumple dos propiedades básicas: es no negativa y el área total que contiene es uno.

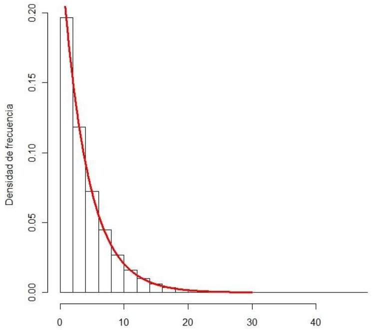
Ejemplo

Un estudiante va todos los días a la facultad en la línea 1 del autobús urbano. Llega a la parada a las 3 de la tarde y cuenta el tiempo (en minutos) que tiene que esperar hasta que llega el autobús. A continuación se muestra el histograma correspondiente al tiempo de espera de los últimos 1000 días. A la vista del histograma, ¿cómo modelizarías el tiempo de espera?



Ejemplo

Un estudiante va todos los días a la facultad en la línea 6 del autobús urbano. Llega a la parada a las 3 de la tarde y cuenta el tiempo (en minutos) que tiene que esperar hasta que llega el autobús. A continuación se muestra el histograma correspondiente al tiempo de espera de los últimos 1000 días. A la vista del histograma, ¿cómo modelizarías el tiempo de espera?



Función de densidad

Una función $f(x)$, definida sobre el conjunto de todos los números reales $\mathbb{R}$, se denomina **función de densidad** si

1. $f(x) \geq 0$.
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

Función de distribución

La **función de distribución** de una variable aleatoria se define como:

$$\begin{array}{cccc} F: & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & x_0 & \longrightarrow & F(x_0) = P(X \leq x_0) \end{array}$$

Variables aleatorias continuas: Función de densidad

La función de densidad expresa probabilidades por áreas.

- La probabilidad de que una variable X sea menor que un determinado valor x_0 se obtiene calculando el área de la función de densidad hasta el punto x_0, es decir,

$$F(x_0) = P(X \leq x_0) = \int_{-\infty}^{x_0} f(x)\, dx,$$

- La probabilidad de que la variable tome un valor entre x_0 y x_1 es,

$$P(x_0 \leq X \leq x_1) = \int_{x_0}^{x_1} f(x)\, dx.$$

Medidas características de una variable aleatoria

- Media poblacional de una variable aleatoria continua: Si X es una variable aleatoria continua y $f(x)$ es su función de densidad, el **valor esperado** de esta variable aleatoria es:

$$\mathbb{E}(X) = \mu = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

- Varianza poblacional de una variable aleatoria continua: Si X es una variable aleatoria continua, $f(x)$ es su función de densidad y μ su media, la **varianza** de esta variable aleatoria es:

$$\text{Var}(X) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx$$

Propiedades

Sea X una variable aleatoria continua con función de densidad $f(x)$. Entonces:

- $\mathbb{E}(a + bX) = a + b\mathbb{E}(X)$
- $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$
- $\text{Var}(a + bX) = b^2 Var(X)$

Principales modelos de distribuciones continuas: Variable Uniforme

Distribución uniforme

Una variable aleatoria continua X se dice que tiene distribución uniforme, y lo denotamos $X \in U(a, b)$, si su función de densidad viene dada por:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in (a, b) \\ 0 & \text{en otro caso} \end{cases}$$

En este caso, la función de distribución es:

$$P(X \leq x) = F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } x \in (a, b) \\ 1 & \text{si } x > b. \end{cases}$$

Características

- $E(X) = \frac{a+b}{2}$
- $Var(X) = \frac{(b-a)^2}{12}$

La probabilidad se reparte de modo uniforme sobre el intervalo (a, b) (densidad constante en dicho intervalo), siendo el modelo de distribución uniforme el utilizado para describir el resultado del experimento consistente en elegir un valor al azar en un intervalo.

Ejemplo: Un estudiante acude todos los días a clase andando. Suponiendo que tarda en llegar 15 minutos, que sale de casa, en un instante aleatorio, entre las ocho cuarenta y las ocho cincuenta y que la clase comienza a las nueve, calcular la probabilidad de que un día determinado llegue tarde.

Sea $X =$"hora a la que sale de casa", X sigue una distribución $U(40, 50)$. Llega tarde cuando la hora a la que sale de casa está en el intervalo $(45, 50]$:

$$P(45 < X \leq 50) = F(50) - F(45) = \frac{50-40}{50-40} - \frac{45-40}{50-40} = \frac{50-45}{50-40} = \frac{1}{2}$$

Ejemplo

Sea X una variable aleatoria y $f(x)$ su función de densidad, dada por:

$$f(x) = \begin{cases} 0.1 & \text{si } 0 \leq x \leq 10 \\ 0 & \text{en otro caso} \end{cases}$$

- Comprueba que f es efectivamente una función de densidad y represéntala.
- Calcula la función de distribución de X.
- Calcula la media y la varianza de X.

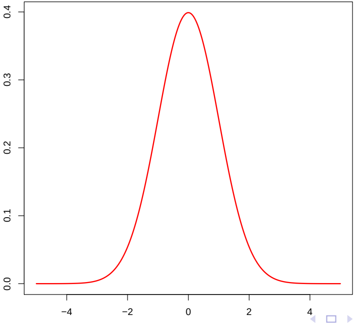
Principales modelos de distribuciones continuas: Variable Normal

- La distribución **normal** es la más importante y de mayor uso de todas las distribuciones continuas de probabilidad.
- Por múltiples razones se viene considerando la más idónea para modelizar una gran diversidad de mediciones de la Física, Química o Biología.
- La normal es una familia de variables que depende de dos parámetros, la media y la varianza.
- Dado que todas están relacionadas entre si mediante una transformación muy sencilla, empezaremos estudiando la denominada **normal estándar** para luego definir la familia completa.

Variable Normal Estándar

Una variable aleatoria continua Z se dice que tiene distribución **normal estándar**, y lo denotamos $Z \in N(0,1)$, si su función de densidad viene dada por:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad \text{si} \quad z \in \mathbb{R}$$



Variable Normal Estándar

Una variable aleatoria continua Z se dice que tiene distribución **normal estándar**, y lo denotamos $Z \in N(0,1)$, si su función de densidad viene dada por:

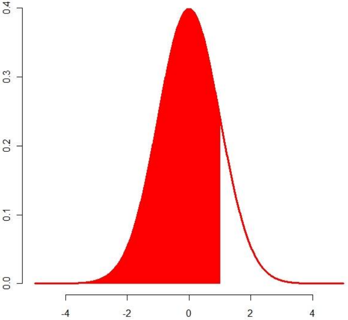
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad \text{si} \quad z \in \mathbb{R}$$

- $Z \in N(0,1)$ toma valores en toda la recta real. $(f(z) > 0 \quad \forall z \in \mathbb{R})$
- f es simétrica en torno a cero.
- Si $Z \in N(0,1)$ entonces $\mu = 0$ y $\sigma^2 = 1$.

Principales modelos de distribuciones continuas: Variable Normal

Supongamos entonces que $Z \in N(0,1)$. ¿Cómo calcularías $P(Z \leq 1)$?

$$P(Z \leq 1) = \int_{-\infty}^{1} f(z)dz = \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$



Supongamos entonces que $Z \in N(0,1)$. ¿Cómo calcularías $P(Z \leq 1)$?

$$P(Z \leq 1) = \int_{-\infty}^{1} f(z)dz = \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

- La probabilidad inducida vendrá dada por el área bajo la densidad.
- Como no existe una expresión explícita para el área existen tablas con algunas probabilidades ya calculadas.
- Las tablas que nosotros utilizaremos proporcionan el valor de la función de distribución, $\Phi(z_0) = P(Z \leq z_0)$, de la normal estándar para valores positivos de z, donde z está aproximado hasta el segundo decimal.

Supongamos que $Z \in N(0,1)$. Calcula usando las tablas de la normal estándar:

- $P(Z \leq 1.64)$
- $P(Z > 1)$
- $P(Z \leq -0.53)$
- $P(Z > -1.23)$
- $P(-1.96 \leq Z \leq 1.96)$
- $P(-1 \leq Z \leq 2)$
- ¿Cuánto vale aproximadamente $P(Z > 4.2)$?

Variable Normal

Efectuando un cambio de localización y escala sobre la normal estándar, podemos obtener una distribución con la misma forma pero con la media y desviación típica que queramos.

Si $Z \in N(0,1)$ entonces

$$X = \mu + \sigma Z$$

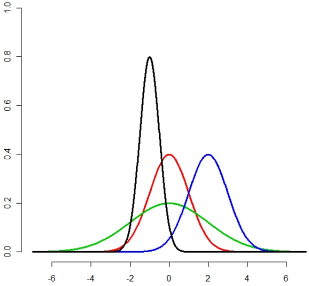
tiene **distribución normal de media** μ **y desviación típica** σ. Denotaremos $X \in N(\mu, \sigma)$.

- Si $X \in N(\mu, \sigma)$ entonces la media de X es μ y su varianza es σ^2.

Variable Normal

Sea $X \in N(\mu, \sigma)$. La función de densidad de una $N(\mu, \sigma)$ es

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$



Funciones de densidad de variables normales con distintas medias y varianzas. En rojo densidad de una $N(0, 1)$

Supongamos entonces que $X \in N(\mu, \sigma)$. ¿Cómo calcularías $P(X \leq 1)$?

$$P(X \leq 1) = \int_{-\infty}^{1} f(x)dx = \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

En la práctica sólo disponemos de la tabla de la distribución normal estándar.

En general, denotaremos por $X \in N(\mu, \sigma)$ la variable normal de media μ y varianza σ^2. Para efectuar cálculos sobre cualquier distribución normal le restamos la media y dividimos por la desviación típica. A este proceso le llamamos **estandarización** de una variable aleatoria.

$$\text{Si } X \in N(\mu, \sigma) \text{ entonces } Z = \frac{X - \mu}{\sigma} \in N(0, 1).$$

Principales modelos de distribuciones continuas: Variable Normal

Ejemplo

Supongamos que $X \in N(5,2)$. ¿Cómo calcularías $P(X \leq 1)$?

$$P(X \leq 1) = P\left(\frac{X-5}{2} \leq \frac{1-5}{2}\right) = P\left(Z \leq -2\right)$$

donde $Z = \frac{X-5}{2} \in N(0,1)$.

Distribuciones asociadas a la normal

Además del modelo normal, existen otros modelos que desempeñan un papel importante en la inferencia estadística. Entre ellos se encuentran

- la distribución χ^2
- la distribución t de Student.

Distribuciones asociadas a la normal: la distribución χ^2

La χ^2_n con n grados de libertad es otro modelo de variable aleatoria continua

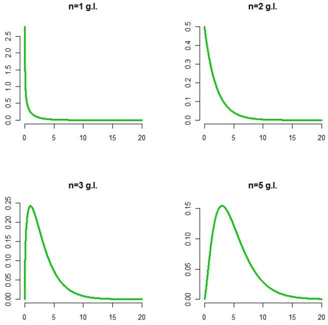


Figure: En verde densidades de variables χ^2_n para distintos valores de n.

Propiedades.

1. La variable Chi-cuadrado toma valores en $[0, +\infty)$.
2. La distribución Chi-cuadrado es asimétrica.

Distribuciones asociadas a la normal: la distribución t de Student

La t de Student con k grados de libertad es otro modelo de variable aleatoria continua

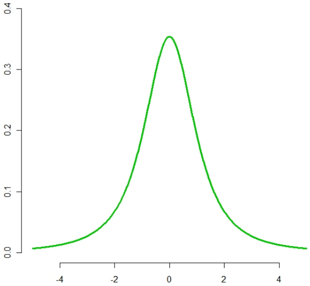


Figure: En verde densidad de una t de Student con 2 grados de libertad

Distribuciones asociadas a la normal: la distribución t de Student

La t de Student con k grados de libertad es otro modelo de variable aleatoria continua como los vistos anteriormente.

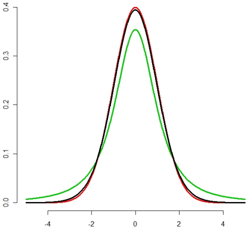


Figure: En verde densidad de una t de Student con 2 grados de libertad, en rojo N(0,1) y en negro densidad de una t de Student con 20 grados de libertad

Distribuciones asociadas a la normal: la distribución t de Student

La t de Student con k grados de libertad es otro modelo de variable aleatoria continua como los vistos anteriormente.

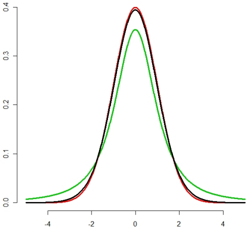


Figure: En verde densidad de una t de Student con 2 grados de libertad, en rojo N(0,1) y en negro densidad de una t de Student con 20 grados de libertad

Propiedades.

1. La variable t de Student toma valores en toda la recta real.
2. La distribución t de Student es simétrica en torno al origen.
3. $t_k \xrightarrow{d} N(0,1)$ cuando $k \to \infty$.

Distribución exponencial

Una variable aleatoria continua X se dice que tiene distribución exponencial de parámetro λ, $\lambda > 0$, y lo denotamos $X \in Exp(\lambda)$, si su función de densidad viene dada por:

$$f(x) = \begin{cases} \frac{\lambda}{e}^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

En este caso, la función de distribución es:

$$P(X \leq x) = F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{en otro caso.} \end{cases}$$

Características

- $E(X) = \frac{1}{\lambda}$
- $Var(X) = \frac{1}{\lambda^2}$

La distribución exponencial "carece de memoria", es decir

$$P(X > x + t | X > x) = P(X > t).$$

Ejemplo: Un Departamento de la Universidad decide adquirir una estación de trabajo. Se sabe que los avances tecnológicos se producen de forma aleatoria de tal manera que, por término medio, surge un nuevo modelo que deja anticuados a los ya existentes cada 7 meses.
¿Cuál es la probabilidad de que la nueva máquina no se quede anticuada durante un período de tiempo comprendido entre 6 meses y 1 año?

Supuesto que la variable $X =$"tiempo, en meses, hasta que la máquina queda anticuada" sigue una distribución $Exp(1/7)$:

$$P(6 < X < 12) = 1 - e^{-1/7*12} - (1 - e^{-1/7*6} = -e^{-1/7*12} + e^{-1/7*6} = 0.244$$

Entonces, la probabilidad de que no quede anticuada en ese periodo es:

$$1 - P(6 < X < 12) = 1 - 0.244 = 0.756$$

Otros modelos de distribuciones continuas

Entre otras...

- Gamma
- Distribución de Pareto
- Beta
- Weibull

Una aproximación a la Normal

Sean X_1, X_2, ..., X_n variables aleatorias independientes e idénticamente distribuidas con media μ y desviación típica σ.

Para n grande, definimos $S_n = X_1 + X_2 + \cdots + X_n$ tiene distribución aproximadamente $N(n\mu, \sigma\sqrt{n})$.

Equivalentemente, para n grande, la distribución de

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \longrightarrow N(0,1).$$

Esto es, se puede aproximar por la distribución $N(0,1)$.

Ejemplo ¿Qué probabilidad de ganar tiene un individuo que apuesta a que el número de caras en cien tiradas de una moneda no sesgada difiere de cincuenta en menos de cuatro?

Solución: Sea X_i la variable que toma el valor 1 si se obtiene cara en el lanzamiento i y 0 si se obtiene una cruz en dicho lanzamiento y sea $X = \sum_{i=1}^{100} X_i$ la variable "número de caras en los 100 lanzamientos de la moneda".

Las variables $X_1, X_2, \ldots, X_{100}$ definidas, son $Be(1/2)$, todas con la misma media ($\mu = 1/2$) y la misma desviación típica ($\sigma = 1/2$).

Aplicando el Teorema Central del Límite, con $n = 100$, la distribución de $Z = \frac{X-50}{5}$ se puede aproximar por la distribución $N(0,1)$.
Utilizando la aproximación anterior, la probabilidad pedida es

$$P(-4 < X - 50 < 4) = P(-4/5 < Z < 4/5) = 0.5762$$