Tema 3

Variables aleatorias. Distribuciones

Variables aleatorias. Distribuciones

- Variables aleatorias:
  - Tipos
  - Distribución
  - Características
- Modelos:
  - Discretos
  - Continuos

Variable aleatoria (v.a.)

- **La <u>variable aleatoria</u> es el modelo matemático de la <u>variable estadística</u> <u>(descriptiva)</u>.**
- Una **<u>variable aleatoria</u>** es una aplicación que asocia números reales a los resultados de un experimento:

$$X:\ \Omega \rightarrow \mathbb{R}$$

Probabilidad asociada a X

- La **probabilidad** de un **subconjunto A de nº's reales (suceso)** se calcula como

$$P(X \in A) = P(\{\omega \in \Omega / X(\omega) \in A\}).$$

- Se llama **función de distribución** de la variable X a la función F real de variable real tal que:

F(x)=P(X $\leq$ x), para cualquier x de IR.

Clasificación de v.a.

- Las variables aleatorias se clasifican en:
  - **Discretas:** si su imagen tiene cardinal finito (o, a lo sumo, numerable).
  - **Continuas:** si tienen como imagen al menos un intervalo de IR.

Ejemplo de v.a.

- Si observamos los resultados al lanzar 5 veces una moneda el espacio muestral es

Ω= {(c, c, c, c, c), (c, c, c, c, +),…, (+, +, +, +, +)}.

- **X = "Número de veces que aparece cara en las 5 tiradas de la moneda"** es una **v. a.**

$$X: \Omega \to IR$$

$$(c, c, c, c, c) \to X(c, c, c, c, c) = 5$$

$$(c, c, c, c, +) \to X(c, c, c, c, +) = 4$$

$$\vdots$$

$$(+,+,+,+,+) \to X(+,+,+,+,+) = 0$$

Ejemplo de v.a.

- La variable X toma los valores {0, 1, 2, 3, 4, 5}, por tanto, **X es una variable aleatoria discreta.**
- Los **sucesos** se expresan en términos de X:

{X < 2} = "salen menos de 2 caras" =

= {(+, +, +, +, +), (+, +, +, +, c),…, (c, +, +, +, +)}.

- $P(\{X<2\}) =$

$=P(\{(+,+,+,+,+), (+,+,+,+,c),\ldots,(c,+,+,+,+)\})=$

$=P(+,+,+,+,+) + P(+,+,+,+,c) + \ldots + P(c,+,+,+,+) =$

$$= \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5 + \ldots + \left(\frac{1}{2}\right)^5 = 6 \cdot \left(\frac{1}{2}\right)^5.$$

Variable aleatoria discreta

- Una v.a. discreta X toma un conjunto **finito o numerable** de valores x_i, con probabilidad asociada $P(X=x_i)=p_i$.
- Se llama **"función de masa de probabilidad"** al conjunto de probabilidades $\{p_i\}_i$, que cumple:
  - $p_i>0$.
  - $\sum_i p_i = 1.$
- El conjunto de pares $(x_i, p_i)_i$ es su **distribución de probabilidad**.

Probabilidades de v.a. discretas

- Si $x \neq x_i$, entonces $P(X=x)=0$.
- Si X es una variable aleatoria discreta y A es un subconjunto de números reales, su probabilidad se calcula como:

$$P(X \in A) = \sum_{x_i \in A} p_i.$$

- La **función de distribución** de una variable discreta X en cualquier real x se calcula como:

$$F(x) = \sum_{x_i \leq x} p_i.$$

Ejemplo v.a. discreta

- En una tienda de repuestos venden, como mucho, 3 ejes al día. El 73% de los días venden algún eje. En concreto, el 42% de los días venden 1 eje y el 22% venden dos ejes.
  a) Halla la función de masa de probabilidad del número de ejes que venden al día.
  b) Si un día a las 12:00 h. ya vendieron un eje, ¿cuál es la probabilidad de llegar al máximo de ventas de ejes al final del día?

Ejemplo v.a. discreta

- X = número de ejes que se venden cada día ={0, 1, 2, 3}.

a)

x_i	p_i
0	0.27 = 1-0.73
1	0.42
2	0.22
3	0.09 =1-(0.27+0.42+0.22)
	1

Ejemplo v.a. discreta

b) Se sabe que a mediodía, ya hay vendido un eje (por tanto, $X \geq 1$). Se nos pide la probabilidad de que se vendan 3 (máximo diario) al cierre de la jornada:

$$P\left(X=3 / X \geq 1\right) = \frac{P(X=3 \cap X \geq 1)}{P(X \geq 1)} =$$

$$= \frac{P(X=3)}{P(X \geq 1)} = \frac{P(X=3)}{1-P(X=0)} = \frac{0.09}{1-0.27} = 0.1233.$$

Esperanza de una v.a. discreta

- **Esperanza matemática o media** de una variable aleatoria discreta:

$$\mu_X = E(X) = \sum_i x_i \cdot p_i.$$

- La esperanza matemática es el **modelo teórico de la media** de variables estadísticas (de hecho, basta cambiar en las variables discretas la probabilidad por la frecuencia relativa para obtener la media).
- Si la variable X se transforma de la **forma lineal** $Y=a\cdot X+b$, entonces $E(Y)=a\cdot E(X)+b$.

Ejemplo E(X)

Un empleado trabaja eventualmente a través de una empresa de trabajo temporal en 2 establecimientos, en uno trabaja en lunes, martes y/o miércoles y en el otro jueves y/o viernes. El 20% de las semanas no le llaman del primer establecimiento para trabajar ningún día, el 30% de las semanas le llaman para trabajar 1 día y el resto de las semanas trabaja 2 días. El segundo establecimiento le llama para trabajar 1 día el 20% de las semanas y el resto de las semanas trabaja 2 días.

Ejemplo E(X)

- X = Número de días que trabaja, semanalmente, el empleado en el establecimiento 1 = {0, 1, 2}
- Y = Número de días que trabaja, semanalmente, el empleado en el establecimiento 2 = {1, 2}
- Las tablas de distribuciones de probabilidad:

x_i	p_i
0	0.2
1	0.3
2	0.5 = 1 - 0.2 - 0.3
	1

y_j	p_j
1	0.2
2	0.8 = 1- 0.2
	1

Ejemplo E(X)

a) ¿Cuántos días a la semana trabajará por término medio en cada establecimiento? ¿y en total?

$$E(X) = 0 \cdot 0.2 + 1 \cdot 0.3 + 2 \cdot 0.5 = 1.3 \text{ días},$$

$$E(Y) = 1 \cdot 0.2 + 2 \cdot 0.8 = 1.8 \text{ días},$$

Si TT= el tiempo total de trabajo semanal = X+Y,

$$E(TT) = E(X+Y) = E(X) + E(Y) = 1.3 + 1.8 = 3.1 \text{ días}.$$

Para un número suficientemente grande de semanas, el número medio de días aprox. que trabaja a la semana en la 1ª empresa es de 1.3, en la 2ª de 1.8 y en ambas empresas, es de 3.1 días.

Ejemplo E(X)

b) Si utiliza el coche sólo para trabajar y para ver a su familia un día la semana, ¿cuántos días a la semana utilizará el coche por término medio?

Z = el número de días que, semanalmente, utiliza el coche = X + Y + 1

$$E(Z) = E(X + Y + 1) = E(X) + E(Y) + 1 = $$
$$= 1.3 + 1.8 + 1 = 4.1 \text{ días},$$

es decir, para un número suficientemente grande de semanas, el nº medio de días a la semana en que utiliza el coche es aprox. de 4.1 días.

Ejemplo E(X)

c) Si cada día que utiliza el coche se gasta en gasolina 3 euros, ¿cuánto gastará en gasolina, por término medio, a la semana?

G = Gasto semanal en gasolina (en euros) = 3·Z

$$E(G) = E(3\cdot Z) = 3\cdot E(Z) = 3\cdot 4.1 = 12.3 \text{ €/semana},$$

es decir, para un número suficientemente grande de semanas, el gasto medio semanal en gasolina es aprox. de 12.3 €.

Varianza v.a. discreta

- **Varianza de una variable aleatoria discreta** es la esperanza matemática de las diferencias con la esperanza elevadas al cuadrado:

$$V(X) = \sigma_X^2 = E\left[(X - E(X))^2\right] = \sum_i (x_i - E(X))^2 \cdot p_i.$$

- El desarrollo de la expresión anterior permite calcular la varianza como **"esperanza del cuadrado menos el cuadrado de la esperanza"**:

$$V(X) = \sigma_X^2 = E\left(X^2\right) - [E(X)]^2 = \sum_i x_i^2 \cdot p_i - [E(X)]^2.$$

Desviación v.a. discreta

- **Desviación típica de una variable aleatoria** discreta: $DT(X) = \sigma_X = +\sqrt{V(X)}.$

  Se expresa en las mismas unidades de X.
- La desviación y la varianza de v.a. son los **modelos teóricos** de la desviación y la varianza de variables estadísticas.
- Si la variable X se transforma de la **forma lineal** Y=aX+b, entonces $V(Y)=a^2\cdot V(X)$ y $DT(Y)=|a|\cdot DT(X)$.

Ejemplo V(X)

- El 24.8% de los ordenadores de cierto organismo tiene 3 años, el 23% tiene 2 años, el 31.6% tiene un año y el resto son nuevos. Una empresa de mantenimiento les ha hecho una oferta para poner al día los ordenadores: plantean un coste fijo de 3 euros por cada ordenador más un coste variable de 1.5 por cada año de antigüedad del aparato. Calcula e interpreta:

a) La variabilidad de la antigüedad de los ordenadores.

b) La dispersión del coste de mantenimiento por ordenador según esa oferta.

Ejemplo V(X)

- X = Antigüedad (en años) de cada ordenador de ese organismo = {0, 1, 2, 3}.

x_i	p_i
0	0.206 = 1 - 0.248 - 0.230 -0.316
1	0.316
2	0.230
3	0.248
	1

Ejemplo V(X)

a) $E(X)=0\cdot0.206+1\cdot0.316+2\cdot0.230+3\cdot0.248=1.52$ años.

$E(X^2)=0^2\cdot0.206+1^2\cdot0.316+2^2\cdot0.230+3^2\cdot0.248=$
$=3.468$ años^2.

$V(X)=3.468-1.52^2=3.468-2.3104=1.1576$ años^2.

$DT(X)=1.0759$ años.

El valor obtenido representa la desviación típica aproximada de la antigüedad (en años) de un ordenador, si se ha tomado una muestra suficientemente grande de ordenadores.

Ejemplo V(X)

b) La dispersión del coste de mantenimiento por ordenador según esa oferta:

C=coste de mantenimiento, por ordenador, según la antigüedad = $1.5\cdot X+3$.

$DT(1.5X+3)=|1.5|\cdot DT(X)=|1.5|\cdot1.0759=1.613$€

es la desviación típica aproximada del coste (en €) de mantenimiento por ordenador, si se ha tomado una muestra suficientemente grande de ordenadores.

V.a. independientes

- Dadas dos variables aleatorias X e Y:
  - E(X+Y) =E(X) + E(Y), es decir, la **esperanza de una suma es <u>siempre</u> la suma de las esperanzas**.
  - Son **<u>independientes</u>** cuando todos los sucesos asociados a ambas variables son independientes: $P[(X \in A) \cap (Y \in B)] = P(X \in A)\, P(Y \in B).$
  - Si las variables X e Y son **<u>independientes</u>**:

    Var (X+Y) = Var (X) + Var (Y).

Modelos discretos

- Hay una serie de distribuciones discretas que sirven para modelizar variables habituales.

  Destacamos:
  - Modelo Bernoulli
  - Modelo Binomial
  - Modelo Poisson

Experimento Bernoulli

- **Experimento de Bernoulli**: Consideramos una experiencia aleatoria en la que el resultado se puede clasificar como **"éxito"** si ocurre el suceso A (**A**), o **"fracaso"** si no ocurre el suceso A.

$$X - \text{¿ocurre}A? = \begin{cases} 1 - \text{Sí} & P(A) = p \\ 0 - \text{No} & P(\overline{A}) = 1 - p = q \end{cases}$$

x_i	p_i
0	1-p=q
1	p
	1

Distribución Bernoulli

- La variable X asociada a este tipo de experiencia se dice que **<u>tiene distribución de Bernoulli de parámetro p</u>**, y se denota por

  X ~ B(p) ó **X ~ B(1,p)**, con <u>$p \in (0,1)$</u>

- La <u>media</u> de la variable es E(X) = p.
- La <u>varianza</u> es V(X)=p·q.

Ejemplo Bernoulli

- Si en una partida de bujías hay una alta proporción de inservibles (20%), la distribución de bujías inservibles se modela mediante la variable aleatoria

$$X - \text{¿una bujía es inservible?} = \begin{cases} 1 - \text{Sí} & P(A) = 0.2 \\ 0 - \text{No} & P(\overline{A}) = 0.8 \end{cases}$$

x_i	p_i
0	0.80
1	0.20
	1

Ejemplo Bernoulli

- En el ejemplo anterior se dice que X sigue una **distribución de Bernoulli de parámetro 0.20** y se escribe **$X \sim B(0.20)$** ó **$X \sim B(1,0.20)$.**
- La media de la variable es E(X) = p=0.2.
- La varianza es V(X)=p·q=0.2·0.8=0.16.

Distribución Binomial

- Repetimos **n** veces un experimento de Bernoulli de forma independiente y definimos la **variable aleatoria X** = el **número de éxitos obtenidos en la realización de los n experimentos de Bernoulli.**
- La **variable aleatoria X** toma los **valores 0** (si no hubo ningún éxito), **1, 2, ..., n** (si todos los experimentos fueron éxito).

Distribución Binomial

- La distribución de probabilidad asociada será:

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \;\; k = 0,1,2,\ldots,n, \;\; p \in (0,1).$$

- El número combinatorio $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$ son las diferentes formas de seleccionar dónde ocurren los k éxitos en la realización de las n experiencias Bernoulli.

Distribución Binomial

- La variable X asociada a este tipo de experiencia se dice que **<u>tiene distribución de Binomial de parámetros n y p</u>**, y se denota por

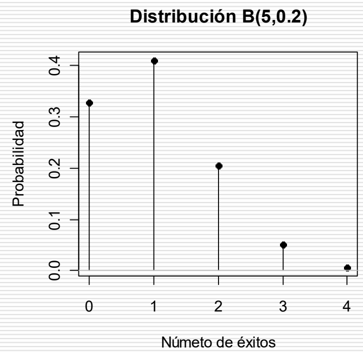
$$\mathbf{X \sim B(n,p)}, \text{ con } p \in (0,1)$$

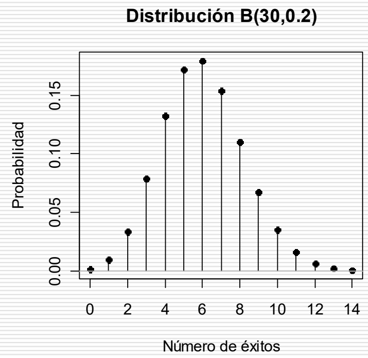
- La <u>media</u> de la variable es $E(X) = n \cdot p$.
- La <u>varianza</u> es $V(X) = n \cdot p \cdot q$.
- **<u>Propiedad</u>:** Si $X \sim B(n,p)$, entonces:

$$X = \sum_{i=1}^{n} X_i$$

con $X_i \sim B(1,p)$, $i=1, 2, \ldots, n$, independientes.

Gráficos Binomial





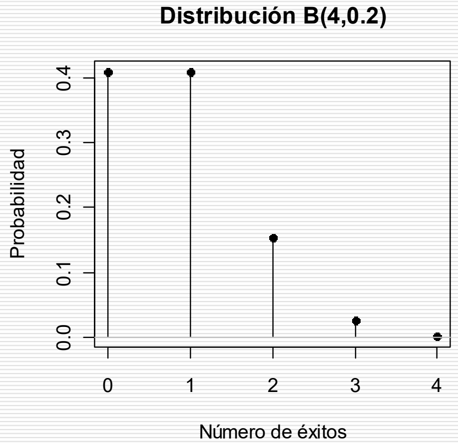
Ejemplo Binomial

- El porcentaje de bujías inservibles de una partida es del 20%. Quiero saber la probabilidad de que en un paquete de 4 bujías, 2 ó más sean inservibles.
  - X = Número de bujías (en un paquete de 4) que son inservibles $\sim B(4,0.2)$.
  - Estoy interesado en calcular $P(X \geq 2)$.

Ejemplo Binomial

x_i	p_i
0	$\binom{4}{0} 0.2^0 \cdot 0.8^4 = 0.8^4 = 0.4096$
1	$\binom{4}{1} 0.2^1 \cdot 0.8^3 = 4 \cdot 0.2 \cdot 0.8^3 = 0.4096$
2	$\binom{4}{2} 0.2^2 \cdot 0.8^2 = 0.1536$
3	$\binom{4}{3} 0.2^3 \cdot 0.8^1 = 4 \cdot 0.2^3 \cdot 0.8 = 0.0256$
4	$\binom{4}{4} 0.2^4 \cdot 0.8^0 = 0.2^4 = 0.0016$
	1

Ejemplo Binomial



Ejemplo Binomial

- La probabilidad pedida (probabilidad de un paquete elegido al azar de 4 bujías tenga 2 ó más bujías inservibles) será, por tanto,

$$P(X\geq 2) = P(X=2)+P(X=3)+P(X=4)=$$

$$=0.1536+0.0256+ 0.0016 = 0.1808,$$

o también,

$$P(X\geq 2)=1–P(X<2)=1–(P(X=0)–P(X=1)=$$

$$=1–0.4096– 0.4096= 0.1808.$$

Reproductividad Binomial

- Si tenemos dos variables X e Y **independientes**,

$$\left.\begin{array}{l} X \sim B(n_1,p) \\ Y \sim B(n_2,p) \end{array}\right\} \Rightarrow X+Y \sim B(n_1+n_2,p)$$

es decir, la suma de dos variables aleatorias independientes y con distribución binomial de parámetro común p, también sigue una distribución binomial.

Ejemplo reproductividad

- Con los datos de las bujías, ¿cuál es la probabilidad de que, elegidos dos paquetes de 4 bujías, haya, entre los dos, menos de 3 elementos inservibles?

X_i = Número de bujías (de un total de 4) que son inservibles en el paquete i (i = 1, 2).

Y= Número de bujías que son inservibles entre los dos paquetes = $X_1 + X_2$.

Ejemplo reproductividad

- X_1 y X_2 son independientes y aplicamos la **reproductividad:**

$$\left.\begin{array}{l} X_1 \sim B(4, 0.2) \\ X_2 \sim B(4, 0.2) \end{array}\right\} \Rightarrow X_1 + X_2 \sim B(4+4, 0.2) = B(8, 0.2)$$

- P(Y < 3) = P(Y=0)+P(Y=1)+P(Y=2)=

$$= \binom{8}{0} \cdot 0.2^0 \cdot 0.8^8 + \binom{8}{1} \cdot 0.2^1 \cdot 0.8^7 + \binom{8}{2} \cdot 0.2^2 \cdot 0.8^6 =$$

=0.1678 + 0.3355 + 0.2936 = 0.7969.

Distribución de Poisson

- La distribución de Poisson representa fenómenos reales en los que se analiza **el número de veces que ocurre cierto suceso en un intervalo** (en general de tiempo).
- **Ejemplos**
  - Desintegración de partículas radiactivas.
  - Número de llamadas a una central telefónica.
  - Número de clientes que acuden a solicitar un cierto servicio.

Experimento de Poisson

- El **experimento** debe cumplir que:
  - Las condiciones son constantes a lo largo de todo el intervalo.
  - Los resultados del experimento deben ser independientes si los intervalos son disjuntos.
  - La probabilidad de que el suceso ocurra una sola vez en un intervalo de amplitud h suficientemente pequeña, debe ser aproximadamente λh.
  - La probabilidad de dos o más ocurrencias del suceso, en un intervalo suficientemente pequeño, debe ser prácticamente cero.

Distribución de Poisson

- La **variable aleatoria X** que representa el **número de ocurrencias de un suceso en un intervalo de tiempo tiene una distribución de Poisson de parámetro** λ. Simbólicamente se denota $\mathbf{X \sim P(\lambda)}$.
- Toma los **valores 0** (si no hubo ninguna ocurrencia), **1, 2, ...**. Teóricamente, seguiríamos hasta el infinito.

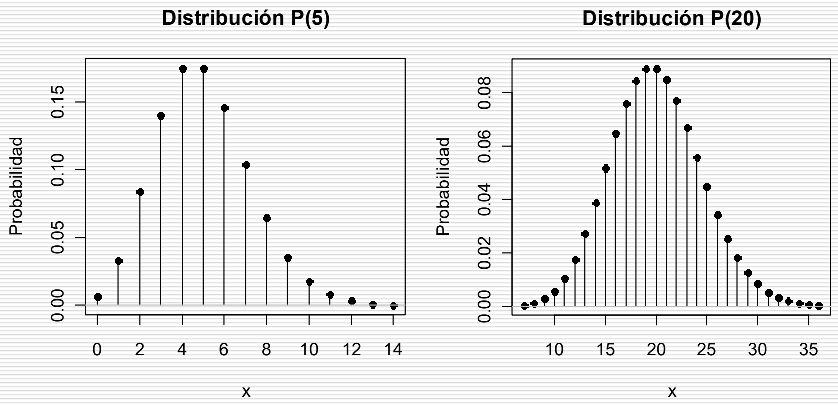
Distribución de Poisson

- La distribución de probabilidad asociada a una variable de este tipo tiene la forma:

$$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}, \text{ con } \lambda>0,\ k=0,1,2,\dots$$

- E(X)=λ.
- V(X)=λ.

Gráficos Poisson



Reproductividad Poisson

- Si tenemos dos variables X e Y independientes

$$\left.\begin{array}{l} X\sim P(\lambda_1) \\ Y\sim P(\lambda_2) \end{array}\right\} \Rightarrow X+Y\sim P(\lambda_1+\lambda_2),$$

es decir, la suma de dos variables aleatorias independientes y con distribución de Poisson λ_1 y λ_2, también sigue una distribución de Poisson.

Ejemplo Poisson

- En Asturias se producen una media de 120 accidentes mensuales.

a) ¿Cuál es la probabilidad de que, en un mes cualquiera, haya 50 accidentes?

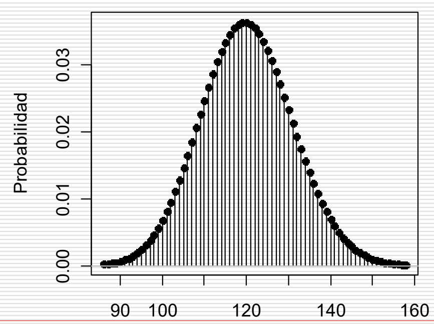
b) ¿Cuál es la probabilidad de que, en un día cualquiera, haya menos de 3 accidentes (supóngase un mes de 30 días)?

c) ¿Cuál es la probabilidad de que en una semana cualquiera haya menos de 3 accidentes?

Gráfica ejemplo

X = Nº de accidentes mensuales en Asturias~P(λ).
E(X) = 120 = λ. Por tanto, X ~P(120).

Distribución P(120)



Ejemplo Poisson

a) ¿Cuál es la probabilidad de que, en un mes cualquiera, haya 50 accidentes?

$$P(X=50)=\frac{e^{-120}\,120^{50}}{50!}=0.$$

b) ¿Cuál es la probabilidad de que, en un día cualquiera, haya menos de 3 accidentes (supóngase un mes de 30 días)?

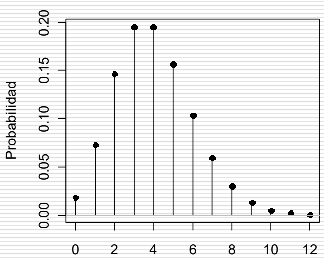
Y = Nº de accidentes diarios en Asturias

Ejemplo Poisson

$$Y\sim P\left(\frac{120}{30}=4\right)\Rightarrow \text{P(Y<3)=P(Y=0)+P(Y=1)+}$$

+P(Y=2)=0.0183+0.0733+0.1465=0.2381.

Distribution P(4)



Ejemplo Poisson

c) ¿Cuál es la probabilidad de que en una semana cualquiera haya menos de 3 accidentes?

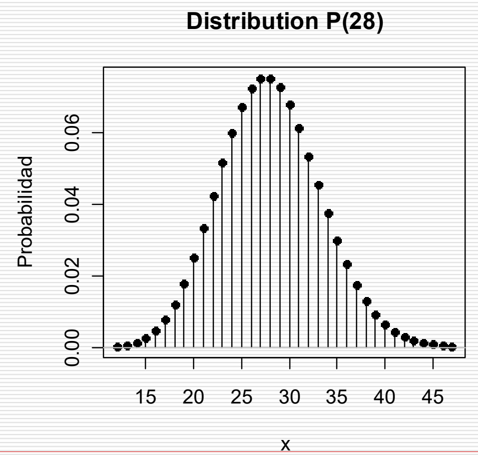
Z = Nº de accidentes semanales en Asturias.

$Z\sim P(4\cdot 7=28).$

P(Z<3)=P(Z=0)+P(Z=1)+P(Z=2)=

$$=\frac{e^{-28}\,28^{0}}{0!}+\frac{e^{-28}\,28^{1}}{1!}+\frac{e^{-28}\,28^{2}}{2!}=0.$$

Ejemplo Poisson



Ejemplo Poisson

d) Si en Cantabria el número medio de accidentes al mes es de 80, ¿cuál es la probabilidad de que haya 7 accidentes diarios entre las dos comunidades, asturiana y cántabra?

T=Nº de accidentes <u>mensuales</u> en Cantabria~ P(80)

T_d=Número de accidentes <u>diarios</u> en Cantabria~ P(80/30=2.6667)

$Y+T_d$=Nº de accidentes diarios entre Asturias y Cantabria ~ P(4+2.6667=6.6667)

$$P(Y+T_d=7)=\frac{e^{-6.6667}\,6.6667^7}{7!}=0.1478.$$

V.a. continuas

- Toman valores en, **al menos, un intervalo real**.
- Su distribución de probabilidad se determina con la <u>"función de densidad"</u>, función f con:
  - $f(x) \geq 0$.
  - $\int_{-\infty}^{+\infty} f(x)\,dx = 1.$

<u>Nota:</u> El valor de la función de densidad, f(x), **no es la probabilidad de x**, sino que representa la importancia o peso de x respecto a los demás valores de la variable X.

Probabilidades de v.a. continua

- **<u>La probabilidad asociada a cualquier subconjunto de números reales A se calcula:</u>**

$$P(X \in A) = \int_A f(x)\,dx,$$

que es el **área encerrada en el dominio A entre la función de densidad y el eje X**.

- **<u>Consecuencia:</u>** en variables continuas

$$P(X = x) = \int_x^x f(t)\,dt = 0.$$

F(x) de una v.a. continua

- La **función de distribución** de una variable aleatoria continua en un número real x se calcula como:

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

que es el **área encerrada en el dominio (-∞, x] entre la función de densidad y el eje X**.

Ejemplo v.a. continua

- Consideramos una variable aleatoria X que toma valores en el intervalo [1, 2] y en la que todos los valores tienen las mismas posibilidades de ocurrir **(distribución uniforme de parámetros 1 y 2, X ~ U(1,2))**.
  - La función de densidad es:

$$f(x) = \begin{cases} 1 & \text{si} \quad 1 \le x \le 2 \\ 0 & \text{otro caso} \end{cases}.$$

Ejemplo v.a. continua

- Una variable aleatoria X tiene función de densidad

$$f(x) = \begin{cases} ax & \text{si} \quad 1 \le x \le 3 \\ 0 & \text{otro caso} \end{cases}.$$

a) Calcula el valor de la constante a.

b) Calcula P(X≤2).

c) Calcula P(2<X<2.7).

Ejemplo v.a. continua

a) Como f es una función de densidad:
  - $f(x) \ge 0$. Por tanto, $a \ge 0$.
  - $1 = \int_{-\infty}^{+\infty} f(x)\,dx = \int_{1}^{3} a \cdot x\,dx = \left.\frac{a \cdot x^2}{2}\right|_1^3 = a \cdot \left(\frac{3^2 - 1^2}{2}\right) = 4 \cdot a$

$\Leftrightarrow \; a = \frac{1}{4}.$

b) $P(X \le 2) = \int_{-\infty}^{2} f(x)\,dx = \int_{1}^{2} \frac{x}{4}dx = \left.\frac{x^2}{8}\right|_1^2 = \frac{2^2 - 1^2}{8} = \frac{3}{8}.$

c) $P(2 < X < 2.7) = \int_{2}^{2.7} \frac{x}{4}dx = \left.\frac{x^2}{8}\right|_2^{2.7} = \frac{2.7^2 - 2^2}{8} = 0.4113.$

Esperanza de una v.a. continua

- **Esperanza matemática o media** de una variable aleatoria continua:

$$\mu_X = E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx.$$

- Como en v.a. discretas:
  - La esperanza matemática es el **modelo teórico de la media.**
  - Si la variable X se transforma de la **forma lineal** Y=a·X+b, entonces E(Y)=a·E(X)+b.

Varianza v.a. continua

- **Varianza de una variable aleatoria continua** es la esperanza matemática de las diferencias con la esperanza elevadas al cuadrado:

$$V(X) = \sigma_X^2 = E\left[(X - E(X))^2\right] = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx.$$

- El desarrollo de la expresión anterior permite calcular la varianza como **"esperanza del cuadrado menos el cuadrado de la esperanza":**

$$V(X) = \sigma_X^2 = E\left(X^2\right) - [E(X)]^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - [E(X)]^2.$$

Desviación v.a. continua

- **Desviación típica de una variable aleatoria** continua:

$$DT(X) = \sigma_X = +\sqrt{V(X)}.$$

  Se expresa en las mismas unidades de X.
- La desviación y la varianza de v.a. son los **modelos teóricos** de la desviación y la varianza de variables estadísticas.
- Si la variable X se transforma de la **forma lineal** Y=aX+b, entonces V(Y)=a^2·V(X) y DT(Y)=|a|·DT(X).

Ejemplo E(X)

- Una variable aleatoria X tiene función de densidad

$$f(x) = \begin{cases} \frac{1}{4}x & \text{si} \quad 1 \le x \le 3 \\ 0 & \text{otro caso} \end{cases}.$$

d) Calcula la esperanza y la varianza de X.

$$\mu_X = E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \int_1^3 \frac{x^2}{4} dx = \left.\frac{x^3}{12}\right|_1^3 = \frac{3^3 - 1^3}{12} = 2.1667.$$

Ejemplo V(X) y DT(X)

Como $V(X) = \sigma_X^2 = E\left(X^2\right) - [E(X)]^2.$

Para ello:

$$E\left(X^2\right) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_1^3 \frac{x^3}{4}dx = \left.\frac{x^4}{16}\right|_1^3 = \frac{3^4 - 1^4}{16} = 5.$$

Por tanto

$$V(X) = E\left(X^2\right) - [E(X)]^2 = 5 - 2.1667^2 = 0.3054,$$

$$DT(X) = +\sqrt{V(X)} = +\sqrt{0.3054} = 0.5526.$$

Modelos continuos

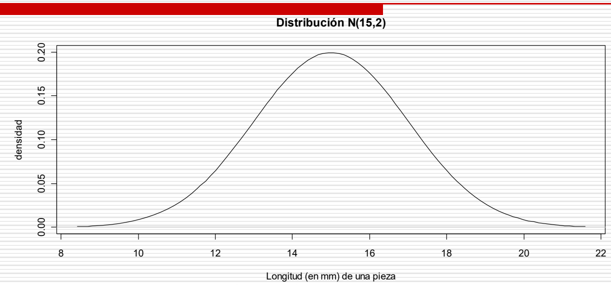
- Hay una serie de distribuciones continuas que sirven para modelizar variables habituales.

  Destacamos:
  - Modelo Normal
  - Modelo Exponencial

Distribución normal

- La distribución normal es la más frecuentemente empleada en la literatura estadística.
- Refleja el comportamiento de variables en las que la mayoría de los individuos toman valores intermedios y hay pocos individuos con valores muy altos o muy bajos.

Ejemplo distribución normal



- El gráfico corresponde a la producción de una máquina en que la mayoría de las piezas tienen una longitud alrededor de 15 mm., y hay muy pocas piezas en que la longitud está alrededor de 11 mm. o alrededor de 19 mm. (valores extremos).

Densidad normal

- Una variable aleatoria **X tiene distribución normal de parámetros μ y σ**, $\mu \in R$ y $\sigma \in R^+$, <u>X~N(μ, σ)</u>, si su distribución de probabilidad viene dada a través de la función de densidad:

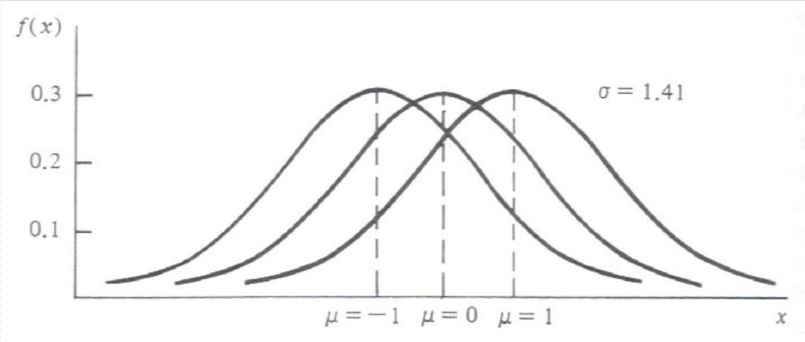
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad \forall\, x \in R$$

- $E(X)=\mu$.
- $V(X)=\sigma^2 \Leftrightarrow DT(X)= \sigma$.

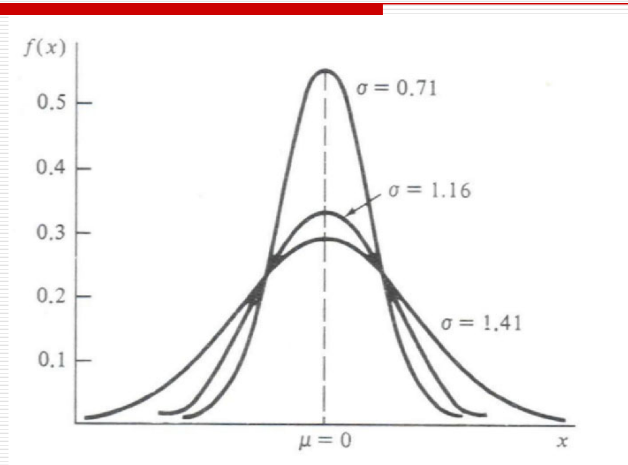
Campana de Gauss

- La representación gráfica de la distribución normal es la llamada "**campana de Gauss**":
  - Es simétrica respecto a la recta $x = \mu$.
  - Tiene máximo en μ.
  - Con puntos de inflexión en $\mu - \sigma$ y $\mu + \sigma$.
  - Son asíntotas los semiejes positivo y negativo de abscisas.

Normales con la misma dispersión



Normales con la misma media



Probabilidades de la normal

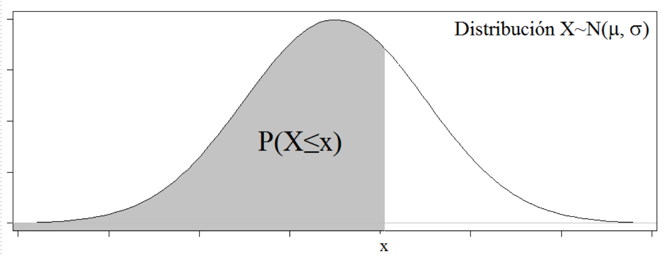
- La probabilidad de que una v.a. X~N(μ, σ) tome valores en un intervalo real (-∞, x] se calcula como:

$$\int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt,$$

esta integral no puede calcularse por métodos elementales por carecer de primitiva, pero puede aproximarse mediante **métodos numéricos**.

Probabilidades de la normal

- Los programas informáticos habitualmente proporcionan la probabilidad P(X≤x) para X~N(μ, σ).
- Gráficamente:



Cambio lineal de la Normal

- Si una v.a. X~N(μ, σ) se transforma de forma lineal **Y=aX+b**, entonces **la v.a. Y también sigue una distribución normal**

  X~N(μ, σ) ⇒ Y = aX+b ~N(a μ+b, |a| σ).

- **Particular:** Si una v.a. X~N(μ, σ) se tipifica se obtiene la N(0,1) (normal estándar o tipificada):

$$X \sim N(\mu,\sigma) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1).$$

Reproductividad de la normal

- Si tenemos dos variables X e Y independientes, con distribución

$$\left.\begin{array}{l} X \sim N(\mu_1,\sigma_1) \\ Y \sim N(\mu_2,\sigma_2) \end{array}\right\} \Rightarrow X+Y \sim N\left(\mu_1+\mu_2, \sqrt{\sigma_1^2+\sigma_2^2}\right),$$

es decir, la suma de dos variables aleatorias independientes y con distribución normal sigue a su vez una distribución normal.

Ejemplo normal

- Una empresa de derivados lácteos envasa sus yogures mediante una máquina de llenado. El contenido de un vasito de yogur se distribuye de acuerdo a una normal de media 125 g y desviación típica 5 g.

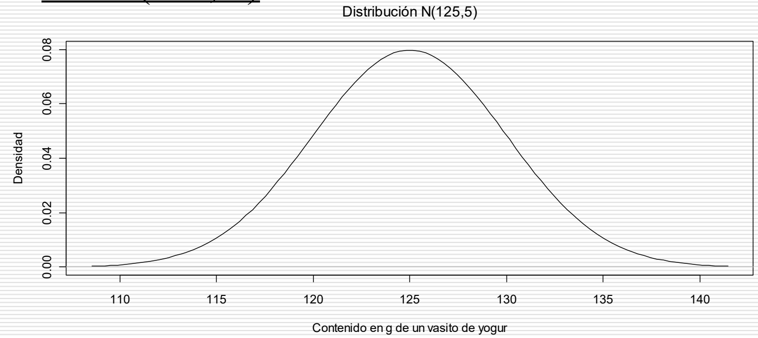
a) ¿Cuál es la probabilidad de que un vasito contenga exactamente 125 g?¿Y la de que contenga menos de 120 g?

b) En un paquete de ocho yogures, ¿cuál es la probabilidad de que, a lo sumo, uno de ellos contenga menos de 120 g?

c) Si cada vasito pesa exactamente 10 g, ¿cuál es la probabilidad de que el peso total de un paquete de ocho yogures sea superior a 1100 g?

Gráfico del ejemplo

- X = Contenido en g de cada vasito de yogur.
- X ~ N(125, 5).



Ejemplo Normal

a) ¿Cuál es la probabilidad de que un vasito contenga exactamente 125 g? ¿y la de que contenga menos de 120 g?

$$P(X=125)=\int_{125}^{125} f_{N(125,5)}(x)\,dx \ =0.$$

P(X<120) = P(N(125,5)<120)=0.1587.

Ejemplo Normal

b) En un paquete de ocho yogures, ¿cuál es la probabilidad de que, a lo sumo, uno de ellos contenga menos de 120 g?

Y = número de vasitos de yogur (entre los 8 del paquete) que tienen menos de 120 g ~ B(8, 0.1587)

$$P(Y\le 1)=P(Y=0)+P(Y=1)=$$

$$=\binom{8}{0}0.1587^0\,0.8413^8+\binom{8}{1}0.1587^1\,0.8413^7=$$

$$=0.6297.$$

Ejemplo Normal

c) Si cada vasito pesa exactamente 10 g, ¿cuál es la probabilidad de que el peso total de un paquete de ocho yogures supere los 1100 g?

Z=Peso total del paquete que contiene 8 yogures=$(X_1+10)+$

$+ (X_2+10) +\ldots+ (X_8+10) = \sum_{i=1}^{8} X_i + 80,$

con X_i = Contenido en g del i-ésimo vasito ~ N(125, 5).

Por la reproductividad: $\sum_{i=1}^{8} X_i \sim N(8\cdot 125 = 1000, \sqrt{8\cdot 25} = 14.1421).$

$$P(Z > 1100) = P\left(\sum_{i=1}^{8} X_i + 80 > 1100\right) = P\left(\sum_{i=1}^{8} X_i > 1020\right) =$$

$$= P\big(N(1000, 14.1421) > 1020\big) = 0.0786.$$

Distribución Exponencial

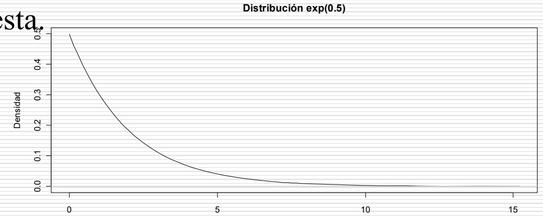
- Una variable aleatoria **<u>X tiene distribución exponencial de parámetro λ</u>, con λ>0**, X~Exp(λ), si su distribución de probabilidad viene por la función de densidad:

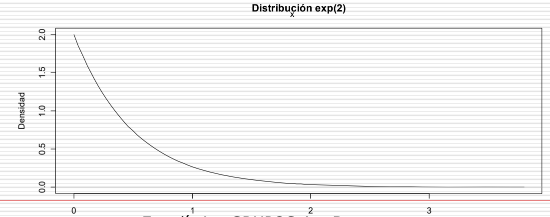
$$f(x) = \lambda\, e^{-\lambda x}, \quad \forall\, x > 0.$$

- E(X) = 1/λ.
- Var(X) = $1/\lambda^2$.
- Las probabilidades se calculan realizando la primitiva de la función de densidad.

Gráfica exponencial

- La función de densidad de una exponencial tiene forma de J transpuesta.





Relación Poisson y Exponencial

- La distribución de Poisson calcula la probabilidad de que ocurran unos sucesos durante un período de tiempo (o espacio) particular.
- En muchas aplicaciones el **tiempo transcurrido (o el espacio) entre la ocurrencia de dichos sucesos** <u>es la variable aleatoria</u> de interés en el estudio:

<u>Propiedad</u>: Si X~P(λ), **<u>la variable que mide el tiempo transcurrido hasta que ocurre el 1^er suceso</u>** (de Poisson), o, el tiempo entre dos sucesos consecutivos, se comporta como una **<u>Exp(λ).</u>**

Falta de memoria de la exponencial

- **<u>Propiedad de pérdida de memoria de la exponencial:</u>**

  Si X~Exp(λ), con s y t son cantidades positivas:

$$P\left(X > t+s \,/\, X > t\right) = P(X > s).$$

Ejemplo Exponencial

- El tiempo que un sistema electrónico funciona sin necesidad de realizarle ninguna reparación sigue una exponencial de media 50 días.

a) Probabilidad de que el sistema no tenga que ser reparado en 60 días, por lo menos.

b) Si en 10 días no se reparó, ¿qué probabilidad hay de que en los siguientes 20 días tampoco tenga que repararse?

c) Se quiere firmar un contrato de mantenimiento con una empresa con la obligación de que se realice una revisión cada n días. ¿Cuál debe ser el valor de n para que el 94% de las veces en que se realice el mantenimiento el sistema todavía no haya necesitado una reparación?

Ejemplo Exponencial

d) ¿Cuál es la distribución de la variable que representa el número de reparaciones que hay que realizar al sistema electrónico en un trimestre (90 días)?

e) Probabilidad de que haya que realizar más de 2 reparaciones en un mes (30 días).

f) ¿En qué momento se espera que se produzca la cuarta reparación del sistema?

Ejemplo Exponencial

- X = Días que funciona el sistema electrónico sin necesidad de ser reparado ~ **exp(λ)** .
- E(X) = 1/λ = 50 ⇒ **λ= 1/50.**
- Función de densidad: $f(x) = \begin{cases} \frac{1}{50} e^{-\frac{1}{50}x}, & \forall x > 0 \\ 0, & \text{en otro caso} \end{cases}.$

Ejemplo Exponencial

a) Probabilidad de que el sistema no tenga que ser reparado en 60 días, por lo menos.

$$P(X \geq 60) = \int_{60}^{+\infty} \frac{1}{50} e^{-\frac{1}{50}x} dx = -e^{-\frac{1}{50}x}\Big|_{60}^{+\infty} = e^{-\frac{60}{50}} = 0.3012.$$

b) Si en 10 días no se reparó, ¿qué probabilidad hay de que en los siguientes 20 días tampoco tenga que repararse?
Por la falta de memoria:

$$P\left(X > 10+20 \middle/ X > 10\right) = P(X > 20) = \int_{20}^{+\infty} \frac{1}{50} e^{-\frac{1}{50}x} dx =$$

$$= -e^{-\frac{1}{50}x}\Big|_{20}^{+\infty} = e^{-\frac{20}{50}} = 0.6703.$$

Ejemplo Exponencial

c) Se quiere firmar un contrato de mantenimiento con una empresa con la obligación de que se realice una revisión cada n días. ¿Cuál debe ser el valor de n para que el 94% de las veces en que se realice el mantenimiento el sistema todavía no haya necesitado una reparación?

$$0.94 = P(X > n) = \int_{n}^{+\infty} \frac{1}{50} e^{-\frac{1}{50}x} dx = -e^{-\frac{1}{50}x}\Big|_{n}^{+\infty} = e^{-\frac{n}{50}},$$

$$\Leftrightarrow \frac{-n}{50} = \text{Ln}\,0.94 \Leftrightarrow n = -50 \cdot \text{Ln}0.94 = 3.0938 \text{ días.}$$

Ejemplo Exponencial

d) ¿Cuál es la distribución de la variable que representa el número de reparaciones que hay que realizar al sistema electrónico en un trimestre (90 días)?

Por la relación entre Exponencial y Poisson:

- Y_d=Número de reparaciones que hay que realizar al sistema electrónico en un día (misma unidad que X) ~ P(λ) =P(1/50).
- Y_t=Número de reparaciones que hay que realizar en un trimestre (90 días) ~ P(90/50)=P(1.8).

Ejemplo Exponencial

e) Probabilidad de que haya que realizar más de 2 reparaciones en un mes (30 días).

- Y_m=Número de reparaciones que hay que realizar al sistema electrónico en un mes (30 días) ~ P(30/50)=P(0.6).
- Por tanto:

$$P(Y_m > 2) = 1 - P(Y_m \leq 2) = 1 - [P(Y_m=0) + P(Y_m=1) +$$

$$+ P(Y_m = 2)] = 1 - e^{-0.6} \cdot \left[\frac{0.6^0}{0!} + \frac{0.6^1}{1!} + \frac{0.6^2}{2!}\right] = 0.0231.$$

Ejemplo exponencial

f) ¿En qué momento se espera que se produzca la cuarta reparación del sistema?

- X_4 = Tiempo (en días) que pasa hasta que es necesaria la cuarta reparación del sistema.
- $X_4=X_1+X_{1\text{-}2}+X_{2\text{-}3}+X_{3\text{-}4}$, todos los sumandos Exp(0.02):
  X1=días hasta la 1ª reparación,
  $X_{1\text{-}2}$=días entre la 1ª y 2ª reparación, …
- $E(X_4)=E(X_1+X_{1\text{-}2}+X_{2\text{-}3}+X_{3\text{-}4})=E(X_1)+E(X_{1\text{-}2})+E(X_{2\text{-}3})+$
  $+E(X_{3\text{-}4})=4\cdot E(Exp(0.02))=4\cdot 50=200$ días.

Función de fiabilidad

- Si T es la v.a. que representa la duración de un dispositivo:
  - La **función de fiabilidad** (reliability) es la complementaria de F(t): R(t)= 1-F(t).
  - La **tasa de fallo media en un intervalo (t_1, t_2)** es:

$$h(t_1,t_2)=\frac{P\left(T\le t_2 \big/ T>t_1\right)}{t_2-t_1}=\frac{R(t_1)-R(t_2)}{(t_2-t_1)R(t_1)}.$$

  - La **tasa instantánea de fallo** o **tasa de riesgo** es:

$$h(t)=\lim_{t_2\to t}h(t,t_2)=\frac{f(t)}{R(t)}.$$

Teorema del Límite Central (TCL)

- **<u>Teorema del límite central</u>:**

  Dadas las variables aleatorias X_1, X_2,..., X_n (con n suficientemente grande) <u>independientes</u> y todas ellas con la **misma distribución,** de **media μ** y **varianza σ²**, la distribución de

$$\sum_{i=1}^{n} X_i \text{ (variable suma o total muestral)}$$

  sigue aproximadamente una distribución

$$N\left(n\cdot\mu, \sigma\cdot\sqrt{n}\right)$$

Notas al TCL

- <u>La aproximación se considera aceptable, en general, para valores de **n ≥ 30**</u>.
- En el teorema se parte de **variables cualesquiera** y se obtiene una **variable normal**, lo que justifica la importancia de tal distribución.
- Como consecuencia inmediata, por las propiedades de la variable normal, en las condiciones del TCL

$$\frac{\sum_{i=1}^{n} X_i}{n} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Ejemplo 1 aplicación TCL

- Dos amigos juegan al ajedrez. Por cada partida ganada, el perdedor le paga 1 € al ganador, y si hacen tablas nadie paga nada. Por sus años de experiencia, saben que el jugador 1 gana el 45% de las partidas, mientras que se hacen tablas en el 25%. Al cabo de 200 partidas:

a) ¿Cuál será la ganancia esperada del jugador 1?

b) ¿Cuál es la probabilidad de que el jugador 1 haya ganado más de 50 €?

Ejemplo 1 aplicación TCL

- Si consideramos X_i= "ganancia del jugador 1 en una partida i-ésima", se tiene que es una variable discreta con: valores 1€ (si gana), 0€ (si hacen tablas) y -1€ (si pierde), y sus probabilidades serán:

x_i	-1	0	1
p_i	0.3	0.25	0.45

Ejemplo 1 aplicación TCL

Si se tienen en cuenta las 200 partidas, tendremos una ganancia total representada por $X_1+\ldots+X_{200}$, suma de $200 \geq 30$ sumandos y verificando:

- Todas las variables tienen la misma distribución.
- Esperanza: $\mu = E(X_i) = -1 \cdot 0.3 + 0 \cdot 0.25 + 1 \cdot 0.45 = 0.15$.
- Varianza: $\sigma^2 = Var(X_i) = E(X_i^2) - [E(X_i)]^2 = (-1)^2 \cdot 0.3 + +1^2 \cdot 0.45 - (0.15)^2 = 0.7275 \Leftrightarrow$ desviación típica $\sigma = 0.8529$.

- Por ello, se puede aplicar el TLC y aproximamos la distribución de $X_1+\ldots+X_{200}$ por

$$N\left(n \cdot \mu, \sigma \cdot \sqrt{n}\right) = N\left(200 \cdot 0.15, 0.8529 \cdot \sqrt{200}\right) = N(30, 12.0618).$$

Ejemplo 1 aplicación TCL

a) ¿Cuál será la ganancia esperada del jugador 1?

$E(X_1+\ldots+X_{200}) = 30$ €.

b) ¿Cuál es la probabilidad de que el jugador 1 haya ganado más de 50 €?

Utilizando la aproximación del TLC:

$P(X_1+\ldots+X_{200}>50) \approx P(N(30,12.0618)>50) = 0.0486$.

Ejemplo 2 aplicación TCL

- En un restaurante, la probabilidad de que una mesa reservada no se ocupe durante la cena es 0.12.

a) Si el restaurante dispone de 60 mesas y se reservan todas, ¿cuál es la probabilidad de que menos de 10 mesas queden sin ocupar?

b) ¿Cuántas mesas habrá que disponer para reserva si se desea que, con probabilidad 0.75, no queden más de 5 mesas sin ocupar?

Ejemplo 2 aplicación TCL

a) Si el restaurante dispone de 60 mesas y se reservan todas, ¿cuál es la probabilidad de que menos de 10 mesas queden sin ocupar?

- Si definimos X= "nº mesas reservadas sin ocupar de entre las 60", su distribución es una B(60,0.12).
- P(X<10)=P(B(60,0.12)<10)=0.8224.

Ejemplo 2 aplicación TCL

b) ¿Cuántas mesas habrá que disponer para reserva si se desea que, con probabilidad 0.75, no queden más de 5 mesas sin ocupar?

Denoto por **"n"** el número de mesas. ¿n?

Y="nº mesas reservadas sin ocupar de entre las n"~B(n, 0.12), por lo que se puede expresar como:

$$Y = \sum_{i=1}^{n} Z_i$$

con $Z_i \sim B(0.12)$, i=1, 2, ..., n, independientes, $E(Z_i)=p=0.12$ y $V(Z_i)=p\cdot q=0.12\cdot 0.88=0.1056$.

Ejemplo 2 aplicación TCL

Asumiendo n≥30, se puede aplicar el TCL:

$$0.75 = P(Y \le 5) \approx P\left(N(n\cdot\mu, \sigma\sqrt{n} \le 5\right) =$$

$$= P\left(N(n\cdot 0.12, \sqrt{0.1056\cdot n} \le 5\right) = P\left(N(0,1) \le \frac{5-0.12\cdot n}{\sqrt{0.1056\cdot n}}\right) \Leftrightarrow$$

$$\Leftrightarrow 0.67 = \frac{5-0.12\cdot n}{\sqrt{0.1056\cdot n}} \Leftrightarrow 0.2177\cdot\sqrt{n} = 5-0.12\cdot n \Leftrightarrow$$

$$\Leftrightarrow 0.12\cdot n + 0.2177\cdot\sqrt{n} - 5 = 0,$$

y resolviendo la ecuación de segundo grado en $n^{1/2}$ se obtiene $n^{1/2}=5.6113$ (la otra solución es negativa), por lo tanto n=31.4867 y **n es 32 mesas**.