Estadística

2º curso del Grado en Ciencias de la Actividad Física y el Deporte

---o0o---

Descripción de las Poblaciones: Distribuciones de Probabilidad

Bioestadística - Facultad de Medicina

Universidad de Granada (España)

http://www.ugr.es/~bioest

PROBABILIDAD Y VARIABLE ALEATORIA

Fenómeno determinístico: Aquel en donde conocidas las condiciones de realización del fenómeno su resultado está determinado, es predecible. (Ej. Velocidad a la que llega un sólido al suelo tras lanzarlo de una altura de 1m; presión a la que está sometido un gas que tiene un determinado volumen; etc.....)

Fenómeno Aleatorio: Aquel en donde conocidas las condiciones de realización del fenómeno su resultado es impredecible, es decir, no podemos asegurar cuál será ese resultado. (Ej. Estado de un paciente tras un tratamiento; presentación de úlceras dérmicas en un paciente tras una estancia prologada en posición decúbito; glucemia de una persona tras una carrera; etc...)

En las Ciencias de la Vida siempre se trabaja con fenómenos aleatorios.

PROBABILIDAD Y VARIABLE ALEATORIA

Probabilidad y V.A. - 3
Resúmenes: 3.1

Ciencias de la Vida → Fenómenos aleatorios

Deberíamos obtener una
medida de la posibilidad de la ocurrencia
de cada uno de los resultados de un experimento aleatorio.
¿Es fácil que ocurra este resultado?

PROBABILIDAD Y VARIABLE ALEATORIA

Probabilidad y V.A. - 4
Resúmenes: 3.1

Primera estrategia (Frecuencista): Repetir el fenómeno aleatorio.

Ejemplo:

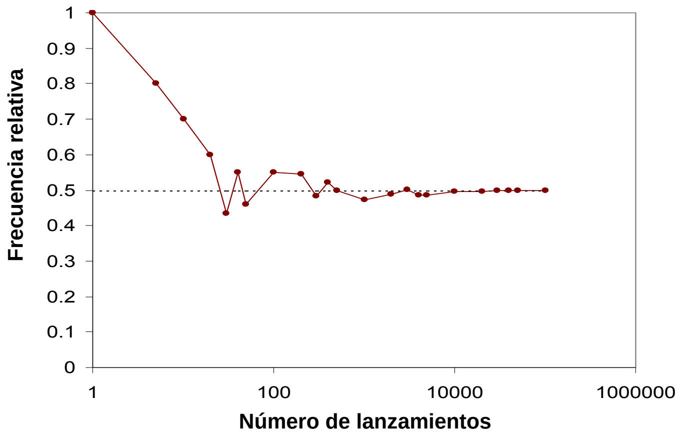
Lanzamos una moneda para intentar determinar una medida de la *posibilidad* de que salga una cara.

Primera medida: Frecuencia absoluta (me han salido 8 caras);
→ No porque aumenta conforme aumenta el número de lanzamientos.

Segunda medida: Frecuencia relativa (de los 20 lanzamientos de la moneda un 55% ha sido caras);
→ No porque cambia de muestra a muestra.

PROBABILIDAD Y VARIABLE ALEATORIA

Las frecuencias relativas cumplen la llamada ***ley de azar*** ó de ***estabilización de las frecuencias relativas.***



PROBABILIDAD Y VARIABLE ALEATORIA

Definición Frecuencista de la Probabilidad: La probabilidad de que ocurra algo, ***un suceso***, es el límite de la frecuencia relativa con que aparece dicho suceso cuando el número de repeticiones del experimento aleatorio es "suficientemente grande".

$$P(A) = \lim_{n \to \infty} h_n(A)$$

Esa definición, en sentido estricto, no sirve para calcular probabilidades, dado que no podemos repetir el experimento infinitas veces.

Pero puede representar una *aproximación* de gran utilidad

PROBABILIDAD Y VARIABLE ALEATORIA

Segunda estrategia (axiomática o Teórica): La probabilidad de que ocurra un suceso es un número que cumple una serie de propiedades (axiomas), teniendo que:

- $0 \le P(A) \le 1$
- Conforme más cerca esté de 1 más "fácil" es que ocurra el suceso
- Conforme más cerca esté de 0 más "difícil" es que ocurra el suceso.

¿Qué pasa si P(A)=1? ¿y si es P(A)=0? ¿Contradice esta definición a la frecuencista?

Todas las afirmaciones que se hagan acerca de los fenómenos aleatorios se harán en términos de la probabilidad de que ocurra algo

PROBABILIDAD Y VARIABLE ALEATORIA

Un nuevo concepto: Cuando los resultados de un experimento aleatorio son numéricos (ó se pueden convertir en números), se dice que el resultado es una ***variable aleatoria*** (porque sus valores numéricos varían al azar): Talla, Peso, tiempo hasta la curación , número de ingresos en urgencia, etc...

Variables Aleatorias:.

- **Discretas:** Aquellas que toman valores aislados. Entre dos cualesquiera valores de la variable no siempre existe otro valor posible de la variable.
- **Continuas:** Entre dos valores cualesquiera de la variable siempre hay otro valor posible de la variable.

Las **variables aleatorias,** como resultados ligados a los fenómenos aleatorios, siempre van afectadas de **probabilidades**: las afirmaciones para ellas **no son nunca seguras.**

PROBABILIDAD Y VARIABLE ALEATORIA

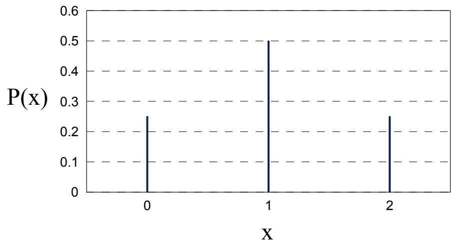
Variables aleatorias Discretas

A cada valor se le asigna una probabilidad → ***función de probabilidad***.

Ejemplo: Número de varones de una familia de dos hijos:

Valores = 0, 1, 2

Probabilidades de esos valores: 0.25, 0.50, 0.25



Para que una variable aleatoria de tipo discreto este perfectamente determinada no basta con conocer los valores que puede tomar, hay que conocer también la probabilidad con que toma esos valores.

PROBABILIDAD Y VARIABLE ALEATORIA

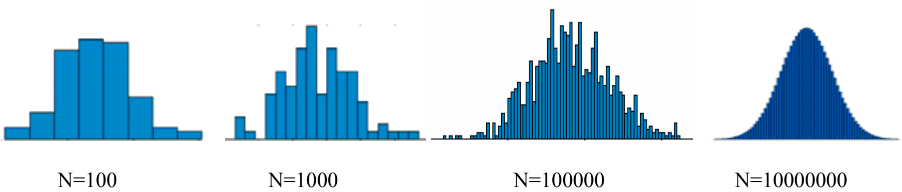
Variables aleatorias Continuas

La función que hace el papel de la función de probabilidad se denomina ***función de densidad*** y juega el mismo papel.

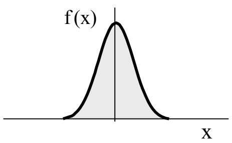
Ejemplo: Índice de Masa Corporal (IMC = Peso(kg)/Talla(m)^2).

- Valores: Cualquier valor real entre (digamos) 10.00 y 60.00.

- Función de densidad: Fórmula ¡compleja! que nos permite calcular probabilidades.



Para **N** suficientemente grande el histograma perfila una curva: la **función de densidad** f(x), que tendrá una expresión matemática que permite reproducirla, obteniendo así un **modelo teórico de distribución de probabilidad**



PROBABILIDAD Y VARIABLE ALEATORIA

Toda variable aleatoria deberá ir acompañada de su función de probabilidad, si es discreta, o de su función de densidad, si es continua.

Tanto la una como la otra serán conocidas cuando se conozcan los parámetros de los que dependen

Existen unas *grandes familias* de variables aleatorias que definen a las más importantes.

PROBABILIDAD Y VARIABLE ALEATORIA

Variable aleatoria Normal o Gaussiana

Tipo: Continuo.

Valores: Cualquier valor entre $-\infty$ y $+\infty$

Función de densidad: (Campana de Gauss)

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

Parámetros de la variable aleatoria:

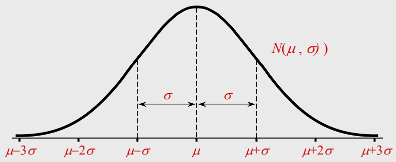
μ : Media, moda y mediana de la variable aleatoria
σ : Desviación típica de la variable aleatoria

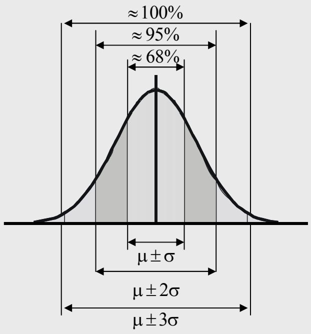
Se indica $x \rightarrow N(\mu, \sigma)$: x sigue una distribución normal de media μ y desviación típica σ

PROBABILIDAD Y VARIABLE ALEATORIA

Variable aleatoria Normal o Gaussiana

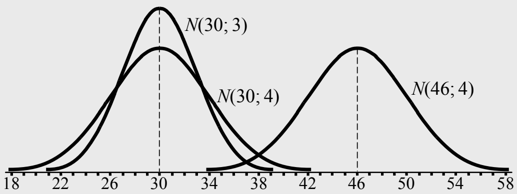
$$x \to N(\mu,\sigma) \;\; \text{si} \;\; f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$





PROBABILIDAD Y VARIABLE ALEATORIA

Probabilidades: área bajo la curva = integral definida es complicado ⇒ para eso existen las tablas.



Representación gráfica de dos distribuciones Normales con igual media pero distinta varianza (las dos de la izquierda) y de otras dos con igual varianza y distinta media (las dos más bajas)

Tablas: tendría que haber ∞ (una para cada pareja de μ´s y σ´s), pero basta con una: la Tabla 1 de la distribución *Normal Típica* o *Normal Estándar*:

$$z \to N(0;\,1) \quad \equiv \quad z \to N\{\mu=0;\sigma=1\}$$

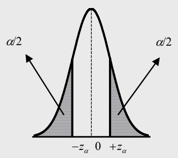
PROBABILIDAD Y VARIABLE ALEATORIA

Tabla de la distribución normal típica o estándar

Construida de un modo útil para las preguntas más frecuentes:

$$P\{-z_\alpha \le z \le +z_\alpha\} = 1-\alpha \;\equiv\; P\{|z| \le z_\alpha\} = 1-\alpha \;\equiv\; P\{|z| \ge z_\alpha\} = \alpha$$

$$P\{z \ge +z_\alpha\} = \alpha/2 \;\;,\;\; P\{z \le -z_\alpha\} = \alpha/2$$



α en el exterior de la Tabla 1

z_α en el interior de la Tabla 1

PROBABILIDAD Y VARIABLE ALEATORIA

Ejemplo: *¿Entre qué dos valores está una N(0;1) el 95% de las veces?*

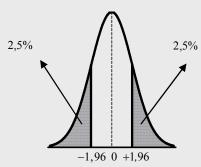
$$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \text{ (dato)} \Rightarrow \text{se pide } \pm z_\alpha = \pm z_{5\%}$$

α	0,00	0,01	0,02	0,03	0,04	**0,05**	0,06	0,07	0,08	0,09
0,0	∞	2,576	2,326	2,170	2,054	**1,960**	1,881	1,812	1,751	1,695

Nota: Para cada valor de α (en el caso de la tabla principal, la suma de la primera columna y de la primera fila) en el interior de la tabla se da el valor z_α tal que a la izquierda de $-z_\alpha$ y a la derecha de $+z_\alpha$ hay un área total de α.

- Respuesta: entre −1.96 y 1.96 $\Rightarrow$

$$P\{-1.96 \le z \le 1.96\} = 95\% \equiv P\{|z| \le 1.96\} = 95\% \equiv P\{|z| \ge 1.96\} = 5\%$$



- General: $-z_\alpha \le z \le +z_\alpha$ el 1−α de las veces (dos colas)

  $-1.96 \le z \le 1.96$ el 95% de las veces.

PROBABILIDAD Y VARIABLE ALEATORIA

Ejemplo: *¿Cuál es el percentil 95? ≡ ¿Qué valor no supera z el 95% de las veces?*

- Percentil 95: $p_{95} \Rightarrow P\{z \le p_{95}\} = 95\% \Rightarrow P\{z \ge p_{95}\} = 5\% = \alpha/2 \Rightarrow \alpha = 10\%$ en Tabla 1 $\Rightarrow \boldsymbol{p_{95} = +z_{10\%} = 1.645}$

α	**0,00**	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,1	**1,645**	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311

- General: $z \le z_{2\alpha}$ el 1−α de las veces (una cola)
  z ≤ +1.645 el 95% de las veces.

Ejemplo: *¿Cuál es el percentil 5? ≡ ¿Qué valor supera z el 95% de las veces?*

- Percentil 5: $p_5 \Rightarrow P\{z \le p_5\} = 5\% = \alpha/2 \Rightarrow \mathrm{p}_5 = -z_{10\%} = -1.645$
- General: $z \ge -z_{2\alpha}$ el 1−α de las veces (una cola)
  z ≥ −1,645 el 95% de las veces.

PROBABILIDAD Y VARIABLE ALEATORIA

Tipificación

- Se demuestra que: si $x \to N(\mu;\sigma) \Rightarrow z = \dfrac{x-\mu}{\sigma} \to N(0;1) \Rightarrow$ *z* v.a. *x* tipificada ≡ *Tipificar x* es hacer $(x-\mu)/\sigma$.

- $-z_\alpha \le z = \dfrac{x-\mu}{\sigma} \le +z_\alpha$ el 1−α de las veces ⇒ operando:
  $\mu - z_\alpha\sigma \le x \le \mu + z_\alpha\sigma$ el 1−α de las veces.

- $\left\{\begin{array}{l} x \in \mu \pm z_\alpha\sigma \text{ el } 1-\alpha \text{ de las veces} \\ x \le \mu \pm z_{2\alpha}\sigma \text{ el } 1-\alpha \text{ de las veces} \\ x \ge \mu - z_\alpha\sigma \text{ el } 1-\alpha \text{ de las veces} \end{array}\right.$ permite obtener intervalos para los valores “normales” de *x*

PROBABILIDAD Y VARIABLE ALEATORIA

Ejemplo:

- Si *x* = estatura español $\to N(\mu=170; \sigma=10)$:

¿Entre qué estaturas está el 95% de los españoles?

- $1-\alpha=0.95 \Rightarrow \alpha=0.05 \Rightarrow z_\alpha=1.96 \Rightarrow x\in 170\pm 1.96\times 10 \Rightarrow$ de 150.4 a 189.6
- 1.96 es muy habitual $(\equiv z_{5\%})$, pero hay otros que se usan:

	$x\in\mu\pm1\times\sigma$	$x\in\mu\pm2\times\sigma$	$x\in\mu\pm3\times\sigma$	
z_α =	1	2	3	en Tabla 1
$1-\alpha \simeq$	68%=2/3	95%=19/20	100%	
contiene a:	la mayoría	casi toda	toda	(la población)

¿Qué estatura no supera el 5% de los españoles? (p_{95})

$p_{95}=170+1.645\times10=186.45 \Rightarrow x\le 186.45$ en el 95% de los españoles.

¿Qué estatura supera el 95% de los españoles? (p_5)

$p_5=170-1.645\times10=153.5 \Rightarrow x\ge 153.5$ en el 95% de los españoles.

Utilidad general

Para dar valores normales de estaturas, glucosa, presión, … (dos colas) o valores mínimos o máximos normales de algo.

PROBABILIDAD Y VARIABLE ALEATORIA

Utilidad extra de la tipificación

- Para comparar dos cantidades no comparables, como en el ejemplo que sigue.
- Si $P(\text{peso})\to N(69;4)$, $E(\text{estatura})\to N(170;6)$ y un individuo tiene P_0=72 y E_0=173: ¿es más pesado que alto?
- Son dos cantidades no comparables, para que lo sean $\Rightarrow$ tipificar:

$$z_P=\frac{72-69}{4}=0.75>z_E=\frac{173-176}{6}=0.50 \Rightarrow$$ la respuesta es que sí

- De igual modo para asignar una beca a dos estudiantes de diferentes estudios.

Comprobación de la normalidad de una variable aleatoria

- Para comprobar si una muestra (de pesos, por ejemplo) proviene de una v.a. Normal, pueden emplearse dos procedimientos:
  - Heurístico (aproximado): verificar si se adapta bien al histograma una curva Normal (se verá en SPSS).
  - Test de Normalidad (exacto): se verá más adelante.
- El *Teorema Central del Límite* permite justificar la abundancia de la distribución Normal: si una variable X (estatura) se puede expresar como la suma de variables independientes:

  $X = X_1$ (efecto del padre) + X_2 (madre) + X_3 (alimentación) + … + X_K

  entonces, aproximadamente, $X \to$ *Normal si K es grande* (a veces eso no pasa con X, pero sí con el logaritmo de X).

Caso especial de $\bar{x}$

- $\bar{x} = (x_1 + x_2 + \ldots + x_n) / n$ es v.a., pues varía con la muestra.
- $\bar{x} = \frac{x_1}{n} + \frac{x_2}{n} + \ldots + \frac{x_n}{n} = X_1 + X_2 + \ldots + X_n \to$Normal si n es grande (por el TCL).
- $\bar{x} \to$ Normal siempre si $x \to$ Normal (se demuestra)

$\Downarrow$

(a) Si x es una v.a. cualquiera de media μ y desviación típica σ, y si $\bar{x}$ es la media de una muestra de tamaño $n \geq 30$, $\bar{x}$ se distribuye aproximadamente como una Normal: $\bar{x} \to N(\mu;\ \sigma/\sqrt{n})$, con $\sigma/\sqrt{n}$ el *error estándar*.

(b) Si x es Normal, lo anterior se verifica exactamente para cualquier valor de n.

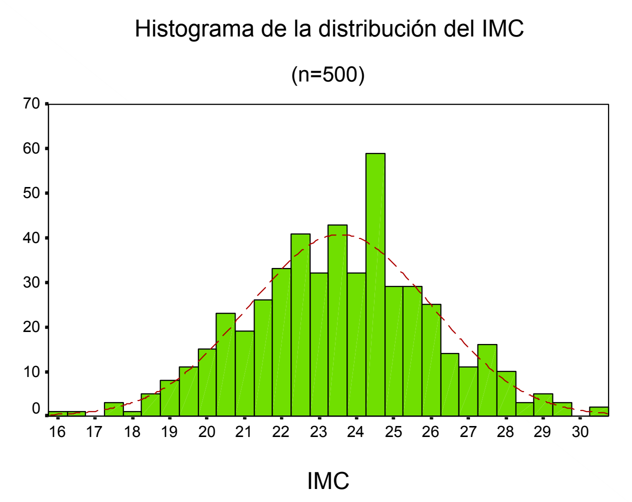
- Desviación estándar (σ) ≠ Error estándar $\left(\sigma/\sqrt{n}\right)$.
- σ mide la variabilidad de x; $\sigma/\sqrt{n}$ mide la de $\bar{x}$.

Utilidad

- x = nº hijos de una familia: no es Normal (por ser discreta).
- Peso $\bar{x}$ $(\text{si } n \geq 30) \to$Normal (aproximadamente).
- ¡Importancia de tomar muestras grandes! para que sea aplicable la distribución Normal a $\bar{x}$: se utilizará con frecuencia más adelante.

PROBABILIDAD Y VARIABLE ALEATORIA

Ejemplo de variable aleatoria Normal



PROBABILIDAD Y VARIABLE ALEATORIA

Variable aleatoria Binomial

(a) Si de una población de tamaño (N) infinito, cuyos individuos verifican una cierta característica dicotómica con probabilidad p, se extrae una muestra de tamaño n, el número x de individuos, de entre los n, que verifican la característica sigue una distribución Binomial; esto se expresa abreviadamente diciendo que $x \to B(n;\, p)$.

(b) Cuando $N<\infty$, x sigue aproximadamente una Binomial si $N > 40$ y n/N (*fracción de muestreo*) $\leq 0{,}10$.

PROBABILIDAD Y VARIABLE ALEATORIA

Ejemplos:

- Número de caras (x) en 100 lanzamientos de una moneda (n) es Binomial pues $N = \infty$ (la moneda se puede lanzar indefinidamente). Aquí p=% de caras.
- Número de varones (x) entre 50 españoles (n) es Binomial pues, si bien $N<\infty$ (hay 44 millones de españoles), la fracción $n/N = 50/(44\times10^6)$ es pequeña. Aquí p=% de varones.
- Número de curados (x) entre 200 enfermos (n) es Binomial pues, si bien $N<\infty$, la fracción n/N es pequeña. Aquí p=% de curados.
- Número de aprobados (x) entre 20 alumnos (n) de una clase de 100 (N) no es binomial: N=100>40 pero n/N = 20/100 = 0,20 > 0,10.
- Número de aprobados (x) entre 5 alumnos (n) de una clase de 20 (N) no es binomial: $N=20 \leq 40$.

(las dos últimas son v.a. *hipergeométricas*).

<u>En general</u>
La Binomial está asociada a los problemas en los que estudiamos el % de individuos que verifican algo

PROBABILIDAD Y VARIABLE ALEATORIA

<u>Variable aleatoria Binomial</u>

Tipo: Discreto.
Valores: 0,1,2,....,n
Función de probabilidad:

$$P(x=r)=\binom{n}{r}p^r(1-p)^{n-r}, \quad r=0,1,2,\ldots,n$$

Parámetros de la variable aleatoria: n, p

Se indica: x→B(n,p)

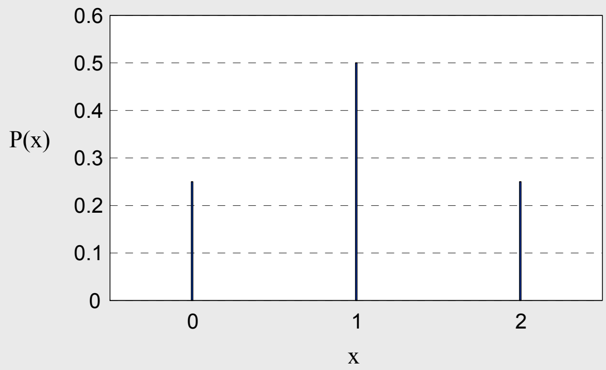
Media: $\mu = np$

Varianza: $\sigma^2 = np(1-p)$

PROBABILIDAD Y VARIABLE ALEATORIA

Ejemplo de variable aleatoria binomial

$$x \to B(2;\,0.5) \;\text{ si }\; P(x=r) = \binom{2}{r} 0.5^{r}(1\text{-}0.5)^{2\text{-}r}, \quad r = 0,1,2$$



PROBABILIDAD Y VARIABLE ALEATORIA

Aproximación a la normal

- Si se asigna un valor 0 (1) al individuo que no (sí) verifica la característica en estudio (varón por ejemplo):
  $x = 0 + 1 + \;\ldots\; + 1 \to$ Normal si *n* grande (por TCL).
- Si $x \to B(n;p) \Rightarrow x \to N\left(p;\sqrt{npq}\right)$ si *n* es grande.
- Como *B* (binomial) es discreta y *N* (Normal) es continua, eso no puede entenderse así:
  $$0 \neq P\{B=4\} = P\{N=4\} = 0$$
  sino como que la Normal redondeada da la Binomial:
  $$P\{B=4\} = P\{3{,}5 \leq N \leq 4{,}5\} = P\{N \in 4 \pm 0{,}5\}$$
  con 0,5 = 1/2 = ”mitad del salto de la Binomial”.
- De modo general, si una variable discreta salta de *h* en *h* (*h* = 1 antes) ha de hacerse una *corrección por continuidad* (cpc) -ella aparece con frecuencia en adelante- consistente en sumar y/o restar la mitad del salto (*h*/2 ó 1/2 = 0,5 de antes).

PROBABILIDAD Y VARIABLE ALEATORIA

Variable aleatoria Poisson

Son variables *x* de Poisson:

i) Aproximadamente, una Binomial con *n* grande y *p* pequeño (nº de enfermos de SIDA entre 1.000 españoles elegidos al azar), de ahí que se la conozca como *la ley de los sucesos raros.*

ii) El número de partículas por unidad de medio, si un gran número de partículas están repartidas al azar en una gran cantidad de medio (nº de granulocitos por milímetro cúbico de sangre; nº de bacterias por milímetro cuadrado de cultivo).

iii) El número de sucesos que ocurren por unidad de tiempo, si estos suceden al azar e independientemente entre sí (nº de urgencias al día en un hospital; nº de partículas α emitidas en un segundo por una sustancia radioactiva).

PROBABILIDAD Y VARIABLE ALEATORIA

Variable aleatoria Poisson

Características de la distribución de Poisson

- $x = 0, 1, 2, \ldots, \infty$.
- Media = $\mu = \lambda$, Varianza = $\sigma^2 = \lambda$.
- Si $x \to P(\lambda) \Rightarrow x \to N(\lambda; \sqrt{\lambda})$ si λ es grande.

PROBABILIDAD Y VARIABLE ALEATORIA

Todos los problemas de Estadística los dejaremos resumidos a intentar conocer 'algo' acerca de esos parámetros, que son generalmente desconocidos, y que nos van a permitir hacer afirmaciones acerca de las variables aleatorias.

Todo problema será expresado en función de una variable aleatoria y de los parámetros que la definen, haciéndose afirmaciones acerca de ellos que nos permitirán contestar a las distintas cuestiones.

Fin del capítulo