Distribuciones

CONTENIDO

- OBJETIVO
- CONCEPTOS
- TIPOS DE DISTRIBUCIONES
- DISTRIBUCIONES TEÓRICAS
- ELECCIÓN DE LA DISTRIBUCIÓN

PROCESOS

- ESTOCÁSTICOS
  - VARIAS POSIBLES SITUACIONES EN T+1
- DETERMINISTAS
  - UNA POSIBLE SITUACIÓN EN T+1 CONOCIDA

OBJETIVO EN LA SIMULACIÓN

- Implementar la aleatoriedad de las variables en la simulación discreta, asignándoles distribuciones estadísticas
  - ejercicio

Acercarse a la realidad

OBJETIVO EN LA SIMULACIÓN

- Usar un modelo probabilístico no te permite predecir el resultado de un experimento individual pero se puede determinar la probabilidad de que una salida caiga dentro de un rango específico de valores

- Los comportamientos particulares o individuales no son importantes frente a los valores medios de comportamiento

CONCEPTOS: PDF

- Función de densidad de probabilidad PDF:
  - La probabilidad de que un valor x ocurra es P(x), expresado con una fracción o un porcentaje
  - La probabilidad de que x tenga un valor entre a y b viene dado por el área bajo la curva de P(x)
  - La probabilidad de que x tenga un valor entre $-\infty$ y $+\infty$ es 1
  - La PDF es cero o positiva

LEY DE LOS GRANDES NÚMEROS

PROBABILIDAD DE QUE AL ARROJAR UN DADO SALGA UN 1

- Experimentalmente: Se tira el dado 1000 veces y se cuenta el número de resultados positivos (1): 149
  - Probabilidad de sacar 1: 149/1000 = 14.9%.
- Teóricamente: cada uno de los 6 números tiene igual probabilidad de salir
  - Probabilidad de sacar 1: 1/6 = 16.7%

Según aumenta el número de experimentos, la probabilidad determinada experimentalmente se aproximará al valor teórico

LEY EXPERIMENTAL DE LOS GRANDES NÚMEROS

EJERCICIO

- SIMULAR MEDIANTE UN GENERADOR DE NÚMEROS ALEATORIOS ENTRE 1 Y 6 EL HECHO DE TIRAR UN DADO

- CALCULAR LA PROBABILIDAD DE QUE SALGA UN NÚMERO IMPAR

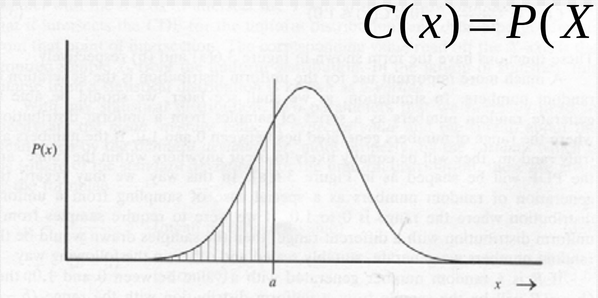
- COMPARAR CON EL VALOR TEÓRICO (50%)

CONCEPTOS: CDF

- Función de densidad acumulativa CDF:
  - La probabilidad de que x tenga un valor entre -∞ y a

$$C(x)\big|_{x=a} = \int_{-\infty}^{a} P(x)dx$$

$$C(x) = P(X \le x)$$



CONCEPTOS: DISTRIBUCIÓN ESTADÍSTICA DE PROBABILIDAD

- Cómo se reparten las posibilidades
- Parámetros de las distribuciones :
  - Forma (tipo)
  - Valor medio, media o valor esperado μ
  - Desviación estándar sobre la media
  - Varianza σ (raíz cuadrada de la desviación estándar)

La elección de una correcta distribución es crucial

EJEMPLO: DATOS PROBABILÍSTICOS

ROTURA DE UNA MÁQUINA ⇨ proceso estocástico

esto no significa que no se pueda decir nada sobre el comportamiento de la máquina en este aspecto

El comportamiento de roturas y fallos se describe por dos variables estocásticas:

- Tiempo de reparación (MTR: mean time to repair)
- Tiempo de operación (MTBF: mean time between failure) o el tiempo que un recurso funciona después de su reparación

Ejemplo: Datos históricos sobre los tiempos en los que una máquina se ha roto en 100 horas:

0,7,15,23,27,33,40,45,54,61,68,76,81,90,98

Intervalo de tiempo entre cada dos roturas consecutivas:

7,8,8,4,6,7,5,9,7,7,8,5,9,8

EJEMPLO: DATOS PROBABILÍSTICOS

Media: 98/14 = 7

Desviación estándar: ((0+1+1+3+1+0+4+4+0+0+1+4+4+1)/(14-1)) = 1.52

Distribución de frecuencias de intervalos entre roturas

valor	4	5	6	7	8	9	10
%	7	14	7	29	29	14	0
Frec.	1	2	1	4	4	2	0

CONCLUSIONES

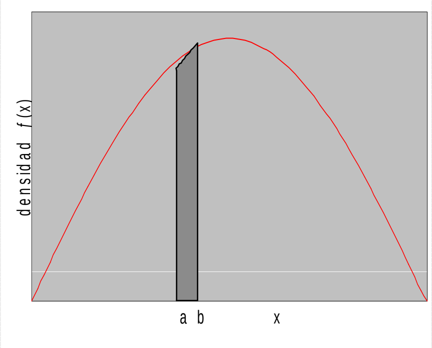
Asumiendo que estos datos dan una buena representación de la realidad, se puede decir que hay una media de 7 horas antes de la próxima rotura, y que la probabilidad de que un fallo ocurra en 6 horas (entre 5.5 y 6.5 horas) es del 7%.

TIPOS DE DISTRIBUCIONES

- CONTINUAS O DISCRETAS
- EMPÍRICAS O TEÓRICAS

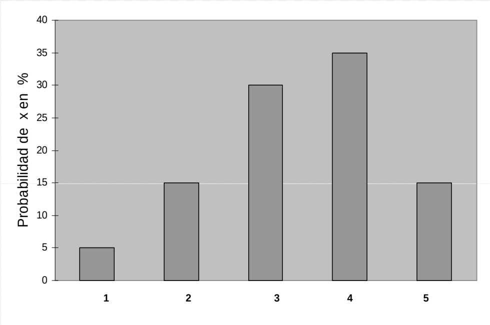
DISTRIBUCIONES CONTINUAS

- Permiten obtener todos los valores en un rango (el nº de valores posibles es ∞)
- Las probabilidades de obtener los valores varían de forma continua
- No se puede dar la probabilidad de un valor determinado



DISTRIBUCIONES DISCRETAS

- Una variable estocástica discreta representa a un número finito de valores posibles
- Un valor en particular tiene una cierta probabilidad de ocurrir

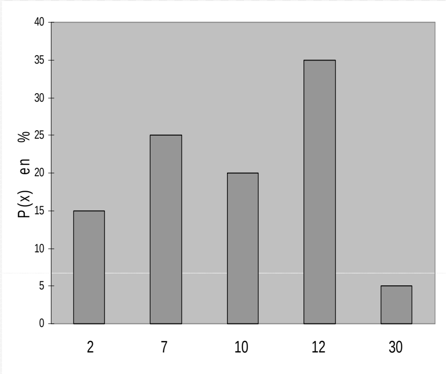


DISTRIBUCIONES EMPÍRICAS

- Basadas en datos medidos
- Continuas y discretas
- Datos para ajustar la forma
  - No suelen estar disponibles (muchos)
- Bondad del ajuste: Tests
- Difíciles y laboriosas
- Descripción incorrecta o idea falsa
- Útiles para horarios, etc. (pocos datos)
  - Llegada de aviones en el día

EJEMPLO: DISTRIBUCIÓN EMPÍRICA

P(x)	Valor
15%	2
25%	7
20%	10
35%	12
5%	30



Histograma

BONDAD DEL AJUSTE

- Dibujar el histograma y comprobar visualmente si tiene la misma forma que la función de densidad teórica
  - La media y la desviación estándar de la función de densidad deben ser similares a los resultados observados

Juzgar la similitud visualmente es muy subjetivo

TEST PARA BONDAD DEL AJUSTE

Medir la similitud entre funciones de probabilidad empíricas y teóricas y expresarla como un valor

TESTS: Chi-square test

Kolmogorov-Smirnov

- Puntos a tener en cuenta:

Una desviación del 10% en la media es mucho peor que una desviación del 10% en la desviación estándar, que a su vez es mucho peor que un pequeño error en la forma de la distribución

DISTRIBUCIONES TEÓRICAS

- Basadas en el conocimiento
- Continuas y discretas
- No datos históricos disponibles
- Conocimiento
  - Intuitivo
  - Medidas
  - Asumiendo suposiciones
- Son mejores para describir la realidad
- Más fáciles de usar (ya construidas en el programa)
- Caracterización simple

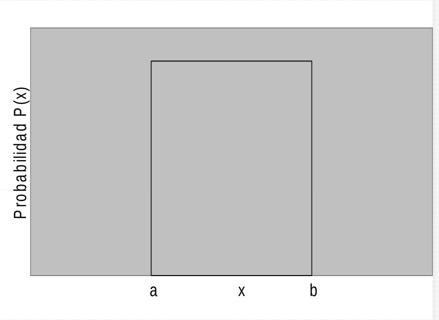
DISTRIBUCIONES TEÓRICAS

DISTRIBUCIÓN UNIFORME

Puede usarse si se tiene información muy general o poco conocimiento sobre la distribución

Simétrica

$$P(x)=\begin{cases}\dfrac{1}{(b-a)+1}, & a \le x \le b\\ 0 & \end{cases}$$

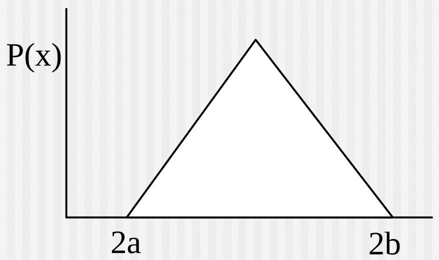


Errores de redondeo en medidas, generación de números aleatorios, llegadas (cuando son independientes y el nº total está determinado)

DISTRIBUCIÓN TRIANGULAR

Si *x*, *y* son variables uniformes,

z = *x* + *y*



DISTRIBUCIÓN NORMAL O GAUSSIANA

Fluctuaciones simétricas alrededor de un valor medio (simétrica)

- Variabilidad en atributos humanos o de cosas: peso, altura, notas de los exámenes, etc., dentro de un grupo
- Medidas de errores angulares o lineales
- Generación de ruido y pequeñas perturbaciones

$$\Phi(x)=\begin{cases}\mu=0\\ \sigma=1\end{cases}$$

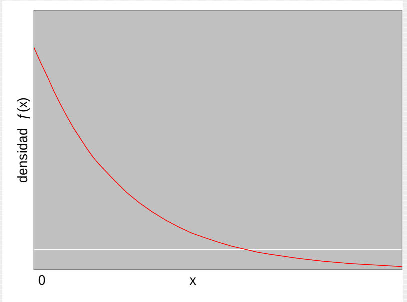
DISTRIBUCIÓN NEGATIVA EXPONENCIAL

Para modelizar procesos independientes

Asimétrica

No tiene memoria

Solo un parámetro (la media)



- Nº de llegadas (de clientes, etc.) o acciones por unidad de tiempo
- Tiempo entre roturas, tamaño de pedidos, tiempos de proceso, etc.

DISTRIBUCIÓN LOGNORMAL

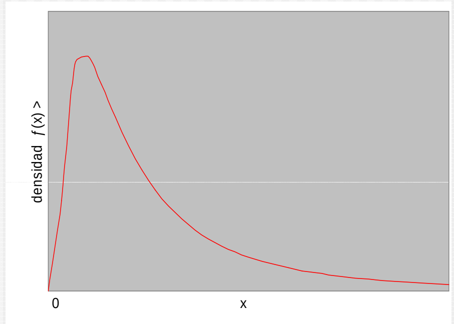
Distribución Normal asimétrica

Tiempos de proceso y de reparación

Una distribución en la cual

$$P[\ln(x)]$$

está normalmente distribuida



Solo para variaciones aleatorias donde los valores medidos son todos positivos (operarios, ingresos, averías)

DISTRIBUCIÓN DE ERLANG

Parámetros: media y un factor K (positivo)

K=1: igual que la distribución exponencial negativa

K -> ∞, se aproxima a una distribución normal

Valores altos de K: cuanto más pequeño es un valor mayor probabilidad tiene de ocurrir.

Por debajo de un cierto límite (entre 0 y la media, dependiendo de K), la frecuencia de ocurrencia baja de nuevo

Asimétrica (si K aumenta, la asimetría disminuye)

DISTRIBUCIÓN GAMMA

La distribución de Erlang con K negativo

La anchura de la distribución de Erlang depende de K:

K=1, la anchura es igual a la media

K -> ∞, la anchura se aproxima a 0 (si la media no cambia)

Tiempos de proceso, tiempos de reparación, tiempos de manipulación, tiempos entre llegadas (con poca aleatoriedad), etc.

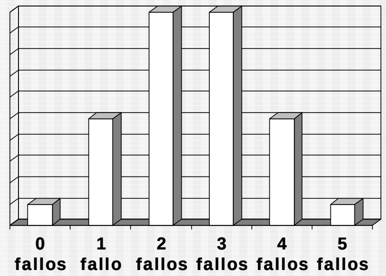
DISTRIBUCIÓN BINOMIAL

MODELA EL NÚMERO TOTAL DE ÉXITOS/FALLOS EN UNA SERIE DE EXPERIMENTOS REPETIDOS

1) Solo dos resultados posibles en cada experimento
2) La probabilidad de éxito en cada intento es constante
3) Los experimentos son independientes entre sí

DISTRIBUCIÓN BINOMIAL

- La probabilidad de que un evento, con probabilidad p, ocurra x veces de n intentos (q=1-p, probabilidad de fallo)
- Media: np
- Varianza: npq
- Se representa como un histograma



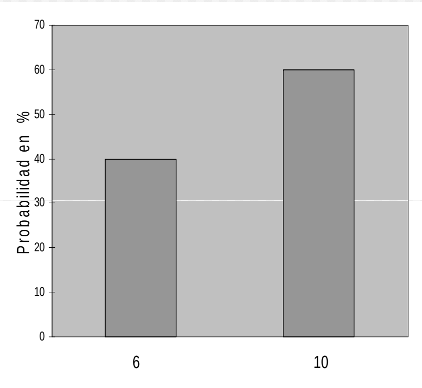
p = ½

q = 1- ½ = ½

$$P(x) = \binom{n}{x} p^x q^{n-x}$$

DISTRIBUCIÓN BERNOULLI

- Parámetros:
  - *probabilidad* en %
  - *resultado1*
  - *resultado2*
- La probabilidad de obtener un *resultado1* es igual a *probabilidad*; *resultado2* tiene un *100-probabilidad (complementarios)*



Bernoulli(40,6,10)

OTRAS DISTRIBUCIONES

- Discretas:
  - Poisson: λ (exponencial)
  - Geométrica
  - Binomial negativa (Pascal)
- Continuas:
  - Beta: [0, 1]
  - Pearson
  - Weibull
  - Chi-Square (χ^2): Gamma b=2

ELECCIÓN DE LA DISTRIBUCIÓN

¿SUCESOS INDEPENDIENTES?

- Si — Exponencial Negativa
- No — ¿Distribución simétrica?
  - Si — ¿Límites superior e inferior fijados?
    - Si — Uniforme
    - No — Normal
  - No — ¿Marcados picos (altos y bajos)?
    - Si — Empírica
    - No — ¿Desviación standar mayor que la media?
      - Si — Lognormal
      - No — Erlang o Lognormal

EJERCICIO

1. Exponencial Negativa 2. Normal 3. Erlang 4. Uniforme 5. Empírica 6. Constante

a. Tiempo de espera después de llamar a un taxi

b. Tiempo requerido para conducir de A a B

c. Tiempo requerido para empaquetar 100 cajas con velas

d. Tiempo entre dos accidentes de coche en un país A

e. Tiempo para coger un producto de una localización aleatoria en un almacén

f. Tiempo entre dos averías de una máquina con muchos elementos sensibles a roturas

g. Tiempo entre dos averías de una máquina con una parte sensible a roturas

h. Tiempo de un AGV (automatically guided vehicle) en recoger un producto

i. Número de clientes en una oficina de correos un lunes