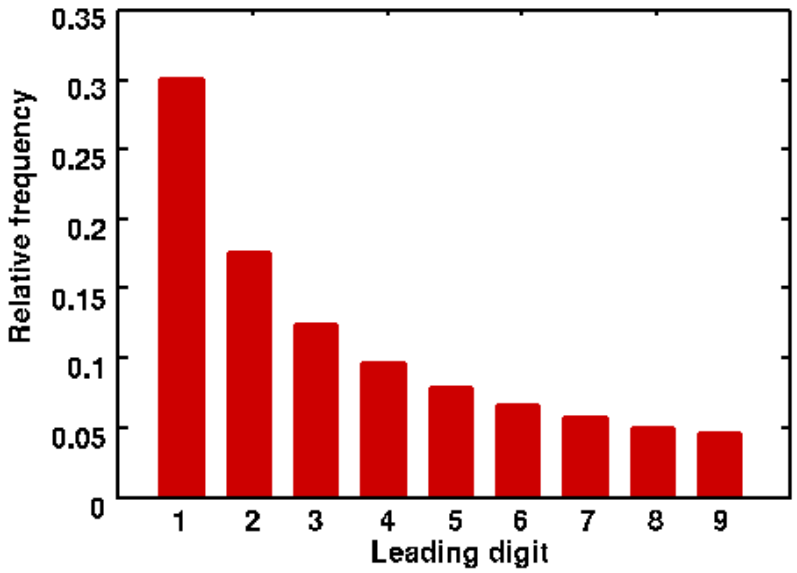
Ley de Benford

CaseWare Analytics

¿Porque el 1 sale más?

$$P(d) = \log_{10}[1 + 1/d],\ d = 1, 2, ..$$



¿No debería de ser cada número igual?

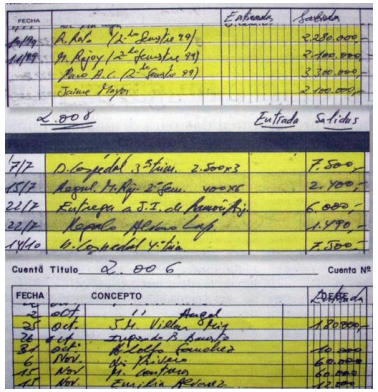
Quizá piensa que P(1) = P(2) = P(3) = ….P(9) = 0.11..
pero…

P(1) = 0.30	P(6) = 0.07
P(2) = 0.18	P(7) = 0.06
P(3) = 0.12	P(8) = 0.05
P(4) = 0.10	P(9) = 0.05
P(5) = 0.08	

Teoría Ley de Benford

$P(d)= \log_{10}[1 + 1/d]$ distribución primer digito
se convierte en
“Ley de Benford”

Resulta más preciso a segunda vista



1° dígito	Bárcenas	Porcentaje	Benford
1	42	50.0%	30.1%
2	8	9.5%	17.6%
3	10	11.9%	12.5%
4	3	3.6%	9.7%
5	3	3.6%	7.9%
6	11	13.1%	6.7%
7	2	2,4%	5.8%
8	0	0.0%	5.1%
9	5	5.9%	4.6%

Probabilidad con etiqueta definida

P(1) = 1/2

P(1) = 1/3

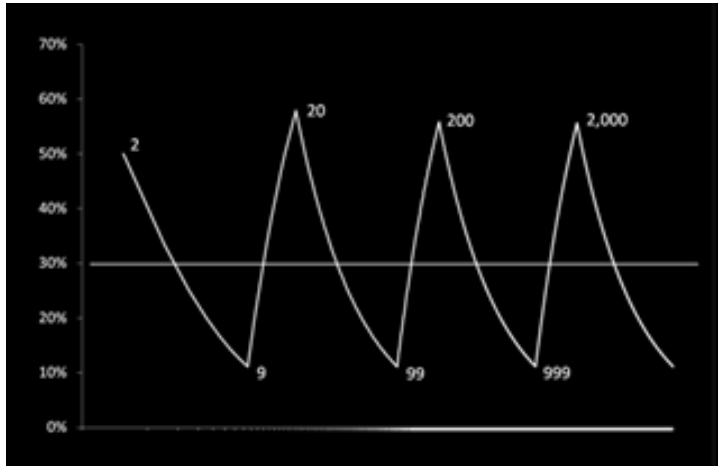
P(1) = 1/9

P(1) = 1/5

Porcentaje de probabilidad

P(1) sube hasta llegar a la etiqueta 19, después baja

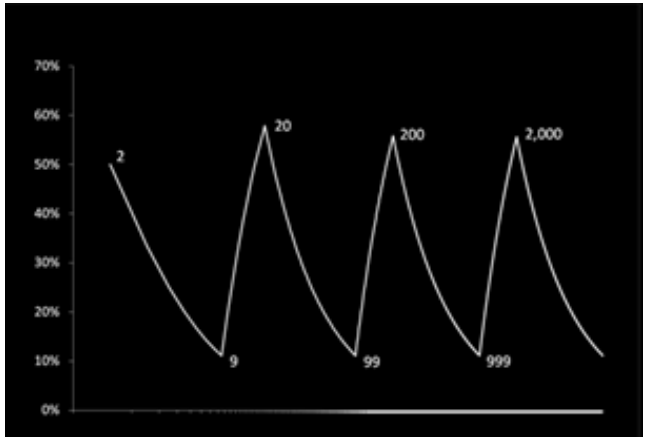
P(1)



Promedio porcentaje infinito

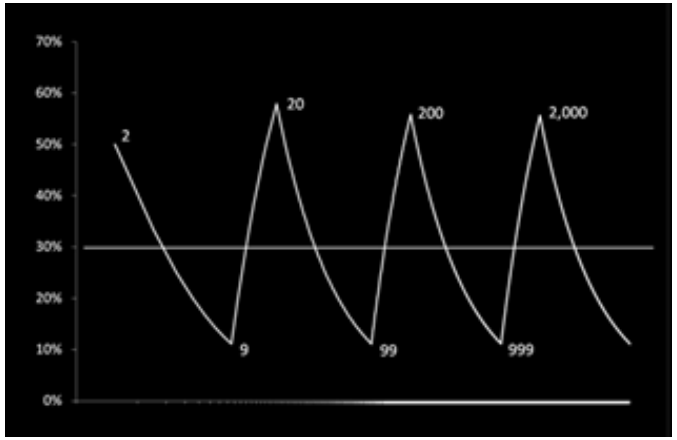
P(1) depende del número de etiquetas

P(1)



Tome el promedio de todos los números posibles de etiquetas

P(1)



←

El promedio es 30.1%

Distribución universal P(x) para los números con unidades

significa que debe ser de escala invariante

$$P(kx) = f(k)P(x)$$

Ya que $\int P(x)dx = 1$ debemos de tener $\int P(kx)dx = 1/k$

asique $1/k = \int P(kx)dx = f(k)\int P(x)dx = f(k)$

Significa $f(k) = 1/k$

d/dk de $P(kx) = f(k)P(x)$

$xdP(kx)/d(kx) \times d(kx)/dk = -P(x)/k^2$

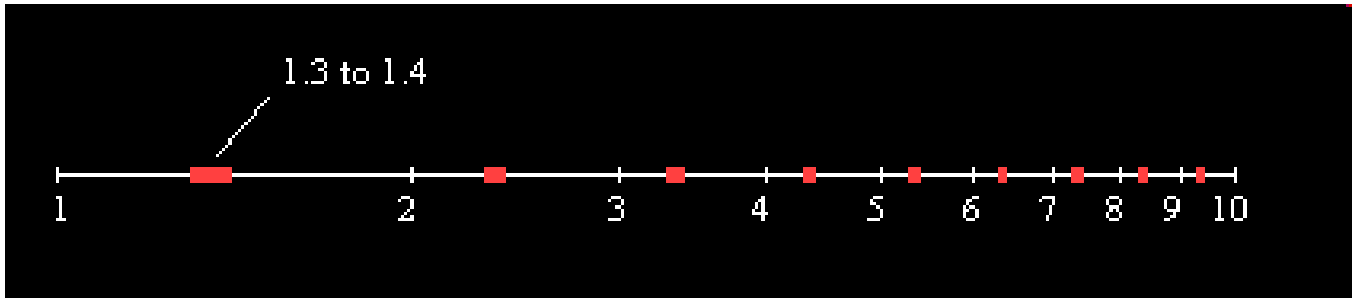
Pone $k = 1$

significa $P(x) = 1/x$

En realidad no se irá a cero o infinito asique no hay que preocuparse de que $\int_0^\infty 1/x\, dx$ sea infinito

$$P(x) = \frac{\int_{x/2}^{2x+1} P(y)\, dy}{\int_{x/2}^{2x} P(y)\, dy} = \lim_{\xi \to 0} \left(1 + \frac{1}{x/2}\right)$$

Por el mismo tipo de análisis se puede determinar la probabilidad de que el *segundo* digito tenga un cierto valor.
Solo hay que tener en cuenta un único orden de la magnitud ya que se repite el patrón en cada orden.
Por ejemplo, en la base 10, la probabilidad del segundo digito sea un "3" es igual que la suma de las probabilidades de "1.3", "2.3", "3.3", ... o "9.3" para los números en el rango de 1 a 10.
Esto indica que son las áreas marcadas de la escala de log aquí abajo:



La fracción en1.4 a 1.3 es $\frac{\ln(1.4)-\ln(1.3)}{\ln(10)-\ln(1)} = \frac{\ln\left(\frac{14}{13}\right)}{\ln(10)}$

Como usar Ley de Benford en la práctica

- Las secuencias de dígitos en una serie de datos siguen un patrón predecible para identificar posibles errores, fraudes potenciales u otras irregularidades. Si existen valores artificiales o ficticios en una base de datos, la distribución de los dígitos dentro de la base de datos puede ser diferente a la establecida al respecto por la Ley de Benford.

- Un análisis de Ley de Benford es más efectivo en:
  - Datos que contienen valores similares para fenómenos similares
  - Datos sin un mínimo y un máximo especificado.
  - Datos sin asignaciones de números, como números de cuentas bancarias y códigos postales
  - Datos que contengan cuatro o más dígitos

¿Sirve para analizar todo?

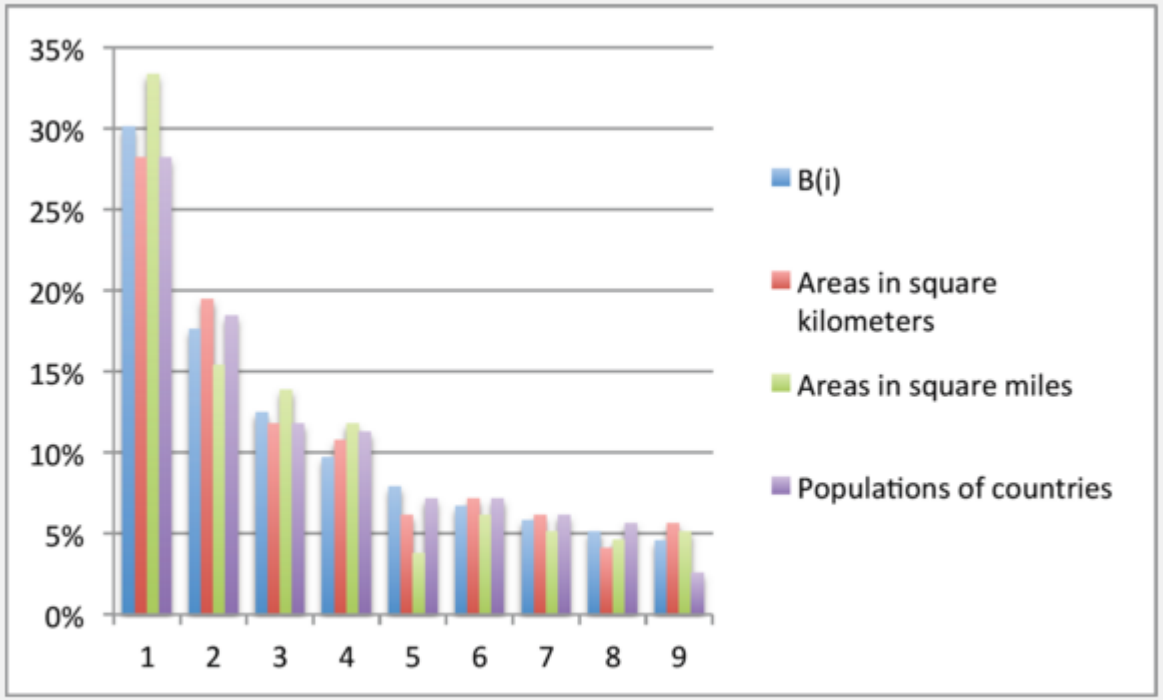


Figura 2: Algunos datos que siguen aproximadamente la ley de Benford: superficies de países en kilómetros cuadrados, áreas de países en millas cuadradas y poblaciones de países.

Ejemplo como detectar Fraude con CAATs

Preguntas

?