Estadística General

Tema 4: Distribución de Probabilidades

Modelos de distribuciones, Bernoulli, Binomial.

Prof. José G. Páez

Algunos modelos de variables aleatorias.

- Hay v.a. que aparecen con frecuencia en las Ciencias de la Salud.
  - Experimentos dicotómicos.
    - Bernoulli

  - Contar éxitos en experimentos dicotómicos repetidos:
    - Binomial
    - Poisson (sucesos raros)

  - Y en otras muchas ocasiones…
    - Distribución normal (gaussiana, campana,…)

- El resto del tema está dedicado a estudiar estas distribuciones especiales.

Distribución de Bernoulli

- Tenemos un experimento de Bernoulli si al realizar un experimentos sólo son posibles dos resultados:
  - X=1 (**éxito**, con probabilidad p)
  - X=0 (**fracaso**, con probabilidad q=1-p)

    - Lanzar una moneda y que salga cara.
      - p=1/2
    - Elegir una persona de la población y que esté enfermo.
      - p=1/1000 = prevalencia de la enfermedad
    - Aplicar un tratamiento a un enfermo y que éste se cure.
      - p=95%, probabilidad de que el individuo se cure

- Como se aprecia, en experimentos donde el resultado es dicotómico, la variable queda perfectamente determinada conociendo el parámetro **p**.

Ejemplo de distribución de Bernoulli.

- Se ha observado estudiando 2000 accidentes de tráfico con impacto frontal y cuyos conductores no tenían cinturón de seguridad, que 300 individuos quedaron con secuelas. Describa el experimento usando conceptos de v.a.

- Solución.
  - La noc. frecuentista de prob. nos permite aproximar la probabilidad de tener secuelas mediante 300/2000=0,15=15%

  - X="tener secuelas tras accidente sin cinturón" es variable de Bernoulli
    - X=1 tiene probabilidad $p \approx 0{,}15$
    - X=0 tiene probabilidad $q \approx 0{,}85$

Ejemplo de distribución de Bernoulli.

- Se ha observado estudiando 2000 accidentes de tráfico con impacto frontal y cuyos conductores sí tenían cinturón de seguridad, que 10 individuos quedaron con secuelas. Describa el experimento usando conceptos de v.a.

- Solución.
  - La noc. frecuentista de prob. nos permite aproximar la probabilidad de quedar con secuelas por 10/2000=0,005=0,5%
  - X="tener secuelas tras accidente usando cinturón" es variable de Bernoulli
    - X=1 tiene probabilidad $p \approx 0{,}005$
    - X=0 tiene probabilidad $q \approx 0{,}995$

Observación

- En los dos ejemplos anteriores hemos visto cómo enunciar los resultados de un experimento en forma de estimación de parámetros en distribuciones de Bernoulli.
    - Sin cinturón: $p \approx 15\%$
    - Con cinturón: $p \approx 0,5\%$

- En realidad no sabemos en este punto si ambas cantidades son muy diferentes o aproximadamente iguales, pues en otros estudios sobre accidentes, las cantidades de individuos con secuelas hubieran sido con seguridad diferentes.

- Para decidir si entre ambas cantidades existen diferencias estadísticamente significativas necesitamos introducir conceptos de estadística inferencial (extrapolar resultados de una muestra a toda la población).

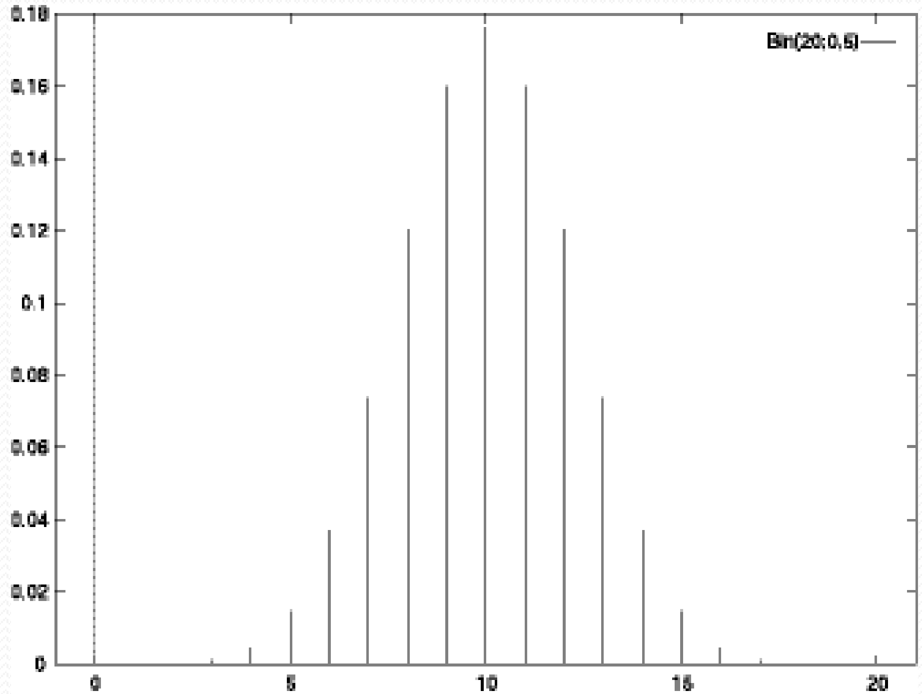
- Es muy pronto para resolver esta cuestión ahora. Esperemos a las pruebas de X^2.

Distribución binomial

- Función de probabilidad

$$P[X=k]=\binom{n}{k}p^k q^{n-k},\ 0\le k\le n$$

  - Problemas de cálculo si n es grande y/o p cercano a 0 o 1.



- Media: $\mu = n\,p$

- Varianza: $\sigma^2 = n\,p\,q$

Distribución Binomial

- Si se repite un número fijo de veces, **n**, un experimento de Bernoulli con parámetro **p**, el número de éxitos sigue una distribución **binomial** de parámetros (n,p).

  - Lanzar una moneda 10 veces y contar las caras.
    - Bin(n=10,p=1/2)
  - Lanzar una moneda 100 veces y contar las caras.
    - Bin(n=100,p=1/2)
    - Difícil hacer cálculos con esas cantidades. El modelo normal será más adecuado.
  - El número de personas que enfermará (en una población de 500.000 personas) de una enfermedad que desarrolla una de cada 2000 personas.
    - Bin(n=500.000, p=1/2000)
      - Difícil hacer cálculos con esas cantidades. El modelo de Poisson será más adecuado.

"Parecidos razonables"

- Aún no conoces la distribución normal, ni de Poisson.

- De cualquier forma ahí tienes la comparación entre valores de p no muy extremos y una normal de misma media y desviación típica, para tamaños de n grandes (n>30).

- Cuando p es muy pequeño es mejor usar la aproximación del modelo de Poisson.