Tema 5

Distribuciones de probabilidad más usuales

En este tema se estudiarán algunas de las distribuciones discretas y continuas más comunes, que se pueden aplicar a una gran diversidad de problemas y campos científicos.

Contenido

5.1. Distribuciones discretas más usuales . . . 19
- 5.1.1. Distribución de Bernoulli . . . 19
- 5.1.2. Distribución binomial . . . 20
- 5.1.3. Distribución geométrica . . . 21
- 5.1.4. Distribución de Poisson . . . 21

5.2. Distribuciones continuas más usuales . . . 22
- 5.2.1. Distribución uniforme . . . 22
- 5.2.2. Distribución normal . . . 23
- 5.2.3. Distribución exponencial . . . 23

5.3. Teorema del Límite Central . . . 25

Comentarios bibliográficos . . . 25

Ejercicios . . . 26

5.1. Distribuciones discretas más usuales

5.1.1. Distribución de Bernoulli

Un experimento aleatorio se dice que es de Bernoulli cuando únicamente puede tener dos resultados mutuamente excluyentes; uno de ellos se denomina “éxito” y el otro “fracaso”.

Ejemplos:

- Los resultados “cara” o “cruz” en el lanzamiento de una moneda.
- Las piezas “defectuosa” o “no defectuosa” en el control de calidad de un producto.

- Resultado "exitoso" o "fallido" de la petición a un servidor.

Sea X una v. a. asociada a un experimento de Bernoulli y que toma los valores:

$$X(\text{éxito}) = 1 \quad X(\text{fracaso}) = 0$$

entonces se dice que X sigue una distribución de Bernoulli $X \equiv B(1, p)$. Su función de probabilidad viene dada por:

$$\Pr(X = 1) = p \qquad \Pr(X = 0) = 1 - p = q$$

Propiedades:

1. $\mu = E(X) = 0 \times q + 1 \times p = p$.
2. $\sigma^2 = \text{Var}(X) = E(X^2) - E^2(X) = p - p^2 = pq$.

5.1.2. Distribución binomial

Una sucesión de n pruebas se dice que es de Bernoulli cuando los experimentos individuales verifican las siguientes condiciones:

1. Las n pruebas son independientes.
2. Cada prueba es de Bernoulli.
3. La probabilidad p de éxito es igual en todas las pruebas.

La variable aleatoria definida como "número de éxitos en n pruebas", $X \equiv B(n, p)$, se dice que sigue una distribución binomial de parámetros n, p.

La variable puede tomar los valores $\{0, 1, 2, \ldots, k, \ldots, n\}$ y su función de probabilidad es la siguiente:

$$\Pr(X = k) = \Pr(k \text{ éxitos en } n \text{ pruebas}) = \binom{n}{k} p^k q^{n-k}$$

donde

$$\binom{n}{k} = \text{número resultados posibles con } k \text{ éxitos}$$

$$p^k q^{n-k} = P(\text{cada resultado con } k \text{ éxitos})$$

Ejemplos:

- Número de "veces" que aparece el resultado cara al lanzar una moneda diez veces.
- Número de éxitos en la recepción de un mensaje enviado a 100 destinatarios.
- Número de ordenadores en una subred que han sido infectados por un virus.

Propiedades:

1. $X = X_1 + X_2 + \cdots + X_n$, donde $X_1, \ldots, X_n$ son variables independientes todas ellas con distribución $B(1, p)$.

2. $\mu = E(X) = np$.

3. $\sigma^2 = \mathrm{Var}(X) = n\,p\,q$.

4. Dadas $X_1 \equiv B(n_1, p)$ y $X_2 \equiv B(n_2, p)$ v.a. independientes, entonces $X_1 + X_2 \equiv B(n_1 + n_2, p)$.

5.1.3. Distribución geométrica

Sea X la variable aleatoria definida como el número de pruebas realizadas hasta que aparece por primera vez el resultado éxito, en pruebas de Bernoulli; entonces se dice que $X \equiv G(p)$ sigue una distribución geométrica de parámetro p.

Los valores que puede tomar esta variable son $\{1, 2, \ldots, k, \ldots\}$ con probabilidades:

$$\Pr(X = k) = \Pr(F_1 \cap F_2 \cap \ldots \cap F_k \cap E) = q^{k-1}p$$

Ejemplos:

- Número de veces que ha sido necesario ejecutar un programa hasta que falla por primera vez.
- Número de veces que ha sido necesario enviar un mensaje hasta que fue recibido por el destinatario.

Propiedades:

1. $\mu = E(X) = \dfrac{1}{p} \qquad \text{si } p > 0.$

2. $\sigma^2 = \mathrm{Var}(X) = \dfrac{q}{p^2}$.

Una forma alternativa de representar este experimento es considerar la variable Y que cuenta el número de fracasos obtenidos hasta que aparece el primer éxito; es evidente que la relación entre las dos variables viene dada por la expresión $X = Y + 1$. La esperanza de Y es $E(Y) = \dfrac{q}{p}$. La varianza, obviamente, es la misma.

5.1.4. Distribución de Poisson

La distribución de Poisson suele emplearse para representar experimentos en los que se analiza el número de veces que ocurre cierto suceso en un intervalo (en general de tiempo). Los requisitos que deben verificar estas experiencias son las siguientes:

1. Las condiciones experimentales deben ser constantes a lo largo de todo el intervalo.

2. Los resultados del experimento deben ser independientes cuando se refieren a intervalos disjuntos.

3. La tasa media de aparición del suceso, en cualquier intervalo de longitud uno, es constante y se representa por λ.

4. La probabilidad de que el suceso ocurra una sola vez en un intervalo de amplitud h suficientemente pequeña, debe ser aproximadamente $\lambda\, h$.

5. La probabilidad de dos o más ocurrencias del suceso, en un intervalo suficientemente pequeño, debe ser prácticamente cero.

Bajo las anteriores condiciones la variable X = número de veces que ocurre el suceso, se dice que sigue una distribución de Poisson de parámetro λ, $X \equiv P(\lambda)$. Los valores de la variable son $\{0, 1, 2, \ldots k \ldots\}$ con probabilidades:

$$\Pr(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad \text{si } \ k = 0, 1, 2, \ldots$$

Ejemplos:

- El número de partículas emitidas por una substancia radiactiva en una hora.
- El número de mensajes que llegan a un servidor de correo durante una hora.

Propiedades:

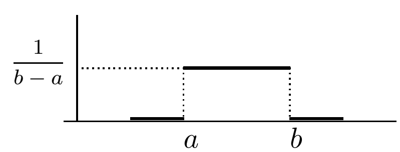
1. $\mu = E(X) = \lambda$.
2. $\sigma^2 = \mathrm{Var}(X) = \lambda$.
3. Sean $X_1 \equiv P(\lambda_1)$ y $X_2 \equiv P(\lambda_2)$ v.a. independientes, entonces $X_1 + X_2 \equiv P(\lambda_1 + \lambda_2)$.

5.2. Distribuciones continuas más usuales

5.2.1. Distribución uniforme

Una variable aleatoria X se dice que sigue una distribución uniforme en [a, b], donde $a \leq b$ son números reales, $X \equiv U(a, b)$, si tiene la siguiente función de densidad:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } a \leq x \leq b, \\ 0 & \text{en otro caso.} \end{cases}$$



Esta distribución se puede emplear en problemas en los que la probabilidad se reparte por igual en todo el intervalo.

Propiedades:

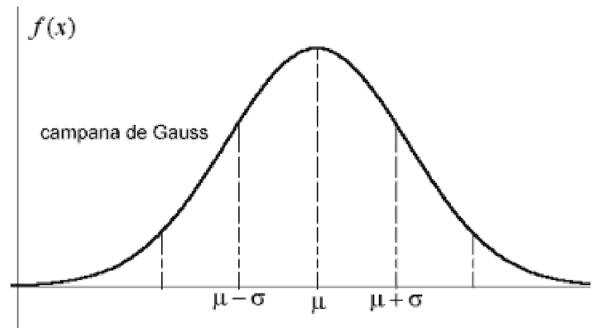
1. $\mu = E(X) = \dfrac{b+a}{2}$ (punto medio del intervalo $[a, b]$).
2. $\sigma^2 = \mathrm{Var}(X) = \dfrac{(b-a)^2}{12}$.

5.2.2. Distribución normal

Es la más importante de las distribuciones continuas ya que permite describir un número muy grande de fenómenos aleatorios, como por ejemplo aquellos en los que intervienen un número elevado de factores no controlables, que actúan de manera independiente y con efectos pequeños.

Una v.a. se dice que sigue una distribución normal $X \equiv N(\mu;\sigma)$, si su función de densidad es:

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



Propiedades:

1. $E(X) = \mu$.
2. $\mathrm{Var}(X) = \sigma^2$.
3. Si $X_1 \equiv N(\mu_1;\sigma_1)$ y $X_2 \equiv N(\mu_2;\sigma_2)$ son v.a. independientes, entonces se verifica que: $X = X_1 + X_2 \equiv N(\mu_1+\mu_2; \sqrt{\sigma_1^2+\sigma_2^2})$.
4. Sea $X \equiv N(\mu;\sigma)$ entonces $Y = a + bX$, con $b \neq 0$, sigue una distribución normal $N(a+b\mu; |b|\sigma)$.
5. La función de distribución de cualquier distribución $N(\mu;\sigma)$ se puede calcular utilizando la distribución $N(0;1)$ o normal tipificada.
$$P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P(Z \leq z)$$
donde $Z = \dfrac{X-\mu}{\sigma}$ es la normal tipificada y $z = \dfrac{x-\mu}{\sigma}$ se denomina puntuación tipificada.
6. La función de densidad de la normal no admite primitiva y su función de distribución se calcula de manera aproximada. En lugar de tener que realizar esos cálculos aproximados para cada par de valores μ y σ, la propiedad anterior permite emplear las tablas de la $N(0;1)$ para obtener la función de distribución de cualquier normal.

5.2.3. Distribución exponencial

En un experimento de Poisson en el que se observa la ocurrencia de un suceso E en un intervalo de tiempo, donde $\lambda > 0$ representa el número medio de sucesos que ocurren por unidad de tiempo, puede ser de interés el tiempo que transcurre entre dos sucesos consecutivos.

En este caso, la v.a. $T =$ "tiempo entre dos sucesos consecutivos" es continua y su distribución se calcula de la siguiente manera:

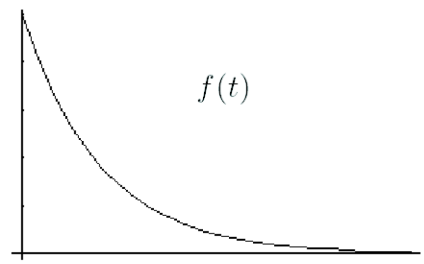
$$P(T \leq t) = 1 - P(T > t) = 1 - P(X = 0) = 1 - e^{-\lambda t}$$

ya que el suceso $(T > t)$ significa que en el intervalo $(0, t]$ no apareció en ninguna ocasión el suceso E. Se dice entonces que T tiene distribución exponencial de parámetro λ, que se denotará por $T \equiv Exp(\lambda)$ y su función de distribución viene dada por:

$$F_T(t) = \begin{cases} 1 - e^{-\lambda t} & \text{si } t > 0, \\ 0 & \text{en otro caso,} \end{cases}$$

y su función de densidad es:

$$f(t) = \begin{cases} \lambda\, e^{-\lambda t} & \text{si } t > 0, \\ 0 & \text{en otro caso.} \end{cases}$$



Ejemplo:

- Tiempo que tarda una cierta cantidad de una substancia radiactiva en reducir su masa a la mitad.
- Tiempo transcurrido entre la llegada de dos clientes consecutivos a una tienda.

Propiedades:

1. $E(T) = \dfrac{1}{\lambda}$.
2. $\text{Var}(T) = \dfrac{1}{\lambda^2}$.
3. $\Pr[T > (t + h) \mid T \geq t] = \Pr[T > h]$.

   Esta probabilidad solo depende de la amplitud del intervalo $[t, t + h)$, pero no del valor concreto de t y nos referiremos a ella diciendo que la distribución exponencial "no tiene memoria".

   Para comprender el significado intuitivo de esta propiedad supongamos que estamos analizando una variable que representa el tiempo transcurrido hasta que un "sistema falla", como por ejemplo el tiempo de semi-desintegración de una materia radiactiva o la vida de una persona. Si esta variable se ajustase a una distribución exponencial, significaría que la probabilidad de que el "sistema" dure diez años más, es independiente de lo antiguo que sea. Esto puede ser cierto para una materia radiactiva, pero suele resultar falso para una persona.

5.3. Teorema del Límite Central

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas tales que $E(X_i) = \mu$ y $\mathrm{Var}(X_i) = \sigma^2$ para todo i. Entonces,

$$S_n = X_1 + \cdots + X_n \overset{\text{aprox.}}{\to} N(\mu n, \sigma\sqrt{n}).$$

o equivalentemente

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n} \overset{\text{aprox.}}{\to} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

La aproximación se considera válida para valores grandes de n. Es importante recordar que según este resultado la suma de un número suficientemente grande de variables independientes del mismo tipo sigue una distribución normal, incluso en el caso de que se desconozca la distribución que siguen los sumandos X_i.

Como consecuencia del Teorema del Límite Central, se obtienen las aproximaciones de las distribuciones binomial y de Poisson por la distribución normal.

- Si $X \equiv B(n,p)$ entonces $X = \sum_{i=1}^{n} X_i$ donde cada $X_i \equiv B(1,p)$. Por tanto, el TLC permite aproximar la distribución de esta variable por la distribución $N(np; \sqrt{npq})$ cuando n es suficientemente grande y p un valor ni próximo a 0 ni próximo a 1. En general se suele trabajar con la aproximación para $n \geq 30$ y $0{,}1 < p < 0{,}9$.

- Teniendo en cuenta la propiedad de reproductividad de la variable de Poisson, el TLC permite aproximar la distribución de Poisson de parámetro λ mediante la distribución $N(\lambda, \sqrt{\lambda})$, para valores grandes de λ. En la práctica, se considera que si $\lambda > 5$ la aproximación a la normal es buena.

Comentarios bibliográficos

El bloque correspondiente al Cálculo de probabilidades, formado por los Temas 3, 4 y 5 del programa de la asignatura es desarrollado en mayor o menor profundidad por la mayoría de los textos de estadística recomendados en la bibliografía. Sirvan como ejemplo los siguientes:

- Temas 3, 4 y 6 de **Introducción a la Estadística y sus aplicaciones**. CAO ABAD, R. *et al. Ediciones Pirámide*, 2001.

- Tema 2, 3 y 4 de **Probabilidad y Estadística para Ingeniería y Ciencias**. DEVORE, J.L. *International Thomson Editores*, 2002.

- Temas 4 y 5 de **Fundamentos de Estadística**. PEÑA SÁNCHEZ DE RIVERA, D. *Alianza Editorial*, 2001.

- Temas 2, 3 y 4 de **Probabilidad y Estadística para Ingeniería**. SCHEAFFER R.L. & McCLAVE J. *Grupo Editorial Iberoamericano*, 1993.