<u>Unidad IV</u>

Distribuciones de Probabilidad Continuas

4.1. Definición de variable aleatoria continúa.

Una variable aleatoria *X* es continua si su función de distribución es una función continua.

En la práctica, se corresponden con variables asociadas con experimentos en los cuales la variable medida puede tomar cualquier valor en un intervalo: mediciones biométricas, intervalos de tiempo, áreas, etc.

Ejemplos

- **Resultado de un generador de números aleatorios entre 0 y 1.** Es el ejemplo más sencillo que podemos considerar, es un caso particular de una familia de variables aleatorias que tienen una distribución uniforme en un intervalo [*a*, *b*]. Se corresponde con la elección al azar de cualquier valor entre *a* y *b*.
- **Estatura de una persona elegida al azar en una población.** El valor que se obtenga será una medición en cualquier unidad de longitud (m, cm, etc.) dentro de unos límites condicionados por la naturaleza de la variable. El resultado es impredecible con antelación, pero existen intervalos de valores más probables que otros debido a la distribución de alturas en la población. Más adelante veremos que, generalmente, variables biométricas como la altura se adaptan un modelo de distribución denominado *distribución Normal* y representado por una campana de Gauss.

Dentro de las variables aleatorias continuas tenemos las **variables aleatorias absolutamente continuas**.

Diremos que una variable aleatoria *X* continua tiene una distribución absolutamente continua si existe una función real *f*, positiva e integrable en el conjunto de números reales, tal que la función de distribución *F* de *X* se puede expresar como

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

Una variable aleatoria con distribución absolutamente continua, por extensión, se clasifica como variable aleatoria absolutamente continua.

En el presente manual, **todas las variables aleatorias continuas con las que trabajemos pertenecen al grupo de las variables absolutamente continuas**, en particular, los ejemplos y casos expuestos.

4.2. Función de densidad y acumulativa.

Una función que ofrece la probabilidad total de obtener un resultado de unavariable aleatoria que varía desde el valor más bajo posible para la variable aleatoria hasta cualquier valor específico de interés. Por ejemplo, podemos preguntar cuál es la probabilidad de que algún tipo de interés (tal como elLIBOR a seis meses) sea del 5% o menos dentro de un año a partir de hoy. Las funciones de densidad acumulativa se derivan de las funciones de densidad deprobabilidad. También se conoce como función de probabilidad acumulativa.

4.3. Valor esperado, varianza y desviación estándar.

Esperanza o media (v.a. continua)

Consideremos una v.a continua *X*.
Llamaremos media o esperanza de *X* a

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

donde *x* son los valores que toma la variable y *f(x)* es la función de densidad

Propiedades:

1. Dada una variable aleatoria continua que toma siempre el mismo valor C, es decir, la variable es constante. Su esperanza es esa misma constante.
   $E[C] = \int_{-\infty}^{\infty} Cf(x)dx = C\int_{-\infty}^{\infty} f(x)dx = C$ ya que $\int_{-\infty}^{\infty} f(x)dx = 1$
2. Si se multiplica una variable aleatoria por una constante, su esperanza se ve multiplicada por esa constante.
   $E[CX] = \int_{-\infty}^{\infty} C \cdot x_i \cdot f(x)dx = C\int_{-\infty}^{\infty} x_i \cdot f(x)dx = C \cdot E[X]$

3. $E[X - E[X]] = 0$, veámoslo:

$$E\left[X - E\left[X\right]\right] = \int_{-\infty}^{\infty} \left(x_i - E\left[X\right]\right) f(x)dx = \int_{-\infty}^{\infty} x_i \cdot f(x)dx - \int_{-\infty}^{\infty} E\left[X\right] f(x)dx =$$

$$\int_{-\infty}^{\infty} x_i \cdot f(x)dx - E\left[X\right]\int_{-\infty}^{\infty} f(x)dx = E\left[X\right] - E\left[X\right] = 0$$

4. Sean *X* e *Y* dos variables aleatorias, *E[X+Y]=E[X]+E[Y]*
5. En general, *E[aX+b]=aE[X]+b*

Ejemplos:

1. El peso (en gramos) de un insecto se distribuye según

$$f(x) = \begin{cases} 1 & si \;\; 1 \le x \le 2 \\ 0 & resto \end{cases}$$

Calcula el peso esperado

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{-\infty}^{1} x \cdot f(x)dx + \int_{1}^{2} x \cdot f(x)dx + \int_{2}^{\infty} x \cdot f(x)dx = \int_{1}^{2} x\, dx = \left[\frac{x^2}{2}\right]_1^2 = 2 - \frac{1}{2} = \frac{3}{2}$$

2. El tiempo de vida (en años) de una determinada componente de un juguete electrónico tiene por función de densidad

$$f(x) = \begin{cases} -6x(x-1) & si \;\; 0 \le x \le 1 \\ 0 & resto \end{cases}$$

El tiempo de vida esperado será

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{0}^{1} x\left(-6x(x-1)\right)dx = \int_{0}^{1} \left(-6x^3 + 6x^2\right)dx = \left[-\frac{3}{2}x^4 + 2x^3\right]_0^1 = 2 - \frac{3}{2} = \frac{1}{2}$$

La duración esperada de la componente es de medio año, es decir, 6 meses.

Varianza (v.a. continua)

Se define la varianza como la esperanza del cuadrado de la diferencia entre la variable y su esperanza:

V(X) = E[(X - E[X])^2], teniendo en cuenta la definición de esperanza, la varianza queda

$$V[X] = E\left[\left(X - E[X]\right)^2\right] = \int_{-\infty}^{\infty} \left(X - E[X]\right)^2 f(x)dx$$

Desarrollando el cuadrado llegamos a una expresión de la varianza más cómoda para operar que la definición

$$V[X] = E[X^2] - E[X]^2 = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - \left(\int_{-\infty}^{\infty} x \cdot f(x)dx \right)^2$$

Desviación típica
Se llama desviación típica o estándar a la raíz cuadrada positiva de la varianza
$\sigma = \sqrt{V(X)}$
Para determinar la varianza hemos elevado los datos al cuadrado, esto implica que las unidades se elevan al cuadrado. La desviación típica consigue tener las desviaciones en la misma unidad que los datos.

Ejemplos:

1. El peso (en gramos) de un insecto se distribuye según

$$f(x) = \begin{cases} 1 & si\ 1 \le x \le 2 \\ 0 & resto \end{cases}$$

Determínese la varianza de *X*
Para determinar la varianza primero calculamos *E[X]* y *E[X²]*

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{-\infty}^{1} x \cdot f(x)dx + \int_{1}^{2} x \cdot f(x)dx + \int_{2}^{\infty} x \cdot f(x)dx =$$

$$\int_{1}^{2} x \cdot dx = \left[\frac{x^2}{2} \right]_1^2 = 2 - \frac{1}{2} = \frac{3}{2}$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_{-\infty}^{1} x^2 \cdot f(x)dx + \int_{1}^{2} x^2 \cdot f(x)dx + \int_{2}^{\infty} x^2 \cdot f(x)dx =$$

$$\int_{1}^{2} x^2 \cdot dx = \left[\frac{x^3}{3} \right]_1^2 = \frac{8}{3} - \frac{1}{3} = \frac{7}{3}$$

V(X)= E[X²] - E[X]²$= \frac{7}{3} - \left(\frac{3}{2} \right)^2 = \frac{7}{3} - \frac{9}{4} = \frac{28-27}{12} = \frac{1}{12}$

2. El tiempo de vida (en años) de una determinada componente de un juguete electrónico tiene por función de densidad

$$f(x) = \begin{cases} -6x(x-1) & si\ 0 \le x \le 1 \\ 0 & resto \end{cases}$$

Halla su varianza
Razonamos de forma análoga al caso anterior

-

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^1 x \cdot (-6x(x-1))dx = \int_0^1 \left(-6x^3 + 6x^2\right)dx =$$

$$\left[-\frac{3}{2}x^4 + 2x^3\right]_0^1 = 2 - \frac{3}{2} = \frac{1}{2}$$

-

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_0^1 x^2 \cdot (-6x(x-1))dx = \int_0^1 \left(-6x^4 + 6x^3\right)dx =$$

$$\left[-\frac{6}{5}x^5 + \frac{3}{2}x^4\right]_0^1 = -\frac{6}{5} + \frac{3}{2} = \frac{-12+15}{10} = \frac{3}{10}$$

V(X)= E[X ^{2}] - E[X] ^{2}= $\frac{3}{10} - \left(\frac{1}{2}\right)^2 = \frac{3}{10} - \frac{1}{4} = \frac{6-5}{20} = \frac{1}{20}$

Variable continua (continuos variable)

Variable que no tiene un número fijo de valores. La variable "ingresos" expresada en pesos, por ejemplo, puede tomar muchos valores diferentes.

Gráficos para variables cuantitativas

Para las variables cuantitativas, consideraremos dos tipos de gráficos, en función de que para realizarlos se usen las frecuencias (absolutas o relativas) o las frecuencias acumuladas:

Diagramas diferenciales:

> Son aquellos en los que se representan frecuencias absolutas o relativas. En ellos se representa el número o porcentaje de elementos que presenta una modalidad dada.

Diagramas integrales:

Son aquellos en los que se representan el número de elementos que presentan una modalidad inferior o igual a una dada. Se realizan a partir de las frecuencias acumuladas, lo que da lugar a gráficos crecientes, y es obvio que este tipo de gráficos no tiene sentido para variables cualitativas.

Según hemos visto existen dos tipos de variables cuantitativas: discretas y continuas. Vemos a continuación las diferentes representaciones gráficas que pueden realizarse para cada una de ellas así como los nombres específicos que reciben.

Gráficos para variables discretas

Cuando representamos una variable discreta, usamos el **diagrama de barras** cuando pretendemos hacer una gráfica diferencial. Las barras deben ser estrechas para representar el que los valores que toma la variable son discretos. El diagrama integral o acumulado tiene, por la naturaleza de la variable, forma de escalera. Un ejemplo de diagrama de barras así como su diagrama integral correspondiente están representados en la figura 1.6.

Ejemplo

Se lanzan tres monedas al aire en 8 ocasiones y se contabiliza el número de caras, *X*, obteniendose los siguientes resultados:

$$X \leadsto 2, 1, 0, 1, 3, 2, 1, 2$$

Representar gráficamente el resultado.

Solución: En primer lugar observamos que la variable *X* es cuantitativa discreta, presentando las modalidades:

$$X \in 0, 1, 2, 3$$

4.4. Distribución Uniforme (continua)

En teoría de probabilidad y estadística, la **distribución uniforme continua** es una familia de distribuciones de probabilidad para variables aleatorias continuas, tales que cada miembro de la familia, todos los intervalos de igual longitud en la distribución en su rango son igualmente probables. El dominio está definido por dos parámetros, *a* y *b*, que son sus valores mínimo y máximo. La distribución es a menudo escrita en forma abreviada como *U*(*a*,*b*).

4.5 Distribución Exponencial.

En estadística la **distribución exponencial** es una distribución de probabilidad continua con un parámetro $\lambda > 0$ cuya función de densidad es:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{para } x \geq 0 \\ 0 & \text{de otro modo} \end{cases}$$

Su función de distribución acumulada es:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{para } x < 0 \\ 1 - e^{-\lambda x} & \text{para } x \geq 0 \end{cases}$$

Donde e representa el número e.

El valor esperado y la varianza de una variable aleatoria X con distribución exponencial son:

$$E[X] = \frac{1}{\lambda}, \qquad V(X) = \frac{1}{\lambda^2}$$

La distribución exponencial es un caso particular de distribución gamma con $k = 1$. Además la suma de variables aleatorias que siguen una misma distribución exponencial es una variable aleatoria expresable en términos de la distribución gamma.

Índice

Ejemplo

Ejemplos para la distribución exponencial es la distribución de la longitud de los intervalos de variable continua que transcuren entre la ocurrencia de dos sucesos, que se distribuyen según la distribución de Poisson.

Calcular variables aleatorias

Se pueden calcular una variable aleatoria de distribución exponencial x por medio de una variable aleatoria de distribución uniforme $u = U(0,1)$:

$$x = -\frac{1}{\lambda} ln(1-u)$$

o, dado que $(1-u)$ es también una variable aleatoria con distribución $U(0,1)$, puede utilizarse la versión más eficiente:

$$x = -\frac{1}{\lambda} ln(u)$$

4.6 Distribución Gamma (Erlang)

En estadística, la **distribución Erlang**, es una distribución de probabilidad continua con dos parámetros k y λ cuya función de densidad para valores $x > 0$ es

$$f(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{k-1}}{(k-1)!} \quad \text{para } x, \lambda \geq 0.$$

La distribución Erlang es el equivalente de la distribución gamma con el parámetro $k = 1, 2 \ldots$ y $\lambda = 1/\theta$. Para $k = 1$eso es la distribución exponencial. Se utiliza la distribución Erlang para describir el tiempo de espera hasta el suceso número $k$en un proceso de Poisson.

Su esperanza viene dada por: $E(X) = k/\lambda$

Su varianza viene dada por: $V(X) = k/\lambda^2$

La función generadora de momentos responde a la expresión: $(1 - t/\lambda)^{-k}$

4.7. Distribución Normal.

Es la distribución de probabilidad más importante, que corresponde a una variable continua. También se la llama distribución gaussiana.

En esta distribución no es posible calcular la probabilidad de un valor exacto, siempre se trabaja con rangos.

En la práctica, muchas variables que se observan tienen distribuciones que sólo se aproximan a la normal. Esto es, las variables tienen propiedades que sólo se acercan a las propiedades teóricas de la distribución normal.

4.7.1 Aproximación de la Binomial a la Normal.

En este caso se estarán calculando probabilidades de experimentos Binomiales de una forma muy aproximada con la distribución Normal, esto puede llevarse a cabo si *n*¥® y *p* = p(éxito) no es muy cercana a *0 y 1*, o cuando *n* es pequeño y *p* tiene un valor muy cercano a ½ ; esto es,

$$P(x,n,p) = {}_nC_x p^x q^{n-x} \cong p\left(z = \frac{x-np}{\sqrt{npq}}\right)$$

Donde:
x = variable de tipo discreto; solo toma valores enteros
m = *np* = media de la distribución Binomial
s = $\sqrt{npq}$ = desviación estándar de la distribución Binomial

Cuando ocurren las condiciones anteriores, la gráfica de la distribución Binomial, es muy parecida a la distribución Normal, por lo que es adecuado calcular probabilidades con la Normal en lugar de con la Binomial y de una forma más rápida.
En resumen, se utiliza la aproximación Normal para evaluar probabilidades Binomiales siempre que *p* no esté cercano a *0 o 1*. La aproximación es excelente cuando *n* es grande y bastante buena para valores pequeños de *n* si *p* está razonablemente cercana a ½. Una posible guía para determinar cuando puede utilizarse la aproximación Normal es tener en cuenta el cálculo de *np* y *nq*. Sí ambos, *np y nq* son *mayores o iguales a 5*, la aproximación será buena.

Antes de empezar a resolver problemas con la aproximación Normal, es bueno aclarar que se están evaluando probabilidades asociadas a una variable

discreta x, con una distribución que evalúa variables de tipo continuo como es la Normal,
Por lo que z sufre un pequeño cambio como se muestra a continuación:

$$z = \frac{(x \pm 1/2) - \mu}{\sigma}$$

4.8. Teorema de Chebyshev.

Si una variable aleatoria tiene una varianza o desviación estándar pequeña,esperaríamos que la mayoría de los valores se agrupan alrededor de la media. Por lo tanto, la probabilidad de que una variable aleatoria tome un valor dentro decierto intervalo alrededor de la media es mayor que para una variable aleatoriasimilar con una desviación estándar mayor si pensamos en la probabilidad entérminos de una área, esperaríamos una distribución continua con un valor grandede que indique una variabilidad mayor y, por lo tanto, esperaríamos que el áreaeste extendida. Sin embargo, una desviación estándar pequeña debería tener lamayor parte de su área cercana a µ.Podemos argumentar lo mismo para una distribución discreta. En el histograma deprobabilidad. El área se extiende mucho más que. Lo cual indica una distribuciónmas variable de mediciones o resultados el matemático ruso P. L. Chebyschev(1821±1894) descubrió que la fracción de área entre cualesquiera dos valoressimétricos alrededor de la media esta relacionada con la desviación estándar.Como el área bajo una curva de distribución de probabilidad, o de un histogramade probabilidad, suma 1, el área entre cualesquiera dos números es laprobabilidad de que la variable aleatoria tome un valor entre estos números