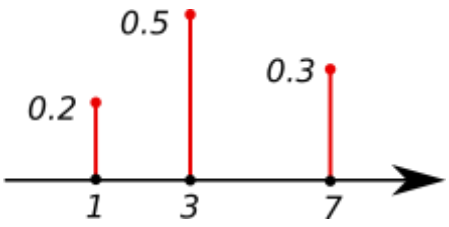
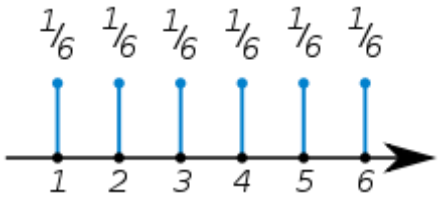
Unidad IV: Distribuciones muestrales

4.1 Función de probabilidad

En teoría de la probabilidad, una función de probabilidad (también denominada función de masa de probabilidad) es una función que asocia a cada punto de su espacio muestral X la probabilidad de que ésta lo asuma.



La gráfica de una función de probabilidad de masa, note que todos los valores no son negativos, y la suma de ellos es igual a 1.



La funcion de masa de probablilidad de un Dado. Todos los numeros tienen la misma probabilidad de aparecer cuando este es tirado.

En concreto, si el espacio muestral, E de la variable aleatoria X consta de los puntos x1, x2, ..., xk, la función de probabilidad P asociada a X es $P(x_i) = p_i$,

donde pi es la probabilidad del suceso X = xi.

Por definición de probabilidad, $\sum_{1}^{k} P(x_i) = 1.$

Hay que advertir que el concepto de función de probabilidad sólo tiene sentido para variables aleatorias que toman un conjunto discreto de valores. Para variables aleatorias continuas el concepto análogo es el de función de densidad.

4.2 Distribución binomial

En estadística, la distribución binomial es una distribución de probabilidad discreta que mide el número de éxitos en una secuencia de n ensayos de Bernoulli independientes entre sí, con una probabilidad fija p de ocurrencia del éxito entre los ensayos.

Un experimento de Bernoulli se caracteriza por ser dicotómico, esto es, sólo son posibles dos resultados. A uno de estos se denomina éxito y tiene una probabilidad de ocurrencia p y al otro, fracaso, con una probabilidad q = 1 - p. En la distribución binomial el anterior experimento se repite n veces, de forma

independiente, y se trata de calcular la probabilidad de un determinado número de éxitos. Para n = 1, la binomial se convierte, de hecho, en una distribución de Bernoulli.

Para representar que una variable aleatoria X sigue una distribución binomial de parámetros n y p, se escribe:

$$X \sim B(n,p)$$

La distribución binomial es la base del test binomial de significación estadística.

4.3 Distribución hipergeométrica

En teoría de la probabilidad la distribución hipergeométrica es una distribución discreta relacionada con muestreos aleatorios y sin reemplazo. Supóngase que se tiene una población de N elementos de los cuales, d pertenecen a la categoría A y N-d a la B. La distribución hipergeométrica mide la probabilidad de obtener x ($0 \leq x \leq d$) elementos de la categoría A en una muestra den elementos de la población original.

Propiedades

La función de probabilidad de una variable aleatoria con distribución hipergeométrica puede deducirse a través de razonamientos combinatorios y es igual a

$$P(X = x) = \frac{\binom{d}{x}\binom{N-d}{n-x}}{\binom{N}{n}},$$

donde N es el tamaño de población, n es el tamaño de la muestra extraída, d es el número de elementos en la población original que pertenecen a la categoría deseada y x es el número de elementos en la muestra que pertenecen a dicha

categoría. La notación $\binom{a}{b}$ hace referencia al coeficiente binomial, es decir, el número de combinaciones posibles al seleccionar b elementos de un total a.

El valor esperado de una variable aleatoria X que sigue la distribución hipergeométrica es

$$E[X] = \frac{nd}{N}$$

y su varianza,

$$Var[X] = \left(\frac{N-n}{N-1}\right)\left(\frac{nd}{N}\right)\left(1-\frac{d}{N}\right).$$

4.4 Distribución de Poisson

En teoría de probabilidad y estadística, la distribución de Poisson es una distribución de probabilidad discreta que expresa, a partir de una frecuencia de ocurrencia media, la probabilidad que ocurra un determinado número de eventos durante cierto periodo de tiempo.

Fue descubierta por Siméon-Denis Poisson, que la dio a conocer en 1838 en su trabajo Recherches sur la probabilista des jugements en matières criminelles et matière civile (Investigación sobre la probabilidad de los juicios en materias criminales y civiles).

Propiedades

La función de masa de la distribución de Poisson es

$$f(k;\lambda) = \frac{e^{-\lambda}\lambda^k}{k!},$$

donde

k es el número de ocurrencias del evento o fenómeno (la función nos da la probabilidad de que el evento suceda precisamente k veces).

λ es un parámetro positivo que representa el número de veces que se espera que ocurra el fenómeno durante un intervalo dado. Por ejemplo, si el suceso estudiado tiene lugar en promedio 4 veces por minuto y estamos interesados en la probabilidad de que ocurra k veces dentro de un intervalo de 10 minutos, usaremos un modelo de distribución de Poisson con λ = 10×4 = 40.

e es la base de los logaritmos naturales (e = 2,71828...)

Tanto el valor esperado como la varianza de una variable aleatoria con distribución de Poisson son iguales a λ. Los momentos de orden superior son polinomios de Touchard en λ cuyos coeficientes tienen una interpretación combinatorio. De hecho, cuando el valor esperado de la distribución de Poisson es 1, entonces según la fórmula de Dobinski, el n-ésimo momento iguala al número de particiones de tamaño n.

La moda de una variable aleatoria de distribución de Poisson con un λ no entero es igual a $\lfloor\lambda\rfloor$, el mayor de los enteros menores que λ (los símbolos $\lfloor\,\rfloor$ representan la función parte entera). Cuando λ es un entero positivo, las modas son λ y λ − 1.

La función generadora de momentos de la distribución de Poisson con valor esperado λ es

$$\mathrm{E}\left(e^{tX}\right) = \sum_{k=0}^{\infty} e^{tk} f(k;\lambda) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k e^{-\lambda}}{k!} = e^{\lambda(e^t-1)}.$$

4.5 Esperanza matemática

Cuando la variable aleatoria es discreta, la esperanza es igual a la suma de la probabilidad de cada posible suceso aleatorio multiplicado por el valor de dicho suceso. Por lo tanto, representa la cantidad media que se "espera" como resultado

de un experimento aleatorio cuando la probabilidad de cada suceso se mantiene constante y el experimento se repite un elevado número de veces. Cabe decir que el valor que toma la esperanza matemática en algunos casos puede no ser "esperado" en el sentido más general de la palabra - el valor de la esperanza puede ser improbable o incluso imposible.

Por ejemplo, el valor esperado cuando tiramos un dado equilibrado de 6 caras es 3,5. Podemos hacer el cálculo

$$\begin{aligned}\mathrm{E}(X) &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= \frac{1+2+3+4+5+6}{6} = 3,5\end{aligned}$$

y cabe destacar que 3,5 no es un valor posible al rodar el dado. En este caso, en el que todos los sucesos son de igual probabilidad, la esperanza es igual a la media aritmética.

4.6 Distribución normal

En estadística y probabilidad se llama distribución normal, distribución de Gauss o distribución gaussiana, a una de lasdistribuciones de probabilidad de variable continua que con más frecuencia aparece aproximada en fenómenos reales.

La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro estadístico. Esta curva se conoce como campana de Gauss y es el gráfico de una función gaussiana.

La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos. Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la enorme cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.

De hecho, la estadística es un modelo matemático que sólo permite describir un fenómeno, sin explicación alguna. Para la explicación causal es preciso el *diseño experimental*, de ahí que al uso de la estadística en psicología y sociología sea conocido como *método correlacional*.

La distribución normal también es importante por su relación con la estimación por mínimos cuadrados, uno de los métodos de estimación más simples y antiguos.

4.7 Distribución T-student

En probabilidad y estadística, la distribución t (de Student) es una distribución de probabilidad que surge del problema de estimarla media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño.

Aparece de manera natural al realizar la prueba t de Student para la determinación de las diferencias entre dos medias muestrales y para la construcción del intervalo de confianza para la diferencia entre las medias de dos poblaciones cuando se desconoce ladesviación típica de una población y ésta debe ser estimada a partir de los datos de una muestra.

Propiedades

La distribución t de Student es la distribución de probabilidad del cociente

$$\frac{Z}{\sqrt{V/\nu}}$$

donde

Z tiene una distribución normal de media nula y varianza 1

V tiene una distribución ji-cuadrado con ν grados de libertad

Z y V son independientes

4.8 Distribución Chi cuadrada

En estadística, la distribución χ² (de Pearson), llamada Chi cuadrado o Ji cuadrado, es una distribución de probabilidad continua con un parámetro k que representa los grados de libertad de la variable aleatoria

$$X = Z_1^2 + \cdots + Z_k^2$$

donde Z_i son variables aleatorias normales independientes de media cero y varianza uno. El que la variable aleatoria X tenga esta distribución se representa habitualmente así: $X \sim \chi_k^2$.

Es conveniente tener en cuenta que la letra griega χ se transcribe al latín como chi y se pronuncia en castellano como ji.

Propiedades

Función de densidad

Su función de densidad es:

$$f(x;k) = \begin{cases} \dfrac{1}{2^{k/2}\Gamma(k/2)}\, x^{(k/2)-1} e^{-x/2} & \text{para } x \geq 0, \\ 0 & \text{para } x < 0 \end{cases}$$

donde Γ es la función gamma.

4.9 Distribución F

Usada en teoría de probabilidad y estadística, la distribución F es una distribución de probabilidad continua. También se le conoce como distribución F de Snedecor (por George Snedecor) o como distribución F de Fisher-Snedecor.

Una variable aleatoria de distribución F se construye como el siguiente cociente:

$$F = \frac{U_1/d_1}{U_2/d_2}$$

donde

U1 y U2 siguen una distribución chi-cuadrado con d1 y d2 grados de libertad respectivamente, y

U1 y U2 son estadísticamente independientes.

La distribución F aparece frecuentemente como la distribución nula de una prueba estadística, especialmente en el análisis de varianza. Véase el test F.

La función de densidad de una F(d1, d2) viene dada por

$$g(x) = \frac{1}{\mathrm{B}(d_1/2, d_2/2)} \left(\frac{d_1\, x}{d_1\, x + d_2}\right)^{d_1/2} \left(1 - \frac{d_1\, x}{d_1\, x + d_2}\right)^{d_2/2} x^{-1}$$

para todo número real x ≥ 0, donde d1 y d2 son enteros positivos, y B es la función beta.

La función de distribución es

$$G(x) = I_{\frac{d_1 x}{d_1 x + d_2}}(d_1/2, d_2/2)$$

donde I es la función beta incompleta regularizada.