Variable Aleatoria Normal

Clase del 8/9/2024

Expresión algebraica

Definición de la f.d.p.

Normal $X \sim N(\mu, \sigma)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in]-\infty, \infty[$$

- μ Es la $E[x]$
- σ, donde $Var[X] = \sigma^2$
- Observa que es simétrica respecto la recta $x = \mu$

Normal standard

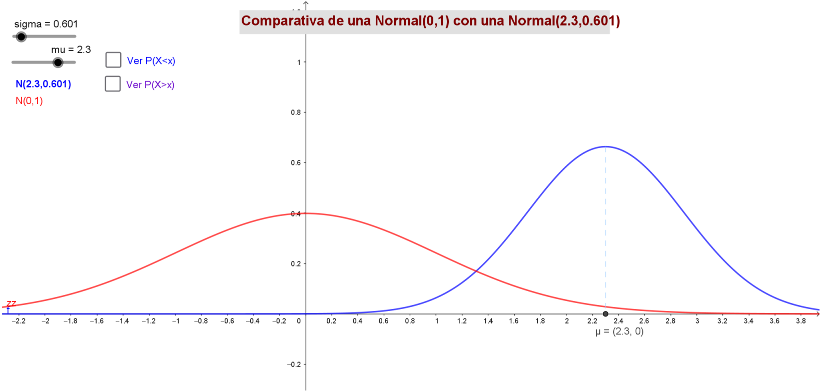
Definición de la f.d.p. de una normal standard

Normal $X \sim N(0,1)$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in]-\infty, \infty[$$

- Los parámetros son $\mu = 0$
- $\sigma = 1$

Gráfica


Comparativa de una Normal(0,1) con una Normal(2.3,0.601)
sigma = 0.601
mu = 2.3
Ver P(X<x)
Ver P(X>x)
N(2.3,0.601)
N(0,1)
μ = (2.3, 0)

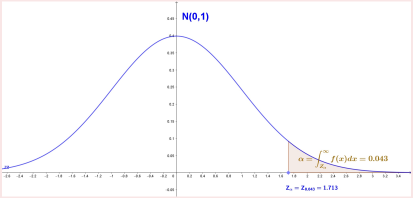
Función de distribución

Notación en la normal tipificada $Z \sim N(0,1)$

La función de distribución se denota como

$$\phi(x) = P(X \leq x)$$

Se denota Z_α el valor tal que $P(X \leq Z_\alpha) = 1 - \alpha$



Propiedades de la simetría

Propiedades

$$\phi(-a) = 1 - \phi(a)$$
$$P(Z \in \left[Z_{-\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}\right] = \alpha$$

Propiedades

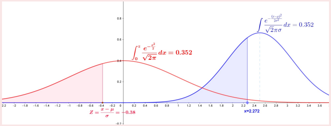
Linealidad

Si $X \sim N(\mu, \sigma)$ y consideramos la v.a. $Y = aX + b$,
donde a y b con $a \neq 0$,
entonces $Y \sim N(a\mu + b, a\sigma)$

Tipificación de la normal

Si $X \sim N(\mu, \sigma)$ y consideramos la v.a. $Y = \frac{X-\mu}{\sigma}$,

entonces $Y \sim N(0, 1)$



Más Propiedades

Suma de normales

Si $X_i \sim N(\mu_i, \sigma_i)$ con $i \in \{1, 2, ..., k\}$ y son independientes, entonces la v.a. $Y = X_1 + X_2 + ... + X_k$ cumple que

$$Y \sim N\left(\mu_1 + \mu_2 + ... + \mu_k, \sqrt{\sigma_1^2 + \sigma_2^2 + ...\sigma_k^2}\right)$$

Más Propiedades

Media muestral de normales

Si $X_i \sim N(\mu, \sigma)$ con $i \in \{1, 2, ..., k\}$ y son independientes, entonces la v.a. $\overline{X}_n = \frac{X_1+X_2+...+X_n}{n}$ cumple que

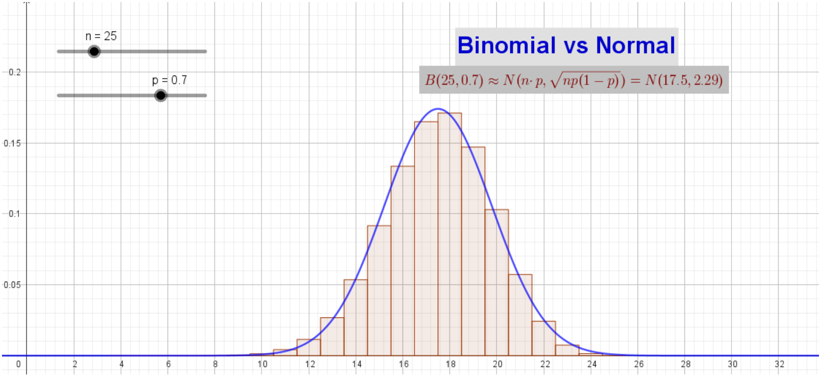
$$\overline{X}_n \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Binomial vs Normal

S

i $X \sim B(n, p)$ con un n *suficientemente* alto y un valor de p bajo, entonces X se puede aproximar por una distribución $N\left(np, \sqrt{bp(1-p)}\right)$. En muchas aplicaciones se suele tomar la aproximación cuando $np \geq 5 \wedge n(1-p) \geq 5$

Binomial vs Normal

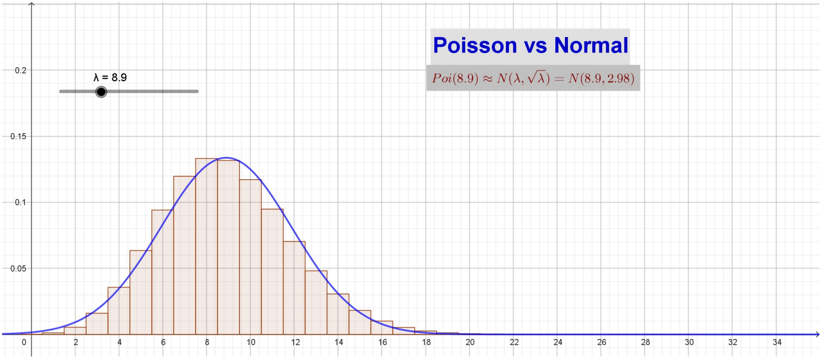


Poisson vs Normal

S

i $X \sim Poi(\lambda)$ con un $\lambda > 15$, entonces X se puede aproximar por una distribución $N\left(\lambda, \sqrt{\lambda}\right)$.

Poisson vs Normal



Introducción

Cuando las variables aleatorias son de distribución desconocida, se pueden sacar ya algunas propiedades que involucran a la *Esperanza* y la *Varianza* de dichas variables. Veamos algunos resultados previos al Teorema Central del Límite o Teorema de los grandes Números.

Teorema de los grandes números (TGN)

Si tenemos una secuencia de variables aleatorias independientes e idénticamente distribuidas $X_1, X_2, \ldots, X_n$, con esperanza finita $\mu = \mathbb{E}(X_i)$, entonces el promedio de estas variables tiende al valor esperado μ cuando el número de observaciones n tiende a infinito. Esto es:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{n\to\infty} \mu$$

Interpretación

El TGN asegura que, al tomar muchas observaciones, el promedio de las variables aleatorias se acerca cada vez más al valor esperado. Cuanto mayor sea la muestra, más confiables serán las estimaciones de la media poblacional.

Teorema de Markov

Sea X una variable aleatoria no negativa (es decir, $X \geq 0$) y sea $a > 0$. Entonces, se cumple que:

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

donde $\mathbb{E}(X)$ es la esperanza (o valor esperado) de X y a es cualquier constante positiva.

Interpretación del Teorema de Markov

Proporciona una cota superior para la probabilidad de que una variable aleatoria no negativa exceda un valor dado, basándose en su valor esperado.
Proporciona una cota para la probabilidad de que una variable aleatoria no negativa sea mayor o igual a un valor a, usando solo el valor esperado de la variable. No se requiere información adicional sobre la distribución de X.

Teorema de Tchebychev

Sea X una variable aleatoria con valor esperado $\mu = \mathbb{E}(X)$ y varianza $\sigma^2 = \mathbb{V}(X)$. Entonces, para cualquier $k > 0$, se cumple que:

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

donde σ^2 es la varianza de X, μ es su esperanza, y k es cualquier número positivo.

Interpretación del Teorema de Tchebychev

El teorema de Tchebychev proporciona una cota para la probabilidad de que una variable aleatoria se desvíe de su valor esperado en más de una cantidad dada, en función de su varianza. Este teorema da una cota superior para la probabilidad de que una variable aleatoria se desvíe de su media en más de k unidades. Es útil para estimar la probabilidad de eventos raros sin necesidad de conocer la distribución exacta de X, solo su media y varianza.

Resumen de los teoremas de Markov y Tchebychev

- Teorema de Markov: Cota para la probabilidad de que una variable no negativa exceda un valor dado, basada en su esperanza.
- Teorema de Tchebychev: Cota para la probabilidad de que una variable se desvíe de su media en más de una cantidad dada, basada en su varianza.

Ambos teoremas son herramientas importantes para acotar probabilidades en situaciones donde no se conoce la distribución completa de la variable aleatoria.

Teorema central del límite

Sea $X_1, X_2, \ldots, X_n$ una sucesión de n variables aleatorias independientes e idénticamente distribuidas (i.i.d.), cada una con media μ y varianza finita σ^2. La media muestral de estas variables ($\overline{X_n}$) cumple que:

Teorema central del límite

A medida que n tiende a infinito, la distribución de la media muestral $\overline{X_n}$, tiende a una $N(\mu, \frac{\sigma}{\sqrt{n}})$, lo que expresamos como:

$$\overline{X_n} \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$$

$$\frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{p} N(0, 1)$$

Comentarios

Este teorema describe el comportamiento del promedio de un gran número de v.a. independientes y con una misma distribución.
En otras palabras, si tomamos muchas muestras independientes de una población, el promedio de esas muestras tiende a seguir una **distribución normal** a medida que aumenta el número de muestras, incluso si los datos originales no están distribuidos normalmente.

Puntos clave del Teorema Central del Límite

1. Independencia: Las variables aleatorias deben ser independientes entre sí.
2. Distribución idéntica: Las variables deben tener la misma distribución.
3. Media y varianza finitas: Las variables deben tener una media μ y varianza σ^2 finitas.
4. Aplicabilidad: A medida que el tamaño de la muestra n crece, la distribución de los promedios se aproxima a una distribución normal.