Universidad de Carabobo
Facultad Experimental de Ciencias y Tecnología

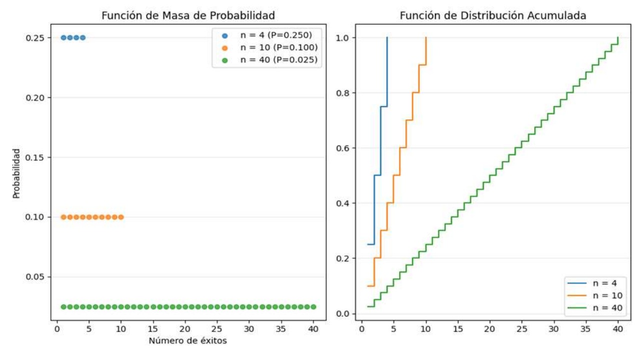
Distribuciones de Probabilidad

Jose Natera
31.185.494
Computación

Distribución Uniforme Discreta

Esta distribución describe variables donde los resultados son números enteros y finitos, y cada uno tiene una probabilidad idéntica.

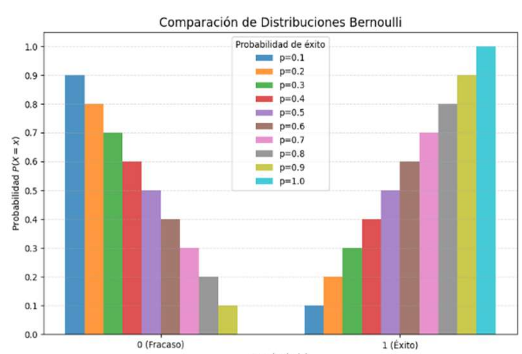
El parámetro principal aquí es el rango, es decir los n elementos. Si se amplía el rango, la probabilidad individual de cada valor disminuye. Como un pastel que se reparte entre más personas; a cada una le toca un trozo más pequeño. Si reduces el rango, la probabilidad de cada punto aumenta, concentrando la certidumbre en menos opciones. La media siempre se mantendrá exactamente en el centro del intervalo, independientemente de qué tan ancho sea.



Aplicaciones técnicas y uso: Se utiliza fundamentalmente en el muestreo aleatorio simple. En informática, es la base de los algoritmos de generación de números pseudoaleatorios cuando se requiere que un programa elija un índice de una lista sin sesgos. En control de calidad, se aplica para seleccionar piezas de un lote numerado donde cada unidad debe tener la misma posibilidad de ser inspeccionada. También es el modelo por defecto en situaciones de incertidumbre máxima, donde no tenemos ninguna razón técnica para creer que un resultado es más probable que otro.

Distribución de Bernoulli

Esta distribución es la base de modelos más complejos. Se define por un único parámetro p, que representa la probabilidad de éxito.



Comportamiento de los parámetros

- **Si p se acerca a 1:** La distribución se sesga totalmente hacia el éxito. La incertidumbre disminuye porque el resultado es casi predecible.
- **Si p se acerca a 0:** El gráfico muestra una barra dominante en el fracaso.
- **Si p = 0.5:** La varianza es **máxima**. Es el punto de mayor incertidumbre, como lanzar una moneda, donde es imposible predecir el resultado con ventaja.
- **Efecto en la varianza:** La varianza se calcula como **p(1 - p)**. Esto significa que si llevas p a los extremos (0 o 1), la varianza desaparece (**0**), confirmando que la aleatoriedad muere cuando el resultado está asegurado.

Aplicaciones técnicas y uso

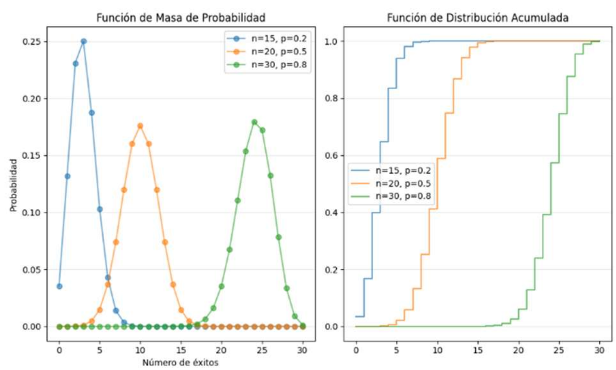
Se utiliza como un **ensayo dicotómico** en procesos donde se requiere una clasificación binaria inmediata. En medicina, se aplica para modelar la efectividad de una dosis única (el paciente sana o no sana). En control de calidad, se usa para la inspección de una sola pieza. En ciencias de la computación, es la base de los **filtros de** decisión y redes neuronales simples donde una neurona se activa o no. Técnicamente, es el bloque de construcción de la Distribución Binomial

Distribución Binomial

Esta distribución se define por dos parámetros principales: **n** (el número total de ensayos) y **p** (la probabilidad de éxito en cada ensayo).

Comportamiento de los parámetros

Jugar con estos valores altera drásticamente la forma de la "curva" de probabilidad:



Si $\mathbf{p = 0.5}$, la distribución es perfectamente simétrica.

Si $\mathbf{p < 0.5}$, la distribución tiene un **sesgo** positivo.

Si $\mathbf{p > 0.5}$, el sesgo es negativo

- **Variación de n:** A medida que el número de ensayos aumenta, la distribución tiende a ensancharse y a reducir su altura máxima, y su forma empieza a aproximarse a una **campana de Gauss**.

Aplicaciones técnicas y uso

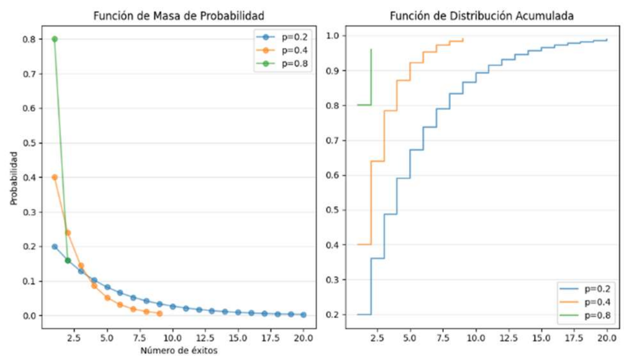
Es el estándar en el **control de calidad de lotes**, donde se calcula la probabilidad de encontrar un número **x** de piezas defectuosas en una muestra de tamaño **n**. En el sector farmacéutico, se utiliza para determinar la eficacia de un tratamiento en un grupo de prueba ej. cuántos pacientes de un grupo de 100 reaccionarán positivamente. También es fundamental en el **marketing digital** para modelar tasas de conversión (clics o ventas) y en ingeniería de telecomunicaciones para estimar el número de paquetes de datos perdidos en una transmisión de longitud fija.

Distribución Geométrica

Esta distribución depende exclusivamente de un parámetro: **p**, la probabilidad de éxito en cada intento.

Comportamiento de los parámetros:

Al ser una distribución que mide el "tiempo de espera" (en términos de intentos) hasta que algo ocurre, su forma es siempre descendente, pero su pendiente cambia según **p**:



- Si **p** es cercano a 1: La probabilidad se concentra fuertemente en el primer intento. La curva cae de forma abrupta; es muy probable que el éxito ocurra de inmediato.

- Si **p** es cercano a 0: La distribución se "estira" hacia la derecha. La caída es mucho más lenta, lo que indica que existe una probabilidad significativa de que se necesiten muchos intentos para lograr el primer éxito.

Aplicaciones técnicas y uso

Es fundamental en el análisis de supervivencia y fiabilidad de sistemas. Se aplica para calcular la probabilidad de que un componente falle por primera vez en un ciclo determinado.Para modelar el número de piezas inspeccionadas hasta encontrar la primera defectuosa. También es clave en la teoría de colas y redes de datos para estimar cuántas veces debe reintentarse la transmisión de un paquete de información hasta que se recibe correctamente.

Distribución Binomial Negativa

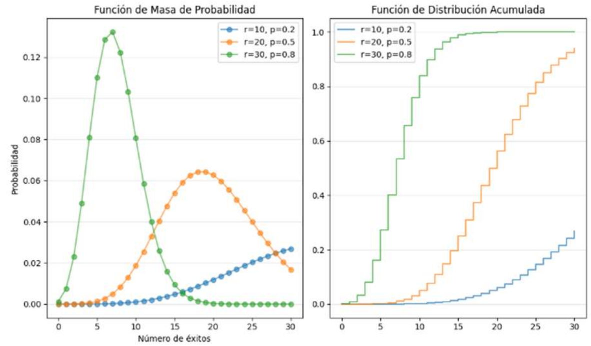
Se define por dos parámetros: **r**, que es el número de éxitos que deseamos alcanzar, y **p**, la probabilidad de éxito en cada ensayo independiente.

Comportamiento de los parámetros

El comportamiento de esta distribución varía según cómo ajustemos sus componentes:

- **Si aumentas r:** La distribución se desplaza hacia la derecha y se vuelve más simétrica. Esto ocurre porque, lógicamente, necesitas más intentos para acumular más éxitos.

- **Si p es pequeño:** La distribución se dispersa mucho. Hay una incertidumbre alta sobre cuándo se alcanzará el objetivo de éxitos, estirando la "cola" de la distribución hacia valores muy altos.



- **Si p es grande:** La distribución se comprime y se desplaza hacia la izquierda, ya que los éxitos ocurren casi de inmediato y el número de intentos necesarios será muy cercano al valor de **r**.
- **Relación con la Geométrica:** Si fijas $\mathbf{r = 1}$, la distribución binomial negativa se convierte exactamente en una distribución geométrica.

Aplicaciones técnicas y uso

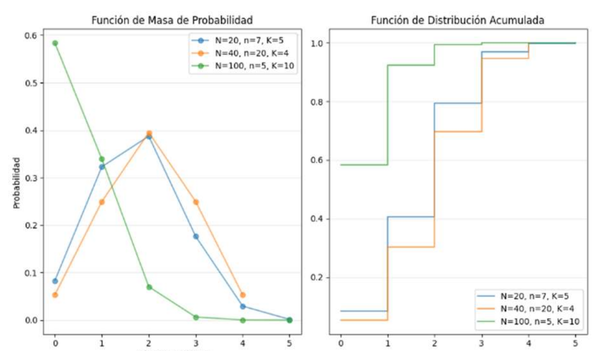
Se utiliza ampliamente en **modelado de datos de conteo con sobre dispersión**, superando a menudo a la distribución de Poisson en eficacia. En **epidemiología**, se aplica para modelar el número de contagios secundarios en brotes de enfermedades. También es crucial en **marketing y análisis de consumo** para predecir la frecuencia de compra de un cliente. En ingeniería, se aplica en pruebas de estrés de materiales para determinar cuántos ciclos de carga resiste un componente antes de alcanzar un número crítico de micro fisuras.

Distribución Hipergeométrica

Esta distribución depende de tres parámetros: **N** siendo el tamaño total de la población, **K** el número de éxitos totales en la población y **n** tamaño de la muestra extraída.

Comportamiento de los parámetros

El comportamiento de la probabilidad cambia drásticamente al ser un proceso sin retorno:



- **Relación n/N:** Si el tamaño de la muestra **n** es muy pequeño respecto a la población total **N**, la distribución se comporta casi como una Binomial. Pero si **n** representa una parte significativa de **N**, la probabilidad cambia rápido tras cada éxito extraído.
- Si **K** se acerca a **N** o a **0:** La varianza se reduce. Si casi todos son "éxitos" o casi todos son "fracasos" en la población original, hay poca incertidumbre sobre lo que obtendrás en la muestra.

- **Efecto del tamaño de muestra** n**:** A medida que **n** aumenta, la distribución se vuelve más estrecha y se desplaza hacia la derecha. Si **n** igualara a **N**, la probabilidad de obtener **K** éxitos sería exactamente 1.

Aplicaciones técnicas y uso

Es la herramienta estándar en el **Control de Calidad de lotes pequeños**, donde no se puede devolver la pieza probada al lote. Se aplica rigurosamente en **Auditoría Fiscal y Contable** para seleccionar una muestra de facturas y determinar la presencia de errores sin revisar todo el archivo. También es la base de las **Pruebas Exactas de Fisher** en estadística clínica para analizar tablas de contingencia con muestras muy reducidas.

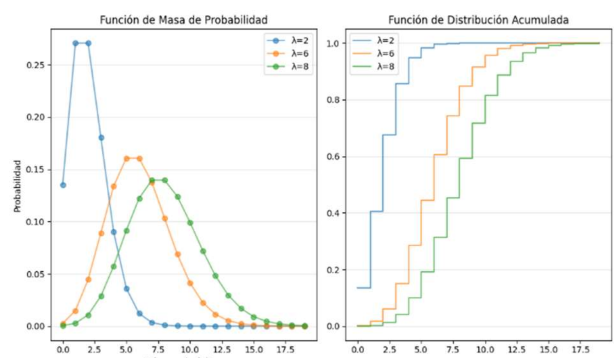
Distribución de Poisson

Esta distribución depende de un único parámetro: lambda λ, que representa el promedio de ocurrencias en el intervalo dado.

Comportamiento de los parámetros

Al jugar con el valor de λ , la forma y dispersión de la distribución cambian significativamente:

- **Si λ es pequeño :** La distribución está fuertemente sesgada a la derecha. El valor más probable es cero, lo más común es que no ocurra nada en ese intervalo.
- **Si λ aumenta:** El pico de la distribución se desplaza hacia la derecha. A medida que el promedio de eventos aumenta, la distribución se vuelve más ancha.
- **Tendencia a la simetría:** Una característica técnica clave es que, a medida que **λ** crece , la distribución de Poisson empieza a perder su asimetría y comienza a parecerse a una **Distribución Normal**.



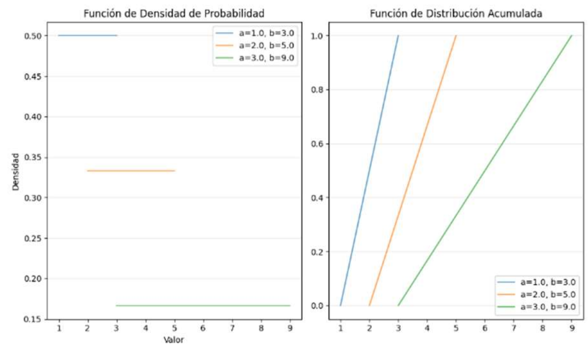
Aplicaciones técnicas y uso

Es el estándar de oro en la **Teoría de Colas** y la ingeniería de telecomunicaciones; se aplica para modelar la llegada de clientes a un banco, el número de llamadas que recibe un conmutador o la llegada de paquetes de datos a un servidor. También es crucial en la **física nuclear** para medir el número de partículas emitidas por una fuente radiactiva en un tiempo dado. En el ámbito digital, se usa para analizar el tráfico de sitios web y detectar anomalías.

Distribución Uniforme Continua

A diferencia de la discreta, aquí la variable puede tomar cualquier valor real dentro de un intervalo **[a,b]**. La probabilidad de un punto exacto es técnicamente cero; lo que medimos es la probabilidad de caer en un subintervalo.

Comportamiento de los parámetros



- Al alejar los límites, la altura de la función de densidad baja.
- Al acercar los límites, la densidad de probabilidad se dispara hacia arriba, indicando una mayor concentración de probabilidad en un espacio físico o temporal más pequeño.
- Si a y b fueran casi iguales, la distribución empezaría a parecerse a un pico muy alto.

Aplicaciones técnicas y uso: En física e ingeniería, se aplica para modelar el error de cuantificación . También es el modelo estándar para representar el tiempo de llegada de un evento cuando solo se sabe que ocurrirá dentro de una ventana de tiempo específica, o para modelar la ubicación de una partícula en un espacio donde no actúan fuerzas externas que la desplacen hacia un lado u otro.

Distribución Exponencial

es la contraparte continua de la distribución geométrica. Se utiliza para modelar el tiempo o el espacio que transcurre entre dos eventos sucesivos que ocurren siguiendo un proceso de Poisson. Se define por un único parámetro **λ**, que representa la tasa de ocurrencia. Su función de densidad es siempre decreciente.

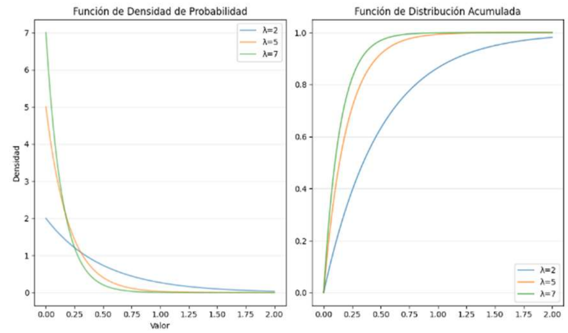
Comportamiento de los parámetros

Al modificar la tasa **λ**, la velocidad con la que la probabilidad decae cambia drásticamente:

- **Si λ es alto:** La curva cae de forma muy pronunciada. Esto indica que los eventos ocurren con mucha frecuencia y, por lo tanto, el tiempo de espera entre ellos es muy corto. La mayoría de la probabilidad se concentra cerca del origen.

- **Si λ es bajo:** La curva se aplana y se extiende hacia la derecha. Esto representa eventos raros o poco frecuentes, donde los tiempos de espera largos son mucho más probables.

- **Relación con la Media:** La media de la distribución es $1/\boldsymbol{\lambda}$ Si duplicas la tasa de eventos, el tiempo de espera promedio se reduce a la mitad.



Aplicaciones técnicas y uso

Es pilar fundamental en la **Ingeniería de Confiabilidad y Fiabilidad**, utilizada para modelar el tiempo hasta la falla de componentes electrónicos que no presentan desgaste químico. En **Teoría de Colas**, se aplica para modelar el tiempo entre la llegada de clientes a un sistema o el tiempo que tarda un servidor en procesar una solicitud. En **hidrología**, se usa para analizar el tiempo entre tormentas o inundaciones de cierta magnitud. También es crucial en la **física cuántica** para describir el tiempo de desintegración de partículas radiactivas.

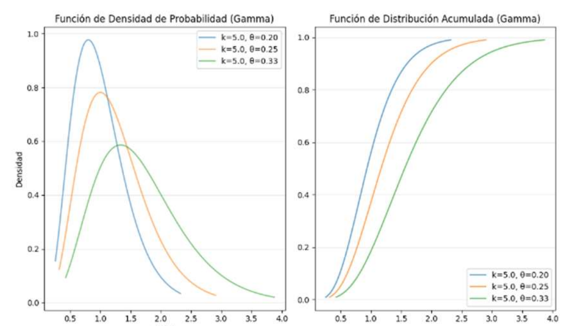
Distribución Gamma

Es una distribución continua de gran versatilidad que generaliza la distribución exponencial. Se utiliza para modelar el tiempo de espera hasta que ocurran **n** eventos independientes en un proceso de Poisson. Se define principalmente por dos parámetros α y λ.

Comportamiento de los parámetros

A diferencia de otras distribuciones, la interacción entre sus parámetros puede cambiar la naturaleza misma de su gráfico:

- Si $\alpha = 1$, la distribución es idéntica a una **Exponencial**
- Si $\alpha < 1$, la curva tiene una pendiente infinita en el origen, concentrándose en valores muy bajos.
- Si $\alpha > 1$, la curva deja de ser siempre decreciente y adquiere una forma de campana asimétrica. A medida que α aumenta, la distribución se vuelve más simétrica y se aproxima a una **Distribución Normal**.



- **Efecto de $\boldsymbol{\beta}$:** Este parámetro estira o comprime la distribución a lo largo del eje horizontal.
- Un valor de escala grande dispersa la probabilidad sobre un rango más amplio, mientras que un valor pequeño concentra la probabilidad cerca del origen.

Aplicaciones técnicas y uso

Es una de las herramientas más potentes en **Meteorología e Hidrología** para modelar la cantidad de precipitación acumulada en un periodo. En **Economía y Seguros**, se aplica para modelar el tamaño total de las reclamaciones o pérdidas agregadas. En **Ciencia de Datos**, se utiliza en estadística bayesiana para estimar tasas de ocurrencia. También es vital en **Ingeniería de Telecomunicaciones** para modelar el tiempo de servicio en sistemas complejos donde el servicio consta de múltiples etapas secuenciales.

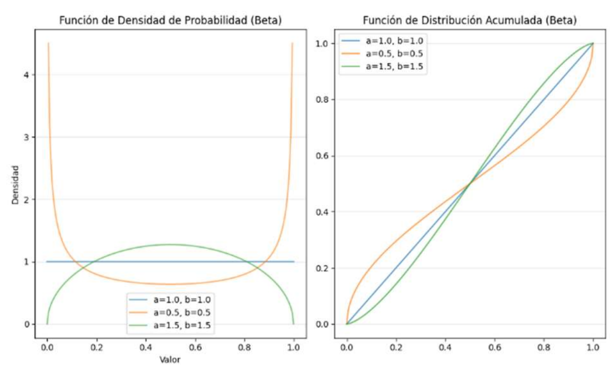
Distribución Beta

Se define mediante dos parámetros de forma: alfa y beta. Ambos deben ser mayores que cero y controlan la densidad de la probabilidad dentro del rango delimitado.

Comportamiento de los parámetros

La versatilidad de esta distribución es su característica técnica más notable, ya que puede adoptar casi cualquier forma dependiendo de cómo se juegue con sus parámetros:

- **Si $\boldsymbol{\alpha} = \boldsymbol{\beta} = 1$:** La distribución se convierte en una **Uniforme Continua**.



- **Si $\boldsymbol{\alpha} > \boldsymbol{\beta}$:** La masa de probabilidad se desplaza hacia la derecha. Se utiliza cuando se espera que la proporción sea alta.
- **Si $\boldsymbol{\alpha} < \boldsymbol{\beta}$:** La masa se desplaza hacia la izquierda, indicando que los valores bajos son más probables.
- **Si ambos son mayores que 1:** La distribución tiene forma de campana. Si $\boldsymbol{\alpha} = \boldsymbol{\beta}$, la campana es perfectamente simétrica en 0.5.
- **Si ambos son menores que 1:** La distribución adquiere forma de "U". Esto modela fenómenos donde los extremos son muy probables; por ejemplo, una opinión polarizada.

Aplicaciones técnicas y uso

Es el estándar en la **Estadística Bayesiana** como la distribución *a priori* conjugada para la distribución Binomial y de Bernoulli. Se aplica para actualizar creencias sobre una tasa de éxito desconocida a medida que se obtienen nuevos datos. En **Gestión de Proyectos**, se utiliza para modelar el tiempo de finalización de una tarea cuando se conocen los tiempos optimista, pesimista y más probable.

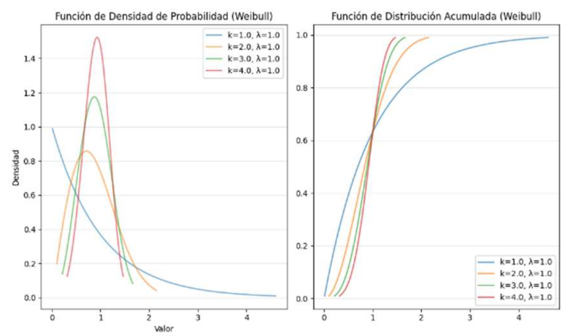
Distribución de Weibull

Se define mediante dos parámetros principales: **k** y λ

Comportamiento de los parámetros

La versatilidad de Weibull reside en cómo el parámetro k altera la naturaleza de la tasa de fallos:

- **Si k $<$ 1:** La tasa de fallos disminuye con el tiempo.
- **Si k $=$ 1:** La distribución se convierte exactamente en una **Distribución Exponencial**. La tasa de fallos es constante.
- **Si k $>$ 1:** La tasa de fallos aumenta con el tiempo.
- **Parámetro $\boldsymbol{\lambda}$:** Determina qué tan estirada está la curva en el tiempo.



Aplicaciones técnicas y uso

Es la herramienta reina en el **Análisis de Supervivencia y Confiabilidad Industrial**. Se utiliza para predecir cuándo fallará un motor, un rodamiento o cualquier componente mecánico sujeto a fatiga. En la **industria de las energías renovables**, es el estándar para modelar la distribución de la **velocidad del viento** en un sitio específico, lo que permite calcular el potencial energético de un parque eólico.

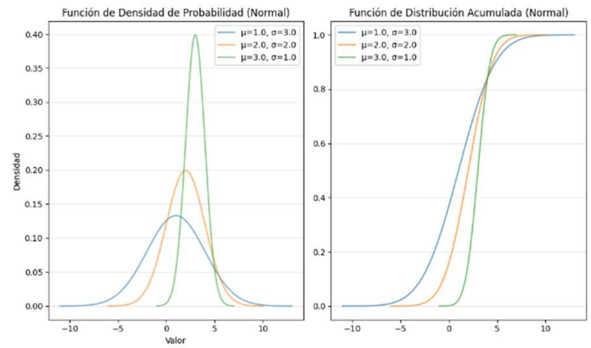
Distribución Normal

También conocida como la Campana de Gauss, es la distribución más importante en estadística. Su relevancia radica en el **Teorema del Límite Central**, que establece que la suma de muchas variables aleatorias independientes tiende a seguir esta distribución, sin importar la forma original de dichas variables. Se define mediante dos parámetros: μ y σ.

Comportamiento de los parámetros:

- **Variación de $\boldsymbol{\mu}$:** Cambiar la media desplaza la campana a lo largo del eje horizontal hacia la derecha o izquierda, pero no altera su forma ni su anchura.
- **Variación de $\boldsymbol{\sigma}$:** Si $\boldsymbol{\sigma}$ es **pequeño**, la campana es alta y estrecha, lo que indica que los datos están muy concentrados cerca de la media.

- Si $\boldsymbol{\sigma}$ es **grande**, la campana se aplana y se ensancha, indicando que los datos están muy dispersos y los valores alejados de la media son más frecuentes.



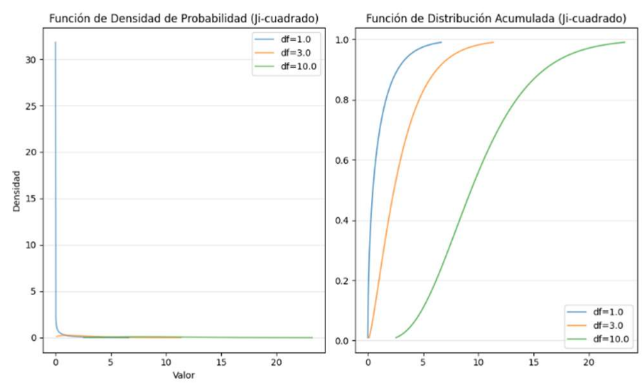
Aplicaciones técnicas y uso

Es la base de la **Inferencia Estadística Paramétrica**. Se utiliza para modelar caracteres morfológicos en biología, errores de medición en experimentos científicos y puntuaciones en pruebas psicométricas. En el **sector financiero**, se aplica para modelar los rendimientos de activos, y en **Control de Calidad**. También es fundamental en el **Machine Learning** para la normalización de datos y en algoritmos como la Regresión Lineal y el Análisis de Discriminante Lineal

Distribución Ji-cuadrada

La **Distribución ji-cuadrada** es una distribución continua que surge de la suma de los cuadrados de variables aleatorias normales estándar independientes. Es una herramienta fundamental para la inferencia estadística y las pruebas de hipótesis.

Comportamiento de los parámetros



Al ser una suma de valores al cuadrado, la distribución solo existe para valores positivos y su forma cambia drásticamente según el valor de **k**:

- **Si k es bajo:** La distribución tiene una forma de "L" o una pendiente descendente muy pronunciada que empieza desde el eje vertical.
- **Si k aumenta:** La distribución se vuelve asimétrica hacia la derecha.
- **Si k es muy grande:** por el Teorema del Límite Central, la distribución ji-cuadrada comienza a perder su asimetría y tiende a aproximarse a una **Distribución Normal**.

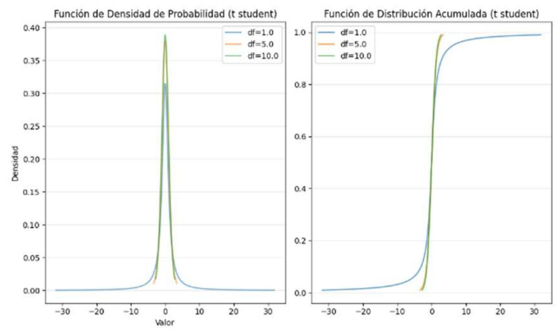
Aplicaciones técnicas y uso

Es el pilar de las **Pruebas de Bondad de Ajuste**, utilizadas para determinar si un conjunto de datos observados sigue una distribución teórica específica (por ejemplo, para saber si un dado está cargado o si los datos siguen una distribución normal). También es crucial en la **Inferencia de la Varianza**, permitiendo construir intervalos de confianza y realizar pruebas de hipótesis sobre la variabilidad de una población normal.

Distribución T

La **Distribución t** es una distribución de probabilidad continua que surge al estimar la media de una población con distribución normal cuando el tamaño de la muestra es pequeño y la desviación estándar poblacional es desconocida.

Su forma depende de un único parámetro n, conocido como **grados de libertad**, que está directamente relacionado con el tamaño de la muestra (n - 1).



Comportamiento de los parámetros

- **Si los grados de libertad son bajos:** La distribución tiene colas más pesadas que la normal. Esto significa que los valores extremos son más probables, reflejando la mayor incertidumbre al trabajar con pocos datos.
- **A medida que n aumenta:** Las colas se vuelven más delgadas y el pico central se eleva. La distribución comienza a estrecharse.
- **Convergencia:** Cuando los grados de libertad tienden al infinito, la distribución se vuelve prácticamente idéntica a la **Distribución Normal Estándar**.

Aplicaciones técnicas y uso

Es la herramienta fundamental para la **Inferencia Estadística con muestras pequeñas**. Se aplica en las **Pruebas de Hipótesis** para comparar si las medias de dos grupos son significativamente diferentes. En ingeniería, se emplea para validar procesos de fabricación donde las pruebas son costosas o destructivas y solo se dispone de unos pocos prototipos para el análisis.

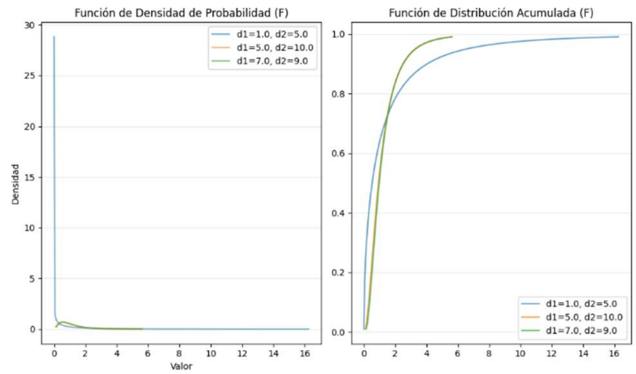
La Distribución F

Es una distribución continua que aparece principalmente al comparar las varianzas de dos poblaciones diferentes. Se define como el cociente de dos distribuciones ji-cuadrada independientes, cada una dividida por sus respectivos grados de libertad.

Distribución F

Esta distribución depende de dos parámetros distintos: d1 y d2.

Comportamiento de los parámetros



- Cuando ambos valores son pequeños, la distribución tiene una asimetría muy marcada y una "cola" muy larga hacia la derecha.
- **Aumento de d2:** A medida que los grados de libertad del denominador aumentan, la variabilidad del denominador disminuye y la distribución se vuelve más estable.
- **Aumento de ambos:** Si tanto a como b crecen significativamente, la distribución tiende a perder su asimetría y comienza a parecerse a una **Distribución Normal**.

Aplicaciones técnicas y uso

Es la pieza central del **Análisis de Varianza**, una técnica estadística que permite comparar las medias de tres o más grupos simultáneamente para ver si al menos una es diferente, por ejemplo, probar si cuatro tipos de fertilizantes producen rendimientos distintos en cultivos. También se aplica en el **Control de Calidad** mediante la Prueba de Identidad de Varianzas para verificar si dos procesos de producción tienen la misma precisión o si un nuevo método ha reducido la variabilidad del sistema. En general, cualquier prueba que busque comparar ruido contra señal en modelos complejos suele apoyarse en esta distribución.