Fundamentos de Inferencia Estadística

Francisco Javier Mercader Martínez

Índice

Tema 1: Muestreo y distribuciones muestrales

1.1) Introducción

El contexto

- Tenemos una pregunta acerca de un fenómenos aleatorio.
- Formulamos un modelo para la varaible de interés X.
- Traducimos la pregunta de interés en términos de uno o varios parámetros del modelo.
- Repetimos el experimento varias veces, apuntamos los valores de X.
- ¿Cómo usar estos valores para extraer información sobre el parámetro?

1.2) Ejemplos

¿Está la moneda trucada?

- Experimento: tirar la modena. X = resultado obtenido.
  $P(X = +) = p, P(X = c) = 1 - p$
  $$¿p = \frac{1}{2}?$$

Sondeo sobre intención de participación en unas elecciones

- Queremos estima la tasa de participación antes de unas elecciones generales.
- Formulamos un modelo:
  → Experimento: "escoger una persona al azar en el censo".
  → X: participación, variable dicotómica ("Sí" o "No"). $p = P(X = \text{Si})$.
- ¿Cuánto vale p?
- Censo: aproximadamente 37 000 000. Escogemos aproximadamente 3000 personas.

Determinación de la concentración de un producto

- Quiero determinar la concentración de un producto.
- Formulo el modelo:
  → Experimento: "llevar a cabo una medición".
  → X: "valor proporcionado por el aparato".
  → $X \sim \mathcal{N}(\mu, \sigma^2)$.
- ¿Qué vale μ?

1.3) Surge una pregunta

En todas estas situaciones donde nos basamos en la repetición de un experimento simple...

- ¿Cómo sabemos que nuestra estimación es fiable?
- ¿Qué confianza tenemos al extrapolar los resultados de una muestra de 3000 personas a una población de 37 millones de personas?

1.4) Esbozo de respuesta: tasa de participación

Para convenceros, un experimento de simulación

- Voy a simular el proceso de extracción de una muestra de 3000 personas en una población de 37 millones de personas.
- Construyo a mi antojo los distintos componentes:
  - → **La población:** defino en mi ordenador un conjunto de 37 000 000 de ceros y unos. (⇔ el censo electoral)
    - ”1” ⇔ ”la persona piensa ir a votar”.
    - ”0” ⇔ ”la persona **no** peinsa ir a votar”
  - → **La tasa de participación ”real”:** Decido que en mi población el 70% piensa ir a votar → 25 900 000 ”1”s.
  - → **La extracción de una muestra:** construyo un pequeño programa que extrae al azar una muestra de 3000 números dentro del conjunto grande.

```r
poblacion <- c(rep(1, 25900000), rep(0, 11100000))
set.seed(314159)
p_muestra <- mean(sample(poblacion, 3000, replace = FALSE))
p_muestra
```

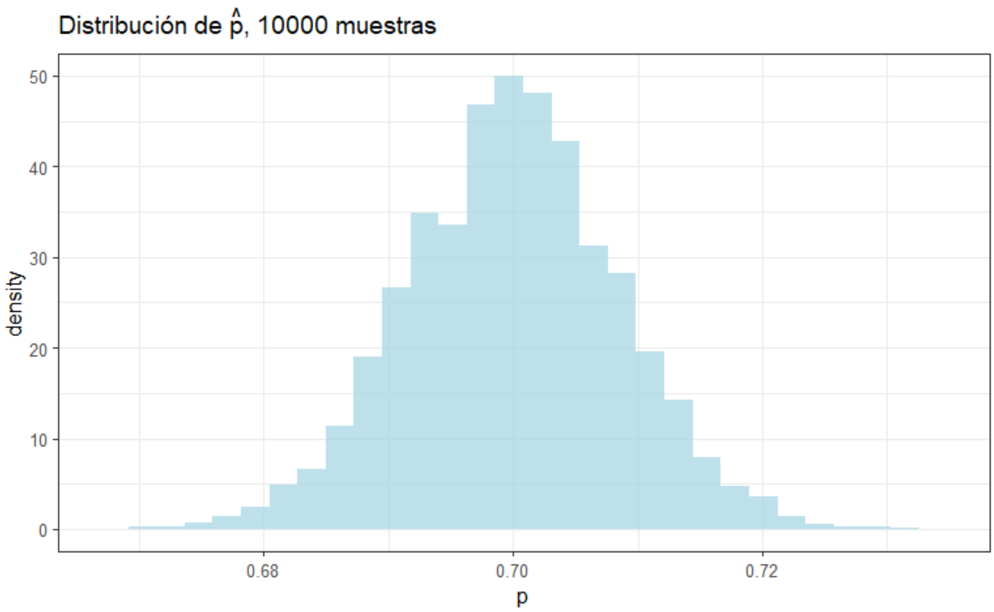
```
## [1] 0.705667
```

Queremos descartar que haya sido suerte. Vamos a repetir muchas veces (10000 veces por ejemplo), la extracción de una muestra de 3000 personas en la población.

```r
library(tidyverse)
lista_muestras <- replicate(
    10000,
    sample(poblacion, 3000, replace = FALSE),
    simplify = FALSE
)
p_muestras <- map_dbl(lista_muestras, mean)
head(p_muestras)
```

```
## [1] 0.6970000  0.7030000  0.7036667  0.7023333  0.7013333  0.7226667
```

Recogemos los valores obtenidos en un histograma.

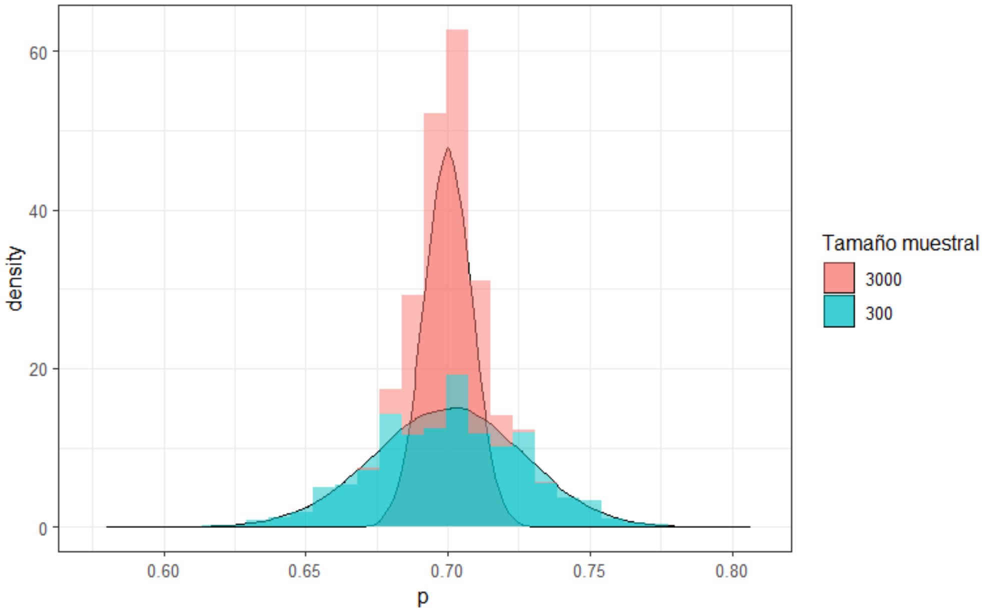


1.5) Realización del experimento: conclusiones

- La enorme mayoría de las muestras de 3000 individuos proporcionan una tasa de partición muy próxima a la de la población.

  → **El riesgo** de cometer un error superior a ±2 puntos, al coger **una** muestra de 3000 individuos es muy pequeño (y asumible. . .)

- Si nos limitamos a muestras de 300 individuos, ¿qué esperáis?



1.6) En la práctica

Usamos las distribuciones muestrales

- Las empresas de sondeos no se basan en simulaciones sino en cálculos teóricos.
- Experimento aleatorio: escoger al azar una muestra de 3000 personas dentro de una población de 37 000 000, con una tasa de participación p.
- Llamamos a $\hat{p}$ la variables aleatoria: proporción de "1"s en la muestra escogida.
- ¿Cuál es la distribución de valores de $\hat{p}$?

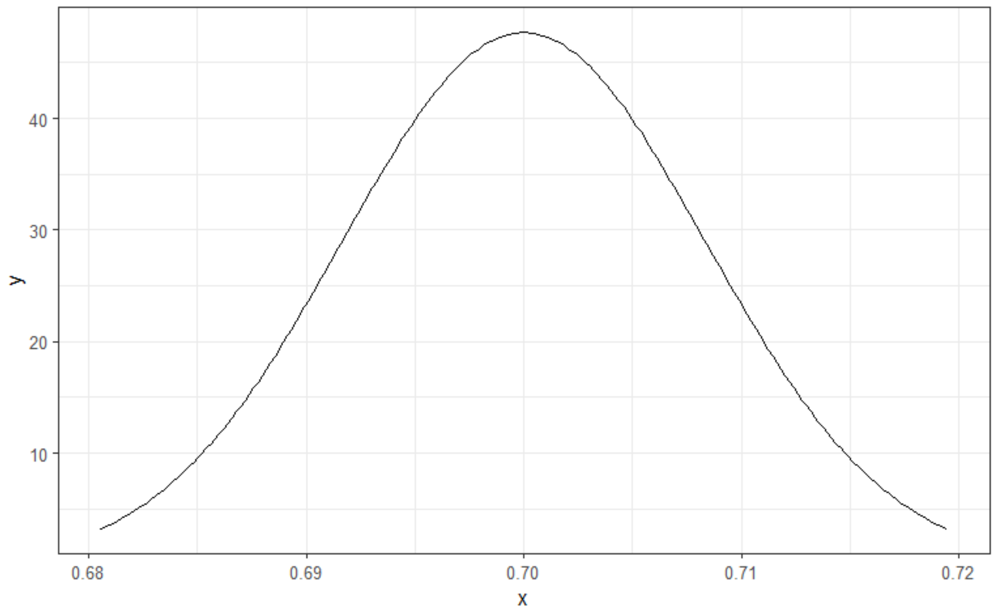
$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

  Es lo qe llamamos la distribución muestral de $\hat{p}$.

1.7) Uso de la distribución muestral

La distribución muestral de $\hat{p}$:

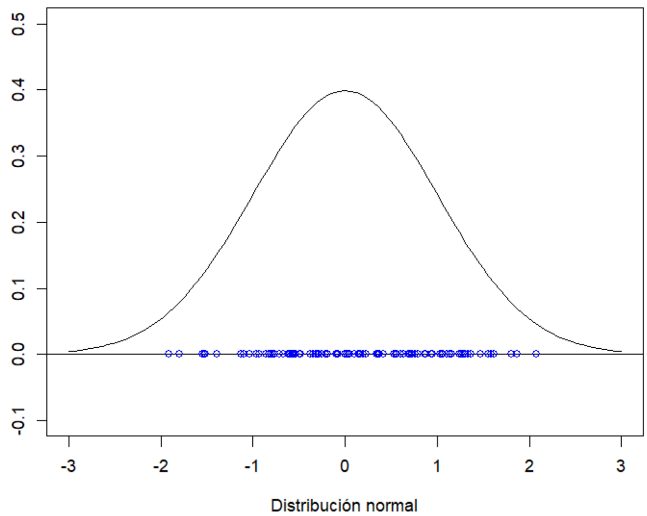
Es la distribución esperada de los valores de $\hat{p}$ respecto a todas las muestras de ese tamaño que podría extraer.

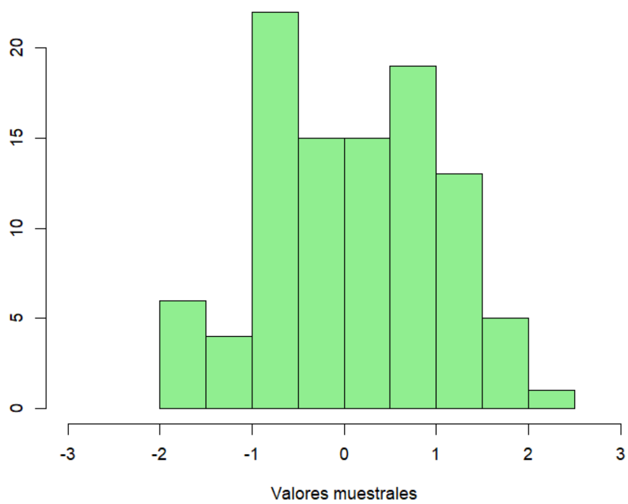


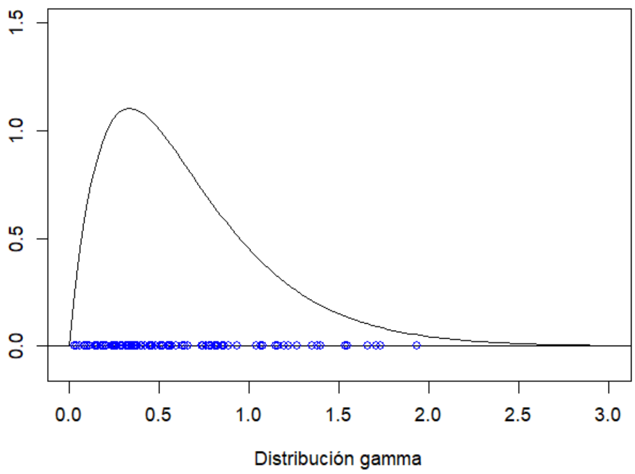
1.8) Antes de extraer una muestra:

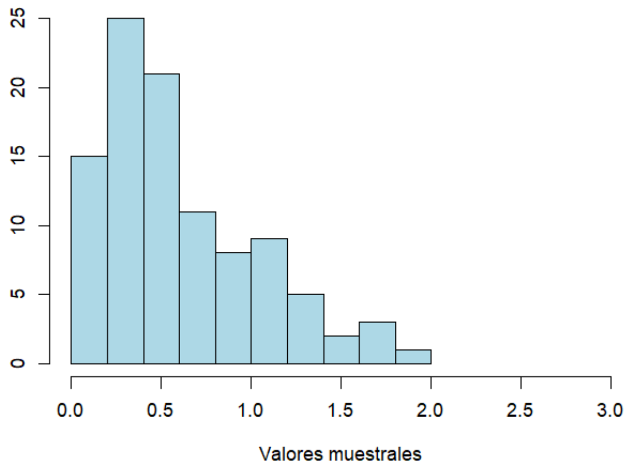
- ¿Es suficiente el tamaño de la muestra para el riesgo asumible y la precisión requerida?
- Una vez extraida la muestra:
  - → ¿Puedo dar un margen de error?
  - → ¿Puedo de decidir si p poblacional es, por ejemplo, mayor que un valor dado?

1.9) Otro ejemplo: valores muestrales de una distribución normal









1.10) Un resultado importante

Ley (débil) de los grandes números

Sea X una variable aletoria y $g(X)$ una variable aleatoria transformada de X, con esperanza y momento de orden 2 finitos. Supongamos $X_1, X_2, \ldots, X_n, \ldots$ una sucesión de variables aleatorias (vv.aa) independientes con la misma distribución que X, entonces

$$\lim_{n \to +\infty} P\left[\left|\frac{\sum_{i=1}^{n} g(X_i)}{n} - E[g(X)]\right| < \varepsilon\right] = 1, \text{ para todo } \varepsilon > 0.$$

1.11) Algunos términos

Definición

- Sea una variable aleatoria X. Consideramos n variables aleatorias independientes e idénticamente distribuidas $X_1, X_2, \ldots, X_n$, que se distribuyen como X. La variable aleatoria multidimensional $(X_1, X_2, \ldots, X_n)$ es una muestra aleatoria simple (m.a.s) de X.
- Cualquier cantidad calculada a partir de las observaciones de un muestra: estadístico.
- Experimento aleatorio: extraer una muestra. Consideramos un estadístico como una variable aleatoria. Nos interesa conocer la distribución del estadístico: distribución muestral.

1.12) Ejemplos de estadísticos

- Proporción muestral: $\hat{p}$
- Media muestral: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
- Desviación típica muestral: $S_X = \sqrt{\frac{1}{n+1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$

1.13) La media muestral

Contexto

Estudiamos una variable X cuantitativa.

- Estamos interesados en μ, el centro de la distribución de X.
- Extraemos una muestra de tamaño n:

$$x_1, x_2, \dots, x_n$$

- Calculamos su media $\overline{x}$ para aproximar μ.
- ¿Cuál es la distribución muestral de $\overline{X}$?

Ejemplo

- Quiero medir una cantidad. Hay variabilidad en las mediciones.
- Introduzco una variable aleatoria X="valor proporcionado por el aparato".
- μ representa el centro de los valores.
- Extraigo una muestra de tamaño 5 del valor de X

1.13.1) Esperanza y varianza de la media muestral

Llamamos $\mu = E[X]$ y $\sigma^2 = \mathrm{Var}(X)$.

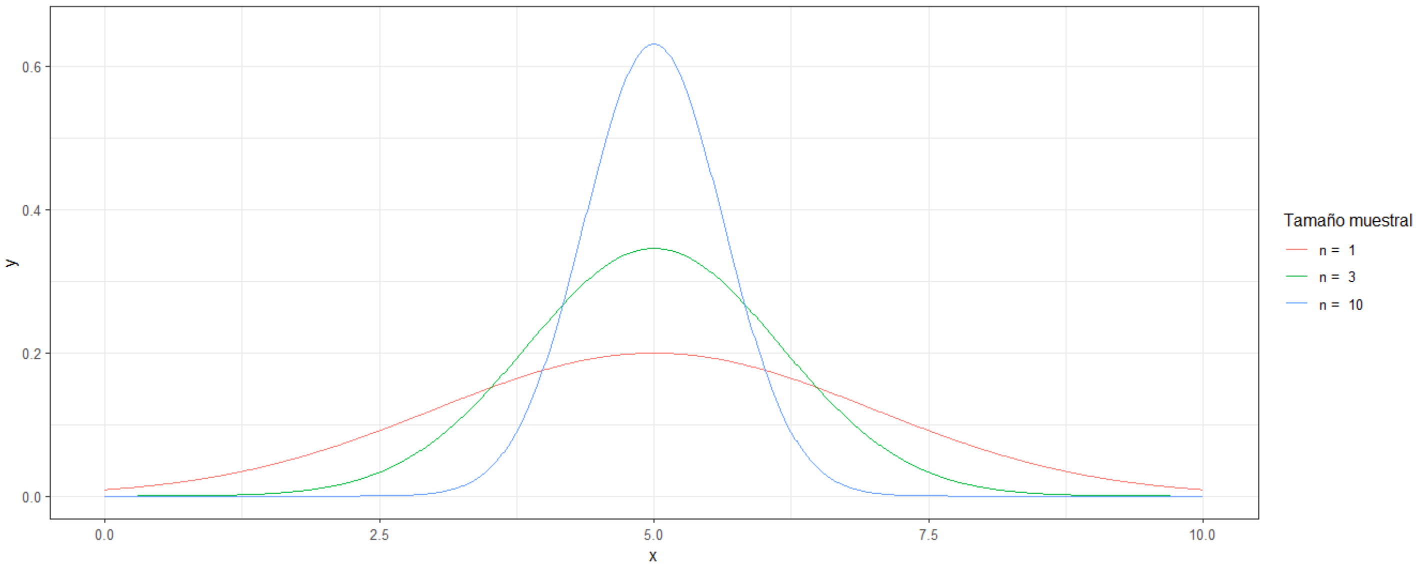
- Tenemos

$$\boxed{E[\overline{X}] = \mu.}$$

  → Es decir que el centro de la distribución muestral de $\overline{X}$ coincide con el centro de la distribución X.

- Tenemos $\mathrm{Var}(\overline{X}) = \dfrac{\sigma^2}{n}$, es decir, la dispersión de la distribución muestral de $\overline{X}$ es $\sqrt{n}$ veces más pequeña que la dispersión inicial de X.

Ilustración: X inicial, $\overline{X}$ con $n = 3, \overline{X}$ con $n = 10$.

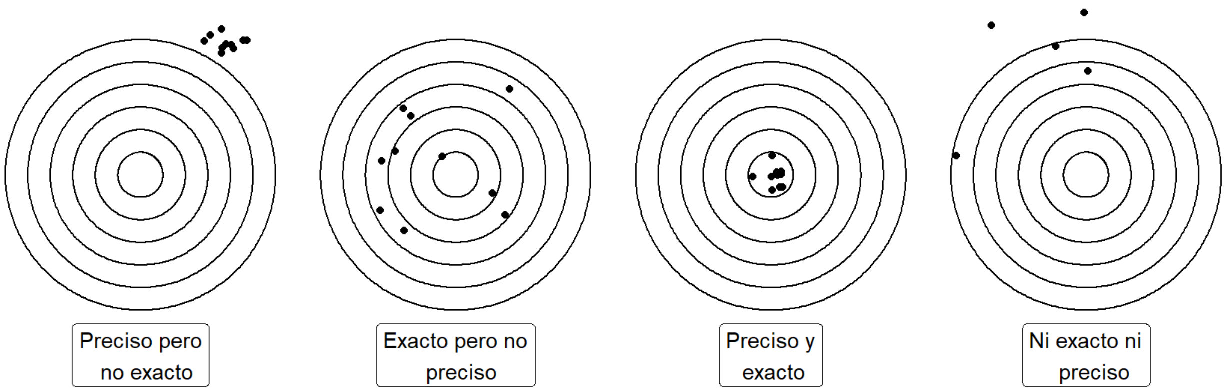


1.14) Consecuencia práctica

Aparato de medición

- Experimento: llevar a cabo una medición con un aparato.
- Variable aleatoria X: "valor propocionado por el aparato".
- $E[X]$: centro de la distribución de los valores proporcionados por el aparato.
  - → Lo deseable: $E[X]$=valor exacto de la cantidad que buscamos medir.
  - → En este caso, decimos: el aparato es exacto.
- σ_X: dispersión de la distribución de los valores proporcionados por el aparato.
  - → Lo deseable: σ_X pequeño.
  - → En este caso, decimos: el aparato es preciso.

1.14.1) Analogía con una diana



1.15) Varianza muestral

Si $(X_1, X_2, \ldots, X_n)$ es una muestra aleatoria simple de X, definimos la varianza muestral S_n^2 como

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2.$$

Fórmula alternativa para S_n^2:

$$S_n^2 = \frac{n}{n-1}\left(\overline{X^2}_n - (\overline{X}_n)^2\right),$$

donde $\overline{X^2}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2$.

1.15.1) Dos apuntes

En algunos textos en castellano:

Se suele llama S_n^2 cuasi-varianza muestra, reservando el término varianza muestral para la cantidad $\dfrac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$.

En estas fórmulas:

Omitimos, si no hay confusión posible, el subíndice n, escribiendo S^2, $\overline{X} = \sum_{i=1}^{n} X_i$ y $\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

1.16) Esperanza de la varianza muestral

Proposición

Si $(X_1, X_2, \ldots, X_n)$ es una muestra aleatoria simple de X con varianza σ_X^2,

$$E[S_n^2] = \sigma_X^2.$$

1.17) Distribuciones muestrales de $\overline{X}$ y S^2

Tened en cuenta

- Los resultados anteriores sobre $E[\overline{X}]$ y $\sigma_{\overline{X}}$ son válidos sea cual sea el modelo escogido para la distribución de X.
- Si queremos decir algo más preciso sobre la distribución de $\overline{X}$ (densidad, etc…) necesitamos especificar la distribución de X.
- En el caso en que la variable X siga una distribución normal, el **teorema de Fisher** analiza cómo se comportan los estadísticos anteriores y nos permiten establecer una serie de consecuencias que serán utilizadas posteriormente en los temas de intervalos de confianza y de constrastes de hipótesis.

1.18) Distribución de $\overline{X}$ y S^2 para una m.a.s. de una distribución normal

Teorema de Fisher

Consideramos una muestra aleatoria simple de una variable aleatoria X con distribución normal $\mathcal{N}(\mu, \sigma^2)$, entonces se verifica:

1) $\overline{X}_n$ y S_n^2 son dos variables aleatorias independientes.
2) $\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$
3) $\dfrac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$.

1.19) Recordatorio: distribución χ^2 con p grados de libertad

La distribución χ^2.

Para $p \in \mathbb{N}^+$, la función de densidad de la distribución χ^2 es igual a

$$\frac{1}{\Gamma\left(\frac{p}{2}\right) 2^{\frac{p}{2}}} \cdot x^{\frac{p}{2}-1} e^{\frac{x}{2}}, \quad \text{si } x > 0,$$

donde Γ denota la función Gamma (Nota: para cualquier real $\alpha > 0, \Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t}\, \mathrm{d}t$).

Caracterización de la χ^2

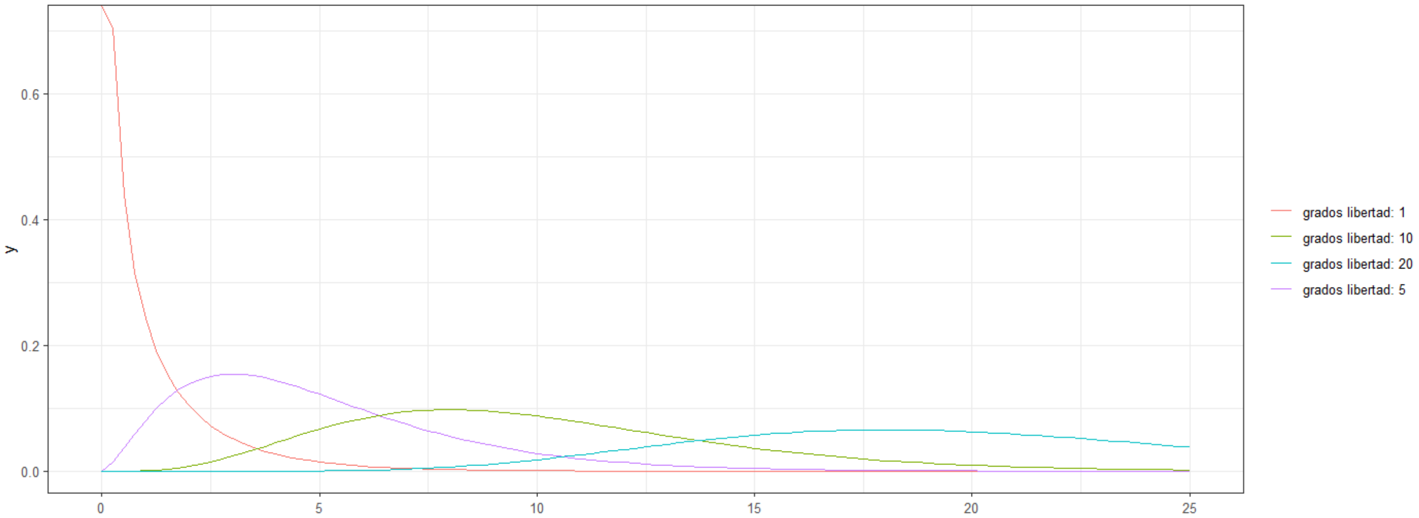
Si $Z_1, \dots, Z_p$ son p variables aleatorias independientes, con $Z_i \sim \mathcal{N}(0,1)$, entonces la variable aleatoria X definida como

$$X = Z_1^2 + \cdots + Z_p^2 = \sum_{i=1}^{p} Z_i^2$$

tiene una distribución χ^2 con p grados de libertad.

- ¿Cómo es su función de densidad?

Depende de los grados de libertad



1.20) Distribución t-Student

Hemos visto, si X es Normal:

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1).$$

Si queremos centrarnos en μ es natural sustituir en ella σ por S_n.

Proposición

Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de una población $\mathcal{N}(\mu, \sigma^2)$,

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

tiene por densidad

$$f_{n-1}(t) \propto \frac{1}{\left(\frac{1+t^2}{n-1}\right)^{\frac{n}{2}}}, \quad -\infty < t < \infty, \tag{1}$$

La distribución que admite esta densidad se llama distribución t-Student con $n-1$ grados de libertad. Escribimos $T \sim t_{n-1}$.

Su densidad

La función de densidad de un t-Student con k grados de libertad:

$$f_k(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \cdot \frac{1}{\sqrt{k\pi}} \cdot \frac{1}{\left(\frac{1\tau^2}{k}\right)^{\frac{k+1}{2}}}, \quad -\infty < t < \infty,$$

donde Γ denota la función Gamma.

Caracterización de la t-Student como cociente

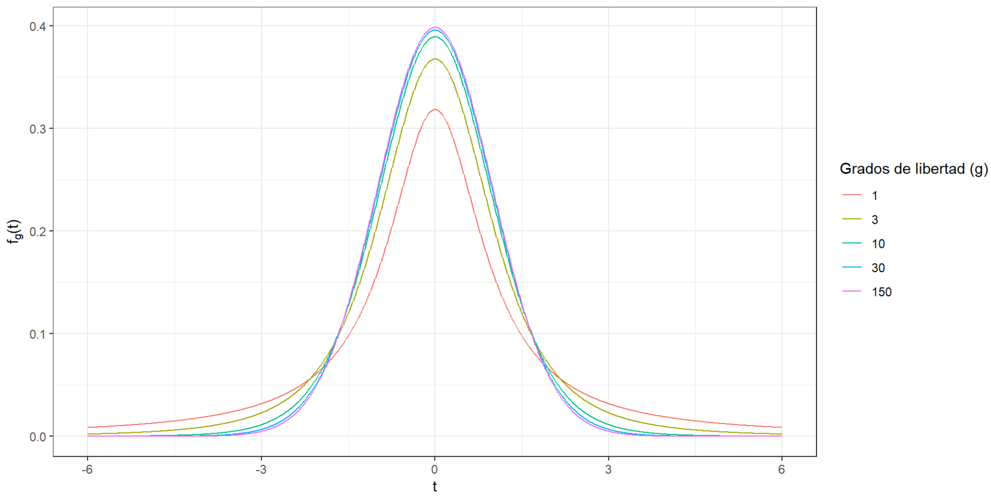
Si Z e Y son dos variables aleatorias independientes, con $Z \sim \mathcal{N}(0,1)$ e $Y \sim \chi^2_p$, el cociente

$$T = \frac{Z}{\sqrt{\frac{Y}{p}}} \sim t_p,$$

donde t_p denota la t-Student con p grados de libertad.

- ¿Cuál es la forma de la densidad de una t-Student?

Tiene colas más pesadas que una normal



1.21) Distribución F de Snedecor para el cociente de varianzas

Proposición

Consideremos U_1 y U_2 dos variables aleatorias independientes con distribución χ^2 con p_1 y p_2 grados de libertad, respectivamente.

El cociente $F = \frac{\frac{U_1}{p_1}}{\frac{U_2}{p_2}}$ admite la densidad

$$f_F(x) = \frac{\Gamma\left(\frac{p_1+p_2}{2}\right)}{\Gamma(p_1)\Gamma(p_2)} \left(\frac{p_1}{p_2}\right)^{p_1} \frac{x^{\frac{p_1}{2}-1}}{\left(1+\frac{p_1}{p_2}x\right)^{\frac{p_1+p_2}{2}}}.$$

Esta distribución se llama F de Snedecor p_1 y p_2 grados de libertad y escribimos $F \sim F_{p_1,p_2}$.

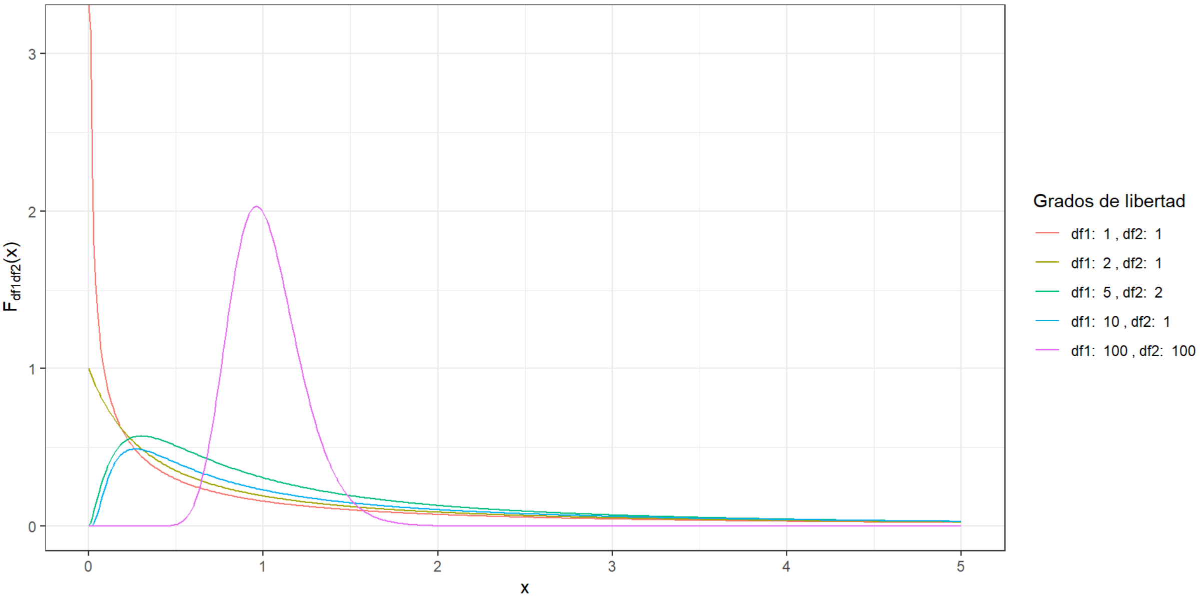
Consecuencia

Consideremos X e Y variables aleatorias normales independientes con varianzas σ_X^2 y σ_Y^2, así como $X_1, \dots, X_{n_x}$ e $Y_1, \dots, Y_{n_Y}$ dos muestras aleatorias simples de X e Y, respectivamente. Deducimos que

$$\frac{\frac{S_X^2}{\sigma_X^2}}{\frac{S_Y^2}{\sigma_Y^2}} \sim F_{n_X-1, n_Y-1}.$$

- ¿Cuál es la forma de la densidad de una F de Snedecor?

Depende mucho de los grados de libertad



1.22) Si la distribución de X no es Normal

No podemos decir nada en general, excepto si n es grande...

Teorema Central del Límite

Si n es *"suficientemente"* grande, se puede aproximar la distribución de $\overline{X}$ por una Normal con media μ y varianza $\frac{\sigma^2}{n}$:

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ aproximadamente.}$$

Formulación matemática

El resultado anterior se taduce por una convergencia de la sucesión de las variables aleatorias $\left(\overline{X}_n\right)_n$ en distribución cuando $n \to \infty$.

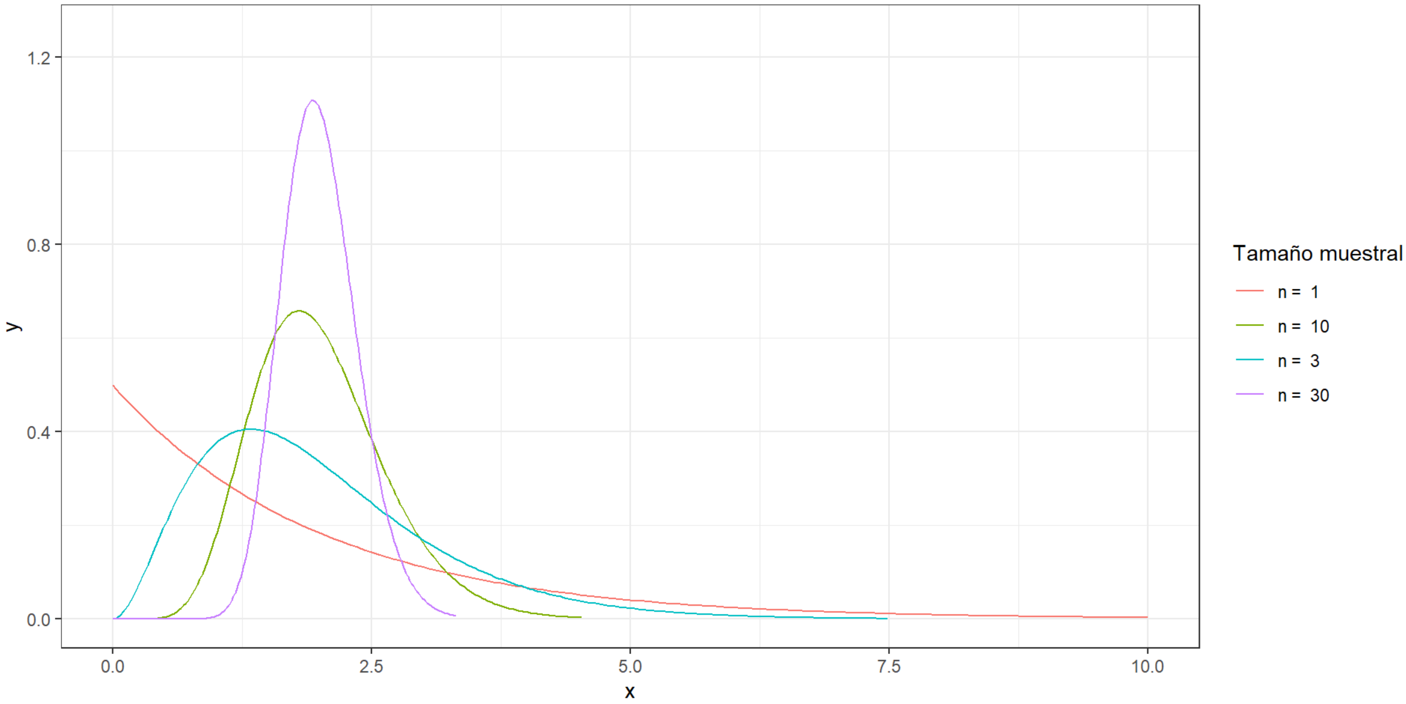
- ¿Cuándo considerar que n es grande?

Depende de la forma de la distribución de X:

- Si X casi Normal: n pequeño es suficiente.
- Si X es muy asimétrico: n mucho más grande necesario.

En general, se suele considerar $n \geq 30$ suficiente...

Ilustración, X inicial $\sim \text{Exp}(\lambda = 0.5)$, $\overline{X}$ **con** $n = 3, 10$ **y** $n = 30$



1.23) Distribución muestral de la proporción muestral

Contexto

- Hay situaciones donde X toma el valor 0 o 1, con probabilidades $1-p$ y p, respectivamente.
- Por ejemplo, el siguiente experimento: escoger una pieza en la producción. $X = 1$ si es defectuosa, $X = 0$ si es correcta.
- Repetimos n veces el experimento. Obtenemos

$$x_1 = 1, x_2 = 0, x_3 = 0, x_4 = 1, x_5 = 0, \ldots, x_n$$

  Contamos N el número de "1"s.
- La proporción muestral es:
$$\hat{p} = \frac{N}{n}.$$

Distribución exacta de $\hat{p}$

¿Cuál es la distribución de N?

- Experimento simple con situación dicotómica, repetimos n veces ...

$$N = \mathcal{B}(n, p).$$

- POdemos usar esta distribución para hacer cálculos exactos...

Distribución aproximada de $\hat{p}$

Tenemos que $N = X_1 + X_2 + \cdots + X_n$, por lo tanto

$$\hat{p} = \frac{N}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \overline{X}.$$

Por el Teorema Central del Límite:

$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right) \text{ aproximadamente.}$$

Podemos usar esta distribución para hacer cálculos aproximados.

1.24) Simulación y método de Monte-Carlo

Motivación

En muchas situaciones, la capacidad de simular valores de las distribuciones de interés puede resultar útil calcular o estimar cantidades relevantes para la inferencia sobre la distribución de $X_1, X_2, \ldots, X_n$.

¿Qué es ser capaz de simular valores de una distribución f?

- Se refiere a la posibilidad de producir, para cualquier tamaño k, conjuntos de valores $u_1, u_2, \ldots, u_k$, cuyo comportamiento imita el de k realizaciones aleatorias independientes de la distribución f.
- Quiere ecir que las propiedades del conjunto generado lo hacer indistinguible, si le aplicamos tests de independencia o de bondad de ajuste, de k realizaciones independientes de f.

Simulación y método de Monte-Carlo

- Como hemos visto en los primeros ejemplos, gráficos como el histograma de frecuencias se comportan como la función de densidad de la variable de la que provienen las observaciones. También se pueden utilizar gráficos como la función de distribución empírica que veremos más adelante.
- Como consecuencia, dado un estadístico, si podemos obtener un número grande de observaciones del mismo podemos a través de alunos gráficos obtener información sobre su distribución.
- Podemos generar esas observaciones a través de lo que se conoce como simulaciones, observaciones generadas mediante algún algoritmo.
- Esta metodología que se puede aplicar en muchas otras situaciones se conoce como el método de Monte-Carlo.

1.24.1) Muestreo de Monte-Carlo para aproximar esperanzas

Ley de los grandes números

Consideremos una muestra aleatoria simple $X_1, X_2, \ldots, X_n$ de una distribución X. Para cualquier función g que cumple $E[g^2(X)] < +\infty$, tenemos que, con probabilidad 1,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} g(X_i) = E[g(X)].$$

Ejemplos

- $g(x) = x : \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = E[X]$, es decir $\overline{X}_n \longrightarrow E[X] = \mu_X$.

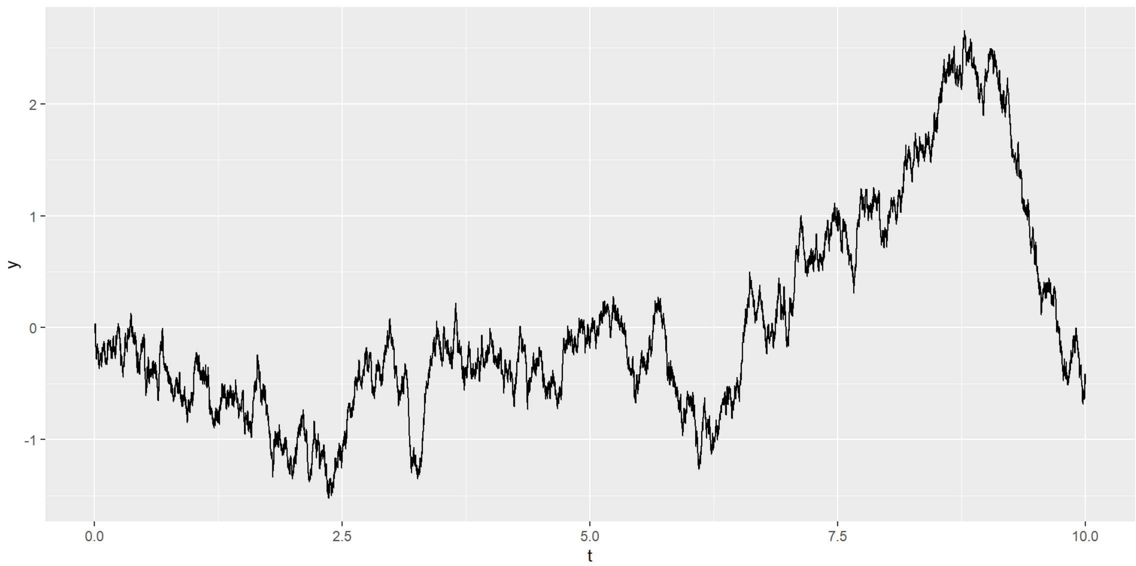
- $g(x) = (x - \mu_X)^2$: $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)^2 = E[(X - \mu_X)^2] = \mathrm{Var}(X)$.
- Si combinamos las dos convergencias anteriores: $\lim_{n\to\infty} \frac{1}{n}(X_i - \overline{X}_n)^2 = \mathrm{Var}(X)$, es decir, $S_n^2 \longrightarrow \mathrm{Var}(X)$.
- $g(x) = 1_{x\leq q}$: $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} 1_{X_i\leq q} = P(X \leq q)$, es decir, que las frecuencias acumuladas relativas convergen hacia la probabilidad asociada.

Aplicaciones

Ejemplo: el movimiento Browniano

- Es proceso muy usado en la predicción de precios (opciones) en matemáticas financieras.
- Una caracterización simplificada: es la suma infinitesimal de pequeñas contribuciones normales independientes.
- Para cualquier $t, W_t \sim \sum_{i=1}^{\frac{t}{h}} \sqrt{h}\cdot Z_i$, donde h es el paso infinitesimal y Z_i son normales estándares independientes e idénticamente distribuidos

1.25) Movimiento Browniano



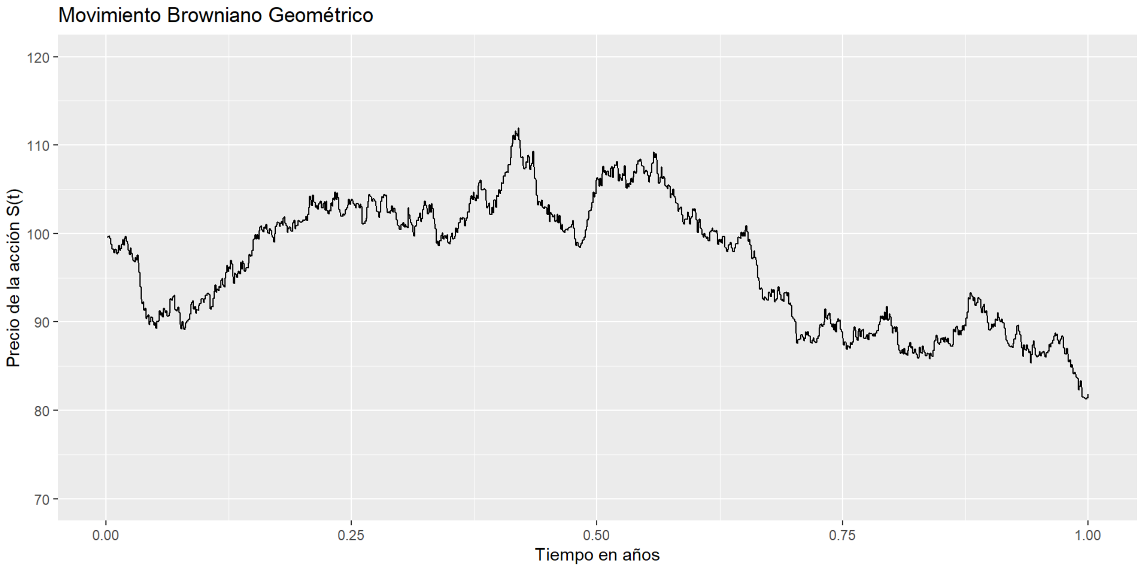
1.26) En finanzas, el modelo de Black-Scholes

El movimiento Browniano Geométrico:

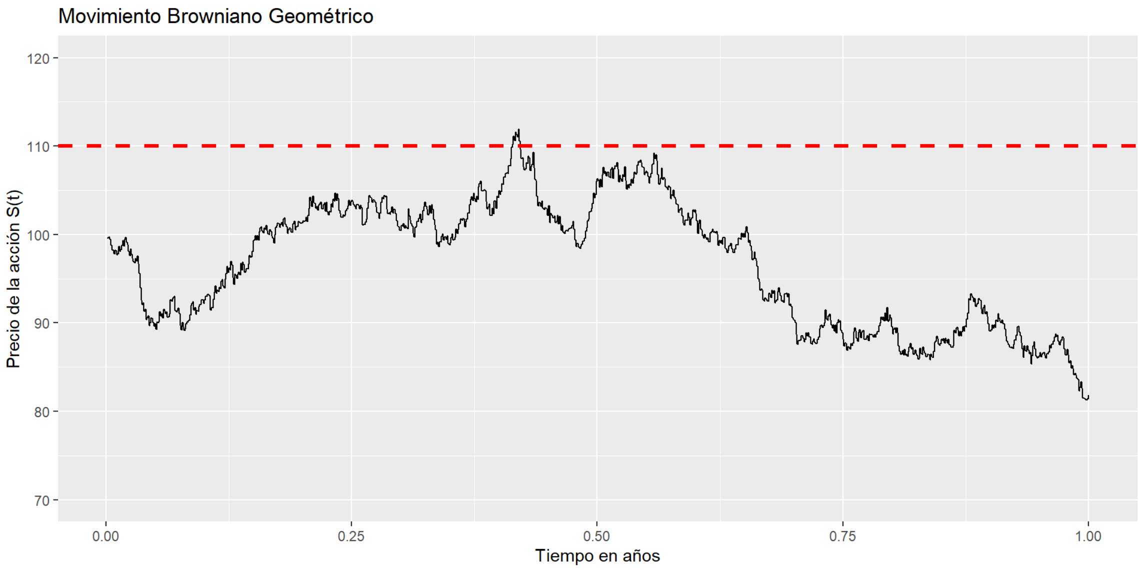
Lo propusieron Merton, Scholes y Black como modelo teórico para precios de acciones:

$$S_t = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right).$$

- S_0: precio inicial de la acción.
- μ: el drift.
- σ^2: la volatilidad

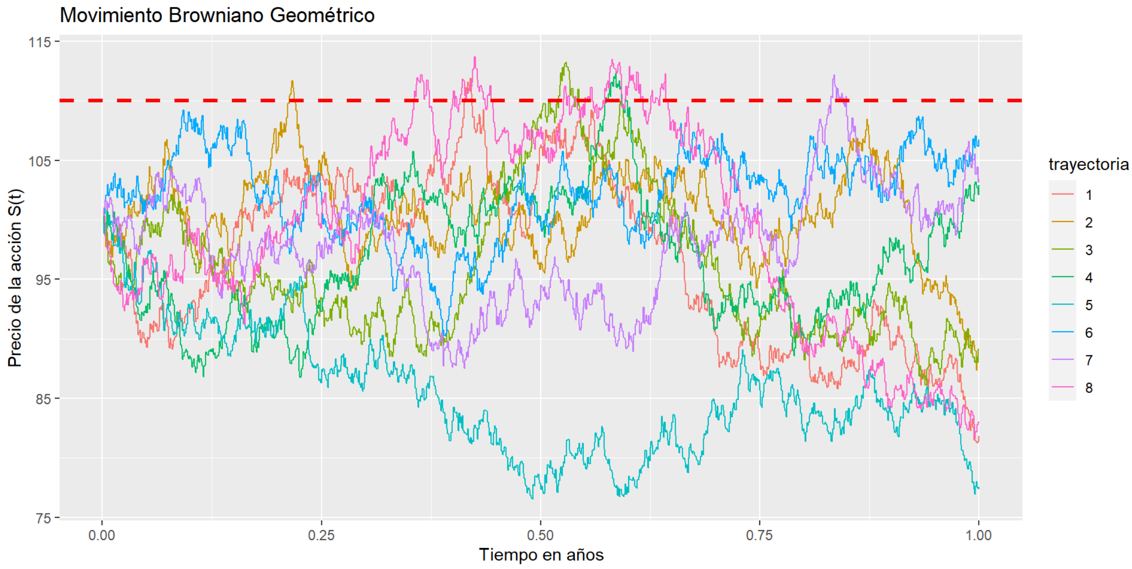


¿Cuán observaremos $S(t) \geq 110$€?



- Podemos simular muchas trayectorias del Movimiento Browniano Geométrico y observa qué pasa con el tiempo en el que supera el umbral 110.
- Esto es Monte-Carlo.
- Luego podremos obtener indicadores de la distribución de este tiempo.

- Representamos las 8 primeras



- ¿Cuál es la probabilidad de que alcance el umbral antes de un año?

Por Monte-Carlo:

Simulamos 1000 trayectorias hata el año y aproximamos la probabilidad que nos interesa por la porporción de trayectorias que alcanzan los 110 euros.

1.27) Simulación y método de Monte-Carlo

Simulación de variables aleatorias

- Lo que nos planteamos ahora es qué algoritmo podemos utilizar para generar observaciones (simulaciones) de una variable aleatoria.
- Para un gran número de varaibles aleatorias y modelos probabilísticos, estas simulaciones ya están incorporadas en los paquetes y lenguajes de programación con enfoque matemático y estadístico, como `R`.
- Aquí describiremo uno de los métodos más básicos, basado en la inversa de la función de distribución.
- El método de la función inversa es uno de los principales métodos de simulación de variables aleatorias.
- Está basado en la inversa de la función de distribución de una variable aleatoria.
- No todas las funciones de distribucion admiten inversa, como la conocemos usualmente. Para ello es necesaria que sea continua y estrictamente creciente.

Inversa generalizada de la función de distribución o función cuantil

Dada una función de distribución F, su función cuantil (inversa generalizada) se define como

$$F^{-1}(p) = \inf\{x|F(x) \geq p\}, \text{ para todo } p \in (0,1).$$

Propiedades de la función cuantil

1) $F(F^{-1}(p)) \geq p$, para todo $p \in (0,1)$.
2) $F^{-1}(F(x)) \leq x$, para todo x donde $F(x) > 0$.
3) Si $U \sim U(0,1)$ entonces $F^{-1}(U)$ tiene como distribución F.
4) Dada una variable aleatoria X, con media finita y función de distribución F se verifica que

$$E[X] = \int_0^1 F^{-1}(u)\,\mathrm{d}u.$$

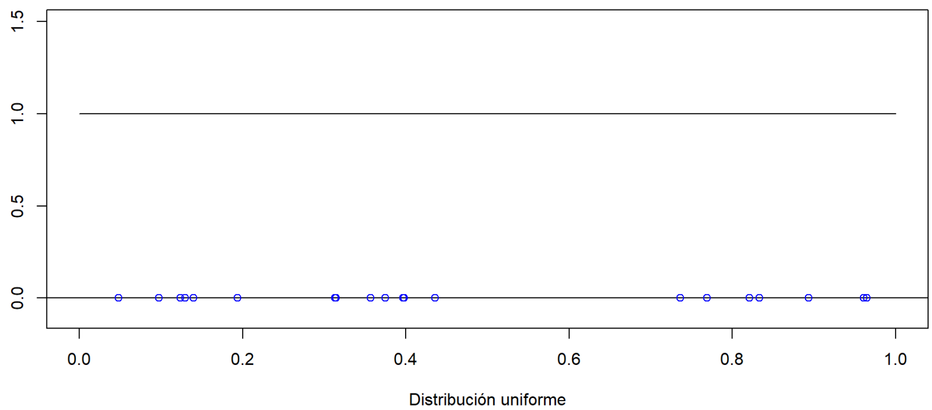
Método de la función inversa

- Dada una variable aleatoria X con función de distribución F, si generamos observaciones independientes $U \sim U(0,1)$, $u_1, u_2, \ldots, u_n$, entonces los valores

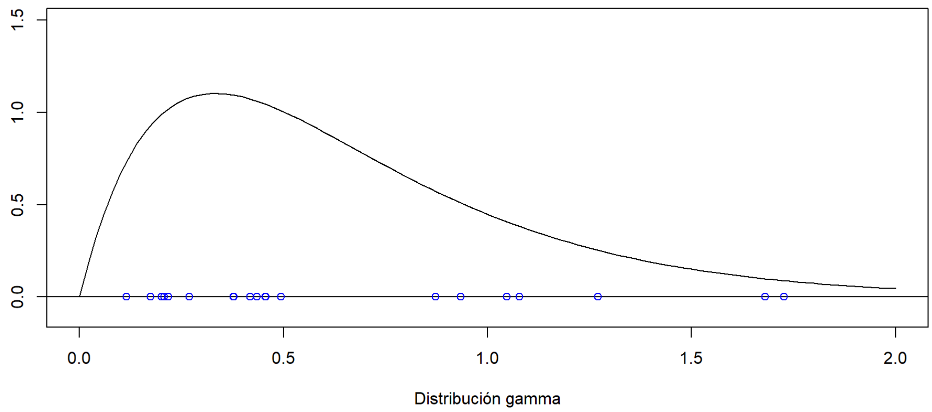
$$x_1 = F^{-1}(u_1), x_2 = F^{-1}(u_2), \ldots, x_n = F^{-1}(u_n)$$

constituyen una observación de la muestra aleatoria simple de tamaño n de la variable aleatoria X.

Generación de valores de una distribución uniforme



Generación de valores de una distribución gamma



Transformación de una variable aleatoria

Planteamiento del problema

En muchas ocasiones, tenemos una variable aleatoria continua X, de la que conocemos la función de densidad, pero nos interesa conocer cómo se comporta una transformación de X, $Y = \varphi(X)$.

Teorema

Sea X una viarble aleatoria con función de densidad f_X definida en un intervalo abierto $(a,b) \subseteq \mathbb{R}$. Sea $\varphi : (a,b) \to \mathbb{R}$ tal que:

- Es continua.
- Es estrictamente creciente o decreciente.
- φ^{-1} es diferenciable.

Entonces, la variable aleatoria $Y = \varphi(X)$ tiene función de densidad

$$f_Y(y) = \begin{cases} f_X(\varphi^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}x}(\varphi^{-1}(y)) \right|, & \text{si } y \in \varphi(a,b) \\ 0 & \text{en otro caso} \end{cases}$$

Función característica

Definición

Sea X una variable aleatoria cualquiera. La función característica de X se define como

$$\phi(t) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} \mathrm{d}F(x)$$

en donde $i = \sqrt{-1}$

Nota: $\phi(t) = \int_{-\infty}^{\infty} xos(tx)\mathrm{d}F(x) + i\int_{-\infty}^{\infty} \sin(tx)\mathrm{d}F(x)$.

Propiedades

- Existencia: $|\phi(t)| \leq 1$, para todo $t \in \mathbb{R}$.
- Si X e Y son variables aleatorias independientes, entonces

  $$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t),$$

  para todo $t \in \mathbb{R}$.

Desigualdades

Desigualdad de Markov

Si Z es una variable aleatoria no negativa con medi finita $E[Z]$ y $\varepsilon > 0$, entonces

$$\varepsilon \Pr[Z \geq \varepsilon] = \varepsilon \int_{[\varepsilon,\infty)} \mathrm{d}F_Z(x) \leq \int_{[\varepsilon,\infty)} x\mathrm{d}F_Z(x) \leq \int_{[0,\infty)} x\mathrm{d}F_Z(x) = E(Z)$$

(donde $F_Z(x) = \Pr[Z \leq x]$ es su función de distribución), es decir

$$\Pr[Z \geq \varepsilon] \leq \frac{E[Z]}{\varepsilon}.$$

Desigualdad de Chebyshev

Si X es una variable aleatoria con media finita $\mu = E[X]$ y varianza $\sigma^2 = \text{Var}(X) > 0$, entonces tomando $Z = \dfrac{(X-\mu)^2}{\sigma^2} \geq 0$ y aplicando la desigualdad de Markov, tenemos

$$\Pr\left[\frac{(X-\mu)^2}{\sigma^2} \geq \varepsilon\right] \leq \frac{1}{\varepsilon}$$

para todo $\varepsilon > 0$.
También se puede escribir como

$$\Pr[(X-\mu)^2 < \varepsilon\sigma^2] \geq 1 - \frac{1}{\varepsilon}$$

o como

$$\Pr[|X-\mu| < r] \geq 1 - \frac{\sigma^2}{r^2},$$

para todo $r > 0$.

Hoja de ejercicios Tema 1: Muestreo y distribuciones muestrales

1) Sea X variable aleatoria con distribución Bernoulli, de parámetro $p(X \sim b(p))$ y sea X_1, X_2, X_3 una muestra aleatoria simple (m.a.s) de X. Se pide:

a) Estudiar la distribución del vector (X_1, X_2, X_3).

Dado que X_1, X_2, X_3 son una muestra aleatoria simple de $X \sim \text{Bernoulli}(p)$, entonces las siguientes propiedades son ciertas:

1) X_1, X_2, X_3 son independientes e idénticamente distribuidas.

2) Cada $X_i \sim \text{Bernoulli}(p)$, es decir, la probabilidad de éxito $P(X_i = 1) = p$ y $P(X_i = 0) = 1 - p$.

El vector (X_1, X_2, X_3) tiene una **distribución multinomial** con 2 posibles resultados (0 o 1) para cada componente. La distribución conjunta es:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = p^{x_1+x_2+x_3}(1-p)^{3-(x_1+x_2+x_3)},$$

donde $x_1, x_2, x_3 \in \{0, 1\}$.

Esto corresponde a la **distribución conjunta** de 3 variables Bernoulli independientes.

b) Estudiar la distribución en el muestreo del estadístico $\dfrac{X_1 + X_2 + X_3}{3}$.

Es estidístico $\overline{X} = \dfrac{X_1 + X_2 + X_3}{3}$ es el **promedio muestral** de las 3 variables. Para analizar su distribución:

1) $S = X_1 + X_2 + X_3$ sigue una **distribución binomial** porque es la suma de $n = 3$ variables Bernoulli independientes:

$$S \sim B(n = 3, p).$$

La función de probabilidad de S es:

$$P(S = k) = \binom{3}{k} p^k (1-p)^{3-k}, \quad k = 0, 1, 2, 3.$$

2) El estadístico $\overline{X} = \dfrac{S}{3}$ simplemente escala los valores posibles de S dividiéndolos por 3. Los valores posibles de $\overline{X}$ son:

$$\overline{X} \in \left\{0, \frac{1}{3}, \frac{2}{3}, 1\right\}.$$

3) La probabilidad de cada valor de $\overline{X}$ es proporcional a la probabilidad de los valores correspondientes de S:

$$P\left(\overline{X} = \frac{k}{3}\right) = P(S = k) = \binom{3}{k} p^k (1-p)^{3-k}, \quad k = 0, 1, 2, 3.$$

Por lo tanto, la distribución de $\overline{X}$ es discreta y está determinada por la distribución binomial de S.

2) Sea X variable aleatoria con distribución $\text{Exp}(\lambda)$. Sea $(X_1, \ldots, X_n)$ una m.a.s de X, estudiar la distribución en el muestre de $S = \sum_{j=1}^{n} X_j$.

Dado que $X \sim \text{Exp}(\lambda)$ (exponencial con parámetro $\lambda > 0$), y que $(X_1, \ldots, X_n)$ es una muestra aleatoria simple (m.a.s) de X, podemos analizar la distribución del estadístico $S = \sum_{j=1}^{n} X_j$.

Propiedades relevantes:

1) **Distribución de la suma de variables exponenciales independientes:** Si $X_1, \ldots, X_n$ son variables aleatorias

independientes e idénticamente distribuidas ($X_i \sim \text{Exp}(\lambda)$), entonces la suma:

$$S = \sum_{j=1}^{n} X_j$$

sigue una distribución **Gamma** con parámetros n y λ. Esto se denota como:

$$S \sim \text{Gamma}(n, \lambda),$$

donde:

- n es el parámetro de forma.
- λ es el parámetro de escala.

Distribución Gamma:

La función de densidad de probabilidad de una variables aleatoria $S \sim \text{Gamma}(n, \lambda)$ está dada por:

$$f_S(s) = \begin{cases} \dfrac{\lambda^n}{\Gamma(n)} s^{n-1} e^{-\lambda s} & s > 0 \\ 0 & s \leq 0 \end{cases}$$

donde:

- $\Gamma(n)$ es la función gamma (para $n \in \mathbb{N}, \Gamma(n) = (n-1)!$).
- s^{n-1} y $e^{-\lambda s}$ controlan la forma y decaimiento de la densidad.

Propiedades del estadístico S:

1) **Esperanza ($E[S]$):**

   Si $S \sim \text{Gamma}(n, \lambda)$, entonces:
   $$E[S] = \frac{n}{\lambda}.$$

2) **Varianza** ($\text{Var}(S)$)**:**

   La varianza de S está dada por:
   $$\text{Var}(S) = \frac{n}{\lambda^2}.$$

3) **Caso especial** ($n = 1$)**:**

   Cuando $n = 1$, la distribución Gamma coincide con la distribución exponencial. Es decir:

   $$\text{Gamma}(1, \lambda) = \text{Exp}(\lambda).$$

3) Sea X variable aleatoria con distribución $\mathcal{N}(\mu, \sigma^2)$. Sea $(X_1, \ldots, X_n)$ una m.a.s de X, estudiar la distribución en el muestreo de $S = \sum_{j=1}^{n} X_j$.

Dado que $X \sim \mathcal{N}(\mu, \sigma^2)$, y que $(X_1, \ldots, X_n)$ es una muestra aleatoria simple (m.a.s) de X, las X_i son independientes e idénticamente distribuidas con $X_i \sim N(\mu, \sigma^2)$. Queremos analizar la distribución en el muestreo de $S = \sum_{j=1}^{n} X_j$.

Propiedades relevantes:

1) **Suma de variables normales independientes:** Si $X_1, \ldots, X_n$ son independientes y $X_i \sim \mathcal{N}(\mu, \sigma^2)$, entonces la suma:
   $$S = \sum_{j=1}^{n} X_j$$

sigue una distribución normal:

$$S \sim N(n\mu, n\sigma^2).$$

Derivación:

1) **Esperanza ($E[S]$):** La esperanza de S es la suma de las esperanzas de las X_i:

$$E[S] = E\left[\sum_{j=1}^{n} X_j\right] = \sum_{j=1}^{n} E[X_j] = \sum_{j=1}^{n} \mu = n\mu.$$

2) **Varianza ($\text{Var}(S)$):** La varianza de S es la suma de las varianzas de las X_i, ya que son independientes:

$$\text{Var}(S) = \text{Var}\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} \text{Var}(X_j) = \sum_{j=1}^{n} \sigma^2 = n\sigma^2.$$

3) **Distribución:** Dado que una combinación lineal de variables normales independientes también sigue una distribución normal, se concluye que:

$$S \sim N(n\mu, n\sigma^2).$$

4) Sea X una variable aleatoria con función de densidad:

$$f(x,\theta) = \frac{2x}{\theta} \exp\left(-\frac{x^2}{\theta}\right) \chi_{(0,+\infty)}(x).$$

Obtener la distribución en el muestreo estadístico:

$$T(X_1, X_2, \ldots, X_n) = \sum_{j=1}^{n} X_j^2.$$

Obtener su media y su varianza.

Paso 1: Verificar la distribución de X

La función de densidad de X es:

$$f(x;\theta) = \frac{2x}{\theta} \exp\left(-\frac{x^2}{\theta}\right) \chi_{(0,\infty)}(x).$$

Esta densidad corresponde a una **distribución Rayleigh generalizada** con parámetro de escala θ. Para esta distribución:

- X^2 sigue una distribución exponencial con parámetro $\lambda = \frac{1}{\theta}$.

Entonces:

$$Y = X^2 \sim \text{Exp}\left(\lambda = \frac{1}{\theta}\right).$$

Paso 2: Distribución del estadístico $T = \sum_{j=1}^{n} X_j^2$

Dado que $Y_j = X_j^2 \sim \text{Exp}\left(\frac{1}{\theta}\right)$, y las Y_j son independientes, la suma de n variables exponenciales independientes sigue una distribución **Gamma**.

Por lo tanto, el estadístico:

$$T = \sum_{j=1}^{n} X_j^2 = \sum_{j=1}^{n} Y_j$$

sigue la distribución:

$$T \sim \text{Gamma}\left(n, \lambda = \frac{1}{\theta}\right),$$

donde:

- n es el parámetro de forma.
- $\lambda = \dfrac{1}{\theta}$ es el parámetro de escala.

La densidad de la distribución Gamma es:

$$f_T(t; n, \lambda) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{\Gamma(n)}, \quad t > 0.$$

Paso 3: Esperanza y Varianza del estadístico T

Para una distribución Gamma con parámetros (n, λ), las propiedades son:

1) Esperanza:

$$E[T] = \frac{n}{\lambda}.$$

2) Varianza:

$$\operatorname{Var}(T) = \frac{n}{\lambda^2}.$$

En este caso, como $\lambda = \dfrac{1}{\theta}$:

$$E[T] = \frac{n}{\frac{1}{\theta}} = n\theta,$$
$$\operatorname{Var}(T) = \frac{n}{\left(\frac{1}{\theta}\right)^2} = n\theta^2.$$

5) Sea X una variable aleatoria con función de densidad:

$$f(x, \theta) = \frac{\theta}{(1+x)^{1+\theta}} \chi_{(0,+\infty)}(x).$$

Obtener la distribución en el muestreo del estadístico:

$$T(X_1, X_2, \ldots, X_n) = \frac{\sum_{j=1}^{n} \ln(1 + X_j)}{n}.$$

Obtener su media y su varianza.

1) Identificación de la distribución de X

La función de densidad de X viene dada por:

$$f(x, \theta) = \frac{\theta}{(1+x)^{1+\theta}} \chi_{(0,+\infty)}(x), \quad \theta > 0.$$

Observemos que, para $x > 0$, la forma $\dfrac{1}{(1+x)^{1+\theta}}$ sugiere una transformación logarítmica conveniente:

$$Y = \ln(1 + X).$$

Vamos a encontrar la distribución de Y.

2) Transformación $Y = \ln(1 + X)$

1) Relación entre X y Y:

$$Y = \ln(1 + X) \longleftrightarrow X = e^Y - 1.$$

2) Soporte:

Dado que $x > 0$, entonces $1 + x > 1$ y por ende $Y > \ln(1) = 0$. De modo que $Y \in (0, +\infty)$.

3) Derivada $\frac{\mathrm{d}x}{\mathrm{d}y}$:

$$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{\mathrm{d}}{\mathrm{d}y}(e^Y - 1) = e^Y.$$

4) Función de densidad de Y:

Partiendo de que $f_X(x, \theta)$ es la densidad de X, la densidad de Y se obtiene mediante:

$$f_X(y) = F_X(x(y)) \cdot \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right|.$$

Sustituyendo $x = e^Y - 1$, obtenemos:

$$f_X(e^Y - 1, \theta) = \frac{\theta}{(1 + (e^Y - 1))^{1+\theta}} = \frac{\theta}{(e^Y)^{1+\theta}} = \theta e^{-(1+\theta)Y}.$$

Por último, multiplicamos por $\frac{\mathrm{d}x}{\mathrm{d}y} = e^Y$:

$$f_Y(y) = \left[\theta e^{-(1+\theta)Y}\right] \cdot e^Y = \theta e^{-(1+\theta)Y} e^Y = \theta e^{-\theta Y}, \quad y > 0.$$

Esta es precisamente la **densidad de una Exponencial** con parámetro θ. Por tanto,

$$Y = \ln(1 + X) \sim \mathrm{Exp}(\theta).$$

3) Distribución del estadístico

$$T(X_1, \ldots, X_n) = \frac{1}{n} \sum_{j=1}^{n} \ln(1 + X_j).$$

Definamos

$$Y_j = \ln(1 + X_j).$$

Cada Y_j es $\mathrm{Exp}(\theta)$ y son independientes e idénticamente distribuidas al provenir de una muestra aleatoria simple de X. Entonces

$$T = \frac{1}{n} \sum_{j=1}^{n} Y_j.$$

- La suma $S = \sum_{j=1}^{n} Y_j$ sigue una Gamma(n, θ).
- El estadístico $T = \frac{1}{n} S$ es simplemente la suma escalada por $\frac{1}{n}$.

1) Si $S \sim \Gamma(n, \theta)$, entonces $T = \frac{1}{n} S$ tiene parámetro de forma n y de **tasa** $n\theta$. En otras palabras,

$$T \sim \Gamma(n, n\theta).$$

Explícitamente, su función de densidad es:

$$f_T(t) = \frac{(n\theta)^n}{\Gamma(n)} t^{n-1} e^{-n\theta t}, \quad t > 0.$$

4) Media y varianza de T

Para una $\Gamma(k, \lambda)$, se sabe que:

$$E[W] = \frac{k}{\lambda}, \qquad \mathrm{Var}(W) = \frac{k}{\lambda^2}.$$

En nuestro caso, $T \sim \Gamma(n, n\theta)$. Luego:

1) Media de T:

$$E[T] = \frac{n}{n\theta} = \frac{1}{\theta}$$

2) Varianza de T:

$$\mathrm{Var}(T) = \frac{n}{(n\theta)^2} = \frac{1}{n\theta^2}.$$

6) Sea X una variable aleatoria con función de densidad en todo $\mathbb{R}$:

$$f(x,\theta) = \exp(-(x-\theta))\exp(-\exp(-(x-\theta))).$$

Obtener la distribución en el muestreo del estadístico:

$$T(X_1, X_2, \ldots, X_n) = \frac{\sum_{j=1}^{n} \exp(-X_j)}{n}.$$

Obtener su media y su varianza.

1) Identificar la distribución de X

La función de densidad que se nos da es, para todo $x \in \mathbb{R}$,

$$f(x,\theta) = \exp(-(x-\theta))\exp\left(-\exp(-(x-\theta))\right).$$

Obsérvese que si definimos

$$Z = X - \theta,$$

la densidad Z queda

$$f_Z(z) = \exp(-z)\exp(-e^{-z}), \quad z \in \mathbb{R}.$$

Esta es la **distribución Gumbel** estándar (con parámetro de localización 0 y escala 1) pa el máximo. Por lo tanto, X se distribuye como una Gumble$(\theta, 1)$ con localización θ y escala 1.

En resumen:

$$X \sim \text{Gumble}(\theta,1) \longleftrightarrow Z = X - \theta \sim \text{Gumble}(0,1).$$

2) Analizar el estadístico

$$T(X_1, \ldots, X_n) = \frac{1}{n}\sum_{j=1}^{n} \exp(-X_j).$$

Definamos

$$Y_j = \exp(-X_j).$$

El objetivo es estudiar la distribución de

$$T = \frac{1}{n}\sum_{j=1}^{n} Y_j = \frac{1}{n}\sum_{j=1}^{n} e^{-X_j}.$$

2.1) Distribución de $Y_j = e^{-X_j}$

Usaremos el hecho de que $Z_j = X_j - \theta$ es Gumbel estándar. Entonces

$$X_j = Z_j + \theta \longrightarrow e^{-X_j} = e^{-\theta}e^{-Z_j}.$$

Definamos $W_j = e^{-Z_j}$. Veamos la distribución de W_j:

- Si $Z_j \sim \text{Gumbel}(0,1)$, la función de distribución acumulativa (CDF) de Z_j es

$$F_{Z_j}(z) = e^{-e^{-z}}, z \in \mathbb{R}.$$

- Entonces
$$W_j = e^{-Z_j} > 0.$$
Para $w > 0$,
$$\{W_j \leq w\} \equiv \{e^{-Z_j} \leq w\} \equiv \{Z_j \geq -\ln w\}.$$
Por tanto,
$$F_{W_j}(w) = \Pr(Z_j \geq -\ln w) = 1 - \Pr(Z_j < -\ln w) = 1 - F_{Z_j}(-\ln w).$$
Usando $F_{Z_j}(z) = e^{-e^{-z}}$, se obtiene
$$F_{Z_j}(-\ln w) = e^{-\exp(-(-\ln w))} = e^{-w}.$$
Por lo tanto,
$$F_{W_j}(-\ln w) = e^{-w}, \quad w > 0,$$
que es la función de distribución acumulativa de una **Exponencial**.

En consecuencia,
$$W_j \sim \mathrm{Exp}(1)$$
Como
$$Y_j = e^{-\theta} W_j,$$
entonces $Y - j$ es simplemente W_j escalada por $e^{-\theta}$.

- Si $W_j \sim \mathrm{Exp}(1)$, entonces la variable $c \cdot W_j$ con $c > 0$ es $\mathrm{Exp}\left(\frac{1}{c}\right)$.
- Aquí $c = e^{-\theta} \longrightarrow \frac{1}{c} = e^{\theta}$.

Por tanto:
$$Y_j = e^{-\theta} W_j \sim \mathrm{Exp}(e^{\theta}).$$
Es decir, cada Y_j tiene tasa e^{θ}.

2.2) Distribución de la media muestral T

Dado que los Y_j son independientes e idénticamente distribuidos $\mathrm{Exp}(e^{\theta})$, la suma
$$S = \sum_{j=1}^{n} Y_j$$
sigue una distribución **Gamma** con forma n y tasa e^{θ}; escribimos
$$S \sim \Gamma(n, e^{\theta}).$$
El estadístico
$$T = \frac{S}{n}$$
es simplemente la suma S escalada por $\frac{1}{n}$. Se conoce la siguiente propiedad de la distribución Gamma:

- Si $S \sim \Gamma(n, \lambda)$, entonces $\alpha S \sim \Gamma\left(n, \frac{\lambda}{\alpha}\right)$.

En nuestro caso, $\alpha = \frac{1}{n}$. Por tanto,
$$T = \frac{S}{n} \sim \Gamma\left(n, ne^{\theta}\right).$$

3) Media y varianza de T

Sea $T \sim \Gamma(k, \lambda)$ con $k = n$ y $\lambda = ne^{\theta}$. Recordemos que:

$$E[\Gamma(k,\lambda)] = \frac{k}{\lambda}, \quad \mathrm{Var}[\Gamma(k,\lambda)] = \frac{k}{\lambda^2}.$$

Por tanto:

1) Media de T:

$$E[T] = \frac{n}{ne^{\theta}} = \frac{1}{e^{\theta}}.$$

2) Varianza de T:

$$\mathrm{Var}(T) = \frac{n}{(ne^{\theta})^2} = \frac{n}{n^2 e^{2\theta}} = \frac{1}{ne^{2\theta}}.$$

7) Sea $X_1, X_2, \ldots, X_n$ una m.a.s de una variable X con distribución $\mathcal{N}(\mu, \sigma^2)$. Sean $\overline{X}$ y S^2 su media y cuasi-varianzas muestrales, respectivamente. Sea X_{n+1} una nueva observación de X independiente de $X_1, X_2, \ldots, X_n$. Obtener la distribución en el muestreo del estadístico:

$$\frac{X_{n+1} - \overline{X}}{S}\sqrt{\frac{n}{n+1}}.$$

1) Distribución de $X_{n+1} - \overline{X}$

1) La media muestral $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$ es independiente de X_{n+1} (porque X_{n+1} es una nueva observación independiente).

2) Cada X_i (incluyendo X_{n+1}) se distribuye como $\mathcal{N}(\mu, \sigma^2)$.

3) Se sabe que

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad X_{n+1} \sim \mathcal{N}(\mu, \sigma^2),$$

y son independientes. Por ende,

$$X_{n+1} - \overline{X} \sim \mathcal{N}\left(0, \sigma^2 + \frac{\sigma^2}{n}\right) = \mathcal{N}\left(0, \sigma^2 \cdot \frac{n+1}{n}\right).$$

Es decir,

$$\mathrm{Var}(X_{n+1} - \overline{X}) = \sigma^2 \cdot \frac{n+1}{n}.$$

2) Relación con el cociente Normal-Chi-cuadrado

Sabemos además que la cuasi-varianza S^2 satisface

$$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1},$$

y es independiente de $\overline{X}$ (y por tanto también independiente de X_{n+1}).

Para simplicar la notación, definamos

$$Y = X_{n+1} - \overline{X}.$$

Entonces $Y \sim \mathcal{N}\left(0, \sigma^2 \cdot \frac{n+1}{n}\right)$. Podemos escribir

$$T = \frac{Y}{S}\sqrt{\frac{n}{n+1}} = \underbrace{\frac{\frac{Y}{\sigma}}{\sqrt{\frac{n+1}{n}}}}_{= U \sim \mathcal{N}(0,1)} \Bigg/ \underbrace{\frac{S}{\sigma}}_{\sqrt{\frac{\chi^2_{n-1}}{n-1}}}.$$

- La variable $U = \dfrac{\frac{Y}{\sigma}}{\sqrt{\frac{n+1}{n}}} \sim \mathcal{N}(0,1)$.
- $\dfrac{S^2}{\sigma^2}$ es $\dfrac{1}{n-1}$ veces una χ^2 con $n-1$ grados de libertad.
- U y $\dfrac{S^2}{\sigma^2}$ son independientes.

La razón

$$\frac{U}{\sqrt{\dfrac{\chi^2_{n-1}}{n-1}}}$$

sigue una distribución **t de Stuent** con $n-1$ grados de libertad.

Por lo tanto,

$$T = \frac{Y}{S}\sqrt{\frac{n}{n+1}} = \frac{U}{\frac{S}{\sigma}} = \frac{U}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} \sim t_{n-1}.$$

8) Sean $X_1, X_2, \dots, X_{n_1}$ e $Y_1, Y_2, \dots, Y_{n_2}$ muestras aleatorias simples independientes de dos poblaciones $X \sim \mathcal{N}(\mu_1, \sigma^2)$ e $Y \sim \mathcal{N}(\mu_2, \sigma^2)$, respectivamente. Obtener la distribución en el muestreo estadístico:

$$\frac{\alpha(\overline{X}-\mu_1)+\beta(\overline{Y}-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2+1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{\alpha^2}{n_1}+\frac{\beta^2}{n_2}}},$$

siendo α y β dos números reales fijos.

Tenemos dos muestras aleatorias simples e independientes:

- $X_1, X_2, \dots, X_{n_1}$ de una población $X \sim \mathcal{N}(\mu_1, \sigma^2)$.
- $Y_1, Y_2, \dots, Y_{n_2}$ de otra población $Y \sim \mathcal{N}(\mu_2, \sigma^2)$.

Se denotan:

- $\overline{X} = \dfrac{1}{n_1}\displaystyle\sum_{i01}^{n_1} X_i$ la media muestral de la primera población.
- $\overline{Y} = \dfrac{1}{n_2}\displaystyle\sum_{j=1}^{n_2} Y_j$ la media muestral de la segunda población.
- S_1^2 la cuasi-varianza muestral de la primera muestra.
- S_2^2 la cuasi-varianza muestral de la segunda muestra.

Sabemos que ambas poblaciones comparten la misma varianza σ^2.

El estadístico a estudiar es:

$$T = \frac{\alpha(\overline{X}-\mu_1)+\beta(\overline{Y}-\mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\dfrac{\alpha^2}{n_1}+\dfrac{\beta^2}{n_2}}},$$

donde $\alpha, \beta \in \mathbb{R}$ son constantes dadas.

1) Distribución del numerador

   Sea

$$Z = \alpha(\overline{X}-\mu_1)+\beta(\overline{Y}-\mu_2).$$

   1) Propiedad de $\overline{X}-\mu_1$:

      $\overline{X}$ es $\mathcal{N}\left(\mu_1, \dfrac{\sigma^2}{n_1}\right)$. Por lo tanto,

$$\overline{X}-\mu_1 \sim \mathcal{N}\left(0, \frac{\sigma^2}{n_1}\right).$$

2) Propiedad de $\overline{Y}-\mu_2$:

$\overline{Y}$ es $\mathcal{N}\left(\mu_2, \dfrac{\sigma^2}{n_2}\right)$. Entonces

$$\overline{Y}-\mu_2 \sim \mathcal{N}\left(0, \frac{\sigma^2}{n_2}\right).$$

3) Independencia de las dos muestras:

Como las dos muestras (de X y de Y) son independientes, también lo son $\overline{X}-\mu_1$ y $\overline{Y}-\mu_2$.

Por consiguiente, la variable

$$Z = \alpha(\overline{X}-\mu_1) + \beta(\overline{Y}-\mu_2)$$

es normal de media 0 y varianza

$$\mathrm{Var}(Z) = \alpha^2(\overline{X}-\mu_1) + \beta^2\mathrm{Var}(\overline{Y}-\mu_2) = \alpha^2\frac{\sigma^2}{n_1} + \beta^2\frac{\sigma^2}{n_2}.$$

Por tanto,

$$Z \sim \mathcal{N}\left(0, \sigma^2\left(\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}\right)\right).$$

2) Distribución de la varianza combinada en el denominador

En el denominador aparece el factor

$$\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}.$$

Este es el **estimador combinado** de la desviación típica σ. Recordemos que si las dos muestras provienen con la **misma** varianza σ^2, entonces

$$\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}.$$

Además, este estimador combinado de la varianza es independiente e las medias muestrales $\overline{X}$ y $\overline{Y}$, y por ende, es independiente del numerador Z.

3) El estadístico T

Reescribimos el estadístico:

$$T = \frac{\alpha(\overline{X}-\mu_1) + \beta(\overline{Y}-\mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\dfrac{\alpha^2}{n_1} + \dfrac{\beta^2}{n_2}}} = \frac{\dfrac{Z}{\sigma}}{\dfrac{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\sigma^2(n_1+n_2-2)}}\sqrt{\dfrac{\alpha^2}{n_1} + \dfrac{\beta^2}{n_2}}}{\sigma}}.$$

- La variable

$$\frac{Z}{\sigma} = \frac{Z}{\sqrt{\mathrm{Var}(Z)}} \cdot \sqrt{\frac{\mathrm{Var}(Z)}{\sigma^2}} \sim \mathcal{N}\left(0, \frac{\mathrm{Var}(Z)}{\sigma^2}\right) \quad \text{con} \quad \mathrm{Var}(Z) = \sigma^2\left(\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}\right).$$

Dividiendo Z por $\sigma\sqrt{\dfrac{\alpha^2}{n_1} + \dfrac{\beta^2}{n_2}}$ se obtiene un $\mathcal{N}(0,1)$.

- La parte

$$\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2} \text{ e independiente de } Z.$$

Por la definición de la distribución t de Student con k grados de libertad,

$$t_k = \frac{\mathcal{N}(0,1)}{\sqrt{\dfrac{\chi^2_k}{k}}},$$

nuestro estadístico T encaja exactamente con esta forma, con $k = n_1 + n_2 - 2$.

4) Conclusión

El estadístico

$$T = \frac{\alpha(\overline{X} - \mu_1) + \beta(\overline{Y} - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}}}$$

sigue, en el muestreo, una distribución **t de Student** con $n_1 + n_2 - 2$ grados de libertad. Es decir,

$$\boxed{T \sim t_{n_1+n_2-2}.}$$

9) Una empresa de agua produce botellas que deberían contener 300ml pero que presentan en la práctica una variabilidad modelada por una distribución Normal con media $\mu = 298$ml y desviación típica $\sigma = 3$ml.

a) ¿Cuál es la probabilidad de que una botella elegida al azar de la producción contenga menos de 295ml?

Tenemos botellas cuyo contenido (X) está modelado por una distribución normal con media $\mu = 298$ml y desviación típica $\sigma = 3$ml. Es decir,

$$X \sim \mathcal{N}(298, 3^2).$$

Queremos calcular

$$P(X < 295).$$

Definimos la variable tipificada

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 298}{3}.$$

Entonces,

$$P(X < 295) = P\left(Z < \frac{295 - 298}{3}\right) = P(Z < -1) = P(Z > 1) = 1 - P(Z < 1) = 1 - 0.8413 = 0.1587.$$

b) ¿Cuál es la probabilidad de que el contenido medio de un paquete de seis botellas sea inferior a 295ml?

Sea $\overline{X}_6$ la media muestral de 6 botellas independientes.

$$\overline{X}_6 = \frac{1}{6}\sum_{i=1}^{6} X_i.$$

Dado que cada $X_i \sim \mathcal{N}(\mu, \sigma^2)$ y son independientes, la media $\overline{X}_6$ sigue una distribución normal con:

- Media $\mu_{\overline{X}_6} = \mu = 298$.
- Desviación típica $\sigma_{\overline{X}_6} = \frac{\sigma}{\sqrt{6}} = \frac{3}{\sqrt{6}}$.

Por lo tanto,

$$\overline{X}_6 \sim \mathcal{N}\left(298, \left(\frac{3}{\sqrt{6}}\right)^2\right).$$

Queremos

$$P(\overline{X}_6 < 295) = P\left(Z < \frac{295 - 298}{\frac{3}{\sqrt{6}}}\right) = P\left(Z < \frac{-3}{\frac{3}{\sqrt{6}}}\right) = P(Z < -\sqrt{6})$$
$$= P(Z > \sqrt{6}) = 1 - P(Z < \sqrt{6}) \simeq 1 - P(Z < 2.45) = 1 - 0.9929 = 0.0071.$$

10) Se realiza una medición de peso en un laboratorio, sabiendo que la desviación típica de las mediciones es $\sigma = 10$mg. La medición se repite 3 veces, se calcula la media $\overline{x}$, y este es el resultado proporcionado como estimación del peso.

a) ¿Cuál es la desviación típica del resultado calculado?

Si cada una de las mediciones X_i tiene varianza $\sigma^2 = 10^2 \text{mg}^2$, y las mediciones son independientes, entonces la varianza de la **media muestral** de n mediciones es

$$\text{Var}(\overline{X}) = \frac{\sigma^2}{n}.$$

En consecuencia, la **desviación típica** de la media (o error estándar de la media) es

$$\sqrt{\text{Var}(\overline{X})} = \frac{\sigma}{\sqrt{n}}.$$

Para $n = 3$ mediciones,

$$\sqrt{\text{Var}(\overline{X})} = \frac{10}{\sqrt{3}} \approx 5.77\text{mg}.$$

b) ¿Cuántas veces se debe repetir la medición para que la desviación típica del valor medio se reduzca a 5?

Queremos que la desviación típica de $\overline{X}$ sea 5 mg.

Usando

$$\sqrt{\text{Var}(\overline{X})} = \frac{10}{\sqrt{n}} = 5,$$

resolvemos para n:

$$\frac{10}{\sqrt{n}} = 5 \longrightarrow \sqrt{n} = 2 \longrightarrow n = 4.$$

Por lo tanto, **necesitamos 4 mediciones** para reducir la desviación típica de la media a 5 mg.

11) El resultado de una encuesta fue que el 59% de la población española opina que el contexto económico es bueno o muy bueno. Supongamos que, extrapolando al conjunto de la población, efectivamente la proporción de todos los españoles que piensan que la situación es buena o muy buena es del 0.59.

Tenemos una población en la que la proporción real de persobas que piensan que la situación económica es buena o muy buena es $p = 0.59$.

Cuando el tamaño de la muestra n es suficientemente grande, podemos aproximar la distribución de $\hat{p}$ mediante una Normal con media p y desviación típica

$$\sqrt{\frac{p(1-p)}{n}}.$$

En este problema:

- $p = 0.59$
- $q = 1 - p = 0.41$
- Desviación típica de $\hat{p}$:

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.59 \cdot 0.41}{n}}.$$

Nos interesa

$$P(0.56 \leq \hat{p} \leq 0.62) = P(\hat{p} - 0.59 \in [-0.03, 0.03]).$$

Definamos la variable tipificada

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \sim \mathcal{N}(0, 1).$$

Entonces

$$P(|\hat{p} - 0.59| \leq 0.03) = P(-0.03 \leq \hat{p} - 0.59 \leq 0.03) = P(-z_0 \leq Z \leq z_0),$$

donde

$$z_0 = \frac{0.03}{\sqrt{\frac{0.59 \cdot 0.41}{n}}}$$

a) Sabemos que las encuentas incluyen márgenes de error que son aproximadamente ± 3 puntos. ¿Cuál es la probabilidad de que una muestra aleatoria de 300 españoles presente una proporción muestral que caiga dentro del intervalo 0.59 ± 0.03?

1) **Cálculo de la desviación típica de $\hat{p}$:**

$$\sqrt{\frac{0.59 \cdot 0.41}{300}} = \sqrt{\frac{0.2419}{300}} \simeq 0.0284.$$

2) **Cálculo de z_0:**

La amplitud del intervalo es 0.03. Dividimos por la desviación típica:

$$z_0 = 0.\frac{03}{0.0284} \simeq 1.06$$

3) **Probabilidad asociada:**

$Z \sim \mathcal{N}(0,1)$. Entonces

$$P(-1.06 \leq Z \leq 1.06).$$

Usando tablas o función de distribución normal,

$$\Phi(1.06) \simeq 0.8554,$$

de modo que

$$P(-1.06 \leq Z \leq 1.06) = \Phi(1.06) - (1 - \Phi(1.06)) = 0.8554 - (1 - 0.8554) = 0.8554 - 0.1446 = 0.7108 \approx 0.71$$

Por lo tanto, **la probabilidad es aproximadamente 0.71** (un 71%).

b) Contesta a la pregunta anterior en el caso en que la muestra consta de 600 personas y cuando la muestra consta de 1200 personas. ¿Cuál es el efecto de aumentar el tamaño de la muestra?

Caso $n = 600$:

1) **Desviación típica de $\hat{p}$:**

$$\sqrt{\frac{0.2419}{600}} \simeq 0.0201.$$

2) **Cálculo de z_0:**

$$z_0 = \frac{0.03}{0.0201} \simeq 1.49$$

3) **Probabilidad:**

$$P(-1.49 \leq Z \leq 1.49).$$

Tenemos que $\Phi(1.49) \approx 0.9319$. Por lo tanto,

$$P(-1.49 \leq Z \leq 1.49) = 0.9319 - (1 - 0.9319) = 0.9319 - 0.0681 = 0.8638 \approx 0.86$$

Por lo tanto, **la probabilidad es aproximadamente 0.86** (un 86%).

Caso $n = 1200$:

1) **Desviación típica de $\hat{p}$:**

$$\sqrt{\frac{0.2419}{1200}} \approx 0.0142.$$

2) **Cálculo de z_0:**

$$z_0 = \frac{0.03}{0.0142} \approx 2.11$$

3) **Probabilidad:**

$$P(-2.11 \leq Z \leq 2.11).$$

Tenemos que $\Phi(2.11) \approx 0.9826$. Entonces,

$$P(-2.11 \leq Z \leq 2.11) = 0.9826 - (1 - 0.9826) = 0.9826 - 0.0174 = 0.9652 \approx 0.965.$$

Por tanto, **la probabilidad es aproximadamente 0.965** (un 96.5%).

12) Un aparato de medición es exacto (el valor proporcionado medio es el valor auténtico de la señal) y la desviación típica del valor medido es 0.1 unidades. La distribución del valor medido es aproximadamente normal. ¿Cuál es la probabilidad de que el valor de una medición se aleje de la señal auténtica en más de 0.1 unidades? ¿Y si se repite la medición 5 veces y se toma la media de los 5 valores obtenidos?

Tenemos un aparato de medición que, al medir una señal (valor esperado μ), proporciona un valor X que es:

- **Exacto en promedio:** $E[X] = \mu$.
- **Normalmente distribuido** con desviación típica $\sigma = 0.1$. Es decir,

$$X \sim \mathcal{N}(\mu, 0.1^2).$$

Queremos calcular:

1) $P(|X - \mu| > 0.1)$, es decir, la probabilidad de que **una medición individual** se aleje más de 0.1 unidades de la señal real.

2) $P(\overline{X}_5 - \mu| > 0.1)$, donde $\overline{X}_5$ es la **media de 5 mediciones independientes**.

1) Probabilidad de que una sola medición difiera de μ es más de 0.1

Sea X la lectura, $X \sim \mathcal{N}(\mu, 0.1^2)$. Entonces

$$P(|X - \mu| > 0.1) = P(X - \mu > 0.1 \text{ o } X - \mu < -0.1).$$

Tipificamos usando $Z = \dfrac{X - \mu}{0.1}$, que sigue $\mathcal{N}(0, 1)$. Así,

$$P(|X - \mu| > 0.1) = P(|Z| > 1) = 2(1 - \Phi(1)).$$

Tenemos que $1 - \Phi(1) \approx 0.1587$. Por tanto,

$$P(|Z| > 1) = 2 \cdot 0.1587 = 0.3174.$$

2) Probabilidad de que la media de 5 mediciones difiera que μ en más de 0.1

Sea $\overline{X}_5 = \dfrac{1}{5} \sum_{i=1}^{5} X_i$. Dado que cada $X_i \sim \mathcal{N}(\mu, 0.1^2)$ y son independientes, se tiene

- $E[\overline{X}_5] = \mu$
- $\mathrm{Var}(\overline{X}_5) = \dfrac{\sigma^2}{5} = \dfrac{0.1^2}{5}$
- $\sigma(\overline{X}_5) = \dfrac{0.1}{\sqrt{5}}$

La probabilidad de interés es

$$P(|\overline{X}_5 - \mu| > 0.1).$$

Tipificamos con

$$Z = \frac{\overline{X}_5 - \mu}{\frac{0.1}{\sqrt{5}}} \sim \mathcal{N}(0,1).$$

Entonces,

$$P(|\overline{X}_5 - \mu| > 0.1) = P\left(|Z| > \frac{0.1}{\frac{0.1}{\sqrt{5}}}\right) = P(|Z| > \sqrt{5}) \approx 1 - \Phi(Z < 2.24) = 0.0127.$$

De modo que

$$P(|Z| > 2.24) = 2 \cdot (1 - \Phi(2.24)) \approx 2 \cdot 0.0127 = 0.0254.$$

13) En condiciones normales, una máquina produce piezas con una tasa de defectuosas del 1%. Para controlar que la máquina sigue bien ajustada, se escogen al azar cada día 100 piezas en la producción y se somete a un test. ¿Cuál es la probabilidad de que, si la máquina está bien ajustada, haya, en una de esas muestras, más del 2% de piezas defectuosas? Si un día, 3 piezas resultan defectuosas, ¿cuáles son las conclusiones que sacaríamos sobre el funcionamiento de la máquina?

Bajo el supuesto de que la máquina está bien ajustada, X (número de piezas defectuosas en la muestra) sigue una distribución **Binomial** con parámetros $n = 100$ y $p = 0.01$.

$$X \sim \mathcal{B}(n = 100, p = 0.01).$$

1) Probabilidad de obtener más del 2% de piezas defectuosas si la máquina está bien

El 2% de 100 piezas son 2 piezas. Por tanto, "más del 2%" significa que en la muestras salen **3 o más defectuosas**. Queremos

$$P(X > 2) = 1 - P(X \leq 2).$$

1.1) Aproximación con distribución Poisson

Dado que $p = 0.01$ es bastante pequeño y $n = 100$ moderado, podemos usar la **aproximación Poisson** con $\lambda = np = 1$.

Entonces $X \approx \text{Poisson}(\lambda = 1)$. Se calcula

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2),$$

donde, para Poisson$(\lambda = 1)$:

$$P(X = k) = \frac{e^{-1} 1^k}{k!}.$$

- $P(X = 0) = e^{-1} \approx 0.3679$
- $P(X = 1) = e^{-1} \cdot \frac{1}{1!} = 0.3679$
- $P(X = 2) = e^{-1} \cdot \frac{1^2}{2!} = 0.1839$

Sumamos:

$$P(X \leq 2) = 0.3679 + 0.3679 + 0.1839 = 0.9197$$

Por tanto,

$$P(X > 2) = 1 - 0.9197 = 0.0803 \approx 0.08.$$

Esto indica que, si la máquina está bien, hay alrededor de un 8% de probabilidad de observar 3 o más piezas defectuosas en una muestra de 100

2) Interpretación en un día con 3 defectuosas observadas

Si en la muestra diaria (de 100 piezas) aparecen **3 defectuosas**, podemos preguntarnos si esto sugiere que la máquina se ha desajustado.

- Hemos visto que, si realmente la máquina sigue produciendo al 1% de defectos, la probabilidad de ver 3 o más defectuosas es alrededor de un 8%.
- Un 8% no es un suceso extremadamente raro. En un test de hipótesis con un nivel de significación típico del 5%, podríamos decir que la **p-value**$\approx$ 0.08 es **mayor** que 0.05, de modo que no **no** rechazariamos la hipótesis de que la máquina está bien.

En otras palabras, **3 piezas defectuosas no es una evidencia suficientemente fuerte** para concluir que la máquina está mal ajustada.

Tema 2: Estimación

2.1) Introducción

- Hemos modelizado un experimento con una variable aleatoria X.
- La estimación hace referencia al proceso de conseguir información sobre la distribución de X a partir de los valores de una muestra, aproximando valores asociados a la distribución mediante el valor de un estadístico en una muestra concreta.

Dos situaciones

- Nuestro modelo supone que la distribución de X pertence a una familia paramétrica de distribuciones: tienen una determinada forma con unos parámetros variables.
  $\rightarrow$ Buscamos información sobre el valor de los parámetros. Estimación paramétrica.
- No limitados la familia de distribuciones a la que pertence nuestro modelo.
  $\rightarrow$ Buscamos información sobre la distribución en sí (función de distribución, de densidad o función puntual de probabilidad). Estimación no paramétrica.
- A lo largo de las prácticas veremos también la estimación de parámetros que no necesitan de una familia paramétrica, como es el caso de la mediana.

2.2) Ejemplos de estimación paramétrica

Sondeo sobre intención de participación en una elecciones

- Queremos estimar la tasa de participación antes de unas elecciones generales.
- Formulamos un modelo: experimento: "escoger una persona al azar en el censo". X: variable dicotómica ("Sí", o "No"). $p = P(X = \text{Si})$
- El modelo pertenece a la familia paramétrica de Bernoulli.

Determinación de la concentración de un producto

- Quiero determinar la concentración.
- Formulo el modelo: experimento="llevar a cabo una medición". X: "valor proporcionado por el aparato". $X \sim \mathcal{N}(\mu, \sigma^2)$.
- El modelo pertence a la familia paramétrica de las distribuciones normales

2.3) Estimación paramétrica: estimación puntual

Ingredientes del modelo

- Experimento aleatoria
- Variable aleatoria X con una distribución f que pertence a una familia paramétrica $\{f_\theta, \theta \in \Theta\}$.
- Disponemos de una muestra de la distribución de X.

Definición

Cualquier estadístico diseñado para aproximar el valor de un parámetro θ del modelo, se llama estimador puntual del parámetro θ.

Ejemplos de estimadores paramétricos

θ	**Estimador**
μ	$\overline{X}$, media muestral
σ^2	S^2, varianza muestral
p	$\hat{p}$, proporción muestral

A tener en cuenta:

- Un estimador es una variable aleatoria, su valor depende de la muestra concreta escogida.
- Para controlar bondad de nuestra estimación, nos basaremos en el estudio de la distribución del estimador.

2.4) Métodos de construcción de estimadores

Para μ, σ^2 o p, es fácil pensar en estimadores naturales, pero para modelos o parámetros más sofisticados, vamos a ver métodos generales.

Veremos dos métodos en este tema:

- El método de los momentos
- El método de la máxima verosimilitud

2.5) Método de los momentos

Contexto

- Experimento con una variable aleatoria X, suponemos $f_X \in \{x \longmapsto f_\theta(x), \theta \in \Theta\}$.
- El parámetro θ tiene dimensión p.
- Consideramos una muestra aleatoria simple $X_1, \ldots, X_n$ de X.

Si θ tiene dimensión p, igualamos los p primeros momentos de f_θ con los equivalentes muestrales.

Sea $\mu_k(\theta)$ el momento de orden k de la distribución $f_\theta, \mu_k = E[X^k]$. Resolvemos:

$$\begin{aligned}
\mu_1(\theta) &= \overline{X}, \\
\mu_2(\theta) &= \overline{X^2}, \\
&\vdots \\
\mu_p(\theta) &= \overline{X^p}.
\end{aligned}$$

Ejemplos

Calculad los estimadores usando el método de los momentos en los dos casos:

- Modelo normal: $X \sim \mathcal{N}(\mu, \sigma^2)$, donde $\theta = (\mu, \sigma^2)$.

$$
\begin{aligned}
&X \rightsquigarrow \mathcal{N}(\mu, \sigma^2) && \mu = E[X] \\
&\theta = (\mu, \sigma^2), \quad p = 2 && \sigma^2 = \mathrm{Var}[X] = E[X^2] - (E[X])^2 \longrightarrow E[X^2] = \sigma^2 + \mu^2 \\
&\mu_1(\theta) = E[X] = \overline{x} && \hat{\mu} = \overline{x} \\
&\mu_2(\theta) = E[X^2] = \overline{x^2} && \hat{\sigma}^2 = \overline{x^2} - (\overline{x})^2
\end{aligned}
$$

- Modelo de Bernoulli: $X \sim \mathrm{Bernoulli}(p)$, donde desconocemos p.

$$
\begin{aligned}
&X \rightsquigarrow \mathrm{Bernoulli}(p), \quad 0 < p < 1 \\
&\theta = p \\
&\mu(\theta) = E[X] = \overline{x} = p \\
&\hat{p} = \overline{x} \text{ (proporción muestral)}
\end{aligned}
$$

2.6) Método de máxima verosimilitud

- El método más utilizado de construicción de un estimador puntual.
- Se basa en lo que se conoce como función de verosimilitud.

Definición

- Sea X una variable aleatoria, con distribución $x \longmapsto f_X(x;\theta)$ (función de densidad o función puntual de probabilidad), donde θ es de dimensión $p : \theta \in \Theta \subset \mathbb{R}^p$.
- Para un valor concreto de una muestra aleatoria simpl $(X_1, \ldots, X_n)$, que denotamos por $(x_1, \ldots, x_n)$, consideramos la función de θ:

$$
L_n : \begin{cases} \Theta \subset \mathbb{R}^p \longrightarrow \mathbb{R}^+ \\ \theta \longmapsto L_n(\theta) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \theta) = \displaystyle\prod_{i=1}^{n} f_X(x_i; \theta). \end{cases}
$$

- La función L_n es la función de verosimilitud. Nos dice lo creíblles (verosímiles) que son las observaciones para ese valor del parámetro.

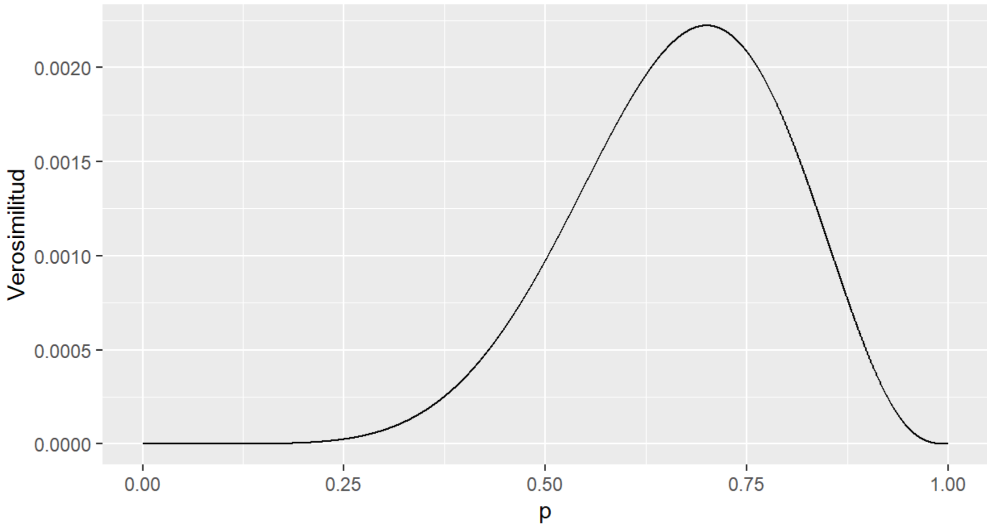
Ejemplo de cálculo de la verosimilitud

- Tiramos 10 veces una moneda (1 es cara, 0 es cruz), y obtenemos: 0,0,1,0,1,1,1,1,1,1.
- La verosimilitud asocia a cada p el valor de

$$
\Pr(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 1, X_6 = 1, X_7 = 1, X_8 = 1, X_9 = 1, X_{10} = 1),
$$

por lo que

$$
L_n(p) = (1-p)(1-p)p(1-p)p^6 = (1-p)^3 \cdot p^7.
$$



2.7) Estimador de máxima verosimilitud

Definición

El estimación de máxima verosimilitud $\hat{\theta}$ de θ es cualquier valor de θ que maximiza $\theta \longmapsto L_n(\theta)$, es decir,

$$\hat{\theta} \underset{\theta \in \Theta}{\operatorname{argmax}} L_n(\theta).$$

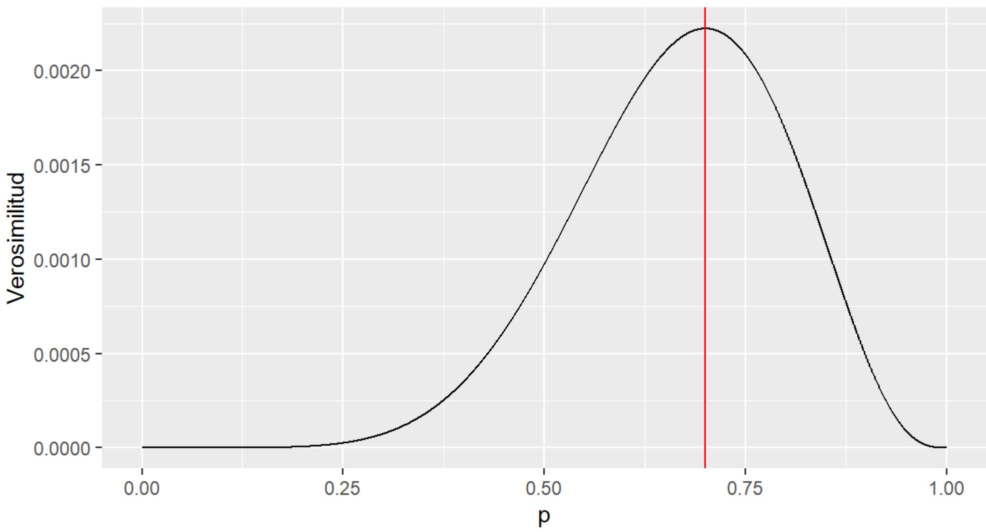
Nota

- La maximización se realiza sobre todos los valores admisibles para el parámetro θ.
- Podría haber de un máximo.

Estimación de la proporción

Retomamos el ejemplo de las 10 monedas:

$$p \longmapsto L_n(p) = (1-p)(1-p)p(1-p)p^6 = (1-p)^3 \cdot p^7.$$



Ejemplos

Ejemplos de estimación por máxima verosimilitud

Calculad los estimadores usando el método de máxima verosimilitud en los dos casos:

- $X \sim \text{Bernoulli}(p)$, donde desconocemos p.
- $X \sim \mathcal{N}(\mu, \sigma^2)$, donde desconocemos $\theta = (\mu, \sigma^2)$.

- En el primer caso anterior, calcular la distribución de Bernoulli, donde desconocemos p.

$X \rightsquigarrow \text{Bernoulli}(p), \quad 0 < p < 1$

$(X_1, \ldots, X_n)$ m.a.s de tamaño n

$$L_n(\theta) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} f_X(x_i; \theta)$$

$$L_n(p) = \prod_{i=1}^{n} f_X(x_i; p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$\log L_n(p) = \log p^{\sum_i x_i} + \log(1-p)^{n-\sum_{i=1}^{n} x_i} = \left(\sum_i x_i\right) \cdot \log p + \left(n - \sum_{i=1}^{n} x_i\right) \cdot \log(1-p)$$

$$\begin{aligned}
\frac{\partial L_n(p)}{\partial p} &= \left(\sum_{i=1}^{n} x_i\right) \cdot \frac{1}{p} + \left(n - \sum_{i=1}^{n} x_i\right) \cdot \left(-\frac{1}{1-p}\right) \\
&= (1-p) \cdot \sum_{i=1}^{n} x_i - \left(n - \sum_{i=1}^{n} x_i\right) \cdot p \\
&= \sum_{i=1}^{n} x_i - \cancel{p \cdot \sum_{i=1}^{n} x_i} - np + \cancel{p \cdot \sum_{i=1}^{n} x_i} = 0 \\
& \boxed{\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}}
\end{aligned}$$

- En el segundo caso anterior, calculad la esperanza del estimador de máxima verosimilitud de σ^2

La función de densidad de una varaible aleatoria X con distribución normal es:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Si tenemos una muestra de tamaño n, es decir, $X_1, X_2, \ldots, X_n$ que son observaciones independientes y distribuidas como $\mathcal{N}(\mu, \sigma^2)$, entonces la función de verosimilitud $L(\mu, \sigma^2)$ es el producto de las funciones de densidad de cada observación

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i|\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

Para simplificar los cálculos, se toma el logaritmo de la función de verosimilitud, lo que da la función de log-verosimilitud $\ell(\mu, \sigma^2)$:

$$\begin{aligned}
\ell(\mu, \sigma^2) = \log L(\mu, \sigma^2) &= \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \sum_{i=1}^{n} \log\left(\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right) \\
&= \sum_{i=1}^{n} \log\left((2\pi\sigma^2)^{-\frac{1}{2}}\right) + \sum_{i=1}^{n} -\frac{(x_i-\mu)^2}{2\sigma^2} = \sum_{i=1}^{n} -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\
&= -\frac{n}{2} \cdot \log(2\pi) - \frac{n}{2} \cdot \log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2
\end{aligned}$$

Para encontrar los estimadores de máxima verosimilitud, derivamos $\ell(\mu, \sigma^2)$ con respecto a μ y σ^2, e igualamos a cero:

$$\frac{\partial \ell(\mu,\sigma^2)}{\partial \mu} = \frac{\cancel{2}}{\cancel{2}\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \longrightarrow \sum_{i=1}^{n}(x_i-\mu) = 0 \longrightarrow \boxed{\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}}$$

$$\frac{\partial \ell(\mu,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \longrightarrow -n\sigma^2 + \sum_{i=1}^{n}(x_i-\mu)^2 = 0 \longrightarrow \boxed{\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2}$$

2.8) Métodos para evaluar un estimador

Recordad

- Un estimador es una variable aleatoria.
- Es valioso disponer de conocimiento sobre la distribución del estimador (su **distribución en el muestreo**) $\longrightarrow$ permite manejar el riesgo y el error que podemos cometer al aproximar θ por $\hat{\theta}$ asociado.

Consideramos dos aspectos de la distribución muestral de $\hat{\theta}$

- Su localización: sesgo.
- Su variabilidad: error cuadrático medio.
- Mencionaremos su comportamiento cuando $n \to \infty$.

2.9) Sesgo

Definición

Consideramos para un estimador $\hat{\theta}$ de un parámetro θ: $E_\theta[\hat{\theta}] - \theta$.
Esta diferencia se llama el sesgo.

Una propiedad deseable para un estimador

Si el sesgo de un estimador es nulo para todo valor de θ, decimos que el estimador **inesgado**.

2.10) Error cuadrático medio

Para medir la variabilidad en el muestreo de un estimador.

Definición

El error cuadrático medio del estimador $\hat{\theta}$ es la función de θ definida por

$$\theta \longmapsto E_\theta[(\hat{\theta}-\theta)^2]$$

Para practicar

Calculad el error cuadrático medio del estimador de máxima verosimilitud (e.m.v.) de μ para una muestra aleatoria simple de $X \sim \mathcal{N}(\mu, \sigma^2)$.

$X \rightsquigarrow \underbrace{\mathcal{N}(\mu, \sigma^2)}_{\text{desconocidos}}$

$\hat{\mu} = \overline{x}$ (estimador basado en los movimientos)

$E[\overline{X}] = \mu \longrightarrow \hat{\mu} = \overline{X}$ es un estimador insesgado para μ

$$\left.\begin{array}{l} \mathrm{Var}[\overline{X}] = \dfrac{\sigma^2}{n} \\ \mathrm{Var}[\overline{X}] = E[(\overline{X}-\mu)^2] \end{array}\right\} \longrightarrow E[(\overline{X}-\mu)^2] = \frac{\sigma^2}{n} \quad \text{(E.C.M)}$$

2.11) Balance entre sesgo y varianza

Sesgo y varianza

El error cuadrático medio se puede descomponer

$$E_\theta[(\hat{\theta}-\theta)^2] = \mathrm{Var}(\hat{\theta}) + [\mathrm{sesgo}(\hat{\theta})]^2$$

En ocasiones, se consigue un menor error cuadrático medio con un estimador sesgado y estamos dispuestos a sacrificar el sesgo, por conseguir una menor varianza.

- Demostración

$$\begin{aligned} E_\theta[(\hat{\theta}-\theta)^2] &= E_\theta[(\hat{\theta}-E_\theta[\hat{\theta}]+E_\theta[\hat{\theta}]-\theta)^2] \\ &= E_\theta[(\hat{\theta})-E_\theta[\hat{\theta}]+(E_\theta[\hat{\theta}]-\theta)^2+2(\hat{\theta}-E_\theta[\hat{\theta}])(E_\theta[\hat{\theta}]-\theta)] \\ &= \mathrm{Var}[\hat{\theta}]+[\mathrm{sesgo}(\hat{\theta})]^2 \end{aligned}$$