

Evaluando hipótesis

En muchos casos es importante evaluar la performance de las hipótesis aprendidas.

1ro: Dada la precisión observada de una hipótesis sobre una muestra de datos limitada, ¿cuán bien estimará la precisión sobre ejemplos adicionales?

2do: Dado que una hipótesis supera a otra sobre alguna muestra de datos, ¿cuán probable es que esta hipótesis sea más precisa en general?

3ro: Cuando los datos son limitados, ¿cuál es la mejor manera de usarlos tanto para aprender la hipótesis como para estimar su precisión?

Debido a que un conjunto limitado de datos pueden distorsionar la distribución general de los datos, estimar la precisión verdadera con tales muestras puede ser engañosa.

Cuando se debe aprender una hipótesis y estimar su futura precisión dado sólo un conjunto limitado de datos, aparecen dos dificultades:

- **sesgo en el estimador**: la precisión observada de la hipótesis aprendida sobre los ejemplos de entrenamiento es frecuentemente un estimador malo de su precisión sobre futuros ejemplos.
 - Debido a que la hipótesis aprendida se derivó de estos ejemplos, ellos darán un estimador **sesgado** de forma optimista de la precisión sobre futuros ejemplos.
 - Para obtener un estimador **insesgado** de la futura precisión, testeamos la hipótesis sobre algún conjunto de ejemplos de testeo elegidos independientemente de los ejemplos de entrenamiento y de la hipótesis.

- **varianza en el estimador**: aún si la precisión de la hipótesis es medida sobre un conjunto insesgado de ejemplos de testeo independientes de los ejemplos de entrenamiento, la precisión medida puede aún variar de la verdadera precisión, dependiendo del particular conjunto de ejemplos de testeo.
 - Cuanto más chico sea el conjunto de ejemplos de testeo, mayor será la varianza esperada.

Estimando la precisión de la hipótesis

- Consideremos por ejemplo el aprendizaje de la función objetivo



“personas que planean comprar un nuevo par de esquíes este año”

dada una muestra de datos entrenamiento recogida en una encuesta entre la gente que llega un centro de esquí.

- En este caso, el espacio de instancias X es el espacio de toda la gente, que puede estar descripto por los atributos como la edad, la ocupación, cuántas veces al año fue a esquiar del año anterior, etc.

- La **distribución \mathcal{D}** especifica para cada persona x es la **probabilidad** de que x **compre esquís nuevos** este año
- La **función objetivo $f: X \rightarrow \{0,1\}$** clasifica a cada persona de acuerdo a si **planea o no comprar** esquís este año.
- Estamos interesados a responder estas dos preguntas:
 - Dada una hipótesis h y una muestra de datos que contiene n ejemplos tomados al azar de acuerdo a la distribución \mathcal{D} , ¿Cuál es el **mejor estimador de la precisión** de h sobre futuras instancias que tengan la misma distribución?
 - ¿Cuál es el **error** probable en esta **estimación de precisión**?

Error muestral y error verdadero

El **error muestral** de una hipótesis con respecto a alguna muestra S de instancias tomadas de X es la **fracción** de S está **mal clasificada**

Definición: El **error muestral**, denotado con $\text{error}_S(h)$ de la hipótesis h con respecto a la función objetivo f y a la muestra S es:

$$\text{error}_S(h) \equiv \frac{1}{n} \sum_{x \in S} \delta(f(x), h(x))$$

donde n es el tamaño de la muestra S , y

$$\delta(f(x), h(x)) = \begin{cases} 1 & f(x) \neq h(x) \\ 0 & f(x) = h(x) \end{cases}$$

El **error verdadero** de una hipótesis es la probabilidad de que ésta clasifique mal a una única instancia al azar con distribución \mathcal{D}

Definición: El **error verdadero**, denotado con $\text{error}_{\mathcal{D}}(h)$ de la hipótesis h con respecto a la función objetivo f y la distribución \mathcal{D} , es la probabilidad de que h clasifique mal a una instancia al azar con distribución \mathcal{D}

$$\text{error}_{\mathcal{D}}(h) = \mathbb{P}_{\mathcal{D}}[f(x) \neq h(x)]$$

- Generalmente queremos conocer cuál es el **error verdadero** $\text{error}_{\mathcal{D}}(h)$ de la hipótesis, porque éste es el **error** que podemos esperar cuando apliquemos la hipótesis a **futuros ejemplos**.
- Sin embargo, todo lo que se puede medir es el **error muestral** $\text{error}_S(h)$ de la hipótesis para la muestra S .
- La cuestión es
¿cuán bien se estima el $\text{error}_{\mathcal{D}}(h)$ a partir del $\text{error}_S(h)$?

Distribución Bernoulli



Daniel Bernoulli
(1700- 1782)

Consideremos un experimento que puede tener dos posibles resultados:

- un evento ocurre (**éxito**) con probabilidad p
- no ocurre (**fracaso**) con probabilidad $1 - p$.

Sea X la variable aleatoria tal que:

$$X = \begin{cases} 1 & \text{si el evento ocurre} \\ 0 & \text{si el evento no ocurre} \end{cases}$$

entonces $X \sim \mathcal{B}(p)$, y la función de probabilidad puntual está dada por:

$$P(X = x) = p^x(1 - p)^{1-x} \quad \text{para } x \in \{0,1\}$$

Además:

$$E(X) = p \quad \text{var}(X) = p(1 - p)$$

Distribución Binomial

Un experimento que satisface:

- El experimento consiste en una secuencia de m ensayos
- Los ensayos son idénticos y cada uno de ellos puede resultar en uno de dos posibles resultados, que notamos por éxito o fracaso
- Los ensayos son independientes, por lo que el resultado de cualquier ensayo particular no influye sobre resultado de cualquier otro ensayo

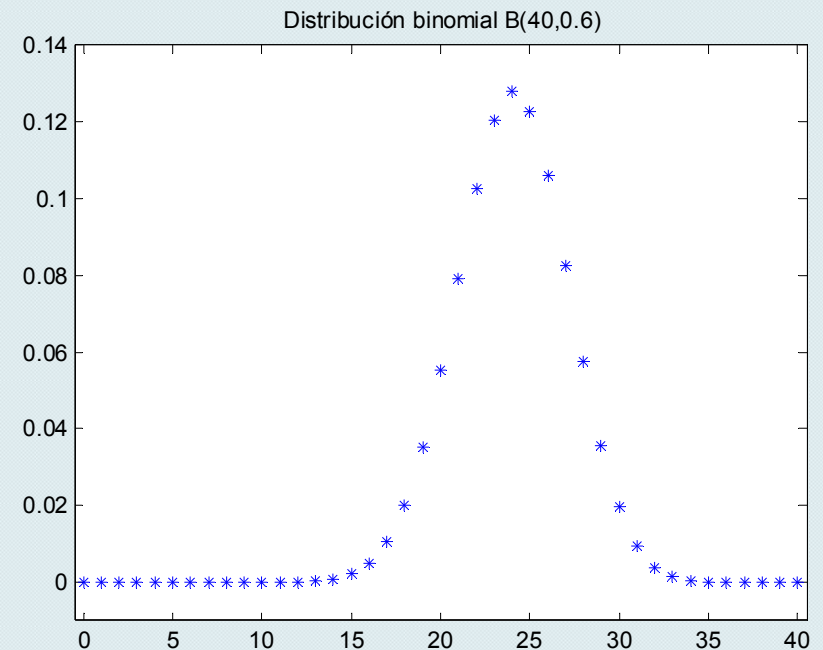
- La probabilidad p de éxito es constante de un ensayo otro se denomina **experimento binominal**.

Si

$X =$ número de éxitos en los n ensayos

entonces $X \sim \text{Bi}(m, p)$, y la función de probabilidad puntual está dada por:

$$P(X = x) = \binom{m}{x} p^x (1 - p)^{m-x} \quad \text{para } x \in \{0, 1, \dots, m\}$$



Distribución multinomial

Consideremos la generalización de la distribución Bernoulli, donde el experimento puede tener K posibles resultados que son mutuamente excluyentes, cada uno de los cuales con probabilidad p_j de modo que $\sum_{j=1}^K p_j = 1$.

Sean x_1, \dots, x_k las variables indicadoras tales que:

$$x_j = \begin{cases} 1 & \text{si se obtiene el resultado } j \\ 0 & \text{si no se obtiene el resultado } j \end{cases}$$

Si $\mathbf{X} = (x_1, \dots, x_k)$ y $\mathbf{p} = (p_1, \dots, p_k)$, entonces $\mathbf{X} \sim \mathcal{M}(K, \mathbf{p})$ y la función de probabilidad puntual está dada por:

$$P(X_1 = x_1, \dots, X_k = x_k) = \prod_{j=1}^K p_j^{x_j} \quad \text{para } \sum_{j=1}^K x_j = 1$$

Distribución normal



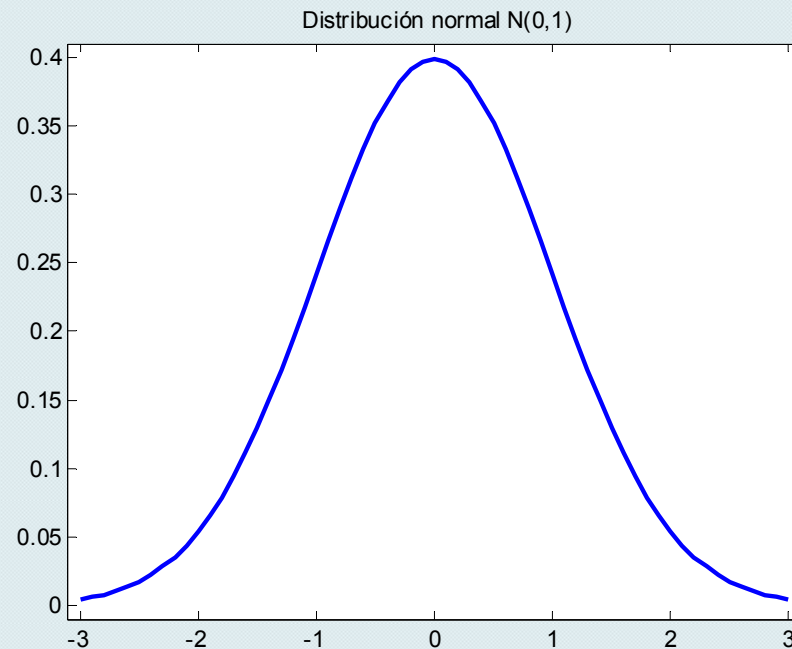
Carl Friedrich Gauss
(1777- 1855)

X es una variable aleatoria gaussiana con media μ y varianza σ^2 , $X \sim \mathcal{N}(\mu, \sigma^2)$ si su función de densidad está dada por:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Además:

$$E(X) = \mu; \quad \text{var}(X) = \sigma^2$$



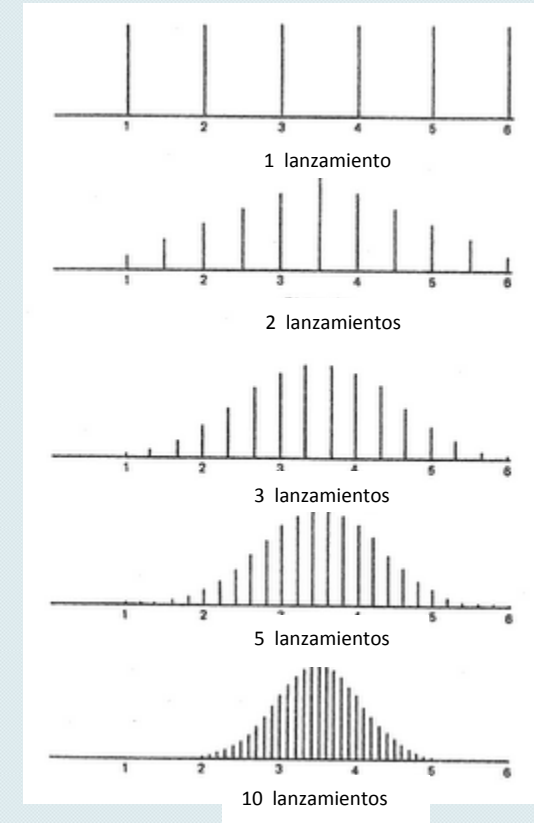
Teorema central del límite

Sean X_1, \dots, X_n variables aleatorias i.i.d. (independientes e idénticamente distribuidas) tales que su distribución tiene media μ y varianza $\sigma^2 \neq 0$, entonces:

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \rightarrow^D \mathcal{N}(0,1)$$

Equivalentemente:

$$\sum_{i=1}^n \frac{X_i}{n} = \bar{X} \rightarrow^D \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$



distribución del lanzamiento de un dado, una vez, dos, tres, cinco y diez veces respectivamente

Estimación

- Dado un parámetro de interés, como por ejemplo una media poblacional μ , el objetivo de la **estimación puntual** es utilizar una **muestra** para calcular un número que represente en algún sentido una buena presunción para el verdadero valor del parámetro.
- El número resultante se llama estimación puntual.
- Se dice que las variables aleatorias X_1, \dots, X_n son una muestra aleatoria de tamaño n si:
 - las X_i son variables aleatorias independientes
 - todas las X_i tienen la misma distribución de probabilidad $p(x|\theta)$

Estimación de máxima verosimilitud



Ronald Fisher
(1890- 1962)

- Este método fue introducido por R. A. Fisher en la década de 1920.
- Supongamos que tenemos una muestra aleatoria $\mathcal{X} = \{x_1, \dots, x_n\}$ correspondiente variables aleatorias X_1, \dots, X_n con distribución $p(x|\theta)$.

Nos preguntamos: ¿para qué valor de θ es más probable que la muestra observada \mathcal{X} haya ocurrido?

- La función de verosimilitud está dada por:

$$\ell(\mathcal{X}|\theta) = \prod_{i=1}^n p(x_i|\theta)$$

- Esta función nos indica **qué tan probable es que una muestra sea observada** como una **función** de los posibles valores del **parámetro**.
- Bajo ciertas condiciones, maximizar la verosimilitud ℓ es equivalente a maximizar $\ln \ell$, es decir:

$$\mathcal{L}(\mathcal{X}|\theta) = \ln \ell(\mathcal{X}|\theta) = \sum_{i=1}^n \ln p(x_i|\theta)$$

- Entonces

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}(\mathcal{X}|\theta)$$

Ejemplos:

Distribución Bernoulli

Dada una muestra aleatoria $\mathcal{X} = \{x_1, \dots, x_n\}$ de tamaño n de variables aleatorias i.i.d. (independientes e idénticamente distribuidas) $\mathcal{B}(p)$, para encontrar el estimador de máxima verosimilitud \hat{p} de p calculamos:

$$\begin{aligned}\mathcal{L}(\mathcal{X}|p) &= \ln \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \\ &= \sum_{i=1}^n x_i \ln p + \left(n - \sum_{i=1}^n x_i \right) \ln (1-p)\end{aligned}$$

Para encontrar \hat{p} que maximiza la verosimilitud hay que resolver $\left. \partial \mathcal{L} / \partial p \right|_{p=\hat{p}} = 0$, es decir:

$$\begin{aligned}
 0 &= \frac{1}{\hat{p}} \left(\sum_{i=1}^n x_i \right) - \frac{1}{1 - \hat{p}} \left(n - \sum_{i=1}^n x_i \right) \\
 &= \frac{n\bar{x}}{\hat{p}} - \frac{n(1 - \bar{x})}{1 - \hat{p}}
 \end{aligned}$$

Despejando \hat{p} obtenemos:

$$\hat{p} = \bar{x} = \sum_{i=1}^n \frac{x_i}{n}$$

Distribución multinomial

Dada una muestra aleatoria $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, donde $\mathbf{x}_j = (x_{1j}, \dots, x_{Kj})$ con

$$x_{ij} = \begin{cases} 1 & \text{si se obtiene el resultado } j \text{ en el experimento } i \\ 0 & \text{si no se obtiene el resultado } j \text{ en el experimento } i \end{cases}$$

de tamaño n de vectores aleatorios i.i.d. $\mathcal{M}(K, \mathbf{p})$, los estimadores de máxima verosimilitud \hat{p}_j de p_j están dados por:

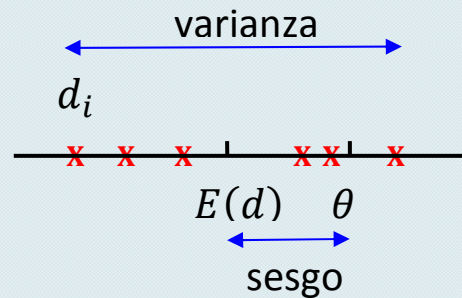
$$\hat{p}_j = \sum_{i=1}^n \frac{x_{ij}}{n}$$

Distribución normal

Dada una muestra aleatoria $\mathcal{X} = \{x_1, \dots, x_n\}$ de tamaño n de variables aleatorias i.i.d. $\mathcal{N}(\mu, \sigma^2)$ los estimadores de máxima verosimilitud de μ y de σ^2 están dados por:

$$\hat{\mu} = \bar{x} = \sum_{i=1}^n \frac{x_i}{n}$$
$$\hat{\sigma}^2 = \sum_{i=1}^n \frac{(x_i - \hat{\mu})^2}{n} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}$$

Evaluando un estimador: sesgo y varianza



θ es el parámetro que hay que estimar

d_i son diferentes estimadores, notados con **x**

- Sea \mathcal{X} una muestra aleatoria de una población especificada por el **parámetro θ** , y sea **$d = d(\mathcal{X})$** un **estimador** de θ .

- Para evaluar la calidad este estimador, queremos medir cuánto difiere de θ , esto es

$$(d(\mathcal{X}) - \theta)^2$$

- Como esta diferencia es una variable aleatoria (depende de la muestra), necesitamos el “promedio” sobre todas las posibles \mathcal{X} , entonces consideramos el **error cuadrático medio** del estimador d definido como:

$$r(d, \theta) = E[(d(\mathcal{X}) - \theta)^2]$$

- El **sesgo** de un estimador está dado por:

$$b_{\theta}(d) = E[d(\mathcal{X})] - \theta$$

Si $b_{\theta}(d) = 0$ entonces d es un **estimador insesgado** de θ .

- El error cuadrático medio puede reescribirse como:

$$\begin{aligned} r(d, \theta) &= \underbrace{E \left[(d - E(d))^2 \right]}_{\text{varianza}} + \underbrace{[E(d) - \theta]^2}_{\text{sesgo}^2} \\ &= \text{var}(d) + [b_\theta(d)]^2 \end{aligned}$$

Ejemplo

- Si $\mathcal{X} = \{x_1, \dots, x_n\}$ es una muestra aleatoria de v.a. i.i.d. $\mathcal{N}(\mu, \sigma^2)$

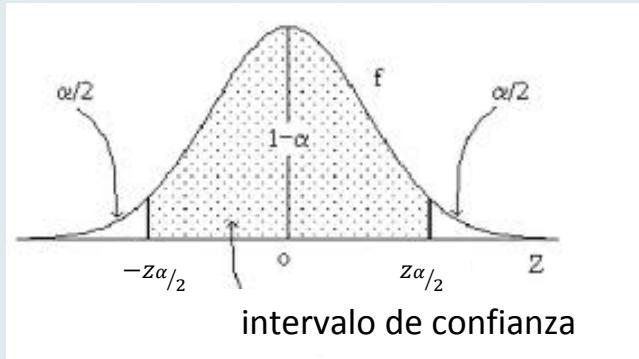
$$E(\hat{\mu}) = E(\bar{x}) = E(X_i) = \mu \quad \Rightarrow \quad b_\mu(\hat{\mu}) = 0$$

- Por lo tanto $\hat{\mu}$ es un estimador insesgado de μ , luego:

$$r(\hat{\mu}, \mu) = \text{var}(\hat{\mu}) = \text{var}(\bar{x}) = \frac{\text{var}(X_i)}{n} = \frac{\sigma^2}{n}$$

- Observemos que el error cuadrático medio decrece al crecer el tamaño de la muestra.

Intervalos de confianza



Un **intervalo de confianza** de nivel $(1 - \alpha)$ para un parámetro θ es un intervalo que contendrá con probabilidad $(1 - \alpha)$ a dicho parámetro.

Ejemplos

Distribución normal

- Si $\mathcal{X} = \{x_1, \dots, x_n\}$ es una muestra aleatoria de v.a. i.i.d. $\mathcal{N}(\mu, \sigma^2)$, el intervalo de confianza de nivel $(1 - \alpha)$ para la media μ cuando σ es conocido está dado por:

$$\left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

- Si $\mathcal{X} = \{x_1, \dots, x_n\}$ es una muestra aleatoria de v.a. i.i.d. $\mathcal{N}(\mu, \sigma^2)$, el intervalo de confianza de nivel $(1 - \alpha)$ para la media μ cuando σ es desconocido está dado por:

$$\left(\bar{x} - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \right)$$

donde

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$$

Distribución Bernoulli

- Si $\mathcal{X} = \{x_1, \dots, x_n\}$ es una muestra aleatoria de v.a. i.i.d. $B(p)$, el intervalo de confianza de *nivel asintótico* $(1 - \alpha)$ para la proporción p está dado por:

$$\begin{aligned} \left(\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right) &= \\ &= \left(\bar{x} - z_{\alpha/2} \sqrt{\frac{\bar{x}(1 - \bar{x})}{n}}, \bar{x} + z_{\alpha/2} \sqrt{\frac{\bar{x}(1 - \bar{x})}{n}} \right) \end{aligned}$$

- La longitud de este intervalo es:

$$L = 2z_{\alpha/2} \sqrt{\frac{\bar{x}(1 - \bar{x})}{n}}$$

- Despejando n resulta:

$$n = \frac{4z_{\alpha/2}^2 \bar{x}(1 - \bar{x})}{L^2}$$

- Pero en la ecuación anterior aparece \bar{x} que depende de n .
- Si consideramos que la función $f(x) = x(1 - x)$ tiene un máximo en $x = 0.5$, entonces:

$$\frac{4z_{\alpha/2}\bar{x}(1 - \bar{x})}{L^2} \leq \frac{z_{\alpha/2}}{L^2}$$

- Luego una forma conservadora de determinar el tamaño de la muestra n sería tomar:

$$n = \frac{z_{\alpha/2}}{L^2}$$

Ejemplo

- En 1983, el episodio de envenenamiento con Tylenol, alertó sobre la conveniencia empacar este producto de modo de que resista el manipuleo indebido.

- Se realizó un estudio de las actitudes de los consumidores hacia el nuevo empaque.
- De los 270 consumidores encuestados, 189 indicaron que estarían dispuestos a pagar más por un empaque resistente al manejo indebido.
- Denotemos con p a la proporción de todos los consumidores que pagarían más por el empaque.
- Entonces

$$\hat{p} = \frac{189}{270} = 0.7$$

y el desvío estándar de \hat{p} es:

$$\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = 0.0279$$

- Un intervalo de confianza asintótico de nivel 0.95 para p está dado por:

$$\hat{p} \pm z_{0.025} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = (0.645, 0.755)$$

Intervalo de confianza para hipótesis de valores discretos

- Queremos contestar la pregunta:
¿cuán bien se estima el $\text{error}_{\mathcal{D}}(h)$ a partir del $\text{error}_{\mathcal{S}}(h)$?
para el caso en el que h sea un hipótesis de valores discretos.
- Supongamos que queremos **estimar el error verdadero** para una hipótesis h de valores discretos mediante **el error muestral observado** a partir de una muestra S , donde:

- la muestra S contiene n ejemplos independientes entre sí e independientes de h , de acuerdo a una distribución \mathcal{D}

- $n \geq 30$

- la hipótesis h comete r errores sobre estos n ejemplos, i.e.

$$\text{error}_S(h) = r/n$$

- Bajo estos supuestos, la teoría estadística nos permite decir lo siguiente:

1. Dada ninguna información adicional, el valor más probable de $\text{error}_{\mathcal{D}}(h)$ es $\text{error}_S(h)$

2. Con una probabilidad aproximada del $(1-\alpha)\%$, el error verdadero $\text{error}_{\mathcal{D}}(h)$ está contenido en el intervalo

$$\left(\text{error}_S(h) - z_{\alpha/2} \sqrt{\frac{\text{error}_S(h)(1 - \text{error}_S(h))}{n}}, \right. \\ \left. \text{error}_S(h) + z_{\alpha/2} \sqrt{\frac{\text{error}_S(h)(1 - \text{error}_S(h))}{n}} \right)$$

- Hay que tener en cuenta este intervalo se aplica sólo al caso de:
 - hipótesis con **valores discretos**
 - se asume que la **muestra S** se toma aleatoriamente usando la **misma distribución** que la de los **futuros datos**

- se asume que los **datos** son **independientes** de la **hipótesis** que está siendo testeada.
- Además esta expresión sólo provee un intervalo de confianza aproximado.
Esta aproximación sobre buena cuando:
 - la muestra contiene al menos 30 ejemplos
 - el $\text{error}_S(h)$ no es un valor demasiado cerca de 0 ó 1.
- Un criterio más preciso para saber si la aproximación anterior es buena es:

$$n \text{error}_S(h)(1 - \text{error}_S(h)) \geq 5$$

Diferencias en error de dos hipótesis

- Consideremos el caso donde tenemos dos hipótesis h_1 y h_2 de alguna función objetivo de valores discretos.
- La hipótesis h_1 se testeó con una muestra S_1 que contenía n_1 ejemplos aleatorios, y h_2 se testeó con una muestra independiente S_2 que contenía n_2 ejemplos aleatorios de la misma distribución.
- Supongamos que queremos estimar la diferencia d entre los verdaderos errores de estas dos hipótesis

$$d \equiv error_D(h_1) - error_D(h_2)$$

- Primero hay que obtener un estimador para d .
Un **estimador** obvio para d sería tomar la **diferencia entre los errores muestrales**

$$\hat{d} \equiv error_{S_1}(h_1) - error_{S_2}(h_2)$$

- Se puede probar que \hat{d} es un estimador **insesgado** de d .
- ¿Cuál es la distribución de la variable aleatoria \hat{d} ?
 - Para n_1 y n_2 “grandes” (ambos ≥ 30), ambos errores siguen una distribución aproximadamente normal.

- Como la diferencia de dos distribuciones normales es también una distribución normal, \hat{d} tiene **distribución aproximada normal** con media d y varianza dada por la suma de las varianzas de $error_{S_1}(h_1)$ y $error_{S_2}(h_2)$, es decir:

$$\begin{aligned}\sigma_{\hat{d}}^2 &= \text{var}(\hat{d}) \\ &\approx \frac{error_{S_1}(h_1)[1 - error_{S_1}(h_1)]}{n_1} \\ &\quad + \frac{error_{S_2}(h_2)[1 - error_{S_2}(h_2)]}{n_2} \quad (\#)\end{aligned}$$

y por lo tanto,

$$\hat{d} \approx \mathcal{N}(d, \sigma_{\hat{d}}^2)$$

- Un intervalo de confianza asintótico para d esta dado por:

$$(\hat{d} - z_{\alpha/2} \sigma_{\hat{d}}, \hat{d} + z_{\alpha/2} \sigma_{\hat{d}})$$

- A pesar de que en el análisis anterior se considera el caso en el que h_1 y h_2 son testeadas con muestras independientes, también el intervalo anterior se puede utilizar en el caso donde h_1 y h_2 son testeadas con una única muestra S .

- En este caso se redefine \hat{d} como

$$\hat{d} = error_S(h_1) - error_S(h_2)$$

- La varianza para este nuevo \hat{d} es distinta a la calculada en $\sigma_{\hat{d}}^2$; si se utiliza la misma, el intervalo de confianza va ser algo conservativo.

Test de hipótesis

- Un **test de hipótesis** es un procedimiento para juzgar si una propiedad que se supone cumple una población estadística es compatible con lo observado en una muestra de dicha población.
- Mediante esta teoría, se aborda el problema estadístico considerando una hipótesis determinada H_0 y una hipótesis alternativa H_a , y se intenta dirimir cuál de las dos es la hipótesis verdadera, tras aplicar el problema estadístico a un cierto número de experimentos.

- Una vez realizado el test, se habrá optado por una de las dos hipótesis, y la decisión escogida coincidirá o no con la que en realidad es cierta.
- Se pueden dar los siguientes cuatro casos:

	H_0 es verdadera	H_0 es falsa
se acepta H_0	no hay error	error de tipo II
se rechaza H_0	error de tipo I	no hay error

$$P(\text{error tipo I}) = P(\text{rechazar } H_0 | H_0 \text{ es verdadera}) = \alpha$$

$$P(\text{error tipo II}) = P(\text{no rechazar } H_0 | H_0 \text{ es falsa}) = \beta$$

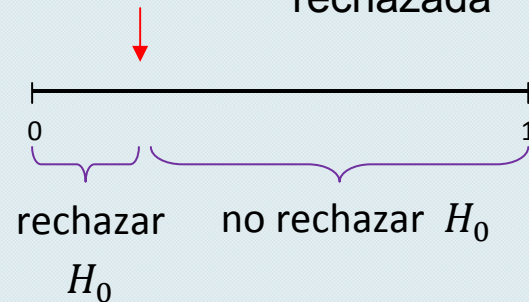
- Al diseñar un test de hipótesis, sería deseable hacerlo de tal manera que las probabilidades de ambos tipos de error fueran tan pequeñas como fuera posible.
- Sin embargo, con una **muestra de tamaño prefijado**, sólo se puede controlar uno de las dos probabilidades, y por lo tanto se elige **controlar**

$$P(\text{error tipo I}) = \alpha = \text{nivel del test}$$

- Para un test de nivel α , la $P(\text{error tipo II})$ se puede disminuir aumentando el tamaño de la muestra.

- El **p-valor** es el mínimo nivel de significación al que H_0 sería rechazada
 - $\text{p-valor} \leq \alpha \Rightarrow$ se rechaza H_0 a nivel α
 - $\text{p-valor} > \alpha \Rightarrow$ no se rechaza H_0 a nivel α

valor p: mínimo nivel al que H_0 puede ser rechazada



Ejemplos

Test asintótico binomial

- Denotemos con p la **proporción de una población** que cumple una propiedad específica.
- Si un individuo con esa propiedad se marca con un éxito, entonces p es la proporción de la población de éxitos.
- Supongamos que se toma una muestra $\mathcal{X} = \{x_1, \dots, x_n\}$ de tamaño n ($n \geq 30$) de esa población, donde

$$x_i = \begin{cases} 1 & \text{si el individuo } i \text{ cumple la propiedad} \\ 0 & \text{en caso contrario} \end{cases}$$

- Entonces, las variables aleatorias X_1, \dots, X_n son i.i.d. con distribución $\mathcal{B}(p)$.
- Vimos que el estimador de máxima verosimilitud de p es $\hat{p} = \bar{x}$ y aplicando el teorema central del límite, se puede ver que

$$\frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1 - p)}} \rightarrow^D \mathcal{N}(0,1)$$

- Supongamos que la hipótesis nula es $H_0: p = p_0$.

Entonces cuando H_0 es verdadera, el estadístico

$$Z = \frac{\sqrt{n}(\hat{p} - p_0)}{\sqrt{p_0(1 - p_0)}}$$

tiene distribución aproximada normal estándar.

- De esta manera, para las distintas hipótesis alternativas de la tabla obtenemos las siguientes regiones de rechazo:

Hipótesis alternativa	Región de rechazo
$H_a: p > p_0$	$Z \geq z_\alpha$
$H_a: p < p_0$	$Z \leq -z_\alpha$
$H_a: p \neq p_0$	$Z \geq z_{\alpha/2} \quad \text{ó} \quad Z \leq -z_{\alpha/2}$

Test para muestras apareadas

- Supongamos que se toma una muestra de n pares seleccionados de manera independiente $\mathcal{X} = \{(x_1, y_1), \dots, (x_n, y_n)\}$ donde suponemos que las variables aleatorias X_1, \dots, X_n e Y_1, \dots, Y_n tienen distribución normal.
- Entonces, si $D_i = X_i - Y_i$ para $i = 1, \dots, n$, las variables aleatorias D_1, \dots, D_n tienen distribución $\mathcal{N}(\mu_D, \sigma_D^2)$ donde
$$\mu_D = E(X_i - Y_i) = \mu_x - \mu_y$$
- Nos interesa estudiar las diferencias entre las medias μ_x y μ_y .

- Supongamos que la hipótesis nula es $H_0: \mu_D = \Delta_0$.

Entonces cuando H_0 es verdadera, el estadístico

$$T = \frac{\sqrt{n}(\bar{d} - \Delta_0)}{S_D}$$

tiene distribución t de Student con $n - 1$ grados de libertad, siendo

$$s_D^2 = \sum_{i=1}^n \frac{(d_i - \bar{d})^2}{n - 1}$$

- De esta manera, para las distintas hipótesis alternativas de la tabla obtenemos las siguientes regiones de rechazo:

Hipótesis alternativa	Región de rechazo
$H_a: \mu_D > \Delta_0$	$T \geq t_{\alpha, n-1}$
$H_a: \mu_D < \Delta_0$	$T \leq -t_{\alpha, n-1}$
$H_a: \mu_D \neq \Delta_0$	$T \geq t_{\alpha/2, n-1} \quad \text{ó} \quad T \leq -t_{\alpha/2, n-1}$