ESTADÍSTICA

· Aplica la probabilidad al estudio de datos experimentales

· Realizamos una medida N veces, obteniendo un conjunto $\{x_i\}_{i \in \{1,\dots,N\}}$ $\Rightarrow$ MUESTRA

· Tanto cada uno de los x_i como el conjunto $\vec{x} = (x_1, \dots, x_N)$ son variables aleatorias (pues tienen una serie de valores posibles, con cierta probabilidad). Nuestro objetivo es podemos hallar su distribución de probabilidad o población.

· Las variables se dicen estadísticamente independientes si la medida de una no afecta a las posteriores. En ese caso, $p(\vec{x}) = \prod_i p_i(x_i)$

Si además todas las p_i son iguales (lo cual no ocurre, por ejemplo, si yo mido con más precisión algunos datos) la muestra se dice aleatoria de tamaño N de la población $p(x)$

· Salvo que se diga lo contrario, nos centraremos en este último caso.

· En general, $p(x_i)$ dependerá de unos parámetros $\vec{a} = (a_1, \dots, a_m)$. Nuestro objetivo es, dado $\{x_i\}$ hallar $\vec{a}$ y determinar $p = p(x_i | \vec{a})$.

i.e., la muestra

ESTADÍSTICOS

· Son funciones de la muestra. Ejemplos:

Son las medias de la muestra, NO de la población

$$\langle x \rangle = \frac{1}{N} \sum x_i \rightarrow \text{Media}$$

$$\overline{x}_{rms} = \sqrt{\langle x^2 \rangle} \rightarrow \text{Error cuadrático medio (root-mean square)}$$

$$s^2 = \frac{1}{N} \sum (x_i - \langle x \rangle)^2 \rightarrow \text{Varianza}$$

$$\sigma = \sqrt{s^2} \rightarrow \text{Desviación típica}$$

$$m_n = \langle x^n \rangle \rightarrow \text{Momentos}$$

$$m_n = \langle (x - \langle x \rangle)^n \rangle \rightarrow \text{Momentos centrales}$$

$$g_n = \frac{m_n}{m_2^{n/2}} \rightarrow \text{Momentos centrales normalizados}$$

· Si yo mido parejas de valores (x_i, y_i), es útil definir:

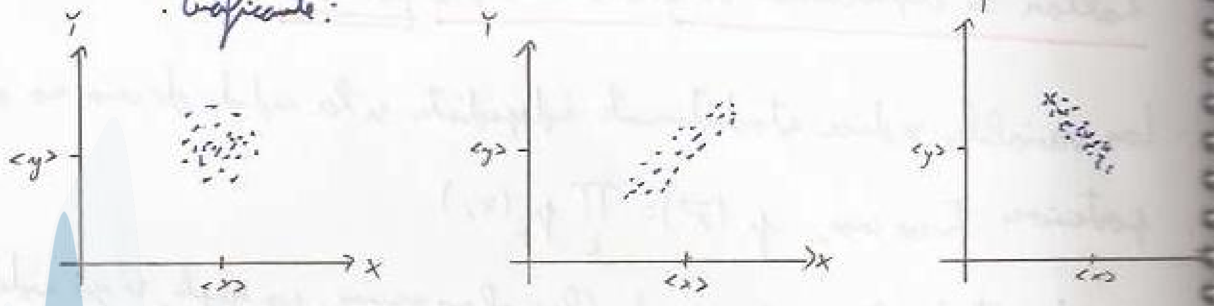
$$V_{xy} = \frac{1}{N}\sum_{i=1}^{N} (x_i - \langle x\rangle)(y_i - \langle y\rangle) = \langle xy\rangle - \langle x\rangle\langle y\rangle \rightarrow \text{Covarianza}$$

$$r_{xy} = \frac{V_{xy}}{\sigma_x \sigma_y} \rightarrow \text{correlación}$$

· $r_{xy} \in [-1, 1]$

· x e y independientes $\Rightarrow r_{xy} = 0$

· Gráficamente:



$r_{xy} = 0$ $\qquad$ $r_{xy} \sim 1$ $\qquad$ $r_{xy} \sim -1$

ESTIMADORES

· Sirven para estimar los parámetros $\{a_i\}$. Se denotan por $\hat{a}_i$

· Son, en general, función de la muestra:

$$\hat{a} = \hat{a}(\vec{x})$$

· Como depende de los datos, es una variable aleatoria $\Rightarrow$ Podemos definir una distribución de probabilidad:

$$p(\hat{a}|a)\, d\hat{a} = p(\vec{x}|a)\, d^N x$$

· Deben cumplir:

1- CONSISTENCIA: $\lim\limits_{N\to+\infty} \hat{a}(\vec{x}) = a$

2- BIAS → exactitud / sesgado : $b(\hat{a}) \equiv \langle\hat{a}\rangle - a$; con $\langle\hat{a}\rangle = \int p(\hat{a}|a)\,\hat{a}\, d\hat{a} = \int p(\vec{x}|a)\,\hat{a}(\vec{x})\, d^N x$

Por consistencia, $\lim\limits_{N\to+\infty} b(\hat{a}) = 0$

3- EFICIENCIA: Sea la varianza $V(\hat{a}) = \langle(\hat{a} - \langle\hat{a}\rangle)^2\rangle$

Si la varianza es pequeña, el estimador se dice eficiente.

Cumple el límite de Fisher:

$$V(\hat{a}) \geq \frac{1 + \frac{\partial b}{\partial a}}{\left\langle -\frac{\partial^2 \ln p}{\partial a^2}\right\rangle} \quad ; \text{con} \quad \begin{array}{l} \cdot\ b \text{ el bias} \\ \cdot\ p = p(\vec{x}|a) \end{array}$$

· Estimadores importantes:

· MEDIA: - Si todas las p_i tienen la misma σ: $\bar{x} \equiv \frac{1}{N}\sum_i x_i$

- Si cada p_i tiene una σ_i: $\bar{x} \equiv \frac{\sum_i x_i/\sigma_i^2}{\sum_i 1/\sigma_i^2}$

· DESVIACIÓN TÍPICA: $\hat{\sigma} \equiv \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N-1}}$

Ejemplo: estimador $\bar{x}$, y distribución subyacente normal: $p(\bar{x}|\mu,\sigma) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$

Para N suficientemente grande, por T.C.L., $p(\bar{x}|\mu,\sigma) = \frac{1}{\sqrt{\frac{2\pi\sigma^2}{N}}} e^{-\frac{(\bar{x}-\mu)^2}{2\sigma^2/N}}$

En este caso, $\forall N$

· Consistencia: $p(\bar{x}|\mu,\sigma) \xrightarrow[N\to+\infty]{} \delta(\bar{x}-\mu)$ (distribución discreta) $\Rightarrow$ $\bar{x}$ sólo puede tomar el valor μ

$$\Downarrow$$

$$\langle \bar{x} \rangle \to \mu$$

· Bias: $\langle \bar{x} \rangle = \mu$ para N suficientemente grande (pues es una Gaussiana centrada en μ)

$$b = \langle \bar{x} \rangle - \mu = 0 \quad \text{(en este caso es una igualdad } \forall N)$$

· Eficiencia: Al ser una Gaussiana, $V(\bar{x}) = \frac{\sigma^2}{N}$. Esto satura el límite de Euler (la eficiencia es máxima):

$$\frac{\partial b}{\partial \mu} = 0$$

$$\frac{\partial^2 \ln p}{\partial \mu^2} = \frac{\partial^2}{\partial \mu^2}\left[-\frac{1}{2}\sum_i \left[\ln(2\pi\sigma^2) + \frac{(x_i-\mu)^2}{\sigma^2}\right]\right] = -\frac{N}{\sigma^2}$$

$$\frac{1+\frac{\partial b}{\partial \mu}}{\left\langle -\frac{\partial^2 \ln p}{\partial \mu^2}\right\rangle} = \frac{\sigma^2}{N}$$

ERRORES EN LOS ESTIMADORES

· El error se suele expresar:

· Si el bias = 0, se usa el error estándar $\boxed{\sigma_{\hat{a}} = \sqrt{V[\hat{a}]}}$

· Si el bias $\neq 0$, se toma el error cuadrático medio $\boxed{\varepsilon_{\hat{a}} = \sqrt{\langle(\hat{a}-a)^2\rangle} = \sqrt{V(\hat{a})^2 + b(a)^2}}$

Demostración:

$$\langle(\hat{a}-a)^2\rangle = \langle(\hat{a}-\langle\hat{a}\rangle+\langle\hat{a}\rangle-a)^2\rangle = \langle(\hat{a}-\langle\hat{a}\rangle)^2 - 2(\hat{a}-\langle\hat{a}\rangle)(\langle\hat{a}\rangle-a) + (\langle\hat{a}\rangle-a)^2\rangle =$$

$$= \underbrace{\langle(\hat{a}-\langle\hat{a}\rangle)^2\rangle}_{V(\hat{a})^2} + 2\underbrace{\langle\hat{a}-\langle\hat{a}\rangle\rangle}_{\langle\hat{a}\rangle-\langle\hat{a}\rangle=0}\langle\langle\hat{a}\rangle-a\rangle + \underbrace{\langle\langle\hat{a}\rangle-a\rangle^2}_{b(a)^2}$$

bias

En el ejemplo de $\bar{x}$, al no tener bias, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$ para N suficientemente grande para poder aplicar T.C.L. Para medidas con distinto error, propagando errores, $\sigma_{\bar{x}} = \frac{1}{\sum_i \frac{1}{\sigma_i^2}}$

INTERVALOS DE CONFIANZA

· Una vez dado el error: $a \in [a_-, a_+]$

$$a = \hat{a}\,^{+d = a_+ - \hat{a}}_{-c = \hat{a} - a_-}$$

$$a = \hat{a} \pm b \quad d=c=b$$

Nos gustaría dar la probabilidad de que a esté en el intervalo dado. A esto se le conoce como nivel de confianza.

· Para ello, definimos $\forall \alpha, \beta \in [0,1]$:

$$\boxed{\begin{array}{l} \hat{a}_\alpha(a) \text{ cumple } P(\hat{a} \leq \hat{a}_\alpha(a)|a) = \alpha \\ \hat{a}_\beta(a) \text{ cumple } P(\hat{a} \geq \hat{a}_\beta(a)|a) = \beta \end{array}}$$

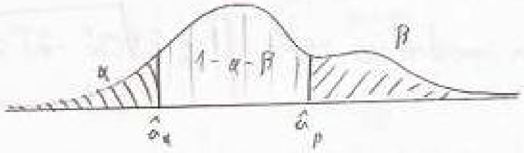
→ Fijado a, la distribución de probabilidad de $\hat{a}$ está fija. Me pregunto por el punto tq el área entre $-\infty$ y ese punto vale α

Son funciones de α, β y a
↓
Pq. la distribución de probabilidad de $\hat{a}$ depende de a

Así:

$$\boxed{P\left(\hat{a}_\alpha(a) \leq \hat{a} \leq \hat{a}_\beta(a)\right) = 1-\alpha-\beta}$$

Representación gráfica



· Supongamos que, dada una muestra, estimo el valor $\hat{a}_{obs}$. Defino $\forall\, \alpha, \beta \in [0,1]$:

$$\boxed{\begin{aligned} a_+ &= \hat{a}_\alpha^{-1}(\hat{a}_{obs}) \\ a_- &= \hat{a}_\beta^{-1}(\hat{a}_{obs}) \end{aligned}}$$

¿Con qué valor de a tengo que fijar mi distribución para que el área entre $-\infty$ y $\hat{a}_{obs}$ sea α?

Cumple:

$$\boxed{\begin{aligned} P(a_+ < a) &= \alpha \\ P(a_- > a) &= \beta \end{aligned}}$$

· Por lo que:

$$\boxed{P(a_- \leq a \leq a_+) = 1-\alpha-\beta}$$

Si $\alpha=\beta$, el intervalo de confianza se dice centrado

Ejemplo: Gaussiana

$$P(\hat{a}|a) = \frac{1}{\sqrt{2\pi\sigma_{\hat{a}}^2}}\, e^{-\frac{(\hat{a}-a)^2}{2\sigma_{\hat{a}}^2}}$$

Así:

$$\alpha = \int_{-\infty}^{\hat{a}_\alpha} d\hat{a}\, \frac{1}{\sqrt{2\pi\sigma_{\hat{a}}^2}}\, e^{-\frac{(\hat{a}-a)^2}{2\sigma_{\hat{a}}^2}}$$

$$\Updownarrow$$

$$\alpha = \phi\left(\frac{\hat{a}_\alpha - a}{\sigma_{\hat{a}}}\right)$$

$$\Updownarrow$$

$$\boxed{\hat{a}_\alpha(a) = a + \sigma_{\hat{a}}\, \phi^{-1}(\alpha)}$$

Análogamente, $\boxed{\hat{a}_\beta(a) = a - \sigma_{\hat{a}}\, \phi^{-1}(\beta)}$

De esta manera:

$$\hat{a}_{obs} = a_+ + \sigma_{\hat{a}}\, \phi^{-1}(\alpha) \;;\quad a_+ = \hat{a}_{obs} - \sigma_{\hat{a}}\, \phi^{-1}(\alpha)$$

$$\hat{a}_{obs} = a_- - \sigma_{\hat{a}}\, \phi^{-1}(\beta) \;;\quad a_- = \hat{a}_{obs} + \sigma_{\hat{a}}\, \phi^{-1}(\beta)$$

Es decir:

$$\boxed{\begin{aligned} a_- &= \hat{a}_{obs} - \sigma_{\hat{a}}\, \phi^{-1}(1-\beta) \\ a_+ &= \hat{a}_{obs} + \sigma_{\hat{a}}\, \phi^{-1}(1-\alpha) \end{aligned}}$$

Así, dados α y β (la probabilidad con la que queremos dar nuestro error), somos capaces de hallar los extremos del error, a_+ y a_-

Viceversa, dados a_+ y a_-, es fácil obtener α y β, sin más que aplicar que

$$\hat{a}_{obs} = \hat{a}_\alpha(a_+) \Longrightarrow \alpha = P(\hat{a} \leq \hat{a}_{obs} | a_+)$$ → Fijado a_+, en su distribución y calculo el área hasta $\hat{a}_{obs}$ (lo mismo para β y a_-)

En la Gaussiana, por último:

$$a = \hat{a}_{obs} \pm \sigma_a \Rightarrow 68\%$$

$$a = \hat{a}_{obs} \pm 2\sigma_a \Rightarrow 95'4\%$$

$$a = \hat{a}_{obs} \pm 3\sigma_a \Rightarrow 99'7\%$$