

# Estimación por intervalos: estimación autosuficiente de intervalos de confianza (bootstrap)

- La **media muestral** es un estimador con una propiedad muy especial:
  - Su **precisión** (inversa de la varianza de la distribución muestral) se **puede conocer a través de la varianza de la muestra**:
  - Recordar que para estimar la varianza de la distribución muestral de la media hacíamos  $E[(\bar{x} - \mu)^2] = Var(\bar{x}) = \sigma^2/n$ .
  - Así si estimamos la varianza  $\sigma^2$  a través de la varianza de la muestra  $\hat{s}^2$  entonces ponemos que  $\hat{s}^2(\bar{x}) = \frac{\hat{s}^2}{n} = \frac{1}{n(n-1)} \sum (x_i - \bar{x})^2$
  - La inversa de este valor,  $1/\hat{s}^2(\bar{x})$ , es la precisión del estimador media muestral.
  - Esta expresión es válida en general, ya que no depende del modelo de distribución de probabilidad que genera la muestra, solo depende de los datos muestrales.
  - Ser conscientes que al estimar cualquier otro parámetro o característica de la población (como varianza, curtosis, asimetría, etc.), la precisión de la estimación depende de la distribución que genera los datos.

# Estimación por intervalos: estimación autosuficiente de intervalos de confianza (bootstrap)

- Si utilizamos un estimador de máxima verosimilitud podemos conocer su varianza asintótica, pero esa medida puede ser poco precisa para muestras pequeñas, y depende de la hipótesis sobre la distribución.
- Existen métodos de simulación por computador como estimación **jackknife** y **bootstrap** que son generales para obtener precisión de un estimador de forma aproximada sin hacer hipótesis respecto a la distribución.
- Estos métodos hoy en día se pueden usar por la potencia y rapidez de los ordenadores digitales.
- **Jackknife:** Quenouille, M. H. (September 1949). "Problems in Plane Sampling". *The Annals of Mathematical Statistics*. **20** (3): 355–375. [doi:10.1214/aoms/1177729989](https://doi.org/10.1214/aoms/1177729989). [JSTOR 2236533](https://www.jstor.org/stable/2236533).
- **Bootstrap:** [Efron, B. \(1979\). "Bootstrap methods: Another look at the jackknife". \*The Annals of Statistics\*. \*\*7\*\* \(1\): 1–26.](#)

# Estimación por intervalos: estimación autosuficiente de intervalos de confianza (bootstrap)

- **Bootstrap se basa en calcular directamente la varianza del estimador considerando la muestra como si fuese toda la población y aplicando el método de Montecarlo (sección 5.7.2 del libro si queréis saber más sobre montecarlo) para obtener réplicas de la muestra.**

# Estimación por intervalos: estimación autosuficiente de intervalos de confianza (bootstrap)

- Siempre se parte de una **muestra de datos**  $(x_1, \dots, x_n)$  y se realiza el siguiente procedimiento:
- **Primero:** Consideramos la muestra como una población de una variable que toma los  $n$  valores  $(x_1, \dots, x_n)$  con probabilidad  $1/n$ . De esta muestra se extrae una m.a.s. de tamaño  $n$  (igual que el tamaño de muestra) mediante el método de Montecarlo. Es decir obtenemos una muestra al azar con reemplazamiento de los valores observados. Notar que esta muestra generada no coincidirá, en general, con la muestra original. Supongamos que la primera muestra obtenida es  $(y_1^1, \dots, y_n^1)$ .

# Estimación por intervalos: estimación autosuficiente de intervalos de confianza (bootstrap)

- **Segundo:** Calculamos en la muestra generada anterior el estimador cuya precisión queremos estimar:
  - $\hat{\vartheta}_1 = \hat{\vartheta}(y_1^1, \dots, y_n^1)$
- **Tercero:** Repetimos los pasos **primero** y **segundo** un número  $B$  de pasos grande (por ejemplo, 1000 veces), así obtenemos una secuencia del estimador  $\hat{\vartheta}_1, \dots, \hat{\vartheta}_B$  en cada uno de las  $B$  m.a.s. Así la estimación de su media y su varianza es:
  - $\hat{\vartheta}_m = 1/B \sum \hat{\vartheta}_i, \quad \text{Var}(\hat{\vartheta}) = 1/B \sum (\hat{\vartheta}_i - \hat{\vartheta}_m)^2, \quad \forall i \in (1, \dots, B)$

# Estimación por intervalos: estimación autosuficiente de intervalos de confianza (bootstrap)

- Se puede demostrar que en condiciones generales, este método obtiene asintóticamente la varianza del estimador  $\hat{\vartheta}$ , y que el intervalo de confianza a un nivel  $1 - \alpha$  se puede obtener de la distribución de  $B$  valores de  $\hat{\vartheta}_i$ .
- Para el intervalo de confianza se obtienen los valores  $\hat{\vartheta}_{\text{INF}}$  y  $\hat{\vartheta}_{\text{SUP}}$  tales que:
  - $P(\hat{\vartheta}_{\text{INF}} \leq \hat{\vartheta}_i \leq \hat{\vartheta}_{\text{SUP}}) = 1 - \alpha$
- Los valores del intervalo de confianza  $\hat{\vartheta}_{\text{INF}}$  y  $\hat{\vartheta}_{\text{SUP}}$  se calculan ordenando los  $\hat{\vartheta}_i$  y tomando los valores situados en las posiciones  $[B \times \alpha/2]$  y  $[B \times (1 - \alpha/2)]$ , donde el símbolo  $[ ]$  significa redondear al entero más cercano.
- Darse cuenta que suponemos que los  $B$  valores cubren el 100% de la distribución de la varianza del estimador  $\hat{\vartheta}$ .
- Ver el ejemplo 8.6 y 8.7, que se puede simular en Python también.

# Ejemplos para realizar en casa

- Realizar para casa los ejemplos del libro:
  - 8-1, 8-2, 8-6 y 8-7.

# Inferencia Estadística: Estimación por intervalos

## Estimación bayesiana

- Introducción
- Distribuciones a priori
  - Distribución conjugada a priori
- Estimación puntual
- Estimación de poblaciones normales con varianza conocida



# Estimación bayesiana: introducción

- Los **métodos de estimación que hemos visto hasta ahora** funcionan **bien** con **muestras grandes**.
- Veamos varios ejemplos con muestras pequeñas:
- **Ejemplo 1:** Supongamos que nos preguntamos por la proporción de estudiantes que leen un determinada novela, y que en una muestra pequeña (30 estudiantes por ejemplo) obtenemos 0 personas que han leído el libro.
- Qué inferencia podemos hacer sobre el parámetro  $p$ , suponiendo un modelo binomial.
- Según lo que hemos visto, uno puede calcular el estimador máximo-verosímil del parámetro de una población binomial.

# Estimación bayesiana: introducción

- En este caso si uno hace los cálculos aprendidos (función de verosimilitud de  $p$ , y luego optimización del parámetro derivando respecto de  $p$  e igualando a cero) la **estimación máximo verosímil del parámetro  $p$  es la frecuencia en la muestra**, que en este caso, nos ha salido cero.
- Así la estimación máximo verosímil del número de estudiantes que han leído esta novela es cero y para cuantificar la precisión de esta estimación nos encontramos que la varianza estimada es cero, o la **precisión infinita**.
- Así no parece razonable aplicar los métodos estudiados.

# Estimación bayesiana: introducción

- El problema es que estamos aplicando un método máximo verosímil, que tiene buenas propiedades en muestras grandes, pero no en pequeñas o medianas muestras.
- **Ejemplo2:** Como segundo ejemplo, supongamos que tratamos de estimar la edad del más veterano de los estudiantes de una universidad mediante esa misma muestra de tamaño 30.
- Como desconocemos la distribución de edades, podríamos estimar como valor máximo de la variable edad el mayor valor observado en la muestra.

# Estimación bayesiana: introducción

- Supongamos que la persona de mayor edad es de 21 años.
- Esta estimación es intuitivamente muy deficiente.
- Es difícil que el más veterano este en la muestra.
- **Ejemplo3:** Como tercer ejemplo supongamos que una moneda de un euro, la tiramos 10 veces y obtenemos 7 caras y 3 cruces. Con esto queremos decir que la probabilidad de cara en esta moneda es 0,7. Claramente no.

# Estimación bayesiana: introducción

- Estos tres ejemplos tienen en común la existencia de cierta información a priori respecto al parámetro que tratamos de estimar, que no se tiene en cuenta en el proceso de inferencia: **información a priori**.
- Ignorar la información inicial que tenemos respecto a un parámetro a estimar no es importante si la muestra es grande, pero puede serlo cuando la muestra es pequeña.
- Es decir si la muestra es muy grande podemos despreciar al información a priori frente a la gran cantidad de información que tenemos en la muestra.

# Estimación bayesiana: introducción

- Sin embargo con una pequeña muestra desperdiciar la información a priori no nos permite hacer una inferencia correcta, ya que la información a priori es significativa frente a los datos (la muestra).
- **Por tanto la inferencia bayesiana es un procedimiento general para combinar nuestra información a priori con la muestra para obtener una inferencia que tenga en cuenta toda la información existente en el problema.**

# Estimación bayesiana: introducción

- En el **enfoque bayesiano** un **parámetro no es una constante** desconocida, **sino una variable aleatoria** sobre la que podemos establecer a priori una distribución de probabilidad que refleje nuestro conocimiento del problema a priori previo a realizar el proceso de inferencia.
- El proceso de inferencia se obtiene aplicando el cálculo del **teorema de Bayes**: combina la **información a priori** con la **información de la muestra** y se obtiene la **distribución del parámetro condicionada a la información disponible**.
- Por tanto, suponemos que **antes de tomar la muestra** se dispone de cierta información respecto al parámetro que se representa mediante una **distribución inicial o a priori**,  $p(\theta)$ .

# Estimación bayesiana: introducción

- Supongamos que  $f(X|\theta)$  es la **probabilidad de obtener una muestra para cada valor posible del parámetro**.
- Después de tomar la muestra  $X = (x_1, \dots, x_n)$ , en la función  $f(X|\theta)$  los datos son fijos, ya que han sido observados, y los parámetros  $\theta$  son variables para esa muestra observada.
- Así **cuando la muestra se observa**  $f(X|\theta) = \ell(\theta|X)$ , que es lo que habíamos dicho que era la función de **verosimilitud**.



# Estimación bayesiana: introducción

- Podemos combinar las cantidades anteriores a través del teorema de Bayes para encontrar la probabilidad a posteriori  $p(\theta|X)$ :

$$p(\theta|X) = \frac{f(X|\theta)p(\theta)}{\int f(X|\theta)p(\theta)d\theta}$$

Recordar TMA Bayes:

$$P(\theta | X) = P(X | \theta)P(\theta) / P(X) = P(X | \theta)P(\theta) / \sum_{\theta} P(X | \theta) P(\theta)$$

- La distribución a posteriori  $p(\theta|X)$  contiene toda la información para hacer inferencias respecto al parámetro.
  - Si se desea un estimador puntual, se tomará la media o la moda de dicha distribución
  - Si se desea un intervalo de confianza, se tomará la zona que encierre una probabilidad fijada en dicha distribución.
- En resumen, una vez obtenida la distribución de probabilidad del parámetro, el problema de estimación queda resuelto de manera precisa.

# Estimación bayesiana: introducción

- Observar que el denominador  $m(X) = \int f(X|\theta)p(\theta)d\theta$  y como función de  $X$  (datos muestrales) representa la distribución marginal de los datos, con independencia de los valores de los parámetros (esta integrado sobre el parámetro).
- Esta distribución se suele denomina **distribución predictiva** y observando la expresión nos damos cuenta que es una media ponderada de las verosimilitudes  $f(X|\theta)$  por las probabilidades que la distribución a priori asignada a los posibles valores del parámetro  $\theta$  a estimar.

# Estimación bayesiana: introducción

- **Cuando observamos la muestra este denominador es una constante**, y el cálculo de la probabilidad a posteriori  $p(\theta|X)$  se simplifica.
- Por lo tanto el denominador es solo es una constante de normalización para que la integral de numerador sea la unidad:
  - De esta manera la densidad de probabilidad  $p(\theta|X)$ , esta bien definida.
- Así podemos poner que
  - $p(\theta|X) = kf(X|\theta)p(\theta) = k\ell(\theta|X)p(\theta)$ .

# Estimación bayesiana: introducción

- Así tenemos  $p(\theta|X) = k\ell(\theta|X)p(\theta)$ , que nos dice que podemos calcular la distribución posteriori multiplicando, para cada valor del parámetro  $\theta$ , su verosimilitud  $\ell(\theta|X)$  por su probabilidad a priori  $p(\theta)$ .
- Observar que la constante  $k$  es irrelevante para la forma de la probabilidad a posteriori  $p(\theta|X)$ , como hemos dicho es un factor de normalización que se puede determinar al final con la condición de que la integral de  $p(\theta|X)$  sea la unidad.
- Así el teorema de Bayes se puede resumir como:
  - $A \text{ posteriori} \propto \text{verosimilitud} \times A \text{ priori}$
- Donde el símbolo  $\propto$  significa proporcional.

# Estimación bayesiana: introducción

- En el caso particular de que  $p(\theta)$  sea aproximadamente constante sobre el rango de valores en los que la verosimilitud no es nula, se dice que la distribución a priori,  $p(\theta)$ , es no informativa, y por tanto la distribución a posteriori  $p(\theta|X)$ , vendrá determinada únicamente por la función de verosimilitud  $\ell(\theta|X)$ .
- Una ventaja adicional del enfoque bayesiano es su facilidad para **procesar información secuencialmente**. Esto es muy importante para afinar la estimación.
- Supongamos que después de calcular la distribución a posteriori con la observación de una muestra  $X$  observamos una nueva muestra de la misma población  $Y$ , independiente de la primera.
- Entonces, la nueva distribución a posteriori final se puede actualizar con la distribución inicial a posteriori,  $p(\theta|X)$ , a través de la expresión:
  - $p(\theta|XY) = k\ell(\theta|Y)p(\theta|X)$ .

# Estimación bayesiana: introducción

- Naturalmente la expresión anterior se puede obtener de la aplicación de la estimación bayesiana a una muestra ampliada  $[X,Y]$  **donde la muestra  $X$  e  $Y$  son independientes:**
  - $p(\theta|XY) = k\ell(\theta|XY)p(\theta) = kf(XY|\theta)p(\theta) = kf(X|\theta)f(Y|\theta)p(\theta) = k\ell(\theta|X)\ell(\theta|Y)p(\theta) = k\ell(\theta|Y)\ell(\theta|X)p(\theta) = k\ell(\theta|Y)p(\theta|X) = p(\theta|XY)$
- Así la estimación bayesiana proporciona pues un procedimiento automático para expresar el aumento de nuestro conocimiento respecto al parámetro a medida que se recibe información adicional.

# Estimación bayesiana: distribuciones a priori

- La **mayor dificultad práctica del enfoque bayesiano** es cómo especificar la **distribución a priori**: normalmente la información de que disponemos es cualitativa y el enfoque bayesiano requiere que establezcamos una distribución de probabilidad sobre sus valores.
- Así se pueden considerar cuatro casos diferentes.
- **Primero:** La distribución a priori proviene de estudios anteriores y se conoce objetivamente.
- **Segundo:** La distribución a priori puede ser importante respecto a la muestral, pero la información existente es subjetiva y no formalizada. Por ejemplo podemos elegir una distribución a priori que refleje globalmente nuestra opinión sobre la distribución a priori del parámetro, en particular la moda a priori y el rango de valores posibles, si la distribución es o no simétrica, etc. (mirar ejemplo en el libro, figura 9.2).

# Estimación bayesiana: distribuciones a priori

- **Tercero:** La información a priori es pequeña con relación a la información muestral. Podemos elegir una distribución a priori que refleje globalmente nuestra opinión, en particular la moda a priori y el rango de valores posibles, pero sin preocuparnos mucho del resto de los detalles. En este caso se suelen escoger distribuciones conjugadas (lo vemos después).
- **Cuarto:** La información a priori es despreciable frente a la información muestral, o no queremos tenerla en cuenta en el proceso de inferencia:
  - En este caso podemos utilizar los métodos clásicos que hemos estudiado anteriormente.
  - O también podemos utilizar el enfoque bayesiano con una distribución a priori *no informativa* o de *referencia* (ver sección 9.2.2).



# Estimación bayesiana: distribución conjugada

- La idea de la **distribución conjugada** es expresar aproximadamente nuestra información a priori con una distribución que facilite el análisis.
- Un ejemplo es utilizar una familia de distribuciones a priori que tiene la misma forma que la verosimilitud, de manera que la posterior pueda calcularse fácilmente al pertenecer a la misma familia que la priori.
- A estas familias se las denomina *conjugadas*.
- Una clase  $\mathcal{C}$  de distribuciones a priori para un parámetro  $\theta$  es conjugada si cuando la *a priori* pertenece a esa clase,  $p(\theta) \in \mathcal{C}$ , entonces también lo hace la *a posteriori*  $p(\theta|X) \in \mathcal{C}$ .
- La distribución conjugada *a priori* se elige tomando como distribución la verosimilitud, y modificando los valores de las constantes para que la función resultante sea una función de densidad y tenga características coincidentes con nuestra información a priori (ver ejemplo para el modelo binomial en el libro en sección 9.2.1).

# Estimación bayesiana: distribución conjugada

- Como segundo ejemplo, supongamos que queremos estimar la media,  $\theta$ , de una población normal de varianza conocida,  $\sigma$ .
- La verosimilitud para una normal de varianza conocida es (ejemplo 7.10)
  - $f(X|\theta) = \ell(\mu|X) = \ell(\theta) = k \exp(-\frac{n}{2\sigma^2}(\bar{x} - \theta)^2)$ , dada una muestra  $(x_1, \dots, x_n)$ .
- Y la a priori conjugada la escogemos como
  - $p(\theta) = k \exp(-\frac{n_0}{2\sigma^2}(\theta - \mu_0)^2) = k \exp(-\frac{1}{2\sigma_0^2}(\theta - \mu_0)^2)$ .
- Siendo  $\mu_0$  un parámetro que determina la media de la distribución y  $n_0$  determina la desviación típica, dando así la forma de la distribución  $p(\theta)$ , para que la función resultante coincida con nuestra opinión sobre la a priori. La constante  $K$  se ajusta para que  $p(\theta)$  este normalizada.

# Estimación bayesiana: estimación puntual

- Si queremos hacer estimación puntual con la a posteriori es necesario elegir un valor único para el parámetro.
- Para ello podemos optar por las diferentes posibilidades:
- **Primera:** Seleccionar el máximo (la moda) de la distribución a posteriori, que es el valor más probable. Cuando la información que nos da la distribución a priori sea pequeña con relación a la proporcionada por la verosimilitud, la a posteriori será análoga a la verosimilitud y su moda es el estadístico máximo-verosímil. Por tanto, en este caso el enfoque bayesiano coincide con el enfoque máximo-verosímil.
- **Segunda:** Definir un criterio de optimalidad y deducir el estimador a partir de él. Esto equivale a definir *una función de pérdida*,  $g(\vartheta, \hat{\vartheta})$  que indique la penalización de tomar  $\hat{\vartheta}$  como estimador cuando el valor real es  $\vartheta$ .

## Estimación bayesiana: estimación puntual

- La función de pérdida más frecuente escogida es la cuadrática  $g(\boldsymbol{\vartheta}, \hat{\boldsymbol{\vartheta}}) = k(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})^2$ .
- El criterio de elección será escoger como estimador aquel valor  $\hat{\boldsymbol{\vartheta}}$  que haga en promedio la pérdida mínima.
- Esto quiere decir  $E[g(\boldsymbol{\vartheta}, \hat{\boldsymbol{\vartheta}})] = kE[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})^2]$ , donde la esperanza se toma respecto a la distribución de  $\boldsymbol{\vartheta}$ ,  $p(\boldsymbol{\vartheta})$ .
- A la pérdida promedio se denomina **riesgo del estimador**.

# Estimación bayesiana: estimación de la media en poblaciones normales con varianza conocida

- Supongamos que se desea estimar la media de una población normal con varianza conocida y que la información inicial respecto a  $\mu$  se traduce en una distribución a priori  $N(\mu_0, \sigma_0)$ , suponiendo una distribución conjugada del tipo que hemos visto anteriormente para la distribución a priori:
  - $f(\mu) = k \exp(-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2)$
- Así la verosimilitud para una normal de varianza  $\sigma^2$ 
  - $\ell(\mu|X) = k' \exp(-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2)$

# Estimación bayesiana: estimación de la media en poblaciones normales con varianza conocida

➤ Y la a posteriori conjugada es entonces:

➤ 
$$f(\mu|X) = k'' \ell(\mu|X) f(\mu) = k' \exp\left(-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2\right) k \exp\left(-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2\right) = k'' \exp\left(-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2 - \frac{1}{2\sigma_0^2} (\mu - \mu_0)^2\right),$$
 (dada la distribución a priori elegida el cálculo analítico es muy sencillo).

➤ Es decir

➤ 
$$f(\mu|X) = K \exp\left(-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2 - \frac{1}{2\sigma_0^2} (\mu - \mu_0)^2\right),$$

➤ que solo depende de la muestra únicamente a través del valor de  $\bar{x}$ , así podemos sustituir  $X$  por  $\bar{x}$ , y por tanto podemos poner:

➤ 
$$f(\mu|X) = f(\mu|\bar{x}), \text{ y } \ell(\mu|X) = \ell(\mu|\bar{x}) = f(\bar{x}|\mu).$$

# Estimación bayesiana: estimación de la media en poblaciones normales con varianza conocida

- Su exponente puede escribirse de la siguiente forma (ejercicio 9.6):

$$\begin{aligned} & \frac{n}{\sigma^2} (\bar{x} - \mu)^2 + \frac{1}{\sigma_0^2} (\mu - \mu_0)^2 \\ &= \left( \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \right) (\mu - \mu_p)^2 + \left( \frac{n}{n\sigma_0^2 + \sigma^2} \right) (\bar{x} - \mu_0)^2 \end{aligned}$$

- Con  $\mu_p = \frac{\frac{n}{\sigma^2}\bar{x} + \frac{1}{\sigma_0^2}\mu_0}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}$

- Así al final tenemos que la expresión para el producto de la verosimilitud ( $\ell(\mu|\bar{x}) = f(\bar{x}|\mu)$ ) por la a priori viene determinada por:  $f(\bar{x}|\mu)f(\mu) = \exp \left( \left( \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \right) (\mu - \mu_p)^2 + \left( \frac{n}{n\sigma_0^2 + \sigma^2} \right) (\bar{x} - \mu_0)^2 \right)$ .

# Estimación bayesiana: estimación de la media en poblaciones normales con varianza conocida

- Por otro lado sabemos por el Teorema de Bayes que  $f(\bar{x}|\mu)f(\mu) = f(\mu|\bar{x})f(\bar{x})$
- Así este segundo miembro está compuesto por la distribución a posteriori ( $f(\mu|\bar{x})$ ) y la distribución **predictiva** ( $f(\bar{x})$ ), que no depende de  $\mu$ , es decir:

- $f(\mu|\bar{x}) = \exp\left(-\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)(\mu - \mu_p)^2\right)$  y  $f(\bar{x}) = \exp\left(-\left(\frac{n}{n\sigma_0^2 + \sigma^2}\right)(\bar{x} - \mu_0)^2\right)$

- Supongamos que definimos por  $p_{\bar{x}}$  y  $p_0$  a las precisiones muestral y a priori, entonces **la distribución a posteriori,  $f(\mu|\bar{x})$ , es una normal con parámetros:**

$$\mu_p = \frac{\frac{n}{\sigma^2}\bar{x} + \frac{1}{\sigma_0^2}\mu_0}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}} = \frac{p_{\bar{x}}\bar{x} + p_0\mu_0}{p_{\bar{x}} + p_0} \text{ y } \frac{1}{\sigma_p^2} = p_p = p_{\bar{x}} + p_0$$

- **Resumiendo la media de la a posteriori es una combinación lineal de la a priori y la muestral, con pesos que dependen de la precisión relativa. La precisión final es la suma de la inicial ( $p_0$ ) y la verosimilitud ( $p_{\bar{x}}$ ). (ver ejemplo 9.2).**



# Bibliografía y lecturas relacionadas:

- Fundamentos de estadística. Daniel Peña Sánchez Ribera. Alianza Editorial, 2001 o 2008.
- Data analysis Statistical and computational methods for scientists and engineers Recurso electrónico 4th.ed. Autor Brandt, Siegmund.
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Springer Second Edition February 2009.
- Probabilidad y Estadística con Python.
- Entre percentiles, cuartiles y cuantiles. Posted on 21 febrero, 2015 by Jesús Soto.
- Cuartiles.
- *NIST/SEMATECH e-Handbook of Statistical Methods*,  
<http://www.itl.nist.gov/div898/handbook/>, 09/22/2015.
- <http://www.itl.nist.gov/div898/handbook/eda/section3/normprpl.htm>
- Workshop on Big Data and Statistics UDC, junio 2015
- <http://bigdata.sgapeio.es/index.php/en/sesion-i>