Trabajo Práctico 1

Métodos Numéricos

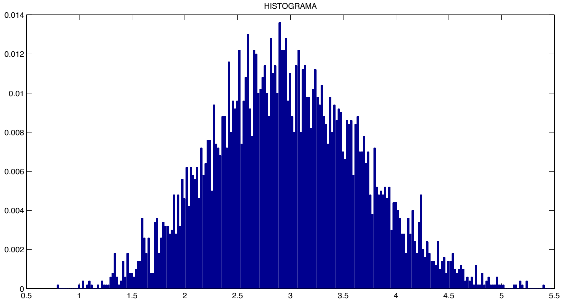
Departmento de Computación
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

22 de mayo de 2013

Estimando parámetros de distribuciones

- Sea $x_1, x_2, \ldots, x_n$ una secuencia de n datos, con $x_i \in \mathbb{R}$
- Provienen de una distribución de la cual no sabemos los parámetros

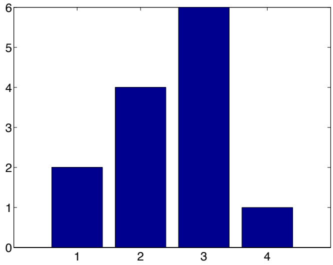
Histograma:

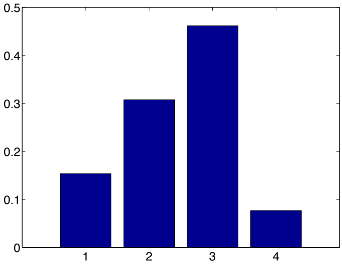


Estimando parámetros de distribuciones

Histograma

- `datos = [1 2 3 2 3 2 1 2 3 4 3 3 3]`





Estimando parámetros de distribuciones

Distribución Γ Generalizada

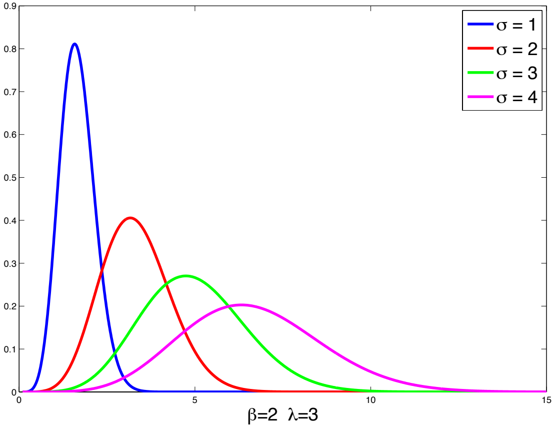
$$f_\Theta(x) = \frac{\beta\, x^{\beta\lambda-1}}{\sigma^{\beta\lambda}\,\Gamma(\lambda)} \exp\left\{-\left(\frac{x}{\sigma}\right)^{\beta}\right\} \qquad \text{con } x \in \mathbb{R}_{>0}$$

donde

- $\Gamma(\cdot)$ función Gamma definida como $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$
- Θ representa a la tupla de parámetros $\Theta = (\sigma, \beta, \lambda)$

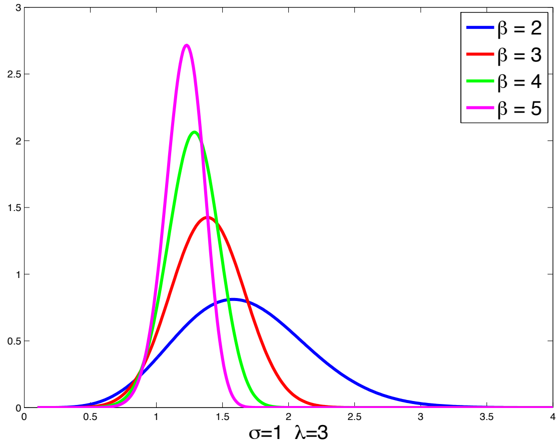
Estimando parámetros de distribuciones

Distribución Γ Generalizada: $f_{\Theta}(x) = \frac{\beta\, x^{\beta\lambda-1}}{\sigma^{\beta\lambda}\,\Gamma(\lambda)} \exp\left\{-\left(\frac{x}{\sigma}\right)^{\beta}\right\}$



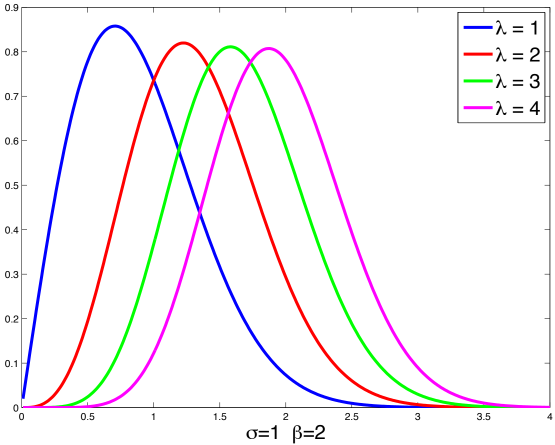
Estimando parámetros de distribuciones

Distribución Γ Generalizada: $f_\Theta(x) = \frac{\beta\, x^{\beta\lambda-1}}{\sigma^{\beta\lambda}\,\Gamma(\lambda)} \exp\left\{-\left(\frac{x}{\sigma}\right)^\beta\right\}$



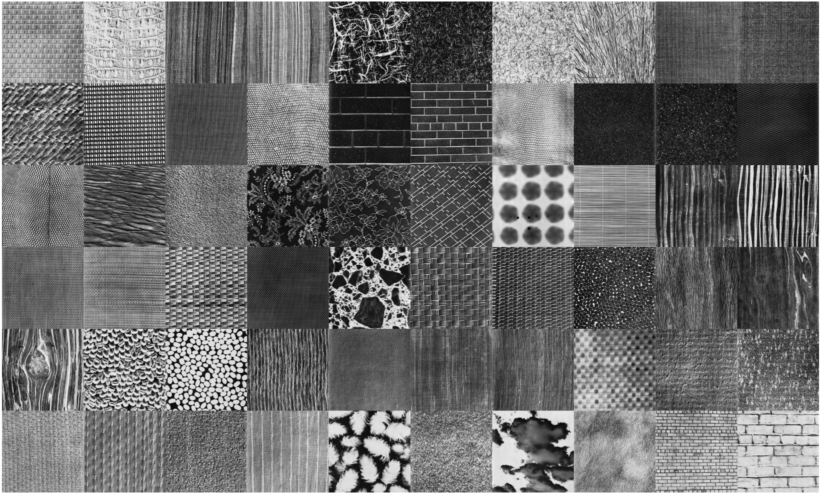
Estimando parámetros de distribuciones

Distribución Γ Generalizada: $f_\Theta(x) = \frac{\beta\, x^{\beta\lambda - 1}}{\sigma^{\beta\lambda}\,\Gamma(\lambda)} \exp\left\{-\left(\frac{x}{\sigma}\right)^{\beta}\right\}$



Ejemplo de Aplicación

Recuperando imágenes de texturas en bases de datos



Ejemplo de Aplicación

Recuperando imágenes de texturas en bases de datos

Problema

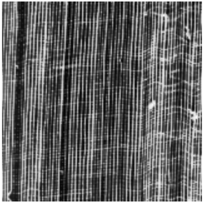
Dada una textura cualquiera, cuáles son las más parecidas en la base de datos?

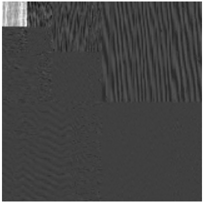
Ejemplo de Aplicación

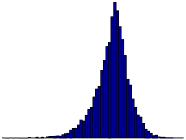
Recuperando imágenes de texturas en bases de datos

Modelado de datos

- Los datos *transformados* de las texturas tienen una distribución similar a una ΓG







Modelado y recuperación de datos

- Las texturas se representan usando sólo los parámetros de la $\Gamma\mathrm{D}$
- Se estiman distancias entre texturas a partir de las distribuciones que las modelan.
- Se logra una recuperación mayor al 95% !

Recordando un poco de *Proba*

Estimadores de máxima verosimilitud

Queremos estimar parámetros Θ de distribución f_Θ a partir de datos $x_1, \ldots, x_n$

- Función de densidad conjunta con muestras i.i.d.: $f(x_1, \ldots, x_n|\Theta) = f(x_1|\Theta) \times \ldots \times f(x_1 x_n|\Theta)$
- Construyo función de verosimilitud en función de los parámetros: $L(\Theta|x_1, \ldots, x_n) = f(x_1, \ldots, x_n|\Theta)$
- $\log(L(\Theta|x_1, \ldots, x_n)) = \sum_{i=1}^{n} \log f(x_i|\Theta)$
- Buscamos el valor de Θ que maximiza $\log(L(\Theta|x_1, \ldots, x_n))$
- En nuestro caso: $\frac{\partial L(\Theta|X)}{\partial \sigma} = \frac{\partial L(\Theta|X)}{\partial \beta} = \frac{\partial L(\Theta|X)}{\partial \lambda} = 0$

Recordando un poco de *Proba*

Estimadores de máxima verosimilitud

Para la distribución Γ Generalizada, los estimadores de máxima verosimilitud deben cumplir:

$$
\begin{aligned}
\tilde{\sigma} &= \left(\frac{\sum_{i=1}^n x_i^{\tilde{\beta}}}{n\tilde{\lambda}} \right)^{1/\tilde{\beta}} \\
\tilde{\lambda} &= \left[\tilde{\beta} \left(\frac{\sum_{i=1}^n x_i^{\tilde{\beta}} \log x_i}{\sum_{i=1}^n x_i^{\tilde{\beta}}} - \frac{\sum_{i=1}^n \log x_i}{n} \right) \right]^{-1} \\
0 &= \frac{\tilde{\beta}}{n} \sum_{i=1}^n \log x_i - \log \sum_{i=1}^n x_i^{\tilde{\beta}} + \log(n\tilde{\lambda}) - \psi(\tilde{\lambda})
\end{aligned}
$$

Recordando un poco de *Proba*

Momentos de una variable aleatoria

- Momento de orden n: $E[X^n]$
- Momentos de diferentes órdenes están relacionados con la media, varianza, skewness (asimetría), curtosis, etc.
- Estimando los dos primeros momentos de una v.a. con distribución Γ Generalizada, se derivan las sig. ecuaciones:

$$\begin{aligned} \log(\mathcal{M}(2\beta)) - 2\log(\mathcal{M}(\beta)) &= \log(1 + \beta(\mathcal{R}(\beta) - \mathcal{R}(0))) \\ \frac{\mathcal{M}(2\beta)}{\mathcal{M}^2(\beta)} &= 1 + \beta(\mathcal{R}(\beta) - \mathcal{R}(0))) \end{aligned}$$

donde $\mathcal{M}$ y $\mathcal{R}$ se calculan a partir de los datos $x_1, \dots, x_n$:
$\mathcal{M}(s) = \frac{1}{n}\sum_{i=1}^{n} x_i^s, \quad \widehat{\mathcal{M}}(s) = \frac{1}{n}\sum_{i=1}^{n} x_i^s \log(x_i),$
$\mathcal{R}(s) = \frac{\widehat{\mathcal{M}}(s)}{\mathcal{M}(s)}$

Recapitulando...

Para estimar los parámetros en el TP:

1. Estimamos $\tilde{\beta}$ resolviendo cualquiera de las sig. 2 ecuaciones:

$$\begin{aligned} \log(\mathcal{M}(2\beta)) - 2\log(\mathcal{M}(\beta)) &= \log(1+\beta(\mathcal{R}(\beta)-\mathcal{R}(0))) \\ \frac{\mathcal{M}(2\beta)}{\mathcal{M}^2(\beta)} &= 1+\beta(\mathcal{R}(\beta)-\mathcal{R}(0))) \end{aligned}$$

2. Una vez estimado $\tilde{\beta}$, despejando de:

$$\begin{aligned} \tilde{\sigma} &= \left(\frac{\sum_{i=1}^n x_i^{\tilde{\beta}}}{n\tilde{\lambda}}\right)^{1/\tilde{\beta}} \\ \tilde{\lambda} &= \left[\tilde{\beta}\left(\frac{\sum_{i=1}^n x_i^{\tilde{\beta}}\log x_i}{\sum_{i=1}^n x_i^{\tilde{\beta}}} - \frac{\sum_{i=1}^n \log x_i}{n}\right)\right]^{-1} \end{aligned}$$

estimamos los restantes parámetros $\tilde{\sigma}$ y $\tilde{\lambda}$.

Finalizando...

Matlab

- Leer `matlab.pdf` en la sección recursos de la web
- Usar `leer_datos.m` para cargar los datos en Matlab
- `dibujarHistyAjuste.m` como ayuda para el ítem 4 del TP

Informe TP

- Leer `pautas.pdf` en la sección downloads de la web
- Seguir los lineamientos allí descriptos para el informe
- También: `TPCheklist.pdf`