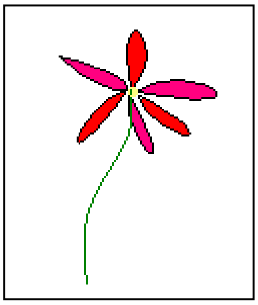
Flor-strap ó el bootstrap visto con los ojos de un niño ♦

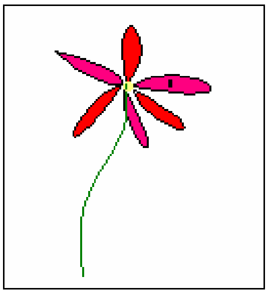
Andrés M. Alonso



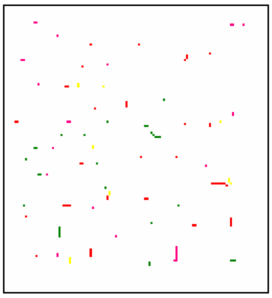
Hola, os presento a una flor especial:

Es especial por sus seis pétalos. Si sois aficionados a la botánica sabréis que no existen flores naturales de seis pétalos. Además, es especial por su nombre *flor-strap*. Tiene múltiples usos: como cura del aburrimiento; como ejemplo de algo “sobrenatural”, y su último uso conocido es como ayuda para introducir las distintas técnicas de remuestreo o bootstrap en datos dependientes.

Como sabréis, primero Tukey (1958) con el *jackknife* y luego Efron (1979) con el *bootstrap* nos enseñaron a estimar varianzas y distribuciones de estadísticos. Estos métodos fueron diseñados para datos independientes, y si los aplicamos a datos como nuestra *flor-strap* poco o nada aprenderemos, como en las figuras siguientes:



(a)



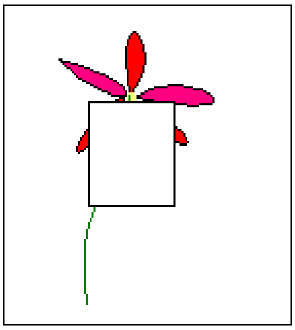
(b)

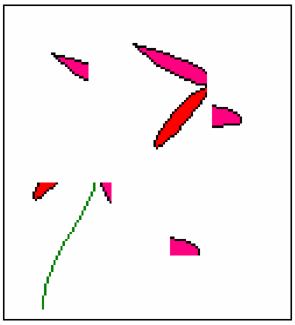
En la Figura (a) aplicamos el *jackknife* de Tukey, hemos omitido un solo punto de la flor. No parece muy distinta de la *flor-strap* original y por tanto no nos permite descubrir la variabilidad en el dibujo. En la Figura (b) aplicamos el *bootstrap* de Efron, aquí toda estructura se rompe y de un gráfico clásico pasamos de golpe a la pintura abstracta. En (b) solo hay una flor porque el pintor nos lo dice en el marco del cuadro.

De manera que, debemos modificar los métodos de remuestreo para que sean aplicables a nuestra *flor-strap* en la que el verde sigue al verde y éste se une con el centro amarillo del cual salen pétalos rojos y rosas, es decir la disposición de los colores tiene estructura de dependencia.

♦ Este trabajo surge de una visión no matemática de mi tesis doctoral titulada “*Métodos de remuestreo y datos omitidos en series temporales*” que fue dirigida por los profesores Daniel Peña y Juan Romo del Departamento de Estadística y Econometría de la Universidad Carlos III de Madrid.

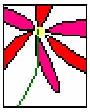
Künsch (1989) dio con una solución: el jackknife por bloques (c) y el bootstrap por bloques (d):


(c)


(d)

Si miramos muchos gráficos como (c) podemos aprender de la variabilidad del gráfico ya que lo que falta en la figura (c) contiene mucha información de la *flor-strap* original. En (d) reconocemos el cubismo, solo con pedazos descubrimos la flor, y muchos (d) son arte y además nos muestran la parte esencial de la *flor-strap*.

Otra solución a este problema es el *subsampling* de Politis y Romano (1994) que ilustramos en (e):


(e)

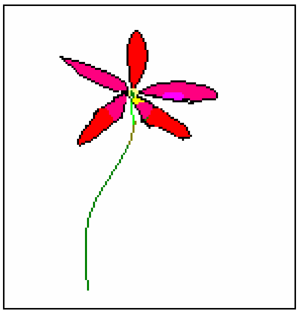
No desprecies a (e) por su tamaño, en modo alguno es una pintura menor. Todo lo contrario, en las miniaturas a veces encontramos más información que en los grandes murales. Mirar a todos los (e) posibles es la manera de entender la pintura de Politis y Romano. Ah! este tipo de métodos es válido para cualquier tipo de flor, mientras que el bootstrap por bloques de Künsch necesita flores de trazos suaves. Ya se sabe, una versión cubista de una pintura cubista converge casi seguro a un cuadro abstracto.

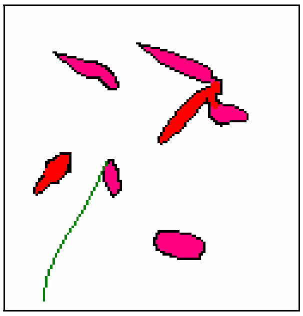
Por último, mostramos el impresionismo del *wild bootstrap* de Franke y Härdle (1993):

Y en este punto, en la tesis, nos preguntamos qué podemos añadir a tales soluciones:

- Restauración del cuadro (c), pues el gran público prefiere los cuadros intactos para tener una visión completa.
- Suavizar el cubismo de (d) completando los pétalos y uniéndolos cuando sea posible.

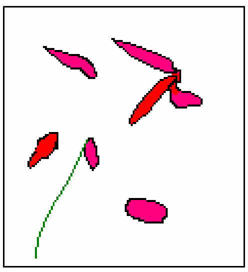
En el Capítulo 1 mostramos la artesanía de la interpolación de los valores omitidos, y que les ilustramos en (g) y (h).


(g)

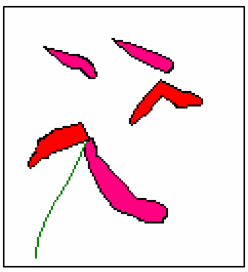

(h)

De más está decir que la interpolación en (g) y (h) no es perfecta, tenemos en cuenta la estructura de la *flor-strap*, pero solo podemos aproximarnos al modelo de dimensión infinita que la genera mediante una sucesión de modelos de flores-*sieve*: flores de 1, 2, 3, 5, 8, … pétalos (filotaxia regular o serie de Fibonacci) ó flores de 3, 4, 7, 11, 18, … pétalos (filotaxia irregular o serie de Lucas); de 1, 2, 3, 4, … colores; de color continuo o moteado, y un largo etcétera de parámetros. Esto nos lleva al Capítulo 2, donde estudiamos la predicción de una característica (color o forma, en nuestro ejemplo) dados los trazos anteriores, y luego generalizamos este problema de predicción al caso de interpolación (restauración) en el que tendremos en cuenta para predecir no solo los trazos anteriores sino las características de los objetos cercanos.

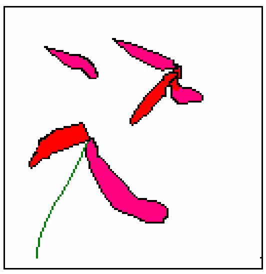
En los resultados del tema anterior, comprendemos que la reconstrucción está condicionada por el modelo seleccionado, y en el Capítulo 3, damos una manera de introducir la incertidumbre sobre el modelo generador de la *flor-strap*. En el gráfico siguiente ilustramos el método implementado en este capítulo: en (i1) reconstruimos (d) con un modelo de siete pétalos y los colores añadidos son una media ponderada de los colores cercanos, en (i2) con un modelo de cinco pétalos y cada pétalo del color predominante, y por último en (i3) damos la predicción que asigna probabilidades iguales a ambos modelos recuperando una característica esencial de la *flor-strap*: sus seis pétalos.


(i1)

+


(i2)

=


(i3)

En el Capítulo 4 concluimos esta tesis con un último apunte sobre la *flor-strap*: a diferencia de la margarita común cuando tire de los pétalos de esta flor conviene empezar con una predicción pesimista.

REFERENCIAS

1. Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, Vol. **7**, 1-26.

2. Härdle, W. y Mammen, E. (1993) Comparing nonparametric versus parametric regression. *The Annals of Statistics*, Vol. **21**, 1926-1947.

3. Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, Vol. **17**, 1217-1241.

4. Politis, D.N. y Romano, J. F. (1994) Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, Vol. **22**, 2031-2050.

5. Tukey, J. (1958) Bias and confidence in not quite large samples. *The Annals of Mathematical Statistics*, Vol. **29**, 614.
