

Capítulo 6

Resultados Importantes

6.1 La casi varianza muestral ante las perturbaciones

Como se ha dicho ya, la relación $E(S_{n-1}^2) = \sigma_{proc}^2$ sólo es válida para procesos estacionarios. Esto hace que deje de ser útil para estimar la varianza del proceso en casos no estacionarios, no solo en el caso de vibraciones del terreno, sino también en los instantes después de un escalón.

En el documento principal de este proyecto se hace un estudio del valor esperado de S_{n-1}^2 a lo largo del tiempo a partir del instante en que se produce un escalón en la media del proceso. El resultado muestra que:

$$E(S_{n-1}^2) = \sigma_{proc}^2 + \frac{t(N-t)}{N} \Delta\mu^2$$

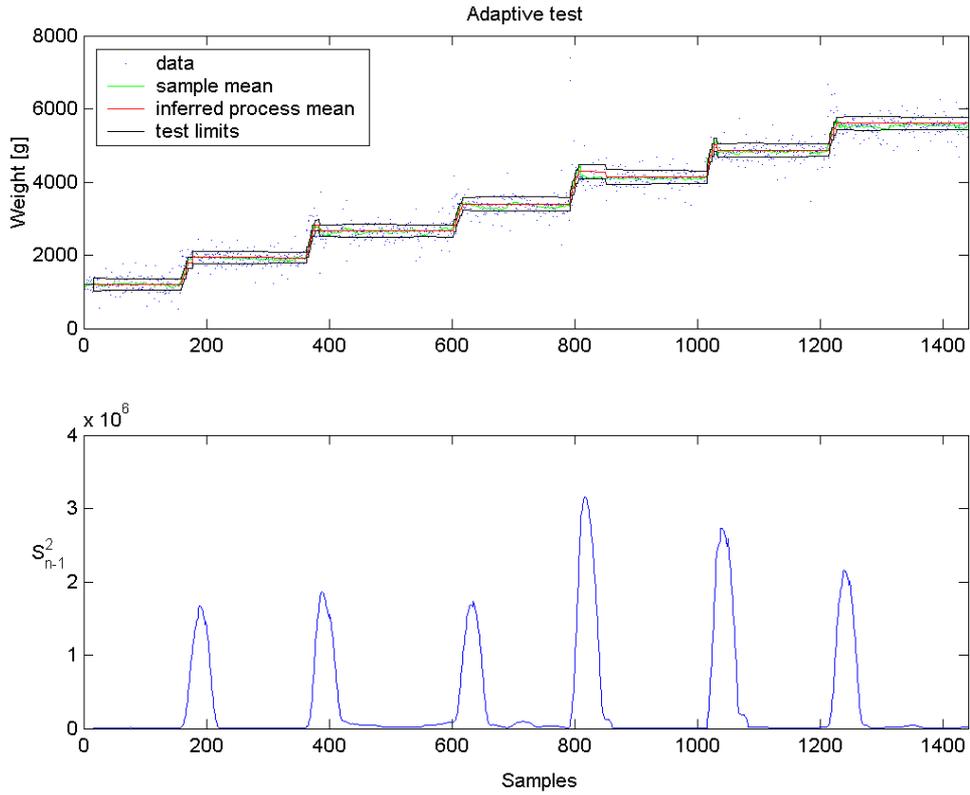
Donde $\Delta\mu$ es el incremento de la media del proceso y t es el tiempo transcurrido desde el escalón medido en muestras.

Este resultado corresponde a una parábola que comienza y termina en σ_{proc}^2 y con un valor máximo esperado de :

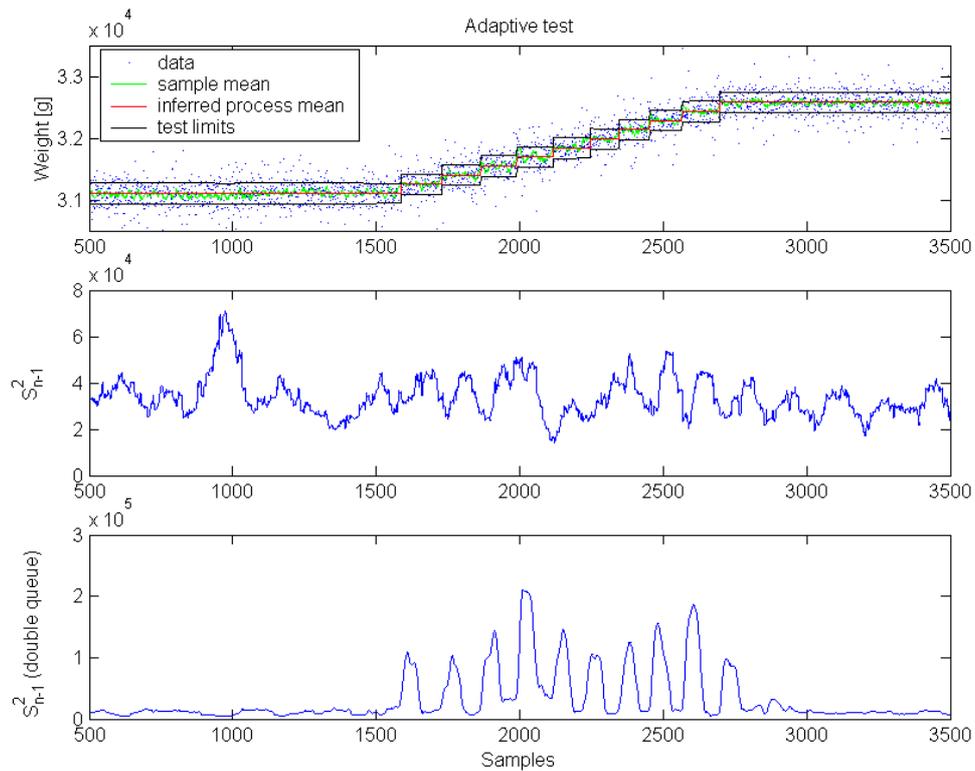
$$S_{n-1}^2 \Big|_{\max} = \frac{1}{4} \frac{N}{(N-1)} \Delta\mu^2$$

en el instante $t = \frac{N}{2}$

Sabiendo esto, el estadístico S_{n-1}^2 no es, ni mucho menos, inservible, sino que resulta un detector de “eventos” (cambios instantáneos de la media de un proceso) idóneo para saltos pequeños. Un seguimiento muestra a muestra (no cada N muestras) de la casi varianza muestral, nos asegura que el estadístico será evaluado en su valor más alto. La siguiente figura muestra con datos reales la parábola predicha:



Si bien los resultados no son siempre tan limpios debido a la incertidumbre de la estimación, el método de doble cola puede ser utilizado, al igual que se ha hecho en el test, para obtener valores más fiables del estadístico como se muestra en la siguiente gráfica:



Este estadístico puede ser idóneo para la detección de saltos de la media, lo que lo hace apropiado para un test de una báscula de cómputo, dónde no cuenta únicamente la variación del peso sino la instantaneidad del cambio.

Así mismo, es apto para la segmentación de series de datos en tramos, facilitando las estimaciones iniciales de los parámetros de un proceso estadístico, especialmente de la media del proceso.

6.2 Estimador de no-estacionareidad del proceso. El estadístico $S_{n-1}^2 - S_{\bar{x}}^2 \cdot N$

Como se ha dicho en el apartado relativo al test de doble cola, las oscilaciones afectan mucho más a la desviación típica muestral del proceso $\hat{\sigma}_{proc}$ que a la desviación típica muestral de la media muestral $\hat{\sigma}_{\bar{x}}$. Eso hace que se deje de cumplir la relación:

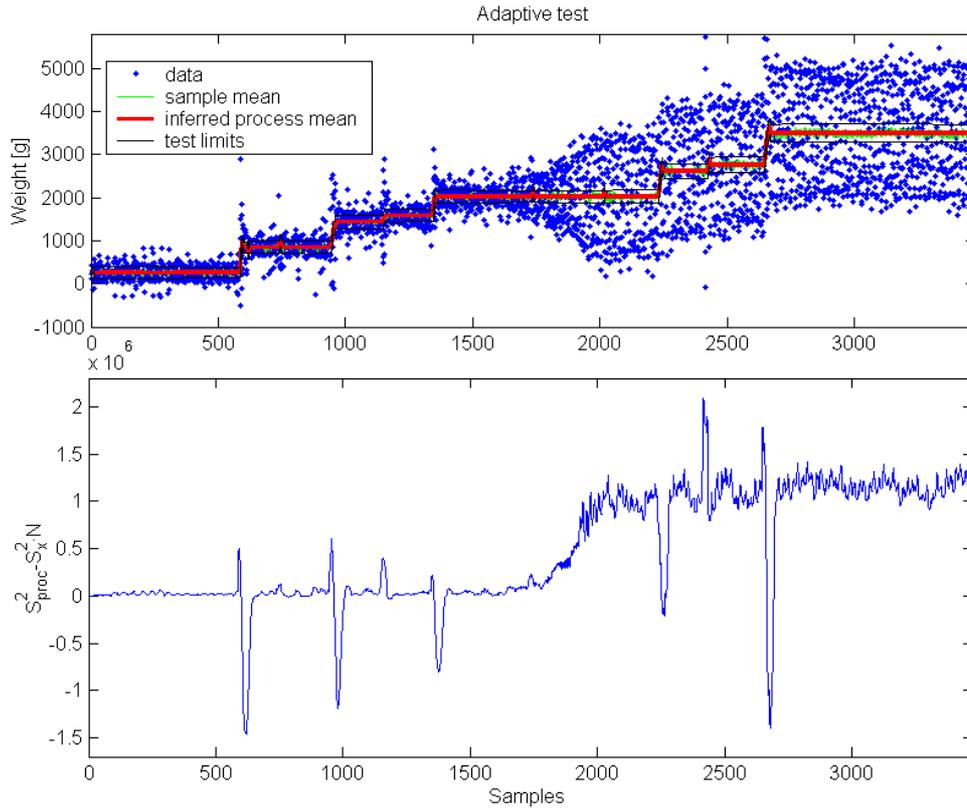
$$E(S_{\bar{x}}^2) = \frac{E(S_{n-1}^2)}{N}$$

Este hecho sugiere que el estadístico $S_{n-1}^2 - S_{\bar{x}}^2 \cdot N$, que tiene un valor esperado nulo para un proceso estacionario, sería buen estimador de no-estacionareidad del proceso (ENE). Su valor esperado en función de la frecuencia de las oscilaciones, de la amplitud de estas y del valor de N (tamaño de la muestra), se estudia en el documento principal de este proyecto.

Este estimador hará notar todo tipo de no-estacionareidades. En el caso de un escalón en la media del proceso, se tiene que el valor de S_{n-1}^2 y el de $S_{\bar{x}}^2$ son afectados prácticamente en la misma medida. Eso hace que, al ir $S_{\bar{x}}^2$ multiplicado por N, el ENE tomará valores siguiendo una parábola con un mínimo en $t=N/2$ de valor esperado:

$$E(ENE)_{\min} = -\frac{N}{4} \Delta\mu^2$$

La gráfica siguiente muestra el valor del ENE calculado para la misma serie de medidas con perturbaciones del terreno utilizada anteriormente. Se observan fácilmente: El incremento del valor medio del ENE al comenzar las oscilaciones y los picos correspondientes a los saltos del valor medio.



Se puede observar que los saltos de la media producen sólo un pequeño pico hacia arriba cuando el escalón es pequeño. Esto es debido a la forma que tiene cada uno de los dos sumandos del estimador y se explica con más detalle en el documento principal del proyecto.

Este estimador puede facilitar la selección de datos de una serie de medidas a la hora de estimar los parámetros de un proceso estadístico que se vea amenazado por perturbaciones, al indicar cuales de esos datos se pueden considerar como pertenecientes a un proceso estacionario. Esto es especialmente apropiado para el caso de la varianza, cuya estimación se ve muy afectada por las oscilaciones.

Por otra parte, puesto que los datos seleccionados aún podrían estar afectados por niveles bajos de oscilaciones, siempre se puede sacar provecho del método de doble cola, que ha demostrado mejorar la estimación de la varianza de forma eficaz.