

## Capítulo 4

### Estudio estadístico de las medidas

#### 4.1 Hipótesis y selección de datos

El estudio estadístico de las medidas se ha realizado bajo las siguientes hipótesis:

- En ausencia de vibraciones del terreno inducidas y mientras permanece constante el número de unidades sobre la báscula, la lectura de las medidas de los sensores de peso constituye un proceso estadístico estacionario. Es decir, su media y varianza permanecen constantes.
- Se acepta sin embargo que la varianza pueda tener variaciones lentas debidas a distintos factores como podría ser una relación con el peso cargado en la báscula. Esta hipótesis se fundamenta en el estudio de vibraciones realizado en el proyecto de *Mohd Syazrulazwan Mohd Zin* y se revisa más adelante en este trabajo. En intervalos de tiempo no muy largos el proceso se considera completamente estacionario y homocedástico.
- Tras cada evento, el proceso pasa por un breve transitorio y vuelve enseguida a ser estacionario entorno a su nueva media y conservando su varianza.

El estudio estadístico de las medidas comienza con el estudio de la varianza del proceso. Para dicho estudio se ha tomado un total de 14706 medidas en el laboratorio (el equivalente a 9 min. y 4 seg. de muestras) en días distintos y con distintos niveles de carga de la báscula. De ellos se han seleccionado 29 secciones de 50 muestras cada una correspondientes a estados estacionarios.

Con esas muestras se han calculado los siguientes valores de la desviación estándar  $\sigma$  :

Valor medio de  $\sigma$  :            194.3683 g  
Desviación típica de  $\sigma$  :    36.9144 g

Por lo que un intervalo de confianza del 95% para el valor de sigma será:

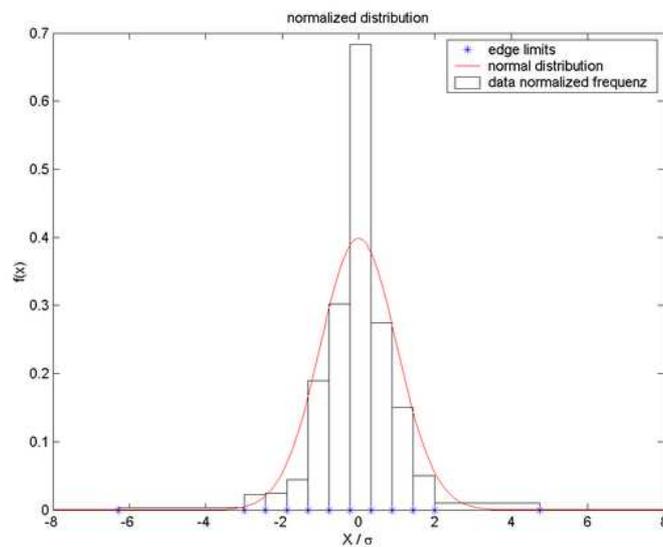
$$I_{95} = \mu_{\sigma} \pm 1.96 \cdot \sigma_{\sigma} = 194.37 \pm 72.35 = [122.02 \dots 266.721].$$

Dicho intervalo se extiende un  $\pm 37\%$  sobre el valor medio, lo que supone un rango demasiado grande que dificulta la elección de parámetros para el test. Esta falta de homogeneidad de las medidas está respaldada por la segunda hipótesis y está causada por los cambios de varianza a largo plazo. Sin embargo, las estimaciones de la varianza en intervalos de tiempo próximos dan resultados mucho más precisos. Esta diferencia ha dado lugar, como se verá más adelante, a una variación del tipo de test a emplear.

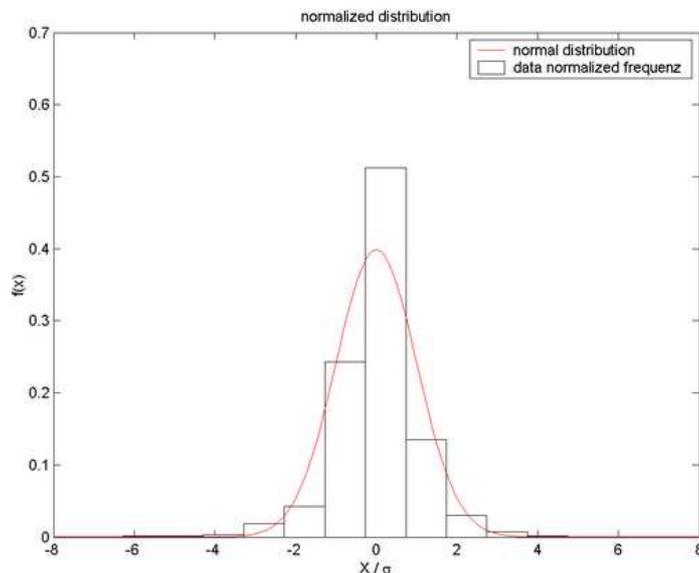
## 4.2 Comparación con la distribución normal

La facilidad en el cálculo de los parámetros necesarios para el test depende de que podamos aproximar la función de distribución por alguna de expresión conocida. En esta sección se realiza la comparación de la distribución real de medidas con una distribución normal. Para ello se ha utilizado el test  $\chi^2$  de bondad de ajuste.

En primer lugar se ha tomado 1479 datos de los utilizados para el estudio de la varianza y se ha segmentado en veinte segmentos equiespaciados a lo largo de su recorrido. La ausencia de datos en algunos de los segmentos en los extremos lleva a la necesidad de unir varios segmentos quedando finalmente 11 segmentos disponibles (Ver figura).



La prueba se ha repetido segmentando en once grupos equiespaciados, como muestra la figura siguiente.



En ambos casos se ha normalizado los datos restando la media y dividiendo entre la desviación típica. El resultado del test ha sido el siguiente:

Degrees of freedom	$\chi^2$	alpha = 0.05	alpha = 0.001	H <sub>0</sub> ?
10	350,42	18,307	29,588	Rejected
10 (equal spaced segments)	45787	18,307	29,588	Rejected

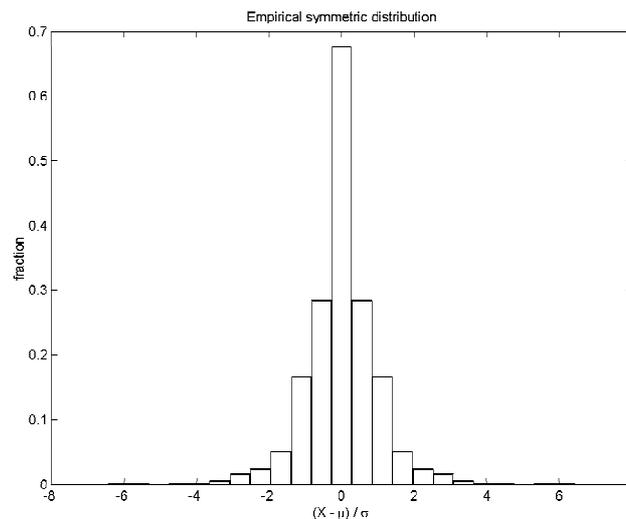
En ambos casos el valor de la función de error  $\chi^2$  es mucho mayor que el permitido de cara a aceptar la hipótesis de distribución normal y como consecuencia se rechaza la hipótesis.

Llegado este punto se decide buscar una aproximación empírica de la función de densidad de probabilidad que sirva de base para la estimación de probabilidades y el cálculo de los parámetros.

### 4.3 Ajuste empírico de la función de densidad

El ajuste de la función también comienza por una segmentación de los datos. En este caso, en lugar de dividir el recorrido de los datos en secciones iguales, se ha elegido segmentar los datos sobre un recorrido centrado en el cero y con un número impar de segmentos (19 segmentos) consiguiendo así que uno de ellos esté centrado en el origen.

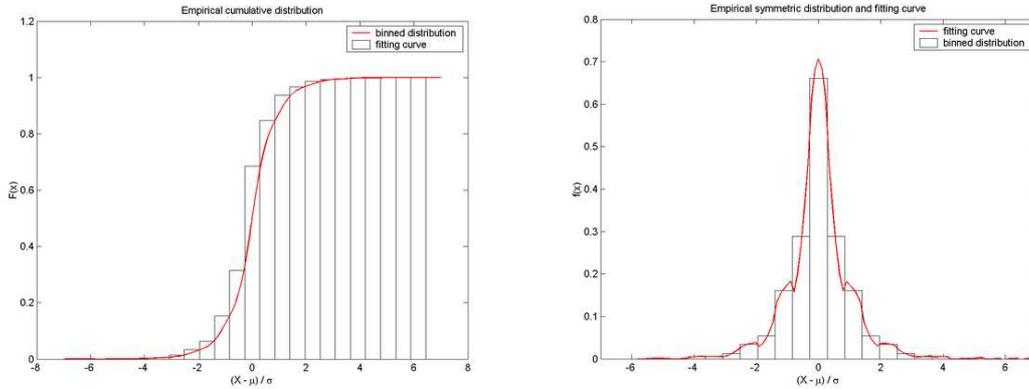
Tras realizar el histograma con esta segmentación se descubre que la distribución es asombrosamente simétrica, por lo que se asume la simetría de la distribución real mientras que las pequeñas asimetrías de la distribución empírica se consideran un fenómeno puramente estocástico.



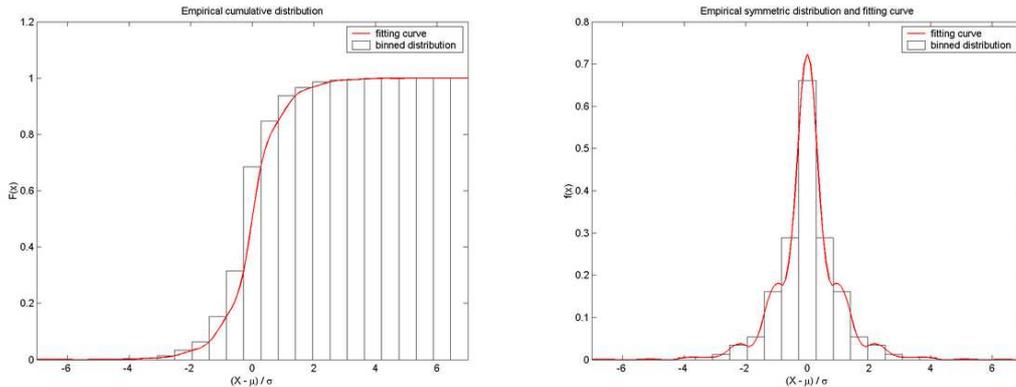
Se comparan dos métodos de ajuste distintos: Un filtro paso-baja y un ajuste mediante líneas spline. Ambos se han sido efectuados sobre los datos de las frecuencias acumuladas. Posteriormente se derivan sólo para visualizar la forma de la función de densidad.

Las siguientes figuras muestran los resultados:

### Low-Pass Filter



### Spline



Los resultados son parecidos, aunque los obtenidos con las curvas spline presentan algunas ventajas, entre ellas que se ha podido imponer una pendiente nula en el “infinito” (Se ha usado  $X=10$  y  $X=-10$  para imponer estas condiciones). Por otra parte la interpolación usando splines es más sencilla y rápida una vez calculados los coeficientes.

Si se comparan los resultados de los dos métodos se comprueba que la diferencia no sobrepasa el 0.25%, y para los valores altos en que será evaluada ( $X > 2$ ) la diferencia no supera el 0.03%.

Por tanto se decide utilizar el método “spline” de cara a facilitar el cálculo.