CAPITULO III

CARACTERIZACION DE FORMAS

ESTIMACION DE PARAMETROS

PROPAGACION DE ERRORES

Caracterización de la forma de una PDF (I)

Estamos suponiendo que nuestros objetos de estudio de procesos aleatorios, que se manifiestan bajo dorma de realizaciones aleatorias a partir de una PDF subyacente

- En la práctica, la verdadera dependencia funcional de una PDF es a menudo desconocida
- La información sobre su forma solamente puede extraerse a partir de una muestra de talla finita (digamos que contiene N eventos), es decir suponemos que la muestra disponible es una realización aleatoria a partir de una PDF desconocida.
- Si consideramos que esa PDF subyacente is de tipo paramétrico, la *caracterización de su forma* es un procedimiento para estimar los valores numéricos de sus parámetros, partiendo de una hipótesis "razonable" sobre la dependencia funcional sobre sus variables.
- Ahora, solamente un número finito de valores de expectación independientes pueden extraerse de una muestra de talla finita.
- No existe una receta única para la elección de los parámetros a ser estimados, con lo que el proceso es intrínsecamente incompleto.
- Se puede sin embargo confirmar en la práctica que el proceso de caracterización de forma es bastante poderoso, si los parámetros seleccionados proveen información útil y complementaria.

Ilustramos estas consideraciones con un ejemplo unidimensional para una variable aleatoria única x.

Caracterización de la forma de una PDF (II)

Ilustramos las consideraciones anteriores con un ejemplo unidimensional para una variable aleatoria única x. Consideremos el promedio empírico $\overline{x}$, definido como

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \ .$$

Mostraremos más adelante que $\overline{x}$ es un buen *estimador* de la media μ de la PDF subyacente $P(x)$.
De modo análogo, la media cuadrática RMS (del inglés "root-mean-squared"), definida como

$$\mathrm{RMS}^2 = \overline{x^2} - (\overline{x})^2 \ , \mathrm{con}\ \overline{x^2} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 \ ,$$

es un estimador razonable de la varianza σ^2 (veremos más adelante una mejor definición).
En términos intuitivos, el promedio y el RMS reflejan información útil y complementaria sobre la "localización" y la "dispersión" de la región de x mayor densidad de eventos, y esta región debe corresponder de manera aproximada a los intervalos de x donde la PDF tiene valores más grandes.
Obviamente, en un caso general esos dos parámetros son insuficientes para caracterizar una PDF más genérica, que require un procedimiento más sistemático.

Caracterización de la forma de una PDF (III)

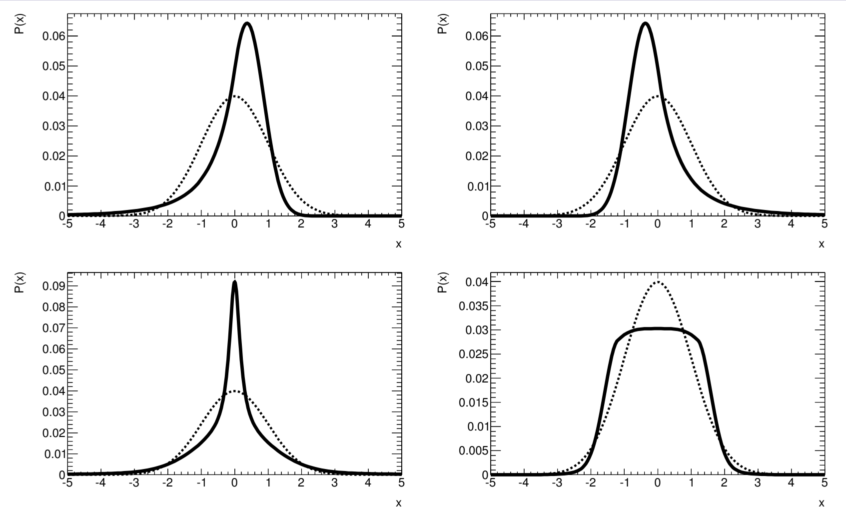
Para un procedimiento más sistemático, transformamos la x-dependencia de la PDF $P(x)$ en una k-dependencia de la *función característica* $C[k]$, definida como

$$C[k] \;=\; E\left[e^{ik\,\frac{x-\mu}{\sigma}}\right] \;=\; \sum_j \frac{(ik)^j}{j!}\mu_j \;.$$

Como se puede notar, la función característica es la transformación de Fourier de la PDF. Los coeficientes μ_j de la expansión are se llaman *momentos reducidos*; por construcción, los primeros momentos son $\mu_1 = 0$ y $\mu_2 = 1$; en términos de la variable reescalada $x' = (x-\mu)/\sigma$, la PDF fue desplazada para tener promedio nulo, y escalada para tener varianza unitaria.
En principio, mientras mayor el número de momentos μ_j sean estimados, más detaillada será la caracterización de la forma de la PDF (pero una muestra finita solamente permite medir un número finito de momentos).
Los momentos 3 y 4 tienen nombres específicos, y sus valores son sencillos de interpretar en términos de la forma:

- el tercer momento es llamado oblicuidad (o *skewness* en inglés)
  - una distribución simétrica tiene skewness nula,
  - un valor negativo (positivo) indica una "anchura" mayor a la izquierda (derecha) de su media.
- el cuarto momento es llamado *kurtosis*
  - cantidad definida positiva, relacionada con cuán "picante" es la distribución
  - un valor grande indica un pico estrecho y "colas" de largo alcance: es una distribución leptokúrtica
  - un valor pequeño indica un pico central ancho y colas poco prominentes: es una distribución platykúrtica

Ejemplos: skewness y kurtosis



Estimación de parámetros (I)

La caracterización de la forma de una PDF a través de una estimación sequencial de parámetros de forma nos permitió introducir de manera cualitativa al concepto de estimación de parámetros (también llamado en inglés "point estimation"). Una receta más general sería la siguiente :
Consideremos una PDF n-dimensional, k-paramétrica,

$$P(\ x_1, x_2, \ldots, x_n\ ;\ \theta_1, \theta_2, \ldots, \theta_k\ )\ ,$$

para la cual queremos estimar los valores $\theta_1, \ldots, \theta_k$ a partir de una muestra de talla finita, utilizando un conjunto de estimadores $\hat{\theta}_1, \ldots, \hat{\theta}_k$.
Esos estimadores son también variables aleatorias, con sus propias medias y varianzas: sus valores diferirían al ser estimados sobre otras muestras.
Esos estimadores deben satisfacer dos propiedades clave:

- *ser consistente* : asegura que, en el límite de una muestra de talla infinita, el estimador converge al verdadero valor del parámetro;
- *ser no sesgado* : la ausencia de sesgo asegura que el valor de expectación del estimador es el verdadero valor del parámetro, para toda talla de la muestra.

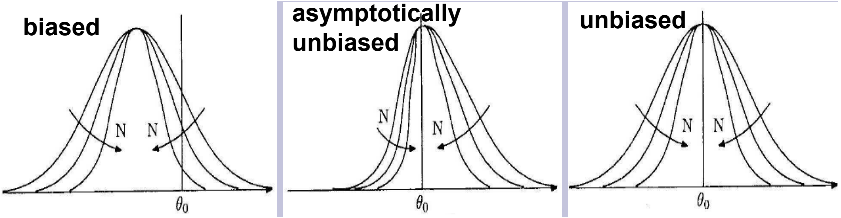
Un estimador sesgado pero consistente (también llamado asintóticamente no-sesgado) es tal que el sesgo disminuye al aumentar la talla de la muestra.

Estimación de parámetros (I bis)

Otros criterios son útiles para caracterizar la calidad de los estimadores; por ejemplo

- eficiencia: un estimador de pequeña varianza es más eficiente que uno de mayor varianza ;
- robusteza: este criterio describe la "sensibilidad" del estimador a incertidumbres en la forma de la PDF. Por ejemplo, el promedio es robusto contra incertidumbres sobre los momentos de orden par, pero es menos robusto contra incertidumbres en los momentos de orden impar.

Nota : estos criterios son en ocasiones mutuamente contradictorios; por razones prácticas, puede ser preferible tener un estimador eficiente pero sesgado, a uno no sesgado pero de pobre convergencia.



El promedio empírico $\overline{x}$ es un estimador convergente, no sesgado de la media μ de la PDF subyacente: $\hat{\mu} = \overline{x}$. Esto se demuestra facilmente, evaluando el valor de expectación y la varianza de $\overline{x}$:

$$\begin{aligned} E\left[\overline{x}\right] &= \frac{1}{N}\sum_{i=1}^{N} E\left[x\right] = \mu \ , \\ V\left[\overline{x}\right] &= E\left[(\overline{x}-\mu)^2\right] = \frac{\sigma^2}{N} \ . \end{aligned}$$

Al contrario, el RMS empírico de una muestra es un estimador sesgado (aunque asintóticamente no-sesgado) de la varianza σ^2. Esto se demuestra facilmente también, reescribiendo su cuadrado en terminos de la media:

$$\mathrm{RMS}^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})^2 = \left[\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2\right] - (\overline{x}-\mu)^2 \ ,$$

de manera que su valor de expectación es

$$E\left[\mathrm{RMS}^2\right] = \sigma^2 - V\left[\overline{x}\right] = \frac{N-1}{N}\sigma^2 \ ,$$

que si bien converge a la verdadera varianza σ^2 en el límite $N \to \infty$, subestima sistemáticamente su valor para muestras de talla finita.

Estimación de parámetros (III)

Si bien converge a la verdadera varianza σ^2 en el límite $N \to \infty$, el RMS subestima sistemáticamente su valor para muestras de talla finita. Pero es inmediato definir un estimador modificado

$$\frac{N}{N-1}\mathrm{RMS}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2 \ ,$$

que es, para muestras de talla finita, un estimador no sesgado de la varianza.
En resumen: para una PDF desconocida, tenemos estimadores consistentes y no sesgados de su media μ y su varianza σ^2, que pueden ser extraídos de muestras de talla finita:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i \ ,$$

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \hat{\mu})^2 \ .$$

Nota: los factores $1/N$ y $1/(N-1)$ para $\hat{\mu}$ y $\hat{\sigma}^2$ se entienden intuitivamente: el promedio empírico puede medirse incluso en la muestra más pequeña posible de un solo evento, mientras que al menos dos eventos son necesarios para estimar la dispersión empírica de una muestra.
Ejercicio : completar los detalles relacionando la media y la varianza con sus estimadores empíricos.

El ejemplo clásico anterior trataba de una única variable aleatoria.
En presencia de múltiples variables aleatorias $\vec{x} = \{x_1, \ldots, x_n\}$, la generalización del resultado anterior lleva a definir la covarianza empírica, cuyos elementos $\hat{C}_{ab}$ son estimados en una muestra de N eventos de la manera siguiente:

$$\hat{C}_{ab} \; = \; \frac{1}{N} \sum_{i=1}^{N} \left(x_{a,i} - \hat{\mu_a} \right) \left(x_{b,i} - \hat{\mu_b} \right) \; .$$

(los índices a, b recorren la lista de variables aleatorias, $1 \leq a, b \leq n$). (Nota: para N pequeños, se debe corregir un sesgo en este estimador, c.f. el ejemplo del RMS)
Suponiendo que la verdadera covarianza es conocida, la varianza de una función arbitraria $f(\vec{x})$ de las variables aleatorias se evalúa a partir de la expansión de Taylor alrededor de las medias de sus parámetros $\hat{\vec{\mu}}$ según

$$f\left(\vec{x}\right) \; = \; f\left(\hat{\vec{\mu}}\right) + \sum_{a=1}^{n} \left. \frac{df}{dx_a} \right|_{\vec{x}=\hat{\vec{\mu}}} \left(x_a - \hat{\mu}_a \right) \; ,$$

en otras palabras, $E\left[f(\vec{x})\right] \; \simeq \; f(\hat{\vec{\mu}})$.

De manera similar,

$$E\left[f^2(\vec{x})\right] \;\simeq\; f^2(\hat{\vec{\mu}}) + \sum_{a,b=1}^{n} \frac{df}{dx_a}\frac{df}{dx_b}\bigg|_{\vec{x}=\hat{\vec{\mu}}} \hat{C}_{ab}\ ,$$

con lo que la varianza de f se estima como

$$\hat{\sigma}_f^2 \;\simeq\; \sum_{a,b=1}^{n} \frac{df}{dx_a}\frac{df}{dx_b}\bigg|_{\vec{x}=\hat{\vec{\mu}}} \hat{C}_{ab}\ .$$

Esta expresión, llamada *fórmula de propagación de incertidumbres*, estima la varianza de una función genérica $f(\vec{x})$ a partir de los estimadores de sus medias y covarianzas.
Ejemplos particulares de propagación de incertidumbres:

- cuando todas las variables aleatorias $\{x_a\}$ son no-correlacionadas, la matriz de covarianza es diagonal, $C_{ab} = \sigma_a^2\delta_{ab}$ y la covariance de $f(\vec{x})$ se reduce a

$$\hat{\sigma}_f^2 \;\simeq\; \sum_{a=1}^{n} \left(\frac{df}{dx_a}\right)^2\bigg|_{\vec{x}=\hat{\vec{\mu}}} \hat{\sigma}_a^2\ .$$

Propagación de incertidumbres (III)

- para la suma de dos variables aleatorias $S = x_1 + x_2$, la varianza es

$$\sigma_S^2 \;=\; \sigma_1^2 + \sigma_2^2 + 2C_{12} \;=\; \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\rho_{12},$$

- la generalización a más de dos variables es:

$$\sigma_S^2 \;=\; \sum_{a,b} \sigma_a\sigma_b\rho_{ab} \;.$$

En ausencia de correlaciones, se dice que los errores absolutos se suman "en cuadratura":

$$\sigma_S \;=\; \sigma_1 \oplus \sigma_2 \;=\; \sqrt{\sigma_1^2 + \sigma_2^2} \;,$$

y si la correlación vale 1, el error en la suma es la suma de los errores.

- para el producto de dos variables aleatorias $P = x_1 x_2$, la varianza relativa es

$$\left(\frac{\sigma_P}{P}\right)^2 = \left(\frac{\sigma_1}{x_1}\right)^2 + \left(\frac{\sigma_2}{x_2}\right)^2 + 2\frac{\sigma_1}{x_1}\frac{\sigma_2}{x_2}\rho_{12} \ ,$$

- la generalización a más de dos variables:

$$\left(\frac{\sigma_P}{P}\right)^2 = \sum_{a,b} \frac{\sigma_a}{x_a}\frac{\sigma_b}{x_b}\rho_{ab} \ .$$

En ausencia de correlaciones, se dice que los errores relativos se suman en cuadratura:

$$\sigma_P/P = \sigma_1/x_1 \oplus \sigma_2/x_2 \ ,$$

y si la correlación vale 1, el error relativo sobre el producto es la suma de los errores relativos.

Propagación de incertidumbres (V)

- para una función genérica en ley de potencia, $Z = x_1^{n_1} x_2^{n_2} \ldots$, si todas las variables son no-correlacionadas, la varianza es

$$\frac{\sigma_Z}{Z} = n_1 \frac{\sigma_1}{x_1} \oplus n_2 \frac{\sigma_2}{x_2} \oplus \ldots$$

En otras palabras: en la suma en cuadratura, los errores relativos son ponderados por los expomentes. Por tanto los términos con exponentes importantes (cuadrados, cubos, etc...) dominarán el *presupuesto de error*, mientras que términos con raíces (cuadradas, cúbicas, etc...) tendrán un impacto subdominante.