

# (#438). TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIIe)

[MONOTEMA] Avanzamos con el quinto apartado del tercer capítulo de Probability Theory and Statistical Inference, de Aris Spanos.

## Parámetros y momentos

Además del histograma de la distribución de datos observados, también disponemos de ciertos números que caracterizan la distribución como la media o la varianza. Esos valores numéricos están relacionados con los momentos de la distribución, que son esperanzas matemáticas de ciertas funciones de la variable aleatoria  $X$ , genéricamente denotadas por  $h(X)$ .

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f_x(x; \theta) dx = g(\theta)$$

Si escogemos diferentes funciones  $g(\theta)$  obtendremos diferentes momentos de la distribución. Por ejemplo:

**Media**  $\mu$

$$h(X) = X, x \in \mathbb{R}_X$$

Para variables continuas:

$$\mu = E[h(X)] = E(X) = \int_{-\infty}^{\infty} x f_x(x; \theta) dx$$

Para variables discretas:

$$\mu = E(X) = \sum_{x_i \in \mathbb{R}_X} x_i f_x(x_i; \theta)$$

**Varianza**  $\sigma^2$

$$h(X) = [X - E(X)]^2, x \in \mathbb{R}_X$$

$$\sigma^2 = h(X) = [X - E(X)]^2 = \int_{-\infty}^{\infty} [x - \mu]^2 f_x(x; \theta) dx$$

Una forma conveniente de calcular los momentos de una distribución es a través de la función generatriz de momentos (mgf), donde  $h(X) = e^{tX}$

$$m_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tX} f(x) dx, t \in (-h, h), h > 0$$

Para variables aleatorias discretas las integrales se vuelven sumatorios.

Por ejemplo, para una variable aleatoria X que sigue una distribución de Poisson:

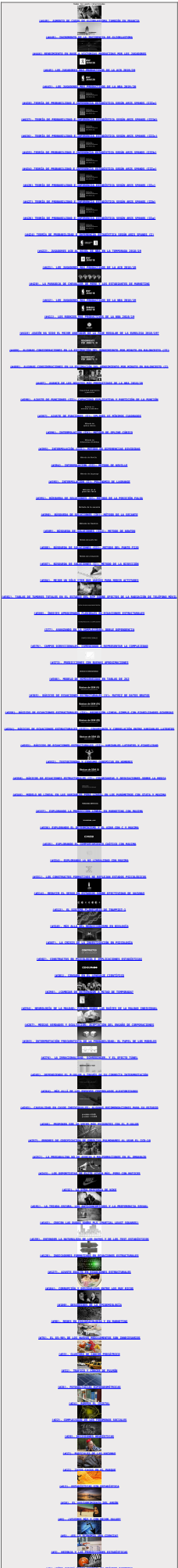
$$m_X(t) = E(e^{tX}) = \sum_{r=0}^{\infty} e^{tX} f(x) dx$$

Dado que:  $f_x = \frac{e^{-\theta} \theta^r}{r!}$

Entonces:

$$m_X(t) = E(e^{tX}) = \sum_{r=0}^{\infty} e^{tX} \left[ \frac{e^{-\theta} \theta^r}{r!} dx \right] = e^{-\theta} \sum_{r=0}^{\infty} \frac{(e^t e^{\theta})^r}{r!} = e^{\theta(e^t - 1)}$$

A partir de los momentos de la distribución se puede estudiar la asimetría y el apuntamiento. De este modo, podemos caracterizar la forma de la distribución a partir de los momentos.



---

# (#437) . TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIId)

[MONOTEMA] En esta cuarta entrega del tercer capítulo de Probability Theory and Statistical Inference, de Aris Spanos, seguimos profundizando en la relación entre espacio de probabilidad y modelo de probabilidad.

Cuando las probabilidades son funciones conocidas de ciertos parámetros desconocidos  $\theta$ , entonces podemos transformar el espacio probabilístico en un modelo de probabilidad definido por:

$$\Phi = \{f(x; \theta), \theta \in \Theta, x \in \mathbb{R}_X\}$$

donde  $\Phi$  es una colección de funciones de densidad que dependen de un conjunto de parámetros  $\theta$  en el espacio paramétrico  $\Theta$ .

Podríamos usar también la función de distribución:

$$\Phi_F = \{F(x; \theta), \theta \in \Theta, x \in \mathbb{R}_X\}$$

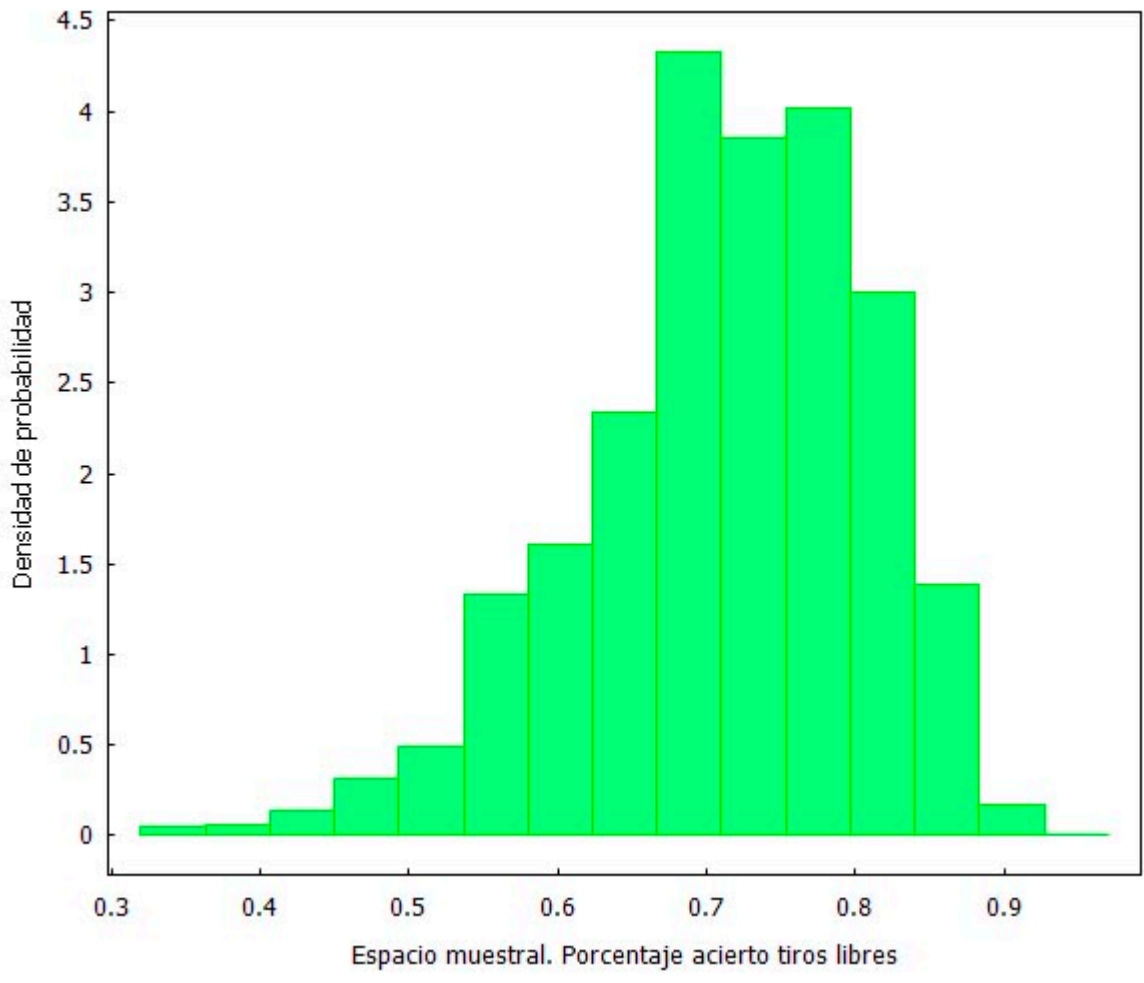
Pongamos un ejemplo usando la distribución Beta como modelo de probabilidad:

$$\Phi = f_x(x; \theta) = \left\{ \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B[\alpha, \beta]}, \theta := (\alpha, \beta) \in \mathbb{R}_+^2, 0 < x < 1 \right\}$$

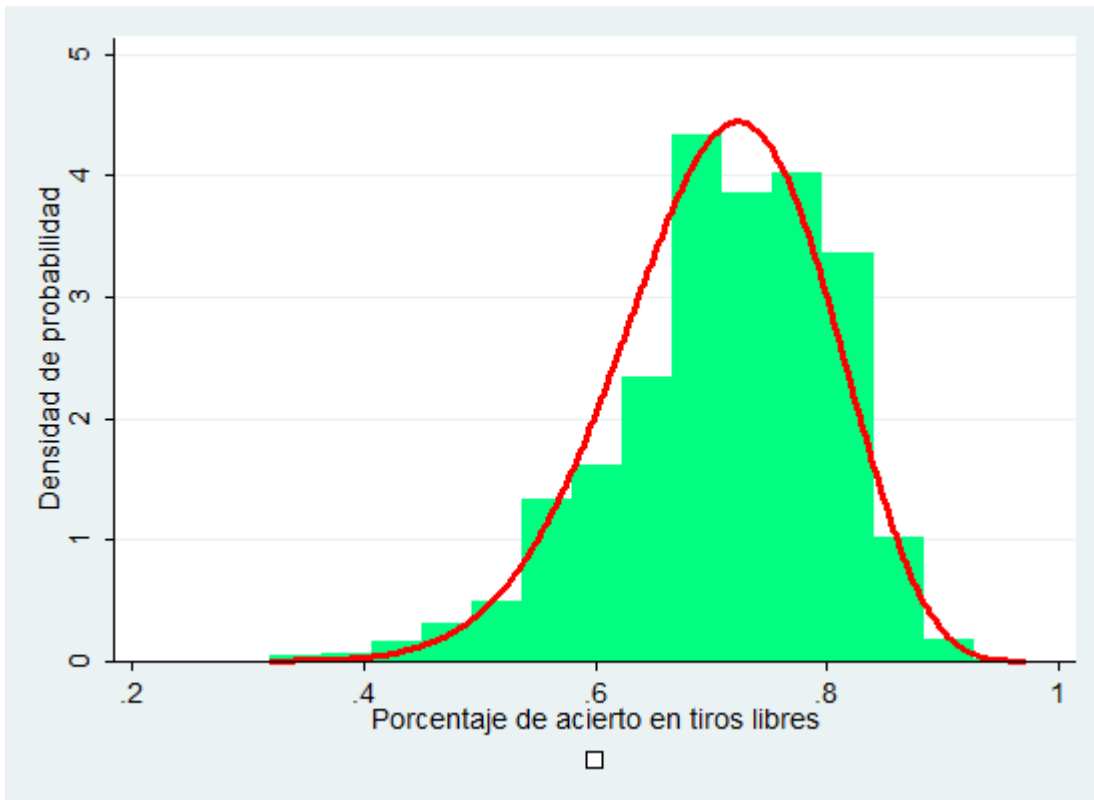
Podemos analizar el porcentaje de acierto en los tiros libres de los jugadores NBA hasta 2015 (el acumulado en sus respectivas carreras), para [aquellos que hubieran lanzado al menos 30 tiros libres.](#)

El histograma de la distribución es el siguiente:

```
data:read_list(file_search("RUTADELARCHIVO.txt"));
datatranspose:transpose(data);
estatura:datatranspose;
histogram (
estatura,
nclasses=15,
frequency=density,
xlabel="Espacio muestral. Porcentaje acierto tiros libres",
ylabel="Densidad de probabilidad",
fill_color=green,
fill_density=0.5);
```

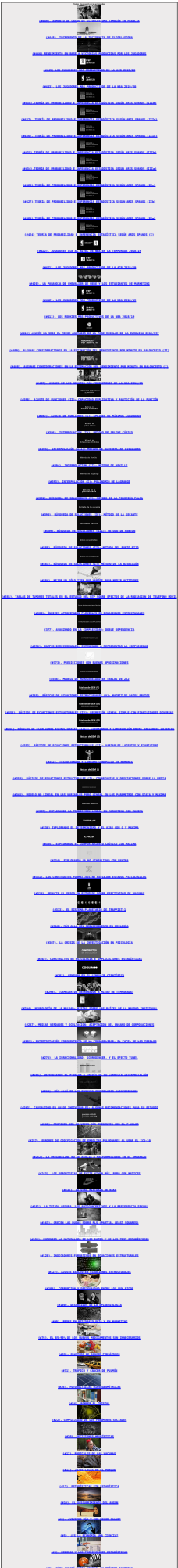


Para ello nos ayudamos de nuevo de Stata 13.0, y estipulamos una distribución Beta de parámetros (18, 7.5).



Es decir, para la modelización empírica debemos postular a priori una familia de densidades que refleje el mecanismo estocástico que da origen a los datos observados. Para ello, tiene especial relevancia el rango de valores de la variable aleatoria.

Estamos todavía al comienzo, pero ya hemos intuido cómo se plantea un modelo de probabilidad.





---

# (#436) . TEORÍA DE PROBABILIDAD E INFERENCIA ESTADÍSTICA SEGÚN ARIS SPANOS (IIIc)

[MONOTEMA] Continuamos con el tercer apartado del tercer capítulo de Probability Theory and Statistical Inference, de Aris Spanos.

Si vemos  $P_X(\cdot)$  como sólo una función del punto final del intervalo  $(-\infty, x]$ , entonces podemos definir la distribución acumulada (cdf):

$$F_X(\cdot) : \mathbb{R} \rightarrow [0, 1], F_X(x) = \mathbb{P}\{s : X(s) \leq x\} = P_X((-\infty, x])$$

Ahora sí hemos generado una función que relaciona el número real asignado a cada posible evento con su probabilidad de ocurrencia. Pero en este caso es la probabilidad acumulada.

Para el caso simple (discreto) tenemos la función de densidad:

$$f_x(\cdot) : \mathbb{R} \rightarrow [0, 1], f_x(x) = \mathbb{P}(X = x), x \in \mathbb{R}_X$$

Por tanto, los espacios probabilísticos pueden simplificarse en el caso de variables aleatorias discretas y continuas a los siguientes:

$$(S, \mathfrak{S}, \mathbb{P}(\cdot)) \xrightarrow{X(\cdot)} (\mathbb{R}, f_x(\cdot))$$

$$(S, \mathfrak{S}, \mathbb{P}(\cdot)) \xrightarrow{X(\cdot)} (\mathbb{R}, F_X(\cdot))$$

Spanos se plantea en este punto si se pueden definir funciones de densidad para variables continuas y funciones de distribución para variables discretas, y la respuesta es que sí.

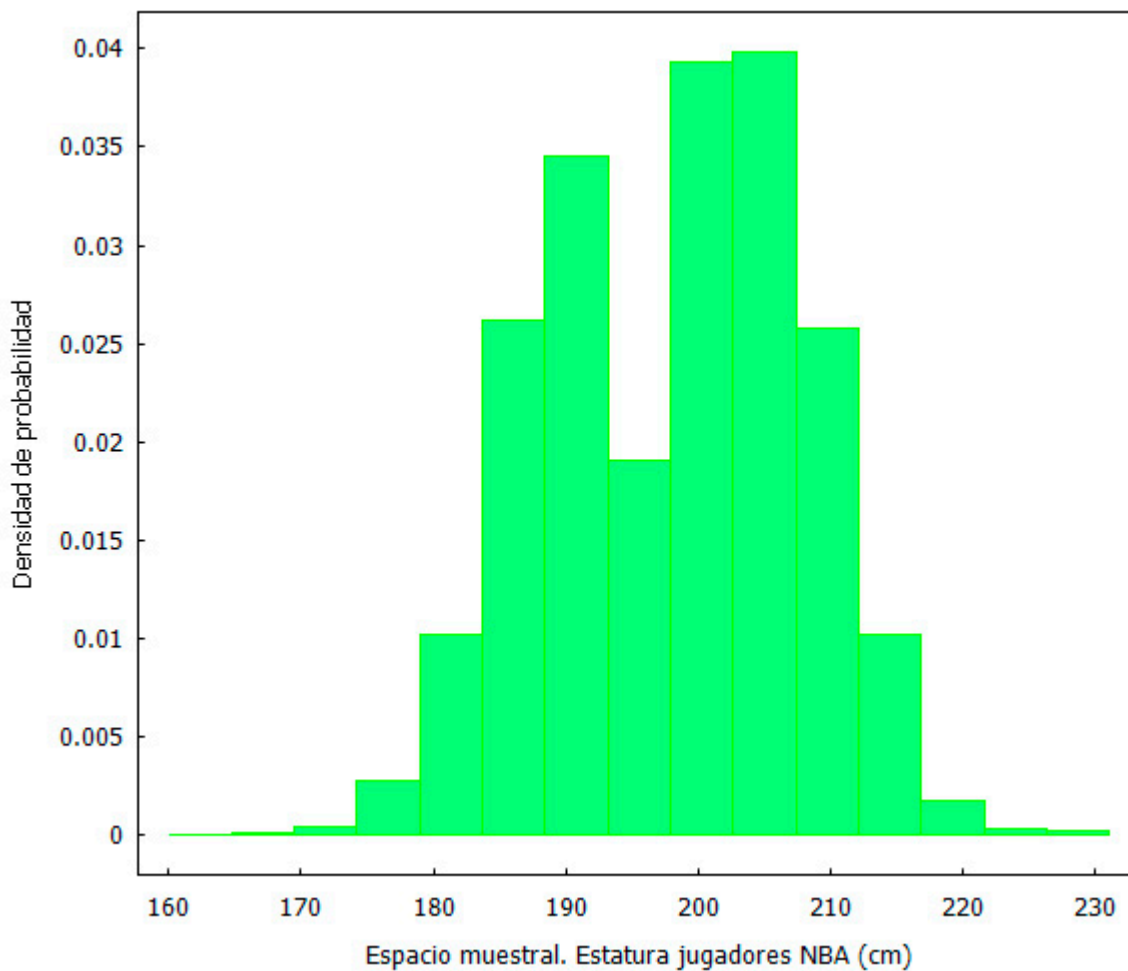
$$F_X(x) = \int_{-\infty}^x f_x(u)du, f_x(u) \geq 0$$

$$\int_a^b f_x(u)du = F_x(b) - F_x(a)$$

La estatura de los jugadores de la NBA se puede considerar como una variable continua. Desde el inicio de la NBA hasta el año 2015, hay 3984 jugadores cuya [estatura se muestra en este archivo](#).

El histograma de la distribución es el siguiente:

```
data:read_list(file_search("RUTADELARCHIVO.txt "));
datatranspose:transpose(data);
estatura:datatranspose;
histogram (
estatura,
nclasses=15,
frequency=density,
xlabel="Espacio muestral. Estatura jugadores NBA (cm)",
ylabel="Densidad de probabilidad",
fill_color=green,
fill_density=0.5);
```



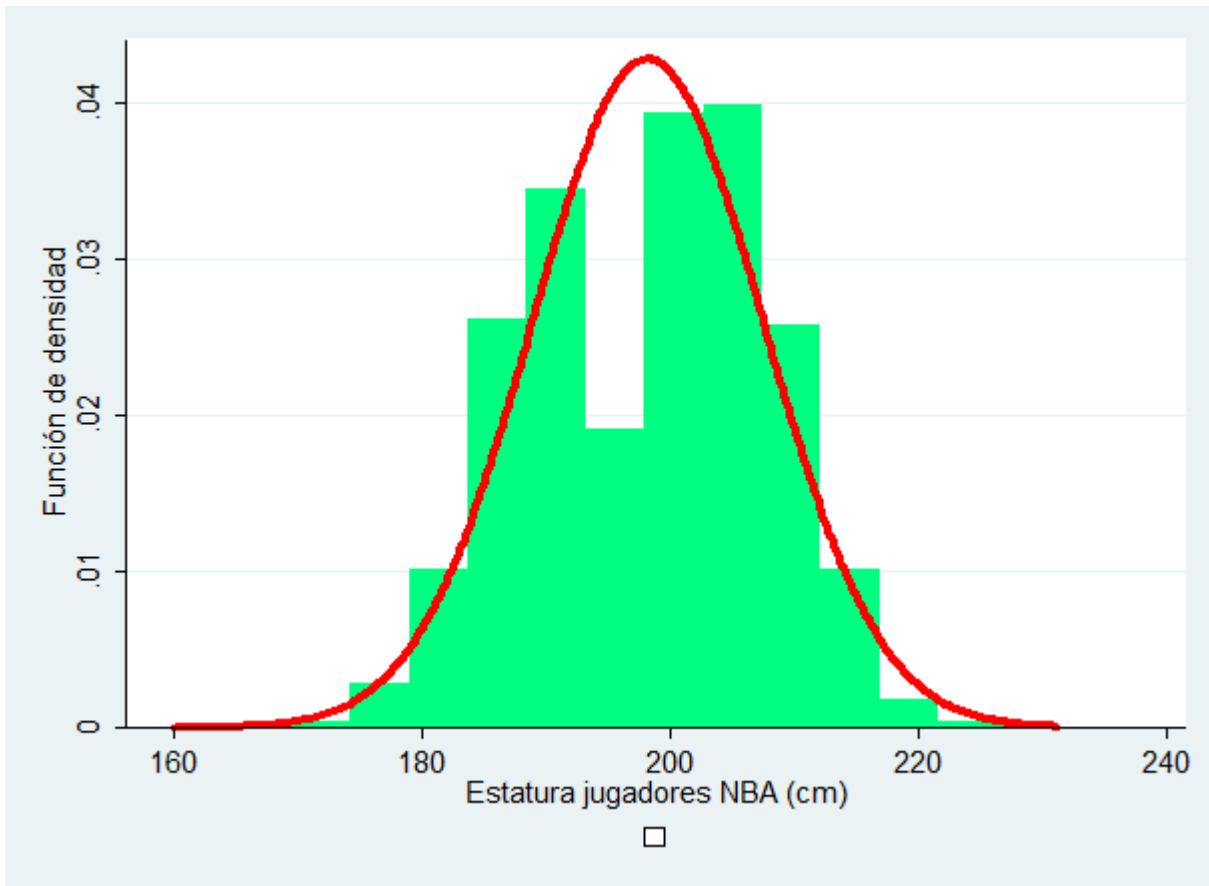
Es una distribución que se aproxima a una Normal, pero que no sabemos realmente si lo es. Recordemos que una distribución Normal tiene como función de densidad:

$$f_x(x; \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \theta := (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+, x \in \mathbb{R}$$

De este modo, podemos tomar como media y desviación típica la de la muestra, como una estimación de los parámetros poblacionales.

$$f_x(x; \theta) = \frac{1}{9.3\sqrt{2\pi}} e^{-\frac{(x-198.2)^2}{2 \cdot 9.3^2}}, \theta := (198.2, 9.3^2) \in \mathbb{R} \times \mathbb{R}_+, x \in \mathbb{R}$$

El resultado, tras emplear Stata 13.0, es el mostrado en el gráfico siguiente:



Sin embargo, otras distribuciones también podrían ajustarse a los datos. Por ejemplo, la distribución Weibull: