Modelado Estocástico

Prof. Fernando Grosz

fgrosz@udesa.edu.ar

Clase 1

Trabajando con Retornos Simples y logarítmicos

Fórmula de Retorno Simple:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$$

donde P_t es el precio de un activo en el momento “t”.

También suele hacerse la distinción entre retorno neto (el de la fórmula) y el retorno bruto, que es $1 + R_t$. Notemos que el retorno bruto, $1 + R_t = \frac{P_t}{P_{t-1}}$.

Fórmula de Retorno Logarítmico:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(P_t) - \ln(P_{t-1})$$
$$= \Delta \ln(P_t)$$

O sea, el retorno logarítmico es el logaritmo del retorno bruto simple.

El retorno que como inversores o administradores de portafolios van a calcular (cuanta plata ganan o pierden) es el **retorno simple**. ¿Qué relación hay entre el retorno simple y el logarítmico?

Aproximaciones por polinomios de Taylor:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!}f''(x_0)(x - x_0)^2 + \cdots$$

Si tomo la aproximación de grado 1, me queda:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

Usando esto para la función logaritmo y tomando $x_0 = 1$, obtenemos:

$$ln(x) \approx ln(1) + \frac{1}{1}(x - 1)$$

De modo que si trabajamos **en un entorno al punto $\boldsymbol{x_0 = 1}$**,

$$ln(x) \approx x - 1$$

Por lo tanto,

$$r_t = ln\left(\frac{P_t}{P_{t-1}}\right) = ln(1 + R_t) \approx R_t$$

Por ejemplo,

$$ln(1.02) = 0.01980263 \approx 0.02$$

Esta aproximación va a ser “buena” para retornos que son “chicos”. Cuando los retornos sean grandes, ya sean positivos o negativos, la diferencia entre trabajar con retornos simples o logarítmicos puede ser importante. Esto hay que tenerlo presente especialmente cuando analizamos riesgos asociados a un portafolio (ya que vamos a mirar la cola izquierda). Por ejemplo,

$$ln(1.25) = 0.22314355, \text{ y}$$

$$ln(0.75) = -0.287682.$$

Es decir, cuando trabajamos con retornos logarítmicos, los retornos negativos se hacen más negativos y los positivos se hacen menos positivos en relación a trabajar con retornos simples.

También, notemos que cuando usamos retornos logarítmicos, el retorno acumulado en un período de varios días es la ***suma*** de los retornos logarítmicos de cada uno de los días de ese período. Por ejemplo, si tenemos los retornos logarítmicos de los 5 días de la semana, r_1, r_2, r_3, r_4, r_5 en donde $r_1 = \ln\left(\frac{P_1}{P_0}\right), r_2 = \ln\left(\frac{P_2}{P_1}\right)$, …, $r_5 = \ln\left(\frac{P_5}{P_4}\right)$, notemos que la suma de los retornos logarítmicos diarios es el retorno logarítmico de la semana:

$$\begin{aligned} r_1 + r_2 + r_3 + r_4 + r_5 &= \ln\left(\frac{P_1}{P_0}\right) + \ln\left(\frac{P_2}{P_1}\right) + \cdots + \ln\left(\frac{P_5}{P_4}\right) \\ &= [\ln(P_1) - \ln(P_0)] \\ &\quad + [\ln(P_2) - \ln(P_1)] + \cdots \\ &\quad + [\ln(P_5) - \ln(P_4)] \\ &= \ln(P_5) - \ln(P_0) = \ln\left(\frac{P_5}{P_0}\right) \end{aligned}$$

Notemos que $\frac{P_5}{P_0}$ es el retorno bruto entre el día (o período) 0 y el día 5.

Ejercicio 1: si un activo financiero presenta cada uno de los 5 días de la semana un retorno logarítmico diario del -2% (o sea, 2% negativo), ¿cuál será el precio del activo financiero final del quinto día si el día 0 valía \$ 130?

Ejercicio 2: un activo financiero presentó un retorno logarítmico promedio diario de 0.177% en un mes de 20 días de trading. En base a esta información, ¿cuál es el retorno simple acumulado en el mes?

Nota al pie: repaso de logaritmos

La función logaritmo

La función logaritmo es la función inversa de la función exponencial. Al subíndice se lo llama "base del logaritmo".

$$log_{10} 100 = 2 \Leftrightarrow \; 10^2 = 100$$

$$log_{10} 10000 = 4 \Leftrightarrow \; 10^4 = 10000$$

$$log_2 8 = 3 \Leftrightarrow \; 2^3 = 8$$

$$ln\, e^5 = 5 \; \Leftrightarrow e^5 = e^5$$

donde $log_e e^5 = ln\, e^5$

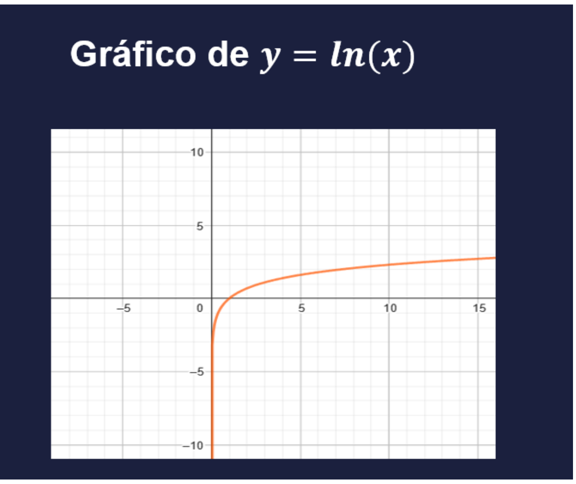
$$e = 2.71828182 \ldots viene\ de:$$

$$e^r = \lim_{T \to \infty} \left(1 + \frac{r}{T}\right)^T$$

Notar que si $\boldsymbol{ln(A) = D}$

Entonces, $\boldsymbol{e^{ln(A)} = e^D}$

Como $\boldsymbol{e^{ln(A)} = A}$, entonces, $\boldsymbol{A = e^D}$



Respuesta al Ejercicio 1:

El retorno logarítmico de la semana es del -10%. De modo que $\ln\left(\frac{P_5}{P_0}\right) = -0.10.$ Entonces, $\ln(P_5) - \ln\,(P_0) = -0.10.$

Como $P_0 = 130,$

$$\ln(P_5) = \ln(P_0) - 0.10 = \ln(130) - 0.10$$

$$\ln(P_5) = 4.7675344505$$

$$P_5 = e^{4.7675344505} = 117.63$$

Respuesta al Ejercicio 2:

Nos dicen que el promedio de los 20 retornos logarítmicos es 0.177%.

Como el promedio es: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Pasando el "n" a la izquierda del igual queda, $n\bar{x} = \sum_{i=1}^{n} x_i$.

De modo que la sumatoria de los 20 retornos logarítmicos diarios tiene que dar 20*0.177% = 3.54%. Pero entonces, el retorno logarítmico acumulado en el mes es 3.54%, y consecuentemente,

$$\ln\left(\frac{P_{20}}{P_0}\right) = 0.0354$$

$$\frac{P_{20}}{P_0} = e^{0.0354} = 1.036$$

De modo que el retorno simple bruto es 1.036 y el retorno simple neto acumulado en el mes es $1.036 - 1 = 0.036 = 3.6\%$.

Modelos de datos y enfoque frecuentista

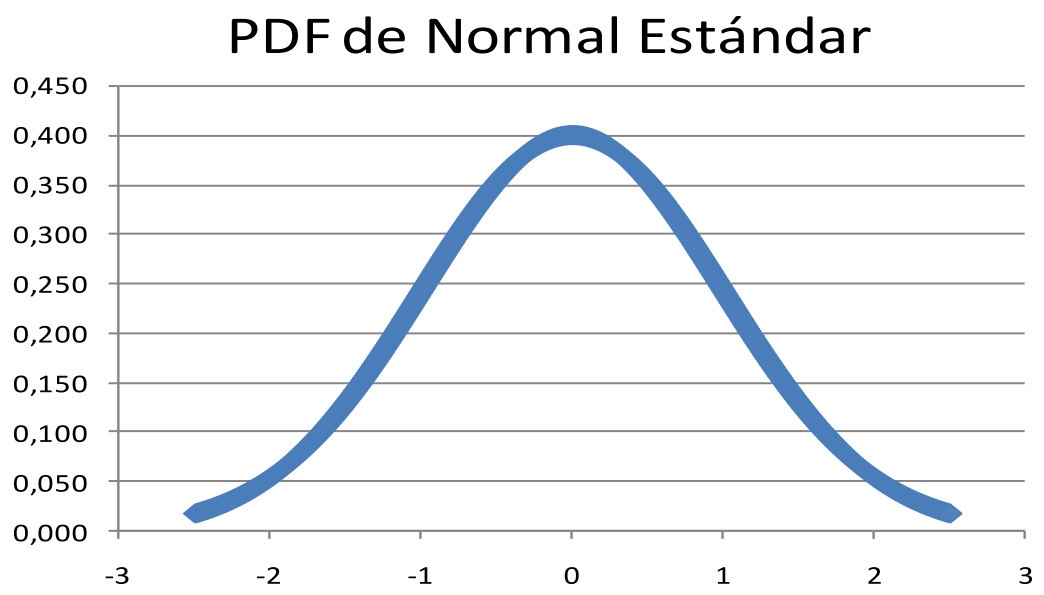
- Serie de tiempo vs histograma
- Consistencia de los modelos de datos
- Analizar si hay o no *clusters* de volatilidad

Volatilidad: Criterio de Media-Varianza

Recordemos la forma funcional de la densidad (pdf) de una variable aleatoria normal, $X \sim N(\mu, \sigma^2)$:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

Si estandarizo X, obtengo $Z \sim N(0,1)$ (una variable aleatoria normal estándar). Estandarizar quiere decir “redefinir” la variable aleatoria de manera que tenga esperanza cero y varianza uno. Para lograrlo, hay que restarle la media y dividir por el desvío estándar.



Recordemos que si a y b son constantes, entonces $E(a+bX) = a + bE(X)$ y que $Var(a+bX) = b^2 Var(X)$.

De modo que si la variable aleatoria X tiene media μ y varianza σ^2, entonces $Z = \frac{X-\mu}{\sigma}$ tendrá media cero y varianza uno.

Cuando trabajamos con series de tiempo largas, la media y varianza vamos a tener que calcularlas a partir de nuestra muestra.

En una muestra, la media muestral es

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Y la varianza muestral se calcula como

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

La raíz cuadrada de la varianza nos da el desvío estándar muestral:

$$Desvío\ Estándar = s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

A este desvío estándar solemos llamarlo volatilidad (del retorno del activo o portafolio en cuestión).

Bajo el supuesto de normalidad de los retornos, por ejemplo, si especificamos un nivel de pérdidas expresado en términos de retornos, r^*, entonces podemos querer determinar la probabilidad de que ocurra un nivel de pérdidas mayor a r^* (siempre bajo el supuesto de distribución normal):

$$P(X \leq r^*) = \int_{-\infty}^{r^*} f(x|\mu, \sigma^2)dx$$

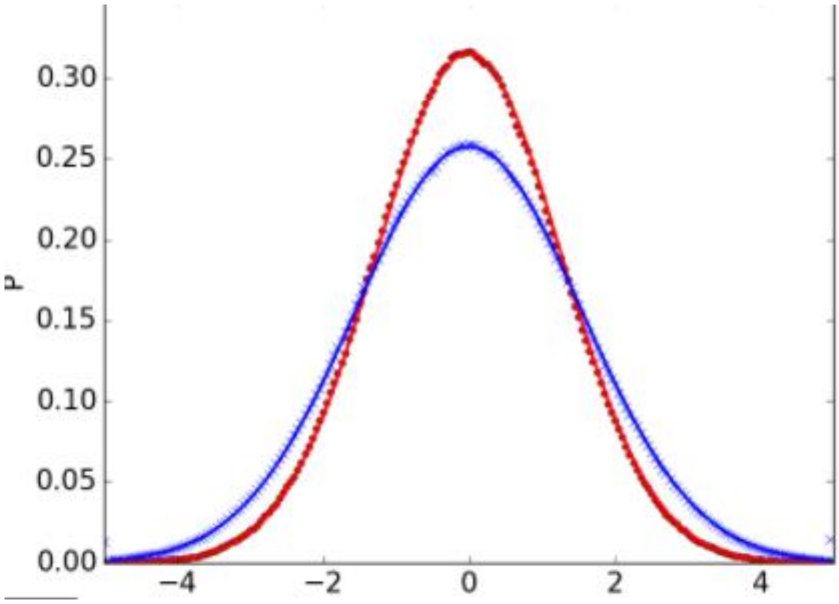
$$= \int_{-\infty}^{r^*} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Alternativamente, podemos especificar un nivel de confianza "cl", y su correspondiente valor crítico α_{cl} para la distribución normal estándar,

$$P(Z \leq \alpha_{cl}) = 1 - \ cl$$

con lo cual, el retorno mínimo que vamos a obtener al $cl * 100\%$ de confianza va a ser $X = \mu + \alpha_{cl}\sigma$. Por ejemplo, si $cl * 100\% = 99\%$, $\alpha_{cl} = -2.33$ y consecuentemente, $X = \mu - 2.33\sigma$

Misma Media, diferente varianza:



Dificultades de la distribución normal:

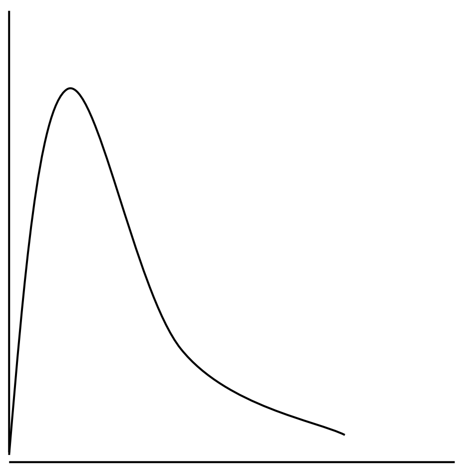
- *Upside* y *Downside Risk* no son equivalentes
- Ignoramos posibles consecuencias de la asimetría y de la curtosis

Recordemos cómo se miden los coeficientes de asimetría (*Skewness*) y curtosis para una muestra de n observaciones:

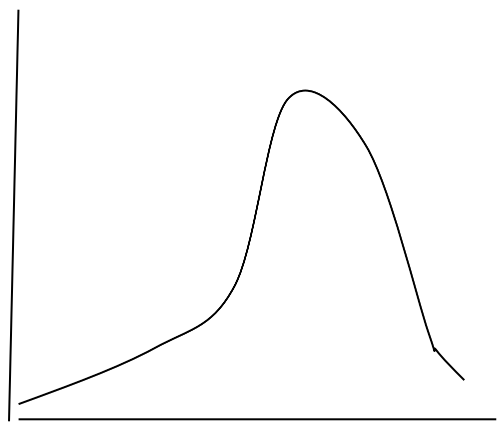
$$Skewness = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$$

donde "s" es el desvío estándar muestral.

Asimetría positiva (*positively skewed*): se da cuando la cola derecha es pesada. La pdf tendrá la forma:



Asimetría negativa (*negatively skewed*): cola izquierda pesada. La pdf tendrá la forma:



Curtosis: este concepto está relacionado al cuarto momento de una variable aleatoria. La curtosis se define como:

$$\mathrm{K} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4$$

La **curtosis en exceso (*excess kurtosis*)** se define como la curtosis, K, menos 3:

$$K_E = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4 - 3$$

En el caso de una variable aleatoria normal, la curtosis es igual a 3 y por lo tanto, $K_E = 0$. Las variables aleatorias con $K_E = 0$ se llaman **mesocúrticas**. Cuando $K_E > 0$, la distribución se llama **leptocúrtica** y se caracteriza por tener las colas más pesadas que una normal. El otro caso, $K_E < 0$, se denomina **platicúrtica** y se caracteriza por tener colas menos pesadas que una normal (la pdf de la variable aleatoria es más plana o con forma de "n").

Note que la simetría y curtosis se miden sobre la variable estandarizada y, por lo tanto, son **medidas de forma** de la densidad (distribución) de la variable aleatoria.

Recordemos:

Para la distribución normal estándar,

$$E(Z) = \int_{-\infty}^{+\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz = 0$$

$$Var(Z) = E(Z^2) = \int_{-\infty}^{+\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz = 1$$

$$E(Z^3) = \int_{-\infty}^{+\infty} z^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz = 0$$

$$E(Z^4) = \int_{-\infty}^{+\infty} z^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz = 3$$

La función generadora de momentos (mgf) para $X \sim N(\mu, \sigma^2)$ es $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$

Validación del supuesto de Normalidad:

Estadístico de Jarque-Bera:

$$JB = \frac{n}{6}\left(Skewness^2 + \frac{K_E^2}{4}\right)$$

donde $Skewness^2$ es el coeficiente de asimetría al cuadrado, es decir, $Skewness = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$, y donde K_E^2 es el coeficiente de curtosis en exceso al cuadrado, o sea, $K_E = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4 - 3$.

El estadístico de JB tiene una distribución asintótica que es chi-cuadrado con 2 grados de libertad. Bajo la hipótesis nula, los datos siguen una distribución normal. Los valores críticos son aproximadamente 9 ó 6 dependiendo de si trabajamos con una probabilidad de error tipo I del 1% ó del 5%.

Enfoque de Q-Q plots (Visual):

1) Dada una muestra de n observaciones ordenamos los datos de menor a mayor, $x_1, \ldots, x_n$, y construimos una función de probabilidad acumulada, $\widehat{P}_i = \sum_{j=1}^{i} \frac{1}{n} = \frac{i}{n}$, en donde cada observación tendrá asignada una probabilidad igual a $\frac{1}{n}$.
2) Dados los $\widehat{P}_i$ calculamos los $\hat{z}_i$ consistentes con una distribución normal estándar que satisfacen que $\hat{z}_i = \Phi^{-1}(\widehat{P}_i)$. Es decir, el valor del soporte de la normal estándar que es consistente con cada una de las probabilidades acumuladas calculadas en el punto (1).
3) Calculamos la media $\bar{x}$ y el desvío estándar s para las observaciones

$x_1, \ldots, x_n$ y reemplazamos cada x_i por su versión estandarizada $z_i = \frac{x_i - \bar{x}}{s}$.

4) Si la hipótesis original (de normalidad) es correcta, deberíamos verificar que $\hat{z}_i = z_i$ y consecuentemente, si realizamos un gráfico de $\hat{z}_i$ en el eje vertical contra z_i en el eje horizontal, bajo la hipótesis de que la distribución que siguen los datos es normal, deberíamos observar que las observaciones se encuentran sobre una recta de 45 grados.

¿Cómo sería el Q-Q plot para el caso de una variable aleatoria con distribución t-Student y cómo sería para el caso de la Uniforme (ambas son simétricas)?