

APUNTES DE SIMULACIONES DE MONTE CARLO

ALEX CARRASCO MARTINEZ
(alex.carmar93@gmail.com)

Estudiante de primer año (maestría) - PUC Rio

Agosto - Octubre, 2017

Nota

El presente apunte recoge las principales enseñanzas de Winkelried (2016) con dos diferencias: a) he intentado dar algo más de detalle en algunas propiedades de convergencia y distribución asintótica, y b) las aplicaciones presentadas en este documento fueron implementadas en MATLAB. *Este documento no es oficial y no tiene ninguna relación adicional con la mencionada universidad.* Todos los errores son enteramente culpa mía.

CONTENTS

1	Introducción: algoritmo general	2
2	Aplicaciones I: muestra aleatoria	3
2.1	Eficiencia relativa - media y mediana	3
2.2	Inferencia en modelos de regresión	3
3	Aplicaciones II: Series de tiempo estacionarias	4
4	Aplicaciones III: Series de tiempo no estacionarias	8

1 INTRODUCCIÓN: ALGORITMO GENERAL

Ver también Hansen (2017). Suponga que $\{z_i\}_{i=1}^n$ es una muestra aleatoria de n vectores aleatorios de orden k obetenida de un proceso generador de datos $F(c|\theta_0) = \mathbb{P}[z_i \leq c|\theta_0]$. Considere un estadístico de la forma $T_n = T(z_1, \dots, z_n; \theta_0)$ cuyas propiedades es de nuestro interés, luego aunque la distribución asintótica puede ser conocida, la distribución de este estimador suele ser muy difícil de obtener para muestras finitas, $G_n(c|F, \theta_0) = \mathbb{P}(T_n \leq c|F, \theta_0)$. Es posible obtener una aproximación de $G_n(\cdot)$ usando *simulaciones de Monte Carlo*,

Algoritmo para simulaciones de Monte Carlo. Defina $F(\cdot)$, θ_0 , $T_n : \mathbb{R}^{nk} \rightarrow \mathbb{R}$, $M \in \mathbb{N}$ e inicie con $m = 1$:

1. Genere una muestra aleatoria de tamaño n , $\{z_i^*\}_{i=1}^n$.
2. Calcule $T_{n,m}$, y guarde.
3. Mientras $n \leq M$: $n = n + 1$ y vuelva al paso 1.

Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ con $\mathbb{E}|g(T_n)| < \infty$, luego por la *ley de grandes números* tenemos la siguiente relación

$$\frac{1}{M} \sum_{m=1}^M g(T_{n,m}) \xrightarrow{p} \mathbb{E}[g(T_n)] \quad (1.1)$$

Con este resultado, se puede estimar cualquier función de $g(\cdot)$ siempre y cuando la media de esta nueva variable aleatoria sea finita. Por ejemplo, (i) *primer momento - media*, $g(t) = t$, (ii) *momento de orden p* $g(t) = t^p$, y (iii) *probabilidades*, $\mathbb{P}(T_n \in \Xi_A)$, donde Ξ_A es un evento medible, puede ser aproximado mediante $g(t) = \mathbb{I}(t \in \Xi_A)$; así la c.d.f. de T_n puede ser aproximado definiendo $\Xi_A = (-\infty, c]$.

Además, sabemos que si $\mathbb{E}|g(T_n)|^2 < \infty$, luego

$$\sqrt{M} \left[\frac{1}{M} \sum_{i=1}^M g(T_{n,m}) - \mathbb{E}[g(T_n)] \right] \xrightarrow{d} \mathcal{N}(0, \sigma_g^2) \quad (1.2)$$

donde $\sigma_g^2 = \text{var}[g(T_n)]$. El *error estándar numérico* es definido como $EEN_M = \frac{\hat{\sigma}_g}{\sqrt{M}}$, luego se puede construir un intervalo al $(1 - \alpha/2)\%$ de confianza para nuestra aproximación: $\frac{1}{M} \sum_{i=1}^M g(T_{n,m}) \pm z_{\alpha/2} EEN_M$, donde $\Phi(z_{\alpha/2}) = 1 - \alpha/2$. El ancho de este intervalo, $2z_{\alpha/2} EEN_M$, da una medida precisa de la calidad de nuestra aproximación. Por tanto, mientras más grande es M las ganancias en precisión de aumentar el número de replicaciones es menor.

2 APLICACIONES I: MUESTRA ALEATORIA

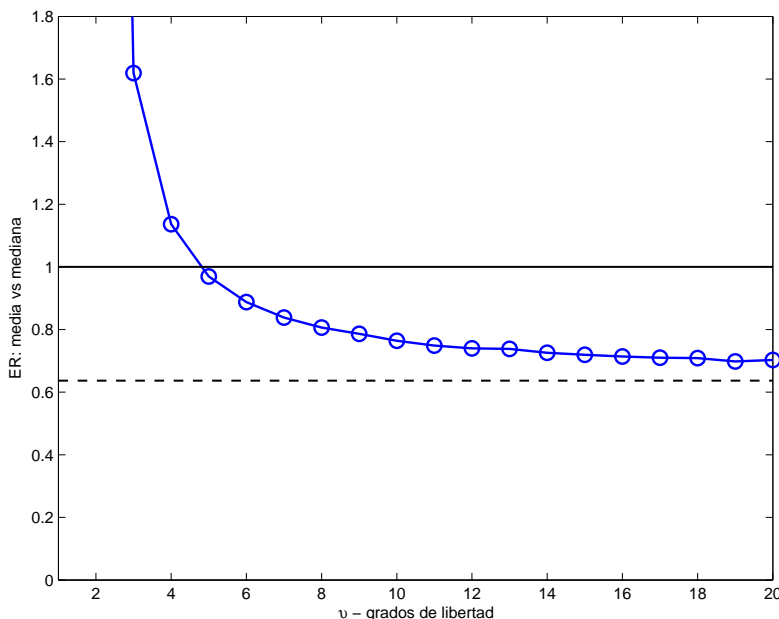
2.1 EFICIENCIA RELATIVA - MEDIA Y MEDIANA

La eficiencia relativa de la media respecto a la mediana se define como

$$ER(\bar{y}, \tilde{y}) = \frac{V(\bar{y})}{V(\tilde{y})} \quad (2.1)$$

El código `MonteCarlo2.m` genera una muestra de 100 observaciones provenientes de la distribución t-Student con v grados de libertad. [Figure 1](#) muestra los resultados con $M = 10^5$ replicas.

Figura 1: Eficiencia relativa: media vs mediana



Notes: `MonteCarlo2.m` - Se genera una muestra de 100 observaciones de una t-Student. El número de replicas fue de $M = 10^5$.

2.2 INFERENCIA EN MODELOS DE REGRESIÓN

Consideremos el siguiente modelo de regresión:

$$y_t = \alpha + \beta x_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2) \quad (2.2)$$

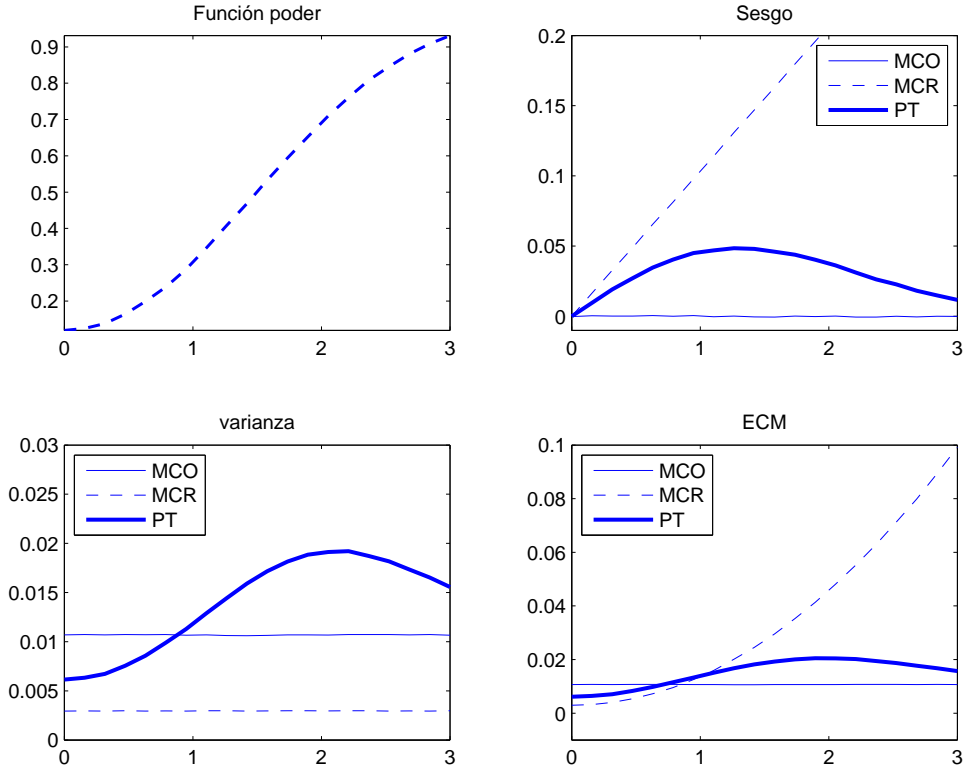
Podemos estimar a β mediante MCO, $\hat{\beta}_{MCO}$, o mediante MCO restringidos ($\alpha = 0$), $\hat{\beta}_{MCR}$. Un tercer estimador es un híbrido entre $\hat{\beta}_{MCO}$ y $\hat{\beta}_{MCR}$:

$$\hat{\beta}_{PT} = \hat{\beta}_{MCR} + \mathbb{I}(|\tau_{\alpha}| > \tau_{crit})[\hat{\beta}_{MCO} - \hat{\beta}_{MCR}]$$

donde $\tau_\alpha = \frac{\hat{\alpha}_{MCO}}{\hat{\sigma}_{\hat{\alpha}}}$. Esto es, el estimador PT está basado en la decisión de rechazar o no la hipótesis nula $H_0 : \alpha = 0$. El código `MonteCarlo3.m` obtiene la distribución de estos estimadores para una muestra finita de $n = 20$ observaciones donde los regresores son fijos. Se realizaron $M = 10^5$ replicas. [Figure 2](#) y [Figure 3](#) muestran los resultados de la simulación.

Un resultado interesante es que el estimador PT siempre es dominado por MCO y MCR mediante el criterio de menor error cuadrático medio.

Figura 2: Propiedades de los estimadores: MCO, MCR y PT



Notes: MonteCarlo3.m - Se genera una muestra de 20 observaciones de $\mathcal{N}(0, 2^2)$. El número de replicaciones fue de $M = 10^5$.

3 APLICACIONES II: SERIES DE TIEMPO ESTACIONARIAS

Las condiciones para estacionariedad de los procesos estocásticos estudiados aquí son idénticas a las requeridas para la ergodicidad, por ello no será necesaria la distinción de uno con el otro. En particular, una condición suficiente para la ergodicidad es que las autocovarianzas sean *absolutamente sumables*

$$\sum_{\tau=0}^{\infty} |\gamma_{\tau}| < \infty \quad (3.1)$$

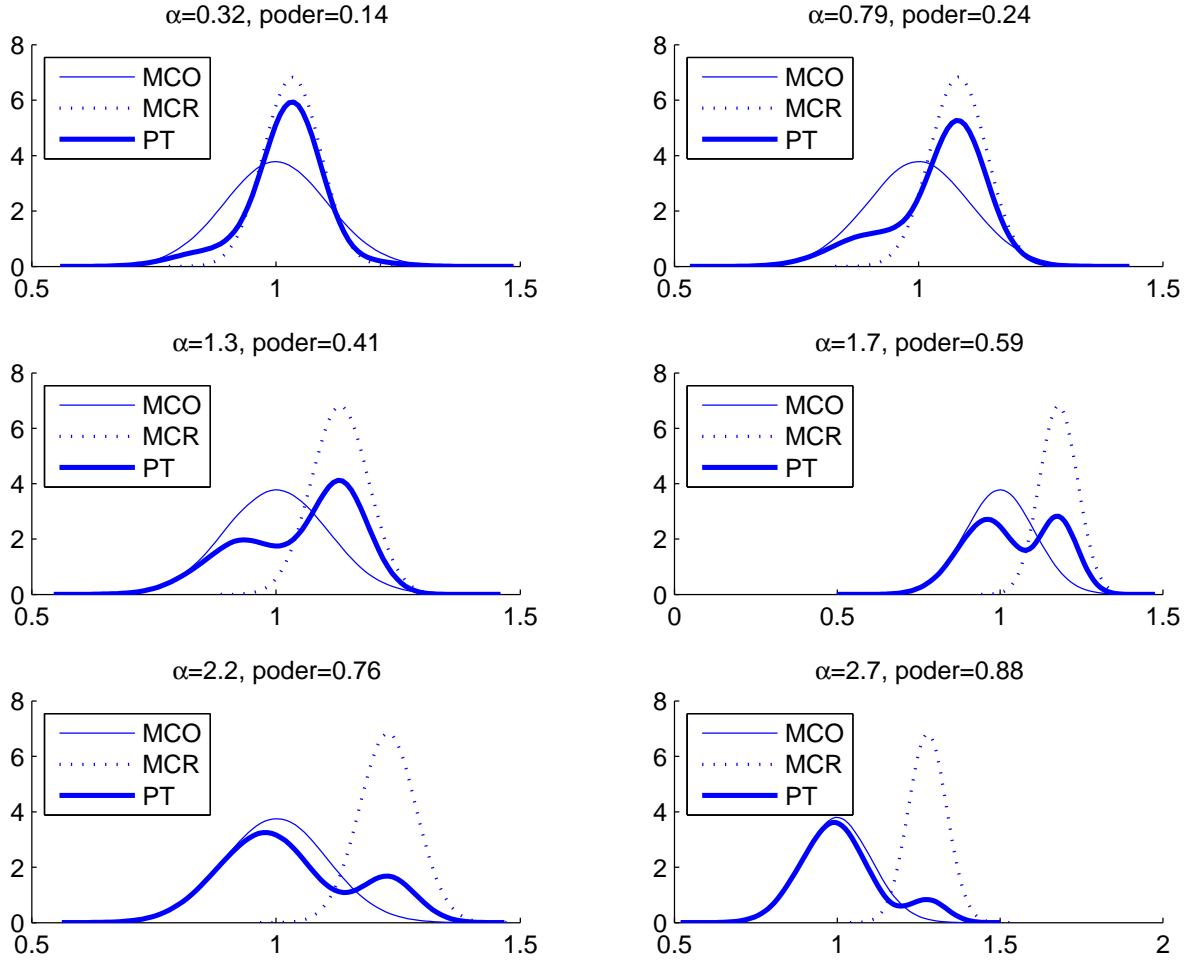
donde $\gamma_{\tau} = \mathbb{C}(y_t, y_{t-\tau}) = \mathbb{E}[(y_t - \mu_t)(y_{t-\tau} - \mu_{t-\tau})]$. Note que esta condición implica $\lim_{t \rightarrow \infty} \gamma_t = 0$. Ahora, para aplicar leyes asintóticas a muestras con observaciones dependientes vale la pena examinar la estructura de la varianza del promedio muestral:

$$\begin{aligned} \mathbb{V}(\bar{y}) &= \frac{1}{T^2} \mathbb{V}\left(\sum_{t=1}^T y_t\right) = \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \mathbb{C}(y_t, y_s) \\ &= \frac{1}{T^2} \left[\sum_{t=1}^T \mathbb{C}(y_t, y_1) + \sum_{t=1}^T \mathbb{C}(y_t, y_2) + \cdots + \sum_{t=1}^T \mathbb{C}(y_t, y_T) \right] \\ &= \frac{1}{T^2} \left[(\gamma_0 + \gamma_1 + \cdots + \gamma_{T-1}) + (\gamma_1 + \gamma_0 + \cdots + \gamma_{T-2}) \cdots + (\gamma_{T-1} + \gamma_{T-2} + \cdots + \gamma_0) \right] \\ &= \frac{1}{T^2} [T\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\gamma_{T-1}] \\ &= \frac{\gamma_0}{T} + \frac{2}{T} \sum_{s=1}^{T-1} \left(1 - \frac{s}{T}\right) \gamma_s \end{aligned}$$

Dado que $s \leq T \rightarrow 1 - s/T \leq 1$, luego

$$\lim_{T \rightarrow \infty} \mathbb{V}(\bar{y}) \leq \lim_{T \rightarrow \infty} \left\{ \frac{\gamma_0}{T} + \frac{2}{T} \sum_{s=1}^{T-1} |\gamma_s| \right\} = 0$$

Figura 3: Distribución de los estimadores: MCO, MCR y PT



Notes: MonteCarlo3.m - Se genera una muestra de 20 observaciones de $\mathcal{N}(0, 2^2)$. El número de replicaciones fue de $M = 10^5$.

Luego

$$\bar{y} - \mathbb{E}(\bar{y}) \xrightarrow{p} 0 \quad (3.2)$$

$$\sqrt{T}[\bar{y} - \mathbb{E}(\bar{y})] \xrightarrow{d} \mathcal{N}(0, \sum_{j=-\infty}^{\infty} \gamma_j) \quad (3.3)$$

Debido que a mayor persistencia de un proceso estocástico aumenta varianza no debe sorprendernos el hecho que, dado un mismo tamaño muestral, la varianza de la media muestral sea mayor¹. Es útil percibir que el teorema ergódico también nos garantiza que

$$(T-i)^{-1} \sum_{t=i+1}^T (y_t - \bar{y})(y_{t-i} - \bar{y}) \xrightarrow{p} \gamma_i \quad (3.4)$$

Si $y_t = \Psi(L)\epsilon_t$ donde $\epsilon_t | \mathcal{F}_{t-1}$ es un diferencias martingale con media cero y varianza σ^2 y $\sum_{j=0}^{\infty} |\psi_j| < \infty$, luego $\mathbb{E}(y_{t-j}\epsilon_t) = 0$ y $\mathbb{E}(y_{t-j}^2\epsilon_t^2) = \sigma^2\gamma_0$ para $j > 0$, por tanto,

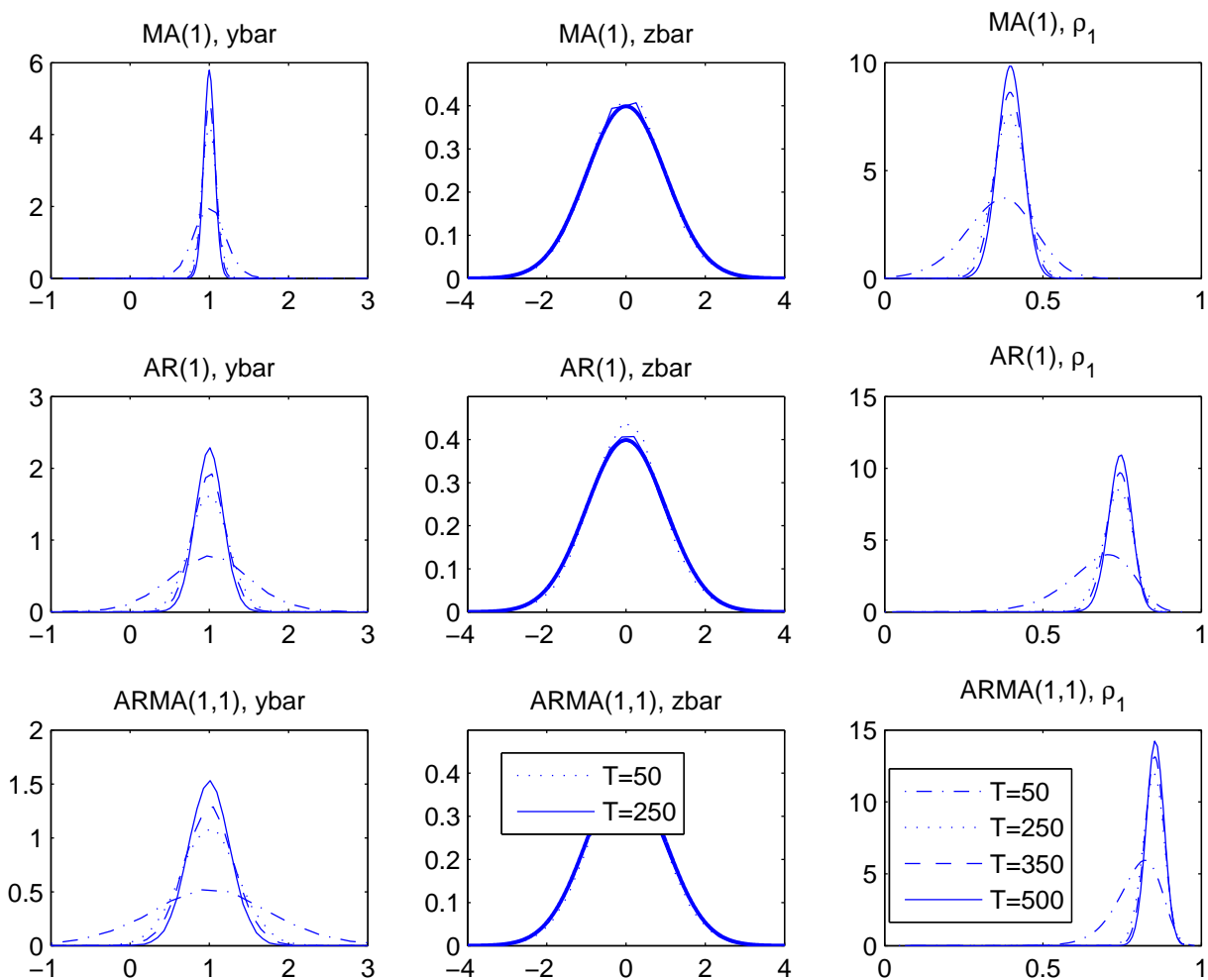
$$T^{-1/2} \sum_{t=1}^T y_{t-j}\epsilon_t \xrightarrow{d} \mathcal{N}(0, \sigma^2\gamma_0) \quad (3.5)$$

¹Una forma fácil de notar este resultado es obtener el ratio de tamaño muestral

$$\frac{T_1}{T_2} = \frac{\Omega_1}{\Omega_2}$$

así si el proceso 1 tiene mayor varianza luego para que la media muestral tenga varianza comparable en ambos procesos el tamaño muestral del proceso 1 deberá ser mayor.

Figura 4: Distribución asintótica de variables dependientes



Notes: `MonteCarlo4.m` - Se genera una muestra de distribuciones MA(1), AR(1) y ARMA(1,1). El número de replicaciones fue de $M = 10^5$.

Ahora, considere

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t = x_t' \phi + \epsilon_t$$

donde $x_t = (y_{t-1}, \dots, y_{t-p})'$ y $\phi = (\phi_1, \dots, \phi_p)'$, luego podemos calcular la distribución asintótica de $\hat{\phi}$

$$\sqrt{T}(\hat{\phi} - \phi) = \left[T^{-1} \sum_{t=1}^T x_t x_t' \right]^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T x_t \epsilon_t$$

Sigue

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T x_t \epsilon_t \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q)$$

$$Q = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix}$$

Por tanto,

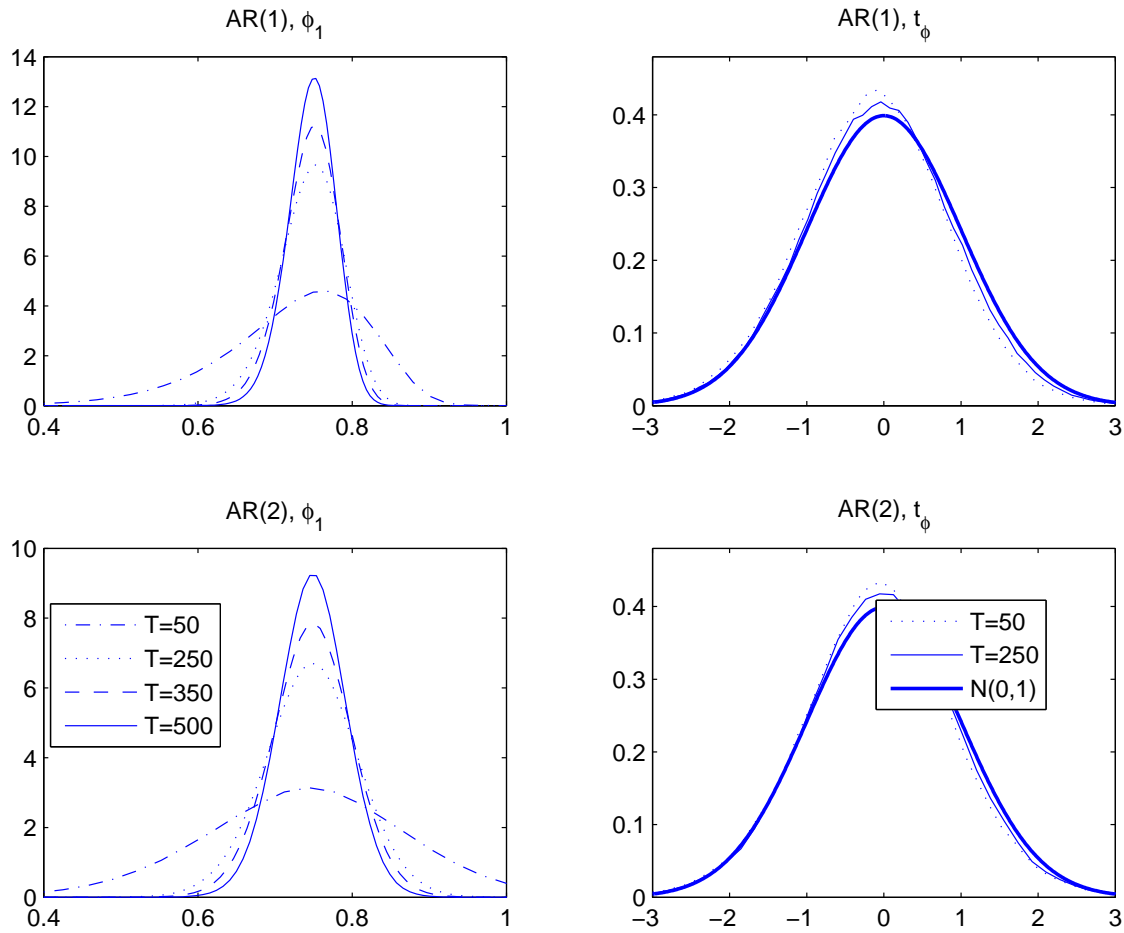
$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q^{-1})$$

Si y_t es efectivamente un $AR(p)$ estacionario luego Q puede ser reescrita como $\sigma^2 \tilde{Q}$ donde \tilde{Q} es independiente de σ , finalmente

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} \mathcal{N}(0, \tilde{Q}^{-1}) \quad (3.6)$$

La distribución de un estadístico $t_{\phi_j} = \frac{\hat{\phi}_j - \phi}{\hat{\sigma}_{\hat{\phi}_j}}$ es una normal estándar debido a que $\sqrt{T}\hat{\sigma}_{\hat{\phi}_j} \xrightarrow{p} \sqrt{\sigma^2 q_{jj}^{(-1)}}$. (ver Figure 5)

Figura 5: Distribución asintótica $\hat{\phi}_1$



Notes: MonteCarlo5.m - Se genera una muestra de distribuciones AR(1) y AR(2). El número de replicaciones fue de $M = 10^6$.

4 APLICACIONES III: SERIES DE TIEMPO NO ESTACIONARIAS

Abordaré simulaciones de Monte Carlo para conseguir caracterizar la distribución de estadísticos que, en su cálculo, involucran series no estacionarias. Específicamente, estudiaré las distribuciones asintóticas de estadísticos en pruebas de raíz unitaria (ver también, [Hamilton \(1994\)](#)).

TEST DICKEY-FULLER

Como punto de partida puede ser útil asumir que el proceso generador de datos (en adelante, PGD) es un random walk,

$$y_t = \alpha + y_{t-1} + u_t \quad (4.1)$$

donde u_t es un ruido blanco.

CASO 1: REGRESIÓN SIN CONSTANTE

Si imponemos la restricción que $\alpha = 0$ y nos enfocamos en testear el parámetro autoregresivo de la ecuación, la regresión estimada es

$$y_t = \rho y_{t-1} + u_t \quad (4.2)$$

$\mathbb{H}_0 : \rho = 1$ vs $\mathbb{H}_1 : |\rho| < 1$. Bajo la hipótesis nula es posible demostrar que

$$T(\hat{\rho}_T - 1) \xrightarrow{d} \frac{(1/2)\{[W(1)]^2 - 1\}}{\int_0^1 [W(r)]^2 dr} \quad (4.3)$$

$$t_T = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} \xrightarrow{d} \frac{(1/2)\{[W(1)]^2 - 1\}}{\left(\int_0^1 [W(r)]^2 dr\right)^{1/2}} \quad (4.4)$$

donde $W(\cdot)$ representa un movimiento Browniano estándar².

CASO 2: REGRESIÓN CON CONSTANTE

Supongamos que $\alpha = 0$, sin embargo, estimamos la ecuación

$$y_t = (1 - \rho)\alpha + \rho y_{t-1} + u_t \quad (4.5)$$

$\mathbb{H}_0 : \rho = 1$ vs $\mathbb{H}_1 : |\rho| < 1$. Bajo la hipótesis nula es posible demostrar que

$$T(\hat{\rho}_T - 1) \xrightarrow{d} \frac{(1/2)\{[W(1)]^2 - 1\} - W(1) \int_0^1 W(r) dr}{\int_0^1 [W(r)]^2 dr - [\int_0^1 W(r) dr]^2} \quad (4.6)$$

$$t_T = \frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}} \xrightarrow{d} \frac{(1/2)\{[W(1)]^2 - 1\} - W(1) \int_0^1 W(r) dr}{\left(\int_0^1 [W(r)]^2 dr - [\int_0^1 W(r) dr]^2\right)^{1/2}} \quad (4.7)$$

donde $W(\cdot)$ representa un movimiento Browniano estándar.

CASO 3: REGRESIÓN CON CONSTANTE Y TENDENCIA LINEAL

Si $\alpha \neq 0$, y estimamos

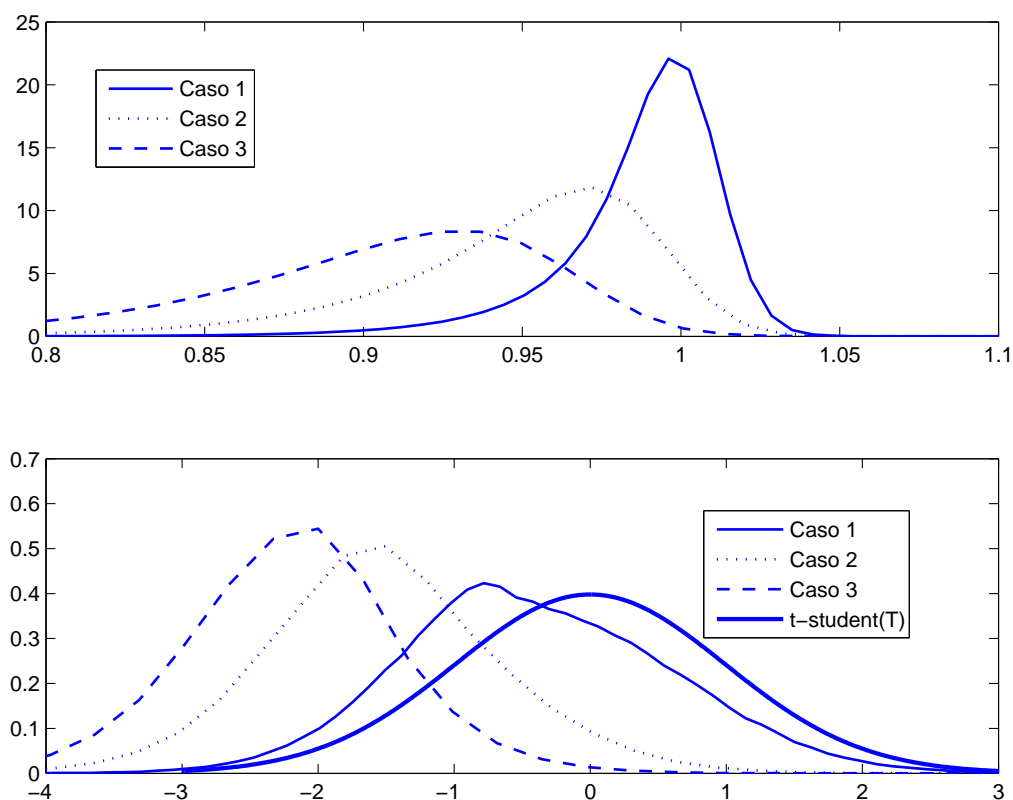
$$y_t = \alpha + \rho y_{t-1} + \delta t + u_t \quad (4.8)$$

$\mathbb{H}_0 : \delta = 0, \rho = 1$ vs $\mathbb{H}_1 : |\rho| < 1, \delta \neq 0$. Bajo la hipótesis nula $T(\hat{\rho}_T - 1)$ y t_T tienen una distribución algo más complicada.

En cualquiera de los tres casos podemos caracterizar estas distribuciones asintóticas por medio de una simulación de Monte Carlo, ver [Figure 6](#).

²Note que $[W(1)]^2 \sim \chi_1^2$, sin embargo la distribución asintótica de $T(\hat{\rho} - 1)$ no tiene forma funcional cerrada.

Figura 6: Distribución Dickey-Fuller



Notes: `MonteCarlo6.m` - Se genera una muestra de distribuciones $t\text{-student}(3)$. El número de replicaciones fue de $M = 10^6$.

REFERENCES

- Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.
- Hansen, B. (2017). *Econometrics (manuscript)*. Personal site.
- Winkelried, D. (2016). *Econometría de series de tiempo: enfoque de Monte Carlo*. Lima: Universidad del Pacífico.