# Change Detection in Autoregressive Time Series

Jakub Pečánka

*j.pecanka@few.vu.nl*

*Department of Probability and Mathematical Statistics, Charles University Prague*
*Department of Stochastics, Vrije Universiteit Amsterdam*

## 1 Autoregressive Model with Change Point

We assume to have a set of observations $Y_1, \dots, Y_n$ from an $\mathsf{AR}(p)$ process, with $p \in \mathbb{N}$ fixed. We define a recursive equation

$$Y_t - \mu \;=\; \varphi_1(Y_{t-1} - \mu) + \ldots + \varphi_p(Y_{t-p} - \mu) + \varepsilon_t, \quad p < t, \tag{1}$$

where $\{\varepsilon_t\}$ is a white noise sequence with variance $\sigma^2$ and $\mu, \sigma^2, \varphi_1, \dots, \varphi_p$ are fixed constants and $Y_1, \dots, Y_p$ are some initial values. This represents the null hypothesis. On the other hand, the equation corresponding to the alternative is

$$Y_t \;=\; \begin{cases} \varphi_0 + \varphi_1 Y_{t-1} + \ldots + \varphi_p Y_{t-p} + \varepsilon_t, & p < t \le k^*, \\ \psi_0 + \psi_1 Y_{t-1} + \ldots + \psi_p Y_{t-p} + \varepsilon_t, & k^* < t, \end{cases} \tag{2}$$

with $p < k^* \le n$. Thus, expressed simply, the pair of hypotheses in question is $H_0 : k^* = n$ and $H_1 : k^* < n$. More precisely, we define the *null hypothesis condition* by

*(H.1) The observations $Y_{p+1}, \dots, Y_n$ follow the model (1) with $k^* = n$; The observations $Y_1, \dots, Y_p$ are independent of the innovations $\varepsilon_{p+1}, \dots, \varepsilon_n$; The characteristic polynomial $\phi(z) = 1 - \varphi_1 z - \ldots - \varphi_p z^p$ has all roots outside the unit ball (*causality, stationary *solution).*

and the *alternative hypothesis condition* we formulate as

*(H.2) The observations $Y_{p+1}, \dots, Y_n$ follow the model (2) with $k^* = \lfloor \tau n \rfloor$, for some fixed $\tau \in (0,1)$; The observations $Y_1, \dots, Y_p$ are independent of the innovations $\varepsilon_{p+1}, \dots, \varepsilon_n$; The polynomials $\phi_1(z) = 1 - \varphi_1 z - \ldots - \varphi_p z^p$ and $\phi_2(z) = 1 - (\varphi_1 + \delta_1)z - \ldots - (\varphi_p + \delta_p)z^p$ have all roots outside the unit ball (*causality*).*

In both cases the initial observations $Y_1, \dots, Y_p$ in the recursive equation (1) are generated from $\{\varepsilon_t\}$ according to

*(H.3) The vector of observations $\boldsymbol{x}_{p+1} = (Y_p, \dots, Y_1)'$ satisfies $\boldsymbol{x}_{p+1} - \mu = \sum_{j=0}^{\infty} \mathbf{B}^j \boldsymbol{\varepsilon}_{p-j}$, where*

$$\mathbf{B} \;=\; \begin{pmatrix} \varphi_1, \dots, \varphi_p \\ \mathbf{I}_{p-1} \quad \mathbf{0} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon}_k = (\varepsilon_k, 0, \dots, 0)',$$

*with $\mathbf{I}_{p-1}$ denoting the $(p-1)$-dimensional unit matrix and $\mathbf{0}$ denoting the $(p-1)$-dimensional zero vector.*

For simplicity we also assume that the $\mathsf{AR}$ process is centered, i.e. $\mu = 0$.

## 2 Theoretical Results

In [2] the asymptotic behavior of several testing statistics based on *partial sums of weighted residuals* is investigated. We define

$$T_n \;=\; \frac{1}{\widehat{\sigma}_n^2} \max_{p<k<n} \left\{ \boldsymbol{S}_k' \boldsymbol{C}_k^{-1} \boldsymbol{C}_n (\boldsymbol{C}_k^0)^{-1} \boldsymbol{S}_k \right\},$$

$$T_n(\epsilon) \;=\; \frac{1}{\widehat{\sigma}_n^2} \max_{n\epsilon \le k \le n(1-\epsilon)} \left\{ \boldsymbol{S}_k' \boldsymbol{C}_k^{-1} \boldsymbol{C}_n (\boldsymbol{C}_k^0)^{-1} \boldsymbol{S}_k \right\},$$

$$T_n(q,\epsilon) \;=\; \frac{1}{\widehat{\sigma}_n^2} \sup_{t \in [\epsilon, 1-\epsilon]} \left\{ \frac{\boldsymbol{S}_{\lfloor (n+1)t \rfloor}' \boldsymbol{C}_n^{-1} \boldsymbol{S}_{\lfloor (n+1)t \rfloor}}{q^2(t)} \right\},$$

$$T_n(q) \;=\; \frac{1}{\widehat{\sigma}_n^2} \sup_{t \in (0,1)} \left\{ \frac{\boldsymbol{S}_{\lfloor (n+1)t \rfloor}' \boldsymbol{C}_n^{-1} \boldsymbol{S}_{\lfloor (n+1)t \rfloor}}{q^2(t)} \right\},$$

where $\epsilon \in (0, \frac{1}{2})$, $q(t)$ is a positive weight function on $(0,1)$ and $\widehat{\sigma}_n^2$ is a suitable estimator of the variance parameter $\sigma^2$ and for $k = p+1, \dots, n$

$$\boldsymbol{S}_k \;=\; \sum_{t=p+1}^{k} \boldsymbol{x}_t (Y_t - \boldsymbol{x}_t' \boldsymbol{\varphi}_n), \qquad \boldsymbol{\varphi}_n \;=\; \boldsymbol{C}_n^{-1} \sum_{t=p+1}^{n} \boldsymbol{x}_t Y_t,$$

$$\boldsymbol{C}_k \;=\; \sum_{t=p+1}^{k} \boldsymbol{x}_t \boldsymbol{x}_t', \qquad \boldsymbol{C}_k^0 \;=\; \sum_{t=k+1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t' \;=\; \boldsymbol{C}_n - \boldsymbol{C}_k.$$

An estimator of $\sigma^2$ which has good properties under $H_0$ and $H_1$ is for example

$$\widehat{\sigma}_n^2(\widehat{k}) \;=\; \frac{1}{n-p} \Big[ \sum_{i=p+1}^{\widehat{k}} \Big( Y_i - \boldsymbol{x}' \boldsymbol{\beta}_{\widehat{k}}^1 \Big)^2 + \sum_{i=\widehat{k}+1}^{n} \Big( Y_i - \boldsymbol{x}' \boldsymbol{\beta}_{\widehat{k}}^2 \Big)^2 \Big],$$

where $\widehat{k}$ is a suitable estimator of the change point $k^*$ and $\boldsymbol{\beta}_{\widehat{k}}^1$, $\boldsymbol{\beta}_{\widehat{k}}^2$ are the least squares estimators of $\boldsymbol{\beta}$ based on $Y_1, \dots, Y_{\widehat{k}}$ and $Y_{\widehat{k}}, \dots, Y_n$, respectively. Suitable estimators of the change point $k^*$ are for example

$$\widehat{k}_1 = \min \Big\{ k \in (p,n) : \boldsymbol{S}_k' \boldsymbol{C}_k^{-1} \boldsymbol{C}_n (\boldsymbol{C}_k^0)^{-1} \boldsymbol{S}_k = \max_{p<h<n} \big( \boldsymbol{S}_h' \boldsymbol{C}_h^{-1} \boldsymbol{C}_n (\boldsymbol{C}_h^0)^{-1} \boldsymbol{S}_h \big) \Big\},$$

$$\widehat{k}_2 = \min \Big\{ k \in (p,n) : \boldsymbol{S}_k' \boldsymbol{C}_n^{-1} \boldsymbol{S}_k = \max_{p<h<n} \big( \boldsymbol{S}_h' \boldsymbol{C}_n^{-1} \boldsymbol{S}_h \big) \Big\}.$$

### 2.1 Asymptotic Distributions

By $\Gamma(x)$ we denote the *gamma function*. The following theorems utilize functions $A(x) = \sqrt{2 \log x}$ and $D_p(x) = 2 \log x + \frac{p}{2} \log \log x - \log \Gamma(p/2)$ and the following two assumptions.

*(I.1) $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ are i.i.d. centered random variables with positive variance $\sigma^2$ and finite moment $\mathsf{E}|\varepsilon_t|^{4+\nu}$, for some $\nu > 0$, with density $f$ such that*

$$\sup_{a \in \mathbb{R}} \frac{1}{|a|} \int_{\mathbb{R}} \big| f(x+a) - f(x) \big| \, \mathrm{d}x \;<\; \infty.$$

*(I.2) $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ are i.i.d. centered random variables with positive variance $\sigma^2$ and finite fourth moment $m_4 = \mathsf{E}|\varepsilon_t|^4$.*

The following theorems are either proved or referenced in [4]. We denote by $\|\cdot\|$ the $p$-dimensional Euclidean norm and $\big\{\boldsymbol{B}(t),\, t \in [0,1]\big\}$ is a $p$-dimensional standard Brownian bridge process with independent components, i.e. $\boldsymbol{B}(t) = \big(B_1(t), \dots, B_p(t)\big)'$ with $\big\{B_i(t),\, t \in [0,1]\big\}$ being a standard Brownian bridge for all $i = 1, \dots, p$.

**Theorem 1 (Asymptotics of $T_n$).** *Let assumptions (H.1) and (H.3) be satisfied and let $\widehat{\sigma}^2$ be an estimator of $\sigma^2$ such that $\widehat{\sigma}^2 - \sigma = o_{\mathsf{P}}((\log \log n)^{-1})$ as $n \to \infty$. Furthermore, let the white noise sequence $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ satisfy (I.1). Then,*

$$\lim_{n \to \infty} \mathsf{P}\Big( A(\log n) T_n^{1/2} - D_p(\log n) \le t \Big) \;=\; \exp\big\{ -2e^{-t} \big\}, \qquad t \in \mathbb{R}.$$

*It also holds for all $t \in \mathbb{R}$*

$$\lim_{n \to \infty} \mathsf{P}\Big( \big[ A^2(\log n) T_n - D_p^2(\log n) \big] \,/\, D_p(\log n) \le t \Big) \;=\; \exp\big\{ -2e^{-t/2} \big\}.$$

**Theorem 2 (Asymptotics of $T_n(\epsilon)$, $T_n(q,\epsilon)$ and $T_n(q)$).** *Let assumptions (H.1) and (H.3) be satisfied and let $\widehat{\sigma}^2$ be a consistent estimator of $\sigma^2$. Furthermore, let $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ be such that it satisfies (I.2). Then, for any $\epsilon \in (0, \frac{1}{2})$, it holds*

$$\sqrt{T_n(\epsilon)} \;\xrightarrow[n \to \infty]{\mathcal{D}}\; \sup_{t \in [\epsilon, 1-\epsilon]} \frac{\|\boldsymbol{B}(t)\|}{\sqrt{t(1-t)}}.$$

*If, in addition, $q(t)$ is a function that satisfies $\inf\{q(t);\, t \in (\epsilon, 1-\epsilon)\} > 0$, then for any $\epsilon \in (0,1)$ it holds*

$$\sqrt{T_n(q,\epsilon)} \;\xrightarrow[n \to \infty]{\mathcal{D}}\; \sup_{t \in [\epsilon, 1-\epsilon]} \frac{\|\boldsymbol{B}(t)\|}{q(t)}.$$

*If, in addition, a function $q_\beta(t)$ is defined as $q_\beta(t) = \big(t(1-t)\big)^\beta$, $t \in (0,1)$, with $\beta \in \big[0, \frac{1}{2}\big)$, then also*

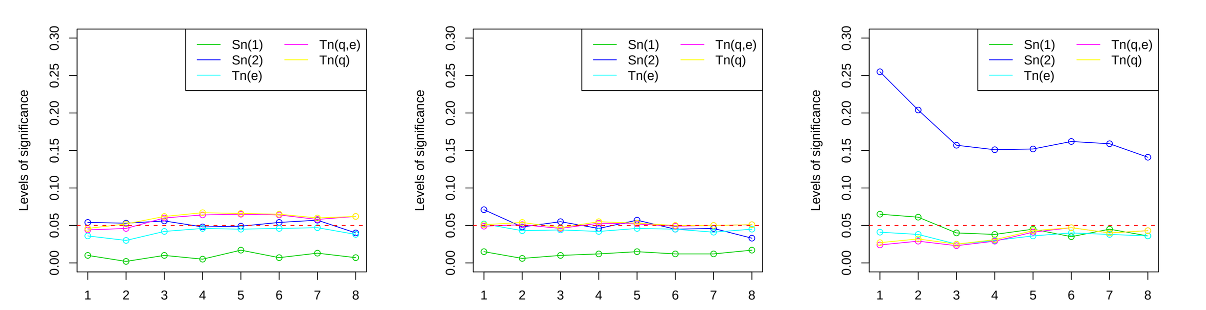
$$\sqrt{T_n(q_\beta)} \;\xrightarrow[n \to \infty]{\mathcal{D}}\; \sup_{t \in (0,1)} \frac{\|\boldsymbol{B}(t)\|}{q_\beta(t)}.$$

## 3 Simulations

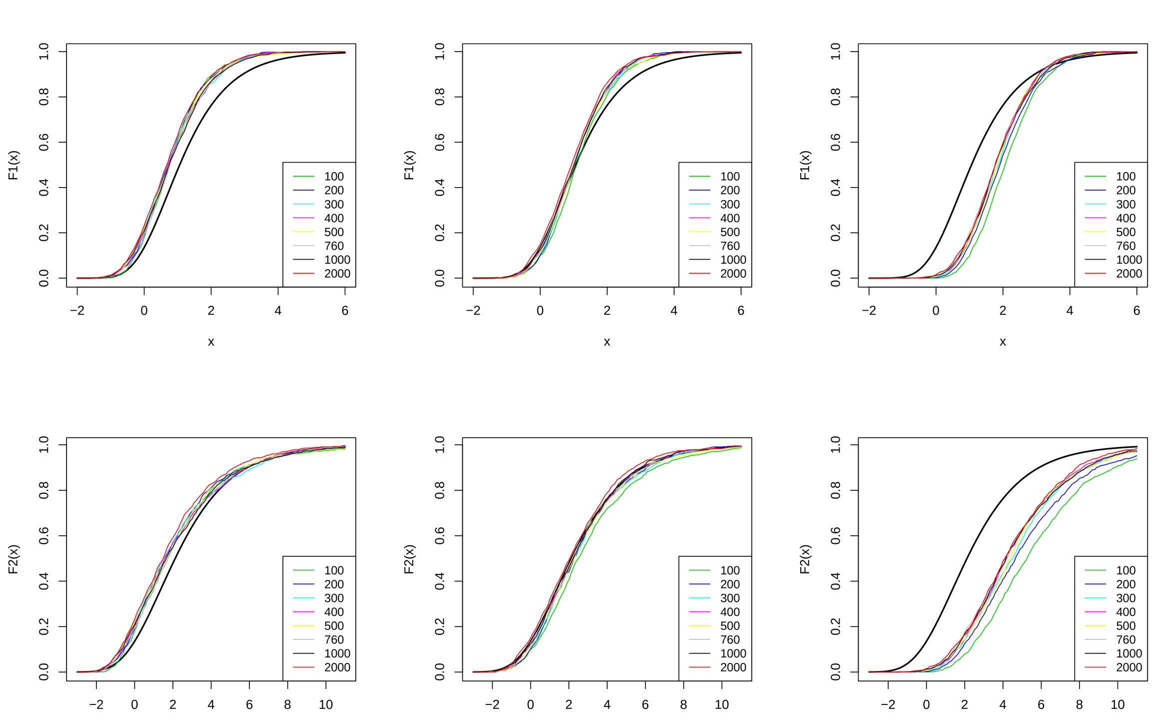
In [4] we simulated large number of $\mathsf{AR}$ processes of 3 different orders of autoregression $p_1 = 1$, $p_2 = 3$ and $p_3 = 5$. For each order we simulated 1000 repetitions of *AR* processes of lengths $n = 100, 200, 300, 400, 500, 760, 1000, 2000$. We performed these simulations under three different distributions of the innovations $\{\varepsilon_t\}$: the standard normal distribution, $t_6$-distribution and the uniform distribution on $[-1, 1]$. For the lack of available space we present here only the normal case. We should like to mention that the results were very similar for all three distributions. We also simulated processes with fixed and randomly chosen coefficients. We only present the case of fixed coefficients, which were fixed at $\varphi_1 = 0.5$ for $p = 1$, and $\varphi_1 = 0.5$, $\varphi_2 = -0.4$, $\varphi_3 = 0.1$ for $p = 3$, and $\varphi_1 = 1$, $\varphi_2 = -0.1$, $\varphi_3 = 0.3$, $\varphi_2 = -0.4$, $\varphi_5 = 0.1$ for $p = 5$. These coefficients were used to generate AR with no change point and the before change portion of AR processes with change point. On the other hand, the the coefficients used to simulate the after change point portion of the data were fixed at the values $\psi_1 = 0.8$ for $p = 1$, and $\psi_1 = 1$, $\psi_2 = -1$, $\psi_3 = 0.1$ for $p = 3$, and $\psi_1 = -0.5$, $\psi_2 = 0.5$, $\psi_3 = 0.3$, $\psi_2 = -0.4$, $\psi_5 = 0.1$ for $p = 5$. To get a broader picture we also used different change points. The location of the change point for each choice of $p$ and each $n$ are defined using $\tau$ as $\tau_1 = 0.25$, $\tau_2 = 0.5$ and $\tau_3 = 0.75$, respectively.

### 3.1 Behavior of Statistics under Null Hypothesis

From these graphs it is apparent that while most of the statistics behaved in accordance with theory, the attained levels of significance were not always near the chosen value of $\alpha = 5\%$ even for large sample sizes. For example, a lower level of significance was attained by the test statistic $S_n(1) = A_n(\log n) T_n^{1/2} - D_p(\log n)$ for the two smaller choices of order $p = 1, 3$, with the results for $p = 3$ being slightly more favorable. On the other hand, already for the case of $p = 5$ the attained level of significance for a test based on $S_n(1)$ is quite close to the asymptotic level of 5%. However, in the case of $S_n(2) = (A_n^2(\log n) T_n - D_p^2(\log n))/D_p(\log n)$ for $p = 5$ the convergence is particularly slow and the attained level of significance even for $n = 2000$ is much higher than 5%. We also conclude that the tests based on $T_n(\epsilon)$, $T_n(q_0, \epsilon)$ and $T_n(q_0)$ have the attained level quite close to the theoretical value of 5%.

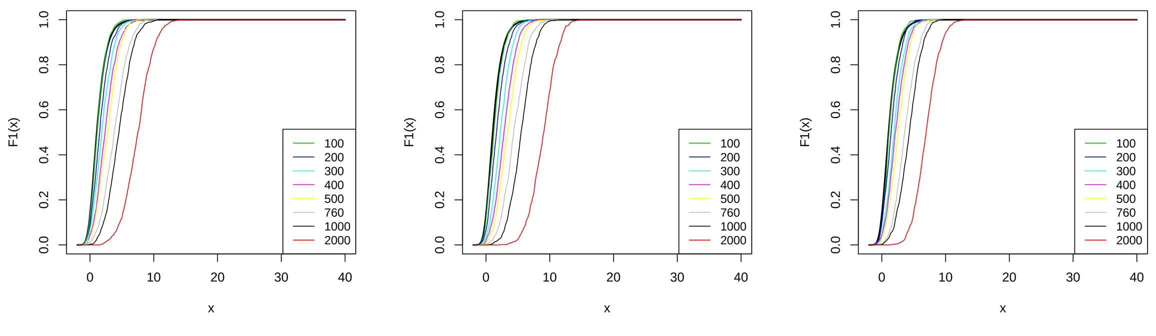


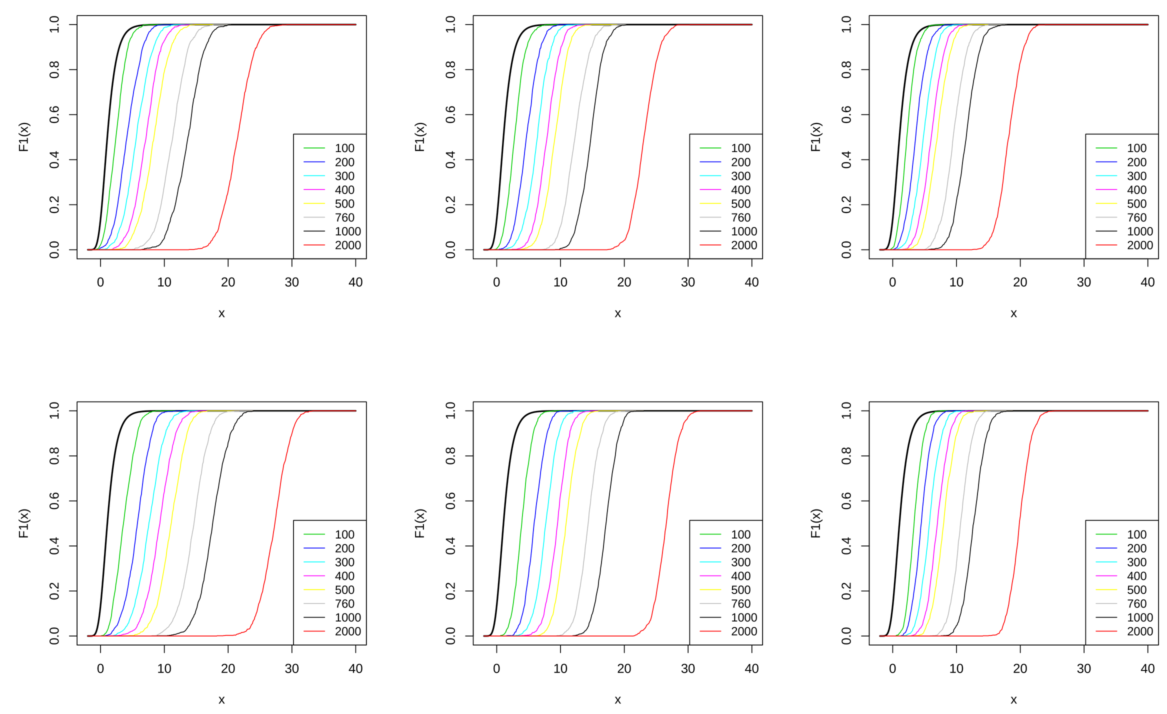

We illustrate the behavior of $S_n(1)$ and $S_n(2)$ by the convergence of their distribution functions. The upper row corresponds to $S_n(1)$, while the lower row corresponds to $S_n(2)$, both with orders increasing from left to right ($p = 1, 3, 5$).



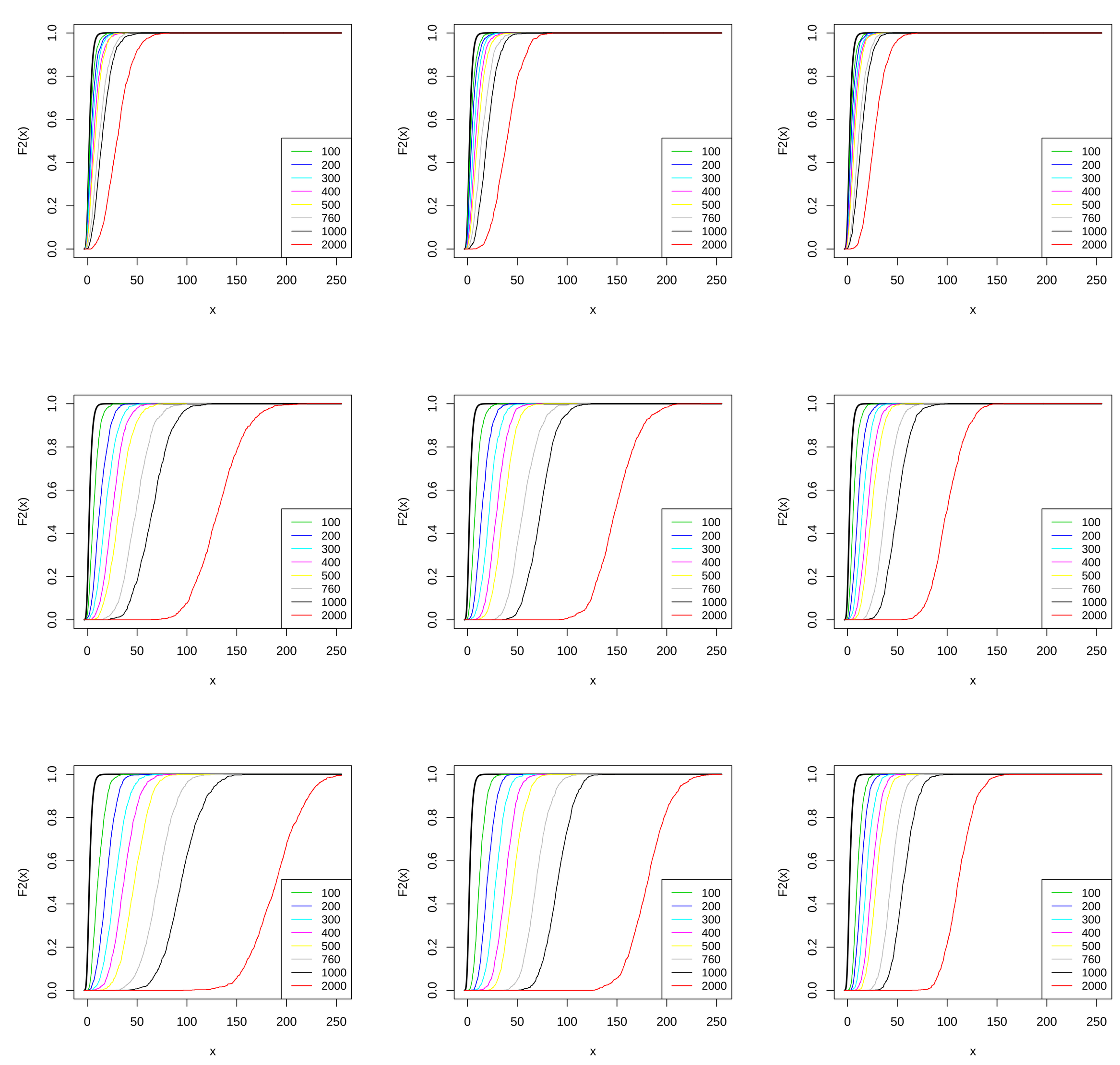

### 3.2 Detection of Change Point under Alternatives

The tests based on statistics statistics $T_n$, $T_n(\epsilon)$, $T_n(q_0, \epsilon)$ and $T_n(q)$ with $q = \big(t(1-t)\big)^\beta$, $t \in (0,1)$, $\beta \in \big[0, \frac{1}{2}\big)$ are consistent (see [2] or [4]). To illustrate this we again simulated $N = 1000$ realizations of $\mathsf{AR}$ processes with a change of autoregression coefficients at time $k^* = \lfloor \tau n \rfloor$. We present here the results for $S_n(1)$ and $S_n(2)$. The distribution of $S_n(1)$ behaved as shown on the following graphs (increasing $\tau = 0.25, 0.5, 0.75$ from left to right, increasing order $p = 1, 3, 5$ from top to bottom).

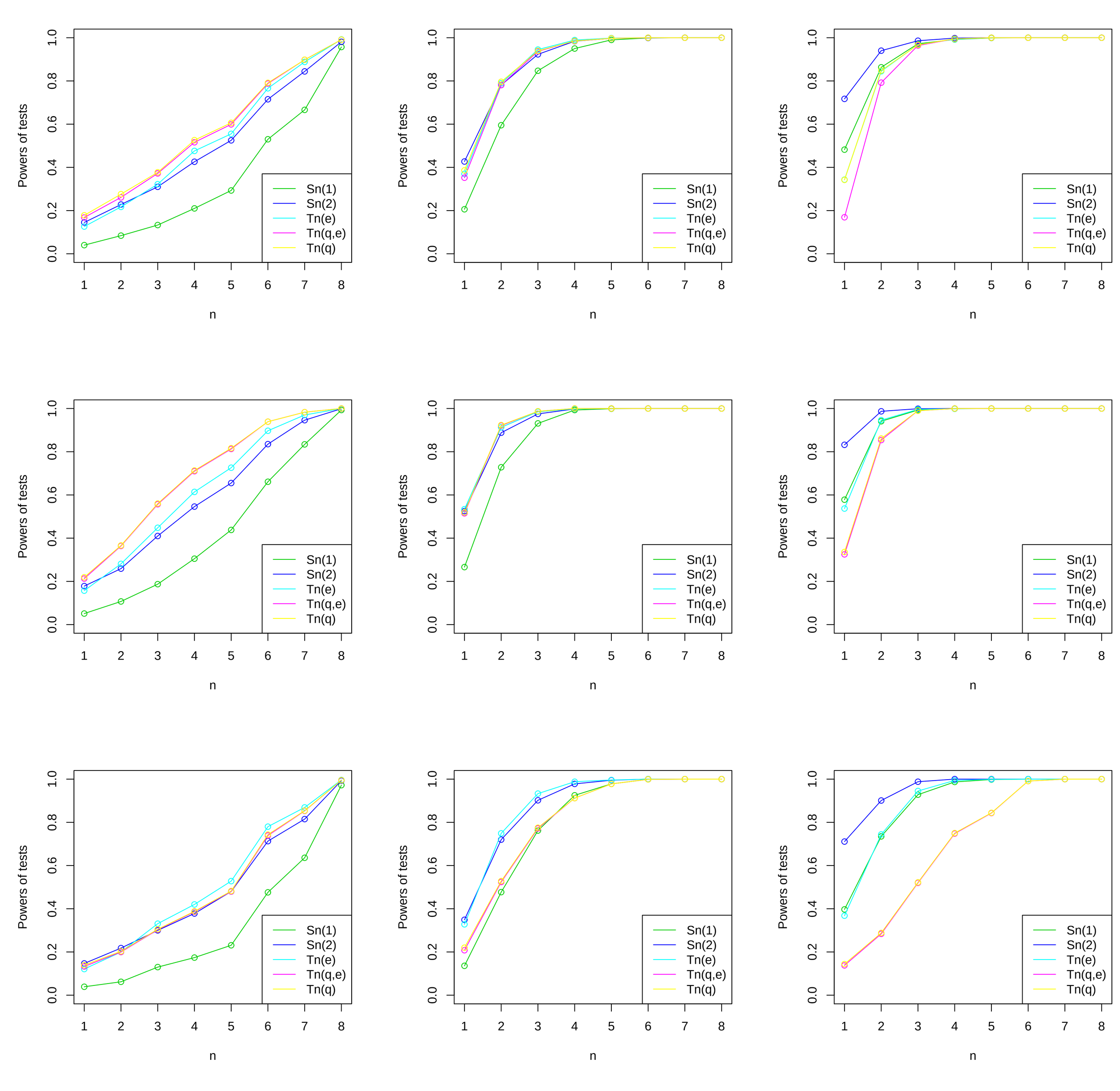



An illustration of the behavior of distribution functions of $S_n(2)$ with the same ordering as in the case of $S_n(1)$ follows.




In our graphs we can observe an apparent shifting of the empirical distribution functions of $S_n(1)$ and $S_n(2)$ towards infinity for all orders $p = 1, 3, 5$ as $n$ increases. Note, that the convergence seems to be faster for higher orders of autoregression. Based on the results in [4] we can also conclude that the convergence of $S_n(1)$ and $S_n(2)$ to infinity appears faster for $\tau = 0.5$ than for either of other two choices $\tau = 0.25$ and $\tau = 0.75$. Finally, we illustrate consistency of all of the presented test statistics $S_n(1)$, $S_n(2)$, $T_n(\epsilon)$, $T_n(q, \epsilon)$ and $T_n(q)$ in terms of empirical power for $p = 1$, $p = 3$ and $p = 5$ with $\tau = 0.5$ and $\epsilon = 0.1$. The simulated results seem to be quite satisfactory.




## References

[1] Davis, R., Huang, D., and Yao, Y.C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *Annals of Statistics*, **23**(11):282–304.

[2] Hušková, M., Prášková, Z., and Steinebach, J. (2007a). On the detection of changes in autoregressive time series, i. asymptotics. *Journal of Statistical Planning and Inference*, **137**(4):1243–1259.

[3] Hušková, M., Kirch, C., Prášková, Z., and Steinebach, J. (2007b). On the detection of changes in autoregressive time series, ii. resampling procedures. *Journal of Statistical Planning and Intference*, **138**(6):1697–1721.

[4] Pečánka, J. (2008). Change Detection in Autoregressive Time Series *Master Thesis*, VU University Amsterdam.