

# Using the Empirical Probability Integral Transformation to Construct a Nonparametric CUSUM Algorithm



Daniel R. Jeske  
University of California, Riverside  
Department of Statistics

Joint work with V. Montes de Oca, W. Bischoff and M. Marvasti

QPRC – Quality Productivity and Research Conference  
June 4, 2009

# Motivation

---

- ❑ An anomalous event is categorized as a **departure from normal operating conditions** due to either an existing or impending failure condition.
- ❑ In network surveillance it is important to **detect anomalous activity as quickly as possible**, and ideally before a major problem develops.
- ❑ There are **1,000's of metrics** to be monitored, therefore automated network management is important.
- ❑ **Example:** By-the-minute measurements of traffic (e.g., bytes or packets) carried over a network link, or passing through a network node.

# ALIVE™

## by Integrien Corporation

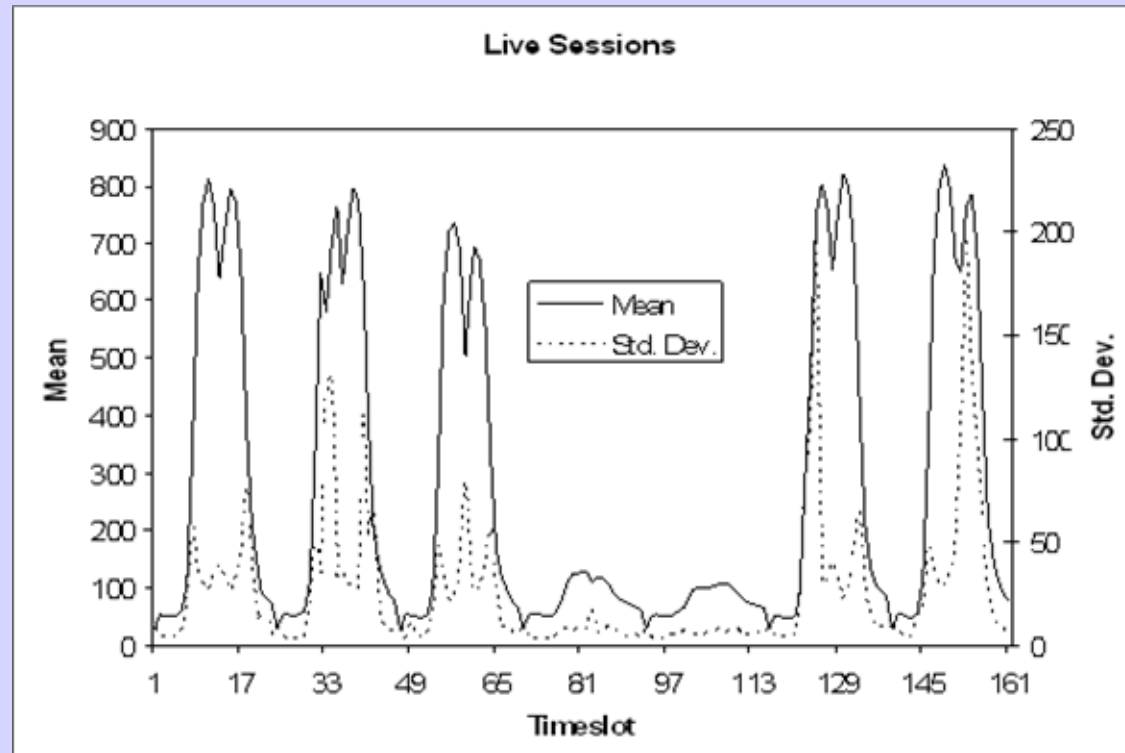
- Alive is a secure network monitoring system running
- Alive integrates monitoring, advanced analytics, root cause analysis, and recovery functions for the full scope of IT system components.

- Through **advanced statistical technology**, Alive detects deviation from normal operations and prevents problems from impacting the business.

Alive Dashboard



# Live Sessions Data



Means and standard deviations for Live sessions obtained from 12 weeks of screened historical data.

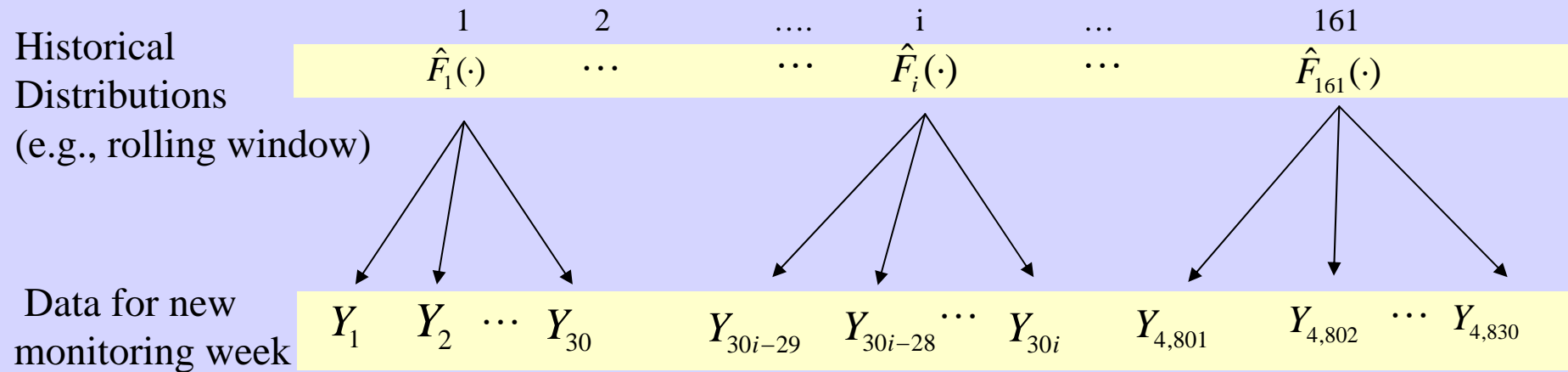
# Requirements

---

- 1) **Versatility** - the need for a **nonparametric** technique so that a wide variety of metrics can be included in the monitoring process
- 2) **Adaptability** - the need to handle **time varying distributions** for the metrics that reflect natural cycles in a work week
- 3) **Efficiency** - the need to be **computationally efficient** with the massive amounts of data that are available for processing

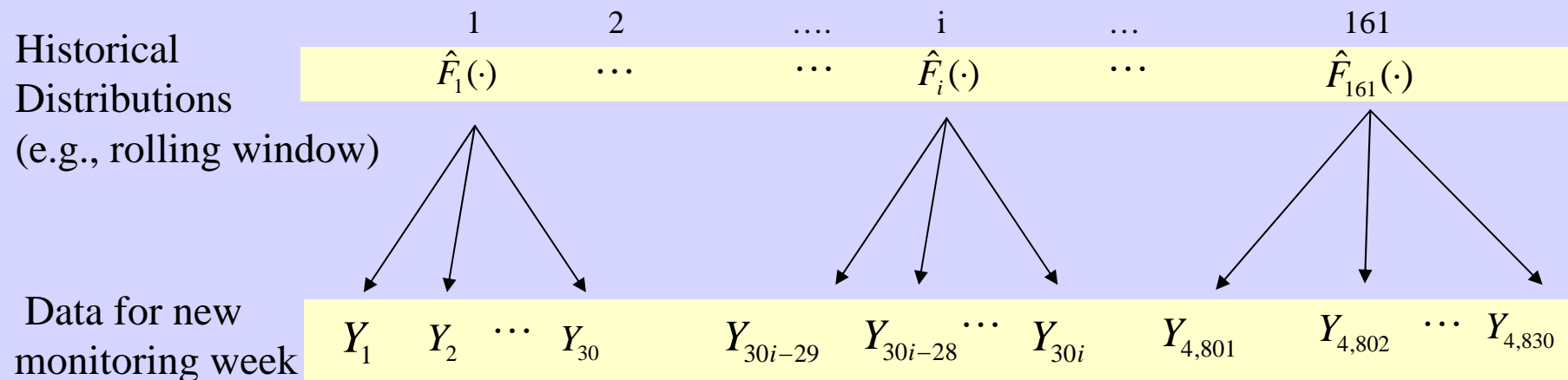
# Structured Timeslot Non-Stationarity

161 Hours of the Week



# Independence Assumptions

## 161 Hours of the Week



- ❑ Observations *within* timeslots (hour) are identically and independently distributed
- ❑ Observations *between* timeslots are independent but *not* identically distributed

# Improving Plausibility of Independence Assumption

$$X_{j1} = Y_{j1}$$

$$X_{j2} = \frac{Y_{j2} - \rho_j Y_{j1}}{\sqrt{1 - \rho_j^2}} + \mu_j \left( 1 - \frac{1 - \rho_j}{\sqrt{1 - \rho_j^2}} \right)$$

⋮

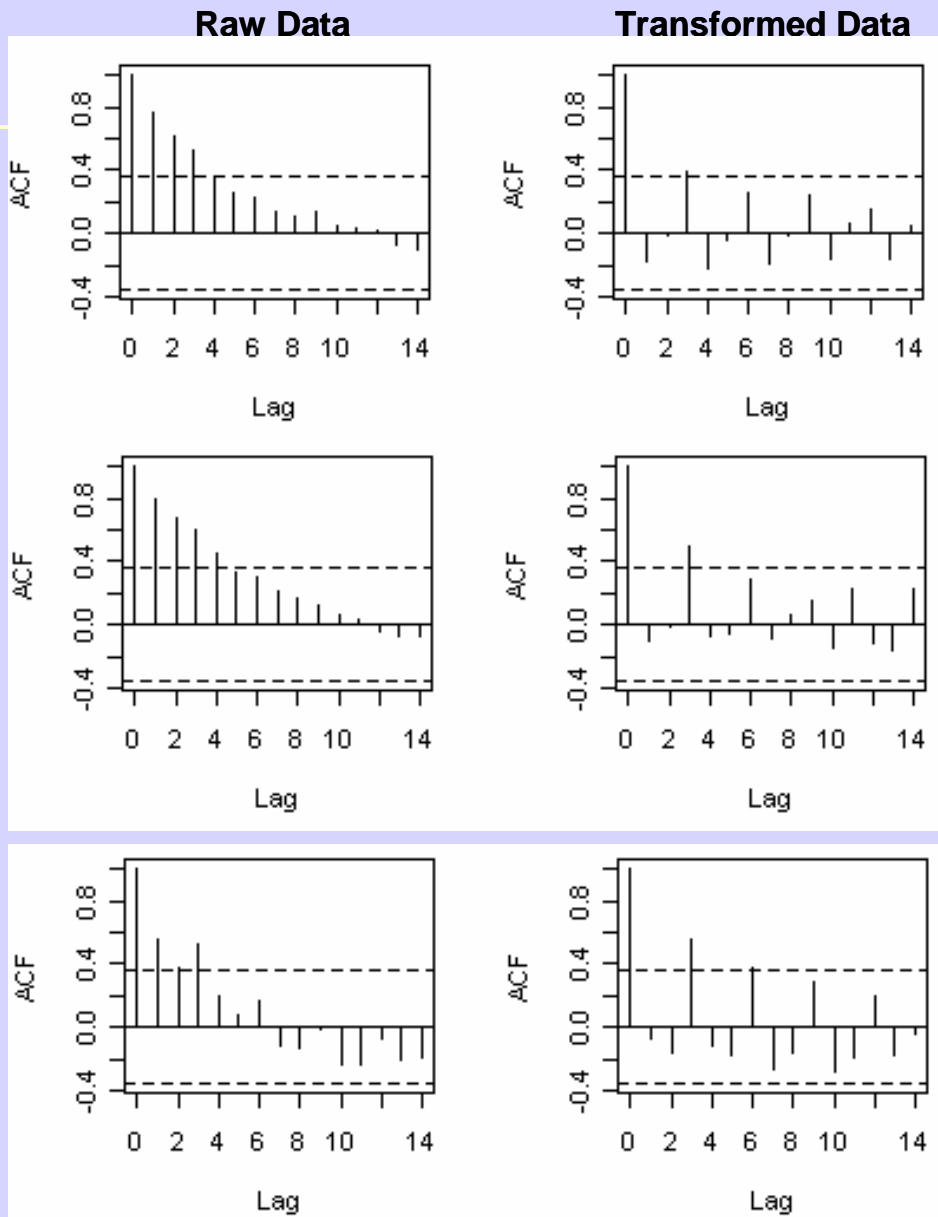
$$X_{jr} = \frac{Y_{jr} - \rho_j Y_{j,r-1}}{\sqrt{1 - \rho_j^2}} + \mu_j \left( 1 - \frac{1 - \rho_j}{\sqrt{1 - \rho_j^2}} \right)$$

Under AR(1) model these observations are uncorrelated with common mean  $\mu_j$  and common variance  $\sigma_j^2$

Other (application-dependent) transformations could be used.

Our starting point is the transformed data.

# Effectiveness of Transformation For Our Application



Autocorrelation Functions of Raw Data (Column 1) and Corresponding Transformed Data (Column 2) for Three Timeslot-Week Combinations (Rows)

# Recent Work in Network Surveillance Applications

---

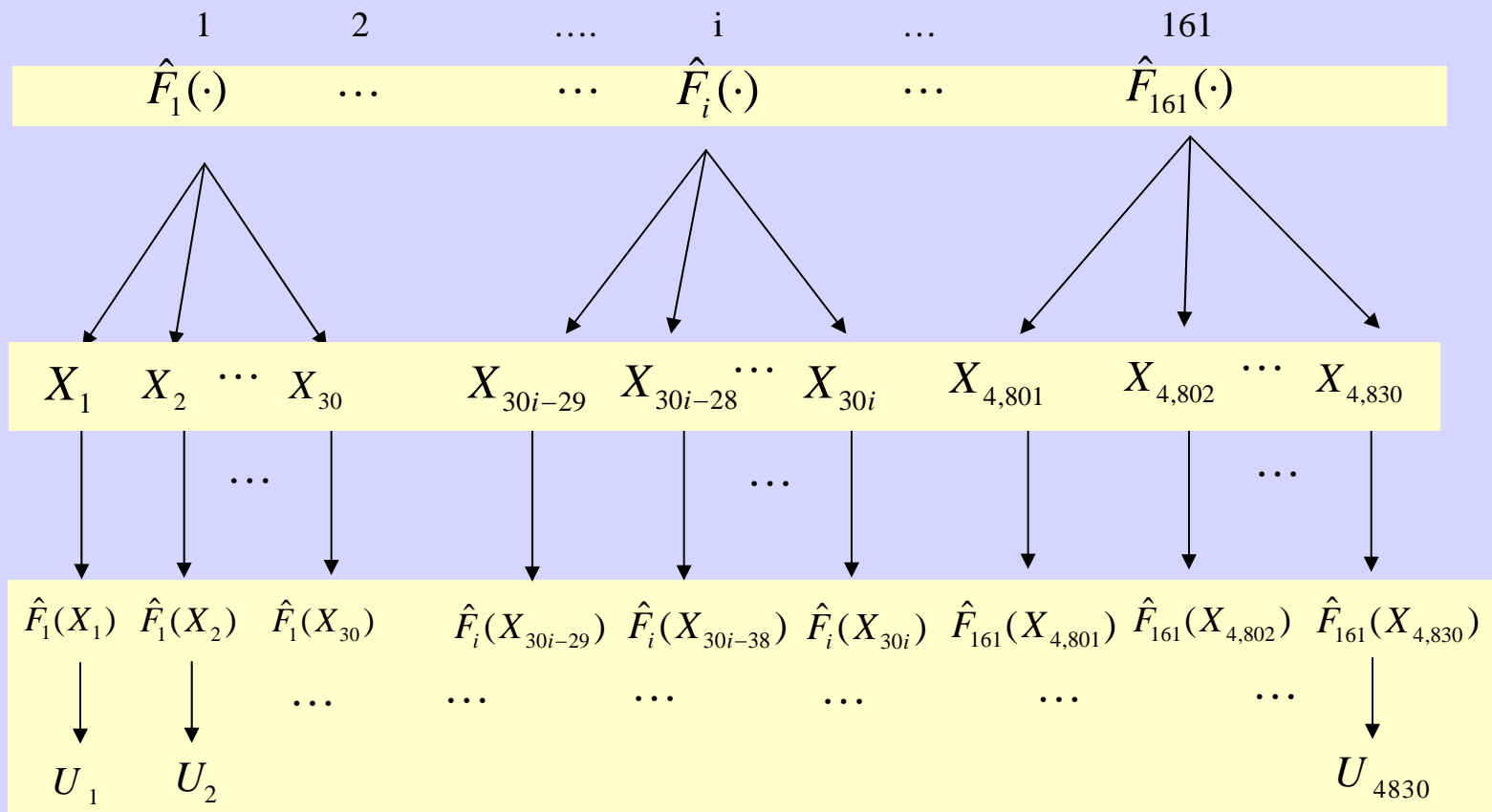
1. Hajji, H. (2005), "Statistical Analysis of Network Traffic for Adaptive Faults Detection," *IEEE Transactions on Neural Networks*, 16, 1053-1063.

*Random regime switching assumption is not suitable for structured timeslot context*

2. Kim, S., Alexopoulos, C., Tsui, K., and Wilson, J. R. (2007), "A Distribution-free Tabular CUSUM Chart for Autocorrelated Data," *IIE Transactions*, 39, 317-330.

*Applies to stationary contexts (which would be one timeslot for us) with known covariance structure*

# Empirical Probability Integral Transformation



# Discrete Probability Integral Transformation

Conditional on the historical data, the  $U$  values for a given timeslot with  $m$  historical values take on  $m+1$  values with probabilities:

$$\begin{pmatrix} 0 & 1/m & \dots & (m-1)/m & 1 \\ F(x_{1:m}) & F(x_{2:m}) - F(x_{1:m}) & \dots & F(x_{m:m}) - F(x_{m-1:m}) & 1 - F(x_{m:m}) \end{pmatrix}$$

Probabilities are spacings of  $m$  observations from a  $U(0,1)$  distribution, so conditional distribution of  $U$ 's is approximately

$$\begin{pmatrix} 0 & 1/m & \dots & (m-1)/m & 1 \\ 1/(m+1) & 1/(m+1) & \dots & 1/(m+1) & 1/(m+1) \end{pmatrix}$$

which depends on this historical data only through sample size  $m$

# Transformed Cusum (TC)

## Cusum Algorithm

$$S_n^+ = \max \left[ 0, S_{n-1}^+ + (F_{\tau_n}(X_n) - \alpha) \right], \quad n = 1, \dots, N, S_0^+ = 0$$

$$S_n^- = \max \left[ 0, S_{n-1}^- + (1 - \alpha - F_{\tau_n}(X_n)) \right], \quad n = 1, \dots, N, S_0^- = 0$$

## Stopping Rule

$S_n^+ > H$  or  $S_n^- > H$  where

$$P_0(S_n^+ > H \text{ or } S_n^- > H) = \gamma$$

$H$  is obtained by simulation

$\gamma$  is the probability of a false alarm

Threshold Values (H)

Weeks of Historical Data	Probability of False Alarm		
	0.01	0.05	0.1
6 (m=180)	0.3611	0.3167	0.2944
12 (m=360)	0.3583	0.3139	0.2917
24 (m=720)	0.3542	0.3083	0.2861
Infinity	0.3500	0.3044	0.2828

\* Infinity corresponds to sampling from  
Continuous Uniform Distribution

# Brownian Motion Approximation

---

Let  $Z_1, Z_2, \dots, Z_N$  be independent, identically distributed random variables with mean 0 and variance 1. Define

$$W_n = \frac{1}{\sqrt{N}} \sum_{i=1}^n Z_i, \quad n = 1, \dots, N$$

Then for large  $N$  we have  $P\left(\max_{1 \leq n \leq N} W_n > H\right) \doteq \gamma$  where  $H = \Phi^{-1}(1 - \gamma/2)$

# Brownian Motion Cusum (BMC)

## Cusum Algorithm

$$U_i = \max \left[ 0, (F_{\tau_i}(X_i) - \alpha) \right] \quad V_i = \max \left[ 0, (1 - \alpha - F_{\tau_i}(X_i)) \right], \quad i = 1, \dots, N$$

$$S_n^+ = \frac{1}{\sqrt{N}} \sum_{i=1}^n \left( \frac{U_i - E(U_i)}{\sqrt{\text{Var}(U_i)}} \right) \quad S_n^- = \frac{1}{\sqrt{N}} \sum_{i=1}^n \left( \frac{V_i - E(V_i)}{\sqrt{\text{Var}(V_i)}} \right), \quad n = 1, \dots, N$$

means and standard deviations are computed as on next slide....

Stopping Rule  $S_n^+ > H$  or  $S_n^- > H$  where  $H = \Phi^{-1}(1 - \gamma/4)$

# Brownian Motion Cusum (BMC)

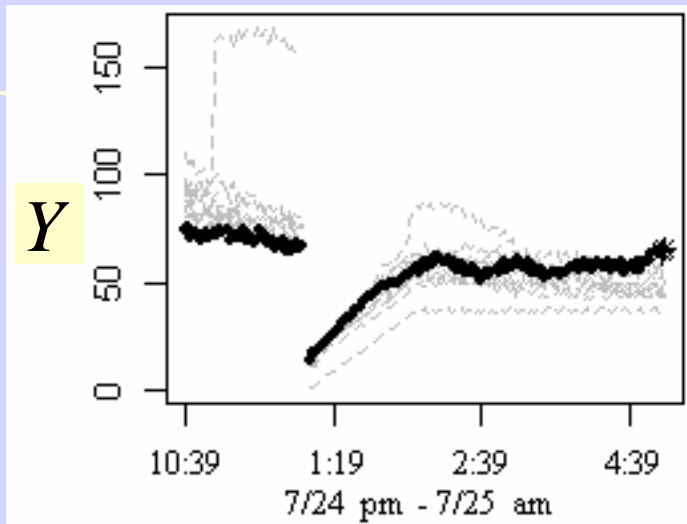
$$E_0(U_l) = E_0(V_l) \doteq \frac{(n_{\tau_l} - \lceil \alpha n_{\tau_l} \rceil + 1)(n_{\tau_l} + \lceil \alpha n_{\tau_l} \rceil - 2\alpha n_{\tau_l})}{2n_{\tau_l}(n_{\tau_l} + 1)}$$

$$V_0(U_l) = V_0(V_l)$$

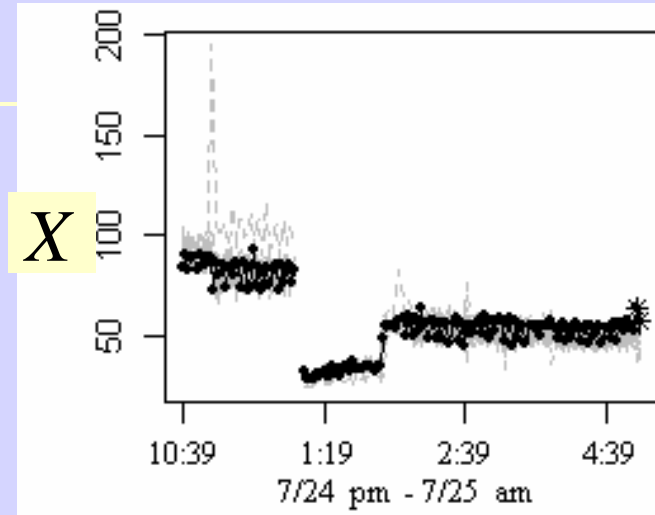
$$\begin{aligned} &\doteq \frac{2(n_{\tau_l} - \lceil \alpha n_{\tau_l} \rceil)(n_{\tau_l}^2 + n_{\tau_l} \lceil \alpha n_{\tau_l} \rceil + \lceil \alpha n_{\tau_l} \rceil^2) + 3(n_{\tau_l}^2 + \lceil \alpha n_{\tau_l} \rceil^2) + (n_{\tau_l} - \lceil \alpha n_{\tau_l} \rceil)}{6n_{\tau_l}^2(n_{\tau_l} + 1)} \\ &\quad + \frac{\alpha(n_{\tau_l} - \lceil \alpha n_{\tau_l} \rceil + 1)(\alpha n_{\tau_l} - n_{\tau_l} - \lceil \alpha n_{\tau_l} \rceil)}{m(m+1)} \\ &\quad - \left[ \frac{(n_{\tau_l} - \lceil \alpha n_{\tau_l} \rceil + 1)(n_{\tau_l} + \lceil \alpha n_{\tau_l} \rceil - 2\alpha n_{\tau_l})}{2n_{\tau_l}(n_{\tau_l} + 1)} \right]^2 \end{aligned}$$

# Illustration of TC and BMC

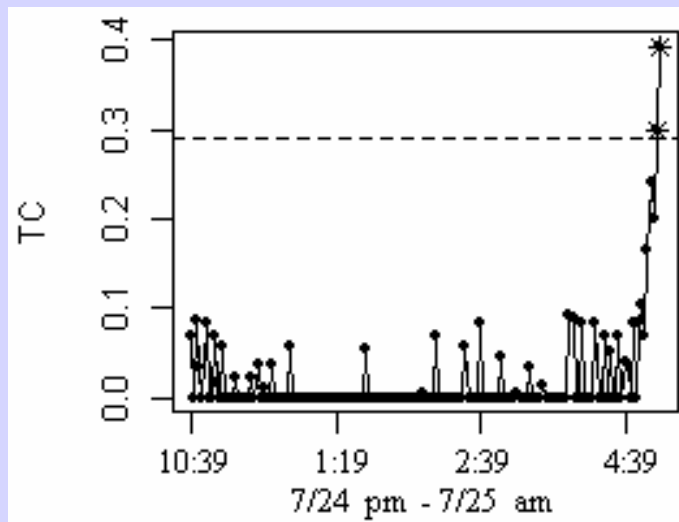
Raw Data



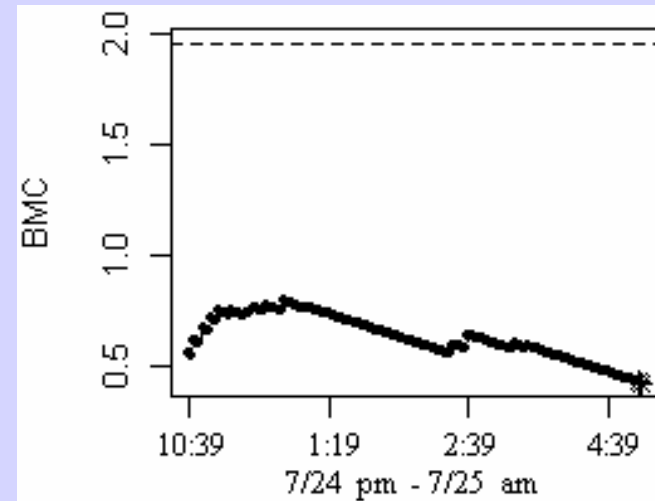
Transformed Data



TC



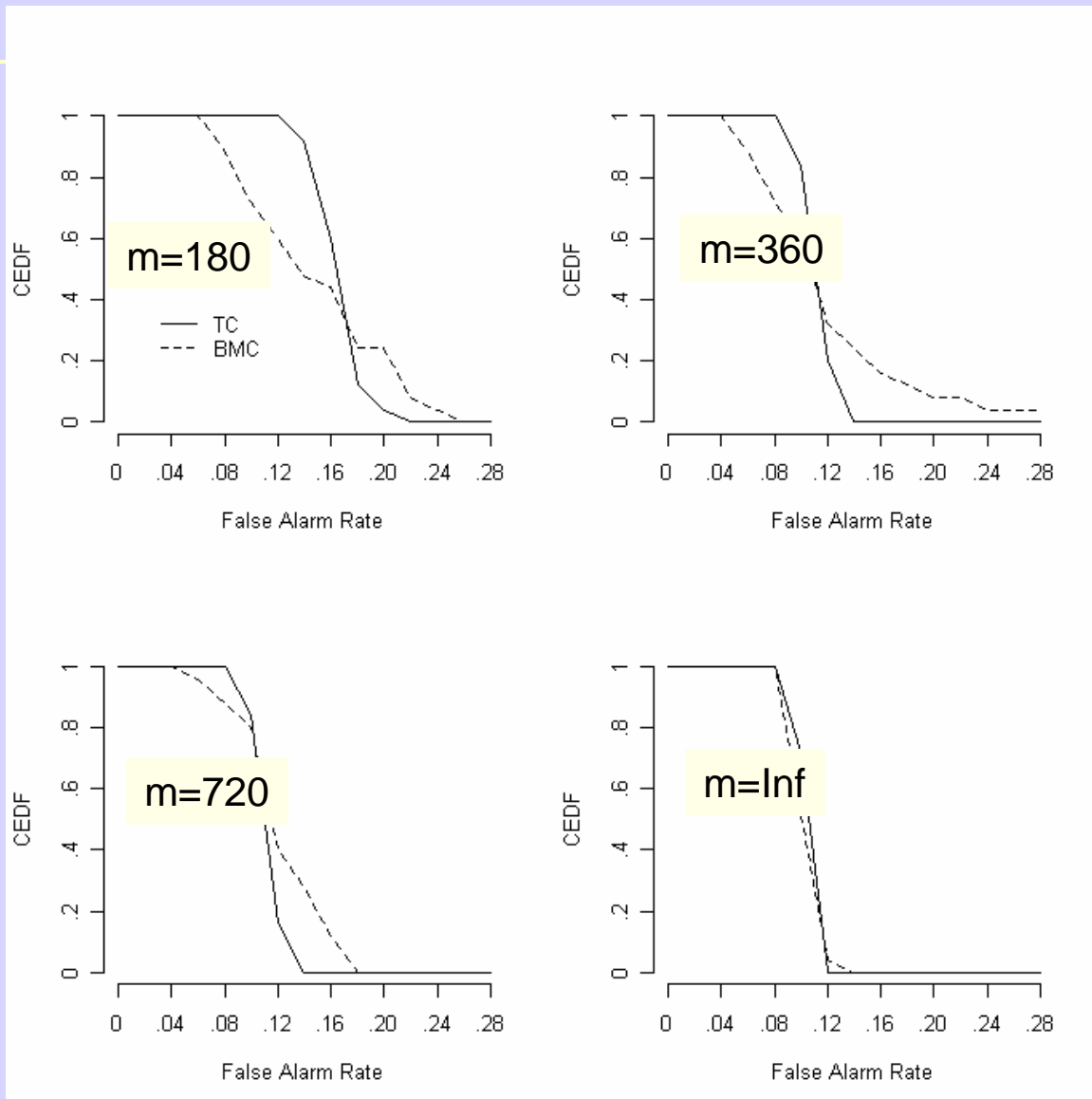
BMC



# Steps to Evaluate Cusum Algorithms

- 
1. Generate historical data from a parametric model with  $m$  observations per timeslot
  2. Generate 1000 monitoring weeks with no faults injected
  3. Evaluate detect times and percentage of faults found
  4. Repeat steps 1-3, 25 times

# Conditional False Alarm Results



For  $n=180$  BMC has a preferred distribution of conditional false alarm rates since TC shows a skew toward high values.

For the other values of  $m$ , both algorithms have conditional false alarm rate distributions that are satisfactorily centered on .10. However, the distributions for TC are more concentrated on .10.

# Unconditional False Alarm Results

---

$n$	$\gamma = 0.01$		$\gamma = 0.05$		$\gamma = 0.1$	
	TC	BMC	TC	BMC	TC	BMC
180	0.025	0.022	0.077	0.081	0.164	0.145
360	0.015	0.014	0.058	0.063	0.111	0.116
720	0.011	0.014	0.057	0.064	0.112	0.119
$\infty$	0.010	0.011	0.053	0.053	0.104	0.102

# Steps to Evaluate Cusum Algorithms

- Outer Loop**
1. Generate 12 weeks of historical data from parametric model
  2. Generate 1000 monitoring weeks with 5 faults of the same type injected per week
- Inner Loop**
- Fault-types are combinations of**  
**Magnitude** – 25%, 50%, 75%, 100%  
**Duration** – 8, 16, 30, 60, 120 min
3. Evaluate detect times and percentage of faults found
  4. Repeat steps 1-3, 25 times

# Fault Injection Study Results

Fault Pattern		Detection Percentage		Average Detect Time (min)	
Mean Increase	Duration (min)	TC	BMC	TC	BMC
25%	8	68.5	0.2	6.6	3.6
	16	81.4	0.4	7.5	9.6
	30	88.4	2.3	8.8	22.4
	60	95.7	45.9	11.4	49.7
	120	99.4	91.6	13.9	65.9
50%	8	90.7	0.3	6.2	4.0
	16	96.6	0.7	6.5	8.9
	30	99.1	3.2	6.9	23.1
	60	100.0	70.2	7.3	49.5
	120	100.0	99.9	7.3	56.2
75%	8	97.4	0.3	6.1	4.2
	16	99.8	0.5	6.2	9.9
	30	100.0	3.4	6.2	23.2
	60	100.0	78.2	6.2	49.5
	120	100.0	100.0	6.3	53.4
100%	8	99.6	0.2	6.02	3.7
	16	100.0	0.5	6.04	10.7
	30	100.0	3.8	6.04	23.3
	60	100.0	80.7	6.04	49.7
	120	100.0	100.0	6.04	52.8

Nominal 10% Two-Sided Cusum Algorithms, Random Fault Injection

# Conclusions

---

- ❑ TC is the recommended algorithm based on fault injection study results
- ❑ Algorithm is being implemented in customer networks to gain field trial experience.
- ❑ Alternative nonparametric cusum algorithms being investigated.