On SPRT and CUSUM Procedures

Mohammad Kazim Khan

khan@ima.umn.edu

Overview

Goal: Get Exact Results for SPRT & CUSUM.

Overview

Goal: Get Exact Results for SPRT & CUSUM. How?:

Overview

Goal: Get Exact Results for SPRT & CUSUM. How?:

- First Consider the SPRT for an iid sequence $\{X_i, i \geq 1\}$ having an mgf.

Overview

Goal: Get Exact Results for SPRT & CUSUM. How?:

- First Consider the SPRT for an iid sequence $\{X_i, i \geq 1\}$ having an mgf. That is, let $S_n := \sum_{i=1}^{n} X_i$, and define a stopping time

$$\tau_{a,h} := \inf\{n \geq 1 : \; S_n \notin (a, h)\}.$$

Overview

Goal: Get Exact Results for SPRT & CUSUM. How?:

- First Consider the SPRT for an iid sequence $\{X_i, i \geq 1\}$ having an mgf. That is, let $S_n := \sum_{i=1}^{n} X_i$, and define a stopping time

$$\tau_{a,h} := \inf\{n \geq 1 : \; S_n \notin (a,h)\}.$$

- Compute joint mgf of the stopped random walk $S_{\tau_{a,h}}$ and $\tau_{a,h}$.

Overview

Goal: Get Exact Results for SPRT & CUSUM. How?:

- First Consider the SPRT for an iid sequence $\{X_i, i \geq 1\}$ having an mgf. That is, let $S_n := \sum_{i=1}^n X_i$, and define a stopping time

$$\tau_{a,h} := \inf\{n \geq 1 : \; S_n \notin (a,h)\}.$$

- Compute joint mgf of the stopped random walk $S_{\tau_{a,h}}$ and $\tau_{a,h}$.
- Link $\tau_{0,h}$ with the CUSUM stopping rule:

$$N := \inf\{n \geq 1 : W_n \geq h\}, \qquad W_n := \max\{0, X_n + W_{n-1}\},$$

where $W_1 = \max\{0, X_1\}$.

Overview (continued)

- From the link of SPRT and CUSUM compute the mgf of N and its relevant statistics.

Overview (continued)

- From the link of SPRT and CUSUM compute the mgf of N and its relevant statistics.
- Present several examples.

Overview (continued)

- From the link of SPRT and CUSUM compute the mgf of N and its relevant statistics.
- Present several examples.
- Some twosided analogs.

Overview (continued)

- From the link of SPRT and CUSUM compute the mgf of N and its relevant statistics.
- Present several examples.
- Some twosided analogs.
- Mention some open problems.

The SPRT

What's known?

The SPRT

What's known?

- A lot is known about Wald's SPRT, (especially in continuous time settings). There are books on the subject.

The SPRT

What's known?

- A lot is known about Wald's SPRT, (especially in continuous time settings). There are books on the subject.
- When time is discrete (as in our case), exact closed form solutions (although known) are less common.

The SPRT

What's known?

- A lot is known about Wald's SPRT, (especially in continuous time settings). There are books on the subject.
- When time is discrete (as in our case), exact closed form solutions (although known) are less common.
- Even when X_i are normal random variables, exact solutions are hard to come by. (Several types of approximations are known, however).

The SPRT

What's known?

- A lot is known about Wald's SPRT, (especially in continuous time settings). There are books on the subject.
- When time is discrete (as in our case), exact closed form solutions (although known) are less common.
- Even when X_i are normal random variables, exact solutions are hard to come by. (Several types of approximations are known, however).

Throughout our setup will be discrete. We will show that under certain conditions a theory can be built.

The SPRT

What's known?

- A lot is known about Wald's SPRT, (especially in continuous time settings). There are books on the subject.
- When time is discrete (as in our case), exact closed form solutions (although known) are less common.
- Even when X_i are normal random variables, exact solutions are hard to come by. (Several types of approximations are known, however).

Throughout our setup will be discrete. We will show that under certain conditions a theory can be built.

SPRT & Stopped Random Walk

The Assumptions:

SPRT & Stopped Random Walk

The Assumptions:

- $\phi(\theta) = E(e^{\theta Y_1})$ exists with $E(Y_1) \neq 0$.

SPRT & Stopped Random Walk

The Assumptions:

- $\phi(\theta) = E(e^{\theta Y_1})$ exists with $E(Y_1) \neq 0$.
- $\phi(\theta) = e^{s\theta}$, $s > 0$, has two solutions, θ_1, θ_2.

SPRT & Stopped Random Walk

The Assumptions:

- $\phi(\theta) = E(e^{\theta Y_1})$ exists with $E(Y_1) \neq 0$.
- $\phi(\theta) = e^{s\theta}$, $s > 0$, has two solutions, θ_1, θ_2.
- There exist $K(\theta), h(x), R(\theta), g(x)$ so that

$$
\begin{aligned}
H(x) &:= E\left(e^{\theta Y_1} I(Y_1 \leq x)\right) = K(\theta)h(x)e^{\theta x}, \quad x < 0, \\
G(x) &:= E\left(e^{\theta Y_1} I(Y_1 \geq x)\right) = R(\theta)g(x)e^{\theta x}, \quad x > 0.
\end{aligned}
$$

SPRT & Stopped Random Walk

Conclusions:

SPRT & Stopped Random Walk

Conclusions:

$$E\left(e^{-s\tau_{a,h}} I(S_{\tau_{a,h}} \leq a)\right)$$
$$= \frac{K(0)\left(R(\theta_2)e^{\theta_2 h} - R(\theta_1)e^{\theta_1 h}\right)}{K(\theta_1)R(\theta_2)e^{\theta_1 a+\theta_2 h} - K(\theta_2)R(\theta_1)e^{\theta_2 a+\theta_1 h}}.$$

SPRT & Stopped Random Walk

Conclusions:

$$E\left(e^{-s\tau_{a,h}} I(S_{\tau_{a,h}} \le a)\right) = \frac{K(0)\left(R(\theta_2)e^{\theta_2 h} - R(\theta_1)e^{\theta_1 h}\right)}{K(\theta_1)R(\theta_2)e^{\theta_1 a+\theta_2 h} - K(\theta_2)R(\theta_1)e^{\theta_2 a+\theta_1 h}}.$$

$$E\left(e^{-s\tau_{a,h}}\right) = \frac{K(0)\left(R(\theta_2)e^{\theta_2 h} - R(\theta_1)e^{\theta_1 h}\right) + R(0)\left(K(\theta_1)e^{\theta_1 a} - K(\theta_2)e^{\theta_2 a}\right)}{K(\theta_1)R(\theta_2)e^{\theta_1 a+\theta_2 h} - K(\theta_2)R(\theta_1)e^{\theta_2 a+\theta_1 h}}$$

SPRT & Stopped Random Walk

Conclusions:

$$E\left(e^{-s\tau_{a,h}} I(S_{\tau_{a,h}} \le a)\right) = \frac{K(0)\left(R(\theta_2)e^{\theta_2 h} - R(\theta_1)e^{\theta_1 h}\right)}{K(\theta_1)R(\theta_2)e^{\theta_1 a+\theta_2 h} - K(\theta_2)R(\theta_1)e^{\theta_2 a+\theta_1 h}}.$$

$$E\left(e^{-s\tau_{a,h}}\right) = \frac{K(0)\left(R(\theta_2)e^{\theta_2 h} - R(\theta_1)e^{\theta_1 h}\right) + R(0)\left(K(\theta_1)e^{\theta_1 a} - K(\theta_2)e^{\theta_2 a}\right)}{K(\theta_1)R(\theta_2)e^{\theta_1 a+\theta_2 h} - K(\theta_2)R(\theta_1)e^{\theta_2 a+\theta_1 h}}$$

$$P(S_{\tau_{a,h}} \ge h) = \frac{R(0)(K(0) - K(\theta^*)e^{\theta^* a})}{K(0)R(\theta^*)e^{\theta^* b} - R(0)K(\theta^*)e^{\theta^* a}},$$

where $\theta_1(s) \to 0$, $\theta_2(s) \to \theta^*$ as $s \to 0^+$.

SPRT & Stopped Random Walk

Conclusions (continued):

$$E(\tau_{0,h})$$

$$= \frac{K(0)\{R'(0) + hR(0)\} - K'(0)\{R(0) - R(\theta^*)e^{\theta^* h}\} - K(\theta^*)\{R}{E(Y)\{K(0)R(\theta^*)e^{\theta^* h} - K(\theta^*)R(0)\}}$$

SPRT & Stopped Random Walk

Conclusions (continued):

$$E(\tau_{0,h}) = \frac{K(0)\{R'(0) + hR(0)\} - K'(0)\{R(0) - R(\theta^*)e^{\theta^* h}\} - K(\theta^*)\{R}{E(Y)\{K(0)R(\theta^*)e^{\theta^* h} - K(\theta^*)R(0)\}}$$

- We can compute higher moments of $\tau_{0,b}$.

SPRT & Stopped Random Walk

Conclusions (continued):

$E(\tau_{0,h})$

$$= \frac{K(0)\{R'(0) + hR(0)\} - K'(0)\{R(0) - R(\theta^*)e^{\theta^* h}\} - K(\theta^*)\{R}{E(Y)\{K(0)R(\theta^*)e^{\theta^* h} - K(\theta^*)R(0)\}}$$

- We can compute higher moments of $\tau_{0,b}$.
- Also we can compute the one sided joint moments of the stopped random walk and the SPRT.
  (Algebra is painful, but the conclusions are not long).

The Degenerate Case

A Slight Glitch:

The Degenerate Case

A Slight Glitch:

- It turns out that for CUSUM, we need statistics for $\tau_{0,h}$.

The Degenerate Case

A Slight Glitch:

- It turns out that for CUSUM, we need statistics for $\tau_{0,h}$.
- But for $\tau_{0,h}$ the last set of results may not exist when $K(\theta), R(\theta)$ are constant functions, (called the degenerate situation).

The Degenerate Case

A Slight Glitch:

- It turns out that for CUSUM, we need statistics for $\tau_{0,h}$.
- But for $\tau_{0,h}$ the last set of results may not exist when $K(\theta), R(\theta)$ are constant functions, (called the degenerate situation).
- The trick is to study a slight modification of the SPRT $\tau_{0,h}$, defined by

$$\tau_{0^-,h} := \inf\{n \geq 1,\ S_n \notin [0,h)\}.$$

The Degenerate Case

A Slight Glitch:

- It turns out that for CUSUM, we need statistics for $\tau_{0,h}$.
- But for $\tau_{0,h}$ the last set of results may not exist when $K(\theta), R(\theta)$ are constant functions, (called the degenerate situation).
- The trick is to study a slight modification of the SPRT $\tau_{0,h}$, defined by

$$\tau_{0^-,h} := \inf\{n \geq 1,\ S_n \notin [0,h)\}.$$

(With this modification, we again can get the exact results under our previous assumptions.)

The Degenerate Case

A Slight Glitch:

- It turns out that for CUSUM, we need statistics for $\tau_{0,h}$.
- But for $\tau_{0,h}$ the last set of results may not exist when $K(\theta), R(\theta)$ are constant functions, (called the degenerate situation).
- The trick is to study a slight modification of the SPRT $\tau_{0,h}$, defined by

$$\tau_{0-,h} := \inf\{n \geq 1,\ S_n \notin [0, h)\}.$$

(With this modification, we again can get the exact results under our previous assumptions.)

For instance, $P(S_{\tau_{0-,h}} \geq h) = (1 - e^{-\theta^*}) \quad / \quad (e^{\theta^* h} - e^{-\theta^*})$.

The CUSUM Procedure

Due to E. S. Page (1954):

The CUSUM Procedure

Due to E. S. Page (1954): Again, let $a < h$ be two given numbers and let $Y_1, Y_2, \cdots$ be a sequence of iid random variables. Set $S_n = Y_1 + Y_2 + \cdots + Y_n$, $n \geq 1$, $S_0 = 0$. Define the following random variables and stopping times.

$$W_{j,k} = \max\{0,\ W_{j,k-1} + Y_{j+k}\}, \qquad W_{j,1} = \max\{0, Y_{j+1}\},$$

for an integer $j = 0, 1, 2, \cdots$.

$$\tau_{a,h} = \inf\{n \geq 1 :\ S_n \notin (a, h)\}.$$

$$N_{0,h-a,j} = \inf\{k \geq 1 :\ W_{j,k} \geq h - a\}, \quad \text{for an integer } j \geq 1.$$

The CUSUM Procedure

Due to E. S. Page (1954): Again, let $a < h$ be two given numbers and let $Y_1, Y_2, \cdots$ be a sequence of iid random variables. Set $S_n = Y_1 + Y_2 + \cdots + Y_n$, $n \geq 1$, $S_0 = 0$. Define the following random variables and stopping times.

$$W_{j,k} = \max\{0,\ W_{j,k-1} + Y_{j+k}\}, \qquad W_{j,1} = \max\{0, Y_{j+1}\},$$

for an integer $j = 0, 1, 2, \cdots$.

$$\tau_{a,h} = \inf\{n \geq 1 :\ S_n \notin (a,h)\}.$$

$$N_{0,h-a,j} = \inf\{k \geq 1 :\ W_{j,k} \geq h - a\}, \quad \text{for an integer } j \geq 1.$$

The stopping rule $N_{0,h,1}$ is the one-sided cusum stopping rule of E. S. Page, which we will sometimes denote by N, or $N_{0,h}$, or N_1 for short.

The CUSUM Procedure – A Recast

We may restate the stopping rule $N_{0,h}$ in another way.

The CUSUM Procedure – A Recast

We may restate the stopping rule $N_{0,h}$ in another way.

- Let $D_n := S_n - \min_{0 \le k \le n} S_k$. Then $N_{0,h} := \inf \{n \ge 1 : \; D_n \ge h\}$. Thus the process, D_n, trigers stopping time as soon as $D_n \ge h$.

The CUSUM Procedure – A Recast

We may restate the stopping rule $N_{0,h}$ in another way.

- Let $D_n := S_n - \min_{0 \leq k \leq n} S_k$. Then $N_{0,h} := \inf \{n \geq 1 : \; D_n \geq h\}$. Thus the process, D_n, trigers stopping time as soon as $D_n \geq h$.
- The questions of interest include finding the distribution, moments, and one sided Laplace transforms of $N_{0,h}$.

The CUSUM Procedure – A Recast

We may restate the stopping rule $N_{0,h}$ in another way.

- Let $D_n := S_n - \min_{0 \leq k \leq n} S_k$. Then $N_{0,h} := \inf \{ n \geq 1 : \ D_n \geq h \}$. Thus the process, D_n, trigers stopping time as soon as $D_n \geq h$.
- The questions of interest include finding the distribution, moments, and one sided Laplace transforms of $N_{0,h}$.
- The original use of $N_{0,h}$ was for process control. Now it has more diverse applications. For instance, it may be used in Finance for studying the so called "trading the line" strategy.

The CUSUM Procedure – A Recast

We may restate the stopping rule $N_{0,h}$ in another way.

- Let $D_n := S_n - \min_{0 \le k \le n} S_k$. Then $N_{0,h} := \inf\{n \ge 1 : \ D_n \ge h\}$. Thus the process, D_n, trigers stopping time as soon as $D_n \ge h$.
- The questions of interest include finding the distribution, moments, and one sided Laplace transforms of $N_{0,h}$.
- The original use of $N_{0,h}$ was for process control. Now it has more diverse applications. For instance, it may be used in Finance for studying the so called "trading the line" strategy.
- It is my feeling that its modifications may also have some use in internet router traffic monitoring.

The CUSUM Procedure – A Recast

We may restate the stopping rule $N_{0,h}$ in another way.

- Let $D_n := S_n - \min_{0 \le k \le n} S_k$. Then $N_{0,h} := \inf\{n \ge 1 : \; D_n \ge h\}$. Thus the process, D_n, trigers stopping time as soon as $D_n \ge h$.
- The questions of interest include finding the distribution, moments, and one sided Laplace transforms of $N_{0,h}$.
- The original use of $N_{0,h}$ was for process control. Now it has more diverse applications. For instance, it may be used in Finance for studying the so called "trading the line" strategy.
- It is my feeling that its modifications may also have some use in internet router traffic monitoring.
- suspicision: perhaps tumor data monitoring (c.f. Michael Newton's last talk)???

The Fundamental Identity

Known Links:

The Fundamental Identity

Known Links: E. S. Page (1954) knew that the SPRT and the CUSUM were related as follows.

$$E(N_{0,h}) = \frac{E(\tau_{0,h})}{P(S_{\tau_{0,h}} \geq h)}.$$

The Fundamental Identity

Known Links: E. S. Page (1954) knew that the SPRT and the CUSUM were related as follows.

$$E(N_{0,h}) = \frac{E(\tau_{0,h})}{P(S_{\tau_{0,h}} \geq h)}.$$

Later it was realized that

$$\begin{aligned} P(N=n) &= \sum_{j=1}^{n-1} P(N=n-j) P(\tau_{0,h}=j, S_{\tau_{0,h}} \leq 0) \\ &\quad + P(\tau_{0,h}=n, S_{\tau_{0,h}} \geq h). \end{aligned}$$

The Fundamental Identity (continued)

Whats the idea?

The Fundamental Identity (continued)

Whats the idea? The above results suggest that perhaps the sample paths of the processes might be linked. It turns out that that is the case and the link between the SPRT and the CUSUM stopping times is as follows.

$$N_{0,h} = \tau_{0,h} + N_{0,h,\tau_{0,h}}\, I(S_{\tau_{0,h}} \leq 0),$$

where $N_{0,h,\tau_{0,h}}$ is another CUSUM independent of $S_{\tau_{0,h}}$ when $\tau_{0,h}$ is given.

The Fundamental Identity (continued)

Whats the idea? The above results suggest that perhaps the sample paths of the processes might be linked. It turns out that that is the case and the link between the SPRT and the CUSUM stopping times is as follows.

$$N_{0,h} = \tau_{0,h} + N_{0,h,\tau_{0,h}} \, I(S_{\tau_{0,h}} \leq 0),$$

where $N_{0,h,\tau_{0,h}}$ is another CUSUM independent of $S_{\tau_{0,h}}$ when $\tau_{0,h}$ is given. (With this result in hand and our earlier exact results for the SPRT, we can find (modulo some algebraic tears) all the relevant exact results concerning the CUSUM procedure.)

The Fundamental Identity (continued)

For instance, the fundamental identity gives:

The Fundamental Identity (continued)

For instance, the fundamental identity gives:

- $E(N) = \frac{E(\tau_{0,h})}{P(S_{\tau_{0,h}} \geq h)}$.
- $P(N = n) = \sum_{j=1}^{n-1} P(N = n - j)P(\tau_{0,h} = j, S_{\tau_{0,h}} \leq 0) + P(\tau_{0,h} = n, S_{\tau_{0,h}} \geq h)$.
- $P(N > n) = \sum_{k=1}^{n} P(S_{\tau_{0,h}} \leq 0,\ \tau_{0,h} = k)P(N > n - k) + P(\tau_{0,h} > n)$.
- $Var(N) = \frac{Var(\tau_{0,h})}{P(S_{\tau_{0,h}} \geq h)} + \frac{2E(\tau_{0,h})E(\tau_{0,h} I(S_{\tau_{0,h}} \leq 0)) - (E(\tau_{0,h}))^2 P(S_{\tau_{0,h}} \leq 0)}{(P(S_{\tau_{0,h}} \geq h))^2}$.
- For any $s > 0$, we have $E(e^{-sN}) = \frac{E(e^{-s\tau_{0,h}}) - E(e^{-s\tau_{0,h}} I(S_{\tau_{0,h}} \leq 0))}{1 - E(e^{-s\tau_{0,h}} I(S_{\tau_{0,h}} \leq 0))}$.

Results for CUSUM

Moreover, under our earlier assumptions, the fundamental identity gives:

Results for CUSUM

Moreover, under our earlier assumptions, the fundamental identity gives:

$$E(e^{-sN}) = \frac{R(0)\{K(\theta_1) - K(\theta_2)\}}{\{K(\theta_1) - K(0)\}R(\theta_2)e^{\theta_2 h} - \{K(\theta_2) - K(0)\}R(\theta_1)e^{\theta_1}}$$

$$E(N) = \frac{1}{E(Y)}\left\{h + \frac{R'(0)}{R(0)} - \frac{K'(0)\{R(0) - R(\theta^*)e^{\theta^* h}\}}{R(0)(K(0) - K(\theta^*))}\right\},$$

where $\theta_1(s) \to 0$ and $\theta_2(s) \to \theta^*$ as $s \to 0$.

Results for CUSUM

Moreover, under our earlier assumptions, the fundamental identity gives:

$$E(e^{-sN}) = \frac{R(0)\{K(\theta_1) - K(\theta_2)\}}{\{K(\theta_1) - K(0)\}R(\theta_2)e^{\theta_2 h} - \{K(\theta_2) - K(0)\}R(\theta_1)e^{\theta_1}}$$

$$E(N) = \frac{1}{E(Y)}\left\{h + \frac{R'(0)}{R(0)} - \frac{K'(0)\{R(0) - R(\theta^*)e^{\theta^* h}\}}{R(0)(K(0) - K(\theta^*))}\right\},$$

where $\theta_1(s) \to 0$ and $\theta_2(s) \to \theta^*$ as $s \to 0$.
When $E(Y) = 0$, we get

$$E(N) = \frac{1}{Var(Y)}\left\{h^2 + h\frac{2R'(0)K'(0) - R(0)K''(0)}{R(0)K'(0)} + \frac{K'(0)R''(0)}{R(0)}\right.$$

Results for CUSUM (Continued)

Under our earlier assumptions:

Results for CUSUM (Continued)

Under our earlier assumptions: The variance, $Var(N)$ is

$$= \frac{1}{(E(Y))^2}\Big\{Var(Y)E(N) + 2(E(Y)E(N) - h)\frac{K'(0)+K'(\theta^*)}{K(0)-K(\theta^*)}$$

$$+ (E(Y)E(N) - h)^2 + \frac{K''(0)\{R(0) - R(\theta^*)e^{\theta^* h}\}}{R(0)(K(0)-K(\theta^*))} - \frac{R''(0)}{R(0)}$$

$$-2\frac{K'(\theta^*)R'(0) - K'(0)R'(\theta^*)e^{\theta^* h}}{R(0)(K(0)-K(\theta^*))} + 2h\frac{K'(0)R(\theta^*)e^{\theta^* h}}{R(0)(K(0)-K(\theta^*))}\Big\}$$

when $\theta_1'(s) + \theta_2'(s) = 0$ in a neighborhood of zero, and $E(Y) \neq 0$.

Results for CUSUM (Continued)

Under our earlier assumptions (degenerate case):

Results for CUSUM (Continued)

Under our earlier assumptions (degenerate case):

$$
\begin{aligned}
E(e^{-sN}) &= \frac{e^{\theta_2} - e^{\theta_1}}{(1-e^{\theta_1})e^{\theta_2(h+1)} - (1-e^{\theta_2})e^{\theta_1(h+1)}}, \qquad s > 0, \\
E(N) &= \frac{1}{E(Y)}\left\{h + \frac{1-e^{\theta^* h}}{1-e^{-\theta^*}}\right\}, \qquad E(Y) \neq 0, \\
E(N) &= \frac{h(h+1)}{Var(Y)}, \qquad E(Y) = 0.
\end{aligned}
$$

Furthermore, when $\theta_1'(s) + \theta_2'(s) = 0$ in a neighborhood of zero, and $E(Y) \neq 0$, the variance, $Var(N)$, simplifies to

$$
= \frac{Var(Y)}{(E(Y))^3}\left\{h + \frac{1-e^{\theta^* h}}{1-e^{-\theta^*}}\right\} + \frac{\{e^{\theta^*(h+1)}+3\}\{e^{\theta^* h}-1\}}{(E(Y))^2(1-e^{-\theta^*})(e^{\theta^*}-1)}.
$$

Results for CUSUM (Continued)

Under our earlier assumptions:

Results for CUSUM (Continued)

Under our earlier assumptions:

- The above results show that, for large h, $E(N) \approx \frac{h}{E(Y)}$ when $\theta^* < 0$, which corresponds to the case when $E(Y) > 0$.

Results for CUSUM (Continued)

Under our earlier assumptions:

- The above results show that, for large h, $E(N) \approx \frac{h}{E(Y)}$ when $\theta^* < 0$, which corresponds to the case when $E(Y) > 0$.
- Furthermore, when $E(Y) > 0$, and $0 < Var(Y) = \sigma^2$, one gets the following asymptotic normality.

$$\lim_{h \to \infty} P\left(\frac{N - (h/\mu)}{\sigma \sqrt{(h/\mu^3)}} \leq x \right) = \Phi(x), \qquad x \in \Re.$$

Results for CUSUM (Continued)

Under our earlier assumptions:

Results for CUSUM (Continued)

Under our earlier assumptions: On the other hand, when $E(Y) < 0$, for any $t > 0$, (in the nondegenerate case) we get

$$\lim_{h\to\infty} E\left(e^{-Nte^{-\theta^* h}}\right) = \frac{\beta}{t+\beta}, \qquad \beta = \frac{E(Y)R(0)(K(0)-K(\theta^*))}{K'(0)R(\theta^*)}.$$

Results for CUSUM (Continued)

Under our earlier assumptions: On the other hand, when $E(Y) < 0$, for any $t > 0$, (in the nondegenerate case) we get

$$\lim_{h \to \infty} E\left(e^{-Nte^{-\theta^* h}}\right) = \frac{\beta}{t+\beta}, \qquad \beta = \frac{E(Y)R(0)(K(0) - K(\theta^*))}{K'(0)R(\theta^*)}.$$

When $E(Y) < 0$, for any $t > 0$, (in the degenerate case)

$$\lim_{h \to \infty} E\left(e^{-Nte^{-\theta^* h}}\right) = \frac{\beta}{t+\beta}, \qquad \beta = -E(Y)(1 - e^{-\theta^*}).$$

Results for CUSUM (Continued)

$E(Y)N/h \stackrel{a.s.}{\to} 1$ is well known, when $E(Y) > 0$.

Results for CUSUM (Continued)

$E(Y)N/h \overset{a.s.}{\to} 1$ is well known, when $E(Y) > 0$.
Under our earlier assumptions:

Results for CUSUM (Continued)

$E(Y)N/h \stackrel{a.s.}{\to} 1$ is well known, when $E(Y) > 0$.
Under our earlier assumptions: Now rate of convergence can be deduced since the collection

$$\left\{ \sqrt{h} \left| \frac{E(Y)N}{h} - 1 \right|, \ h > 0 \right\},$$

turns out to be uniformly integrable, when $E(Y) > 0$.

Approximations for CUSUM

For $-\delta < 0 < h$,

Approximations for CUSUM

For $-\delta < 0 < h$,

$$E(e^{-sN}) \approx \frac{e^{-\delta\theta_1} - e^{-\delta\theta_2}}{e^{\theta_2 h}(e^{-\delta\theta_1} - 1) - e^{\theta_1 h}(e^{-\delta\theta_2} - 1)},$$

$$P(S_{\tau_{0-,h}} \geq h) \approx \frac{1 - e^{-\delta\theta^*}}{e^{\theta^* h} - e^{-\delta\theta^*}}, \quad E(N) \approx \frac{1}{E(Y)} \left\{ h + \frac{\delta(1 - e^{\theta^* h}}{1 - e^{-\delta\theta^*}} \right.$$

Approximations for CUSUM

For $-\delta < 0 < h$,

$$E(e^{-sN}) \;\approx\; \frac{e^{-\delta\theta_1} - e^{-\delta\theta_2}}{e^{\theta_2 h}(e^{-\delta\theta_1} - 1) - e^{\theta_1 h}(e^{-\delta\theta_2} - 1)},$$

$$P(S_{\tau_{0-,h}} \geq h) \;\approx\; \frac{1 - e^{-\delta\theta^*}}{e^{\theta^* h} - e^{-\delta\theta^*}}, \quad E(N) \;\approx\; \frac{1}{E(Y)} \left\{ h + \frac{\delta(1 - e^{\theta^* h}}{1 - e^{-\delta\theta^*}} \right.$$

When Y is a continuous r.v., letting $\delta \to 0$,

$$E(e^{-sN}) \;\approx\; \frac{\theta_2 - \theta_1}{\theta_2 e^{\theta_1 h} - \theta_1 e^{\theta_2 h}}, \qquad s > 0, \qquad E(N) \;\approx\; \frac{h - \frac{e^{\theta^* h} - 1}{\theta^*}}{E(Y)}.$$

This approximation of $E(N)$ is the usual Wald's approximation. Similarly, one can now propose an approximation for $Var(N)$

Two Sided CUSUM

(Two sided Cusum can also be handled):

Two Sided CUSUM

(Two sided Cusum can also be handled): Just define:

$$
\begin{aligned}
N_1 &= \inf\{n \geq 1 : S_n - \min_{0 \leq i \leq n} S_i \geq h\} \\
N_2 &= \inf\{n \geq 1 : \max_{0 \leq i \leq n} S_i - S_n \geq h\} = \inf\{n \geq 1 : S_n^* - \min_{0 \leq i \leq} \\
\tau_{a,h} &= \inf\{n \geq 1 : S_n \notin (a, h)\}, \\
\tau_{a,h^+} &= \inf\{n \geq 1 : S_n \notin (a, h]\}.
\end{aligned}
$$

Two Sided CUSUM

(Two sided Cusum can also be handled): Just define:

$$\begin{aligned} N_1 &= \inf\{n \geq 1: S_n - \min_{0\leq i\leq n} S_i \geq h\} \\ N_2 &= \inf\{n \geq 1: \max_{0\leq i\leq n} S_i - S_n \geq h\} = \inf\{n \geq 1: S_n^* - \min_{0\leq i\leq n} \\ \tau_{a,h} &= \inf\{n \geq 1: S_n \notin (a,h)\}, \\ \tau_{a,h^+} &= \inf\{n \geq 1: S_n \notin (a,h]\}. \end{aligned}$$

The cusum N_2 is linked with $\tau_{-h,0}$, (or $\tau_{-h,0^+}$ in the degnerate case). When the distribution of Y_1 obeys our earlier conditions then so will be the case with $Y_1^* := -Y_1$,

$$\begin{aligned} E\left(e^{\theta Y_1^*} I(Y_1^* \leq x)\right) &= E(e^{-\theta Y_1} I(Y_1 \geq -x)) = R(-\theta)g(-x)e^{\theta x}, \\ E\left(e^{\theta Y_1^*} I(Y_1^* \geq x)\right) &= E(e^{-\theta Y_1} I(Y_1 \leq -x)) = K(-\theta)h(-x)e^{\theta x}, \end{aligned}$$

Two Sided CUSUM (Continued)

So, the K, R functions switch role as well as the argument changes sign, and the earlier results apply for N_2.

Two Sided CUSUM (Continued)

So, the K, R functions switch role as well as the argument changes sign, and the earlier results apply for N_2.
The two sided CUSUM stopping rule is just

$$N^* := \min\{N_1, N_2\}.$$

Two Sided CUSUM (Continued)

So, the K, R functions switch role as well as the argument changes sign, and the earlier results apply for N_2.
The two sided CUSUM stopping rule is just

$$N^* := \min\{N_1, N_2\}.$$

When all is put together for the two sided CUSUM, we get

$$\begin{aligned} \phi(s) &= E(e^{-sN^*}) = \frac{\phi_1(s) + \phi_2(s) - 2\phi_1(s)\phi_2(s)}{1 - \phi_1(s)\phi_2(s)}, \qquad s > 0. \\ E(N^*) &= \frac{E(N_1)E(N_2)}{E(N_1) + E(N_2)}, \quad \text{take } \alpha := \frac{E(N_2)}{E(N_1) + E(N_2)}, \\ Var(N^*) &= Var(N_1)\alpha^2 + Var(N_2)(1-\alpha)^2 - (E(N^*))^2. \end{aligned}$$

Examples

(Enough is enough!): Show me some examples?

Examples

(Enough is enough!): Show me some examples?

(Comparing Bernoulli Processes) Consider independent Bernoulli processes $\{X_{1n}\}$, $\{X_{2n}\}$, where $P(X_{1n} = 1) = p_1 = 1 - P(X_{1n} = 0)$, and $P(X_{2n} = 1) = p_2 = 1 - P(X_{2n} = 0)$. Consider the case when the observed data is $Y_n = X_{1n} - X_{2n}$. In this case Y_1 takes three values, $-1, 0, 1$, with respective probabilities $q := q_1 p_2$, $p_1 p_2 + q_1 q_2$ and $p := p_1 q_2$.

Examples

(Enough is enough!): Show me some examples?

(Comparing Bernoulli Processes) Consider independent Bernoulli processes $\{X_{1n}\}$, $\{X_{2n}\}$, where $P(X_{1n} = 1) = p_1 = 1 - P(X_{1n} = 0)$, and $P(X_{2n} = 1) = p_2 = 1 - P(X_{2n} = 0)$. Consider the case when the observed data is $Y_n = X_{1n} - X_{2n}$. In this case Y_1 takes three values, $-1, 0, 1$, with respective probabilities $q := q_1 p_2$, $p_1 p_2 + q_1 q_2$ and $p := p_1 q_2$.

Note that the classical Gambler's ruin problem is a special case when $P(Y_1 = 0) = 0$.

Examples

(Enough is enough!): Show me some examples?
(Comparing Bernoulli Processes) Consider independent Bernoulli processes $\{X_{1n}\}$, $\{X_{2n}\}$, where $P(X_{1n} = 1) = p_1 = 1 - P(X_{1n} = 0)$, and $P(X_{2n} = 1) = p_2 = 1 - P(X_{2n} = 0)$. Consider the case when the observed data is $Y_n = X_{1n} - X_{2n}$. In this case Y_1 takes three values, $-1, 0, 1$, with respective probabilities $q := q_1 p_2$, $p_1 p_2 + q_1 q_2$ and $p := p_1 q_2$.
Note that the classical Gambler's ruin problem is a special case when $P(Y_1 = 0) = 0$.

In this case our assumptions are met and all relevant information about the SPRT and the CUSUM can be deduced in closed form. (Gory details on next slide)

Examples (continued)

(Comparing Bernoulli Processes)

Examples (continued)

(Comparing Bernoulli Processes) The vital statistics are:

$$E(e^{-sN}) = \frac{e^{\theta_2} - e^{\theta_1}}{(1-e^{\theta_1})e^{\theta_2(h+1)} - (1-e^{\theta_2})e^{\theta_1(h+1)}}, \qquad s > 0,$$

$$E(N) = \frac{h}{p-q} - \frac{q}{(p-q)^2}\{1-(q/p)^h\}, \qquad p \neq q,$$

$$Var(N) = \frac{p(1-p)+q(1-q)-2pq}{(p-q)^3}\left\{h + \frac{1-(q/p)^h}{1-(p/q)} + \frac{pq\{(q/p)^{h+1}+3\}\{(q/p)^h-1\}}{(p-q)^4}\right\}, \qquad p \neq q$$

Examples (continued)

(Comparing Bernoulli Processes) The vital statistics are:

$$E(e^{-sN}) = \frac{e^{\theta_2} - e^{\theta_1}}{(1-e^{\theta_1})e^{\theta_2(h+1)} - (1-e^{\theta_2})e^{\theta_1(h+1)}}, \qquad s > 0,$$

$$E(N) = \frac{h}{p-q} - \frac{q}{(p-q)^2}\{1-(q/p)^h\}, \qquad p \neq q,$$

$$Var(N) = \frac{p(1-p)+q(1-q)-2pq}{(p-q)^3}\left\{h + \frac{1-(q/p)^h}{1-(p/q)} + \frac{pq\{(q/p)^{h+1}+3\}\{(q/p)^h-1\}}{(p-q)^4}\right\}, \qquad p \neq q$$

The Laplace transform is due to Kennedy (1976) by a martingale argument. The mean is by Munford (1980). When $E(Y) < 0$, $Ne^{-\theta^* h} \to \text{Exponential}$ r.v. with mean $q/(p-q)^2$.

Examples (Continued)

(Comparing Geometric Proportions)

Examples (Continued)

(Comparing Geometric Proportions)

Consider the problem of monitoring two proportions p_1, p_2. For instance, let p_1 be the cure rate of a disease by an old drug, and let p_2 be the cure rate of a new drug. An equally illustrative example involves p_1 and p_2 as the proportions of defectives by an old and new machine respectively.

Examples (Continued)

(Comparing Geometric Proportions)

Consider the problem of monitoring two proportions p_1, p_2. For instance, let p_1 be the cure rate of a disease by an old drug, and let p_2 be the cure rate of a new drug. An equally illustrative example involves p_1 and p_2 as the proportions of defectives by an old and new machine respectively.

The question is: "is the new drug (or the new machine) better than the old one"? Given the sequential nature of associated data, and the possibility of deteoriating quality or the mere possiblity of improved quality of the new drug (machine), it seems that detecting a change $\delta = p_1 - p_2$ is more appropriate. Moreover, it is also natural to observe the number of items (or patients) until the first defective (or the first cure/treatement), and this cycle of observations is repeated sequentially.

Examples (Continued)

(Comparing Geometric Proportions)

Consider the problem of monitoring two proportions p_1, p_2. For instance, let p_1 be the cure rate of a disease by an old drug, and let p_2 be the cure rate of a new drug. An equally illustrative example involves p_1 and p_2 as the proportions of defectives by an old and new machine respectively.

The question is: "is the new drug (or the new machine) better than the old one"? Given the sequential nature of associated data, and the possibility of deteoriating quality or the mere possiblity of improved quality of the new drug (machine), it seems that detecting a change $\delta = p_1 - p_2$ is more appropriate. Moreover, it is also natural to observe the number of items (or patients) until the first defective (or the first cure/treatement), and this cycle of observations is repeated sequentially.

Examples (Continued)

(Comparing Geometric Proportions)

Examples (Continued)

(Comparing Geometric Proportions)

Thus we consider the geometric distribution, $P(\xi_i = j) = p_i q_i^{j-1}$, $j = 1, 2, \cdots$ ($i = 1, 2$), as the discrete waiting time distributions of the cycles, and let $Y = U - V$.

Examples (Continued)

(Comparing Geometric Proportions)

Thus we consider the geometric distribution, $P(\xi_i = j) = p_i q_i^{j-1}$, $j = 1, 2, \cdots$ ($i = 1, 2$), as the discrete waiting time distributions of the cycles, and let $Y = U - V$. Again our assumptions are met and the vital statistics for the SPRT and the CUSUM can be deduced.

Examples (Continued)

(Comparing Geometric Proportions)
Thus we consider the geometric distribution, $P(\xi_i = j) = p_i q_i^{j-1}$, $j = 1, 2, \cdots$ ($i = 1, 2$), as the discrete waiting time distributions of the cycles, and let $Y = U - V$. Again our assumptions are met and the vital statistics for the SPRT and the CUSUM can be deduced.

(Gory details should be omitted ... as a gesture of goodwill)

Examples (Continued)

(Comparing Poisson Processes)

Examples (Continued)

(Comparing Poisson Processes)

Consider two independent Poisson processes $\{M_1(t),\ t \geq 0\}$, with rate μ and $\{M_2(t),\ t \geq 0\}$ with rate λ. The observable process is the difference of their respective inter-arrival times, $U_i \sim Exp(\mu)$ and $V_i \sim Exp(\lambda)$. The aim is to detect as early as one can if $\mu \neq \lambda$.

Examples (Continued)

(Comparing Poisson Processes)

Consider two independent Poisson processes $\{M_1(t),\ t \geq 0\}$, with rate μ and $\{M_2(t),\ t \geq 0\}$ with rate λ. The observable process is the difference of their respective inter-arrival times, $U_i \sim Exp(\mu)$ and $V_i \sim Exp(\lambda)$. The aim is to detect as early as one can if $\mu \neq \lambda$.

It is easy to check that our assumptions about the mgf are met and all vital statistics of the CUSUM and the SPRT can be deduced.

Examples (Continued)

(Comparing Poisson Processes)
Consider two independent Poisson processes $\{M_1(t),\ t \geq 0\}$, with rate μ and $\{M_2(t),\ t \geq 0\}$ with rate λ. The observable process is the difference of their respective inter-arrival times, $U_i \sim Exp(\mu)$ and $V_i \sim Exp(\lambda)$. The aim is to detect as early as one can if $\mu \neq \lambda$.
It is easy to check that our assumptions about the mgf are met and all vital statistics of the CUSUM and the SPRT can be deduced.

(Gory details should be omitted ... as a second gesture of goodwill)

Open Problems

(Where is the weakness in the theory?)

Open Problems

(Where is the weakness in the theory?)

The assumptions we made about the mgf of Y are too strong.

The most useful case when $Y_i \overset{iid}{\sim} N(\mu, 1)$ does not fall within the theory. No closed form results are known to date.

Open Problems

(Where is the weakness in the theory?)

The assumptions we made about the mgf of Y are too strong.

The most useful case when $Y_i \overset{iid}{\sim} N(\mu, 1)$ does not fall within the theory. No closed form results are known to date.
The assumption that Y_i should be univariate random variables is also too restrictive. One should consider multivariate analogs. (I just have not had the time to devote to this problem. This just might work out.)

Open Problems

(Where is the weakness in the theory?)

The assumptions we made about the mgf of Y are too strong.

The most useful case when $Y_i \overset{iid}{\sim} N(\mu, 1)$ does not fall within the theory. No closed form results are known to date.

The assumption that Y_i should be univariate random variables is also too restrictive. One should consider multivariate analogs. (I just have not had the time to devote to this problem. This just might work out.)

If one wants to be bold, (at least via simulations) one may want to relax the iid assumption. (My computer skills are just too limited to perform this task.)

Open Problems

(Where is the weakness in the theory?)
The assumptions we made about the mgf of Y are too strong.

The most useful case when $Y_i \overset{iid}{\sim} N(\mu, 1)$ does not fall within the theory. No closed form results are known to date.
The assumption that Y_i should be univariate random variables is also too restrictive. One should consider multivariate analogs. (I just have not had the time to devote to this problem. This just might work out.)
If one wants to be bold, (at least via simulations) one may want to relax the iid assumption. (My computer skills are just too limited to perform this task.)

Find some real applications of the theory to the current IMA theme.