Imperial College London

Sequential Implementation of Monte Carlo Tests with Uniformly Bounded Resampling Risk

Axel Gandy

Department of Mathematics
Imperial College London
a.gandy@imperial.ac.uk

useR! 2009, Rennes
July 8-10, 2009

Introduction

- Test statistic T, reject for large values.
- Observation: t.
- p-value:

$$p = \mathrm{P}(T \geq t)$$

  Often not available in closed form.
- Monte Carlo Test:

$$\hat{p}_{\text{naive}} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(T_i \geq t),$$

  where $T, T_1, \ldots T_n$ i.i.d.
- Examples:
  - Bootstrap,
  - Permutation tests.
- Goal: Estimate p using few X_i

  Mainly interested in deciding if $p \leq \alpha$ for some α.

Introduction

- Test statistic T, reject for large values.
- Observation: t.
- p-value:

$$p = \mathrm{P}(T \geq t)$$

  Often not available in closed form.
- Monte Carlo Test:

$$\hat{p}_{\text{naive}} = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\mathbb{I}(T_i \geq t)}_{=:X_i \sim B(1,p)},$$

  where $T, T_1, \dots T_n$ i.i.d.
- Examples:
  - Bootstrap,
  - Permutation tests.
- Goal: Estimate p using few X_i

  Mainly interested in deciding if $p \leq \alpha$ for some α.

Introduction

- Test statistic T, reject for large values.
- Observation: t.
- p-value:

$$p = \mathrm{P}(T \geq t)$$

Often not available in closed form.
- Monte Carlo Test:

$$\hat{p}_{\mathrm{naive}} = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\mathbb{I}(T_i \geq t)}_{=:X_i \sim B(1,p)},$$

where $T, T_1, \dots T_n$ i.i.d.
- Examples:
  - Bootstrap,
  - Permutation tests.
- Goal: Estimate p using few X_i

Mainly interested in deciding if $p \leq \alpha$ for some α.

Introduction

- Test statistic T, reject for large values.
- Observation: t.
- p-value:

$$p = \mathrm{P}(T \geq t)$$

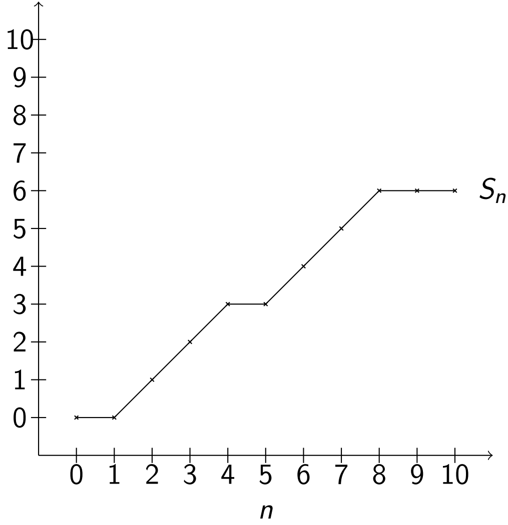
  Often not available in closed form.
- Monte Carlo Test:

$$\hat{p}_{\text{naive}} = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\mathbb{I}(T_i \geq t)}_{=:X_i \sim B(1,p)},$$

  where $T, T_1, \ldots T_n$ i.i.d.
- Examples:
  - Bootstrap,
  - Permutation tests.
- Goal: Estimate p using few X_i

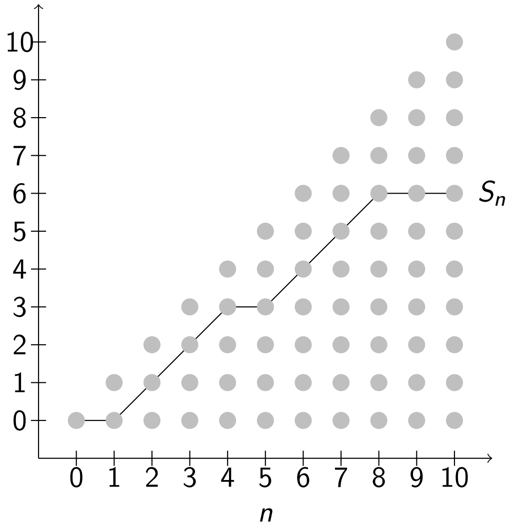
Mainly interested in deciding if $p \leq \alpha$ for some α.

Sequential approaches based on $S_n = \sum_{i=1}^n X_i$



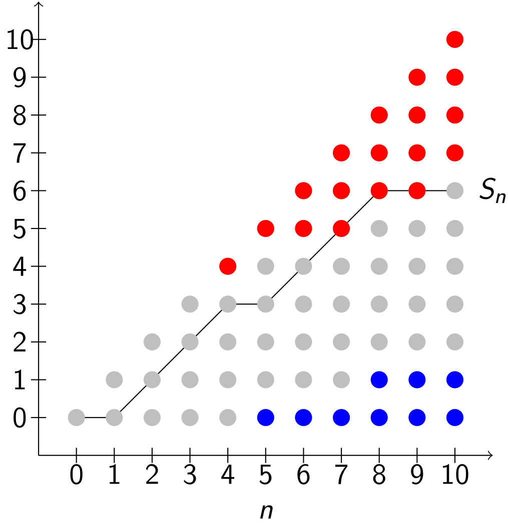
- Stop once $S_n \geq U_n$ or $S_n \leq L_n$
- τ: hitting time
- Compute $\hat{p}$ based on S_τ and τ.
- Hit B_U: decide $p > \alpha$,
- Hit B_L: decide $p \leq \alpha$,

Sequential approaches based on $S_n = \sum_{i=1}^{n} X_i$



- Stop once $S_n \geq U_n$ or $S_n \leq L_n$
- τ: hitting time
- Compute $\hat{p}$ based on S_τ and τ.
- Hit B_U: decide $p > \alpha$,
- Hit B_L: decide $p \leq \alpha$,

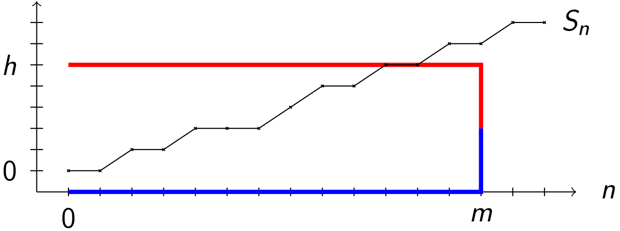
Sequential approaches based on $S_n = \sum_{i=1}^n X_i$



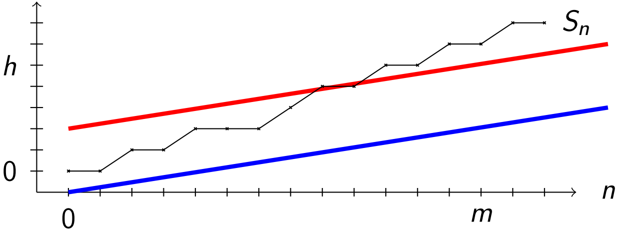
- Stop once $S_n \geq U_n$ or $S_n \leq L_n$
- τ: hitting time
- Compute $\hat{p}$ based on S_τ and τ.
- Hit B_U: decide $p > \alpha$,
- Hit B_L: decide $p \leq \alpha$,

Previous Approaches

- Besag & Clifford (1991):



- (Truncated) Sequential Probability Ratio Test, Fay et al. (2007)



- R-package MChtest.

What do we really want?

Is $p \leq \alpha$?

Two individuals using the same statistical method on the same data should arrive at the same conclusion.
First law of applied statistics, Gleser (1996)

Consider the resampling risk

$$\mathrm{RR}_p(\hat{p}) \equiv \begin{cases} \mathrm{P}_p(\hat{p} > \alpha) & \text{if } p \leq \alpha, \\ \mathrm{P}_p(\hat{p} \leq \alpha) & \text{if } p > \alpha. \end{cases}$$

Want:

$$\sup_{p \in [0,1]} \mathrm{RR}_p(\hat{p}) \leq \epsilon$$

for some (small) $\epsilon > 0$.
For Besag & Clifford (1991), SPRT: $\sup_p \mathrm{RR}_P \geq 0.5$

What do we really want?

Is $p \leq \alpha$?

Two individuals using the same statistical method on the same data should arrive at the same conclusion.
First law of applied statistics, Gleser (1996)

Consider the resampling risk

$$\mathrm{RR}_p(\hat{p}) \equiv \begin{cases} \mathrm{P}_p(\hat{p} > \alpha) & \text{if } p \leq \alpha, \\ \mathrm{P}_p(\hat{p} \leq \alpha) & \text{if } p > \alpha. \end{cases}$$

Want:

$$\sup_{p \in [0,1]} \mathrm{RR}_p(\hat{p}) \leq \epsilon$$

for some (small) $\epsilon > 0$.
For Besag & Clifford (1991), SPRT: $\sup_p \mathrm{RR}_P \geq 0.5$

What do we really want?

Is $p \le \alpha$?

Two individuals using the same statistical method on the same data should arrive at the same conclusion.
First law of applied statistics, Gleser (1996)

Consider the resampling risk

$$\mathrm{RR}_p(\hat{p}) \equiv \begin{cases} \mathrm{P}_p(\hat{p} > \alpha) & \text{if } p \le \alpha, \\ \mathrm{P}_p(\hat{p} \le \alpha) & \text{if } p > \alpha. \end{cases}$$

Want:

$$\sup_{p \in [0,1]} \mathrm{RR}_p(\hat{p}) \le \epsilon$$

for some (small) $\epsilon > 0$.
For Besag & Clifford (1991), SPRT: $\sup_p \mathrm{RR}_P \ge 0.5$

What do we really want?

Is $p \leq \alpha$?

Two individuals using the same statistical method on the same data should arrive at the same conclusion.

First law of applied statistics, Gleser (1996)

Consider the resampling risk

$$\mathrm{RR}_p(\hat{p}) \equiv \begin{cases} \mathrm{P}_p(\hat{p} > \alpha) & \text{if } p \leq \alpha, \\ \mathrm{P}_p(\hat{p} \leq \alpha) & \text{if } p > \alpha. \end{cases}$$

Want:

$$\sup_{p \in [0,1]} \mathrm{RR}_p(\hat{p}) \leq \epsilon$$

for some (small) $\epsilon > 0$.

For Besag & Clifford (1991), SPRT: $\sup_p \mathrm{RR}_P \geq 0.5$

What do we really want?

Is $p \leq \alpha$?

Two individuals using the same statistical method on the same data should arrive at the same conclusion.

First law of applied statistics, Gleser (1996)

Consider the resampling risk

$$\mathrm{RR}_p(\hat{p}) \equiv \begin{cases} \mathrm{P}_p(\hat{p} > \alpha) & \text{if } p \leq \alpha, \\ \mathrm{P}_p(\hat{p} \leq \alpha) & \text{if } p > \alpha. \end{cases}$$

Want:

$$\sup_{p \in [0,1]} \mathrm{RR}_p(\hat{p}) \leq \epsilon$$

for some (small) $\epsilon > 0$.

For Besag & Clifford (1991), SPRT: $\sup_p \mathrm{RR}_P \geq 0.5$

What do we really want?

Is $p \leq \alpha$?

Two individuals using the same statistical method on the same data should arrive at the same conclusion.

First law of applied statistics, Gleser (1996)

Consider the resampling risk

$$\mathrm{RR}_p(\hat{p}) \equiv \begin{cases} \mathrm{P}_p(\hat{p} > \alpha) & \text{if } p \leq \alpha, \\ \mathrm{P}_p(\hat{p} \leq \alpha) & \text{if } p > \alpha. \end{cases}$$

Want:

$$\sup_{p \in [0,1]} \mathrm{RR}_p(\hat{p}) \leq \epsilon$$

for some (small) $\epsilon > 0$.

For Besag & Clifford (1991), SPRT: $\sup_p \mathrm{RR}_P \geq 0.5$

Recursive Definition of the Boundaries

Want:

$$\sup_p \mathrm{RR}_p(\hat{p}) \leq \epsilon$$

Suffices to ensure

$$\mathrm{P}_\alpha(\text{hit } B_U) \leq \epsilon$$
$$\mathrm{P}_\alpha(\text{hit } B_L) \leq \epsilon$$

Recursive definition:
Given $U_1, \dots, U_{n-1}$ and $L_1, \dots, L_{n-1}$, define

- U_n as the minimal value such that

$$\mathrm{P}_\alpha(\text{hit } B_U \text{ until } n) \leq \epsilon_n$$

- and L_n as the maximal value such that

$$\mathrm{P}_\alpha(\text{hit } B_L \text{ until } n) \leq \epsilon_n$$

where $\epsilon_n \geq 0$ with $\epsilon_n \nearrow \epsilon$ (spending sequence).

Recursive Definition of the Boundaries

Want:

$$\sup_p \mathrm{RR}_p(\hat{p}) \leq \epsilon$$

Suffices to ensure

$$\mathrm{P}_\alpha(\text{hit } B_U) \leq \epsilon$$
$$\mathrm{P}_\alpha(\text{hit } B_L) \leq \epsilon$$

Recursive definition:

Given $U_1, \dots, U_{n-1}$ and $L_1, \dots, L_{n-1}$, define

- U_n as the minimal value such that

$$\mathrm{P}_\alpha(\text{hit } B_U \text{ until } n) \leq \epsilon_n$$

- and L_n as the maximal value such that

$$\mathrm{P}_\alpha(\text{hit } B_L \text{ until } n) \leq \epsilon_n$$

where $\epsilon_n \geq 0$ with $\epsilon_n \nearrow \epsilon$ (spending sequence).

Recursive Definition of the Boundaries

Want:

$$\sup_p \mathrm{RR}_p(\hat{p}) \leq \epsilon$$

Suffices to ensure

$$\mathrm{P}_\alpha(\text{hit } B_U) \leq \epsilon$$

$$\mathrm{P}_\alpha(\text{hit } B_L) \leq \epsilon$$

Recursive definition:

Given $U_1, \dots, U_{n-1}$ and $L_1, \dots, L_{n-1}$, define

- U_n as the minimal value such that

$$\mathrm{P}_\alpha(\text{hit } B_U \text{ until } n) \leq \epsilon_n$$

- and L_n as the maximal value such that

$$\mathrm{P}_\alpha(\text{hit } B_L \text{ until } n) \leq \epsilon_n$$

where $\epsilon_n \geq 0$ with $\epsilon_n \nearrow \epsilon$ (spending sequence).

Recursive Definition of the Boundaries

Want:

$$\sup_p \mathrm{RR}_p(\hat{p}) \leq \epsilon$$

Suffices to ensure

$$\mathrm{P}_\alpha(\text{hit } B_U) \leq \epsilon$$

$$\mathrm{P}_\alpha(\text{hit } B_L) \leq \epsilon$$

Recursive definition:

Given $U_1, \dots, U_{n-1}$ and $L_1, \dots, L_{n-1}$, define

- U_n as the minimal value such that

$$\mathrm{P}_\alpha(\text{hit } B_U \text{ until } n) \leq \epsilon_n$$

- and L_n as the maximal value such that

$$\mathrm{P}_\alpha(\text{hit } B_L \text{ until } n) \leq \epsilon_n$$

where $\epsilon_n \geq 0$ with $\epsilon_n \nearrow \epsilon$ (spending sequence).

Recursive Definition - Example

- $\alpha = 0.2$, $\epsilon_n = 0.4\frac{n}{5+n}$.
- U_n=the minimal value such that

$$\mathrm{P}_\alpha(\text{hit } B_U \text{ until } n) \leq \epsilon_n$$

- L_n = maximal value such that

$$\mathrm{P}_\alpha(\text{hit } B_L \text{ until } n) \leq \epsilon_n$$

	$n =$
$\mathrm{P}_\alpha(S_n = k, \tau \geq n)$	0
k= 3	
k= 2	
k= 1	
k= 0	1
ϵ_n	0
U_n	1
L_n	-1

Recursive Definition - Example

- $\alpha = 0.2$, $\epsilon_n = 0.4\frac{n}{5+n}$.
- U_n=the minimal value such that

$$\mathrm{P}_\alpha(\text{hit } B_U \text{ until } n) \leq \epsilon_n$$

- L_n = maximal value such that

$$\mathrm{P}_\alpha(\text{hit } B_L \text{ until } n) \leq \epsilon_n$$

	$n =$	
$\mathrm{P}_\alpha(S_n = k, \tau \geq n)$	0	1
k= 3		
k= 2		
k= 1		.2
k= 0	1	.8
ϵ_n	0	.07
U_n	1	2
L_n	-1	-1

Recursive Definition - Example

- $\alpha = 0.2$, $\epsilon_n = 0.4\frac{n}{5+n}$.
- U_n=the minimal value such that

$$P_\alpha(\text{hit } B_U \text{ until } n) \le \epsilon_n$$

- L_n = maximal value such that

$$P_\alpha(\text{hit } B_L \text{ until } n) \le \epsilon_n$$

	$n =$		
$P_\alpha(S_n = k, \tau \ge n)$	0	1	2
k= 3			
k= 2			.04
k= 1		.2	.32
k= 0	1	.8	.64
ϵ_n	0	.07	.11
U_n	1	2	2
L_n	-1	-1	-1

Recursive Definition - Example

- $\alpha = 0.2$, $\epsilon_n = 0.4\frac{n}{5+n}$.
- U_n=the minimal value such that

$$\mathrm{P}_\alpha(\text{hit } B_U \text{ until } n) \leq \epsilon_n$$

- L_n = maximal value such that

$$\mathrm{P}_\alpha(\text{hit } B_L \text{ until } n) \leq \epsilon_n$$

	$n =$			
$\mathrm{P}_\alpha(S_n = k, \tau \geq n)$	0	1	2	3
k= 3				
k= 2			.04	.06
k= 1		.2	.32	.38
k= 0	1	.8	.64	.51
ϵ_n	0	.07	.11	.15
U_n	1	2	2	2
L_n	-1	-1	-1	-1

Recursive Definition - Example

- $\alpha = 0.2$, $\epsilon_n = 0.4\frac{n}{5+n}$.
- U_n=the minimal value such that

$$P_\alpha(\text{hit } B_U \text{ until } n) \leq \epsilon_n$$

- L_n = maximal value such that

$$P_\alpha(\text{hit } B_L \text{ until } n) \leq \epsilon_n$$

	$n =$				
$P_\alpha(S_n = k, \tau \geq n)$	0	1	2	3	4
k= 3					
k= 2			.04	.06	.08
k= 1		.2	.32	.38	.41
k= 0	1	.8	.64	.51	.41
ϵ_n	0	.07	.11	.15	.18
U_n	1	2	2	2	3
L_n	-1	-1	-1	-1	-1

Recursive Definition - Example

- $\alpha = 0.2$, $\epsilon_n = 0.4\frac{n}{5+n}$.
- U_n=the minimal value such that

$$P_\alpha(\text{hit } B_U \text{ until } n) \leq \epsilon_n$$

- L_n = maximal value such that

$$P_\alpha(\text{hit } B_L \text{ until } n) \leq \epsilon_n$$

	$n=$					
$P_\alpha(S_n = k, \tau \geq n)$	0	1	2	3	4	5
k= 3						.02
k= 2			.04	.06	.08	.14
k= 1		.2	.32	.38	.41	.41
k= 0	1	.8	.64	.51	.41	.33
ϵ_n	0	.07	.11	.15	.18	.20
U_n	1	2	2	2	3	3
L_n	-1	-1	-1	-1	-1	-1

Recursive Definition - Example

- $\alpha = 0.2$, $\epsilon_n = 0.4\frac{n}{5+n}$.
- U_n=the minimal value such that

$$P_\alpha(\text{hit } B_U \text{ until } n) \leq \epsilon_n$$

- L_n = maximal value such that

$$P_\alpha(\text{hit } B_L \text{ until } n) \leq \epsilon_n$$

	$n =$						
$P_\alpha(S_n = k, \tau \geq n)$	0	1	2	3	4	5	6
k= 3						.02	.03
k= 2			.04	.06	.08	.14	.20
k= 1		.2	.32	.38	.41	.41	.39
k= 0	1	.8	.64	.51	.41	.33	.26
ϵ_n	0	.07	.11	.15	.18	.20	.22
U_n	1	2	2	2	3	3	3
L_n	-1	-1	-1	-1	-1	-1	-1

Recursive Definition - Example

- $\alpha = 0.2$, $\epsilon_n = 0.4\frac{n}{5+n}$.
- U_n=the minimal value such that

$$\mathrm{P}_\alpha(\text{hit } B_U \text{ until } n) \leq \epsilon_n$$

- L_n = maximal value such that

$$\mathrm{P}_\alpha(\text{hit } B_L \text{ until } n) \leq \epsilon_n$$

	$n=$							
$\mathrm{P}_\alpha(S_n = k, \tau \geq n)$	0	1	2	3	4	5	6	7
k= 3						.02	.03	.04
k= 2			.04	.06	.08	.14	.20	.24
k= 1		.2	.32	.38	.41	.41	.39	.37
k= 0	1	.8	.64	.51	.41	.33	.26	.21
ϵ_n	0	.07	.11	.15	.18	.20	.22	.23
U_n	1	2	2	2	3	3	3	3
L_n	-1	-1	-1	-1	-1	-1	-1	0

Recursive Definition - Example

- $\alpha = 0.2$, $\epsilon_n = 0.4\frac{n}{5+n}$.
- U_n=the minimal value such that

$$\mathrm{P}_\alpha(\text{hit } B_U \text{ until } n) \leq \epsilon_n$$

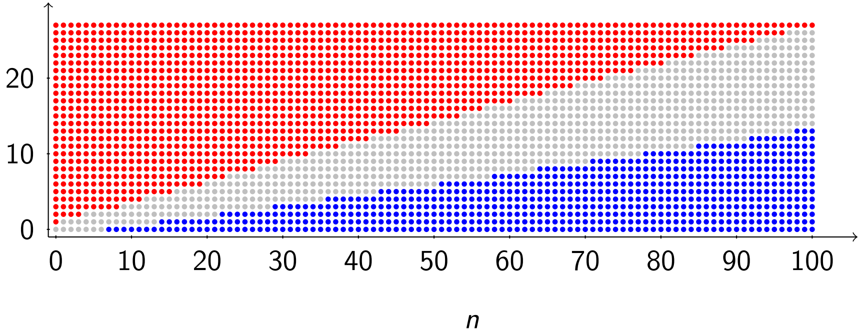
- L_n = maximal value such that

$$\mathrm{P}_\alpha(\text{hit } B_L \text{ until } n) \leq \epsilon_n$$

	$n =$								
$\mathrm{P}_\alpha(S_n = k, \tau \geq n)$	0	1	2	3	4	5	6	7	8
k= 3						.02	.03	.04	.05
k= 2			.04	.06	.08	.14	.20	.24	.26
k= 1		.2	.32	.38	.41	.41	.39	.37	.29
k= 0	1	.8	.64	.51	.41	.33	.26	.21	
ϵ_n	0	.07	.11	.15	.18	.20	.22	.23	.25
U_n	1	2	2	2	3	3	3	3	3
L_n	-1	-1	-1	-1	-1	-1	-1	0	0

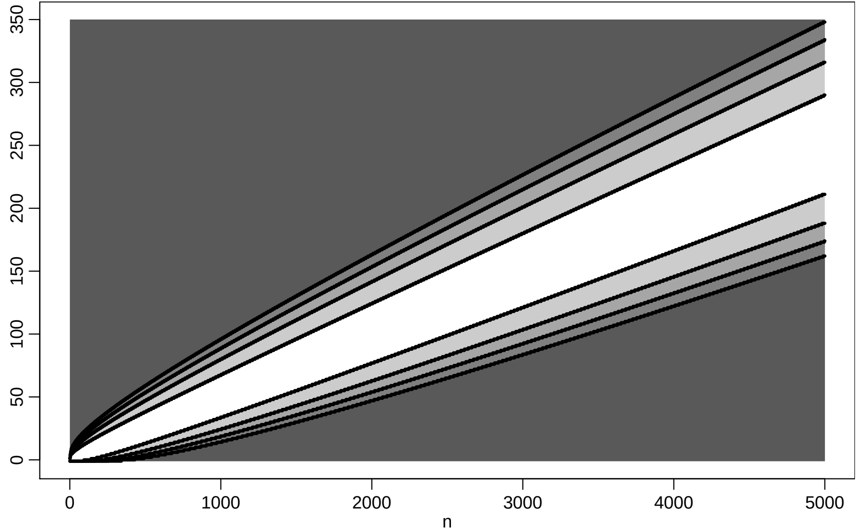
Sequential Decision Procedure - Example

$\alpha = 0.2$, $\epsilon_n = 0.4\frac{n}{5+n}$.



Influence of ϵ on the stopping rule

$\epsilon = 0.1,\ 0.001,\ 10^{-5},\ 10^{-7};\ \epsilon_n = \epsilon\frac{n}{1000+n}$



Sequential Estimation based on the MLE

$$\hat{p} = \begin{cases} \frac{S_\tau}{\tau}, & \tau < \infty \\ \alpha, & \tau = \infty, \end{cases}$$

- One can show:
  - hitting the upper boundary implies $\hat{p} > \alpha$,
  - hitting the lower boundary implies $\hat{p} < \alpha$.

  Hence,

$$\sup_p \mathrm{RR}_p(\hat{p}) \leq \epsilon$$

- Furthermore, $\exists$ random interval I_n s.t.
  - I_n only depends on $X_1, \dots, X_n$,
  - $\hat{p} \in I_n$.

Example - Two-way sparse contingency table

1	2	2	1	1	0	1
2	0	0	2	3	0	0
0	1	1	1	2	7	3
1	1	2	0	0	0	1
0	1	1	1	1	0	0

- H_0: variables are independent.
- Reject for large values of the likelihood ratio test statistic T
- $T \xrightarrow{d} \chi^2_{(7-1)(5-1)}$ under H_0. Based on this: $p = 0.031$.
- Matrix sparse - approximation poor?
- Use parametric bootstrap based on row and column sums.
- Naive test statistic $\hat{p}_{naive}$ with $n = 1{,}000$ replicates:
  $p = 0.041 < 0.05$.
  Probability of reporting $p > 0.05$: roughly 0.08.

Example - Two-way sparse contingency table

1	2	2	1	1	0	1
2	0	0	2	3	0	0
0	1	1	1	2	7	3
1	1	2	0	0	0	1
0	1	1	1	1	0	0

- H_0: variables are independent.
- Reject for large values of the likelihood ratio test statistic T
- $T \xrightarrow{d} \chi^2_{(7-1)(5-1)}$ under H_0. Based on this: $p = 0.031$.
- Matrix sparse - approximation poor?
- Use parametric bootstrap based on row and column sums.
- Naive test statistic $\hat{p}_{naive}$ with $n = 1{,}000$ replicates:
  $p = 0.041 < 0.05$.
  Probability of reporting $p > 0.05$: roughly 0.08.

Example - Two-way sparse contingency table

1	2	2	1	1	0	1
2	0	0	2	3	0	0
0	1	1	1	2	7	3
1	1	2	0	0	0	1
0	1	1	1	1	0	0

- H_0: variables are independent.
- Reject for large values of the likelihood ratio test statistic T
- $T \xrightarrow{d} \chi^2_{(7-1)(5-1)}$ under H_0. Based on this: $p = 0.031$.
- Matrix sparse - approximation poor?
- Use parametric bootstrap based on row and column sums.
- Naive test statistic $\hat{p}_{naive}$ with $n = 1{,}000$ replicates:
  $p = 0.041 < 0.05$.
  Probability of reporting $p > 0.05$: roughly 0.08.

Example - Bootstrap and Sequential Algorithm

```
> dat <- matrix(c(1,2,2,1,1,0,1, 2,0,0,2,3,0,0, 0,1,1,1,2,7,3, 1,1,2,0,0,0,1,
+                 0,1,1,1,1,0,0), nrow=5,ncol=7,byrow=TRUE)
> loglikrat <- function(data){
+   cs <- colSums(data);rs <- rowSums(data); mu <- outer(rs,cs)/sum(rs)
+   2*sum(ifelse(data<=0.5, 0,data*log(data/mu)))
+ }
> resample <- function(data){
+   cs <- colSums(data);rs <- rowSums(data); n <- sum(rs)
+   mu <- outer(rs,cs)/n/n
+   matrix(rmultinom(1,n,c(mu)),nrow=dim(data)[1],ncol=dim(data)[2])
+ }
> t <- loglikrat(dat);
> library(simctest)
> res <- simctest(function(){loglikrat(resample(dat))>=t},maxsteps=1000)
> res
No decision reached.
Final estimate will be in [ 0.02859135 , 0.07965451 ]
Current estimate of the p.value: 0.041
Number of samples: 1000
> cont(res, steps=10000)
p.value: 0.04035456
Number of samples: 8574
```

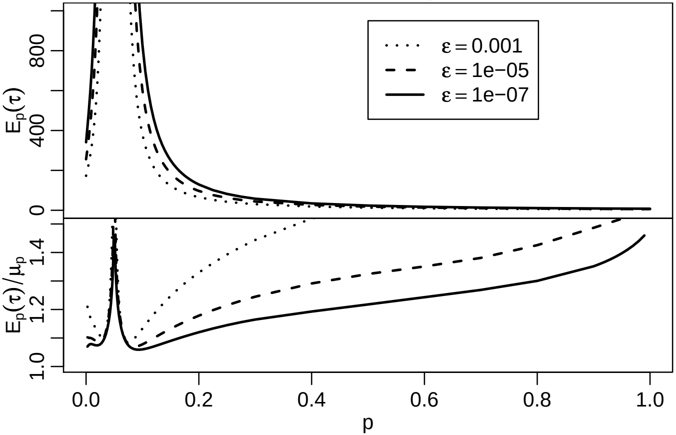
Further Uses of the Algorithm

- Simulation study to evaluate whether a test is liberal/conservative.
- Determining the sample size to achieve a certain power.
- Iterated Use:
  - Determining the power of a bootstrap test.
  - Simulation study to evaluate whether a bootstrap test is liberal/conservative.
  - Double bootstrap test.

Expected Hitting Time

Result: $E_p(\tau) < \infty \ \forall p \neq \alpha$

Example with $\alpha = 0.05$, $\epsilon_n = \epsilon \frac{n}{1000+n}$:



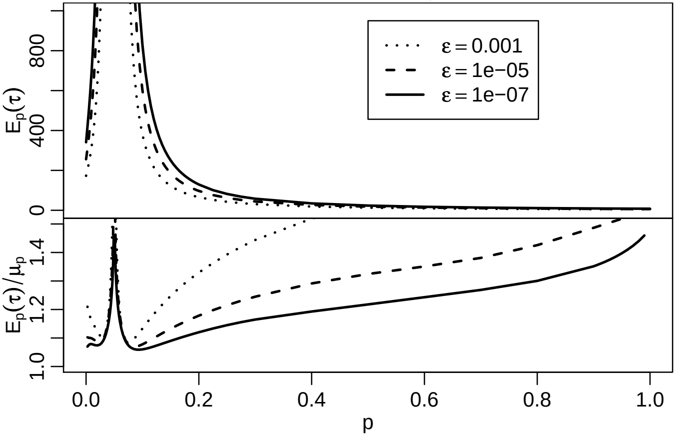
μ_p = theoretical lower bound on $E_p(\tau)$.

- Note: $\int_0^1 \mu_p dp = \infty$;
- for iterated use: Need to limit the number of steps.

Expected Hitting Time

Result: $\mathrm{E}_p(\tau) < \infty \; \forall p \neq \alpha$

Example with $\alpha = 0.05$, $\epsilon_n = \epsilon \frac{n}{1000+n}$:



μ_p = theoretical lower bound on $\mathrm{E}_p(\tau)$.

- Note: $\int_0^1 \mu_p dp = \infty$;
- for iterated use: Need to limit the number of steps.

Summary

- Sequential implementation of Monte Carlo Tests and computation of p-values.
- Useful when implementing tests in packages.
- After a finite number of steps:
  - $\hat{p}$ or
  - interval $[\hat{p}_n^L, \hat{p}_n^U]$ in which $\hat{p}$ will lie.
- Guarantee (up to a very small error probability):

$$\hat{p} \text{ is on the ``correct side'' of } \alpha.$$

- R-package simctest available on CRAN.
  (efficient implementation with C-code)
- For details see Gandy (2009).

References

Besag, J. & Clifford, P. (1991). Sequential Monte Carlo p-values. *Biometrika* **78**, 301–304.

Davison, A. & Hinkley, D. (1997). *Bootstrap methods and their application*. Cambridge University Press.

Fay, M. P., Kim, H.-J. & Hachey, M. (2007). On using truncated sequential probability ratio test boundaries for Monte Carlo implementation of hypothesis tests. *Journal of Computational & Graphical Statistics* **16**, 946 – 967.

Gandy, A. (2009). Sequential implementation of Monte Carlo tests with uniformly bounded resampling risk. Accepted for publication in JASA.

Gleser, L. J. (1996). Comment on *Bootstrap Confidence Intervals* by T. J. DiCiccio and B. Efron. *Statistical Science* **11**, 219–221.