The Maximum of a Random Walk
Reflected at a General Barrier

DYNSTOCH workshop, Copenhagen 2004

Niels Hansen
The Maximum of a Random Walk Reflected at a General Barrier

DYNSTOCH workshop, Copenhagen 2004

Niels Hansen

Definitions and precise results about the (tail of) the distribution of a random walk reflected at general barriers.
The Maximum of a Random Walk Reflected at a General Barrier

DYNSTOCH workshop, Copenhagen 2004

Niels Hansen

- Definitions and precise results about the (tail of) the distribution of a random walk reflected at general barriers.
- Motivation and applications from structural biology to the search for specific structures in genomes.
Reflection of a random walk

Let $X_1, X_2, \ldots$ be iid real valued random variables with $\mathbb{E}(X_1) < 0$. Let

$$S_n = \sum_{k=1}^{n} X_k$$

and for $g : \mathbb{N} \rightarrow (-\infty, 0]$ given

$$W_n^g = \max\{W_{n-1}^g + X_n, g(n)\} \quad (W_0^g = 0).$$
Reflection of a random walk

Let $X_1, X_2, \ldots$ be iid real valued random variables with $\mathbb{E}(X_1) < 0$. Let

$$S_n = \sum_{k=1}^{n} X_k$$

and for $g : \mathbb{N} \rightarrow (-\infty, 0]$ given

$$W_n^g = \max\{W_{n-1}^g + X_n, g(n)\} \quad (W_0^g = 0).$$

We call $(W_n^g)_{n \geq 0}$ the reflection of the random walk at the barrier given by $g$. It is a non-homogeneous Markov chain.
Examples

With $g \equiv 0$ we obtain the reflection at the zero barrier.

A queueing theoretical interpretation of $W^g_n$ is the waiting time for costumer $n$ at the time of arrival.
Examples

With $g \equiv 0$ we obtain the reflection at the zero barrier.

A queueing theoretical interpretation of $W_n^g$ is the waiting time for customer $n$ at the time of arrival.

We are more interested in:

- The linear barrier $g(n) = -\alpha n$ for $\alpha > 0$.
- The logarithmic barrier $g(n) = -\rho \log n$ for $\rho > 0$.
- General barriers $g$ with $g(n) \to -\infty$ for $n \to \infty$. 
Three examples

\[ g(n) = 0 \]

\[ g(n) = -15 \log(n) \]

\[ g(n) = -n \]
Questions

- When is the maximum of the reflected random walk finite?
- If finite, what is the tail behaviour of the distribution of the maximum?
Questions

- When is the maximum of the reflected random walk finite?
- If finite, what is the tail behaviour of the distribution of the maximum?

We will assume that

$$\mathbb{E}(\exp(\theta X_1)) = 1$$

has a solution $\theta^* > 0$.

We define the exponentially tilted probability measure $\mathbb{P}^*$ under which $X_1, X_2, \ldots$ are iid with

$$\mathbb{P}^*(X_1 \in A) = \int_A \exp(\theta^* x) X_1(\mathbb{P})(dx).$$
Define

\[ M^g = \max_{n \geq 0} W^g_n \]

and

\[ D = \max_{n \geq 0} \{g(n) - S_n\}. \]
Define

\[ M^g = \max_{n \geq 0} W^g_n \]

and

\[ D = \max_{n \geq 0} \{ g(n) - S_n \} \].

**Theorem 1** It holds that

\[ \mathbb{P}(M^g > u) \leq \exp(-\theta^* u) \mathbb{E}^*(\exp(\theta^* D)) \],

and \( \mathbb{P}(M^g < \infty) = 1 \) if and only if

\[ \mathbb{E}^*(\exp(\theta^* D)) < \infty \].
Define the ascending ladder height distribution under $\mathbb{P}^*$ by

$$G^*_+(x) = \mathbb{P}^*(S_{\tau_+} \leq x)$$

with $\tau_+ = \inf\{n \mid S_n > 0\}$. 
Define the ascending ladder height distribution under $\mathbb{P}^{*}$ by

$$G^{*}_+(x) = \mathbb{P}^{*}(S_{\tau_{+}} \leq x)$$

with $\tau_{+} = \inf\{n \mid S_n > 0\}$.

**Theorem 2** If $\mathbb{E}^{*}(\exp(\theta^{*}D)) < \infty$, if the distribution of $X_1$ is nonarithmetic, and if $B$ is a stochastic variable with distribution

$$\mathbb{P}^{*}(B \leq x) = \frac{1}{\mathbb{E}^{*}(S_{\tau_{+}})} \int_{0}^{x} 1 - G^{*}_+(y) \, dy,$$

then

$$\mathbb{P}(\mathcal{M}^g > u) \sim \exp(-\theta^{*}u)\mathbb{E}^{*}(\exp(\theta^{*}D))\mathbb{E}^{*}(\exp(-\theta^{*}B))$$

for $u \to \infty$. 
Outline of proof

First, under $\mathbb{P}^*$

$$B_u := W^{g}_{\tau(u)} - u \overset{\mathcal{D}}{\to} B$$

with $\tau(u) = \inf\{n \geq 0 \mid W^{g}_n > u\}$. 
Outline of proof

First, under $\mathbb{P}^*$

$$B_u := W^g_{\tau(u)} - u \xrightarrow{\mathcal{D}} B$$

with $\tau(u) = \inf\{n \geq 0 \mid W^g_n > u\}$. Second,

$$D_u := W^g_{\tau(u)} - S_{\tau(u)} = \max_{0 \leq k \leq \tau(u)} \{g(k) - S_k\} \nearrow D$$

for $u \to \infty$ under $\mathbb{P}^*$. 
Outline of proof

First, under $\mathbb{P}^*$

$$B_u := W^g_{\tau(u)} - u \overset{D}{\to} B$$

with $\tau(u) = \inf\{n \geq 0 \mid W^g_n > u\}$. Second,

$$D_u := W^g_{\tau(u)} - S_{\tau(u)} = \max_{0 \leq k \leq \tau(u)} \{g(k) - S_k\} \uparrow D$$

for $u \to \infty$ under $\mathbb{P}^*$. Finally, $D$ and $B$ are asymptotically independent. Changing measure to $\mathbb{P}^*$ gives a proof of Theorems 2:

$$\mathbb{P}(\mathcal{M}^g > u) = \exp(-\theta^* u) \mathbb{E}^*(\exp(-\theta^*(S_{\tau(u)} - u)))$$
Outline of proof

First, under $\mathbb{P}^*$

$$B_u := W^g_{\tau(u)} - u \xrightarrow{\mathcal{D}} B$$

with $\tau(u) = \inf\{n \geq 0 \mid W_n^g > u\}$. Second,

$$D_u := W^g_{\tau(u)} - S_{\tau(u)} = \max_{0 \leq k \leq \tau(u)} \{g(k) - S_k\} \xrightarrow{\mathcal{D}} D$$

for $u \to \infty$ under $\mathbb{P}^*$. Finally, $D$ and $B$ are asymptotically independent.

Changing measure to $\mathbb{P}^*$ gives a proof of Theorems 2:

$$\mathbb{P}(M^g > u) = \exp(-\theta^* u) \mathbb{E}^*(\exp(-\theta^*(S_{\tau(u)} - u)))$$

$$= \exp(-\theta^* u) \mathbb{E}^*(\exp(\theta^* D_u) \exp(-\theta^* B_u))$$
Outline of proof

First, under $\mathbb{P}^*$

$$B_u := W^g_{\tau(u)} - u \overset{\mathcal{D}}{\to} B$$

with $\tau(u) = \inf\{n \geq 0 \mid W^g_n > u\}$. Second,

$$D_u := W^g_{\tau(u)} - S_{\tau(u)} = \max_{0 \leq k \leq \tau(u)} \{g(k) - S_k\} \overset{\mathbb{P}^*}{\to} D$$

for $u \to \infty$ under $\mathbb{P}^*$. Finally, $D$ and $B$ are asymptotically independent. Changing measure to $\mathbb{P}^*$ gives a proof of Theorems 2:

$$\mathbb{P}(\mathcal{M}^g > u) = \exp(-\theta^* u) \mathbb{E}^*(\exp(-\theta^* (S_{\tau(u)} - u)))$$

$$= \exp(-\theta^* u) \mathbb{E}^*(\exp(\theta^* D_u) \exp(-\theta^* B_u))$$

$$\sim \exp(-\theta^* u) \mathbb{E}^*(\exp(\theta^* D)) \mathbb{E}^*(\exp(-\theta^* B)).$$
Examples - continued

When is $E^*(\exp(\theta^* D)) < \infty$?
Examples - continued

When is $\mathbb{E}^*(\exp(\theta^* D)) < \infty$?

Example 1: If $g(n) = -\alpha n$ is a linear barrier with $\alpha < 0$, $D$ is the maximum of a random walk with drift $\alpha - \mathbb{E}^*(X_1) < 0$. Then

$$D \overset{\text{D}}{=} \sum_{k=1}^{\rho} U_k$$

with $\rho$ a geometric random variable and $U_n$'s iid independent of $\rho$ (the Pollaczeck-Khinchine formula). In this case $\mathbb{E}^*(\exp(\theta^* D)) < \infty$ always holds.
Example 2: If $g$ is super-linearly decaying, $g(n) \leq -\alpha n$ for some $\alpha > 0$ like for an affine barrier $g(n) = -\beta - \alpha n$ or a quadratic barrier $g(n) = -\alpha n^2$, then $\mathbb{E}^*(\exp(\theta^* D)) < \infty$ always holds.
Example 2: If $g$ is super-linearly decaying, $g(n) \leq -\alpha n$ for some $\alpha > 0$ like for an affine barrier $g(n) = -\beta - \alpha n$ or a quadratic barrier $g(n) = -\alpha n^2$, then $E^*(\exp(\theta^* D)) < \infty$ always holds.

Example 3: If $g(n) = -\rho \log(n)$ it is possible to show that $M^g = \infty$ a.s. if $\rho < 1/\theta^*$ and $E^*(\exp(\theta^* D)) < \infty$ (hence $M^g < \infty$ a.s) if $\rho > 1/\theta^*$. 
RNA molecules are transcribed from the DNA-genome. Some folds into functionally important structures.

An RNA-molecule is represented as a sequence of letters from the alphabet \{A, C, G, U\}. 
RNA molecules are transcribed from the DNA-genome. Some folds into functionally important structures.

An RNA-molecule is represented as a sequence of letters from the alphabet \{A, C, G, U\}.

An example RNA-molecule from the nematode *C. elegans*. 
If $y = y_1 \ldots y_n$ is a (very, very long) sequence from $\{A, C, G, U\}$, we search for \textit{local folding structures}. We proceed as follows:

Choose a function $f : \{A, C, G, U\}^2 \rightarrow \mathbb{R}$. 


If $y = y_1 \ldots y_n$ is a (very, very long) sequence from $\{A, C, G, U\}$, we search for local folding structures. We proceed as follows:

- Choose a function $f : \{A, C, G, U\}^2 \rightarrow \mathbb{R}$.
- Pick a pair $(i, j)$ with $i < j$. 
If \( y = y_1 \ldots y_n \) is a (very, very long) sequence from \( \{A, C, G, U\} \), we search for \textcolor{red}{local folding structures}. We proceed as follows:

- Choose a function \( f : \{A, C, G, U\}^2 \to \mathbb{R} \).
- Pick a pair \((i, j)\) with \( i < j \).
- For \( 1 \leq m \leq \min\{i - 1, n - j\} \) compute

\[
S^{i,j}_m = \sum_{k=0}^{m} f(y_{i-k}, y_{j+k}).
\]
If $y = y_1 \ldots y_n$ is a (very, very long) sequence from $\{A, C, G, U\}$, we search for local folding structures. We proceed as follows:

- Choose a function $f : \{A, C, G, U\}^2 \rightarrow \mathbb{R}$.
- Pick a pair $(i, j)$ with $i < j$.
- For $1 \leq m \leq \min\{i - 1, n - j\}$ compute

$$S_{m}^{i,j} = \sum_{k=0}^{m} f(y_{i-k}, y_{j+k}).$$

- Find $M_n = \max_{i,j,m} S_{m}^{i,j}$. 
We define the score matrix \((W_{ij})_{i \leq j}\) by

\[
W_{ii} = W_{i,i+1} = 0
\]

and recursively for \(i + 1 \leq j - 1\)

\[
W_{ij} = \max\{W_{i+1,j-1} + f(y_i, y_j), 0\}.
\]

Then

\[
M_n = \max_{i,j} W_{ij}.
\]
The score matrix

We define the score matrix \((W_{ij})_{i \leq j}\) by

\[ W_{ii} = W_{i,i+1} = 0 \]

and recursively for \(i + 1 \leq j - 1\)

\[ W_{ij} = \max\{W_{i+1,j-1} + f(y_i, y_j), 0\}. \]

Then

\[ M_n = \max_{i,j} W_{ij}. \]

If \(y_1 \ldots y_n\) are iid, the diagonals of \((W_{ij})_{i \leq j}\) are (dependent) random walks reflected at the zero barrier.
If $\mathbb{E}(f(y_1, y_2)) < 0$, if $\theta^*$ solves

$$\mathbb{E}(\exp(\theta f(y_1, y_2))) = 1,$$

and some technical conditions then

$$\mathbb{P}(M_n \leq t_n) \simeq \exp(-\exp(-x))$$

for $n \to \infty$ with

$$t_n = \frac{\log K^* + \log n^2/2 + x}{\theta^*}$$

and $K^*$ some constant.
Problem: The loop size $j - i$ can become arbitrarily large.

We penalise loop size by a function $g : \mathbb{N} \rightarrow (-\infty, 0]$ and consider

$$M_n^g = \max_{i,j,m} S_{i,j}^m + g(j - i).$$
Penalty on loop size

Problem: The loop size $j - i$ can become arbitrarily large.

We penalise loop size by a function $g : \mathbb{N} \rightarrow (-\infty, 0]$ and consider

$$M_n^g = \max_{i,j,m} S_{m}^{i,j} + g(j - i).$$

Defining $(W_{i,j}^g)_{i \leq j}$ by

$$W_{ii}^g = W_{i,i+1}^g = 0$$

and recursively for $i + 1 \leq j - 1$

$$W_{ij}^g = \max\{W_{i+1,j-1}^g + f(y_i, y_j), g(j - i)\},$$

then

$$M_n^g = \max_{i,j} W_{i,j}^g.$$
An expected result about $M_n^g$

The diagonals of $(W_{ij}^g)_{i \leq j}$ are (dependent) random walks reflected at barriers given in terms of $g$ – one barrier for odd and one for even loop sizes.
An expected result about $M_{n}^{g}$

The diagonals of $(W_{ij}^{g})_{i \leq j}$ are (dependent) random walks reflected at barriers given in terms of $g$ – one barrier for odd and one for even loop sizes.

We conjecture that if $\mathbb{E}(\exp(\theta^{*}D)) < \infty$ (+ technical assumptions) then $M_{n}^{g}$ will behave as the maximum of $2n$ maxima of independent random walks implying that

$$
\mathbb{P}(M_{n}^{g} \leq t_{n}) \simeq \exp(-\exp(-x))
$$

for $n \to \infty$ with

$$
t_{n} = \frac{\log \tilde{K} + \log 2n + x}{\theta^{*}}
$$

and $\tilde{K} = \mathbb{E}^{*}(\exp(\theta^{*}D))\mathbb{E}^{*}(\exp(-\theta^{*}B))$. 
We have generalised a classical result about the maximum of a random walk to random walks reflected at general barriers.
Concluding remarks

We have generalised a classical result about the maximum of a random walk to random walks reflected at general barriers.

We have gained good, precise results for barriers of potential interest.
We have generalised a classical result about the maximum of a random walk to random walks reflected at general barriers.

We have gained good, precise results for barriers of potential interest.

The results provide an important tool for handling loop size penalised searches for potential RNA-structures in DNA-genomes.
Concluding remarks

We have generalised a classical result about the maximum of a random walk to random walks reflected at general barriers.

We have gained good, precise results for barriers of potential interest.

The results provide an important tool for handling loop size penalised searches for potential RNA-structures in DNA-genomes.

We have obtained good insight into the difference between penalised and unpenalised local structure search.