The Contraction Method on $\mathcal{C}([0,1])$ and Donsker's Theorem

Henning Sulzbach
J. W. Goethe-Universität Frankfurt a. M.

YEP VII Probability, random trees and algorithms
Eindhoven, March 12, 2010

joint work with Ralph Neininger

Quickselect / Find

Task: Given a list of n different elements S , FIND the element of rank k .

Quickselect / Find

Task: Given a list of n different elements S , FIND the element of rank k .

Algorithm:

- Choose one element x at random (*pivot*),

Quickselect / Find

Task: Given a list of n different elements S , FIND the element of rank k .

Algorithm:

- Choose one element x at random (*pivot*),
- Comparing all other elements with x gives sublists $S_{\leq}, S_{>}$,

Quickselect / Find

Task: Given a list of n different elements S , FIND the element of rank k .

Algorithm:

- Choose one element x at random (*pivot*),
- Comparing all other elements with x gives sublists $S_{\leq}, S_{>}$,
- If $I_n = |S_{\leq}| \geq k$, search recursively in $S_{\leq}$ for the element of rank k, otherwise search recursively in$S_{>}$ for element of rank $k - I_n$.

Quickselect / Find

Task: Given a list of n different elements S , FIND the element of rank k .

Algorithm:

- Choose one element x at random (*pivot*),
- Comparing all other elements with x gives sublists $S_{\leq}, S_{>}$,
- If $I_n = |S_{\leq}| \geq k$, search recursively in $S_{\leq}$ for the element of rank k, otherwise search recursively in$S_{>}$ for element of rank $k - I_n$.
- Stop, if the considered set is of size 1.

Quickselect / Find

For simplicity $k = 1$.

X_n = number of key comparisons,
I_n = size of left sublist

Quickselect / Find

For simplicity $k = 1$.

X_n = number of key comparisons,
I_n = size of left sublist

This leads to

$$X_n \stackrel{d}{=} X_{I_n} + n - 1,$$

where $(X_j), I_n$ are independent, I_n is distributed uniformly on $\{0, \ldots, n-1\}$.

Quickselect / Find

For simplicity $k = 1$.

X_n = number of key comparisons,
I_n = size of left sublist

This leads to

$$X_n \stackrel{d}{=} X_{I_n} + n - 1,$$

where $(X_j), I_n$ are independent, I_n is distributed uniformly on $\{0, \dots, n-1\}$.
Scaling gives

$$Y_n := \frac{X_n}{n} \stackrel{d}{=} \frac{I_n}{n} Y_{I_n} + \frac{n-1}{n}$$

.

Recursive equation

$$Y_n := \frac{X_n}{n} \stackrel{d}{=} \frac{I_n}{n} Y_{I_n} + \frac{n-1}{n}.$$

Recursive equation

$$Y_n := \frac{X_n}{n} \stackrel{d}{=} \frac{I_n}{n} Y_{I_n} + \frac{n-1}{n}.$$

A possible limit Y should satisfy

$$Y \stackrel{d}{=} UY + 1, \tag{1}$$

for ind. U, Y with U uniform on $[0,1]$.

Recursive equation

$$Y_n := \frac{X_n}{n} \stackrel{d}{=} \frac{I_n}{n} Y_{I_n} + \frac{n-1}{n}.$$

A possible limit Y should satisfy

$$Y \stackrel{d}{=} UY + 1, \tag{1}$$

for ind. U, Y with U uniform on $[0, 1]$.

Observe: $\mu \in \mathcal{M}(\mathbb{R})$ satisfies (1), iff $F(\mu) = \mu$ for

$$F : \mathcal{M}(\mathbb{R}) \to \mathcal{M}(\mathbb{R}), \quad F(\mu) = \mathcal{L}(UY + 1)$$

with U, Y ind. and $\mathcal{L}(Y) = \mu$.

Contraction

Contraction

For $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$ let

$$\ell_1(\mu, \nu) = \inf_{X,Y:\mathcal{L}(X)=\mu,\mathcal{L}(Y)=\nu} \mathbb{E}[\|X - Y\|]$$

Contraction

For $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$ let

$$\ell_1(\mu, \nu) = \inf_{X,Y:\mathcal{L}(X)=\mu,\mathcal{L}(Y)=\nu} \mathbb{E}[|X - Y|]$$

ℓ_1 is a complete metric on $\mathcal{M}_1(\mathbb{R})$ and

$$\ell_1(\mu_n, \mu) \to 0 \Rightarrow \mu_n \xrightarrow{w} \mu.$$

Contraction

For $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$ let

$$\ell_1(\mu, \nu) = \inf_{X,Y:\mathcal{L}(X)=\mu,\mathcal{L}(Y)=\nu} \mathbb{E}[|X - Y|]$$

ℓ_1 is a complete metric on $\mathcal{M}_1(\mathbb{R})$ and

$$\ell_1(\mu_n, \mu) \to 0 \Rightarrow \mu_n \xrightarrow{w} \mu.$$

Show: F is a contraction according to ℓ_1 in $\mathcal{M}_1(\mathbb{R})$.

Contraction

Proof: Let X, Y s.t. $\mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu$ and

$$\mathbb{E}[|X - Y|] \leq d(\mu, \nu) + \varepsilon.$$

Contraction

Proof: Let X, Y s.t. $\mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu$ and

$$\mathbb{E}[|X - Y|] \leq d(\mu, \nu) + \varepsilon.$$

Then

$$\begin{aligned} \ell_1(F(\mu), F(\nu)) &\leq \mathbb{E}[|UX + 1 - (UY + 1)|] = \mathbb{E}U\mathbb{E}[|X - Y|] \\ &\leq \mathbb{E}U(\ell_1(\mu, \nu) + \varepsilon). \end{aligned}$$

Contraction

Proof: Let X, Y s.t. $\mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu$ and

$$\mathbb{E}[|X - Y|] \leq d(\mu, \nu) + \varepsilon.$$

Then

$$\begin{aligned} \ell_1(F(\mu), F(\nu)) &\leq \mathbb{E}[|UX + 1 - (UY + 1)|] = \mathbb{E}U\mathbb{E}[|X - Y|] \\ &\leq \mathbb{E}U(\ell_1(\mu, \nu) + \varepsilon). \end{aligned}$$

This gives

$$\ell_1(F(\mu), F(\nu)) \leq \mathbb{E}U\ell_1(\mu, \nu)$$

□

The stochastic fixed-point equation

$$Y \stackrel{d}{=} UY + 1$$

has a unique solution in $\mathcal{M}_1(\mathbb{R})$

The stochastic fixed-point equation

$$Y \stackrel{d}{=} UY + 1$$

has a unique solution in $\mathcal{M}_1(\mathbb{R})$ and it is easy to show

$$\ell_1(Y_n, Y) \to 0$$

for $\mathcal{L}(Y) = \mu$.

Process version

- Consider all $1 \leq k \leq n$ uniformly in the same algorithm $X_n(k,)$
- Rescaling gives a the r.v. $Y_n\left(\frac{k}{n}\right) = \frac{X_n(k)}{n}$ defined on lattice points $\frac{k}{n}$
- Fit Y_n to a stepfunction in $D([0,1])$
- Grübel, Rösler ('95) prove

$$Y_n \xrightarrow{d} Y$$

in $(D([0,1]), d_{Sk})$

Donsker's Theorem

Let $X_1, X_2, \ldots$ be iid r.v. with $\mathbb{E}X_1 = 0, \mathbb{E}X_1^2 = 1$ and $\mathbb{E}|X_1|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$.

Donsker's Theorem

Let $X_1, X_2, \ldots$ be iid r.v. with $\mathbb{E}X_1 = 0, \mathbb{E}X_1^2 = 1$ and $\mathbb{E}|X_1|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$.

$$S^n = (S_t^n)_{t\in[0,1]}$$

with

$$S_t^n = \frac{1}{\sqrt{n}}\left(\sum_{k=1}^{\lfloor nt\rfloor} X_k + (nt - \lfloor nt\rfloor)X_{\lfloor nt\rfloor+1}\right), \quad t \in [0,1]$$

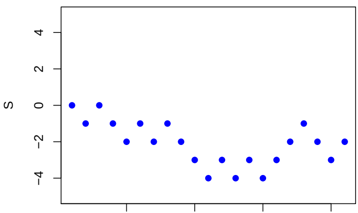
Donsker's Theorem

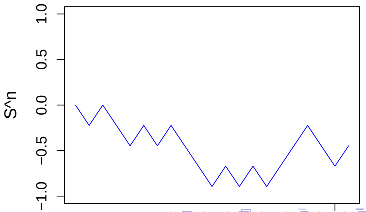
Let $X_1, X_2, \ldots$ be iid r.v. with $\mathbb{E}X_1 = 0, \mathbb{E}X_1^2 = 1$ and $\mathbb{E}|X_1|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$.

$$S^n = (S_t^n)_{t \in [0,1]}$$

with

$$S_t^n = \frac{1}{\sqrt{n}} \left(\sum_{k=1}^{\lfloor nt \rfloor} X_k + (nt - \lfloor nt \rfloor) X_{\lfloor nt \rfloor + 1} \right), \quad t \in [0,1]$$





Donsker's Theorem

Donsker's Theorem

$$S_t^n = \frac{1}{\sqrt{n}} \left(\sum_{k=1}^{\lfloor nt \rfloor} X_k + (nt - \lfloor nt \rfloor) X_{\lfloor nt \rfloor + 1} \right), \quad t \in [0,1]$$

Theorem (Donsker, 1951)

$S^n \xrightarrow{d} B$ in $(\mathcal{C}([0,1]), ||\cdot||_{\sup})$, where B is a standard Brownian Motion.

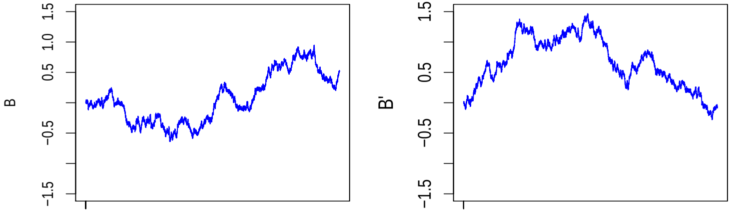
Decomposition of Brownian Motion

Decomposition of Brownian Motion

In space:

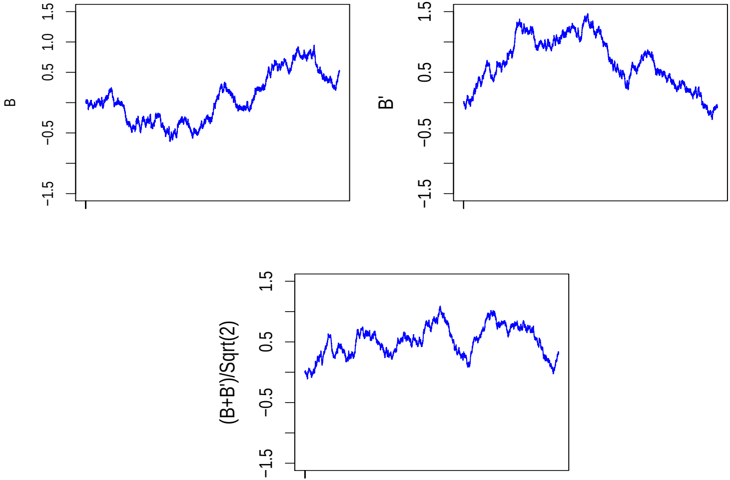
Decomposition of Brownian Motion

In space:



Decomposition of Brownian Motion

In space:



Decomposition of Brownian Motion

It holds

$$(B_t)_t \stackrel{d}{=} \left(\frac{B_t + B_t'}{\sqrt{2}} \right)_t \tag{2}$$

for two independent standard Brownian Motions B, B'.

Decomposition of Brownian Motion

It holds

$$(B_t)_t \stackrel{d}{=} \left(\frac{B_t + B'_t}{\sqrt{2}} \right)_t \qquad (2)$$

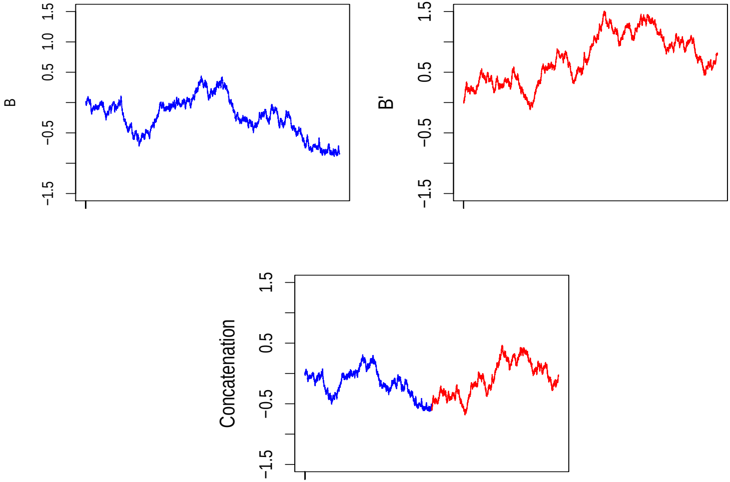
for two independent standard Brownian Motions B, B'.
In general, there is no spacian decomposition of S^n.
Equation (2) is satisfied by any gaussian process.

Decomposition of Brownian Motion

In time:

Decomposition of Brownian Motion

In time:



Decomposition of Brownian Motion

$$
\begin{aligned}
(B_t)_t &\stackrel{d}{=} \left(\frac{1}{\sqrt{2}} \left(\mathbf{1}_{\{t \le 1/2\}} B_{2t} + \mathbf{1}_{\{t > 1/2\}} B_1 \right) + \frac{1}{\sqrt{2}} \mathbf{1}_{\{t > 1/2\}} B'_{2t-1} \right)_t \\
&= \frac{1}{\sqrt{2}} \varphi_2(B) + \frac{1}{\sqrt{2}} \psi_2(B'), \qquad (3)
\end{aligned}
$$

Decomposition of Brownian Motion

$$
\begin{aligned}
(B_t)_t &\stackrel{d}{=} \left(\frac{1}{\sqrt{2}} \left(\mathbf{1}_{\{t \le 1/2\}} B_{2t} + \mathbf{1}_{\{t > 1/2\}} B_1 \right) + \frac{1}{\sqrt{2}} \mathbf{1}_{\{t > 1/2\}} B'_{2t-1} \right)_t \\
&= \frac{1}{\sqrt{2}} \varphi_2(B) + \frac{1}{\sqrt{2}} \psi_2(B'),
\end{aligned} \tag{3}
$$

with (continuous and linear) functions $\varphi_\beta, \psi_\beta : \mathcal{C}([0,1]) \to \mathcal{C}([0,1])$

$$
\begin{aligned}
\varphi_\beta(f)(t) &= \mathbf{1}_{\{t \le 1/\beta\}} f(\beta t) + \mathbf{1}_{\{t > 1/\beta\}} f(1), \\
\psi_\beta(f)(t) &= \mathbf{1}_{\{t \le 1/\beta\}} f(0) + \mathbf{1}_{\{t > 1/\beta\}} f\left(\frac{\beta t - 1}{\beta - 1} \right).
\end{aligned}
$$

Decomposition of Brownian Motion

$$
\begin{aligned}
(B_t)_t &\stackrel{d}{=} \left(\frac{1}{\sqrt{2}}\left(\mathbf{1}_{\{t\le 1/2\}}B_{2t}+\mathbf{1}_{\{t>1/2\}}B_1\right)+\frac{1}{\sqrt{2}}\mathbf{1}_{\{t>1/2\}}B'_{2t-1}\right)_t \\
&= \frac{1}{\sqrt{2}}\varphi_2(B)+\frac{1}{\sqrt{2}}\psi_2(B'), \qquad (3)
\end{aligned}
$$

with (continuous and linear) functions $\varphi_\beta, \psi_\beta : \mathcal{C}([0,1]) \to \mathcal{C}([0,1])$

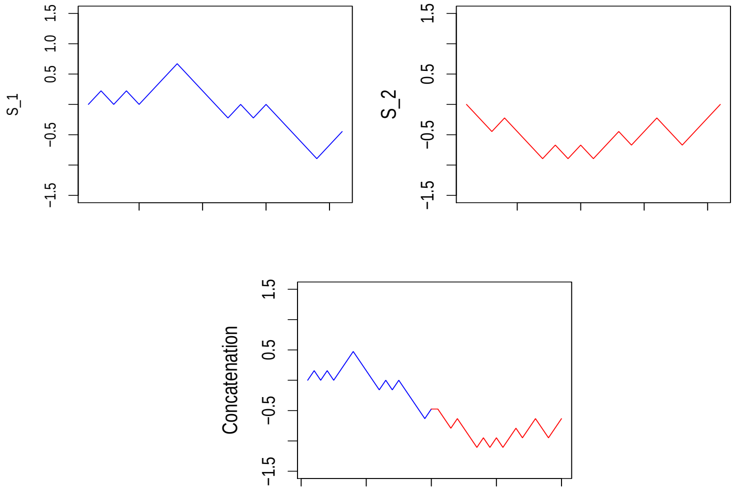
$$
\begin{aligned}
\varphi_\beta(f)(t) &= \mathbf{1}_{\{t\le 1/\beta\}}f(\beta t)+\mathbf{1}_{\{t>1/\beta\}}f(1), \\
\psi_\beta(f)(t) &= \mathbf{1}_{\{t\le 1/\beta\}}f(0)+\mathbf{1}_{\{t>1/\beta\}}f\left(\frac{\beta t-1}{\beta-1}\right).
\end{aligned}
$$

Theorem (Uniqueness - Characterization of Brownian Motion)

Let $X = (X_t)_{t\in[0,1]}$ be a continuous real-valued stochastic process satisfying (3). Then there exists a constant $\sigma \geq 0$ such that σX is a standard Brownian Motion.

Time-decomposition of the random walk

Time-decomposition of the random walk



Time-decomposition of the random walk

It holds:

$$S^n \stackrel{d}{=} \sqrt{\frac{\lceil n/2 \rceil}{n}} \varphi_{\frac{n}{\lceil n/2 \rceil}} \left(S^{\lceil n/2 \rceil} \right) + \sqrt{\frac{\lfloor n/2 \rfloor}{n}} \psi_{\frac{n}{\lceil n/2 \rceil}} \left(\widehat{S}^{\lfloor n/2 \rfloor} \right), \quad (4)$$

Time-decomposition of the random walk

It holds:

$$S^n \stackrel{d}{=} \sqrt{\frac{\lceil n/2 \rceil}{n}} \varphi_{\frac{n}{\lceil n/2 \rceil}} \left(S^{\lceil n/2 \rceil} \right) + \sqrt{\frac{\lfloor n/2 \rfloor}{n}} \psi_{\frac{n}{\lceil n/2 \rceil}} \left(\widehat{S}^{\lfloor n/2 \rfloor} \right), \qquad (4)$$

Now

$$\sqrt{\frac{\lceil n/2 \rceil}{n}} \to \frac{1}{\sqrt{2}}, \quad \sqrt{\frac{\lfloor n/2 \rfloor}{n}} \to \frac{1}{\sqrt{2}}.$$

suggests weak convergence $S^n \to B$.

The Zolotarev metric on $\mathcal{C}([0,1])$

The Zolotarev metric on $\mathcal{C}([0,1])$

Let $\mathcal{M}(\mathcal{C}([0,1]))$ be the set of probability measures on $\mathcal{C}([0,1])$. For $\mu, \nu \in \mathcal{M}(\mathcal{C}([0,1]))$ and $s > 0$ define the Zolotarev distance of μ and ν by

$$\zeta_s(\mu, \nu) = \sup_{f \in \mathcal{F}_s} |\mathbb{E}[f(X) - f(Y)]|,$$

with $\mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu$

The Zolotarev metric on $\mathcal{C}([0,1])$

Let $\mathcal{M}(\mathcal{C}([0,1]))$ be the set of probability measures on $\mathcal{C}([0,1])$. For $\mu, \nu \in \mathcal{M}(\mathcal{C}([0,1]))$ and $s > 0$ define the Zolotarev distance of μ and ν by

$$\zeta_s(\mu,\nu) = \sup_{f \in \mathcal{F}_s} |\mathbb{E}[f(X) - f(Y)]|,$$

with $\mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu$ and for $s = m + \alpha$ with $0 < \alpha \leq 1$ and $m := \lceil s \rceil - 1 \in \mathbb{N}_0$

$$\mathcal{F}_s := \{f \in C^m(\mathcal{C}([0,1]),\mathbb{R}) : \|D^m f(x) - D^m f(y)\| \leq \|x-y\|^\alpha, \quad x, y \in \mathcal{C}$$

The Zolotarev metric on $\mathcal{C}([0,1])$

Let $\mathcal{M}(\mathcal{C}([0,1]))$ be the set of probability measures on $\mathcal{C}([0,1])$. For $\mu, \nu \in \mathcal{M}(\mathcal{C}([0,1]))$ and $s > 0$ define the Zolotarev distance of μ and ν by

$$\zeta_s(\mu, \nu) = \sup_{f \in \mathcal{F}_s} |\mathbb{E}[f(X) - f(Y)]|,$$

with $\mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu$ and for $s = m + \alpha$ with $0 < \alpha \leq 1$ and $m := \lceil s \rceil - 1 \in \mathbb{N}_0$

$$\mathcal{F}_s := \{f \in C^m(\mathcal{C}([0,1]), \mathbb{R}) : \|D^m f(x) - D^m f(y)\| \leq \|x - y\|^\alpha, \quad x, y \in \mathcal{C}$$

Set $\zeta_s(X, Y) = \zeta_s(\mathcal{L}(X), \mathcal{L}(Y))$.

Properties of the Zolotarev distance ζ_s

Properties of the Zolotarev distance ζ_s

It holds $\zeta_s(X, Y) < \infty$, if

$$\mathbb{E}||X||^s, \mathbb{E}||Y||^s < \infty, \quad \mathbb{E}[g(X, \dots, X)] = \mathbb{E}[g(Y, \dots, Y)]$$

for all $k \leq m$ and multilinear, bounded functions $g : \mathcal{C}([0,1])^k \to \mathbb{R}$.
In the following assume finiteness of the considered ζ_s-distances

Properties of the Zolotarev distance ζ_s

Lemma

ζ_s is $(s,+)$ - ideal, i.e.

$$\zeta_s(\varphi(X),\varphi(Y)) \leq ||\varphi||^s \zeta_s(X,Y)$$

for any continuous and linear function $\varphi : \mathcal{C}([0,1]) \to \mathcal{C}([0,1])$ with

$$||\varphi|| = \sup_{f \in \mathcal{C}([0,1]), ||f||=1} ||\varphi(f)||.$$

Properties of the Zolotarev distance ζ_s

Lemma

ζ_s is $(s,+)$ - ideal, i.e.

$$\zeta_s(\varphi(X),\varphi(Y)) \leq ||\varphi||^s \zeta_s(X,Y)$$

for any continuous and linear function $\varphi : \mathcal{C}([0,1]) \to \mathcal{C}([0,1])$ with

$$||\varphi|| = \sup_{f \in \mathcal{C}([0,1]), ||f||=1} ||\varphi(f)||.$$

Furthermore

$$\zeta_s(X_1+X_2, Y_1+Y_2) \leq \zeta_s(X_1,Y_1) + \zeta_s(X_2,Y_2)$$

for (X_1,Y_1) and (X_2,Y_2) independent.

Proof: Zolotarev ('76)

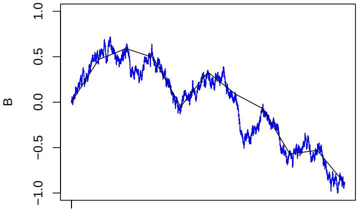
Proof of Donsker's Theorem (Sketch)

We are in the case $2 < s < 3$ and have to consider $\mathbb{E}[f(X,X)]$ for continuous, bilinear functions $f : \mathcal{C}([0,1])^2 \to \mathbb{R}$.
This is done by controlling the covariance function.

Proof of Donsker's Theorem (Sketch)

We are in the case $2 < s < 3$ and have to consider $\mathbb{E}[f(X,X)]$ for continuous, bilinear functions $f : \mathcal{C}([0,1])^2 \to \mathbb{R}$.
This is done by controlling the covariance function.
Since S^n and B do not share their covariance function (of course $\mathbb{E}[S^n_s, S^n_t] \to \min(s,t)$) we also consider the process B^n defined by

$$B^n_t = B_{\frac{\lfloor nt \rfloor}{n}} + (nt - \lfloor nt \rfloor)\left(B_{\frac{\lfloor nt \rfloor + 1}{n}} - B_{\frac{\lfloor nt \rfloor}{n}}\right).$$



Proof of Donsker's Theorem (Sketch)

An contraction argument shows

$$\zeta_{2+\varepsilon}(S^n, B^n) = O(n^{-\delta})$$

for $\delta < \varepsilon/2$.

Proof of Donsker's Theorem (Sketch)

An contraction argument shows

$$\zeta_{2+\varepsilon}(S^n, B^n) = O(n^{-\delta})$$

for $\delta < \varepsilon/2$. Together with

$$||B^n - B|| \to 0 \quad a.s.$$

Proof of Donsker's Theorem (Sketch)

An contraction argument shows

$$\zeta_{2+\varepsilon}(S^n, B^n) = O(n^{-\delta})$$

for $\delta < \varepsilon/2$. Together with

$$||B^n - B|| \to 0 \quad a.s.$$

Donsker's Theorem follows with the following properties of Zolotarev's distance.

Properties of ζ_s in $\mathcal{C}([0,1])$

Properties of ζ_s in $\mathcal{C}([0,1])$

Theorem

Let $B = \mathcal{C}([0,1])$ *and* $0 < s \leq 3$, *i.e,* $m \in \{0,1,2\}$. *Let* $(X_n)_{n\geq 1}$, X *be random variables in* $(\mathcal{C}([0,1]), \|\cdot\|)$ *where, for each* $n \geq 1$, X_n *is piecewise linear with intervals of length at least* r_n. *If*

$$\zeta_s(X_n, X) = o\left(\log^{-m}\frac{1}{r_n}\right), \quad \textit{as} \quad n \to \infty,$$

then $X_n \to X$ *in distribution.*

Properties of ζ_s in $\mathcal{C}([0,1])$

Theorem

Let $B = \mathcal{C}([0,1])$ and $0 < s \leq 3$, i.e, $m \in \{0,1,2\}$. Let $(X_n)_{n\geq 1}$, X be random variables in $(\mathcal{C}([0,1]), \|\cdot\|)$ where, for each $n \geq 1$, X_n is piecewise linear with intervals of length at least r_n. If

$$\zeta_s(X_n, X) = o\left(\log^{-m}\frac{1}{r_n}\right), \quad \text{as} \quad n \to \infty,$$

then $X_n \to X$ in distribution.

Proof: This basically follows from a result of Barbour from the context of Stein's method.

Properties of ζ_s in $\mathcal{C}([0,1])$

Theorem

Suppose $0 < s \leq 3$, *i.e,* $m \in \{0,1,2\}$. *Let* $(X_n)_{n\geq 1}$, $(Y_n)_{n\geq 1}$, Z *be random variables in* $(\mathcal{C}([0,1]), \|\cdot\|)$ *where, for each* $n \geq 1$, X_n, Y_n *are piecewise linear with intervals of length at least* r_n. *If* $Y_n \to Z$ *in distribution and*

$$\zeta_s(X_n, Y_n) = o\left(\log^{-m}\frac{1}{r_n}\right), \quad \text{as} \quad n \to \infty,$$

then $X_n \to Z$ *in distribution.*