Weak convergence and Brownian Motion

(telegram style notes)

P.J.C. Spreij

this version: December 8, 2006

1 The space $C[0,\infty)$

In this section we summarize some facts concerning the space $C[0,\infty)$ of real valued continuous functions defined on $[0,\infty)$. For $x_1, x_2 \in C[0,\infty)$ we define

$$\rho(x_1,x_2) = \sum_{n\geq 1} 2^{-n}(\max\{|x_1(t)-x_2(t)| : 0 \leq t \leq n\} \wedge 1). \tag{1.1}$$

Then ρ defines a metric on $C[0,\infty)$ (which we use throughout these notes) and we have

Proposition 1.1 *The metric space $(C[0,\infty),\rho)$ is complete and separable.*

Later on we need the relatively compact subsets of $C[0,\infty)$. To describe these we introduce the modulus of continuity m^T. For each $x \in C[0,\infty)$, $T, \delta > 0$ we define

$$m^T(x,\delta) = \max\{|x(t)-x(s)| : s,t \in [0,T], |s-t| \leq \delta\}. \tag{1.2}$$

It holds that $m^T(\cdot,\delta)$ is continuous and $\lim_{\delta\downarrow 0} m^T(x,\delta) = 0$ for each x and T. The following characterization is known as the Arzelà-Ascoli theorem.

Theorem 1.2 *A set A in $C[0,\infty)$ is relatively compact (has compact closure) iff (i) $\sup\{|x(0)| : x \in A\} < \infty$ and (ii) for all $T > 0$ $\lim_{\delta\downarrow 0} \sup\{m^T(x,\delta) : x \in A\} = 0$.*

Under requirement (ii) in this proposition, the functions in A are uniformly equicontinuous.

Cylinder sets of $C[0,\infty)$ have the typical form $\{x : (x(t_1),\ldots,x(t_k)) \in A\}$, where $A \in \mathcal{B}(\mathbb{R}^k)$ for some $k \geq 1$ and $t_1,\ldots,t_k \in [0,\infty)$. A finite dimensional projection on $(C[0,\infty),\rho)$ is by definition of the following type: $\pi_{t_1,\ldots,t_k}(x) = (x(t_1),\ldots,x(t_k))$, where the t_i are nonnegative real numbers. It is easy to see that any finite dimensional projection is continuous ($\mathbb{R}^k$ is endowed with the ordinary metric). Note that cylinder sets are invere images under finite dimensional projections of Borel sets of $\mathbb{R}^k$ ($k \geq 1$). Let $\mathcal{C}$ be the collection of all cylinder sets and $\mathcal{B}$ the Borel σ-algebra on $C[0,\infty)$ induced by the metric ρ. Let $(\Omega,\mathcal{F})$ be a measurable space. A map $X : \Omega \to C[0,\infty)$ is called a random element of $C[0,\infty)$ if it is $\mathcal{F}/\mathcal{B}$-measurable. It follows that $\pi_{t_1,\ldots,t_k} \circ X$ is random vector in $\mathbb{R}^k$, for any finite dimensional projection $\pi_{t_1,\ldots,t_k}$, and it is usually denoted by $(X_{t_1},\ldots,X_{t_k})$. One can prove that $\mathcal{B} = \sigma(\mathcal{C})$ and thus that X is a random element of $C[0,\infty)$, if all X_t are real random variables. Moreover, if $\mathbb{P}$ is a probability measure on $(\Omega,\mathcal{F})$ and X a random element of $C[0,\infty)$, then the distribution $\mathbb{P}^X$ of X on $(C[0,\infty),\mathcal{B})$ is completely determined by the distributions of all k-tuples $(X_{t_1},\ldots,X_{t_k})$ on $\mathbb{R}^k$ ($k \geq 1, t_i \in [0,\infty)$).

2 Weak convergence on metric spaces

Let (S, ρ) be a metric space and $P, P^1, P^2, \ldots$ be probability measures on the Borel σ-algebra $\mathcal{B}(S)$. Like in the real case we say that P^n converges weakly to P (notation $P^n \stackrel{w}{\to} P$) iff for all $f \in C_b(S)$ one has $\lim P^n f = Pf$. If $X, X^1, X^2, \ldots$ are random variables defined on probability spaces $(\Omega, \mathcal{F}, P)$ and $(\Omega^n, \mathcal{F}^n, \mathbb{P}^n)$ $(n \geq 1)$ with values in one and the same (S, ρ), we say that X^n converges in distribution to X ($X^n \stackrel{w}{\to} X$) if the laws P^n of X^n converge weakly to the law P of X, equivalently, iff $P^n f(X^n) \to Pf(X)$ for all $f \in C_b(S)$.

A family of probability measures Π on $\mathcal{B}(S)$ is called tight if for every $\varepsilon > 0$, there is a compact subset K of S such that $\inf\{P(K) : P \in \Pi\} > 1-\varepsilon$. One can show that any single probability measure on $\mathcal{B}(S)$ is tight if (S, ρ) is a separable and complete metric space (a Polish space). A family of random variables with values in a metric space is called tight if the family of their distributions is tight. Like in the real case (but much harder to prove here) there is equivalence between relative compactness (in this context it means that every sequence in a set of probability measures has a weakly converging subsequence) and tightness, known as Prohorov's theorem.

Theorem 2.1 *A family Π of probability measures on a complete separable metric space is tight iff it is relatively compact.*

We will also need the following perturbation result.

Proposition 2.2 *Let $X^1, X^2, \ldots$ and $Y^1, Y^2, \ldots$ be random sequences in a metric space (S, ρ) and defined on a single probability space. If $X^n \stackrel{w}{\to} X$ and $\rho(Y^n, X^n) \stackrel{P}{\to} 0$, then $Y^n \stackrel{w}{\to} X$.*

If we take $S = C[0, \infty)$ with the metric ρ of the previous section, we get the following 'stochastic version' of the Arzelà-Ascoli theorem.

Theorem 2.3 *Let $P^1, P^2, \ldots$ be a sequence of probability measures on the space $(C[0, \infty), \mathcal{B})$. This sequence is tight iff*

$$\lim_{\lambda \uparrow \infty} \sup\{P^n(x : |x(0)| > \lambda) : n \geq 1\} = 0 \tag{2.3}$$

and

$$\lim_{\delta \downarrow 0} \sup\{P^n(x : m^T(x, \delta) > \varepsilon) : n \geq 1\} = 0, \forall T, \varepsilon > 0. \tag{2.4}$$

Proof. If the sequence is tight, the result is a straightforward application of theorem 1.2. For every $\varepsilon > 0$ we can find a compact K such that $\inf_n P^n(K) > 1 - \varepsilon$. But then we can find $\lambda > 0$ such that for all $x \in K$ we have $|x(0)| < \lambda$ and we can similarly find for given $T > 0$ and $\eta > 0$ a $\delta_0 > 0$ such that for all $0 < \delta < \delta_0$ we have on K that $m^T(x, \delta) < \eta$.

Conversely, assume (2.3) and (2.4) and let $\varepsilon, T > 0$, T integer, be given. Choose λ_T such that $\sup_n P^n(x : |x(0)| > \lambda_T) \leq \varepsilon 2^{-T-1}$. For each $k \geq 1$ we can also find δ_k such that $\sup_n P^n(x : m^T(x, \delta_k) > 1/k) \leq \varepsilon 2^{-T-k-1}$. Notice that the sets $A_{T,k} = \{x : m^T(x, \delta_k) \leq 1/k\}$ and $A_{T,0} = \{x : |x(0)| \leq \lambda_T\}$ are closed and so is their intersection over both k and (integer) T, call it K. From theorem 1.2 we obtain that K has compact closure and it is thus compact itself. Finally we compute $P^n(K^c) \leq \sum_{T\geq 1} \mathbb{P}(A^c_{T,0}) + \sum_{k\geq 1} \mathbb{P}(A^c_{T,k}) \leq \varepsilon$. □

3 Finite dimensional convergence

We have seen that any finite dimensional projection is continuous. Hence, if $X, X_1, X_2, \ldots$ are random elements of $(C[0,\infty), \mathcal{B})$ and if we assume that $X_n \xrightarrow{w} X$, then also $(X^n_{t_1}, \ldots, X^n_{t_k})$ considered as random elements in $\mathbb{R}^k$ converge in distribution to $(X_{t_1}, \ldots, X_{t_k})$. This is then true for any finite set of t_i's and we say that all finite dimensional distributions converge weakly. The converse does not hold in general, unless one assumes tightness.

Theorem 3.1 *Let $X_1, X_2, \ldots$ be random elements of $C[0,\infty)$. Assume that their collection $\{P^1, P^2, \ldots\}$ of distributions is tight and that all finite dimensional distributions of the X_n converge weakly. Then there exists a probability measure P on $(C[0,\infty), \mathcal{B})$ such that $P^n \xrightarrow{w} P$.*

Proof. Every subsequence of (P^n) is tight as well and thus has a convergent subsequence. Different subsequences have to converge to the same limit, call it P, since the finite dimensional distributions corresponding to these sequences converge. Hence, if (P^n) has a limit, it must be P. Suppose therefore that the P^n don't converge. Then there is bounded and continuous f and an $\varepsilon > 0$ such that $|P^{n_k} f - Pf| > \varepsilon$ along a subsequence (P^{n_k}). No further subsequence of this can have P as a limit which contradicts what we just showed. □

4 An invariance principle

Throughout this section we work with a real valued *iid* sequence $\xi_1, \xi_2, \ldots$ with zero mean and variance $\sigma^2 \in (0,\infty)$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $S_k = \sum_{i=1}^k \xi_i$ and for each integer n and $t \geq 0$

$$X^n_t = \frac{1}{\sigma\sqrt{n}}(S_{[nt]} + (nt - [nt])\xi_{[nt]+1}. \tag{4.5}$$

The processes X^n have continuous paths and can be considered as random elements of $C[0,\infty)$. Notice that the increments $X^n_t - X^n_s$ of each X^n over intervals (s,t) with $s = \frac{k}{n}, t = \frac{l}{n}$, $k < l$ integers, are independent. Since for these values of t and s we have $\mathrm{Var}\,(X^n_t - X^n_s) = t - s$, the central limit theorem should be helpful to understand the limit behaviour.

Theorem 4.1 *Let $0 = t_0 < t_1 < \cdots < t_k$. Then the k-vector of increments $X^n_{t_j} - X^n_{t_{j-1}}$ with $j = 1, \ldots, k$ converges in distribution to a random vector with independent elements N_j, where each N_j has a $N(0, t_j - t_{j-1})$ distribution.*

Proof. Since the term in (4.5) with the $\xi_{[nt]}$ tends to zero in probability, we can ignore it as a consequence of proposition 2.2. But then the conclusion follows from the ordinary Central Limit Theorem.

Denote by P^n the law of X^n. We have the following important result.

Theorem 4.2 *The sequence of probability measures P^n is tight.*

Combined with theorems 3.1 and 4.1 one obtains

Theorem 4.3 *There exists a probability measure P^* on $(C[0,\infty),\mathcal{B})$ such that $P^n \xrightarrow{w} P^*$.*

Any process defined on some probability space that has continuous paths, that starts in zero and that has independent increments over non-overlapping intervals (s,t) with a $N(0,t-s)$ distribution is called a Wiener process, also called a Brownian motion. Let W denotes the *coordinate mapping process* on $\Omega = C[0,\infty)$, i.e. W is defined by $W_t(\omega) = \omega(t)$ for all $t \geq 0$. Under the measure P^* this process has independent increments over non-overlapping intervals (s,t) and these increments have a $N(0,t-s)$ distribution. Since by definition W is a random element of $(C[0,\infty),\mathcal{B})$, W is thus a Wiener process and the measure P^* is called Wiener measure. Notice that P^* is also the law of W.

We can rephrase theorem 4.3 as

Theorem 4.4 *The processes X^n of this section converge in distribution to a Wiener process W.*

Both theorems 4.3 and 4.4 are known as Donsker's invariance principle. What we have done in this section can be summarized by saying that we have shown the existence of a Wiener process and we have given a Functional Central Limit Theorem.

5 The proof of theorem 4.2

Consider the process S_n of section 4. To prove theorem 4.2 we use the following results.

Lemma 5.1 *Let $\gamma > 0$, $n \geq 1$, $N \geq n$ and $\eta \geq \sigma\sqrt{2(n-1)}$. The following inequalities are valid.*

$$\mathbb{P}(\max_{j\leq n}|S_j| > \gamma) \leq 2\mathbb{P}(|S_n| > \gamma - \eta) \tag{5.6}$$

$$\mathbb{P}(\max_{\substack{1\leq j\leq n\\ 0\leq k\leq N}} |S_{j+k} - S_k| > \gamma) \leq (\frac{N}{n} + 2)\mathbb{P}(\max_{j\leq n}|S_j| > \gamma/3). \tag{5.7}$$

Proof. Assume $\eta < \gamma$. Let $\tau = \min\{j : |S_j| > \gamma\}$. Then we have to consider $\mathbb{P}(\tau \leq n)$. Split this probability up as

$$\mathbb{P}(\tau \leq n, |S_n| > \gamma - \eta) + \mathbb{P}(\tau \leq n, |S_n| \leq \gamma - \eta) \tag{5.8}$$

and work on the second probability. It can be written as $\sum_{j=1}^{n-1} \mathbb{P}(\tau = j, |S_n| \leq \gamma - \eta)$ and each of the probabilities in the sum is less than or equal to $\mathbb{P}(\tau = j, |S_n - S_j| > \eta) = \mathbb{P}(\tau = j)\mathbb{P}(|S_n - S_j| > \eta)$. The second factor is by Chebychev's inequality less than $\frac{1}{\eta^2}(n-1)\sigma^2 \leq \frac{1}{2}$, by the assumption on η. Therefore $\mathbb{P}(\tau \leq n, |S_n| \leq \gamma - \eta) \leq \frac{1}{2}\mathbb{P}(\tau \leq n-1)$. From (5.8), we then get $\mathbb{P}(\tau \leq n) \leq \mathbb{P}(|S_n| > \gamma - \eta) + \frac{1}{2}\mathbb{P}(\tau \leq n)$ and the inequality (5.6) follows.
To prove (5.7) we argue as follows. Let $m = [\frac{N}{n}]$ and consider the 'intervals' $\{pn, \ldots, (p+1)n - 1\}$, for $p = 0, \ldots, m$. N belongs to the last one. Consider j and k for which the maximum is bigger than γ. If $k + j$ belongs to the same interval as k, the one starting with pn, say, we certainly have $|S_{np} - S_k| > \gamma/3$ or $|S_{np} - S_{k+j}| > \gamma/3$ and so in this case there is $p \leq m$ such that $\max_{j \leq n} |S_{np} - S_j| > \gamma/3$. If $k+j$ lies in the interval starting with $(p+1)n$, we must have $|S_{np} - S_k| > \gamma/3$ or $|S_{n(p+1)} - S_{k+j}| > \gamma/3$ or $|S_{n(p+1)} - S_{np}| > \gamma/3$. Both cases are contained in the event $\bigcup_{0 \leq p \leq m+1}\{\max_{j \leq n} |S_{np} - S_{np+j}| > \gamma/3\}$, whose probability is less than or equal to $\sum_{p=0}^{m+1} \mathbb{P}(\max_{j \leq n} |S_{np} - S_{np+j}| > \gamma/3)$. By the *iid* assumption all probabilities in this sum are equal to the first one and thus the sum is equal to $(m+2)\mathbb{P}(\max_{j \leq n} |S_j| > \gamma/3)$, which yields the result. □

With this lemma we prove 4.2 as follows. According to theorem 2.3 it is sufficient to show that

$$\lim_{\delta \downarrow 0} \sup_{n \geq 1} \mathbb{P}(\max_{\substack{|s-t| \leq \delta \\ 0 \leq t,s \leq T}} |X_t^n - X_s^n| > \varepsilon) = 0 \text{ for all } T, \varepsilon > 0. \tag{5.9}$$

But since we only need tightness for all but finitely many n, we can as well replace the 'sup' by a 'lim sup'. Let $Y_t = \sigma\sqrt{n}X^n_{t/n}$. Each of the probabilities in (5.9) is less than

$$\mathbb{P}(\max_{\substack{|s-t| \leq [n\delta]+1 \\ 0 \leq t,s \leq [nT]+1}} |Y_t - Y_s| > \varepsilon\sigma\sqrt{n}).$$

But, since Y is piecewise linear between the integer values of its arguments, the max is attained at integer numbers. Hence we consider

$$\mathbb{P}(\max_{\substack{0 \leq j \leq [n\delta]+1 \\ 0 \leq k \leq [nT]+1}} |S_{j+k} - S_k| > \varepsilon\sigma\sqrt{n}). \tag{5.10}$$

Now we apply inequality (5.7) and bound this probability by

$$(\frac{[nT]+1}{[n\delta]+1} + 2)\mathbb{P}(\max_{j \leq [n\delta]+1} |S_j| > \varepsilon\sigma\sqrt{n}/3). \tag{5.11}$$

In view of (5.6) (take $\eta = \sigma\sqrt{2[n\delta]}$) the probability in (5.11) is less than

$$\mathbb{P}(|S_{[n\delta]+1}| > \varepsilon\sigma\sqrt{n}/3 - \sigma\sqrt{2[n\delta]}).$$

Now we apply the central limit theorem: $\frac{1}{\sigma\sqrt{[n\delta]}} S_{[n\delta]+1} \xrightarrow{w} Z$, where Z has a $N(0,1)$ distribution. So for $n \to \infty$ the last probability tends to $\mathbb{P}(|Z| > \frac{\varepsilon}{3\sqrt{\delta}} - \sqrt{2})$ which is less than $\frac{\delta^2}{(\varepsilon/3-\sqrt{2\delta})^4}\mathbb{E}\, Z^4$. Hence the lim sup in (5.11) for $n \to \infty$ is less than $\frac{T\delta+2\delta^2}{(\varepsilon/3-\sqrt{2\delta})^4}\mathbb{E}\, Z^4$, from which we obtain (5.9). $\square$.