Solutions for $\mathbb{E}$xercise sheet 7: Some more martingales & Donsker's theorem (v2)

Solution 1 — *A weaker condition for the first Wald's lemma.* (1) Define $\tau := \min\{k : 4^k \geq T\}$. Set $M(t) := \max_{[0,t]} B$ and

$$\mathbb{E}[X_{k+1} - X_k \mid \mathcal{F}_{4^k}] = \mathbb{E}[M(4^{k+1}) - M(4^k) \mid \mathcal{F}_{4^k}] - 4 \times 2^k.$$

Since we know that almost surely $M(4^{k+1}) - M(4^k) \leq |B_{4^{k+1}} - B_{4^k}|$ which is independent of $\mathcal{F}_{4^k}$ and distributed like $|B_{4^{k+1}-4^k}|$, then

$$\mathbb{E}[X_{k+1} - X_k \mid \mathcal{F}_{4^k}] \leq \mathbb{E}[|B_{4^{k+1}-4^k}|] - 4 \times 2^k = \sqrt{3 \times 4^k}\,\mathbb{E}[|B_1|] - 4 \times 2^k.$$

A simple application of Cauchy-Schwarz or Jensen gives $\mathbb{E}[|B_1|] \leq \sqrt{\mathbb{E}[|B_1|^2]} = 1$, and the expectation above is bounded by 0.

If we consider τ, we have the equality of events $\{\tau \leq k\} = \{4^k \geq T\}$, which belongs to $\mathcal{F}_{4^k}$. So τ is a $(\mathcal{F}_{4^k})_k$-stopping time.

(2) Let $n \geq 0$. $\mathbb{E}[M(4^\tau \wedge 4^n)] = \mathbb{E}[X_{\tau\wedge n}] + \mathbb{E}[2^{\tau\wedge n+2}] \leq \mathbb{E}[X_0] + 8\,\mathbb{E}[T^{1/2}]$, where we have used the supermartingale property at the bounded stopping time $\tau \wedge n$ and the fact that $4^\tau \leq 4T$. By monotone convergence $M(4^\tau)$ is integrable so $\max_{[0,T]} B \leq M(4^\tau)$ too. By reversal, $-\min_{[0,T]} B$ is integrable also, and this provides an integrable random variable that bounds $B_{t\wedge T}$ for every t. So the optional stopping theorem applies and $\mathbb{E}[B_T] = 0$.

(3) If $\alpha < 1/2$, then $t^\alpha \times t^{-3/2} e^{-1/(2t)}$ is $o(e^{-1/(2t)})$ (so it's integrable) near 0, and is $O(t^{-1-(1/2-\alpha)})$ near infinity, so is integrable too.

Solution 2 — *An application of Donsker's invariance principle.*
Let $\Phi : \mathcal{C}([0,1]) \to \mathcal{C}([0,1])$ that reflects a continuous function after its last 0. Let $S_n : [0,1] \to \mathbb{R}$ be a properly rescaled and linearly interpolated simple random walk, so that $S_n \xrightarrow{d} B$ in $\mathcal{C}([0,1])$. Since flipping a length n random-walk at its last zero is a bijective (involutive!) operation on the finite set of n-length random walks, we get that $\Phi(S_n) \overset{d}{=} S_n$. Now if we had $\Phi(S_n) \xrightarrow{d} \Phi(B)$, we would get $\Phi(B) \overset{d}{=} B$. To that end we need to show that B is a continuity point of Φ.

Denote by $Z(f)$ the position of the last zero of f. We already know that almost surely, B is a continuity point of Z. Let f be a continuity point of Z and $f_n \to f$ in $\mathcal{C}([0,1])$. Then for $x \in [0,1]$,

$$|\Phi(f_n)(x) - \Phi(f)(x)| \leq |f_n(x) - f(x)|\,\mathbb{1}_A(x) + |f_n(x) + f(x)|\,\mathbb{1}_{A^\complement}(x)$$

Where A denotes the set of x such that "$Z(f_n)$ and $Z(f)$ are on the same side of x". But now if $x \in A^\complement$, x is a distance at most $|Z(f_n) - Z(f)|$ of $Z(f)$. So $|f(x)| = |f(x) - f(Z(f))| \leq$

$m_{[0,1]}(f, |Z(f) - Z(f_n)|)$. Moreover, $|f_n(x) - f(x)|$ is always bounded by $\|f_n - f\|$. We get

$$\begin{aligned} |\Phi(f_n)(x) - \Phi(f)(x)| &\leq |f_n(x) - f(x)| \mathbb{1}_A(x) + |f_n(x) - f(x)| \mathbb{1}_{A^{\complement}}(x) + 2|f(x)| \mathbb{1}_{A^{\complement}}(x) \\ &\leq \|f_n - f\| + \|f_n - f\| + 2m_{[0,1]}(f, |Z(f) - Z(f_n)|) \end{aligned}$$

We know, since f is continuous and a continuity point of Z, that this last quantity goes to 0. Since it's independent of x, we have shown $\Phi(f_n) \to \Phi(f)$ in $\mathcal{C}([0,1])$.

Solution 3 — *First arcsine law.* (1) Draw the picture to understand that this gives a bijective transformation of the set of length-n walks.

(2) Once again a consequence of the drawing.

(3) Using the fact mentioned, you only need to show that the functional $\Phi : \mathcal{C}([0,1]) \to [0,1], \Phi(f) = \inf\{t \in [0,1], f(t) = \max_{[0,1]} f\}$, is continuous at every f that reaches its maximum at a unique point. Indeed $B_n/n = \Phi(t \mapsto \frac{1}{\sqrt{n}} R^n_{nt})$ (understood as being suitably interpolated), and Donsker's theorem says that $(\frac{1}{\sqrt{n}} R^n_{nt})_t \to B$ in distribution.

Let f be a function that reaches its maximum at a unique point $m \in [0,1]$, and $f_n \to f$. Let $\epsilon > 0$. By continuity and compactness, the maximum of f on $[0, m-\epsilon] \cup [m+\epsilon, 1]$ is reached in some point $y \neq m$, and by assumption $f(y) < f(m)$. Hence we can find η be such that $f(m) - \max_{x \notin (m-\epsilon, m+\epsilon)} f_n > 2\eta$. Then for n such that $\|f_n - f\| < \eta$, we have that $f_n(m) > \max_{x \notin (m-\epsilon, m+\epsilon)} f_n$. So $\Phi(f_n) \in (m - \epsilon, m + \epsilon)$, and $|\Phi(f_n) - \Phi(f)| < \epsilon$. This shows continuity.

So $B_n/n \to \Phi(B)$ which is arcsine distributed.

(4) We can show that A_n/n is equal to $\text{Leb}\{t \in [0,1], \frac{1}{\sqrt{n}} S^n_{nt} \geq \frac{1}{2\sqrt{n}}\}$. (Once again $\frac{1}{\sqrt{n}} S^n_{nt} \geq \frac{1}{2\sqrt{n}}$ is understood as being suitably interpolated). Let now define $\Phi(f) = \text{Leb}\{t \in [0,1], f(t) \geq 0\}$. We have $A_n/n = \Phi((\frac{1}{\sqrt{n}} S^n_{nt} - \frac{1}{2\sqrt{n}})_t$. By Slutsky's lemma and Donsker's invariance principle, we have $(\frac{1}{\sqrt{n}} S^n_{nt} - \frac{1}{2\sqrt{n}})_t \to B$, and showing that B almost surely is a continuity point of Φ suffices to get $A_n/n \xrightarrow{d} P = \Phi(B)$.

Now suppose that f is such that $\text{Leb}\{t : |f(t)| \leq \epsilon\} \xrightarrow[\epsilon \to 0]{} 0$. Then f is a continuity point of Φ. Indeed if $f_n \to f$, fix $\epsilon > 0$. Then for n large enough, $\|f_n - f\| \leq \epsilon$. Then $|\Phi(f) - \Phi(f_n)| \leq \text{Leb}\{t \in [0,1] : f_n(t) f(t) \leq 0\}$. But $f_n(t) f(t) \leq 0$ implies $|f(t)| < \epsilon$. So $|\Phi(f) - \Phi(f_n)| \leq \text{Leb}\{t : |f(t)| \leq \epsilon\}$, which could have been taken arbitrarily close to 0 by choosing ϵ small enough. So $\Phi(f_n) \to \Phi(f)$.

We are left to show that almost surely, $U_\epsilon := \text{Leb}\{t : |B_t| \leq \epsilon\} \xrightarrow[\epsilon \to 0]{} 0$. But $\mathbb{E}[U_\epsilon] = \int_0^1 \mathbb{P}(-\epsilon/\sqrt{t} \leq B_1 \leq \epsilon/\sqrt{t}) dt \leq \frac{2\epsilon}{\sqrt{2\pi}} \int_0^1 dt/\sqrt{t} \leq \epsilon$. So U_ϵ goes to 0 in L^1, hence almost surely there exists a subsequence of ϵ that goes to 0 along which $U_\epsilon \to 0$, but as almost surely $\epsilon \mapsto U_\epsilon$ is decreasing, we get $U_\epsilon \to 0$.

So $B_n/n = A_n/n \to P$ and P is arcsine-distributed.

Solution 4 — *Convergence in distribution of random continuous functions.* (1) (a) You know that the measure μ_n gives you a continuous linear form f_n on the set

$\mathcal{C}_c(E)$, of norm 1. Banach-Alaoglu's theorem tells you that you can extract a weak-$\star$-convergent subsequence $f_{a_n} \to f \in B_{C_c(E)'}(0,1)$, i.e. such that $f_{a_n}(\phi) \to f(\phi)$ for every $\phi \in \mathcal{C}_c(E)$. Now f is positive (since for $\phi \geq 0$, $f(\phi) = \lim_n f_{a_n}(\phi) = \lim_n \mu_{a_n}\phi \geq 0$, so by Riesz' representation theorem, it can be represented by a positive Borel measure μ, and we precisely have vague convergence $\mu_{a_n} \to \mu$.

(b) The fact that ($\mu_n \to \mu$ narrowly) $\iff$ ($\mu_n \to \mu$ vaguely and $\mu(E) = 1$) is standard and the proof is rather easy (relies only on the fact that $\mathcal{R}^d$ is σ-compact).
Now suppose a tight sequence μ_n. It admits a vaguely convergent subsequence $\mu_{a_n} \to \mu$. Now we only need to show that $\mu(E) = 1$. For every $\epsilon > 0$, we can find K_ϵ so that $\mu_n(K_\epsilon) > 1-\epsilon$ for every n. Then we can find a function $\phi \in \mathcal{C}_c(E)$ with $\mathbb{1}_{K_\epsilon} \leq \phi \leq 1$. We get $\mu(E) \geq \mu\phi = \lim_n \mu_{a_n}\phi \geq 1-\epsilon$. So $\mu(E) = 1$.

(c) If (μ_n) is tight in $\mathbb{R}^{\mathbb{N}}$, then it is a simple matter that the sequences of f.d.m's $(\mathrm{proj}_{I\star}\mu_n)_n$, which are sequences of probability measures on $\mathbb{R}^{\#I}$, are tight too. So by diagonal extraction, we can find a_n and μ_I for each finite I so that $\mathrm{proj}_{I\star}\mu_{a_n} \to \mu_I$ narrowly. When $J \subset I$, we have

$$\mathrm{proj}_{J\star}\mathrm{proj}_{I\star}\mu_{a_n} = \mathrm{proj}_{J\star}\mu_{a_n}.$$

The left-hand side goes to $\mathrm{proj}_{J\star}\mu_I$ by continuous mapping. The right-hand side goes to μ_J. Hence the family of probability measures μ_I verifies the consistency condition $\mathrm{proj}_{J\star}\mu_I = \mu_J$. So there exists a probability measure μ on $\mathbb{R}^{\mathbb{N}}$, with $\mathrm{proj}_I\,\mu = \mu_I = \lim \mathrm{proj}_I\,\mu_{a_n}$ for every I finite. Remark that this implies narrow convergence $\mu_n \to \mu$. Indeed we metrize $\mathbb{R}^{\mathbb{N}}$ by the distance $d(x,y) = \sum_n 2^{-n-1}(|x_n - y_n| \wedge 1)$. For a fixed k and $x \in \mathbb{R}^{\mathbb{N}}$, the distance between x and $\phi_k(x) = (x_1,\ldots,x_n,0,0,0,\ldots)$ is less than 2^{-k}. But the finite dimensional convergence entails that $\phi_{k*}\mu_n \to \phi_{k*}\mu$. Now for $h \in \mathcal{C}_c(\mathbb{R}^{\mathbb{N}})$, we get $|\phi_{k*}\mu_n h - \mu_n h| = |\mu_n(h\circ\phi_k - h)| \leq m(h,2^{-k})$. Similarly, $|\phi_{k*}\mu h - \mu h| \leq m(h,2^{-k})$. Since k is arbitrary, this implies $\mu_n h \to \mu h$ hence vague convergence. And since μ is a probability measure, the convergence is narrow.

(d) Every Polish space E is homeomorphic to $S \subset [0,1]^{\mathbb{N}}$ through $\varphi(x) = (d(x,u_1)\wedge 1, d(x,u_2)\wedge 1,\ldots)$, where $(u_i)_i$ is a dense sequence. If $(\mu_n)_n \in \mathcal{P}(E)^{\mathbb{N}}$ is tight, then $\varphi_*\mu_n$ is too, hence we can find a sequence a_n and a probability measure π on $\mathbb{R}^{\mathbb{N}}$ such that $\varphi_*\mu_{a_n} \to \pi$. Now we just need to check that π is supported by S, i.e $\pi(S) = 1$. But if we look at K_ϵ, we have that $\pi(S) \geq \pi(\varphi(K_\epsilon)) \geq \lim_n \mu_n(K_\epsilon) \geq 1-\epsilon$. So $\pi(S) = 1$. Now we take $\mu = (\varphi_{-1})_*\pi$, and μ is a probability measure on E that appears as a sub-limit of μ_n.

(2) Let $X^{(n)}$ be a sequence of random variables in $\mathcal{C}(\mathbb{R}_+)$ such that

(a) $\sup_n \mathbb{P}(|X^{(n)}(0)| > M) \xrightarrow[M\to\infty]{} 0$

(b) for every $\eta > 0$, $T > 0$, we have $\sup_n \mathbb{P}(m_{[0,T]}(X^{(n)},\delta) > \eta) \xrightarrow[\delta\to 0]{} 0$

Let $\epsilon > 0$. Let M be such that $\mathbb{P}(|X^{(n)}| > M) < \epsilon/2$. For every $k \geq 1$, $m \geq 1$ let $\delta_{k,p} > 0$ be such that

$$\sup_n \mathbb{P}(m_{[0,p]}(X^{(n)}, \delta_{k,p}) > 2^{-k}) < \epsilon 2^{-k-m-100}.$$

Then for every n, with probability over $1-\epsilon$, we have that $X^{(n)}$ belongs to the set of functions f such that $|f(0)| < M$ and for every $m \geq 1, k \geq 1$, $m_{[0,p]}(f, \delta_{k,p}) < 2^{-k}$. This set is relatively compact thanks to Arzela-Ascoli's theorem and a diagonal argument. So for every ϵ, we found a compact set that contains $X^{(n)}$ with probability over $1-\epsilon$ and tightness is proved.

(3) If $(X^{(n)})_n$ is tight and f.d.m's converge to those of X, then fix a subsequence a_n. By tightness and Prokhorov's theorem you can find a further subsequence a_{b_n} and Y so that $X^{(a_{b_n})} \to Y$ in distribution. Now by continuous mapping, the f.d.m's of $X^{(a_{b_n})}$ converge to those of Y. But they also converge to those of X by assumption. So the f.d.m's of X and Y are equal, and $X \stackrel{d}{=} Y$. We have shown that for every subsequence a further subsequence exists on which $X^{(n)} \to X$, proving full convergence.