Some Possible Additions to Section 2.6

Jon A. Wellner

27 October 2006

Theorem: The uniform quantile process can be written as

$$\mathbb{V}_n = -\mathbb{U}_n(\mathbb{G}_n^{-1}) + \sqrt{n}(\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I)\,, \tag{1}$$

and hence for the specially constructed $\mathbb{U}_n$ of Theorem 4.2 it follows that, with $\mathbb{V} \equiv -\mathbb{U} \stackrel{d}{=} \mathbb{U}$,

$$\|\mathbb{V}_n - \mathbb{V}\|_\infty \to_{a.s.} 0 \qquad \text{as} \qquad n \to \infty\,. \tag{2}$$

Proof: First we prove the identity (1):

$$\begin{aligned}
\mathbb{V}_n &= \sqrt{n}(\mathbb{G}_n^{-1} - I) \\
&= \sqrt{n}(\mathbb{G}_n^{-1} - \mathbb{G}_n(\mathbb{G}_n^{-1})) + \sqrt{n}(\mathbb{G}_n(\mathbb{G}_n^{-1}) - I) \\
&= -\mathbb{U}_n(\mathbb{G}_n^{-1}) + \sqrt{n}(\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I)\,.
\end{aligned}$$

Now $\|\mathbb{G}_n^{-1} - I\|_\infty \to_{a.s.} 0$ by Proposition 6.1, and

$$\|\mathbb{G}_n \circ \mathbb{G}_n^{-1} - I\|_\infty = \sup_{0 \le t \le 1} |\mathbb{G}_n(\mathbb{G}_n^{-1}(t)) - t| = \frac{1}{n}\,. \tag{3}$$

Hence

$$\begin{aligned}
\|\mathbb{V}_n - \mathbb{V}\|_\infty &\le \|\mathbb{U}_n(\mathbb{G}_n^{-1}) - \mathbb{U}\|_\infty \\
&\le \|\mathbb{U}_n(\mathbb{G}_n^{-1}) - \mathbb{U}(\mathbb{G}_n^{-1})\|_\infty + \|\mathbb{U}(\mathbb{G}_n^{-1}) - \mathbb{U}\|_\infty + \frac{1}{\sqrt{n}} \\
&\le \|\mathbb{U}_n - \mathbb{U}\|_\infty + \|\mathbb{U}(\mathbb{G}_n^{-1}) - \mathbb{U}\|_\infty + \frac{1}{\sqrt{n}} \\
&\to_{a.s.} 0 + 0 + 0 = 0\,.
\end{aligned}$$

since $\mathbb{U}$ is a continuous (and hence uniformly continuous) function on $[0,1]$. □

On the other hand, the uniform empirical process can be written in terms of the uniform quantile process $\mathbb{V}_n$ as

$$\mathbb{U}_n = -\mathbb{V}_n(\mathbb{G}_n) + \sqrt{n}(\mathbb{G}_n^{-1} \circ \mathbb{G}_n - I)$$

This is easily seen by noting that

$$\begin{aligned}\mathbb{U}_n(t) &= \sqrt{n}(\mathbb{G}_n(t) - t) \\ &= \sqrt{n}(\mathbb{G}_n(t) - \mathbb{G}_n^{-1}(\mathbb{G}_n(t)) + \mathbb{G}_n^{-1}(\mathbb{G}_n(t)) - t) \\ &= -\mathbb{V}_n(\mathbb{G}_n(t)) + \sqrt{n}(\mathbb{G}_n^{-1}(\mathbb{G}_n(t)) - t).\end{aligned}$$

Note that $\mathbb{G}_n^{-1}(\mathbb{G}_n(t)) = \xi_{(i)}$ for $\xi_{(i)} \le t < \xi_{(i+1)}$, and hence

$$\|\mathbb{G}_n^{-1}(\mathbb{G}_n(t)) - t\|_\infty = \max_{1\le i\le n}(\xi_{(i+1)} - \xi_{(i)}) \equiv D_n.$$

with $\xi_{(n+1)} \equiv 1$, the *maximal spacing* corresponding to the uniform spacings $\delta_i \equiv \xi_{(i+1)} - \xi_{(i)}$.

Proposition: With $D_n = \max_{1\le i\le n}(\xi_{(i+1)} - \xi_{(i)})$, we have

$$D_n = O_p\left(\frac{\log n}{n}\right), \tag{4}$$

$$\limsup_{n\to\infty} \frac{nD_n}{\log n} \le 2 \qquad \text{a.s., and} \tag{5}$$

$$\sqrt{n}\|\mathbb{G}_n^{-1}(\mathbb{G}_n) - I\|_\infty = O(n^{-1/2}\log n) \quad a.s. \tag{6}$$

Proof. Let $\delta_{n,i} \equiv \xi_{(i)} - \xi_{(i-1)} \equiv \xi_{n:i} - \xi_{n:i-1}$ for $i = 1, \ldots, n+1$ be the spacings (with $\xi_{(0)} \equiv 0$, $\xi_{(n+1)} \equiv 1$. Then for every $\lambda > 0$

$$\begin{aligned}P(nD_n > \lambda) &= P(\max_{1\le i\le n+1} \delta_{n,i} > \lambda/n) \\ &= P(\cup_{i=1}^{n+1}[\delta_{n,i} > \lambda/n]) \\ &\le \sum_{i=1}^{n+1} P(\delta_{n,i} > \lambda/n) \\ &= (n+1)P(\delta_{n,1} > \lambda/n) = (n+1)P(\xi_{(1)} > \lambda/n) \\ &= (n+1)P(\xi_1 > \lambda/n, \ldots, \xi_n > \lambda/n) \\ &= (n+1)(1-\lambda/n)^n \le (n+1)\exp(-\lambda) \le 2n\exp(-\lambda).\end{aligned}$$

Thus

$$P(nD_n > M\log n) \le 2n\exp(-M\log n) = 2n^{-(M-1)} \to 0$$

for any $M > 1$. Thus (4) holds. If we chose $M = 2 + \epsilon$ with $\epsilon > 0$, then $\sum_{n=1}^\infty P(nD_n > (2+\epsilon)\log n) < \infty$ and hence $P(nD_n > (2+\epsilon)\log n \ \ i.o.) = 0$ by the Borel-Cantelli lemma, and thus (5) holds. This also proves (6). □

Note that Ferguson (1996), section 13, proves asymptotic normality of two sample quantiles by first proving asymptotic normality of two quantiles of uniform order statistics, say $\mathbb{G}_n^{-1}(p_1)$ and $\mathbb{G}_n^{-1}(p_2)$ by exploiting the connection between uniform order statistics and partial sums of exponential(1) random variables:

Proposition: If $Y_1, Y_2, \ldots$ are i.i.d. exponential(1) random variables, and $S_j \equiv \sum_{i=1}^{j} Y_i$ for $j \geq 1$, then

$$\frac{1}{S_{n+1}}(S_1, S_2, \ldots, S_n) \stackrel{d}{=} (\xi_{(1)}, \xi_{(2)}, \ldots, \xi_{(n)}).$$

Thus if we define

$$\widetilde{\mathbb{G}}_n^{-1}(t) \equiv S_i/S_{n+1} \qquad \text{for } \frac{i-1}{n} < t \leq \frac{i}{n}, \quad i = 1, \ldots, n,$$

and

$$\widetilde{\mathbb{V}}_n(t) \equiv \sqrt{n}(\widetilde{\mathbb{G}}_n^{-1}(t) - t),$$

then $\widetilde{\mathbb{G}}_n^{-1} \stackrel{d}{=} \mathbb{G}_n^{-1}$, and $\widetilde{\mathbb{V}}_n \stackrel{d}{=} \mathbb{V}_n$. On the other hand, the study of $\widetilde{\mathbb{V}}_n$ reduces to the study of partial sums of exponential(1) random variables.

If we want to get back to $\mathbb{U}_n$ by this route, with $\tilde{\mathbb{G}}_n \stackrel{d}{=} \mathbb{G}_n$ and $\tilde{\mathbb{U}}_n \stackrel{d}{=} \mathbb{U}_n$ defined in terms of $\{S_j/S_{n+1},\ j = 1, \ldots, n\}$, then we need to deal with the remainder term $\sqrt{n}(\mathbb{G}_n^{-1} \circ \mathbb{G}_n(t) - t)$.

To study this remainder term in a bit more detail we will also consider the following processes. Noting that $\mathbb{G}_n^{-1} \circ \mathbb{G}_n(t) \leq t$, leads to defining a corresponding process that is always above the identity, namely

$$\mathbb{G}_n^{-1}(\mathbb{G}_n(t) + 1/n) = \sum_{i=1}^{n+1} \xi_{(i)} 1_{[\xi_{(i-1)}, \xi_{(i)})}(t)$$

where $\xi_{(0)} \equiv 0$ and $\mathbb{G}_n^{-1}(t) = 1$ for $t > 1$. we define

$$\mathbb{R}_n^L(t) \equiv n(\mathbb{G}_n^{-1} \circ \mathbb{G}_n(t) - t) \leq 0,$$
$$\mathbb{R}_n^U(t) \equiv n(\mathbb{G}_n^{-1}(\mathbb{G}_n(t) + 1/n) - t) \geq 0.$$

Then $\mathbb{R}_n \equiv (\mathbb{R}_n^U + \mathbb{R}_n^L)/2$ is the process which equals $n[(\xi_{(i)} + \xi_{(i-1)})/2 - t]$ on the interval $[\xi_{(i-1)}, \xi_{(i)})$, $i = 1, \ldots, n+1$ (with $\xi_{(n+1)} \equiv 1$). As it turns out, $\mathbb{R}_n$ is a *Laplace noise process* in the following sense: for any $0 < t_1 < \cdots < t_k < 1$ we have

$$(\mathbb{R}_n(t_1), \ldots, \mathbb{R}_n(t_k)) \to (Z_1, \ldots, Z_k)$$

where $Z_1, \ldots, Z_k$ are i.i.d. with common Laplace (or double exponential) density $f_Z(z) = \exp(-2|z|)$ for $z \in R$.

Here are some plots of the processes $\mathbb{G}_n$, $\mathbb{G}_n^{-1}$, $\mathbb{G}_n^{-1} \circ \mathbb{G}_n(t)$, $\mathbb{G}_n^{-1}(\mathbb{G}_n(t) + 1/n)$, $\mathbb{U}_n$, $\mathbb{R}_n$, and $\mathbb{R}_n^U$ for various sample sizes.

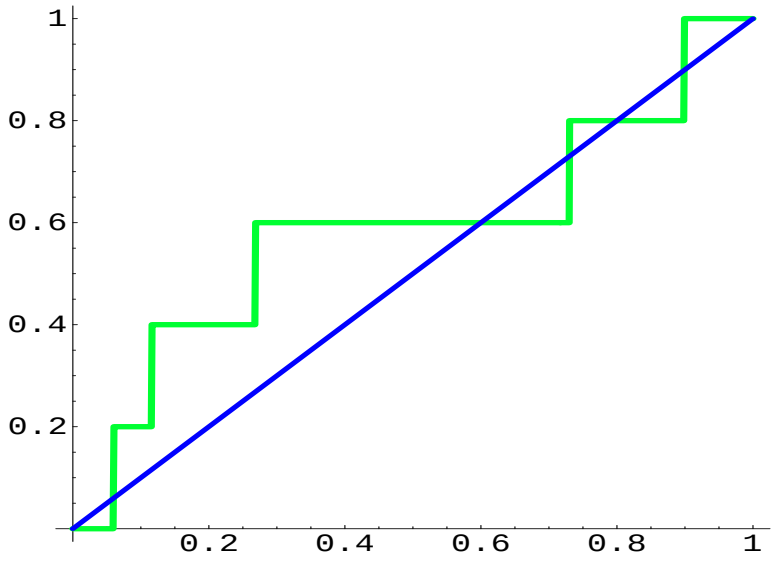

Figure 1: Uniform empirical distribution function $\mathbb{G}_n$, $n = 5$

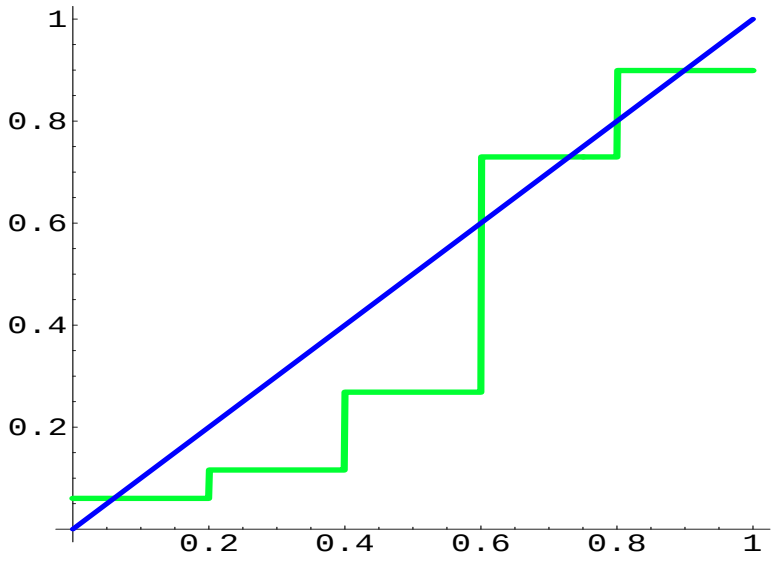

Figure 2: Uniform empirical quantile function $\mathbb{G}_n^{-1}$, $n = 5$

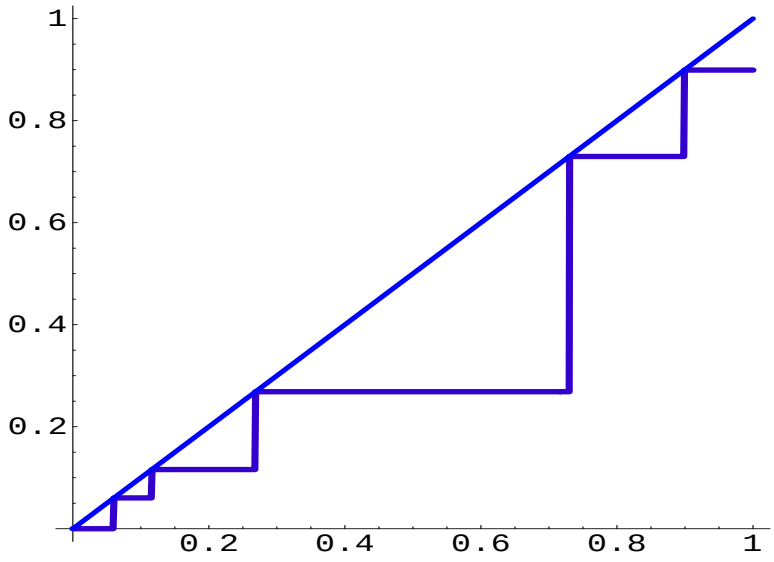

Figure 3: $\mathbb{G}_n^{-1}(\mathbb{G}_n(t))$, $n = 5$

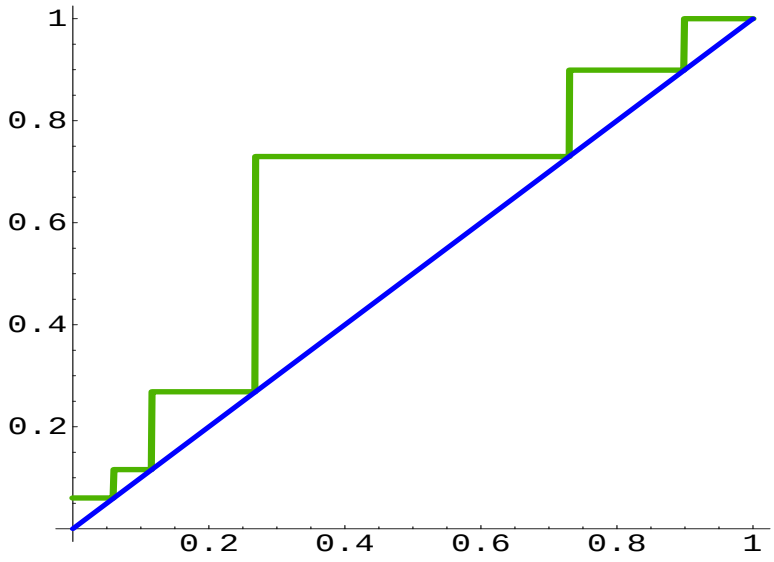

Figure 4: $\mathbb{G}_n^{-1}(\mathbb{G}_n(t) + 1/n)$, $n = 5$

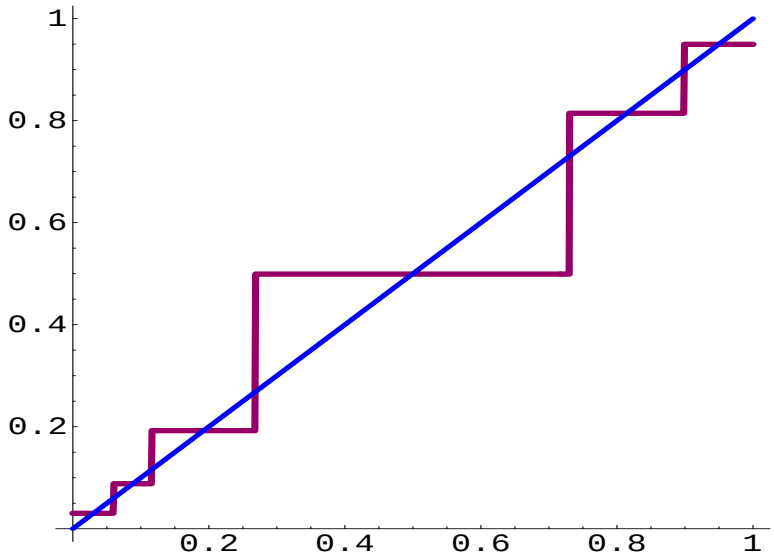

Figure 5: $(1/2)[\mathbb{G}_n^{-1}(\mathbb{G}_n(t)) + \mathbb{G}_n^{-1}(\mathbb{G}_n(t) + 1/n)]$, $n = 5$

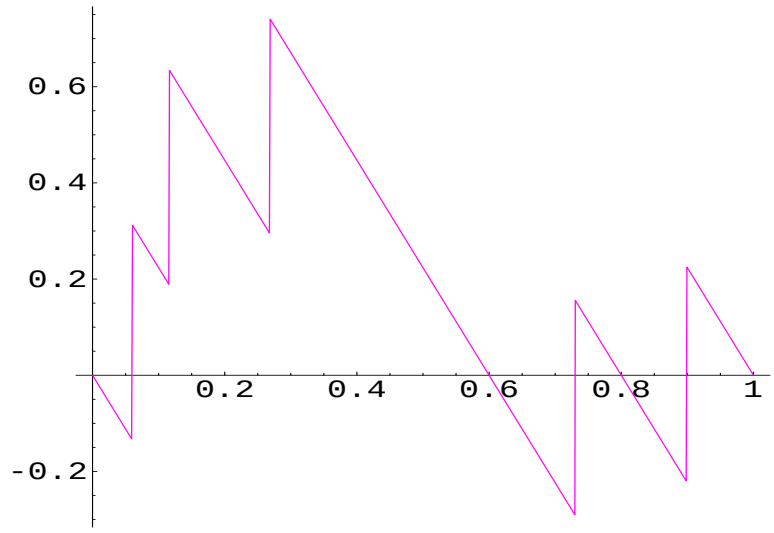

Figure 6: uniform empirical process $\mathbb{U}_n$, $n = 5$

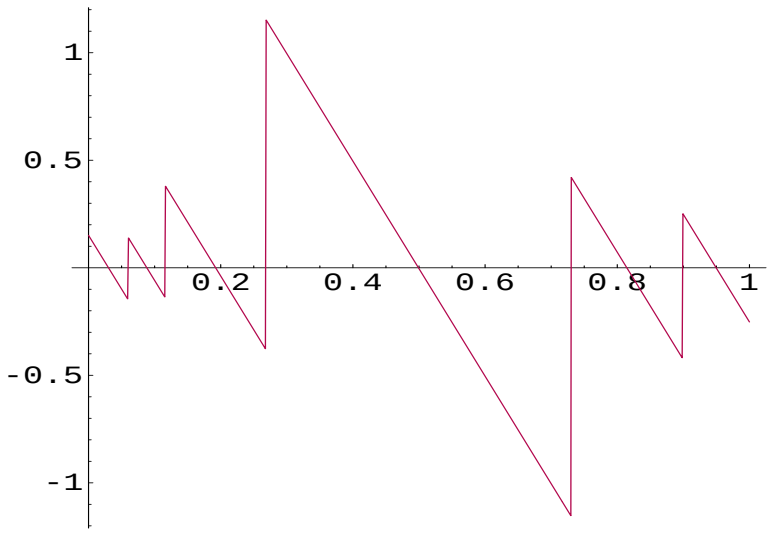

Figure 7: R_n process, $n = 5$

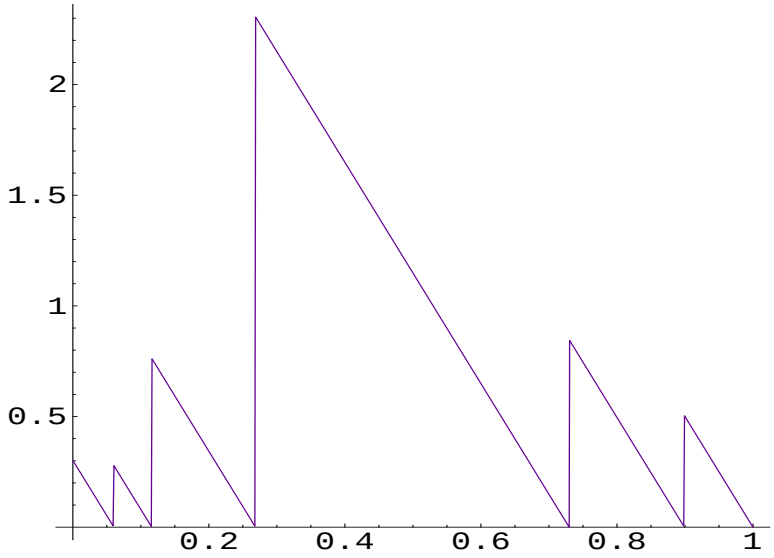

Figure 8: R_n^U process, $n = 5$

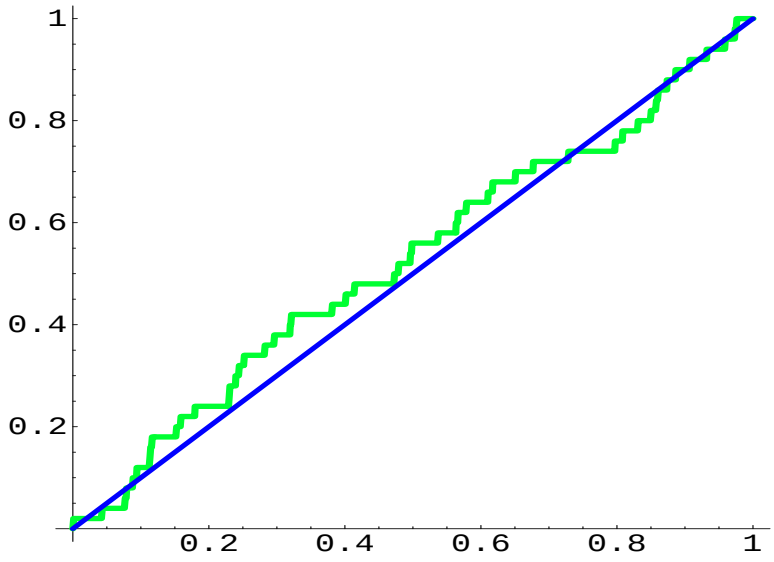

Figure 9: Uniform empirical distribution function $\mathbb{G}_n$, $n = 50$

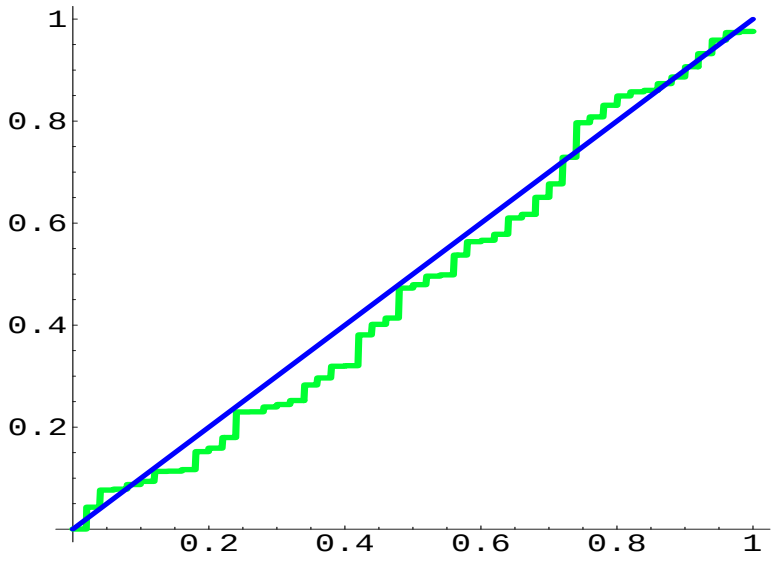

Figure 10: Uniform empirical quantile function $\mathbb{G}_n^{-1}$, $n = 50$

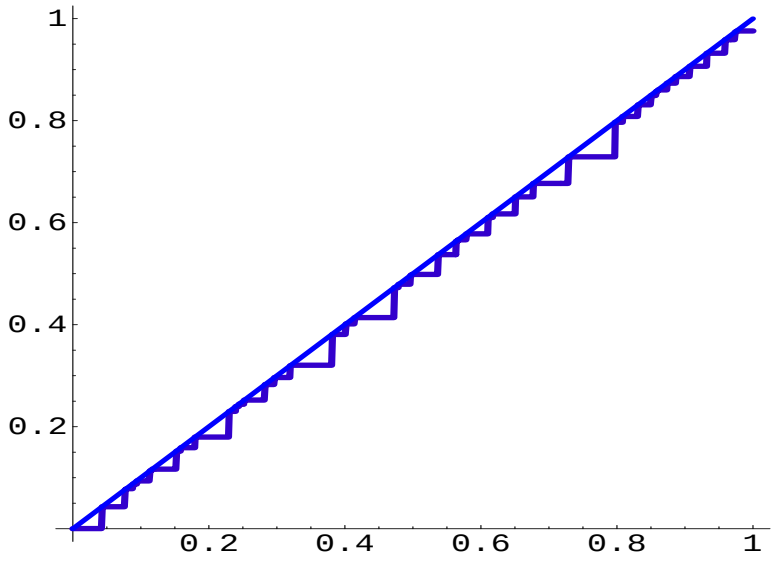

Figure 11: $\mathbb{G}_n^{-1}(\mathbb{G}_n(t))$, $n = 50$

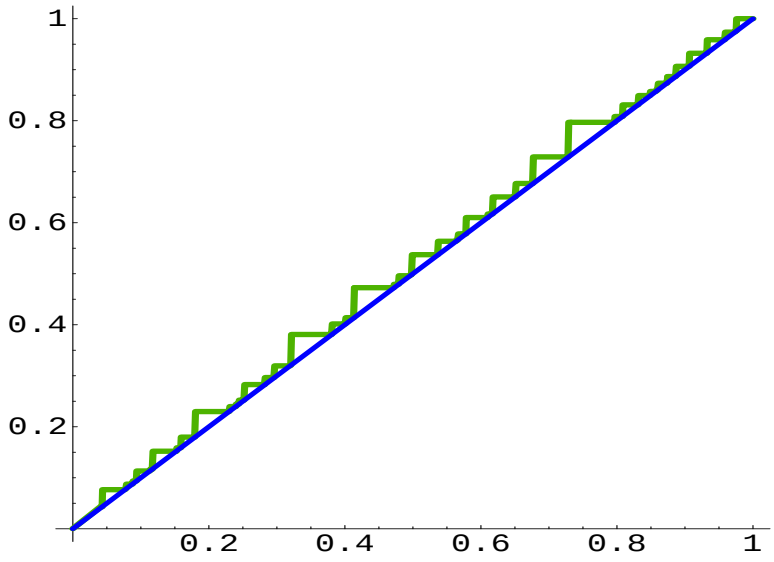

Figure 12: $\mathbb{G}_n^{-1}(\mathbb{G}_n(t) + 1/n)$, $n = 50$

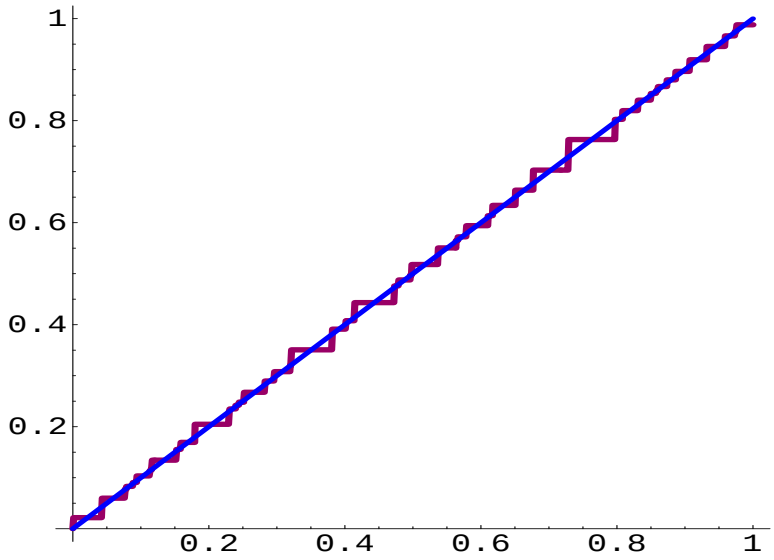

Figure 13: $(1/2)[\mathbb{G}_n^{-1}(\mathbb{G}_n(t)) + \mathbb{G}_n^{-1}(\mathbb{G}_n(t) + 1/n)]$, $n = 50$

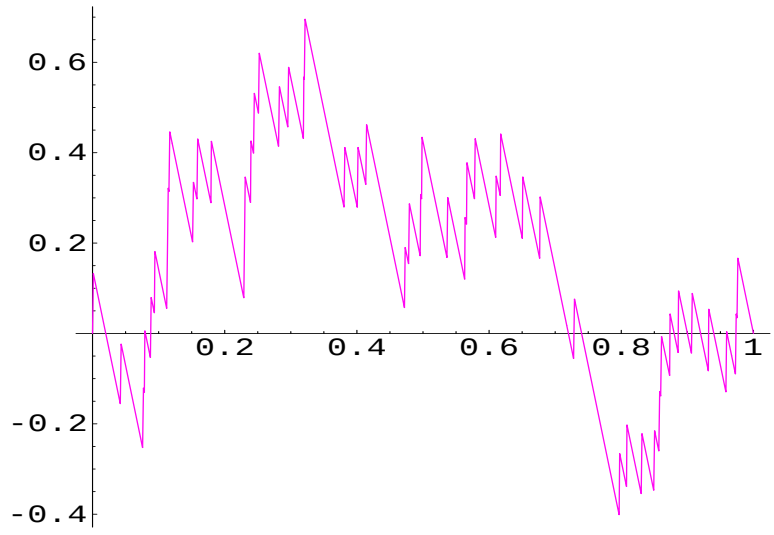

Figure 14: uniform empirical process $\mathbb{U}_n$, $n = 50$

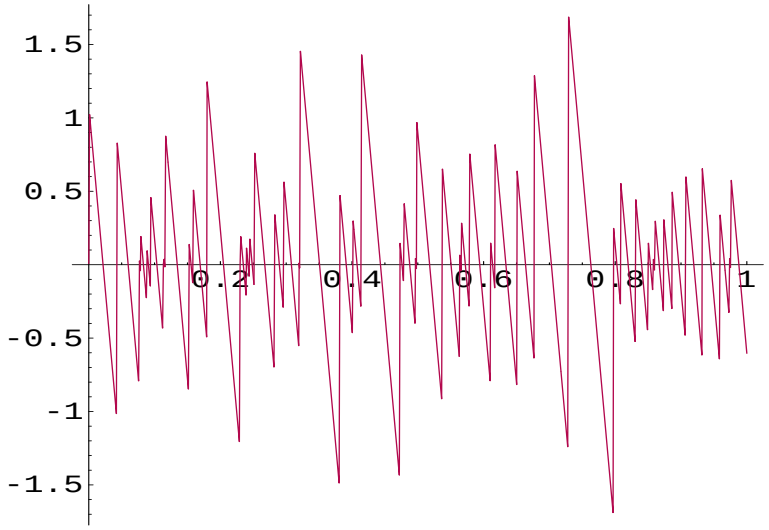

Figure 15: R_n process, $n = 50$

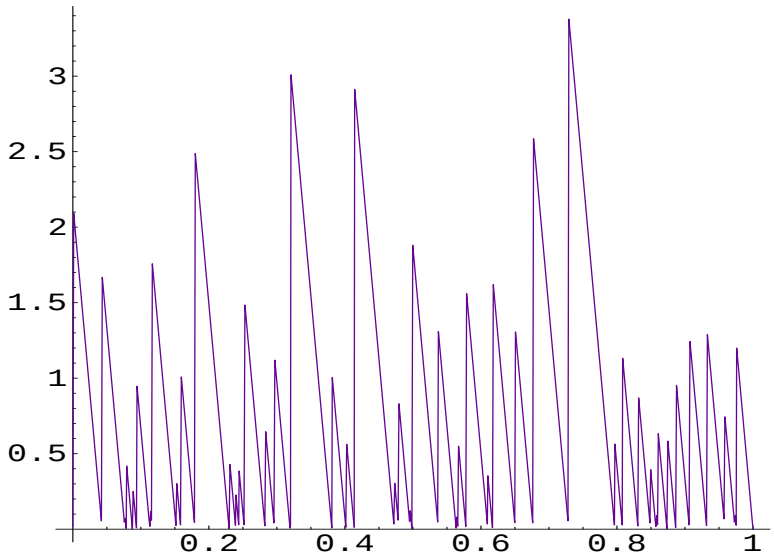

Figure 16: R_n^U process, $n = 50$

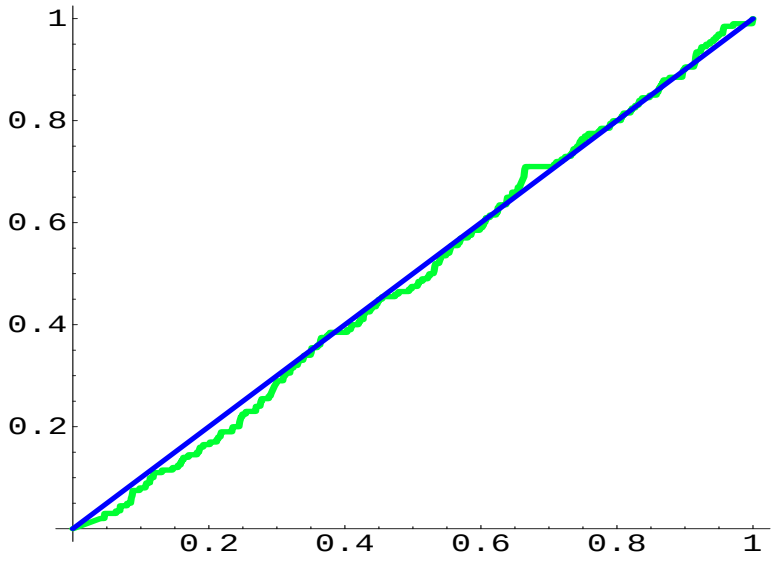

Figure 17: Uniform empirical distribution function $\mathbb{G}_n$, $n = 200$

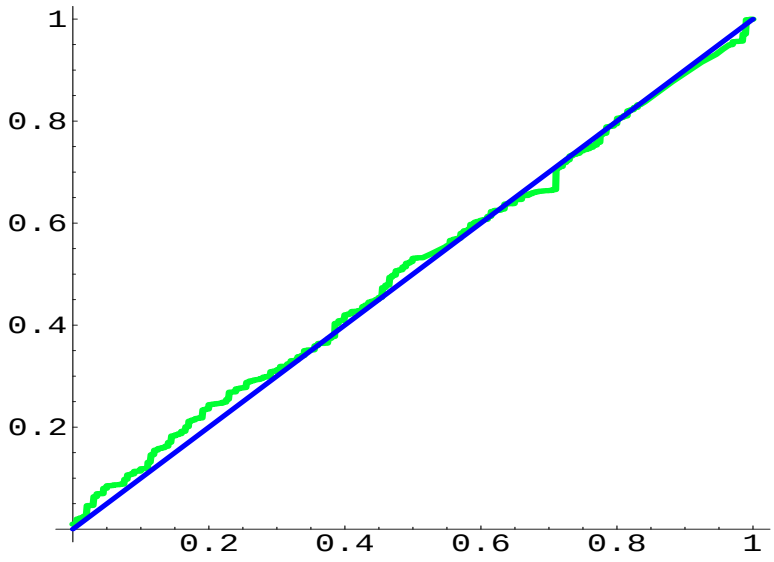

Figure 18: Uniform empirical quantile function $\mathbb{G}_n^{-1}$, $n = 200$

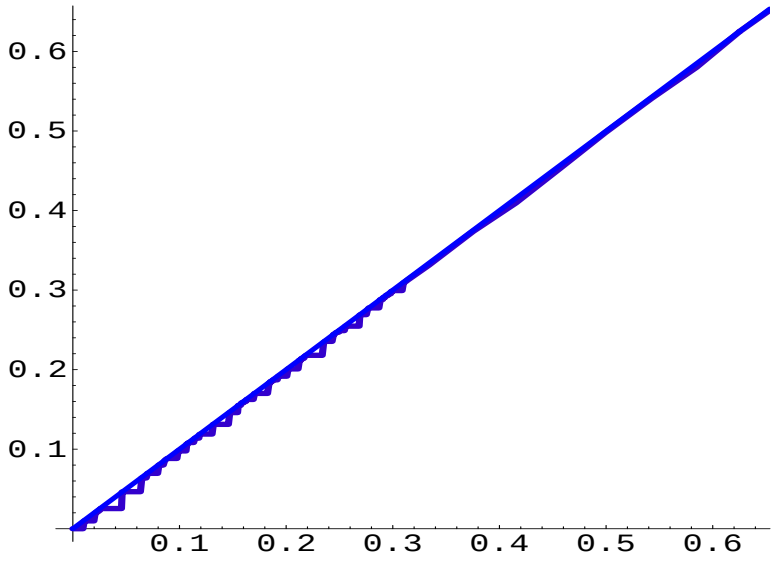

Figure 19: $\mathbb{G}_n^{-1}(\mathbb{G}_n(t))$, $n = 200$

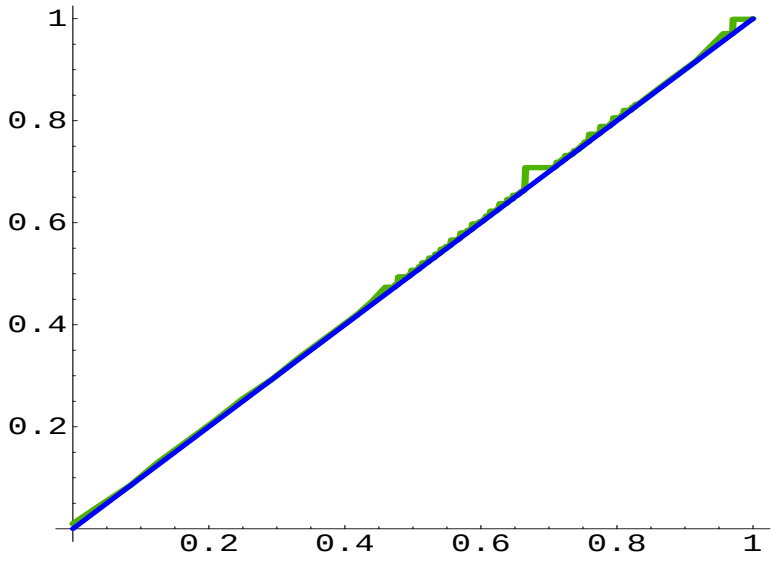

Figure 20: $\mathbb{G}_n^{-1}(\mathbb{G}_n(t) + 1/n)$, $n = 200$

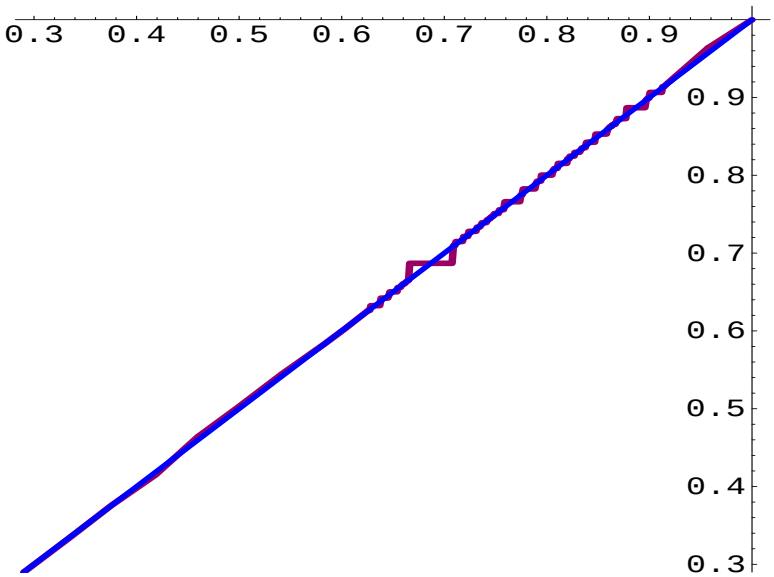

Figure 21: $(1/2)[\mathbb{G}_n^{-1}(\mathbb{G}_n(t)) + \mathbb{G}_n^{-1}(\mathbb{G}_n(t) + 1/n)]$, $n = 200$

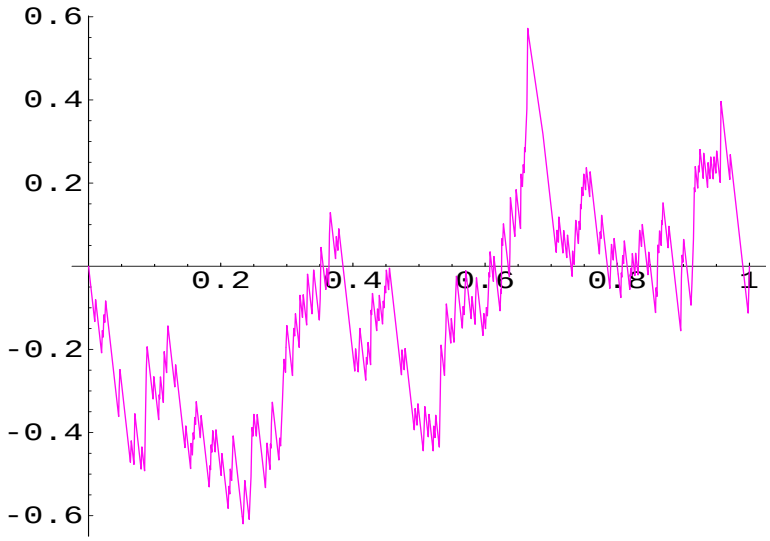

Figure 22: uniform empirical process $\mathbb{U}_n$, $n = 200$

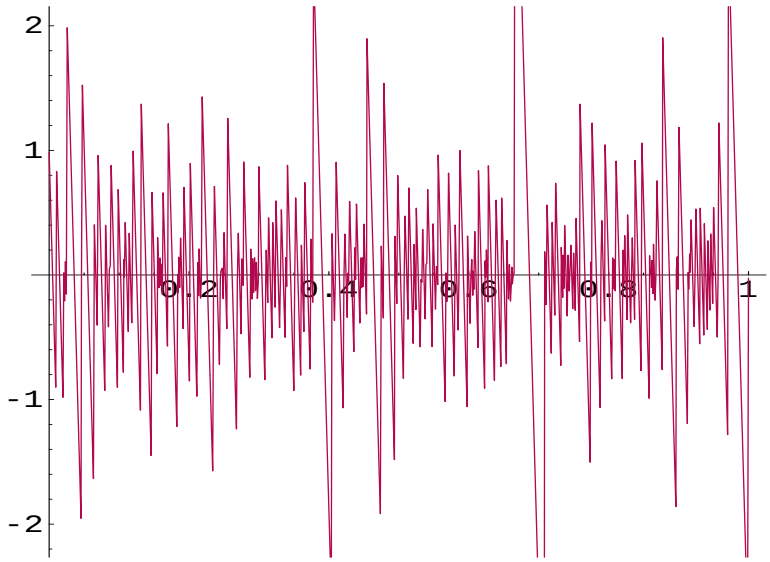

Figure 23: R_n process, $n = 200$

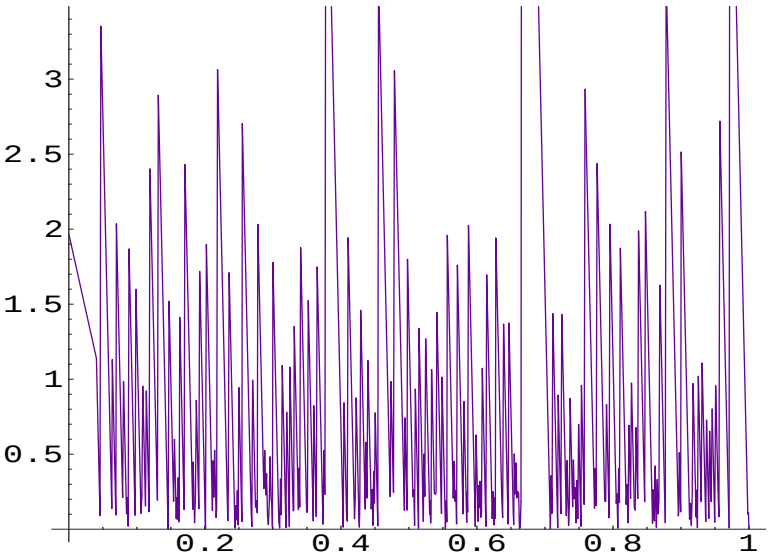

Figure 24: R_n^U process, $n = 200$

Finally, here are some plots of $\mathbb{U}_n$, $\mathbb{V}_n$, $\mathbb{D}_n \equiv \mathbb{U}_n + \mathbb{V}_n$, and $\mathbb{R}_n$ for $n = 200$ and $n = 500$.

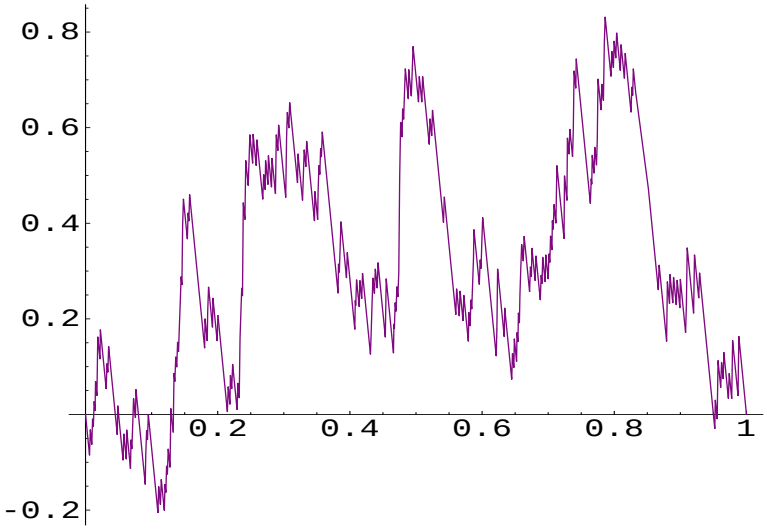

Figure 25: Uniform empirical process $\mathbb{U}_n$, $n = 200$

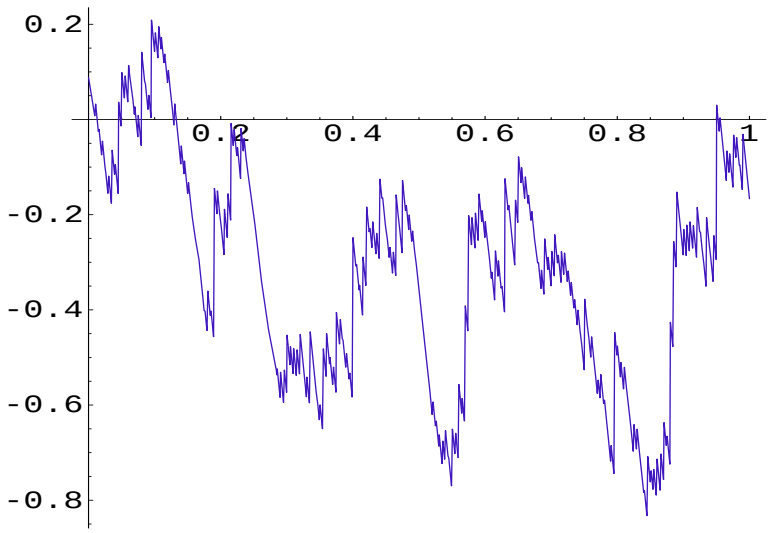

Figure 26: Uniform quantile process $\mathbb{V}_n$, $n = 200$

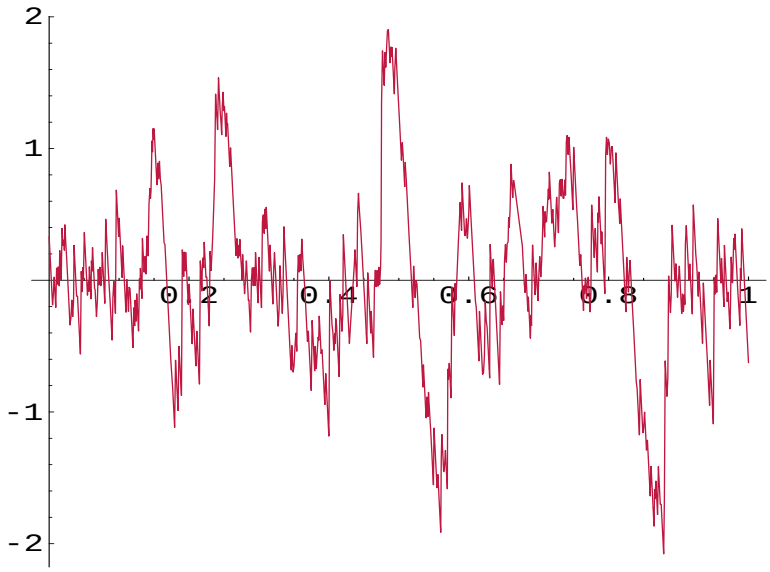

Figure 27: Bahadur-Kiefer process $\mathbb{D}_n$, $n = 200$

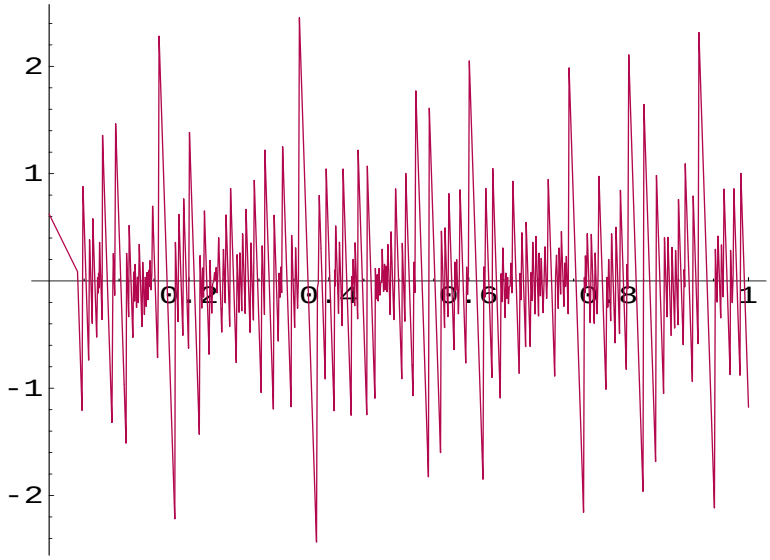

Figure 28: $\mathbb{R}_n$ process, $n = 200$

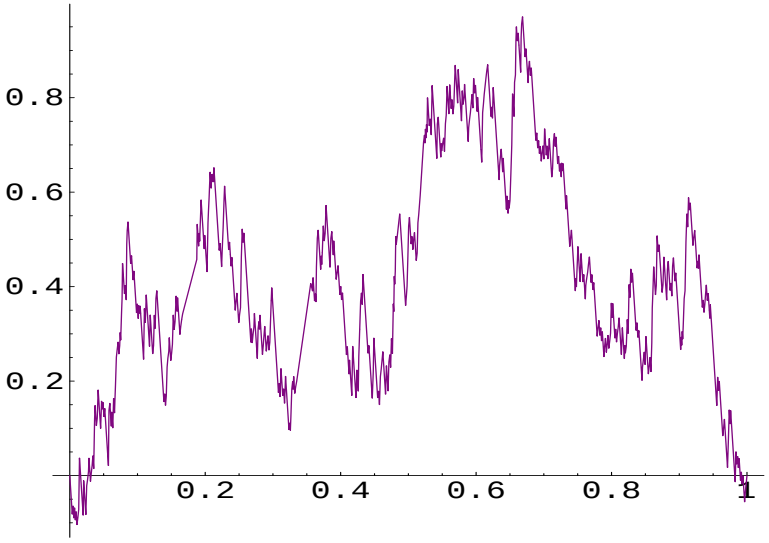

Figure 29: Uniform empirical process $\mathbb{U}_n$, $n = 500$

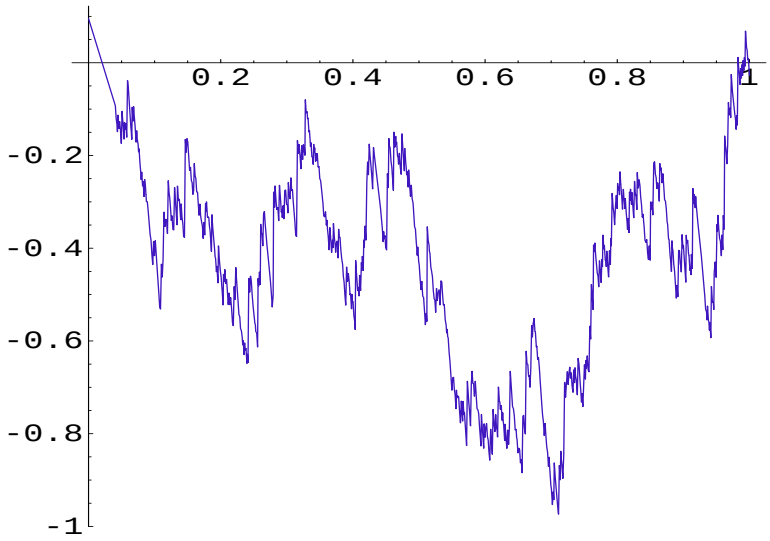

Figure 30: Uniform quantile process $\mathbb{V}_n$, $n = 500$

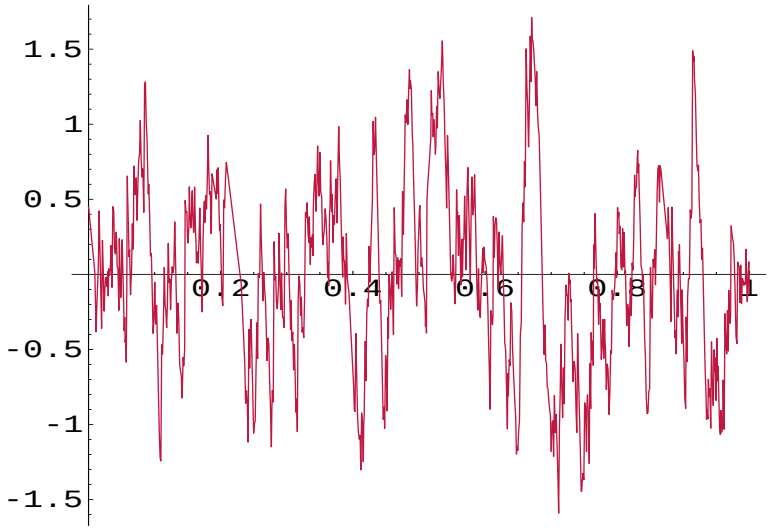

Figure 31: Bahadur-Kiefer process $\mathbb{D}_n$, $n = 500$

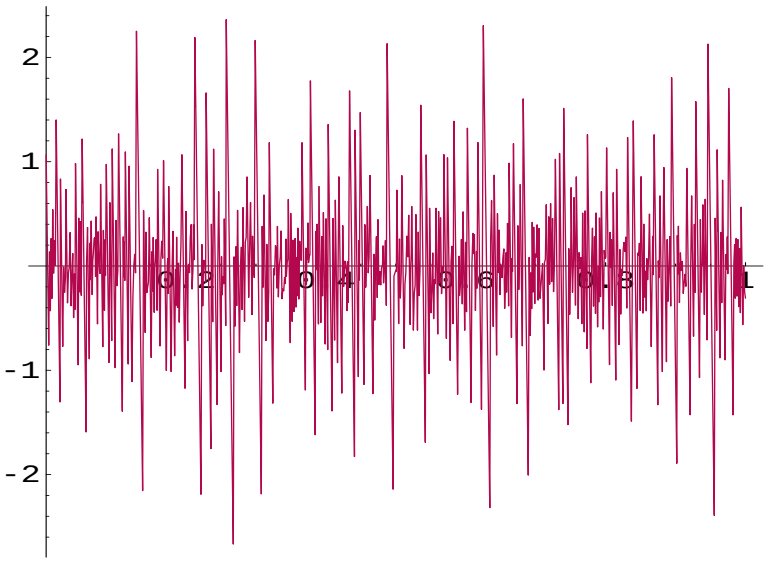

Figure 32: $\mathbb{R}_n$ process, $n = 500$