Chapter 29

A discussion of Willem van Zwet's probability papers

David M. Mason

Abstract I discuss five papers of Willem van Zwet: [19], [20], [13], [21] and [15].

I shall begin my discussion of Willem's probability papers with his 1978 paper on the Kakutani conjecture. Willem tells me that this is his favorite paper. I can see why. It is not only a fine piece of mathematics; it also displays very well a common feature of many of Willem's best papers, namely, it begins with a key insight, which lights the way to the solution of a knotty problem.

The full story of Willem's involvement with the Kakutani conjecture appears in the *Statistical Science* interview of him by Rudy Beran and Nick Fisher (Beran and Fisher (2009)). Therefore I shall not repeat it here. I met Kakutani in Luntenen in 1984, when he was a speaker and I was a participant. By that time his conjecture was solved. He did not lecture about it and we did not discuss it. On the other hand, I have heard Willem talk about it a number of times at Oberwolfach Meetings and other occasions.

Here is a description of the conjecture and a sketch of Willem's beautiful solution. The Kakutani interval splitting is as follows: Let V_1 be a Uniform $(0, 1)$ random variable. Now proceed sequentially, choosing for each integer $n \geq 1$, V_{n+1} uniformly from the largest of the $n + 1$ subintervals into which the previously constructed variables $V_1, \dots, V_n$ partition $(0, 1)$. Consider the empirical distribution function F_n based on the sample $V_1, \dots, V_n$ at stage n . Kakutani conjectured, that with probability 1, as $n \rightarrow \infty$,

$$\sup_{0 \leq t \leq 1} |F_n(t) - t| \rightarrow 0. \quad (29.1)$$

He was, of course, motivated by the Glivenko-Cantelli theorem, which implies that if G_n is the empirical distribution function of $U_1, \dots, U_n$, i.i.d. Uniform $(0, 1)$, then, with probability 1, as $n \rightarrow \infty$,

David M. Mason

Statistics Program, University of Delaware, 213 Townsend Hall Newark, DE 19716
e-mail: davidm@udel.edu This research was partially supported by an NSF Grant.

$$\sup_{0 \leq t \leq 1} |G_n(t) - t| \rightarrow 0. \quad (29.2)$$

It turns out that the Kakutani conjecture is correct.

In van Zwet (1978), Willem devised an ingenious proof of (29.1). Essential to his approach was a stopping time formed as follows: For each integer $n \geq 1$ let $0 = V_{0,n} \leq V_{1,n} \leq \dots \leq V_{n,n} \leq V_{n+1,n} = 1$ denote the order statistics of $V_1, \dots, V_n$, and introduce the $n+1$ spacings $D_{i,n} = V_{i,n} - V_{i-1,n}$, $i = 1, \dots, n+1$. For $t > 0$, define the stopping time $N_t = \min \{n : \max_{1 \leq i \leq n+1} D_{i,n} \leq t\}$. We shall write N for the process N_t , $t > 0$. Notice that N_t is the smallest integer such that all of the subintervals have length at most t . Crucial to his proof was the distributional identity

$$N_t = {}_d N_{t/V}^{(1)} + N_{t/(1-V)}^{(2)} + 1, \quad (29.3)$$

where $N^{(1)}, N^{(2)}$ and V are independent, with $N^{(1)} = {}_d N^{(2)} = {}_d N$ and V being Uniform $(0, 1)$. The distributional identity (29.3), in turn, leads to differential equations whose solutions give that for some $c > 0$,

$$\mu(t) := EN_t = 2/t - 1 \text{ for } 0 < t < 1 \text{ and } \sigma^2(t) := \text{Var}(N_t) = c/t \text{ for } 0 < t < 1/2. \quad (29.4)$$

Here is a synopsis of Willem's proof of (29.1). For any $x \in (0, 1]$ let $N_t(x) = N_t F_{N_t}(x)$. It is easy to argue that for any $x \in (0, 1]$ there exist distributionally equivalent versions of $N'_{t/x}$ of $N_{t/x}$ and $N'_{t/(1-x)}$ of $N_{t/(1-x)}$ such that

$$N'_{t/x} \leq N_t(x) \leq N_t - N'_{t/(1-x)}. \quad (29.5)$$

Set $t_m = m^{-2}$ for $m \geq 1$. Using Chebyshev's inequality and the identities in (29.4) he shows that for any sequences $\{N'_{t_m}\}$, $\{N'_{t_m/x}\}$ and $\{N'_{t_m/(1-x)}\}$, where for $m \geq 1$, $N'_{t_m} = {}_d N_{t_m}$, $N'_{t_m/x} = {}_d N_{t_m/x}$ and $N'_{t_m/(1-x)} = {}_d N_{t_m/(1-x)}$, with probability 1,

$$t_m N'_{t_m}/2 \rightarrow 1, \quad t_m N'_{t_m/x}/2 \rightarrow x \text{ and } t_m N'_{t_m/(1-x)}/2 \rightarrow 1 - x.$$

This implies by the inequalities in (29.5) that, with probability 1, $F_{N_{t_m}}(x) \rightarrow x$, from which (29.1) follows by a routine argument, noting that, with probability 1, $N_{t_m}/N_{t_{m+1}} \rightarrow 1$.

In Pyke and van Zwet (2004), Ron Pyke and Willem considered the question of the weak convergence of the *Kakutani empirical process*,

$$U_n(x) = \sqrt{n} \{F_n(x) - x\}, \quad x \in [0, 1].$$

They prove that U_n converges weakly to σB , written $U_n \rightarrow_d \sigma B$, where B is a standard Brownian bridge on $[0, 1]$ with $\sigma = \sqrt{4 \ln 2 - 5/2} = .5221003$, which is approximately one half as large as in the i.i.d. case. Recall that the *uniform empirical process*

$$\alpha_n(x) = \sqrt{n} \{G_n(x) - x\}, \quad x \in [0, 1], \quad (29.6)$$

converges weakly to B . This shows that though the Kakutani sampling procedure does not differ from the usual i.i.d. situation at the Glivenko-Cantelli level in the sense that both (29.2) and (29.1) hold, it does at the asymptotic distributional level. Roughly speaking, this says that the Kakutani scheme leads to more regularly spaced points than in the i.i.d. uniform model.

Once again the stopping time N_t is a basic ingredient in the proof. Now higher moments are required and these are obtained through the cumulants of $N_t + 1$. They extend the method in van Zwet (1978) used to derive (29.4) to show that for $m \geq 1$, the m -th cumulant of $N_t + 1$ is

$$k_m(t) = c_m/t \text{ for } 0 < t \leq 1/m, \text{ with } mc_m = k_m(1/m).$$

It was also necessary to ferret out the underlying process that drives the weak convergence of U_n , which they found to be the following: For $0 < t < s < 1$ let

$$S_t(x; s) = (t/2)^{1/2} \sum_{V_{i,N_s} \leq x} \left\{ N_{t/D_{i,N_s}}^{(i)} - \mu(t/D_{i,N_s}) \right\}, \quad x \in [0, 1],$$

where $N^{(1)}, N^{(2)}, \dots$, are i.i.d. N .

Conditioned on D_{i,N_s} , $i = 1, \dots, N_s + 1$, $S_t(\cdot; s)$ is clearly a partial sum process on $[0, 1]$ based on independent summands, with jumps at the V_{i,N_s} . By a careful analysis, applying a classical weak convergence result, they prove that as long as $t(s) = o(s)$ as $s \searrow 0$ at a proper rate, the process $S_t(\cdot; s)$ converges weakly, both conditionally and unconditionally, to σW , where W is a standard Brownian motion. After that is established, they complete the proof by verifying that random jump times can be replaced by deterministic ones and the appropriate pieces can be fit together properly to conclude that $U_n \rightarrow_d \sigma B$.

Pyke (1980) had shown that the empirical distribution of the normalized spacings $(n+1)D_{i,n}$, $i = 1, \dots, n+1$, of the Kakutani sample converges to the Uniform $(0, 2)$ distribution, whereas in the i.i.d. case it goes to the exponential distribution with mean 1. Pyke and van Zwet (2004) establish that the corresponding empirical process converges weakly to a mean zero Gaussian process with a complicated covariance function. They also treat the empirical process of the relative spacings $D_{i,n}/M_n$, where $M_n = \max\{D_{i,n} : i = 1, \dots, n+1\}$. They do this by adapting the methods that they used to prove $U_n \rightarrow_d \sigma B$. Sadly this was Ron Pyke's last paper.

Willem and I first met at an *N.S.F. Regional Conference* in 1982 in Eugene, Oregon, where he was the featured speaker. However, my first professional contact with Willem was in 1979, when I wrote him for a preprint of his paper van Zwet (1980), which is on a strong law of large numbers for linear functions of orders statistics. These are statistics of the form

$$\mathbb{L}_n = \sum_{i=1}^n g(U_{i,n}) \int_{(i-1)/n}^{i/n} J_n(u) du,$$

where g is a measurable function on $(0, 1)$ and $\{J_n\}$ is a sequence of measurable functions on $(0, 1)$ that converge in an appropriate sense to a function J . At that

time, proving central limit theorems and strong laws for $\mathbb{L}_n$ was still in vogue. The game was to balance smoothness conditions on g with those on J .

Willem came up with an elegant solution to the strong law problem for $\mathbb{L}_n$ in van Zwet (1980). To state his main results, introduce the centering constants

$$\mu_n = \int_0^1 J_n g d\lambda \text{ and } \mu = \int_0^1 J g d\lambda.$$

(We denote Lebesgue measure by λ and assume here whatever conditions needed to insure that the integrals are finite.) The following result is a combination of Theorem 2.1 and Corollary 2.1 of van Zwet (1980).

Theorem 1 *Let $1 \leq p \leq \infty$, $p^{-1} + q^{-1} = 1$, and suppose that $J_n \in L_p(d\lambda)$ for $n \geq 1$, and $g \in L_q(d\lambda)$. If either*

- (i) $1 < p \leq \infty$ and $\sup_n \|J_n\| < \infty$, or
- (ii) $p = 1$ and $\{J_n\}$ is uniformly integrable,

then with probability 1

$$\mathbb{L}_n - \mu_n \rightarrow 0. \quad (29.7)$$

Moreover, if there is a $J \in L_p(d\lambda)$ such that $\int_0^t J_n g d\lambda \rightarrow \int_0^t J g d\lambda$ for every $t \in (0, 1)$. Then we can replace μ_n by μ in (29.7).

The key technical result used in his proof was the following lemma based on Lusin's theorem: With $[x]$ denoting the integer part of x , define for each integer $n \geq 1$,

$$g_n(t) = g(U_{[tn]+1,n}) \text{ for } t \in (0, 1).$$

Lemma 1 *With probability 1, g_n converges to g in Lebesgue measure λ on $(0, 1)$, i.e. for all $\delta > 0$,*

$$\lim_{n \rightarrow \infty} \lambda \{t : |g_n(t) - g(t)| > \delta\} = 0.$$

Armed with this result, the proof of (29.7) is short and sweet. Some classical criteria for weak convergence in L_p , $1 \leq p < \infty$, and weak* convergence in L_∞ play a role. Willem then shows how this result and its extensions imply the strong laws for $\mathbb{L}_n$ of Helmers (1977), Wellner (1977) and Sen (1978).

I published my own strong law for $\mathbb{L}_n$ in Mason (1982), in which for certain subclasses of linear functions of order statistics I provided necessary and sufficient conditions for (29.7) to hold. I also considered $\mathbb{L}_n$ whose g functions are not in $L_q(d\lambda)$ for any $q \geq 1$.

I next discuss the most significant of my two joint papers with Willem. This is Mason and van Zwet (1987), where we obtained the following refinement of the Komlós, Major and Tusnády [KMT] (1975) Brownian bridge approximation to the uniform empirical process α_n . Recall the definition of α_n given in (29.6).

Theorem 2 *There exists a probability space $(\Omega, \mathcal{A}, P)$ with independent Uniform $(0, 1)$ random variables $U_1, U_2, \dots$, and a sequence of Brownian bridges $B_1, B_2, \dots$, such that for all $n \geq 1$, $1 \leq d \leq n$, and $-\infty < x < \infty$,*

$$P \left\{ \sup_{0 \leq t \leq d/n} |\alpha_n(t) - B_n(t)| \geq n^{-1/2}(a \log d + x) \right\} \leq b \exp(-cx) \quad (29.8)$$

and with the same inequality holding with the supremum is taken over $1 - d/n \leq t \leq 1$, where a, b and c are suitable positive constants independent of n, d and x .

Setting $d = n$ into these inequalities yields the original KMT inequality. Our result drew a fair amount of interest and has been applied in a number of papers. I note that as of this writing google scholar lists 54 citations to our paper, which is pretty good for a theoretical paper. Rio (1994) has computed values for the constants in (29.8). Later, Castelle, and Laurent-Bonvalot (1998) obtained a similar refinement to the KMT Kiefer process approximation to α_n .

M. Csörgő, S. Csörgő, Horváth and Mason [Cs-Cs-H-M](1986) had earlier constructed a probability space on which an analog to these inequalities holds with α_n replaced by β_n (the uniform quantile process). We shall not define β_n in this discussion.

Willem and I did much of the work on our paper during the month of March 1985, while I was visiting him in Leiden. At the time, I was on extended leave from the University of Delaware and was in the middle of the first year of a two year stay at the University of Munich supported by an Alexander von Humboldt stipend.

Our progress was severely hampered by the fact that KMT (1975) had only provided a bare sketch of the proof of their original inequality. We spent an enormous amount of time filling in the missing details. This became an obsession to me. After two weeks in Leiden watching me consumed with this project night and day, my wife at the time got fed up with me and returned to Munich.

Complete proofs are now available. Mason and van Zwet (1987) combined with Mason (2001a), which contains additional details based on Leiden notes, provides a proof. Also consult Bretagnolle and Massart (1989), Major (1999) and Dudley (2000). Bretagnolle and Massart (1989) determined values of the constants in the original KMT inequality. For more on the history of proofs refer to Mason (2007).

Mason and van Zwet (1987) pointed out that their inequality leads to the following useful weighted approximation: For any $0 \leq v < 1/2$, $n \geq 1$, and $1 \leq d \leq n - 1$ let

$$\Delta_{n,v}(d) := \sup_{d/n \leq t \leq 1-d/n} \frac{n^v |\alpha_n(t) - B_n(t)|}{(t(1-t))^{1/2-v}}. \quad (29.9)$$

On the probability space of Theorem 2, one has

$$\Delta_{n,v}(1) = O_p(1). \quad (29.10)$$

Cs-Cs-H-M (1986) had obtained a version of this result on their probability space under the restriction that $0 \leq v < 1/4$. Later, motivated by a question of Evarist Giné, I derived the following improved version of the Mason and van Zwet weighted approximation (29.10). I published this in the van Zwet Festschrift (see Mason (2001b)).

Theorem 3 *On the probability space of Theorem 2, for every $0 \leq v < 1/2$ there exist positive constants A_v and C_v such that for all $n \geq 2$, $1 \leq d \leq n-1$ and $0 \leq x < \infty$,*

$$P\{\Delta_{n,v}(d) \geq x\} \leq 2A_v \exp(d^{1/2-v}C_v) \exp\left(-d^{1/2-v}cx/4\right). \quad (29.11)$$

One of the key inequalities in Mason and van Zwet (1987) that was essential to derive the refinement of KMT was the following: Let $(X_1, \dots, X_{k+1})$ with $k \geq 1$ be a multinomial random vector with parameters $n, p_1, \dots, p_{k+1}$. Assume that $p_i > 0$ for $i = 1, \dots, k$ and let $s = \sum_{i=1}^k p_i \in (0, 1]$.

Lemma 2 *For every $C > 0$ and $\delta > 0$, there exist positive numbers a, b and λ such that for all $n \geq 1, k \geq 1$ and positive $s, p_1, \dots, p_k$ satisfying $z \leq Cn \min\{p_1, \dots, p_k\}$ and $s \leq 1 - \delta$*

$$P\left(\sum_{i=1}^k \frac{(X_i - np_i)^2}{np_i} > z\right) < a \exp(bk - \lambda z). \quad (29.12)$$

This is Lemma 3 of Mason and van Zwet (1987). It was just what we needed to complete the proof of inequality (29.8) and it required us some time to formulate. (The final result was Willem's doing.) Therefore I was surprised when I found it quoted in the 1996 monograph on empirical processes by Aad van der Vaart and Jon Wellner (van der Vaart and Wellner (1996)), with a remark that it follows from Talagrand's general empirical process inequality.

Willem and I also published a nice paper on the strong approximation to the renewal process. It was also begun during my Leiden visit. Unfortunately it had a controversial history, which I will not discuss here.

Finally I must say something about the remaining paper in the collection of Willem's papers that are classified as 'probability'. This is van Zwet (1994). In this little gem can be found an elegant solution to the following problem:

Consider n point charges, each with charge $1/n$, in electrostatic equilibrium on the surface S of a conducting sphere. Show that, as n tends to infinity, the distribution of the total charge 1 on S tends to the uniform distribution on S .

Though this is an entirely deterministic result, the proof is probabilistic in nature. I will conclude this discussion with some informative and personal remarks by Willem about the history of this paper:

Let me also say something about the paper classified as 'probability', to wit the paper on point charges on a conducting sphere. In 1975 Herbert Robbins spoke at Lunteren. At the end of his talk he proposed a few problems, one of these being: prove that n electrons become uniformly distributed on the sphere as n tends to infinity. He said this problem had been bothering him for years.

Some time later I saw the same problem again, this time as a prize problem of the Dutch Mathematical Society. I found a simple proof that is a funny mixture of potential theory and probability, and sent it to the mathematics society. They awarded me the prize, but the jury report mentioned that the problem had been solved before by a well-known Dutch analyst called Jaap Korevaar.

It turned out that Korevaar had suggested the problem to the mathematics society. (He also heard it from Robbins.) They had used it for a prize problem for a number of years without receiving a solution. Then Korevaar found a solution himself, but forgot to tell the mathematics society about this, so they continued running the problem as a prize question.

They were pretty embarrassed, and though Korevaar's proof was totally different from mine (lots of calculations), we both buried our proofs in a desk drawer. However, 17 years later I still liked my proof and thought 'what the hell' and spoke about it at the Purdue Symposium in 1992, mentioning Korevaar's proof too, of course. David Siegmund was present and after the meeting he told Robbins about my talk. Robbins' – as usual cynical – comment was 'My god, did it take him 20 years to solve this?'

This is Willem, the inimitable storyteller. It has been a pleasure writing these pages.

References

1. BERAN, R. J. and FISHER, N. I. (2009). An evening spent with Bill van Zwet. *Statistical Science* **24** 87–115.
2. BRETAGNOLLE, J. and MASSART, P. (1989). Hungarian constructions from the nonasymptotic viewpoint. *Ann. Probability* **17** 239–256.
3. CASTELLE, N. and LAURENT-BONVALOT, F. (1998). Strong approximations of bivariate uniform empirical processes. *Ann. Inst. Henri Poincaré* **34** 425–480.
4. CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. and MASON, D. M. (1986). Weighted empirical and quantile processes. *Ann. Probability* **14** 31–85.
5. DUDLEY, R. M. (2000). An exposition of Bretagnolle and Massart's proof of the KMT (1975) theorem for the uniform empirical process. In: Notes on empirical processes, lectures notes for a course given at Aarhus University, Denmark, August 2000.
6. HELMERS, R. (1977). A strong law of large numbers for linear combinations of order statistics. Report SW 50/77, Mathematisch Centrum, Amsterdam.
7. KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** 111–131.
8. MAJOR, P. (1999). The approximation of the normalized empirical distribution function by a Brownian bridge, <http://www.renyi.hu/~major/probability/empir.html>
9. MASON, D. M. (1982) Some characterizations of strong laws for linear functions of order statistics. *Ann. Probability* **10** 1051–1057.
10. MASON, D. M. (2001a) Notes on the KMT Brownian bridge approximation to the uniform empirical process. *Asymptotic methods in probability and statistics with applications* (St. Petersburg, 1998), 351–369, Stat. Ind. Technol., Birkhäuser, Boston, Boston, MA.
11. MASON, D. M. (2001b) An exponential inequality for a weighted approximation to the uniform empirical process with applications. *State of the art in probability and statistics* (Leiden, 1999), 477–498, IMS Lecture Notes Monogr. Ser., 36, Inst. Math. Statist., Beachwood, OH.
12. MASON, D. M. (2007). Some observations on the KMT dyadic scheme. (Special issue on Nonparametric Statistics and Related topics in honor of M. L. Puri.) *J. Statist. Plann. Inf.* **137** 895–906.
13. MASON, D. M. and VAN ZWET, W. R. (1987). A refinement of the KMT inequality for the uniform empirical process. *Ann. Probability* **15** 871–884.
14. PYKE, R. (1980). The asymptotic behavior of spacings under Kakutani's model for interval subdivision. *Ann. Probability* **8** 157–163.

15. PYKE, R. and VAN ZWET, W. R. (2004). Weak convergence results for the Kakutani interval splitting procedure *Ann. Probability* **32** 380–423
16. RIO, E. (1994). Local invariance principles and their application to density estimation. *Probab. Theory Related Fields* **98** 21–45.
17. SEN, P. K. (1978). An invariance principle for linear combinations of order statistics. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **42** 327–340.
18. VAART, VAN DER, A.W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
19. VAN ZWET, W. R. (1978). A proof of Kakutani's conjecture on random subdivision of longest intervals. *Ann Probability* **6** 133–137.
20. VAN ZWET, W. R. (1980). A strong law for linear functions of uniform order statistics. *Ann Probability* **8** 986–990.
21. VAN ZWET, W. R. (1994). The asymptotic distribution of point charges on a conducting sphere. *Statistical Decision Theory and Related Topics*, V, S. S. Gupta and J. O. Berger editors, Springer New York 427–430.
22. WELLNER, J. A. (1977). A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics. *Ann. Statist.* **5** 473–480.