Chapter 4
Discussion of three statistics papers by Willem van Zwet

Jon A. Wellner

Abstract I discuss three statistics papers of Willem van Zwet: [12], [5], and [15].

4.1 Introduction

I first met Willem at the 2nd conference on *Statistical decision theory and related topics*, held at Purdue University in May 1976. After a brief discussion over dinner on the topic of my dissertation (concerning certain limit theorems for linear combinations of order statistics), Willem tactfully pointed out that perhaps I had missed some interesting problems of a somewhat more fundamental nature concerning strong laws for such linear combinations. This brief conversation lead to [16]. Willem himself beautifully improved my results in [14] as discussed by David Mason elsewhere in this volume.

Beyond giving good advice, Willem is well-known to many for his story-telling abilities, both in his papers and over a beer in a corner at Oberwolfach. The three papers discussed here provide ample evidence of the former (with hints of the latter, especially in [15]). The reader interested in more of the latter should consult [1].

4.2 Paper 1.

The first of these three papers, *Convex transformations: a new approach to skewness and kurtosis*, is based on [13]. It gives a wonderfully clear exposition of partial orderings for distribution functions (or their associated random variables) which

Jon A. Wellner
Department of Statistics, University of Washington, Seattle, WA 98195-4322 e-mail: jaw@stat.washington.edu Supported in part by NSF Grant DMS-0804587, and by NI-AID grant 2R01 AI291968-04

S. van de Geer and M. Wegkamp (eds.), *Selected Works of Willem van Zwet*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1314-1_4,

"cover our intuitive ideas about skewness and kurtosis", and have a variety of further statistical applications.

Briefly, for distribution functions F, F^* which are twice continuously differentiable on some interval I with $F'(x) > 0$ on I, $F \overset{c}{<} F^*$ if and only if G^*F is convex on I where G^* is the inverse function of F^* defined by $G^*F^*(x) = x$. Similarly, for the subclass of all distribution functions as above which are symmetric about some point x_0, $F \overset{s}{<} F^*$ if and only if G^*F is convex for $x \ge x_0$, $x \in I$, where x_0 is the (common) point of symmetry. It is easily seen from the forward and inverse probability integral transformations that if X has distribution function F, then $X^* \equiv G^*F(X)$ has distribution function F^*, and hence it is natural to write $X \overset{c}{<} X^*$ whenever $F \overset{c}{<} F^*$.

For $X \sim F$ and a positive integer k let

$$\begin{aligned}
\gamma_{2k+1}(F) &\equiv \gamma_{2k+1}(X) \equiv \frac{E(X-EX)^{2k+1}}{\sigma^{2k+1}(F)}, \\
\gamma_{2k}(F) &\equiv \gamma_{2k}(X) \equiv \frac{E(X-EX)^{2k}}{\sigma^{2k}(F)}, \quad \text{where} \\
\sigma^2(F) &\equiv \sigma^2(X) \equiv E(X-EX)^2,
\end{aligned}$$

assuming that the expectations exist. Thus $\gamma_1(F)$ is the classical skewness of F and $\gamma_2(F)$ is the kurtosis of F. The paper [12] starts with the basic results

$$\begin{aligned}
&\gamma_{2k+1}(X) \le \gamma_{2k+1}(\varphi(X)) \ \text{ for any convex function } \ \varphi, \ \text{ and} \\
&\gamma_{2k}(X) \le \gamma_{2k}(\varphi(X)) \ \text{ for any convex, odd about } x_0, \text{ function } \ \varphi \\
&\qquad\qquad\qquad\qquad\qquad \text{if } \ X \sim F \ \text{ symmetric about } \ x_0.
\end{aligned}$$

These results are discussed heuristically and used to motivate the definitions of $F \overset{c}{<} F^*$ and $F \overset{s}{<} F^*$. A natural choice of φ is exactly G^*F.

Willem himself writes about this paper:

> This is a short summary of my dissertation. Over the years, the Centrum voor Wiskunde en Informatica (Center for Mathematics and Computer Science) at Amsterdam has sold 800 copies. The reason is that the topic is revisited every ten years or so. Among other things, the thesis deals with a partial ordering of one-dimensional probability distributions that produces an increasing skewness to the right, and discusses a few simple consequences of this ordering. Nowadays I would formulate this as an ordering in terms of the fatness of the tail rather than in terms of skewness. The subject will doubtless enjoy yet another lifetime due to the current interest in heavy tails by queuing theory folks, financial mathematics people, et cetera. The thesis is now out of print, but it should be available in the libraries of some statistics departments.

Despite the recent comprehensive book [10], Willem's paper and his thesis [13] remain gems of the stochastic orderings literature.

Here is a conjecture related to van Zwet's $\overset{s}{<}$ ordering:

Conjecture: Let X have Chernoff's distribution as described in [6]; this distribution arises as the limit distribution in a variety of problems involving monotone nonparametric function estimation. Let Z be a random variable with a standard normal distribution (with mean zero, variance 1). Both X and Z have distributions symmetric about 0. I conjecture that $X \overset{s}{<} \sigma Z$ and that $f_X(t) = h(t)\varphi(t/\sigma)/\sigma$ with $h-$log-concave if $\sigma \geq .52$.

4.3 Paper 2.

The "two-armed bandit problem", apparently introduced in [8], is as follows: you are presented with a slot machine with two arms. One arm yields a payoff of \$1 with probability α and the other arm yields a payoff of \$1 with probability β. The rub is that you do not know which arm is connected with these probabilities, and you also don't know the values of α and β. The goal is to maximize your expected winnings in N successive pulls of one or the other of the two arms. Alternatively, if you are very patient and have lots of time to play the machine, you may have the goal of maximizing your limiting average expected winnings as N is allowed to become large. It has been known since [8] that there exist strategies achieving the latter goal: if X_k denotes the winnings from play k, then there is decision rule or strategy for choosing one or the other of the two arms so that

$$\frac{1}{N}\sum_{k=1}^{N} X_k \to \max\{\alpha,\beta\} \quad \text{as} \quad N \to \infty$$

with probability one; see e.g. [7] and [4]. Finding optimal strategies for finite N is somewhat more difficult, but perhaps more important for a variety of real problems. If you have played both arms by step $m < N$, then playing the arm which has yielded the smaller winnings so far results in sub-optimal winnings in the next step, but a strategy involving always "playing the winner" can also be sub-optimal, as was shown by [3]. The results of these authors prompt Fabius and van Zwet to write:

> Though these relations may seem intuitively evident, one does well to remember that the two-armed bandit problem has been shown to defy intuition in many aspects (cf. [3]).

Fabius and van Zwet formulate the two-armed bandit problem in a general decision theoretic setting allowing randomized decision rules and an arbitrary prior distribution for (α,β) on $[0,1]^2$. They proceed by characterizing the class of all Bayes rules, and show (Theorem 4) that every admissible strategy is Bayes agains a "non-marginal prior distribution" π. They give an explicit example showing that "... there is an essentially unique and hence admissible Bayes strategy against π which violates (the monotonicity requirements) (i) and (ii) (of "play the winner" rules)... ", thereby reconfirming the results of [3]. Fabius and van Zwet go on to provide wonderfully explicit calculations of minimax symmetric rules and risk for $N = 3$ and $N = 4$.

For further development of these problems and themes, see [4], [9], and the survey by [7].

4.4 Paper 3.

In this delightful historical article Willem reviews the work of the Dutch astronomer Van de Hulst on the behavior of trimmed means, and the wonderful interactions between Van de Hulst and the imminent Dutch mathematician and statistician D. van Dantzig.

In connection with this paper Willem writes:

> At some time during the early 1980s I gave a talk for a general sciences audience. As many scientists routinely remove outliers from their data, I thought it might be useful to speak about trimmed means and what happens if you use them. In the talk I showed them the derivation of the asymptotic variance of the trimmed mean. There was a spirited discussion afterwards. To my utter surprise, Van de Hulst – a world-famous astronomer from Leiden – shows up in my office a few days later carrying a small notebook written in 1942 that contained precisely this asymptotic result. From a mathematical point of view, the proof left something to be desired, but the right ideas were all there. In 1942 he apparently knew all about M-estimators too, and this knowledge goes back to Nobel laureate Zernike in 1928. In his 1942 notes Van de Hulst also showed that what is now known as Huber's estimator has the same asymptotic variance as the trimmed mean. After Huber's estimator had been introduced, statisticians first believed that its asymptotic variance would coincide with that of the Winsorized mean until Bickel proved Van de Hulst's result in 1965 ([2]). Van de Hulst is justifiably pleased by the recognition provided in this paper and has shown it to all of his astronomy friends!.

Willem's article sketches the theory of "M-estimators" that was apparently well-known in the Dutch astronomy community in the 1930's and 1940's and that was used as a starting point by Van de Hulst in his investigations. A proof of the theorem concerning the asymptotic variance of an "M-estimator" was not given in the known reference, so Van de Hulst provided one. But van Dantzig felt that Van de Hulst's proof was not "rigorous". Willem provides a fascinating commentary on the interactions between the two scientists, with a very readable introduction to the theory (translated into modern notation and terminology), including connections between "M-estimators" (or "Z- estimators" as they are renamed slightly in [11]) and trimmed means via Bahadur's representation theorem for quantiles. In the last section of the paper Willem's intimate familiarity with second order expansions and correction terms comes into play in an elegant and subtle re-analysis of the results of Van de Hulst and empirical data concerning trimmed means from Hertzsprung.

I commend the article to the reader as a superb example of Willem at his storytelling best!

References

1. Beran, R. J. and Fisher, N. I. (2009). An evening spent with Bill van Zwet. *Statist. Sci.* **24** 87–115.
2. Bickel, P. J. (1965). On some robust estimates of location. *Ann. Math. Statist.* **36** 847–858.
3. Bradt, R. N., Johnson, S. M. and Karlin, S. (1956). On sequential designs for maximizing the sum of n observations. *Ann. Math. Statist.* **27** 1060–1074.
4. Duflo, M. (1996). *Algorithmes Stochastiques, Mathématiques & Applications* **23**, Springer-Verlag, Berlin.
5. Fabius, J. and van Zwet, W. R. (1970). Some remarks on the two-armed bandit. *Ann. Math. Statist.* **41** 1906–1916.
6. Groeneboom, P. and Wellner, J. A. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10** 388–400.
7. Pemantle, R. (2007). A survey of random processes with reinforcement. *Probab. Surv.* **4** 1–79 (electronic).
8. Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58** 527–535.
9. Rosenberger, W. F. (1996). New directions in adaptive designs. *Statistical Science* **11** 137–149.
10. Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders.* Springer Series in Statistics, Springer, New York.
11. van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics* Springer Series in Statistics, Springer-Verlag, New York.
12. van Zwet, W. R. (1964a). Convex transformations: A new approach to skew- ness and kurtosis. *Statistica Neerlandica* **18** 433–441.
13. van Zwet, W. R. (1964b). *Convex transformations of Random Variables.* Vol. 7 of Mathematical Centre Tracts. Mathematisch Centrum, Amsterdam.
14. van Zwet, W. R. (1980). A strong law for linear functions of order statistics. *Ann. Probab.* **8** 986–990.
15. van Zwet, W. R. (1985). van de Hulst on robust statistics: a historical note. *Statist. Neerlandica* *39* 81–95.
16. Wellner, J. A. (1977). A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics. *Ann. Statist.* **5** 473–480.