Minimum Support ICA Using Order Statistics. Part II: Performance Analysis

Frédéric Vrins and Michel Verleysen*

Université catholique de Louvain, Machine Learning Group,
Place du Levant, 3, 1380 Louvain-la-Neuve, Belgium
{vrins, verleysen}@dice.ucl.ac.be

Abstract. Linear instantaneous independent component analysis (ICA) is a well-known problem, for which efficient algorithms like FastICA and JADE have been developed. Nevertheless, the development of new contrasts and optimization procedures is still needed, e.g. to improve the separation performances in specific cases. For example, algorithms may exploit prior information, such as the sparseness or the non-negativity of the sources. In this paper, we show that support-width minimization-based ICA algorithms may outperform other well-known ICA methods when extracting bounded sources. The output supports are estimated using symmetric differences of order statistics.

1 Introduction and Motivation

Most of ICA researchers and practitioners agree with the idea that it does not exist a unique ICA algorithm outperforming all alternatives, and making the other methods useless. Obviously, certain approaches, like e.g. FastICA [11] or JADE [12] yield remarkable separation performances while simultaneously being fast. Nevertheless, at least three arguments for developing new ICA contrasts can be emphasized, even for the simplest (but also most widely used) linear, instantaneous and noise-free mixture scheme [10]. First, to extend the field of application of BSS techniques (specific procedures have been derived to deal with e.g. structured gaussian sources). Second, some contrasts can be handled easier than others; for example, the convexity property simplifies the optimization step. Third, the contrast performances may vary with the source densities, so that the separation performances depend on the cost function and on the application.

For example, we can cite BSS methods exploiting the non-negativity [9] or sparseness [8] of the sources, as well as their temporal dependency [13], etc. The minimum support approach has been independently suggested by Cruces & Duran [14] and Vrins et al. [1], to extract bounded sources in a deflation way. The theoretical framework has been well established; this approach has relationship to zero Renyi's entropy, and also with the Young and Brunn-Minkowski inequalities. On the other hand, this approach benefits from the discriminacy property, i.e. all the local optima of the theoretic criterion are relevant for ICA.

* Michel Verleysen is Research Director of the Belgian F.N.R.S.

J. Rosca et al. (Eds.): ICA 2006, LNCS 3889, pp. 270–277, 2006.

This property gives confidence in the solution obtained when the optimum is reached using gradient techniques. This is not the case for example when separating multimodal sources by minimizing the entropy or the mutual information [4]. It is interesting to note that the boundness prior on the sources have been used by Theis and Gruber to establish separability results in postnonlinear mixture schemes [7]. In addition, this approach can also be used to separate signals being correlated in some specific way, such as landscape images [3]. Finally, a related symmetric method with geometrical interpretation can be found in the nice paper of Pham [5].

However, the performances of the minimum-support ICA method on bounded source signals have not been detailed and compared to other methods. Similarly, the support estimation problem, which is a crucial issue though, is not discussed in this context. In this paper, we compare the performances of FastICA and maximum absolute kurtosis maximized using [6] (AKMICA) to minimum support algorithms called XSICA, OSICA and AVOSICA. In the three last algorithms, the support measure criterion are estimated in different ways, and is minimized using the optimization technique for non-differentiable criteria presented in [6]. We also analyze the performances of JADE, though it is a rather different method (highly limited by the number of sources, symmetric, algebraic and thus non-iterative).

We show that in the instantaneous noise-free and noisy cases, AVOSICA benefits from interesting signal interference ratio (SIR) performances results in comparison to other ICA algorithms, without added complexity.

2 The XSICA, OSICA and AVOSICA Algorithms

The recent minimum support approach to ICA requires support estimation; in [1,3] the statistical range is used, i.e. the output supports are estimated by the difference of the output extreme values. When this criterion is minimized using [6], we call this algorithm XSICA (extreme statistics ICA). Nevertheless, extreme values can be unreliable in the noisy case, so that alternative ways to estimate bounded support widths have to be derived. This can be easily done by using order statistics differences. The i-th order statistic of an observed sequence $\mathcal{X}_N = \{x_1, ..., x_N\}$ is noted $x_{(j)}$ and is the j-th largest observed sample, i.e. $\{x_{(1)} \leq \cdots \leq x_{(N)}\}$ [17]. The latter sequence is no other than an ordered version of the set $\mathcal{X}_N$. If we note by $R_m(X)$ ($1 \leq m < \lfloor N/2 \rfloor$, $m \in \mathbb{Z}$) the quasi-range defined by $x_{(N-m+1)} - x_{(m)}$, both the quantities $R_m(X)$ and $\langle R_m(X) \rangle \triangleq 1/m \sum_{i=1}^{m} R_m(X)$ can be seen as support width estimators, where m is a tuning parameter. Combining those criteria with the optimization procedure [6], the OSICA (order statistics ICA) and AVOSICA (average order statistics ICA) are obtained. Note that XSICA, OSICA and AVOSICA are equivalent when setting $m = 1$. In $\langle R_m(X) \rangle$, m equals twice the number of sample points used in the support estimation. The estimation of the support convex hull width by $\langle R_m(X) \rangle$ is analyzed in [2]; it is shown to be preferred to $R_m(X)$, but the performances of those practical criterion in terms of SIR are not discussed. In addition, no specific information about how to choose the tuning parameter m in $\langle R_m(X) \rangle$

is given. A small value of m cancels the regularization induced by the average, so that the criterion could be highly sensitive to noise; on the other hand, an excessive value of m may lead the algorithm to totally fail. Furthermore, even if only a small error is observed for large N and small m when estimating the support of a given random variable (r.v.), the shape of the function that links the variance of the estimator to m depends of the (unknown) pdf of the r.v.; the variance can either increases or decreases with m [2].

In the remaining of this paper, a meaningful procedure for choosing a satisfactory value for m given N is derived in Section 3. The performances of AVOSICA are then pointed out, in comparison to XSICA, OSICA, AKMICA, JADE and FastICA using the *gauss* non-linearity, for robustness purpose [10] (the *tanh* non-linearity gives similar results).

3 Towards a Meaningful Choice of m with Fixed N

In this section, we derive a procedure to set a default value for the tuning parameter m, for fixed N. We propose to find the maximum value m_0 of m given N, ensuring that the positive error $\mu[\Omega(X)] - \langle R_m(X)\rangle$ is lower than an error threshold $\mathcal{E}$ with a high probability, whatever is the density of X. In other words, we try to find m_0 such that for all $m \leq m_0$:

$$\Pr\left[\mu[\Omega(X)] - \langle R_m(X)\rangle \leq \mathcal{E}\right] \geq \mathcal{L}(m_0) \ , \tag{1}$$

where $\mathcal{L}(m_0)$ is a threshold ideally close to, but lower than one.

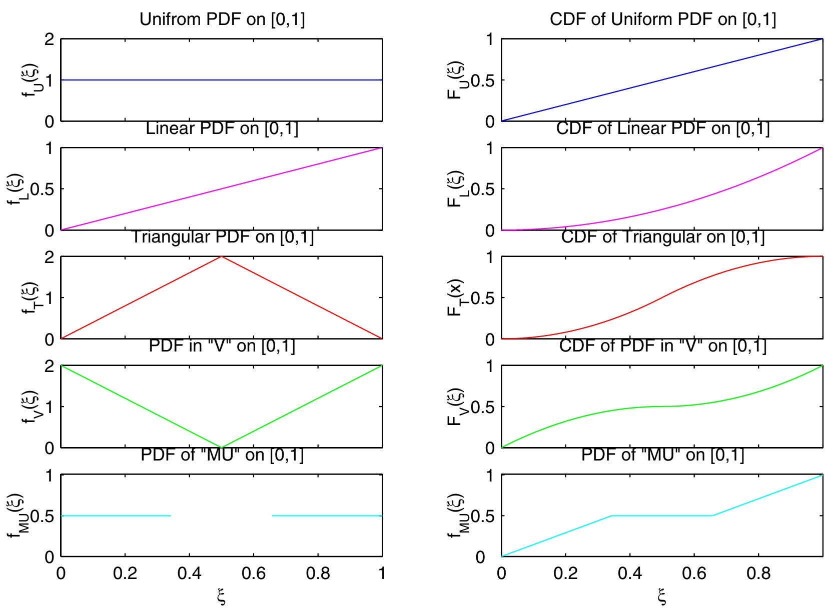

Fig. 1. Densities and (cumulative) distributions of the 5 sources

The main problem of this approach is that if $\mathcal{E}$ is a constant, we are not able to find an expression for $\mathcal{L}(m_0)$ that is i) useful, and ii) *distribution-free*, in the sense that it does not depends on f_X. For instance, the probability in (1) can be written as $1 - F_{\langle R_m(X)\rangle}(\mu[\Omega(X)] - \mathcal{E})$ where $F_{\langle R_m(X)\rangle}$ is the (cumulative) distribution of $\langle R_m(X)\rangle$, which depends on f_X through the order statistic densities $f_{X_{(i)}}$. The point is thus to include the density dependency into the error term $\mathcal{E}$. Let us approximate the support measure by using quantile differences, and define the error term as

$$\mathcal{E}(X) \triangleq \mu[\Omega(X)] - (\xi_q - \xi_p) \ , \tag{2}$$

where ξ_q and ξ_p ($0 \leq p < q \leq 1$) are the q-th and p-th quantiles of F_X, respectively. Note that $\mathcal{E}(X)$ is positive and tends to 0 for increasing q and decreasing p, whatever is the density of X, but at a various rate. For example, with $q = .95$ and $p = 1 - q$ we have $\mathcal{E}(T) = 31.6\%$ and $\mathcal{E}(V) = 5\%$ (see Fig. 1).

Observe that any lower bound of $\Pr[R_m(X) \geq \mu[\Omega(X)] - \mathcal{E}]$ can be used in the right hand side of eq. 1:

$$\begin{aligned}\Pr[\langle R_m(X)\rangle \geq \mu[\Omega(X)] - \mathcal{E}] &= \Pr\Big[\langle R_m(X)\rangle \geq \mu[\Omega(X)] - \mathcal{E} | R_m(X) \geq \mu[\Omega(X)] - \mathcal{E}\Big] \\ &\quad \times \Pr[R_m(X) \geq \mu[\Omega(X)] - \mathcal{E}] \\ &\quad + \Pr\Big[\langle R_m(X)\rangle \geq \mu[\Omega(X)] - \mathcal{E} | R_m(X) < \mu[\Omega(X)] - \mathcal{E}\Big] \\ &\quad \times \Pr[R_m(X) < \mu[\Omega(X)] - \mathcal{E}] \\ &\geq \Pr[R_m(X) \geq \mu[\Omega(X)] - \mathcal{E}] \ ,\end{aligned} \tag{3}$$

where the inequality result from the fact that $\langle R_m(X)\rangle \geq R_m(X)$ with probability one.

On the other hand, using the confidence interval for quantiles derived in [16], noting that $\Pr\Big[R_m(X) \geq R_{m_0}(X) | m \leq m_0\Big] = 1$ and setting $p = 1 - q$ in (2), the following inequality holds for for all $m \leq m_0$:

$$\Pr[R_m(X) \geq \xi_q - \xi_{1-q}] \geq \underbrace{\sum_{i=m_0}^{N} \binom{N}{i} q^{N-i}(1-q)^i - \sum_{i=N-m_0+1}^{N} \binom{N}{i} q^i(1-q)^{N-i}}_{\triangleq \mathcal{L}(q, m_0, N)}$$

and consequently, using inequality (3) and $\mathcal{E}(X)$ given by (2):

$$\begin{aligned}\Pr\Big[\mu[\Omega(X)] - \langle R_m(X)\rangle \leq \mathcal{E}(X)\Big] &= \Pr[\langle R_m(X)\rangle \geq \xi_q - \xi_{1-q}] \\ &\geq \mathcal{L}_+(q, m_0, N) \ ,\end{aligned} \tag{4}$$

with $\mathcal{L}_+(q, m_0, N) \triangleq \max(0, \mathcal{L}(q, m_0, N))$. The latter inequality can be understood as follows: if q is chosen close enough to one, $\langle R_m(X)\rangle$ *nearly covers* the true support, with a probability higher than $\mathcal{L}_+(q, m_0, N)$. Note that q has to be chosen close enough to one, so that $\mathcal{E}(X)$ is small; otherwise the bound $\mathcal{L}_+$ in (4) is no more related to support estimation quality. The terms *close enough to one* depends of the cdf F_X. In practice however, if no information on the source

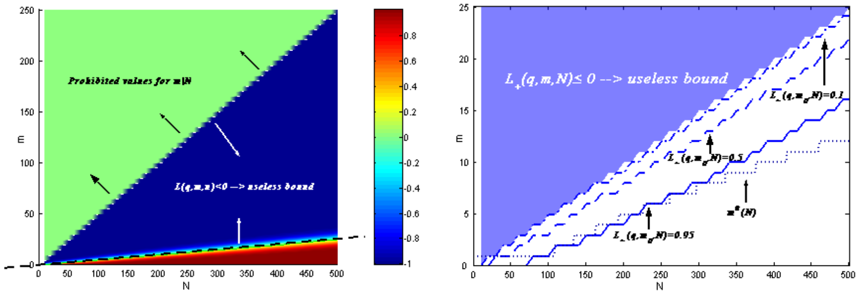

Fig. 2. Left: $\mathcal{L}(q,m,N)$, useful only for couples (m,N) below the dashed line; Right: selected iso-$\mathcal{L}_+$ curves for $q=0.95$. The curve $m^\sharp$ given by eq. (5) vs N has been also plotted.

densities is available, q can be a priori taken equal to e.g. 0.95. The value of m is thus fixed once the quantile number q and the probability threshold are fixed: we take m_0 as default value for m so that for a given quantile number q and $p=1-q$, the probability lower bound $\mathcal{L}_+(q,m_0,N)$ is higher than or equal to a positive threshold lower than 1.

The single parameter m has thus been replaced by two parameters, but the proposed approach has two advantages, though. First, the new parameters have a concrete interpretation; q is related to the support estimation, and the bound $\mathcal{L}_+$ tells us the confidence that we can have in the support estimation. Second, in practice, q and $\mathcal{L}_+$ can be fixed, so that a direct relation between m and N is found, which can be used to set a default value for m.

Figure 2(a) shows $\mathcal{L}(q,m,N)$ versus m and N. The valid values of m for a given N are $m \leq \lfloor N/2 \rfloor$. The bound is useful only for couples (m,N) below the dashed line illustrating $\mathcal{L}(q,m,N)=0$. In Figure 2(b) we plot the maximum value m_0 of m so that the quantity $\mathcal{L}_+(q,m_0,N)$ equals various fixed values (indicated on the related curve) with respect to N. Null values for m_0 indicate that it does not exist m_0 such that $\mathcal{L}_+(q,m,N) \geq 0.95$ for fixed N, $q=.95$ and all $m \leq m_0$. In other words, each couple (m,N) located under these curves ensure that inequality (4) holds. Observe that for sufficiently large N, small m and for a given q, $\mathcal{L}_+(q,m,N)$ tends to one.

It must be stressed that some attention must be paid when evaluating $\mathcal{L}_+$ for large N; numerical problems may arise when dividing two factorial expressions of large numbers. Therefore, we suggest to use the logarithms when computing the binomial coefficients, i.e. $\binom{N}{i} = \exp\left[\sum_{j=1}^{N}\log j - \sum_{j=1}^{N-i}\log j - \sum_{j=1}^{i}\log j\right]$. If one desires to speed up the method, the following empirical law can be used for selecting a default value for m; we can take

$$m^\sharp(N) = \max\left(1, \Re\left(\overline{\left[\left(\frac{N-18}{6.5}\right)^{0.65} - 4.5\right]}\right)\right) , \tag{5}$$

where $\overline{\alpha}$ denotes the nearest integer to α (see Fig. 2(b)).

4 Performances Comparison

In this section, we compare the extraction performances of 5 ICA algorithms: FastICA, JADE, AKMICA and three minimum-support approaches, OSICA (support estimated by $R_m(X)$), XSICA (support estimated by $R_1(X)$) and AVOSICA (support estimated by $\langle R_m(X) \rangle$). The default value for the parameter m was chosen equal to $m^\sharp$, given by (5). The algorithms have been tested on the extraction of 5 bounded and white sources from 5 mixtures. The pdf and cdf of the five sources (matched to (0,1)) are illustrated in Fig. 1. The mixing matrix is built from 25 random coefficients uniformly distributed in $(0, 1)$.

Figure 3 compares the histograms of the SIR for each extracted source in the noise-free case for $N = 2000$ and $m = m^\sharp(N)$. Remind that after having processed the permutation indetermination, the SIR criterion of the i-th source s_i reduces to $\mathrm{SIR}(s_i) = \sum_{j \neq i} |\mathbf{c}_i(j)|/|\mathbf{c}_i(i)|$. We can observe in Figure 3 that AVOSICA and XSICA give the most interesting results, in comparison to OSICA, AKMICA, JADE and FastICA (gauss), especially for the separation of sources with linear and triangular pdf. It must be stressed that even if AVOSICA and OSICA perform quite satisfactory for small values of N, the performances are improved for large N.

Table 1 summarizes the global SIR performance of ICA algorithms for various noise levels. Since we deal with SIR, the performance results are analyzed from the mixing matrix recovery point of view; the source denoising task is not

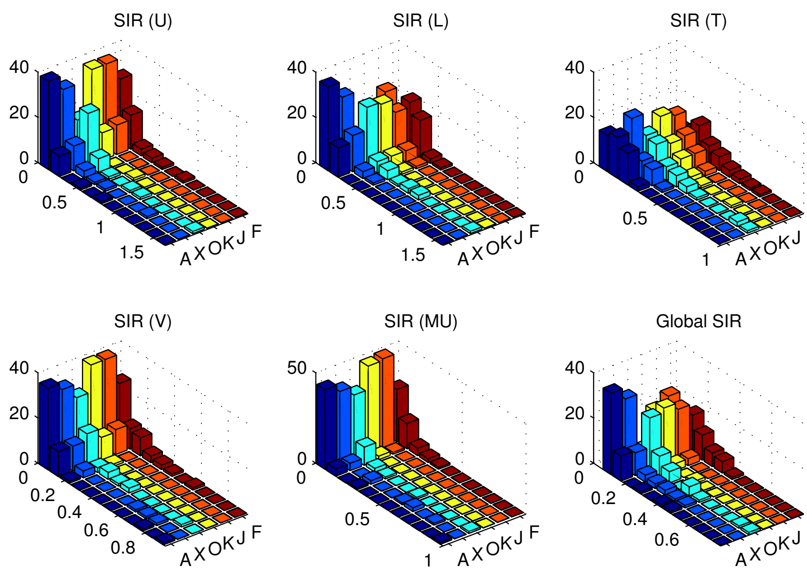

Fig. 3. 12-bins histograms of SIR for each extracted source, for 50 trials, $N = 2000$, and $m = m^\sharp(N) = 37$. The analyzed algorithms are AVOSICA ('A'), XSICA ('X'), OSICA ('O'), AKMICA ('K'), JADE ('J') and FastICA ('F'). The *global SIR* is the averaged SIR computed from the individual source SIRs for a given trial.

Table 1. 100-trials empirical means and variances of global SIR performances of several ICA algorithms (*global SIR* is the averaged SIR computed from the individual source SIRs for a given trial); $m = m^\sharp(N)$. Gaussian noise with standard deviation σ_n has been added to the whitened mixtures (so that for a given σ_n, the mixture SNRs equal $-10 \log \sigma_n^2$; they do not vary between trials, and do not depend of the mixing weights).

σ_n^2	N	AVOSICA	AKMICA	JADE	FastICA	XSICA	OSICA
0	500	**.106** (**.005**)	.135 (.006)	.127 (.006)	.194 (.023)	.139 (.025)	.208 (.03)
	2000	**.05** (.004)	.065 (.0012)	.06(**.0007**)	.097 (.004)	.082 (.02)	.087 (.003)
0.01	500	**.102** (**.003**)	.13 (.005)	.12 (.004)	.187 (.02)	.127 (.02)	.189 (.021)
	2000	**.047** (**.0006**)	.066 (.001)	.059 (.0008)	.105 (.005)	.08 (.02)	.085 (.0024)
0.05	500	**.105** (**.0027**)	.13 (.0039)	.122 (.0032)	.184 (.015)	.144 (.012)	.176 (.013)
	2000	**.051** (.0012)	.067 (.0012)	.06 (**.0006**)	.112 (.007)	.1 (.0225)	.09 (.006)

considered here. The global SIR, for a given trial, is obtained by computing the mean of the extracted sources SIR. The good results of AVOSICA can be observed, despite the fact that the value of m has not been chosen to optimize the results, i.e. we always have taken $m = m^\sharp(N)$. It must be stressed that the value of the parameter m is not critical when chosen around $m^\sharp(N)$.

JADE is a very good alternative when the dimensionality of the source space is low. The computational time of FastICA is its main advantage, contrarily to AKMICA.

5 Conclusion

In addition to existing results regarding the theoretical framework of minimum-support ICA and their specific advantages when separating sources correlated in a specific way, we have shown that these methods also yield competitive results in comparison to other ICA algorithms for extracting bounded sources in the noise-free and noisy cases. This is shown in the particular situation where the support measure is estimated using averaged quasi-ranges. We have further derived a rule to choose a default value for the tuning parameter m, for given sample size N. This choice is related to the confidence of support estimation quality. Numerical results illustrate that the proposed default value of m yield interesting SIR performances, that are comparable for m close to the suggested value.

References

1. Vrins, F., Jutten, C. & Verleysen, M. (2005) SWM : a Class of Convex Contrasts for Source Separation. In proc. ICASSP, *IEEE Int. Conf. on Acoustics, Speech and Sig. Process.*: V.161-V.164, Philadelphia (USA).
2. Vrins, F. & Verleysen, M. (2006) Minimum Support ICA Using Order Statistics. Part I: Quasi-Range Based Support Estimation. In proc. ICA, *Int. Conf. on Ind. Comp. Anal. and Blind Sig. Sep.*, Charleston SC (USA).

3. Vrins, F., Lee, J.A. & Verleysen, M. (2005) Filtering-free Blind Separation of Correlated Images. In proc. IWANN, *Int. Work-Conf. on Art. Neur. Net.*: 1091-1099, Barcelona (Spain).
4. Pham, D.T. & Vrins, F. (2005) Local Minima of Information-Theoretic Criteria in Blind Source Separation. *IEEE Sig. Process. Lett.*, **12**(11): 788-791.
5. Pham, D.-T. (2000) Blind Separation of Instantaneous Mixture of Sources Based on Order Statistics. *IEEE Trans. Sig. Process.*, **48**(2): 363-375.
6. Lee, J.A., Vrins, F. & Verleysen, M. (2005) A Simple ICA Algorithm for Non-Differentiable Contrasts. In proc. EUSIPCO, *Eur. Sig. Process. Conf.*: cr.1412.1-4, Antalya (Turkey).
7. Theis, F.J. & Gruber, P. (2004) Separability of Analytic Postnonlinear Blind Source Separation with Bounded Sources. In proc. ESANN, *Eur. Symp. Art. Neur. Net.*: 217-222, Bruges (Belgium).
8. Theis, F.J., Georgiev P. & Cichocki, A. (2004) Robust Overcomplete Matrix Recovery for Sparse Sources Using a Genralized Hough Transform. In proc. ESANN, *Eur. Symp. Art. Neur. Net.*: 343-348, Bruges (Belgium).
9. Plumbley, M.D. (2003) Algorithms for Nonnegative Independent Component Analysis. *IEEE Trans. Neur. Net.* **4**(3): 534-543.
10. Hyvarinen, A., Karhunen, J. & Oja, E. (2001). "Independent Component Analysis." Wiley series on adaptive and learning systems for signal processing, Wiley, New York.
11. Hyvarinen, A. & Oja, E. (1997) A Fast Fixed-point Algorithm Independent Component Analysis. *Neur. Comp.* **9**(7): 1483–1492.
12. Cardoso, J.F. & Souloumiac, A. (1993) Blind Beamforming for Non-Gaussian Signals. *IEE Proc.-F* **140**(6): 362–370.
13. Molgedey, J. & Schuster, H.G. (1994) Separation of a Mixture of Independent Signals Using Time Delayed Correlations. *Phys. Rev. Lett.* **72**: 3634–3636.
14. Cruces, S. & Duran, I. (2004) The Minimum Support Criterion for Blind Source Extraction: a Limiting Case of the Strengthened Young's Inequality. In proc. ICA, *Int. Conf. on Ind. Comp. Anal. and Blind Sig. Sep.*: 57–64, Granada (Spain).
15. Hoel, P.G. (1975) "Introduction to Mathematical Statistics." Wiley series in probability and mathematical statistics, Wiley, New York.
16. Chu, J.T. (1957). "Some Uses of Quasi-Ranges." Ann. Math. Statist. **28**: 173–180.
17. David, H.A. (1970). "Order Statistics." Wiley series in probability and mathematical statistics, Wiley, New York.