NOVELTY DETECTION WITH MULTIVARIATE EXTREME VALUE THEORY, PART II: AN ANALYTICAL APPROACH TO UNIMODAL ESTIMATION

Samuel Hugueny[*], *David A. Clifton*[†] *and Lionel Tarassenko*

Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Roosevelt Drive, Oxford, OX3 7DQ, UK
samuel.hugueny@eng.ox.ac.uk, {davidc, lionel}@robots.ox.ac.uk

ABSTRACT

Extreme Value Theory (EVT) describes the distribution of data considered extreme with respect to some generative distribution, effectively modelling the tails of that distribution. In novelty detection, we wish to determine if data are "normal" with respect to some model of normality. If that model consists of generative distributions, then EVT is appropriate for describing the behaviour of extrema generated from the model, and can be used to separate "normal" areas from "abnormal" areas of feature space in a principled manner. In a companion paper, we show that existing work in the use of EVT for novelty detection does not accurately describe the extrema of multimodal, multivariate distributions and propose a numerical method for overcoming such problems. In this paper, we introduce an analytical approach to obtain closed-form solutions for the extreme value distributions of multivariate Gaussian distributions and present an application to vital-sign monitoring.

1. INTRODUCTION

Extreme Value Theory (EVT) is a branch of statistics that is concerned with extreme deviations from the median of some data-generating distribution; i.e., abnormally high or low values in the tails of these distributions. Classical EVT, for which the theory is well-established [1, 2], is mainly concerned with the statistics of the largest (or smallest values) of univariate distributions. While this is useful in fields such as hydrology, insurance, and finance, it is inadequate for multivariate machine learning problems, e.g outlier or novelty detection. In the latter, we are interested not only in data of abnormal magnitude but in all areas of data space of abnormally low probability, including those that may be located between the modes of the distributions. In applications such as structural health monitoring [3, 4] or vital-sign monitoring, data in these areas can indicate a potential malfunction of the system or deterioration of patient condition.

[*]Supported by the EPSRC LSI Doctoral Training Centre, Oxford
[†]Supported by the NIHR Biomedical Research Centre, Oxford

[5, 6] proposed an extension of classical EVT to mixtures of multivariate Gaussians, with three applications in biomedical engineering. This work recommends, for a given sample $\mathbf{x}$, that only the extreme value distribution (EVD) associated with the kernel closest to $\mathbf{x}$ (in the Mahalanobis sense) is considered, which may be calculated using the known EVD of the single-sided univariate Gaussian distribution. In a companion paper [7] we argued the need for a multivariate Extreme Value Theory (mEVT) in machine learning, identified some of the limitations of the existing approach, and proposed a numerical scheme for multimodal multivariate estimation.

In this paper, we offer an analytical approach to mEVT, restricted to single Gaussian distributions; i.e., for dimensionality $n \in \mathbb{N}^*$, distributions F_n with probability density functions of the form:

$$f_n(\mathbf{x}) = \frac{1}{C_n} \exp\left(-\frac{M(\mathbf{x})^2}{2}\right) \tag{1}$$

where $M(\mathbf{x}) = \left((\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)^{1/2}$ is the Mahalanobis distance, $C_n = (2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}$ the normalisation coefficient, $\boldsymbol{\mu}$ the centre, and $\boldsymbol{\Sigma}$ the covariance matrix. We term $\mathcal{D} = \mathbb{R}^n$ the data space, and $\mathcal{P} = f_n(\mathcal{D}) = \left]0, \frac{1}{C_n}\right]$, the associated probability space. A unimodal approach is of interest because the existing method [5] yields significant errors when estimating EVD parameters, which can be solved directly by the use of analytically-derived estimates. Furthermore, the numerical approach presented in the companion paper [7] requires that we generate extrema from a multivariate distribution, which can be time-consuming as the sample size is increased. While a fully analytical mEVT may not be possible, the analytical study presented here paves the way to more elaborate numerical schemes with no need for sampling extrema.

978-1-4244-4948-4/09/$25.00 © 2009 IEEE

2. EXTREME VALUE THEORY

2.1. Classical univariate theory

Consider $\{X_m\}$, a set of m independent and identically distributed random variables (iid rvs), where $X_i \in \mathbb{R}$ is drawn from a distribution F, and $M_m = \max(X_1, X_2, \ldots, X_m)$. The basis of EVT is the Fisher-Tippett theorem [8]:

Theorem 1. *(Fisher-Tippett theorem)*
Let $\{X_m\}$ be a sequence of iid rvs and M_m the maximum of the sequence $\{X_m\}$. If there exist norming constants $d_m \in \mathbb{R}$, $c_m > 0$ and some non-degenerate distribution function (df) H such that

$$c_m^{-1}(M_m - d_m) \xrightarrow{d} H, \tag{2}$$

then H belongs to the type of one of the following three distribution functions:

$$\textit{Gumbel:} \quad H_1^+(x) = \exp\left(-e^{-x}\right),\ x \in \mathbb{R}.$$

$$\textit{Fréchet:} \quad H_2^+(x,\alpha) = \begin{cases} 0, & x \le 0 \\ \exp\left(-x^{-\alpha}\right), & x > 0 \end{cases},$$

$$\textit{Weibull:} \quad H_3^+(x,\alpha) = \begin{cases} \exp\left(-(-x)^{\alpha}\right), & x \le 0 \\ 1, & x > 0 \end{cases},$$

for $\alpha > 0$.

In other words, if the distribution over M_n is to be stable in the limit $m \to \infty$, then it can only converge, under positive affine transformation, to three types of distributions. d_m is the location parameter and c_m is the scale parameter. For the distributions in which it appears, α is the shape parameter and is not explicitly made dependent on m in the theorem. We show in section 4 that α can be made dependent on m for practical purposes. The superscript '+' refers to the fact that these distributions are distributions of maxima. The theorem holds for the distribution of minima, as minima of $\{X_m\}$ are maxima of $\{-X_m\}$. EVDs of minima are therefore the same as EVDs of maxima, with a reverse axis. In section 4, we are interested in the Weibull-type EVD of minima, or minimal Weibull, for which the attractor is (see [2]):

$$H_3^-(x,\alpha) = \begin{cases} 0, & x < 0 \\ 1 - \exp\left(-(x)^{\alpha}\right), & x \ge 0 \end{cases}. \tag{3}$$

2.2. Multivariate EVT: redefining extrema

Classical univariate EVT (uEVT) cannot be directly applied to the estimation of multivariate EVDs. For the multivariate case, we no longer wish to answer the question "how is the sample of greatest magnitude distributed?", which would require choosing an ad-hoc multivariate distance, but rather "how is the most improbable sample distributed?". To this end, we use an alternate definition of extremum:

Definition 1. *Let $m \in \mathbb{N}^*$ and $\{X_m\}$ be a sequence of (possibly multivariate) iid rvs, drawn from a distribution F with probability density function f. We define the extremum to be the random variable $E_m = \textit{argmin}\{f(X_1), \ldots, f(X_m)\}$.*

This different perspective on EVT can be related to classical EVT in two ways:

- away from the distribution modes, the pdf monotonically decreases with increasing distance to the modes. Extrema in magnitude are therefore also minima in probability density values,
- selecting the most improbable sample with respect to f is equivalent to selecting the sample of minimal magnitude with respect to the df over $f(X)$. uEVT can therefore be applied to samples drawn in the probability space.

However, using this definition is better-suited to problems where all improbable events are of interest, including those that could occur between modes of f. The aim of the next two sections is to find the form of the df over E_m.

3. PROBABILITY DISTRIBUTION OF PROBABILITY DENSITY VALUES

3.1. Sampling in the data space is equivalent to sampling in the image probability space

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ be samples (vectors of $\mathcal{D}$) drawn from the distribution F_n for $k \in \mathbb{N}$ and $f_n(\mathbf{x}_1)$, $f_n(\mathbf{x}_2), \ldots$, $f_n(\mathbf{x}_k)$ be the corresponding pdf values of these samples. The probability of obtaining a given value in the probability space by drawing a sample in the data space is strongly related to the form of f. Assuming that X is a random variable distributed according to F_n, our aim in this section is to determine the form of the df G_n according to which $f_n(X)$ is distributed on $\mathcal{P}$.

3.2. Distribution function over $f_n(X)$

We define a distribution over $Y = f_n(X)$ as follows:

$$\forall y \in \mathcal{P}, \quad G_n(y) = \int_{f^{-1}(]0,y])} f_n(\mathbf{x}) d\mathbf{x} \tag{4}$$

where $f^{-1}(]0, y])$ is the preimage of $]0, y]$ under f_n.

To take advantage of the ellipsoidal symmetry of the problem we rewrite f in a Mahalanobis n-dimensional spherical polar coordinate system. Then, $x_v = (r, \boldsymbol{\theta})$ such that $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{n-1})$, $r = M(\mathbf{x})$, $\theta_i \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right[$ for $i \le n-2$ and the base angle θ_{n-1} ranges over $[0, 2\pi]$. The Jacobian of the transformation is $|J| = |\boldsymbol{\Sigma}|^{1/2} r^{n-1} \prod_{i=0}^{n-3} (\cos\theta_i)^{n-i}$ (see, for instance, [9]).

$$G_n(y) = \int_{f_n^{-1}(]0,y])} \frac{1}{C_n} \exp\left(-\frac{M(\mathbf{x})^2}{2}\right) d\mathbf{x}, \quad (5)$$

$$= \int_{f_n^{-1}(]0,y])} \frac{|J|}{C_n} \exp\left(-\frac{r^2}{2}\right) \mathrm{d}r\mathrm{d}\boldsymbol{\theta}, \quad (6)$$

$$= \Omega \int_{M^{-1}(y)}^{+\infty} \frac{r^{n-1}}{(2\pi)^{n/2}} \exp\left(-\frac{r^2}{2}\right) \mathrm{d}r, \quad (7)$$

$$= \Omega|\boldsymbol{\Sigma}|^{1/2} \int_0^y \left[-2\ln\left(C_n u\right)\right]^{(n-2)/2} \mathrm{d}u. \quad (8)$$

$M^{-1}(y) = \sqrt{-2\ln\left(C_n y\right)}$ is the unique Mahalanobis distance associated with the pdf value y. Eq. 6 is obtained by rewriting eq. 5 in the spherical polar coordinate system. Integrating out the angles yields eq. 7, where $\Omega = \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}$ is the total solid angle subtended by the unit n-sphere. Eq. 8 is obtained after making the substitution $u = \frac{1}{C_n}\exp\left(-\frac{r^2}{2}\right)$. Note that eq.7 and 8 hold for $n = 1$.

The integrand in 8 is the pdf of G_n, which we aim to find,

$$g_n(y) = \Omega_n|\boldsymbol{\Sigma}|^{1/2}\left[-2\ln\left(C_n y\right)\right]^{n-2/2} \quad (9)$$

If $n = 1$, the integration of eq.7 yields $G_1(y) = \mathrm{erfc}\left(\sqrt{-\ln\left(C_1 y\right)}\right)$ where $\mathrm{erfc}\left(.\right)$ is the complementary error function. If $n \geq 2$, the integration of 8 is possible using a recursive integration by parts, which yields two cases:

$$G_{2p}(y) = y\sum_{k=0}^{p-1} A_{2p}^k \left(-2\ln\left(C_{2p} y\right)\right)^{(p-k-1)} \quad (10)$$

$$G_{2p+1}(y) = y\sum_{k=0}^{p-1} A_{2p+1}^k \left(-2\ln\left(C_{2p+1} y\right)\right)^{p-k-1/2} + \mathrm{erfc}\left(\sqrt{-\ln\left(C_{2p+1} y\right)}\right) \quad (11)$$

for all $p \in \mathbb{N}^*$, where $A_{2p}^k = \Omega_{2p}|\boldsymbol{\Sigma}|^{1/2}\frac{2^k(p-1)!}{(p-1-k)!}$ and $A_{2p+1}^k = \Omega_{2p+1}|\boldsymbol{\Sigma}|^{1/2}\frac{(2p-1)!(p-k)!}{2^{k-1}(p-1)!(2p-2k)!}$.

G_n and g_n are plotted for $n = 1$ to 5 in figure 1, together with simulated data. Perhaps counter-intuitively, we observe that, relative to the right endpoint of G_n, the probability mass shifts towards 0 as the dimensionality n increases, which indicates that the probability mass in the data space moves away from the centre of the distribution, as is noted by Bishop in [10] (ex.1.4, p.29).

4. EVD FOR PROBABILITY DENSITY VALUES

Our approach is based on the idea that pdf values of the EVD in the data space must be equal on a level set (or contour) of F; i.e., the EVD is obtained by applying a weighting function to the level sets of F. Consequently, determining the EVD in the data space can be performed by determining the EVD of G in the probability space. We therefore reduce an n-dimensional problem (finding an EVD in $\mathcal{D}$) to a simpler one-dimensional case (finding an EVD in $\mathcal{P}$). In this section, our aim is to determine the Extreme Value Distribution of minima for G_n and estimate its parameters, using uEVT.

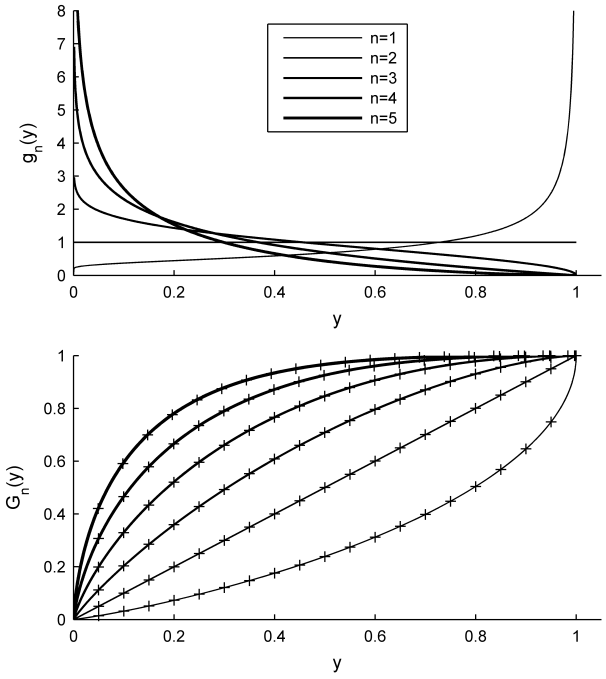

Fig. 1. Analytical and simulated G_n and g_n for various values of n. The x-axis is scaled so that all distributions have the same right endpoint. Crosses are the result of drawing 10^6 samples in the data space and computing the cumulative histograms of their probabilities. F_n is the multivariate standard normal distribution.

4.1. Maximum domain of attraction of the Weibull distribution

The Fisher-Tippett theorem effectively defines a three-class (Gumbel, Fréchet, and Weibull) equivalence relation on the set of non-degenerate univariate distributions. [1] gives the characterizations for each class. Theorem 3.3.12 in [1] characterizes the maximum domain of attraction (MDA) of the maximal Weibull distribution. We adapt it to the MDA of the minimal Weibull distribution:

Theorem 2. *(Maximum domain of attraction of H_3^-) The df F belongs to the maximum domain of attraction of the minimal Weibull distribution ($\alpha > 0$), if and only if*

$x_F > -\infty$ and $F(x_F + x^{-1}) = x^{-\alpha}L(x)$ for some slowly varying function L. If $F \in$ MDA(H_3^-), then $c_m^{-1}(E_m - x_F) \xrightarrow{d} H_1^+$, where the norming constants c_m, d_m can be chosen to be $c_m = x_F + F^{\leftarrow}(m^{-1})$ and $d_m = x_F$.

x_F is the left endpoint of the df F, $F^{\leftarrow}(p)$ is the p-quantile of F, and L a slowly varying function at ∞; i.e., a positive function that obeys

$$\forall t > 0, \lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1. \tag{12}$$

From eq.12, it may be seen that $y \mapsto -\ln(1/y), y > 1$ is slowly varying, as is $y \mapsto -\ln(1/y)^{\beta}, y > 1$, for all $\beta \in \mathbb{R}$. Therefore $y\,G_{2p}(1/y)$ is a sum of slowly varying functions, which is itself slowly varying. Theorem 2 can therefore be applied to G_{2p}. A similar process can be followed to show that G_{2p+1} is in the MDA of H_3^-. Consequently, G_n is in the MDA of H_3^- for all values of n.

4.2. Parameter estimation

If G_n is in the MDA of H_3^-, theorem 2 gives the minimal Weibull parameters:

$$d_m = 0, \quad \alpha_m = 1, \tag{13}$$

$$c_m = G_n^{\leftarrow}\left(\tfrac{1}{m}\right). \tag{14}$$

The scale parameter can be easily estimated numerically from eq.14 to arbitrary accuracy, as G_n is a strictly increasing function over finite support. Figure 2 shows that eq.14 is a very close approximation to the values of the scale parameter d_m obtained via maximum likelihood estimation (MLE). However, the value of the shape parameter α_m, although theoretically guaranteed to converge to 1 in the limit $m \to \infty$ seems to decrease significantly as the dimensionality of the data space increases, and it is overestimated even for large values of m. To address this issue, we note that the class of equivalence of H_3^- contains all the distributions with a power law behaviour at the finite left endpoint [1]. Therefore the tail of G_n is, in the limit $y \to 0$, equivalent to a power law; i.e., $G_n(y) \sim Ky^s$. Here, s can be estimated locally by noting that, in this case, $g_n(y) \sim sKy^{s-1}$; i.e., $s = y\frac{G_n(y)}{g_n(y)}$. We therefore propose the following formula for the shape parameter

$$\alpha_m = c_m \frac{g_n(c_m)}{G_n(c_m)} \tag{15}$$

Figure 2 shows that eq.15, although still inaccurate for very small values of m, gives values closer to the MLE estimates as m and n increase.

Finally, the EVD of G_n is:

$$G_n^e(y) = 1 - \exp\left(-(y/c_m)^{\alpha_m}\right), \tag{16}$$

where c_m and α_m are given by eq.14 and 15, respectively.

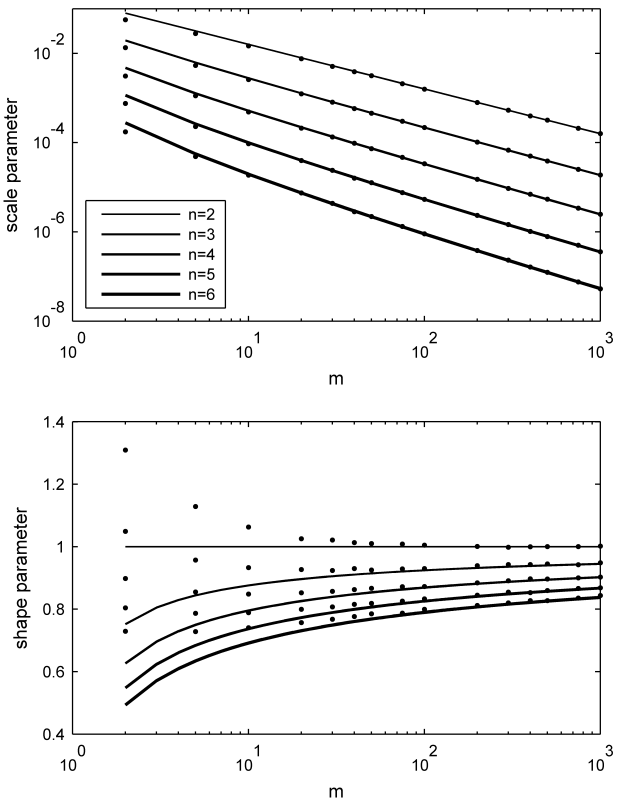

Fig. 2. Comparison between results of maximum likelihood estimation of the scale parameter c_m (top) and the shape parameter α_m (bottom) parameter, and values obtained using formulae 15 and 14 for increasing m. F_n is again the multivariate standard normal distribution. The dots are obtained by taking the means of 10 MLEs, each with 10^4 simulated extrema. Error bars are too small to be visible at this scale.

4.3. Novelty Scores

In novelty detection, extrema are regarded as potentially abnormal data. Assuming a distribution for the normal data, if we observe m samples for which the extremum has pdf value y_m, the probability of drawing an extremum of lower probability is given by $G_n^e(y_m)$. Therefore, the probability of drawing an extremum of higher probability is $1-G_n^e(y_m)$ and our extremum is abnormal with probability $1-G_n^e(y_m)$. Therefore we define a novelty score in the data space as being the probability of obtaining an extemum closer to the centre of the distribution (in the Mahalanobis sense):

$$F_n^e(\mathbf{x}) = 1 - G_n^e(f_n(\mathbf{x})), \tag{17}$$

$$= \exp\left(-\left(\frac{1}{C_n c_m}\mathrm{e}^{-\frac{M(\mathbf{x})^2}{2}}\right)^{\alpha_m}\right). \tag{18}$$

In the limit $m \to \infty$, $F_n^e(\mathbf{x})$ can be interpreted as a Mahalanobis-radial cdf of extrema.

5. APPLICATION TO A VITAL-SIGN MONITORING PROBLEM

In this section, we present an application of mEVT to a vital-sign monitoring problem. Continuous real-time patient monitoring in hospital is usually based on single vital-sign channel alarms, which yield an unusably large number of false alarms. Recently, efforts have been made to take advantage of the correlation between vital-sign channels. In [11, 12], the authors adopt a novelty detection approach, whereby the model of normality is a 4-dimensional pdf constructed using data from a high-risk adult population. The authors then test for abnormal data by comparing the unconditional probabilities $f(\mathbf{x})$ of new measurements obtained in real-time to a predefined threshold.

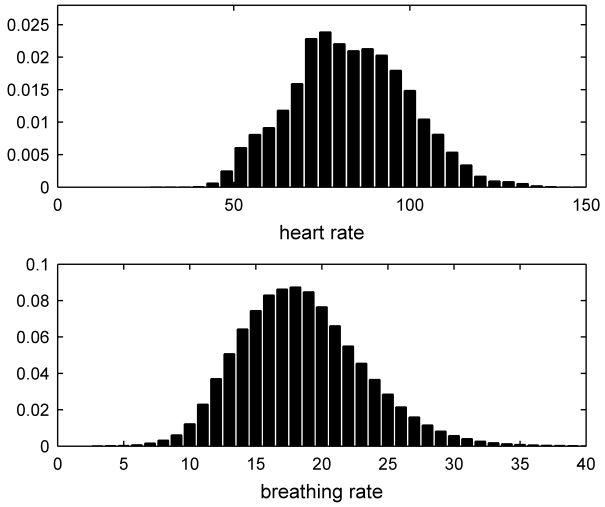

Fig. 3. Normalised histograms of the heart rate and breathing rate values for all patients in the training group. The means and standard deviations are, 84.41 and 18.45 bpm for the heart rate, and 16.39 and 4.60 rpm for the breathing rate. The covariance is 14.75.

Here, we introduce the use of mEVT to address the same problem, limited to the use of two vital-sign channels, heart rate (HR) and breathing rate (BR). The data set was collected during the first phase of a trial conducted at the University of Pittsburgh Medical Center [11, 13, 14], which is composed of the recordings of 332 high-risk adult patients, totalling over 18,000 hours of data. Vital-sign measurements are available every second for all patients. Recordings are labeled with "crisis events"; i.e., events that should have caused an emergency call to clinical staff to be made on the patient's behalf, as they are indicative of potentially adverse events. The cause of each event (high/low HR, high/low BR, etc.) is given with its start-time and end-time. 46 of 113 events are caused by an abnormally high or low heart rate or breathing rate (approximately 19 hours of data). We split patients into three groups: a *test* group, composed of the 28 patients who suffered at least one cardio-respiratory crisis over the course of their stay (approximately 1000 hours of data), a *training* group, and a *control* group, these latter two each composed of 154 randomly-assigned patients who did not suffer a cardio-respiratory crisis (approximately 8500 hours of data each). Figure 3 shows histograms of the training data. A bivariate Gaussian distribution F_2 is fitted to the data of the control group. To ease their graphical interpretation, we define novelty scores to be:

$$Z_1(\mathbf{x}_M) = -\ln\left(1 - F_2^e(\mathbf{x}_m)\right). \tag{19}$$

where $\mathbf{x}$ is the extremum of m samples and F_2^e is defined by eq.18. Note that $Z_1(\mathbf{x})$ takes low values if $\mathbf{x}$ is close to the centre of the distribution and increases as $\mathbf{x}$ becomes more and more "abnormal". Novelty scores are subsequently assigned to the entire recordings of the patients in all groups. At time t, the highest novelty score of the last m HR and BR measurements is returned. The value of the parameter m is empirically chosen to be 30 for illustration.

We compare the mEVT-based approach with the classical thresholding approach; i.e., setting a threshold in the probability space to which each measurement is individually compared. To make the comparison with the mEVT-based method easier, we deduce from eq.6 that $\Pr(f_2(\mathbf{x}) \geq f_2(\mathbf{x}_0)) = 1 - \exp\left(-\frac{M(\mathbf{x}_0)^2}{2}\right)$ and therefore define the novelty score for the classical thresholding approach to be:

$$Z_2(\mathbf{x}_0) = -\ln\left(1 - f_2(\mathbf{x}_0)\right) = \frac{M(\mathbf{x}_0)^2}{2}. \tag{20}$$

For a sample $\mathbf{x}$, Z_2 answers the question: "What was the probability of drawing a sample of smaller magnitude?". Z_1 answers the question: "Considering $\mathbf{x}$ and the m-1 samples observed before it, what was the probability of drawing m samples with a more probable extremum?" For both approaches, we define $\tau_{training}(q)$, $\tau_{control}(q)$, $\tau_{test}(q)$ and $\tau_{crisis}(q)$ to be the fraction of the total recording time that novelty scores are higher than the threshold q for the training group, the control group, the test group in the absence of a crisis, and the test group during crises, respectively.

Figure 4 shows the evolution of these fractions as q is increased for the mEVT-based method and the thresholding method. The heterogeneity of the measurements in the crisis windows means that we cannot expect a $\tau_{crisis} = 100\%$. However, it is important to detect as much of the crisis data as possible to avoid false negatives (where the novelty score is below the threshold during a crisis). If we

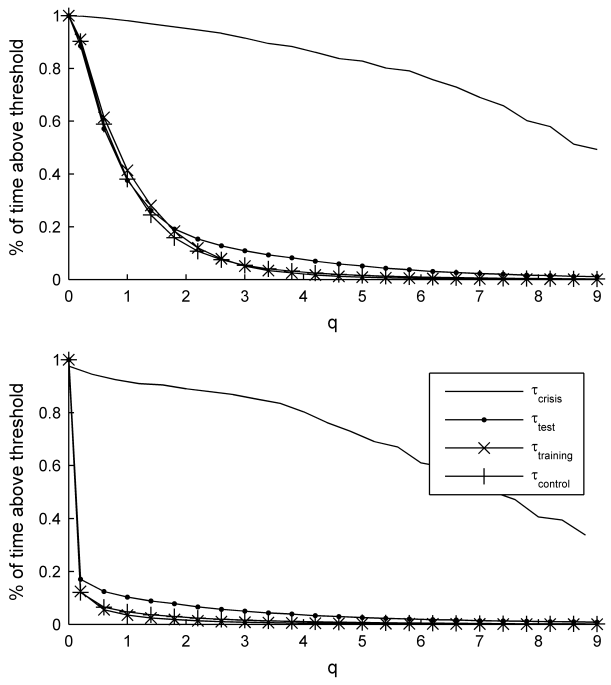

Fig. 4. $\tau_{training}$, $\tau_{control}$, τ_{test} and τ_{crisis} for the thresholding method (upper plot) and the mEVT-based method (lower plot). A warning system should aim to trigger an alarm only during crises; i.e., we must set the threshold so that for an acceptable value of τ_{crisis}, $\tau_{control}$ is minimal.

allow τ_{crisis} to be 90%, then $\tau_{training}$, $\tau_{control}$, τ_{test} equal 1.50%, 2.59%, 7.03% for the mEVT method, respectively, and 3.39%, 4.01%, and 9.34% for the thresholding method. This is a 35.4% reduction of false-positive alert time (where the novelty score is above the threshold in the absence of a crisis) for the control group. This reduction becomes 42.1%, and 29.2%, for τ_{crisis} =85% and 80%, respectively. Therefore, the mEVT-based method consistently brings a significant improvement to the false-positive time while yielding the same true-positive detection rate.

6. DISCUSSION

Novelty detection can benefit from a comprehensive multivariate Extreme Value Theory. Here, by giving an alternate definition of extrema and closed-form solutions for the distribution function over pdf values, we show that we can obtain accurate estimates of the EVDs of multivariate Gaussian kernels. We applied our formulae to actual patient vital-sign data and showed that the use of EVD is a significant improvement over the conventional thresholding method. Obtaining these formulae relies on our ability to proceed from eq.5 to eq.8; i.e., our ability to parameterise the level sets of the generative probability distribution and integrate the resulting parameterization. While this is relatively easy for multivariate Gaussian distributions, it is not possible for arbitrarily complex, non-symmetrical distributions and fully-analytical closed-form extreme value distributions are not to be expected. However, depending on our ability to estimate the distribution function over the pdf values, accurate estimates of its minimal EVD can be obtained without any sampling of extrema, which would be a great improvement over the existing numerical approach presented in the companion paper [7].

7. REFERENCES

[1] P. Embrechts, C Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Springer, 1997.

[2] E. Castillo, Ali S. Hadi, N. Balakrishnan, and J.M Sarabia, *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley Interscience, 2005.

[3] K. Worden, G. Manson, and D. Allman, "Experimental validation of a structural health monitoring methodology: Part I. novelty detection on a laboratory structure," *Journal of Sound and Vision*, vol. 259, no. 2, pp. 323–343, 2003.

[4] H. Sohn, D.W. Allen, K. Worden, and C.R. Farrar, "Structural damage classification using extreme value statistics," *Journal of Dynamic Systems, Measurement, and Control*, vol. 127, no. 1, pp. 125–132, 2005.

[5] S.J. Roberts, "Novelty detection using extreme value statistics," *IEE. Proc.-Vis. Image Signal Process*, vol. 146, no. 3, June 1999.

[6] S.J. Roberts, "Extreme value statistics for novelty detection in biomedical data processing," *IEE Proc.-Sci. Meas. Technol.*, vol. 147, no. 6, pp. 363–367, Nov. 2000.

[7] D. Clifton, S. Hugueny, and L. Tarassenko, "Novelty detection with multivariate extreme value theory, part I: A numerical approach to multimodal estimation," *Proceedings of IEEE Machine Learning in Signal Processing, In press.*, Sept. 2009.

[8] R.A Fisher and L.H.C Tippett, "Limiting forms of the frequency distributions of the largest or smallest members of a sample," *Proc. Camb. Philos. Soc.*, , no. 24, pp. 180–190, 1928.

[9] D.W. Scott, *Multivariate density estimation: theory, practice and visualization*, Wiley-Interscience, 1992.

[10] C. M. Bishop, *Neural networks for Pattern Recognition*, Oxford University Press, 1995.

[11] A. Hann, *Multi-parameter Monitoring for Early Warning of Patient Detoriation*, Ph.D. thesis, University of Oxford, 2008.

[12] L. Tarassenko, A. Hann, and D. Young, "Integrated monitoring and analysis for early warning of patient deterioration," *British Journal of Anaesthesia*, vol. 97, no. 1, pp. 64–68, 2006.

[13] M. Hravnak, L. Edwards, A. Clontz, C. Valenta, M. A. Devita, and M. R. Pinsky, "Defining the incidence of cardiorespiratory instability in patients in step-down units using an electronic integrated monitoring system.," *Archives of internal medicine*, vol. 168, no. 12, pp. 1300–1308, June 2008.

[14] M. Hravnak, L. Edwards, A. Clontz, C. Valenta, M.A. DeVita, and M.R. Pinsky, "Impact of electronic integrated monitoring system upon the incidence and duration of patient instability on a step down unit," *Proceedings of 4th International Medical Emergency Team Conference, Toronto*, 2008.