# Stochastic Measure of Informativity and its Application to the Task of Stable Extraction of Features 

## Alexander Lepskiy

National Research University - Higher School of Economics, Moscow, Russia

6th International Conference on Soft Methods in Probability and Statistics, October 4-6, 2012, University of Konstanz, Konstanz, Germany

## Problem of Features Extraction in Pattern Recognition



ת
$X=\left\{x_{1}, \ldots, x_{n}\right\}$ - initial discrete contour

$\Omega=\left\{\omega_{1}, \ldots, \omega_{n}\right\}$ - set of features (curvatures, $\Rightarrow \mu(A)$ - aggregating features for $A$ $\varpi_{k}=\varpi\left(x_{k}, A\right) \quad$ variance of length)

$A=\left\{x_{i}, \ldots, x_{i_{m}}\right\}$ - subcontour (measure of informativeness)

## Problem

Find a representation $A$ which satisfies certain conditions of optimality

## Monotone Geometrical Measure of Informativeness

Let $X=\left\{x_{1}, \ldots, x_{n}\right\}$ be an initial discrete closed contour, where $x_{i} \in R^{2}$, $i=1, \ldots, n$.
A geometrical measure of informativeness $\mu: 2^{X} \rightarrow[0,1]$ is a set function that has to obey the following properties:
(1) $\mu(\emptyset)=0, \mu(X)=1$;
(2) $A, B \in 2^{X}$ and $A \subseteq B$ implies $\mu(A) \leq \mu(B)$;
(3) let $B=\left\{x_{i_{1}}, \ldots, x_{i_{k-1}}, x_{i_{k}}, x_{i_{k+1}}, \ldots, x_{i_{m}}\right\} \in X$ and neighbouring points $x_{i_{k-1}}, x_{i_{k}}, x_{i_{k+1}}$ belong to a straight line in the plane, then $\mu(B)=\mu\left(B \backslash\left\{x_{i_{n}}\right\}\right) ;$

- $\mu$ is invariant w.r.t. affine transformations in the plane such as parallel transferring, rotation and scaling.


## Examples of Monotone Geometrical Measure of Informativeness

- Let $L(A)$ be a length of subcontour $A \in 2^{X}$. Then $\mu_{L}(A)=\frac{L(A)}{L(X)}$ is a measure of informativeness by length.
- Suppose that the domain limited by $X$ is convex, and a function $S(B)$ determines the area bounded by an subcontour $A \in 2^{X}$. Then an measure of informativeness by area is $\mu_{S}(A)=\frac{S(A)}{S(X)}$.
- Let $w(x, A)$ be an estimate of information value of the part of a contour in a neighbourhood of point $x \in A \subseteq X$. Then an average measure of informativeness is defined by $\mu(A)=\frac{\sum_{x \in A} w(x, A)}{\sum_{x \in X} w(x, X)}$, $\mu(\emptyset)=0$ by definition.

The geometrical measure of informativeness $\mu_{L}$ and $\mu_{S}$ can be considered as average measure of informativeness:

- for $\mu_{L}$ function $w(x, A)=\left|x-x_{+}\right|$, where $x_{+}$is the next neighbouring point in contour $A$;
- $\mu_{S}$ function $w(x, A)=S\left(O, x, x_{+}\right)$, where $O$ is the centroid of area, bounded by contour $A$, and $S\left(O, x, x_{+}\right)$is the area of triangle with vertices in points $O, x, x_{+}$.



## Stochastic Average Monotone Measure of Informativeness

Let the values $w(x, A)$ are corrupted by a probabilistic noise. We get random values $W(x, A)$ and the stochastic measure of informativeness $\mathrm{M}(A)=\frac{\sum_{x \in A} W(x, A)}{\sum_{x \in X} W(x, A)}$.

## Problem

Find the most stable and informative representation $A \in 2^{X}$ of the pattern $X$ for which the expectation $\mathbf{E}[\mathrm{M}(A)]$ will be maximize and the variance $\sigma^{2}[\mathrm{M}(A)]$ will be minimize.

If $W(x, A)=W(x)$ and random values $W(x), x \in X$, are independent random variables then the $\mathrm{M}(A)$ is additive.
Let value $W(x, A)$ depends on only two neighbouring points. For example, the geometrical measures of informativeness $\mu_{L}$ and $\mu_{S}$ are satisfied this condition.

## Some Notations and Assumptions

- $X=\left\{x_{1}, \ldots, x_{n}\right\}$ be an original contour, $A=\left\{x_{i_{1}}, \ldots, x_{i_{m}}\right\} \in 2^{X}$;
- $x_{k}(A)=x_{i_{k}}$ if $k \in\{1, \ldots, m\}$;
- we suppose that $W\left(x_{k}(A), A\right)=W\left(x_{k}(A), x_{k+1}(A)\right), k=1, \ldots,|A|$, i.e. the value $W\left(x_{k}(A), A\right)=W\left(x_{k}(A), x_{k+1}(A)\right)=W_{k, k+1}(A)$ depends on two neighbouring points $x_{k}(A), x_{k+1}(A)$;
- random variables $W_{k, k+1}(A), W_{l, l+1}(A)$ are independent if $|l-k|>1$.

Then an average monotone measure and stochastic average monotone measure have a view

$$
\begin{equation*}
\mu(A)=\frac{\sum_{k=1}^{|A|} w_{k, k+1}(A)}{\sum_{j=1}^{|X|} w_{k, k+1}(X)}, \mathrm{M}(A)=\frac{\sum_{k=1}^{|A|} W_{k, k+1}(A)}{\sum_{j=1}^{|X|} W_{k, k+1}(X)} \tag{1}
\end{equation*}
$$

correspondingly.

## Auxiliary Lemma

We have $\mathrm{M}(A)=\frac{\xi}{\eta}$, where $\xi=\sum_{k=1}^{|A|} W_{k, k+1}(A)$ and $\eta=\sum_{j=1}^{|X|} W_{k, k+1}(X)$.

## Lemma

Let $\xi$ and $\eta$ be random variables that taking values in the intervals $l_{\xi}, l_{\eta}$ respectively on positive semiaxis and $l_{\eta} \subseteq((1-\delta) \mathbf{E}[\eta],(1+\delta) \mathbf{E}[\eta])$, $l_{\xi} \subseteq(\mathbf{E}[\xi]-\delta \mathbf{E}[\eta], \mathbf{E}[\xi]+\delta \mathbf{E}[\eta])$. Then

$$
\begin{gather*}
\mathbf{E}\left[\frac{\xi}{\eta}\right]=\frac{\mathbf{E}[\xi]}{\mathbf{E}[\eta]}+\frac{\mathbf{E}[\xi]}{\mathbf{E}^{3}[\eta]} \sigma^{2}[\eta]+\frac{1}{\mathbf{E}^{2}[\eta]} \mathbf{C o v}[\xi, \eta]+r_{1},  \tag{2}\\
\sigma^{2}\left[\frac{\xi}{\eta}\right]=\frac{1}{\mathbf{E}^{2}[\eta]} \sigma^{2}[\xi]+\frac{\mathbf{E}^{2}[\xi]}{\mathbf{E}^{2}[\eta]} \sigma^{2}[\eta]-\frac{2 \mathbf{E}[\xi]}{\mathbf{E}^{3}[\eta]} \mathbf{C o v}[\xi, \eta]+r_{2}, \tag{3}
\end{gather*}
$$

where $\operatorname{Cov}[\xi, \eta]$ is a covariation of random variables $\xi$ and $\eta ; r_{1}, r_{2}$ are the residuals. It being known that $\left|r_{1}\right| \leq \frac{\mathrm{E}(\xi]+\mathrm{E}[\eta]}{(1-\delta) \mathbf{E}[\eta]} \delta^{3},\left|r_{2}\right| \leq C \delta^{3}$.

## Stochastic Measure of Informativeness by Contour Length. Assumption about noise and features

- original contour is corrupted by noise: $X=\left\{x_{k}+\mathbf{n}_{k}\right\}_{k=1}^{m}, x_{k} \in R^{2}$ and $\mathbf{n}_{k}=\left(\xi_{k}, \eta_{k}\right)$ are random variables;
- $\xi_{k}, \eta_{k}, k=1, \ldots, m$, are independent, normally distributed and such that $\mathbf{E}\left[\xi_{k}\right]=\mathbf{E}\left[\eta_{k}\right]=0, \sigma^{2}\left[\xi_{k}\right]=\sigma^{2}\left[\eta_{k}\right]=\sigma^{2}, k=1, \ldots, m$;
- the random length satisfied approximately to conditions of Lemma 1. Suppose that $W_{k, k+1}(X), k=1, \ldots, m$, are independent random variables.


## Numerical Characteristics of Random Variable $W_{k, k+1}(A)$

## Proposition 1

The following asymptotic equalities are valid

$$
\begin{gathered}
\mathbf{E}\left[W_{k, k+1}(A)\right]=l_{k}\left(1+\frac{\sigma^{2}}{l_{k}^{2}}+\frac{\sigma^{4}}{2 l_{k}^{4}}+O\left(\frac{\sigma^{6}}{l_{k}^{6}}\right)\right), \\
\sigma^{2}\left[W_{k, k+1}(A)\right]=2 \sigma^{2}\left(1-\frac{\sigma^{2}}{l_{k}^{2}}+O\left(\frac{\sigma^{4}}{l_{k}^{4}}\right)\right),
\end{gathered}
$$

where $l_{k}=\left|x_{k+1}(A)-x_{k}(A)\right|$.

Let $\mathbf{l}_{i}=\mathbf{l}_{i}(A)=x_{i+1}(A)-x_{i}(A)$ be a i-s segment-vector of polygon $A$
and $\alpha_{i}=\alpha\left(x_{i}\right)=\left(\mathbf{l}_{i-1}, \mathbf{l}_{i}\right)$.

## Proposition 2

We have

$$
\operatorname{Cov}\left[W_{k-1, k}(A), W_{k, k+1}(A)\right]=
$$

$$
=-\sigma^{2} \cos \alpha_{k}\left(1-\left(\frac{1}{l_{k-1}^{2}}+\frac{\cos \alpha_{k}}{2 l_{k-1} l_{k}}+\frac{1}{l_{k}^{2}}\right) \sigma^{2}+o\left(\frac{\sigma^{2}}{l^{2}}\right)\right),
$$

where $l_{k}=l\left(x_{k}\right)=\left|x_{k+1}(A)-x_{k}(A)\right|, l=\min \left\{l_{k-1}, l_{k}\right\}$.

## The Numerical Characteristics of Stochastic Measure of

 Informativeness by LengthLet $\alpha(x)(\beta(x)$ ) be an inner angle of polygon $A$ (polygon $X$ ) in vertex $x$, $\gamma(x)$ be an angle between the vectors $x_{+1}(A)-x, x_{+1}(X)-x$, where $x_{+1}(A)\left(x_{+1}(X)\right)$ is the next point w.r.t. $x$ in the contour $A$ (contour $\left.X\right)$.


## Asymptotic Equality for Expectation

## Theorem

The asymptotic equality

$$
\begin{equation*}
\tilde{\mathbf{E}}[\mathrm{M}(A)]=\frac{L(A)}{L(X)}+\frac{C_{1}(A)}{L^{2}(X)} \sigma^{2}+o\left(\frac{\sigma^{2}}{\Delta^{2}(A)}\right), A \in 2^{X}, \tag{4}
\end{equation*}
$$

is true, where

$$
\begin{aligned}
C_{1}(A)= & -L(A) \sum_{x \in X}\left|\mathbf{1}_{x}\right|^{-1}+L(X) \sum_{x \in A}\left|\mathbf{1}_{x}\right|^{-1}+4 \frac{L(A)}{L(X)} \sum_{x \in X} \cos ^{2} \frac{\beta(x)}{2}- \\
& -4 \sum_{x \in A} \cos \frac{\alpha(x)}{2} \cos \frac{\beta(x)}{2} \cos \left(\gamma(x)+\frac{1}{2} \alpha(x)-\frac{1}{2} \beta(x)\right) .
\end{aligned}
$$

and where $L(A)=\sum_{k=1}^{|A|}\left|x_{k+1}(A)-x_{k}(A)\right|$ is the length of contour $A$ without an influence of noise.

## Asymptotic Equality for Variance

## Theorem

The asymptotic equality

$$
\tilde{\sigma}^{2}[\mathrm{M}(A)]=4 \frac{C_{2}(A)}{L^{2}(X)} \sigma^{2}+o\left(\frac{\sigma^{2}}{\Delta^{2}(A)}\right), \quad A \in 2^{X}
$$

is true, where

$$
\begin{gathered}
C_{2}(A)=\sum_{x \in A} \cos ^{2} \frac{\alpha(x)}{2}+\frac{L^{2}(A)}{L^{2}(X)} \sum_{x \in X} \cos ^{2} \frac{\beta(x)}{2}- \\
-2 \frac{L(A)}{L(X)} \sum_{x \in A} \cos \frac{\alpha(x)}{2} \cos \frac{\beta(x)}{2} \cos \left(\gamma(x)+\frac{1}{2} \alpha(x)-\frac{1}{2} \beta(x)\right) .
\end{gathered}
$$

## Optimization Task for Stochastic Measure. Problem 1

The value of variance of stochastic informational measure characterizes the degree of stability of informational measure of curve with respect to level of curve noise.

## Problem 1

Find a polygonal representation of fixed cardinality $A \in 2^{X},|A|=k$, which minimized the value of variance of stochastic informational measure by length.

The polygonal representation

$$
A=\arg \min _{A \in 2^{x},|A|=k} C_{2}(A)
$$

is a solution of indicated problem for not great level of curve noise $\sigma$.

## Example

Let $X=\left\{x_{1}, \ldots, x_{6}\right\}$ be an ordered set of vertexes of regular 6-gon with length of segment is equal 1 .
Calculate the value $C_{2}(A)$ for various polygonal representations $A$ of cardinality $|A|=3: A_{1}=\left\{x_{1}, x_{3}, x_{5}\right\}, A_{2}=\left\{x_{1}, x_{2}, x_{4}\right\}$, $A_{3}=\left\{x_{1}, x_{2}, x_{3}\right\}:$
$C_{2}\left(A_{1}\right)=1.125, C_{2}\left(A_{2}\right)=1.25, C_{2}\left(A_{3}\right) \approx 1.66$.
Thus the contour $A_{1}$ is a most stable contour to noise pollution with respect to informational measure by length among of contours of cardinality is equal 3.


## Optimization Tasks for Stochastic Measure. Problem 2

The value of $\tilde{\mathbf{E}}[\mathrm{M}(A)]$ characterizes the overall information content polygonal representation.
The greater value of $\tilde{\sigma}^{2}[\mathrm{M}(A)]$ corresponds to greater relative value in representation of points of high informativeness (with sharp corners).

## Problem 2

Find a polygonal representation which maximize values $\tilde{\mathbf{E}}[\mathrm{M}(A)]$ and $\tilde{\sigma}^{2}[\mathrm{M}(A)]$.

## Example

Let $X$ - regular $2^{n}$-gon with length of segment is equal $b$ that inscribed in a circle of radius $R ; A=A_{m}$ - regular $2^{m}$-gon ( $m \leq n$ ) with length of segment is equal $a$. Then $\alpha(x)=\pi\left(1-2^{1-m}\right), \beta(x)=\pi\left(1-2^{1-n}\right)$, $a=b \sin \left(\pi 2^{-m}\right) \sin ^{-1}\left(\pi 2^{-n}\right), \gamma(x)=\frac{\beta(x)-\alpha(x)}{2}$. We have

$$
\begin{gathered}
\tilde{\mathbf{E}}[\mathrm{M}(A)] \approx \mu(A)\left(1+\left(\frac{\sigma}{a}\right)^{2}-\left(\frac{\sigma}{b}\right)^{2}\right), \\
\tilde{\sigma}^{2}[\mathrm{M}(A)] \approx \mu^{2}(A)\left(\frac{|X|}{|A|}-1\right)\left(\frac{\sigma}{R}\right)^{2}
\end{gathered}
$$

For example, for criteria

$$
k(m)=(1-\alpha) \tilde{\mathbf{E}}\left[\mathrm{M}\left(A_{m}\right)\right]+\alpha \tilde{\sigma}\left[\mathrm{M}\left(A_{m}\right)\right]
$$

and $n=6$ we have $\arg \max k(m)=6$ if $\alpha=0.5$ and $\arg \max k(m)=3$ if $\alpha=0.75$.

## Conclusion

- we recieved the asymptotic formulas that expresses the numerical characteristic of stochastic measure of informativeness by length through the geometrical relations of discrete contour;
- there are different formulations of problem about finding of the "best"representation of discrete contour with help of stochastic measure of informativeness;
- we considered the finding of a polygonal representation of fixed cardinality, which minimized the value of variance of stochastic informational measure by length;
- for example, the other problem consist to finding of polygonal representation which maximize the variance of stochastic measure of informativeness. Because the some informational features (such as curvature) may to strongly change under the influence of a noise.


# Thanks for you attention 

alex.lepskiy@gmail.com

