Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Master of Science in Geospatial Technologies

Geostatistics
Assessment of Local Uncertainty with Indicator Geostatistics II

Carlos Alberto Felgueiras
cfelgueiras@isegi.unl.pt

Assessment of Local Uncertainty

Contents

Assessment of Local Uncertainty

- **Introduction**

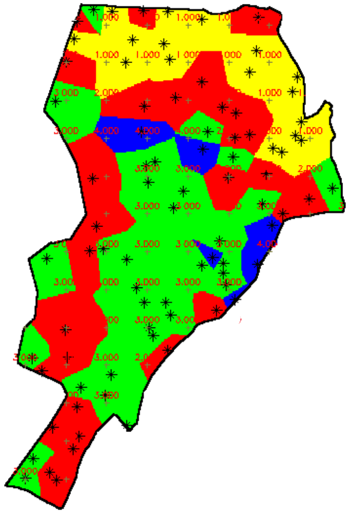
- **Spatial Models: Continuous x Categorical Variables**

- **The Kriging (simple kriging for example)** allows the estimation of continuous attributes from the equation:

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) \cdot Z(\mathbf{u}_\alpha) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})\right] \cdot m$$

- How to works on categorical variables? Is it possible to use geostatistics to assess the uncertainty model of a categorical Random Variable?

- Deterministic solution: Nearest neighbors estimator. What are the drawbacks of this model? What about the quality of these maps?



Assessment of Local Uncertainty

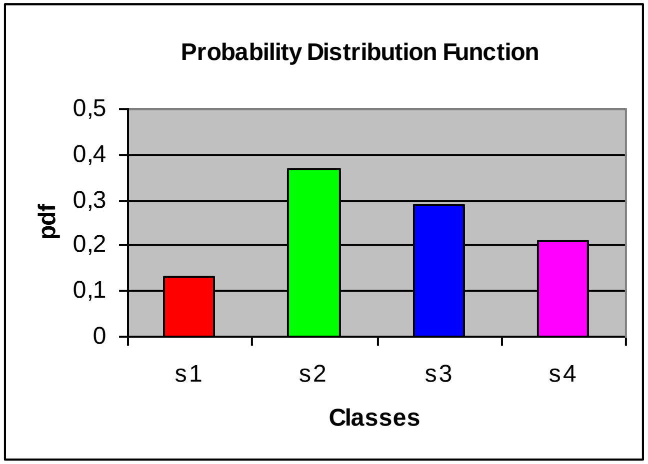
- **Introduction**
- **Categorical Random Variables**

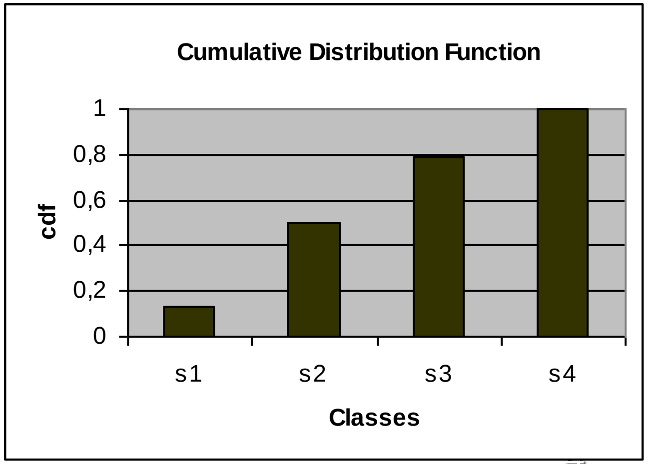
- *Definition:* A categorical Random Variable $S(\mathbf{u})$ is composed by a set of categorical (discrete) values that are associated to an uncertainty model (probability distribution function) representing the probability distribution of its values.

- *Properties*
    - Each one of the s_k values, or classes, (A, B, C or D) is associated to a probability of occurrence $p_k \in [0,1]$.
    - The summation of the class (or value) probabilities at a location **u** is equal 1

$$\sum_{k=1}^{K} p_k(\mathbf{u}) = \sum_{k=1}^{K} p(\mathbf{u}; s_k) = 1$$

- The categorical pdf can be transformed in a cdf (cumulative distribution function) considering an a priori order among the classes.





Assessment of Local Uncertainty

- **Indicator Approach for categorical variables**
  - **Indicator transformation and properties**

Instead of the Variable $S(\mathbf{u})$, consider its binary indicator transform $I(\mathbf{u};z_k)$ as defined by the relation:

$$I(\mathbf{u};s_k) = \begin{cases} 1, & \text{if} \quad S(\mathbf{u}) = s_k \\ 0, & \text{otherwise} \end{cases}$$

Kriging of the indicator R.V. $I(\mathbf{u};z)$ provides an estimate that is also the best LS estimate of the conditional expectation of $I(\mathbf{u};z)$. Now the conditional expectation of $I(\mathbf{u};z)$ is equal to the local pdf of $Z(\mathbf{u})$; indeed:

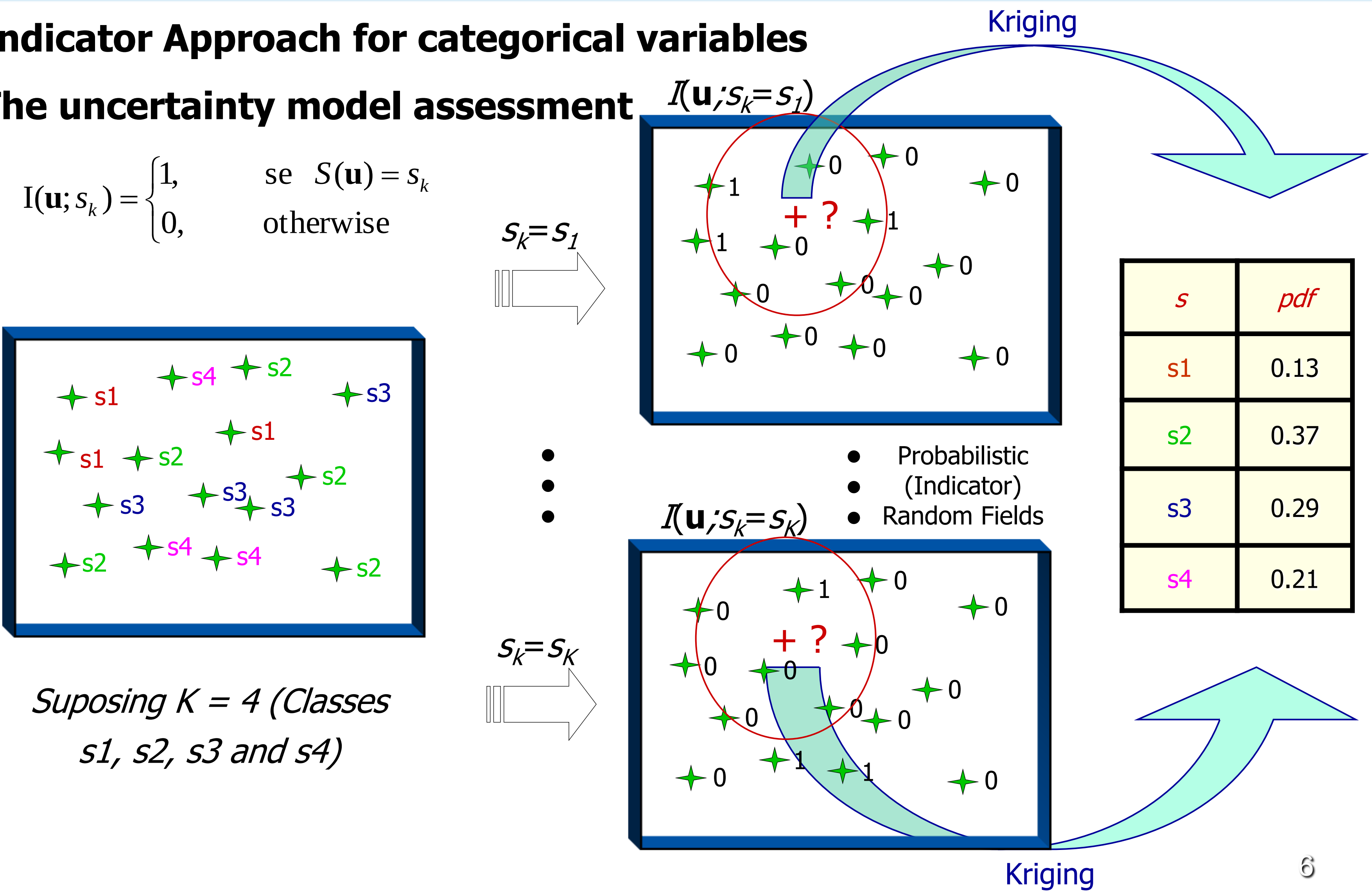
$$\begin{aligned} \mathrm{E}\{I(\mathbf{u};s_k)\,|\,(n))\} &= 1\cdot\mathrm{Prob}\{I(\mathbf{u};s_k)=1\,|\,(n)\} + 0\cdot\mathrm{Prob}\{I(\mathbf{u};s_k)=0\,|\,(n)\} \\ &= 1\cdot\mathrm{Prob}\{I(\mathbf{u};s_k)=1\,|\,(n)\} = \mathrm{p}(\mathbf{u};s_k\,|\,(n)) \end{aligned}$$

Important: So the indicator kriging is not aimed at estimating the unsampled value $z(\mathbf{u})$; or its indicator transform $i(\mathbf{u};z)$ but at providing a pdf model of uncertainty about $z(\mathbf{u})$.

Assessment of Local Uncertainty

- **Indicator Approach for categorical variables**
- **The uncertainty model assessment**

$$I(\mathbf{u}; s_k) = \begin{cases} 1, & \text{se} \quad S(\mathbf{u}) = s_k \\ 0, & \text{otherwise} \end{cases}$$

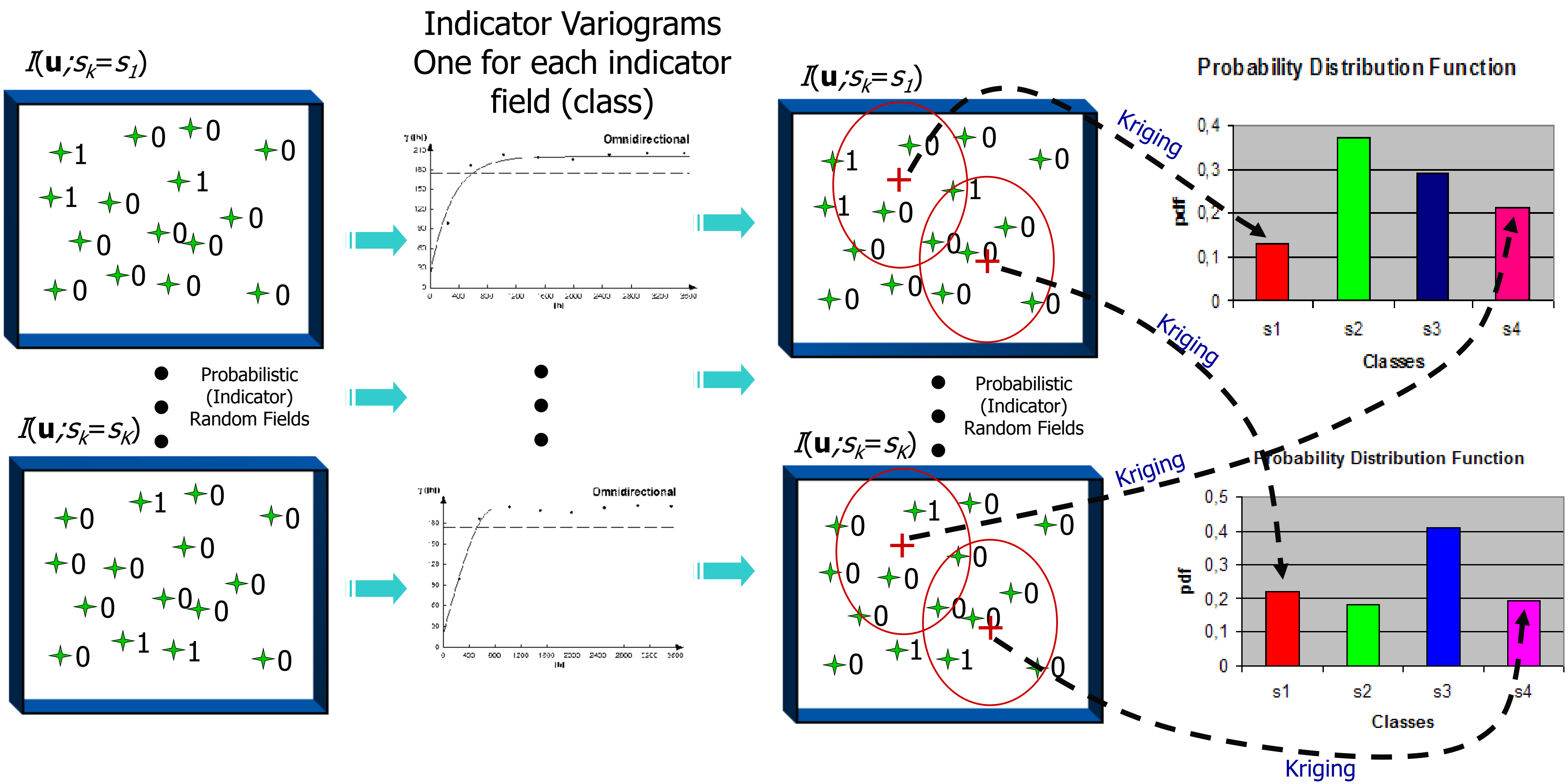


s	*pdf*
s1	0.13
s2	0.37
s3	0.29
s4	0.21

Suposing K = 4 (Classes s1, s2, s3 and s4)

Assessment of Local Uncertainty

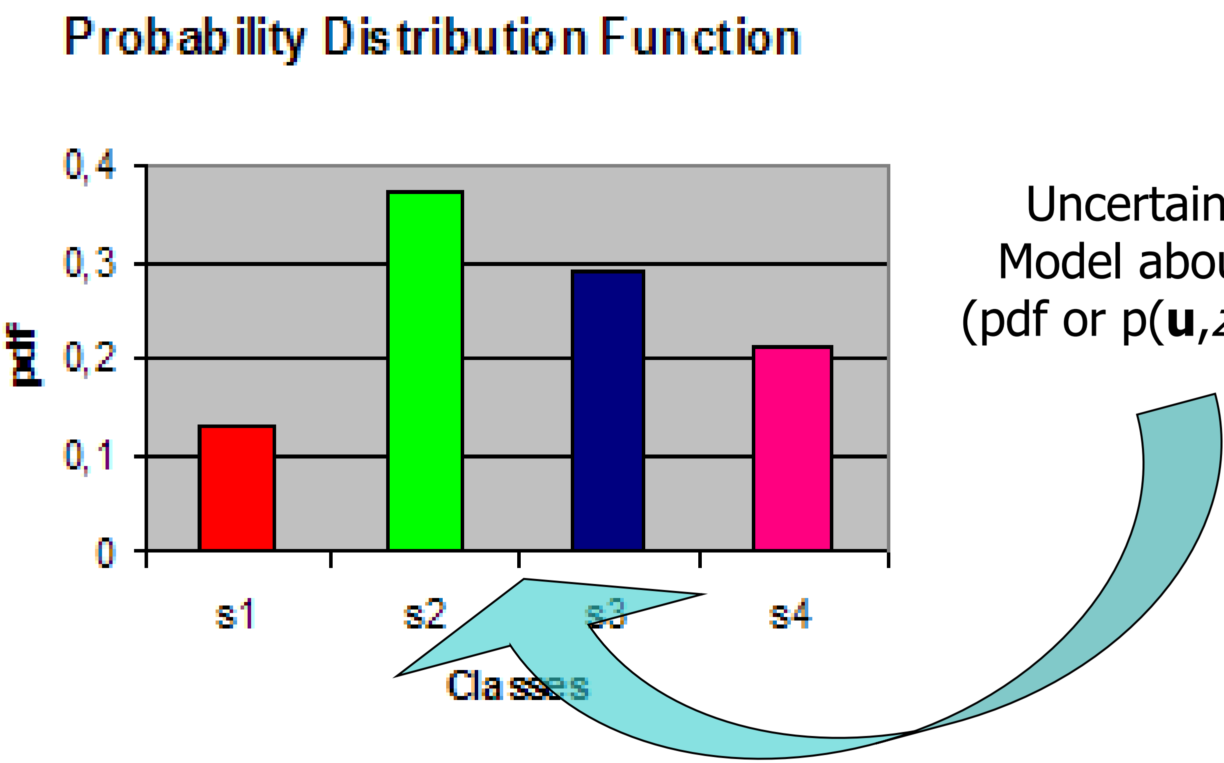
- **Indicator Approach for categorical variables**
  - **The uncertainty model assessment (more details)**



Assessment of Local Uncertainty

- **Indicator Approach for categorical variables**
  - **The uncertainty model assessment (Ilustration) :** Using *K* cutoffs, or thresholds, values (in this case, for 4 classes, *K*=4)

s	*pdf*
s1	0.13
s2	0.37
s3	0.29
s4	0.21

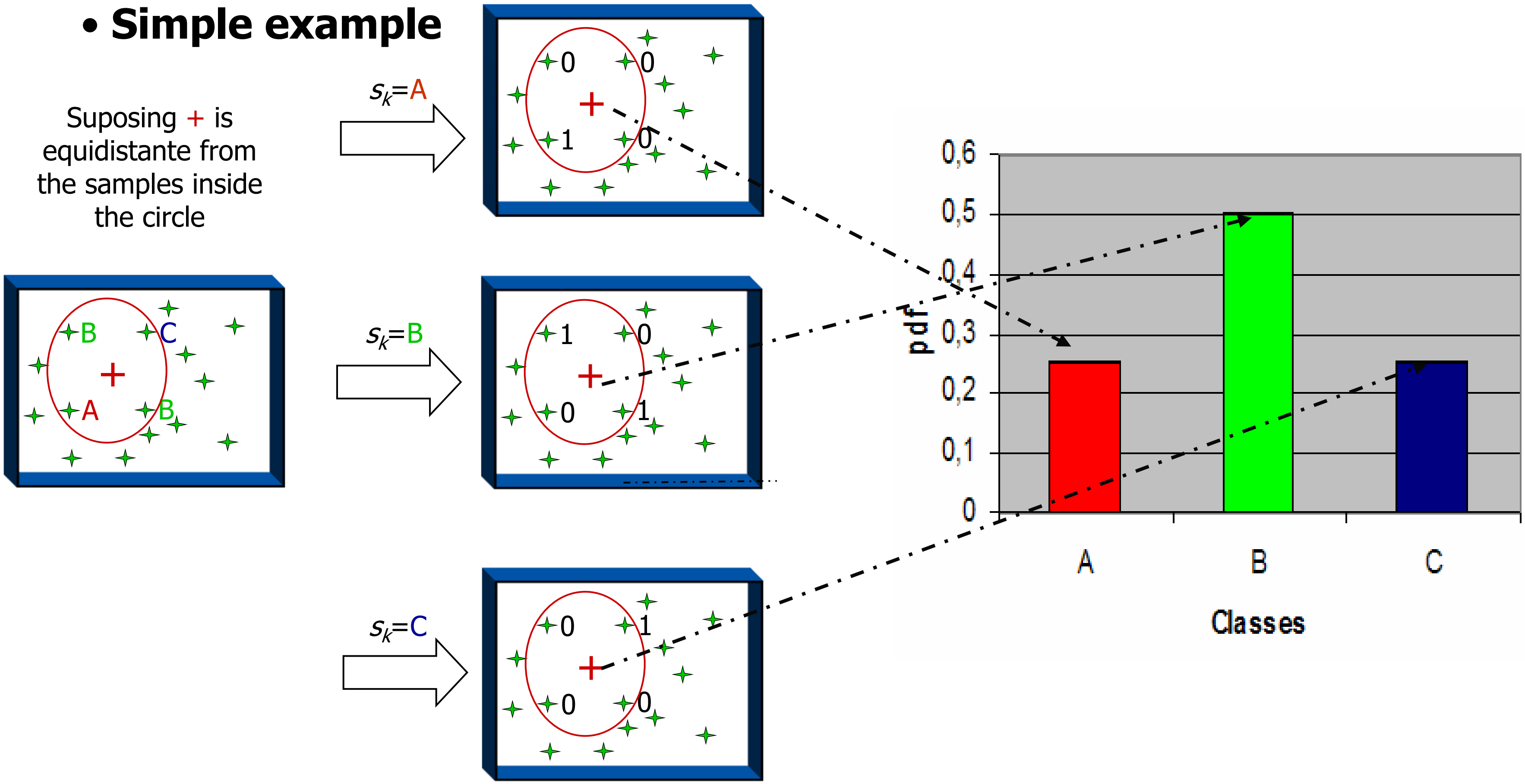


Uncertainty Model about *s* (pdf or p(**u**,*z*|(*n*))

Assessment of Local Uncertainty

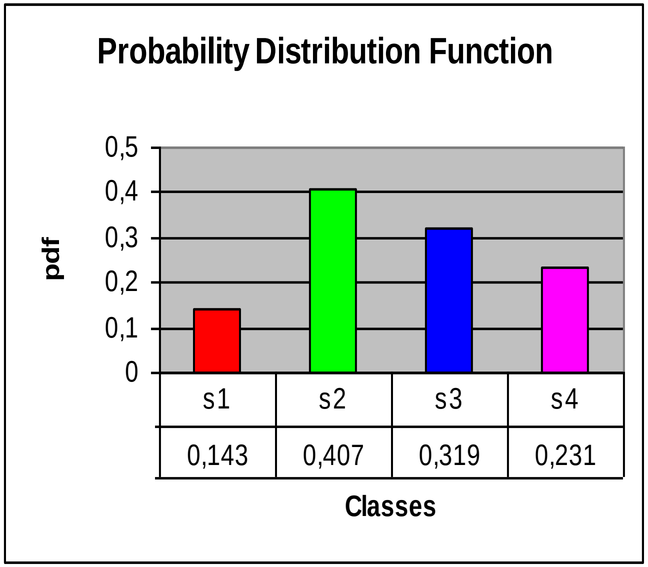
- **Indicator Approach for categorical variables**
  - **Simple example**

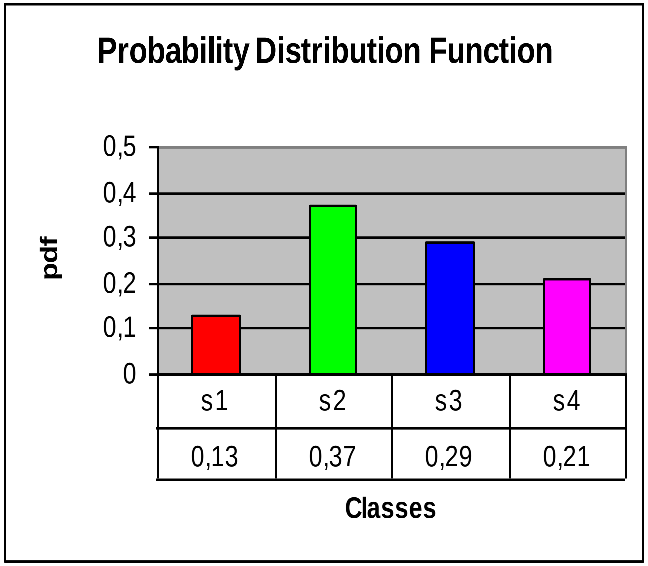
Suposing + is equidistante from the samples inside the circle



Assessment of Local Uncertainty

- **Indicator Approach for categorical variables**
  - **Correcting for order relation deviations (Goovaerts, 1997) –** the final summation of pdf values, $p(\mathbf{u},S|(n))$, must be equal 1. So it may be necessary corrections for order relation deviations to guarantee this property





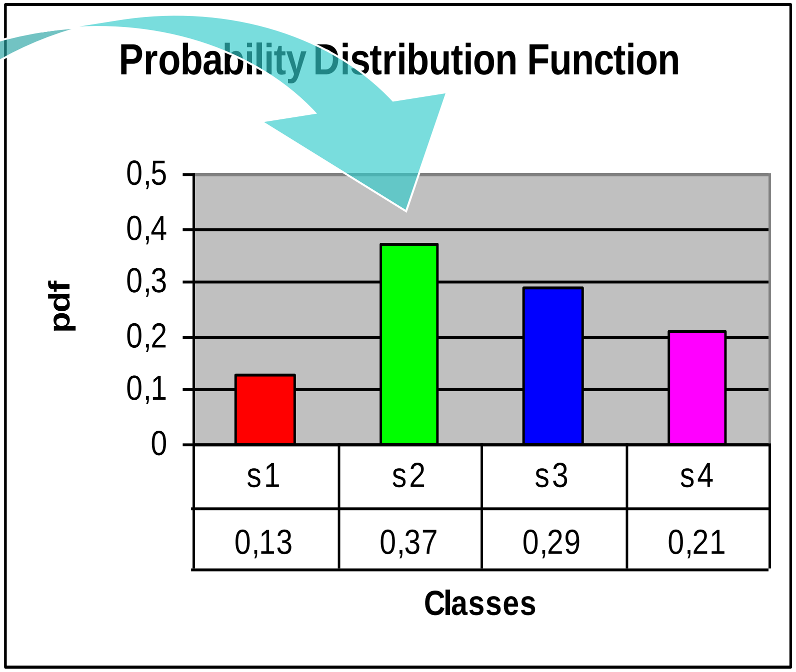
$$\sum_{k=1}^{K} p_k(\mathbf{u}) = \sum_{k=1}^{K} p(\mathbf{u}; s_k) > 1$$

$$\sum_{k=1}^{K} p_k(\mathbf{u}) = \sum_{k=1}^{K} p(\mathbf{u}; s_k) = 1$$

Assessment of Local Uncertainty

- **Indicator Approach for categorical variables**
  - **Estimating RV parameters** (***Mode or Maximum Probability Criterion***)
  - The ***categorical value*** can be estimated, at a spatial location **u**, from a pdf representation as:

$$s^*(\mathbf{u}) = \mathrm{Max}_{k=1}^{K}\left(p(\mathbf{u}; s_k)\right) = \mathrm{Max}_{k=1}^{K}\left(p_k(\mathbf{u})\right)$$



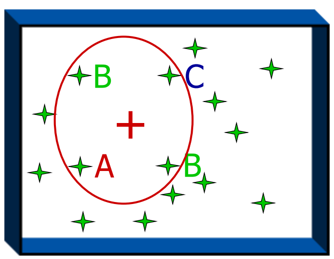
In this example $s^*(\mathbf{u})$ = s2

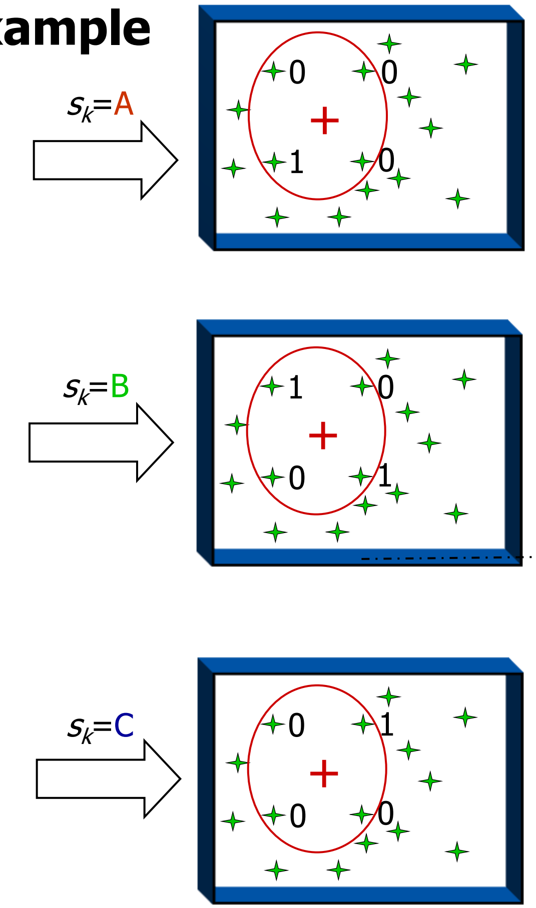
Any other estimator ?

Assessment of Local Uncertainty

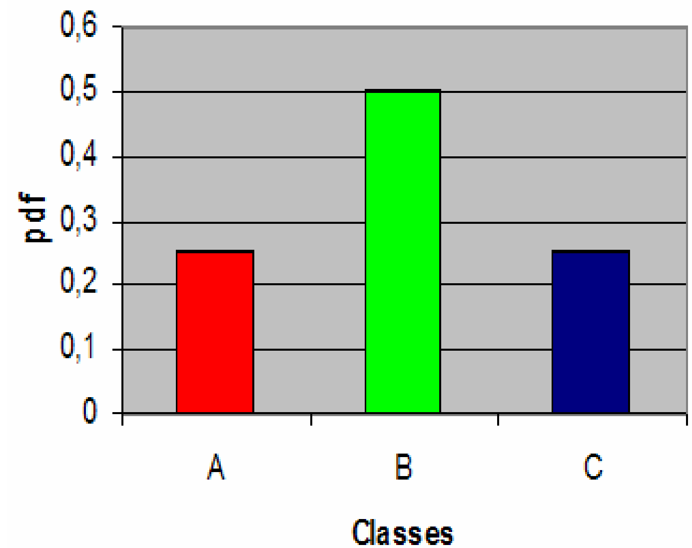
- **Indicator Approach for categorical variables**
  - **Simple example**

Suposing + is equidistante from the samples inside the circle





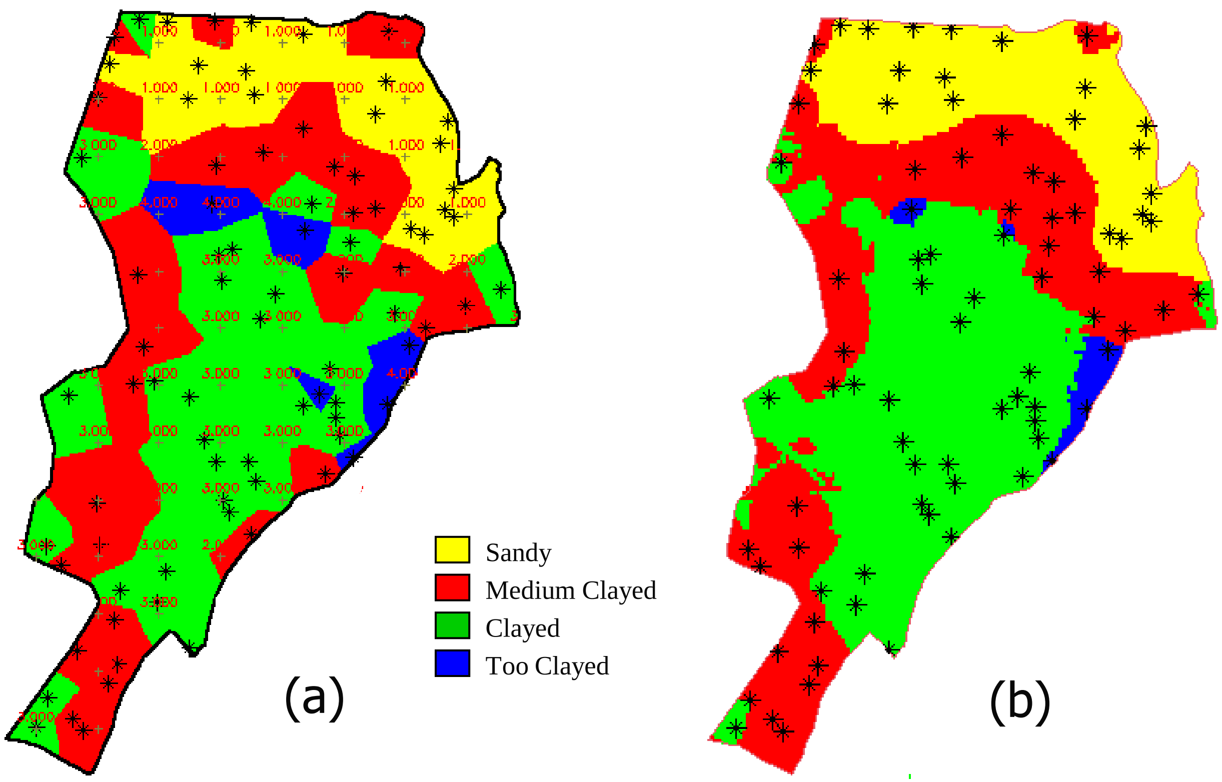
s	*pdf*
A	0.25
B	0.5
C	0.25



In this example *s(u) = B**

Assessment of Local Uncertainty

- **Indicator Approach for categorical variables**
  - Examples of Prediction Maps ((a) Dirichlet and (b) Maximum Probability



Classes of soil textures

Assessment of Local Uncertainty

- **Indicator Approach for categorical variables**
  - **Uncertainties evaluations from the probability distribution function**
    - ***Mode***: based on the estimator of maximum probability criterion.

$$\mathrm{Unc}(\mathbf{u}) = 1 - p(\mathbf{u}; s^*(\mathbf{u})) = 1 - \mathrm{Max}_{k=1}^{K}(p_k(\mathbf{u}))$$

    - ***Shannon Entropy***: a measure of the disorder or randomness in a closed system

$$\mathrm{Unc}(\mathbf{u}) = -\sum_{k=1}^{K} p_k(\mathbf{u}).\ln(p_k(\mathbf{u}))$$

Shannon Entropy main properties:

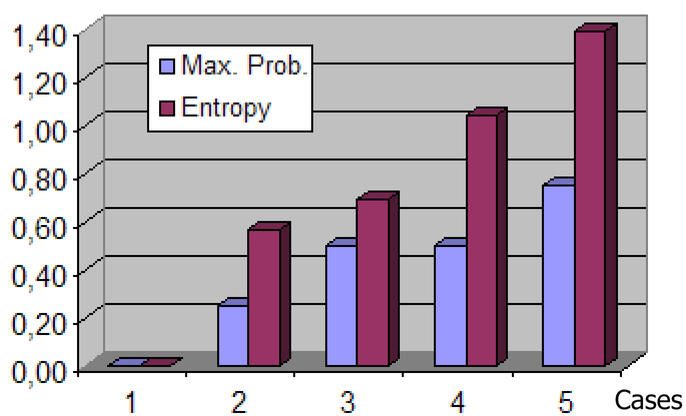
- It is a non negative value
- It is zero when one of the probabilities is 1 (the others are 0)
- It is maximum ocurrs when the distribution is uniform (maximum confusion)

Assessment of Local Uncertainty

- **Indicator Approach for categorical variables**
  - **Uncertainties evaluations from the probability distribution function**
  - **Example: Uncertainty by Max Prob. x Shannon Entropy**

Cases	p(u,s1)	p(u,s2)	p(u,s3)	p(u,s4)	Max Prob	Shannon
1	0,00	0,00	0,00	1,00	0,00	0,00
2	0,00	0,00	0,25	0,75	0,25	0,57
3	0,00	0,00	0,50	0,50	0,50	0,69
4	0,00	0,25	0,25	0,50	0,50	1,04
5	0,25	0,25	0,25	0,25	0,75	1,39

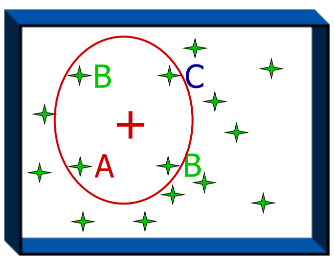
Table: Probability configuration examples with uncertainty measures

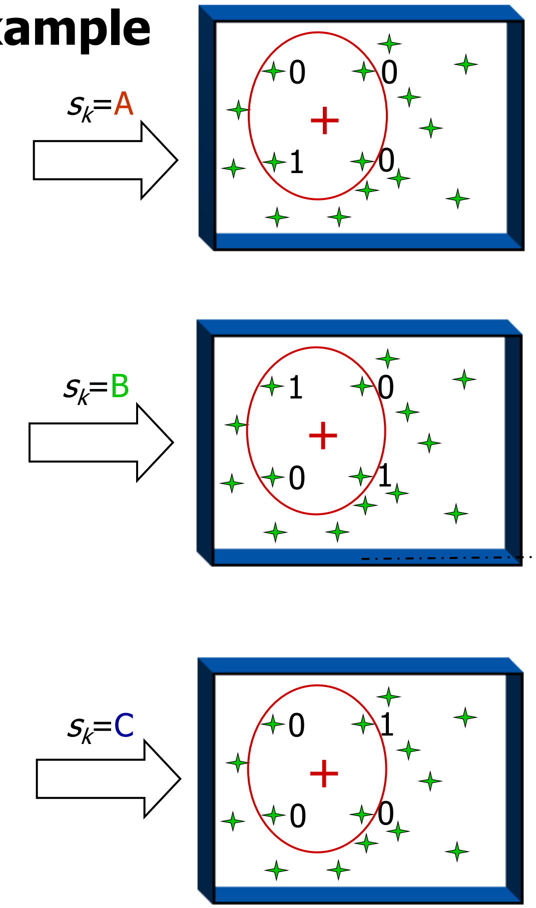


Assessment of Local Uncertainty

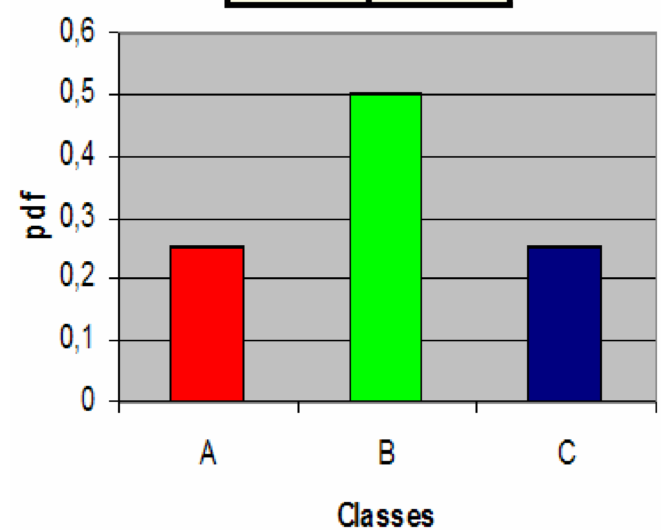
- **Indicator Approach for categorical variables**
  - **Simple example**

Suposing + is equidistante from the samples inside the circle





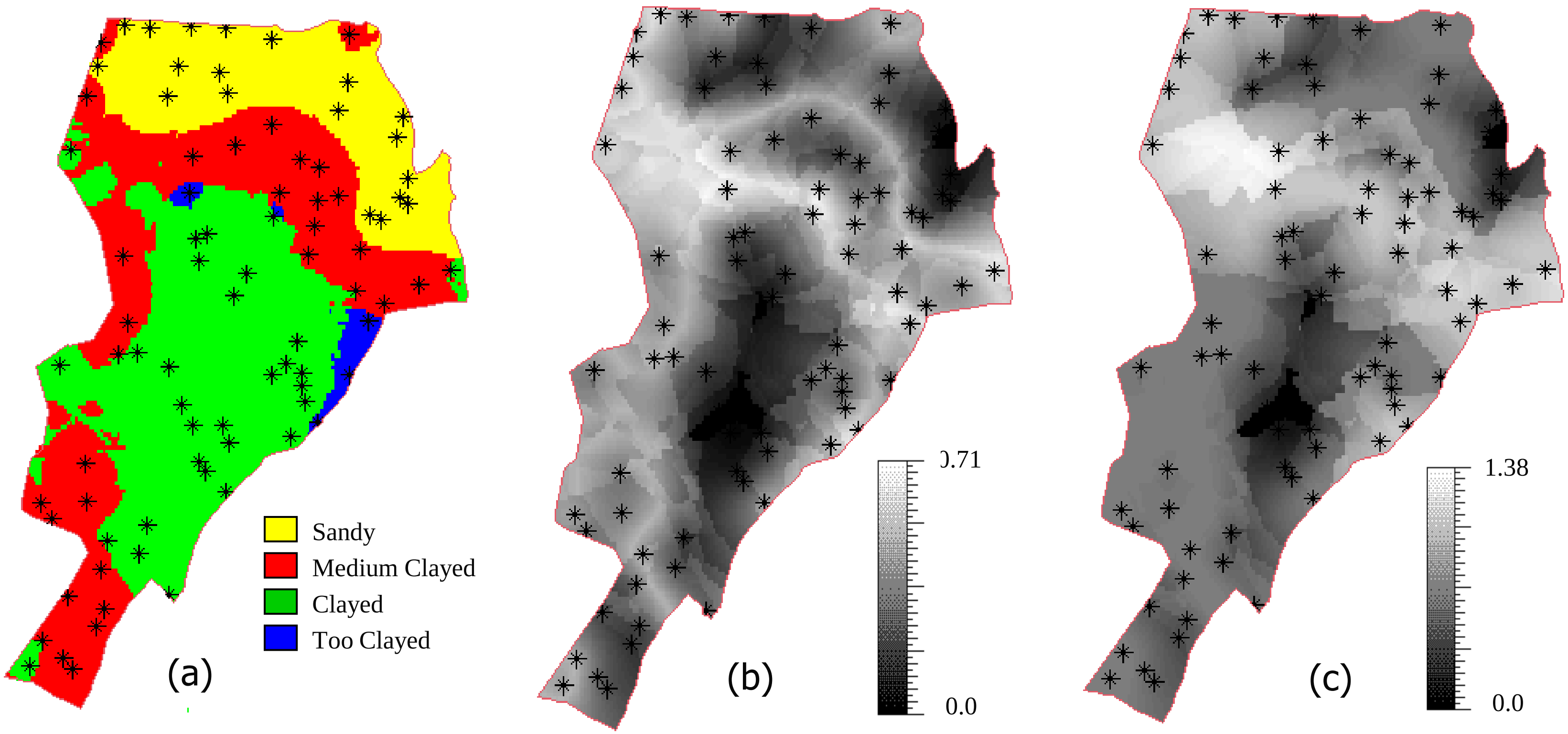
s	*pdf*
A	0.25
B	0.5
C	0.25



In this example Unc(**u**) = .25 (MaxProb) and Unc(**u**)= 1.04 (Shanon)

Assessment of Local Uncertainty

- **Indicator Approach for categorical variables**
  - Maps: (a) Estimates by Maximum Probability, Uncertainties by (b) Maximum Probability and (c) Shannon Entropy



Assessment of Local Uncertainty

- **Problems with indicator geostochastic procedures**

The main drawback of using indicator geostatistic approaches is the need of work on variogram generations and fittings for each cutoff . This work is interactive and requires from the user knowledge of the main concepts related to basics of the geostatistics, and indicator approaches, in order to obtain reliable results.

Indicator approach for categorical attributes requires the number of cutoffs be equal to the number of classes presented in the sample data. The user must generate and fit one variogram for each indicator field related to each specific class.

Assessment of Local Uncertainty

- **Advantages on using indicator geostochastic procedures**
  - All the advantages of the geostatistic approaches because of the use of:
    - variograms to represent the variation of the attribute.
    - kriging to estimate the values considering covariance between samples and between the samples and the point to be estimated
  - Allows the assessment of the local uncertainty model at any **u** spatial location that can be used for getting:
    - estimate maps using different distribution parameters as mean, median or any quantil when the attribute values are continuous
    - uncertainty maps based on confidence intervals of standard deviation or quantils when the attribute values are continuous
    - estimate maps for categorical attributes using the pdf of each R.V.
    - uncertainty maps for categorical attributes based in the probability of the estimate or in the Shannon Entropy using all the information of the R. V. probability distribution function.

Assessment of Local Uncertainty

Summary and Conclusions

- If the value to be estimated is the expected value (mean), standard krigings (simple, ordinary, cokriging, ….) are a priori the preferred algorithm.
- Otherwise, the ***indicator kriging*** provides tools for constructing an approximation of the uncertainty model (pdf or ccdf) of the attribute for any spatial location **u**.
- Indicator approaches can be applied to continuous variables and to categorical variables.
- For continuous attributes, the ccdf model allow the creation of estimate maps, other than the mean value, and uncertainty maps based on confidence intervals.
- For categorical attributes, the pdf model allow the creation of estimate maps (based in the maximum probability) as well as the creation of uncertainty maps based in the maximum probability or in the Shannon entropy.
- The uncertainties can be used to qualify the estimation at each spatial location **u** considered.

Assessment of Local Uncertainty

Exercises

1. Run the Lab7 to be available in the geostatistics course area of ISEGI online.
2. Given the estimated pdf (right) of a categorical variable:

   2.1 Plot the pdf of the variable (after order correction)

   2.2 Estimate a class value for this pdf.

   2.3 Calculate the uncertainty of the above estimate using the maximum probability criterion.

   2.4 Calculate the uncertainty of the above estimate using the Shannon entropy criterion.
3. Send a report to the professor about the above exercises, before 06/12/2007

s	*pdf*
A	.12
B	.17
C	.38
D	.32
E	.11
F	.1

Assessment of Local Uncertainty

Test 2

1. Using the data set of the first test apply the indicator approach for continuous attributes for:

   4.1 Generate mean and median estimate maps. Compare these maps between them and with the estimate maps of the ordinary kriging.

   4.2 Generate uncertainty value maps based in standard deviation and decil confidence intervals. Compare these maps between them and with the ordinary kriging variance maps.

   4.3 Prepare an "article" about the above work and send it to the professor before 13/12/2007.

   4.4 Prepare a presentation about the above work. To be presented (in 10-15 minutes) in the class of 13/12/2007.

Observation: We should have an extra class on Tuesday 18/12/2007 to replace the holyday of 01/11/2007. Can it be before this data? Maybe extend the class of 13/12 up to 19:00hs?

Predictions with Deterministic Procedures

END

of Presentation