(I) Statistical estimation of Population size for Brown Bear in the Rhodope mounting

(II) Density modelling using Monte Carlo method

Gurov T.[1], Atanassov E.[1], Serbezov R.[2], Spassov N.[3]

[1] *Institute of Information and Communication Technologies*

[2]*National Nature Protection Service, Ministry of Environment and Water*

[3] *National Museum of National History*

A. Karaivanova, T. Gurov, S.Ivanovska

IICT-BAS, www.iict.bas.bg

HP-SEE

High-Performance Computing Infrastructure for South East Europe's Research Communities

The HP-SEE initiative is co-funded by the European Commission under the FP7 Research Infrastructures contract no. 261499

OTLINE

- Motivation
- National Monitoring in the West Rhodope
- Statistical estimates using data of national monitoring
- Estimate population size of the brown bear in Rhodope Mountains
- Density modelling using Monte Carlo
- Numerical rezults
- Conclusion and future work

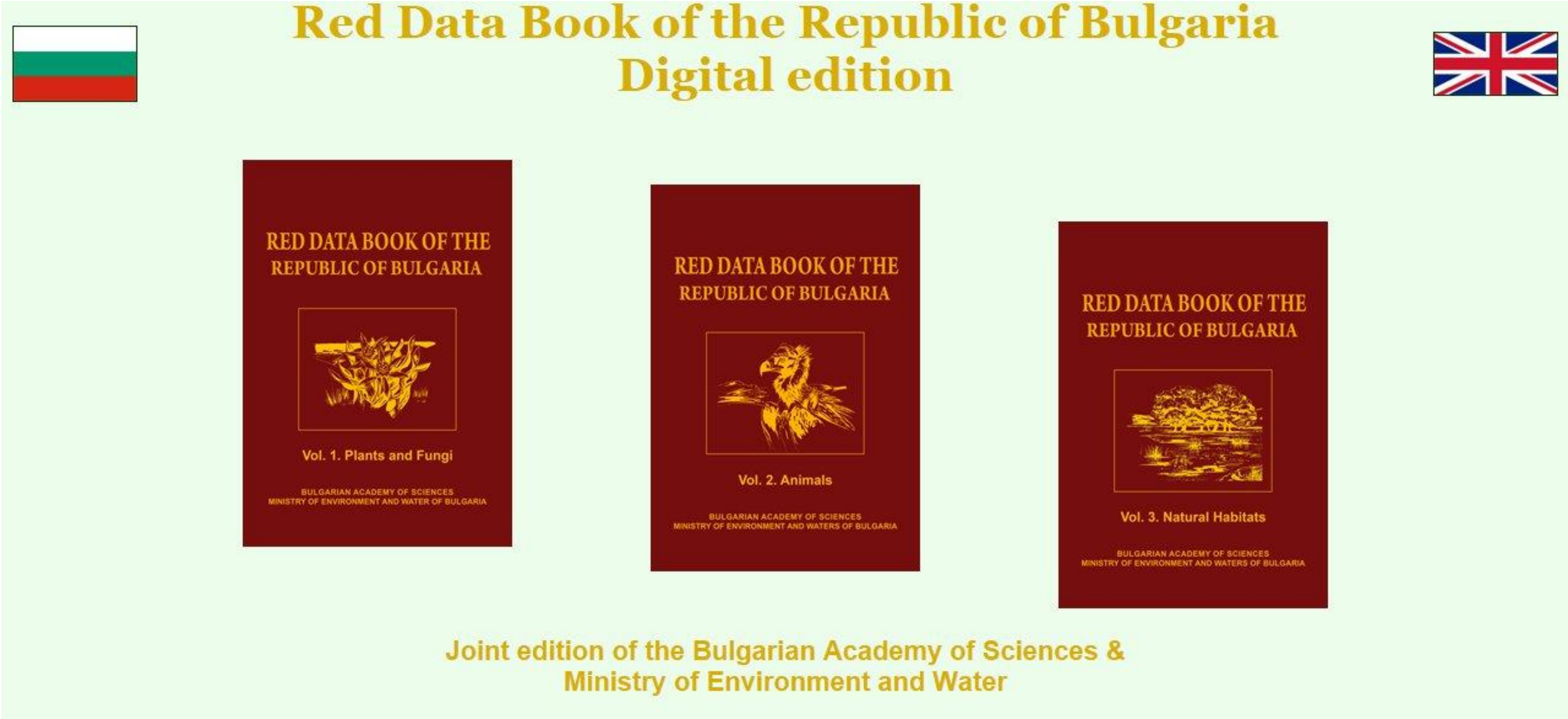
Motivation(1/2)

- One of the best brown bear (**Ursus arctos**) habitats in Europe are located in Bulgaria. They are situated in the mountain massifs – Rhodope, Stara planina, Rila, Pirin, Vitosha ;
- **Legal status**
  - Until 1992 the bear had been a game target. By Order №1023 dated 31.12.1992 of the Ministry of Environment and Water **(MoEW)** the species has been declared protected, in compliance with the Nature protection act. This status has been kept also after the Biodiversity act has passed in 2002.
- The Habitat directive requires a strict protection of the species and declaration of special protected areas for conservation of its habitats.
- **Special program**
  - **ACTION PLAN FOR THE BROWN BEAR IN BULGARIA - 2008**
- The main habitats of the bear in Bulgaria are included in the ecological network NATURA 2000. For the purposes of protection of the habitats and the management of the network NATURA 2000 a mapping and determination of their environmental status is carried out in the frame of project under Operating program environment. The acquired information are used for elaboration of plans for management of the protected areas, populations of the species as well as for regulation of the investment projects therein.

Motivation(2/2)

- Brown Bear (Ursus arctos) e a priority species for conservation of mammals in the European Union. Conservation status: in Bulgaria endangered EN [C2a (i)], BA-II, III, International: Beck-II; CITES-II; DH-II, IV
- Red book in Bulgaria, Vol.2 – Animals, Sofia, 2011, 2011.

  http://e-ecodb.bas.bg/rdb/bg/



National monitoring in the West Rhodope(1/2)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- ***Transect method:*** collection of traces of brown bear on predefined set of routes (transects) and analysis to determine the unique trace. More popular and cheaper method for monitoring the bear population;
- **Quantitative metric traits:**
  Traces - Length and width of the front paw;
  Traces - Length and width of the paw;
  The distance between the bear and the man at the meeting;
  Winter den - width and height of the entrance; Width and height of the den;
  Total area of suitable habitat unfragmented;
  Area of habitat suitable for placing dens;
  Quality nominal features:
  Color: tan, brown, black, gold, other;
  Sex: male, female, indeterminate;
  Attacks for the following reasons: protection of offspring, protection of prey at aggression with / at the bear; impaired critical distance; **Types of food** ;

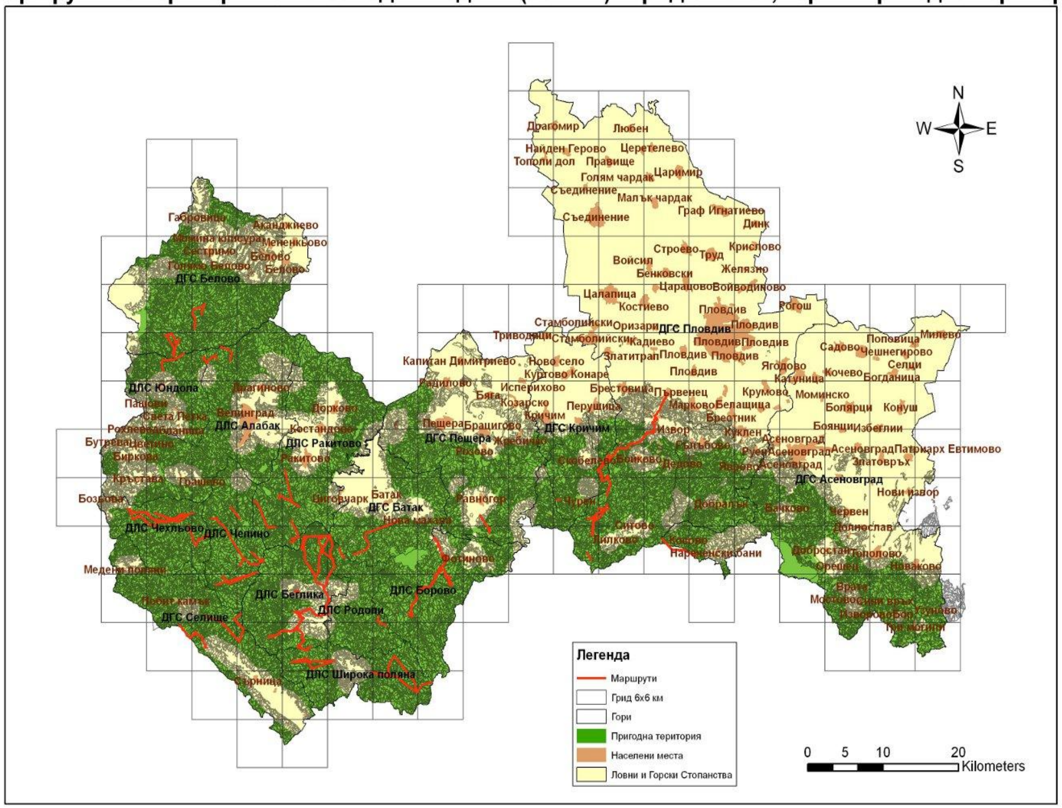
Ordinal qualitative features:
Size: bear, a small bear, small bear, a medium-sized bear, big bear, very large bear;
Behavior when meeting with person: Avoid immediately; away calmly, threatening; attacks.

National monitoring in the West Rhodope (2/2)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- **National monitoring – Autumn 2011.**
- **Number of transects – 48**
- **Forestry areas** (Forest administrative unit (forest farm)) **- 14**
- Determining a unique sample of traces of a brown bear:
- **quantitative indicator: width of the front paw of the bear**
- **Additional indicators: width/length of the paw of the bear**

Маршрути на територията на Западни Родопи (2011 г.) - грид 6 - 6 км, гори и пригодна територия



Statistical estimates using data of national monitoring (1/2)

Distribution of the bear's traces on the forestry areas and on the transects

FORESTRY AREA: (administrative unit or forest farm)	Asenovgrad			Chekeritsa					Batak		Belovo			
Transects	3			5					2		3			
Number of traces	0	0	0	0	4	1	1	1	0	1	0	0	2	
FORESTRY AREA:	Peshtera			Selishte					Alabak		Beglika			
Transects	3			3					2		3			
Traces	1	-	-	2	0	0			2	0	2	3	1	
FORESTRY AREA:	Borovo			Rakitovo					Rodopi		Shiroka polyana			
transects	3			2					2		3			
Traces	1	2	2	3	0				2	1	1	1	1	
FORETRY AREA:	Chepino/Chehlyovo											Yundolau		
Number of transects	11											3		
Number of bear's traces	1	0	1	0	1	2	4	0	1	2	0	0	0	0
Total transects :	46 from 48			traces:								47		

Statistical estimates using data of national monitoring (2/2)

- Mean value and variance on non-grouping data:

$$\overline{x} = \sum_{i=1}^{48} x_i / 46, \quad (S)^2 = \sum_{i=1}^{48} (x_i - \overline{x})^2 / 45 .$$

$$\bar{x} = 1.0217391 \quad \text{и} \quad (s)^2 = 1.1541,$$

- Standard deviation: *s =1.0742885*
- *Table2:* Number of unique traces by confidence interval β

Deviation	Min deviation	Mean value	Max deviation	Round to a whole number	Level of significance $\mathbf{x}_{\beta}$	Confidence interval β
1.0742885	45.89	49.04	52.19	**45-53**	3.00	99.7%
1.0742885	47.21	49.04	50.79	**47-51**	1.67	95%

***Statistical estimate depend on the choice of the confidence interval*.**

Mahalanobis distance (1/2)

- The **Mahalanobis distance** is a descriptive statistic that provides a relative measure of a data point's distance (residual) from a common point.

In general, if you have a normal (Gaussian) distribution X with variance

S=1 and mean $\mu = 0$

then any other normal distribution, R, can be defined

from X by the equation

$$R = \mu_1 + \sqrt{S_1}X.$$

Conversely, to recover a normalized distribution from any distribution, you can typically solve for

$$X = (R - \mu_1)/\sqrt{S_1}.$$

If you then square both sides, and take the square-root, you get an equation for a metric that looks a lot like the Mahalanobis distance:

$$D = \sqrt{X^2} = \sqrt{(R - \mu_1)^2/S_1} = \sqrt{(R - \mu_1)S_1^{-1}(R - \mu_1)}.$$

The resulting magnitude is always positive and varies with the distance of the data from the mean, attributes that are convenient when trying to define a model for the data.

Mahalanobis distance (2/2)

HP-SEE
-Performance Computing Infrastructure
outh East Europe's Research Communities

Formally, the Mahalanobis distance of a multivariate vector

$$x = (x_1, x_2, x_3, \dots, x_N)^T$$

from a group of values with mean

$$\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_N)^T$$

and covariance matrix S is defined as:

$$D_M(x) = \sqrt{(x-\mu)^T S^{-1} (x-\mu)}.$$

Mahalanobis distance (or "generalized squared interpoint distance" for its squared value) can also be defined as a dissimilarity measure between two random vectors $\vec{x}$ and $\vec{y}$ of the same distribution with the covariance matrix S:

$$d(\vec{x},\vec{y}) = \sqrt{(\vec{x}-\vec{y})^T S^{-1} (\vec{x}-\vec{y})}.$$

If the covariance matrix is the identity matrix, the Mahalanobis distance reduces to the Euclidean distance. If the covariance matrix is diagonal, then the resulting distance measure is called the *normalized Euclidean distance*:

$$d(\vec{x},\vec{y}) = \sqrt{\sum_{i=1}^{N} \frac{(x_i - y_i)^2}{s_i^2}},$$

where s_i is the standard deviation of the x_i and y_i over the sample set.

Estimation of population size (1/4)

- Geographical Information System (GIS model) of the fit areas bear habitats.
  - The model is created using distance classifier based on a statistical method (*Mahalanobis distance* – D^2).
  - For this purpose, 412 GPS locations of brown bears are used (traces marked trees, burrows, observations, genetic samples collected from sites in the hair bear habitat, excrement, etc.).

- 4 groups woodland :
  - „other land cover" (like blackberries, shrubs, forest herbs – mean value of D^2 - 11.4
  - coniferous forest - mean value of D^2 - 6.8
  - mixed forests - mean value of D^2 - 3.6
  - deciduous forest - mean value of D^2 - 5.56

Estimation of population size (2/4)

- 4 fit areas bear habitats in square kilometres (km^2), grouped by regions of the Rhodope mountain

Fit areas bear habitats	„other land cover”	coniferous forest	mixed forests	deciduous forest	Total
14 Forestry areas with transects	271.72	835.99	399.51	343.73	1850.95
Other forestry areas (Pazardjik and Plovdiv) without transects	351.619	896.9119	452.226	448.6128	2149.369
all fit areas bear habitats in Smolyan and Kardjalii	671.28	988.36	609.9237	447.5292	2717.094

Estimation of bear population size (3/4)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

Consider the equation:

$$x_1*271.72+x_2*835.99+x_3*399.51+x_4*343.73 = 49$$

$x_1= \mu*(27.4 =11.4+6.8+3.6 +5.6)/11.4$;
$x_2= \mu*27.4/6.8$;
$x_3= \mu*27.4/3.6$;
$x_4 = \mu*27.4/5.6$

Compute $\mu=0.005613$

- Thus we obtain for the weights:

$x_1 = 0.0135105$, $x_2 = 0.022387$, $x_3 =0.042783$, $x_4 = 0.027701$.

Estimation brawn bear's population size (4/4)

Results for population size of the brown bear I Rhodope mounting

Fit areas bear habitats	Other land cover	coniferous forest	mixed forests	deciduous forest	Total
14 Forestry areas with transects ; Other forestry areas (Pazardjik and Plovdiv) without transects	4.75	20.08	19.35	12.43	56.61
Round to a whole number	**5**	**21**	**20**	**13**	**57/59**
all fit areas bear habitats in Smolyan and Kardjalii	9.07	22.13	26.09	12.40	69.69
Round to a whole number	**10**	**23**	**27**	**13**	**70/73**
Total size population	**15**	**44**	**47**	**26**	**127/132**

Statistical error is about 8%; Max population size should be; 137 - 143

Conclusion and future work

HP-SEE
High-Performance Computing Infrastructure
for South East Europe's Research Communities

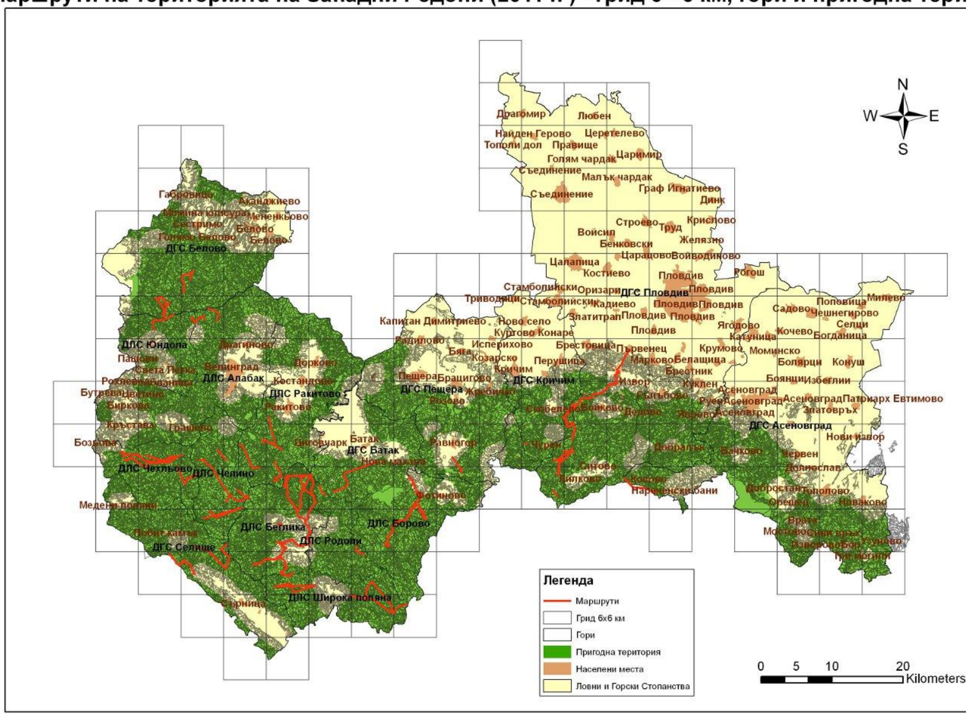
Coarse statistical estimation for population size of the brown bear is obtained in Rhodope using data from the National monitoring, 25-26 Oct 2011.

- **Recommendations:**
- The transects should be optimized;
- Repeatability of crawling routes along with the inclusion of new ones.

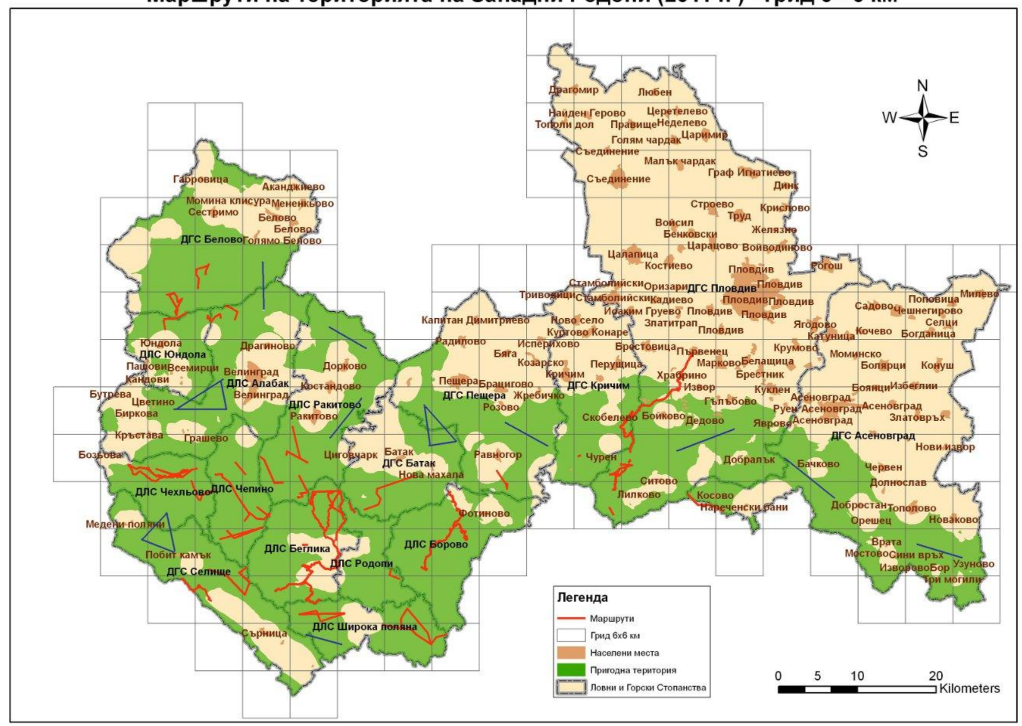
- **Future work:**
- Estimate population size in the whole country;
- Estimate the trend of development / growth or reduce / a brown bear population in the country, using the actual data from several consistent monitoring
- Our ambition is to create a program product for solving the above problems

Transects

Маршрути на територията на Западни Родопи (2011 г.) - грид 6 - 6 км, гори и пригодна тери



Маршрути на територията на Западни Родопи (2011 г.) - грид 6 - 6 км



Monte Carlo method

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

Let us compute the following multidimensional integral by MC method

$$I(f) = \int_{I^s} f(x)dx = \int_0^1 \cdots \int_0^1 f(x^{(1)}, \ldots, x^{(s)}) dx^{(1)} \ldots dx^{(s)}$$

$\xi = f(x_i)$, x_i is s –dimension uniform distributed vector belong to I^s
The domain I^s can be different from s – dimensional cube

MC method:

$$\xi : \quad E\xi = I(f) \qquad \xi^{(1)}, \ldots, \xi^{(N)}$$

$$\bar{\xi} = \frac{1}{N} \sum_{i=1}^{N} \xi^{(i)} \xrightarrow{P} I(f) \qquad \textbf{(1)}$$

The "law of three sigmas" gives the rate of convergence:

$$P\left(|\bar{\xi} - I(f)| < 3\frac{\sqrt{Var(\xi)}}{\sqrt{N}} \right) \approx 0.997$$

$O(N^{-1/2})$.

Variance: $Var(\xi) = E\xi^2 - E^2\xi$

Modeling function (sampling rule) $F(\beta_1, \beta_2, \ldots,)$ **(2)**

$$\beta_1, \beta_2, \ldots, \in (0,\ 1)$$

MC algorithm – (1) and (2)

Reconstruction of densities

- Consider the inverse problem:

to estimate an unknown density with given realizations of the random variable

- To solve the problem we use:
  - approximation with B – spline,
  - Least square method
  - Monte Carlo for solving the integrals

Consider *p(x) defines in (a ,b) with given N realization of the r.v.* $\{\xi_i\}_{i=1}^{N} \in [a, b]$
Consider the grid of points with mesh h=(b-a)/m,

$$\omega_m = \{a = x_0 < x_1 < \ldots < x_m = b\}$$

*and add 2*k new points:*

$$T_n = \{t_1 < t_2 < \ldots < t_{k+1} = x_0 < \ldots < x_m = t_{n-k} < t_{n-k+1} < \ldots < t_n\},$$

$$n = 2k + m + 1.$$

Approximation with B-spline

- *Consider the following approximation of the unknown density with B-spline and an error $O(h^k)$*

$$p(x) = \sum_{i=1}^{L} c_i B_{i,k}(x), \quad x \in [a,b], \quad L = n-k-1$$

- *and approximation coefficients c_i , i-th B-spline of degree k is defined as* divided difference of cut speed function, x is fixed.

$$B_{i,k}(x) = (\cdot - x)_+^k [t_i, \ldots, t_{i+k+1}], \quad i = 1, \ldots, L,$$

$$(t-x)_+^k = \begin{cases} (t-x)^k, & t > x \\ 0, & t \le x. \end{cases}$$

$$B_{i,k}(x) = \sum_{s=i}^{i+k+1} \frac{(t_s - x)_+^k}{\omega'_{i,k}(t_s)},$$

$$\omega_{i,k}(t) = (t-t_i)\ldots(t-t_{i+k+1}).$$

Least Square method

- *To find the coefficients c_i use the Lest Square method*

$$U = \int_a^b \left(p(x) - \sum_{i=1}^{L} c_i B_{i,k}(x) \right)^2 \mathrm{d}x$$

$$U = U(c_1, \ldots, c_L) = \int_a^b \left(p(x) - \varphi(x; c_1, \ldots, c_L)\right)^2 \mathrm{d}x$$

$$U = \int_a^b p^2(x)\,\mathrm{d}x - 2\sum_{i=1}^{L} c_i(p, B_{i,k}(x)) + \int_a^b \left(\sum_{i=1}^{L} c_i B_{i,k}(x) \right)^2 \mathrm{d}x.$$

$$\frac{\partial U}{\partial c_i} = -2(p, B_{i,k}(x)) + 2\int_a^b \left(\sum_{j=1}^{L} c_j B_{j,k}(x) \right) B_{i,k}(x)\,\mathrm{d}x = 0$$

$$\sum_{j=1}^{L} (B_{i,k}(x), B_{j,k}(x)) c_j = (p(x), B_{i,k}(x)) \quad i = 1, \ldots, n-k-1,$$

Monte Carlo method

- *The right side we solve with Monte Carlo:*

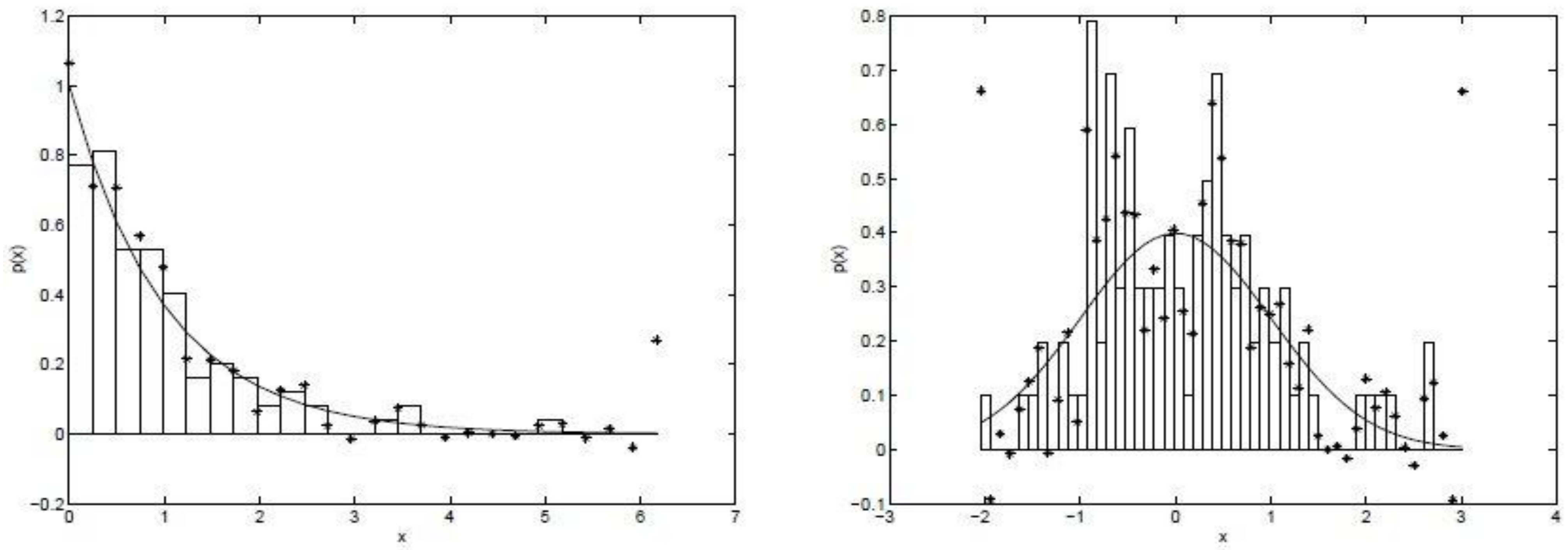
$$(f, g) = \int_a^b f(x) g(x) \, \mathrm{d}x.$$

$$(p(x), B_{i,k}(x)) = \int_a^b p(x) B_{i,k}(x) \, \mathrm{d}x = \mathrm{E} B_{i,k}(\xi),$$

$$(p(x), B_{i,k}(x)) \approx \frac{1}{N} \sum_{j=1}^{N} B(\xi_j) = \bar{\xi} = \frac{1}{N} \sum_{i=1}^{N} \xi^{(i)}$$

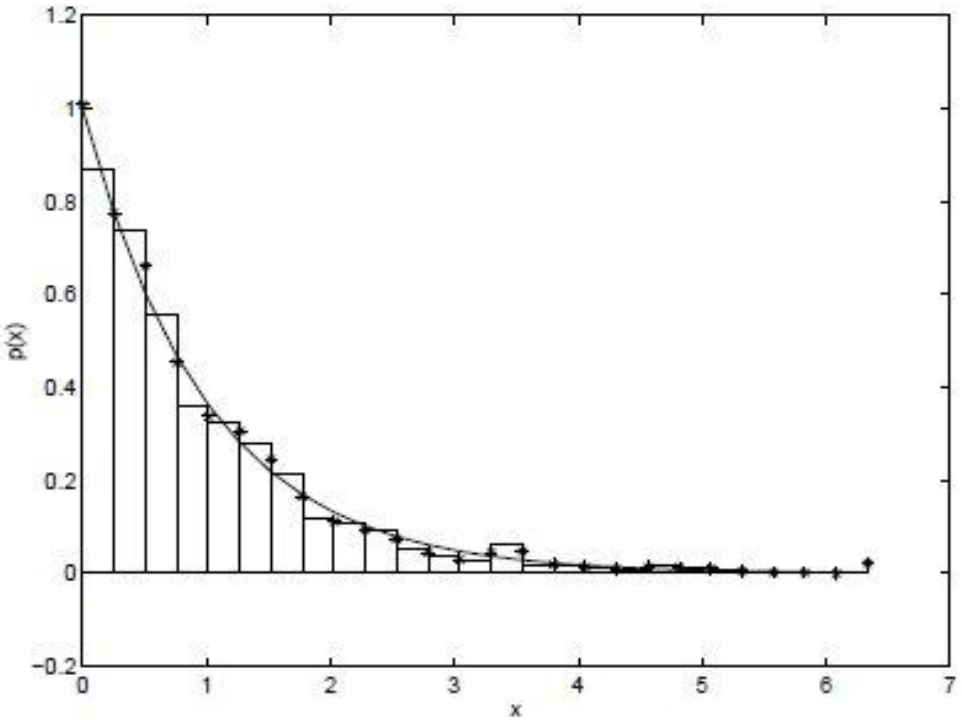
- *Where* $\{\xi_i\}_{i=1}^{N} \in [a, b]$ *is given saple of the r.v.*

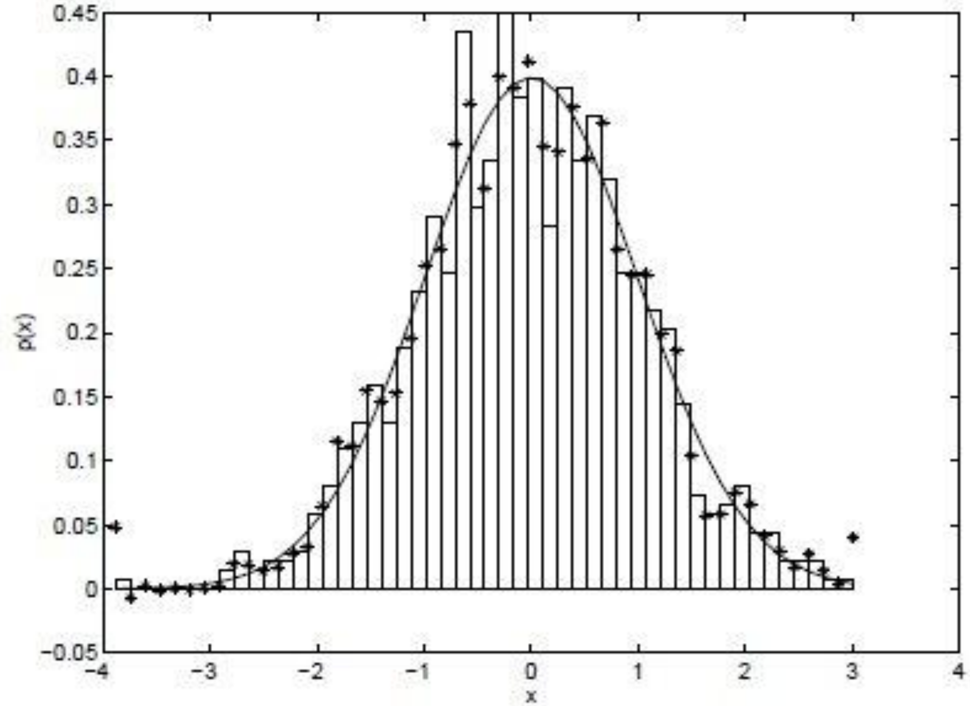
Numerical Results



- Estimation of exponential and normal distribution densities with the MC method and the method of histograms
- **Sample size: N =100**
- Number of intervals: 25 (exp. distribution); 50 (normal distribution)

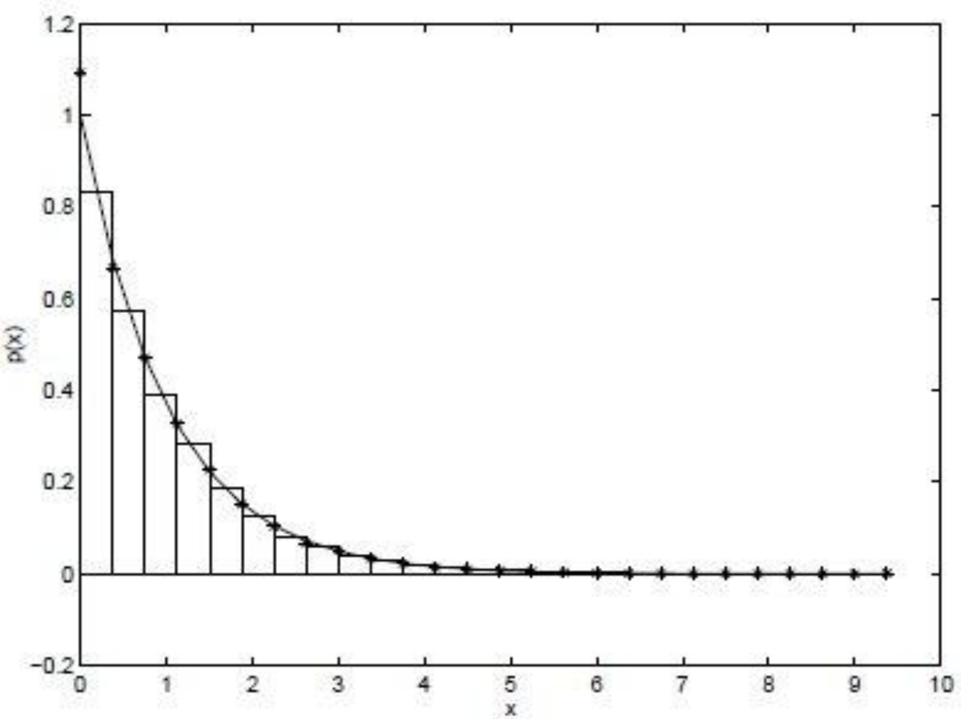
Numerical Results

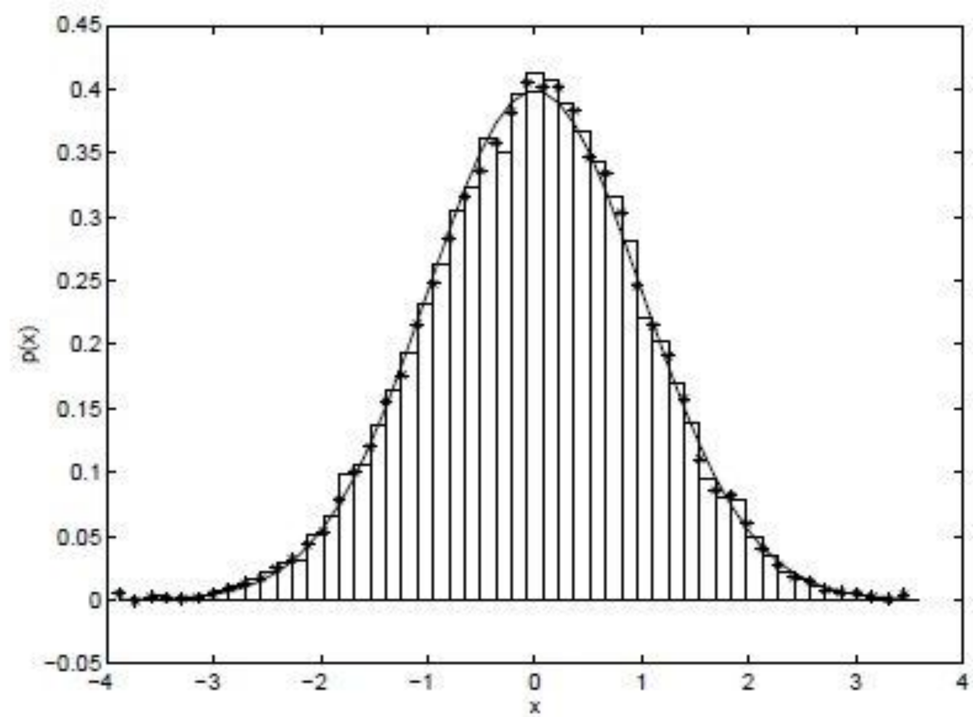




- Estimation of exponential and normal distribution densities with the MC method and the method of histograms
- **Sample size: N = 1000**
- Number of intervals: 25 (exp. distribution); 50 (normal distribution)

Numerical Results





- Estimation of exponential and normal distribution densities with the MC method and the method of histograms
- **Sample size: N = 10000**
- Number of intervals: 25 (exp. distribution); 50 (normal distribution)

Conclusion

Approximation of the unknown density function is done using B-spline, LS method and MC method.

Q&R