On scoring rules

Ksenia N. Kyzyurova
kseniak.ucoz.net, ksenia.kyzyurova@gmail.com

December 7, 2019

Abstract

The subjectivity of the scoring rules have been demonstrated in (Kyzyurova 2019). For brevity the focus has been on the logarithmic score and frequency predictive measures. Here we give additional illustrative examples of other scoring rules. The inadequacy of the scores is demonstrated.

Scoring rules

Consider the following scoring rules Gneiting & Raftery (2007) for a normal predictive distribution p with mean μ and standard deviation σ and evaluated with respect to the true value x.

Quadratic score

$$\mathrm{QS}(p,x) = 2p(x) - \frac{1}{2\sqrt{\pi}\sigma}\,. \tag{1}$$

Spherical score

$$\mathrm{sphS}(p,x) = p(x)\sqrt{2\sqrt{\pi}\sigma}\,. \tag{2}$$

Continuous ranked probability score

$$\mathrm{CRPS}(p,x) = \sigma\left(\frac{1}{\sqrt{\pi}} - 2\phi\left(\frac{x-\mu}{\sigma}\right) - \left(\frac{x-\mu}{\sigma}\right)\left(2\Phi\left(\frac{x-\mu}{\sigma}\right) - 1\right)\right)\,, \tag{3}$$

where ϕ and Φ denote the probability density function and the cumulative distribution function of a standard Gaussian variable.

Linear score

$$\mathrm{linS}(p,x) = p(x)\,. \tag{4}$$

Probability score

$$\mathrm{PS}(p,x) = \int_{x-1}^{x+1} p(y)dy\,. \tag{5}$$

Predictive model choice criterion

$$\mathrm{PMCC}(p,x) = -(x-\mu)^2 - \sigma^2\,. \tag{6}$$

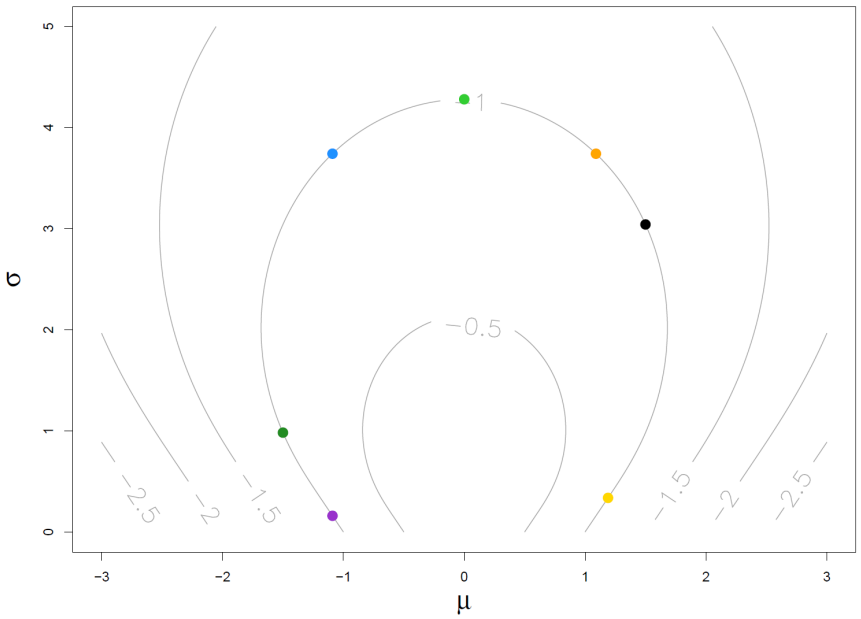

Figure 1: Contourlines of CRPS score.

Example

Consider the continuous ranked probability scoring rule for a normal predictive distribution p with mean μ and standard deviation σ. With respect to the true value $x = 0$. The contourlines of the score are given in Figure 1. Seven coloured points which belong to the same contourline of CRPS score equal to -1 correspond to seven distributions whose densities with corresponding colors are shown in Figure 2. We see that the shapes of the distributions are very different. Numerical evaluation of seven predictive distributions with seven scores are summarized in the tables.

Different scores lead to different representation and understanding of results of comparison of the models using these scores. For example, seven models are found to be equivalent using CRPS score. At the same time, the second distribution is preferred by logarithmic score, linear score and probability score. Model with the 6th predictive distribution is preferred by spherical and quadratic scores. The first model is preferred by PMCC score. The fact that such different distributions are assigned the same value of a score makes us question the adequacy of a chosen scoring rule.

Appendix: contourlines of other scores

The plots of contourlines of the scores other than CRPS are given for illustrative purposes. All figures support the conclusion on subjectivity of the scores and inadequcy of their behaviour.

References

Gneiting, T. & Raftery, A. E. (2007), 'Strictly proper scoring rules, prediction, and estimation', *Journal of the American Statistical Association* **102**(477), 359–378.

Kyzyurova, K. N. (2019), *Analysis of scientific computer models. Methodology in numerical simulator data analysis.*

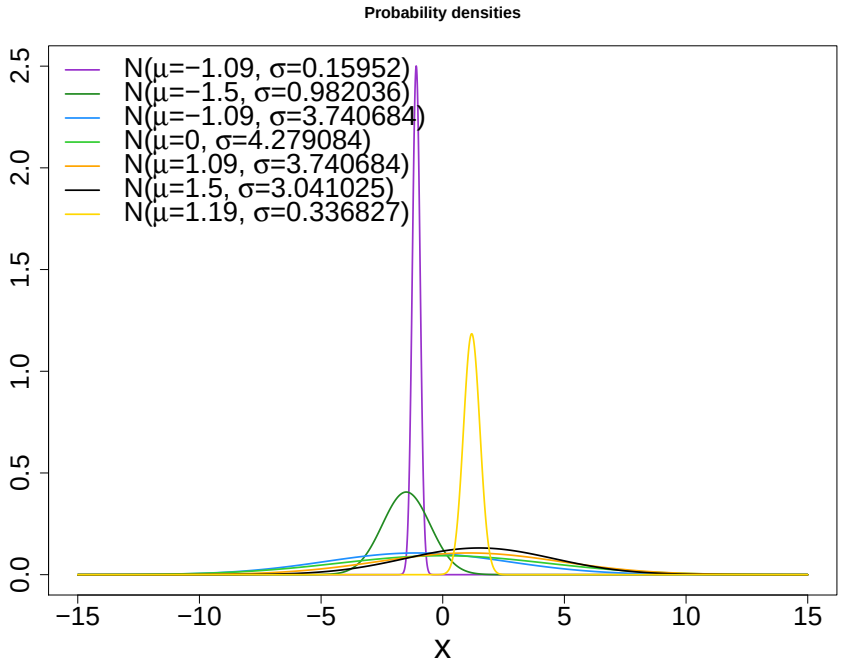

Figure 2: Forecasts from seven models.

Table 1: Results from evaluation of the scores for all the predictive distributions among considered. CRPS score assigns the same value of -1 to each model. Other scores have different preferences in choosing their best model.

Distr./score	CRPS	Log. score	Sph. score	Quad. score	Prob. score	Lin. score	PMCC score
1	-1	-22.428	0.000	-1.768	0.286	0.000	**-1.214**
2	-1	**-2.067**	0.236	-0.034	**0.300**	**0.127**	-3.214
3	-1	-2.281	0.372	0.129	0.202	0.102	-15.181
4	-1	-2.373	0.363	0.121	0.185	0.093	-18.311
5	-1	-2.281	0.372	0.129	0.202	0.102	-15.181
6	-1	-2.153	**0.381**	**0.140**	0.229	0.116	-11.498
7	-1	-6.072	0.003	0.832	0.286	0.002	-1.530

Table 2: Summary of which model contains the true value in the 95%-CIs.

	1	2	3	4	5	6	7
$x = 0 \in CI_{95\%}$	No	Yes	Yes	Yes	Yes	Yes	No

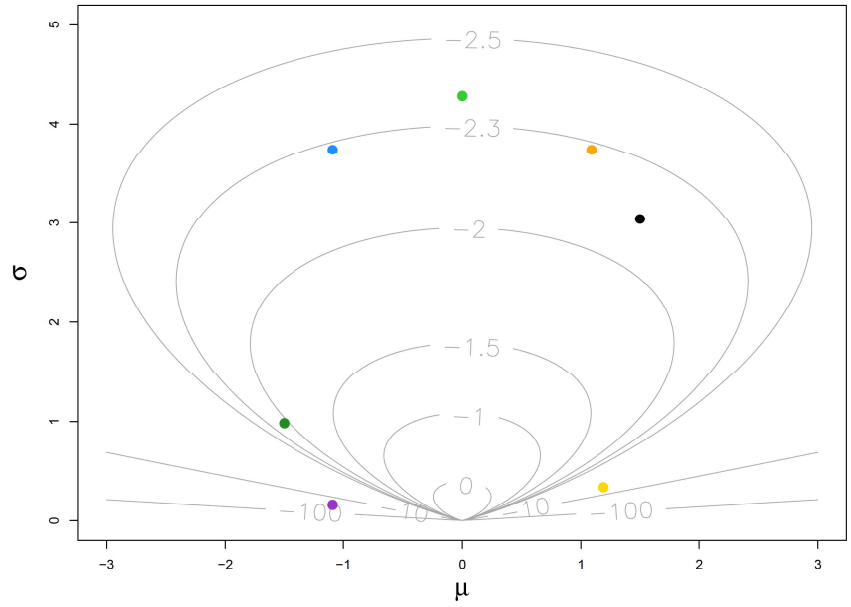

Figure 3: Contourlines of the log-score.

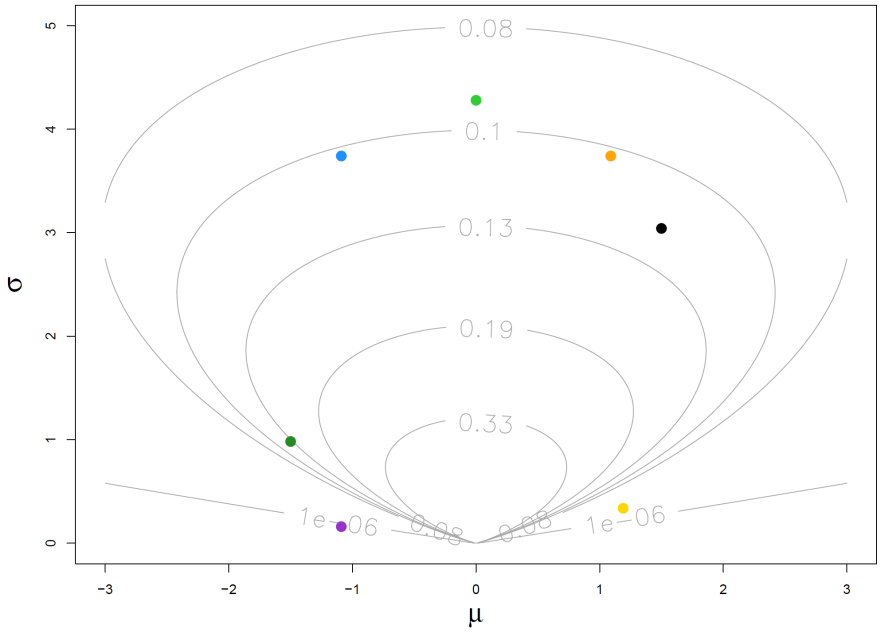

Figure 4: Contourlines of the linear score for normal distribution parameterized by mean μ and standard deviation σ.

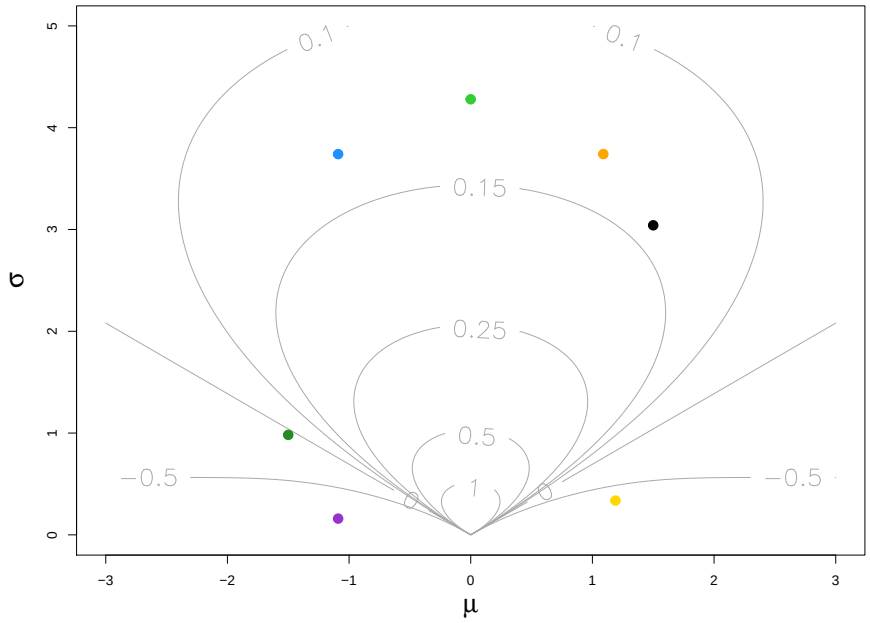

Figure 5: Contourlines of the quadratic score for normal distribution parameterized by mean μ and standard deviation σ.

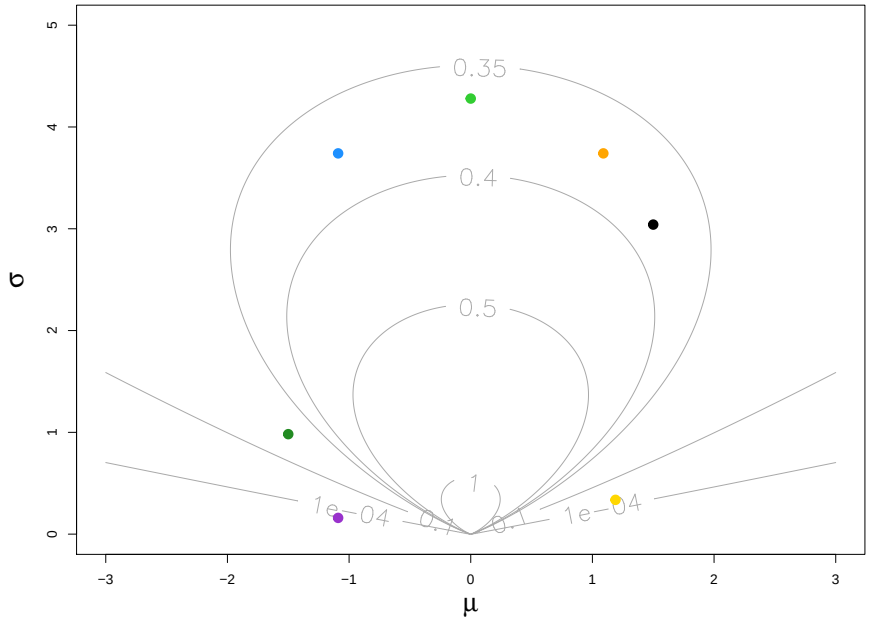

Figure 6: Contourlines of the spherical score for normal distribution parameterized by mean μ and standard deviation σ.

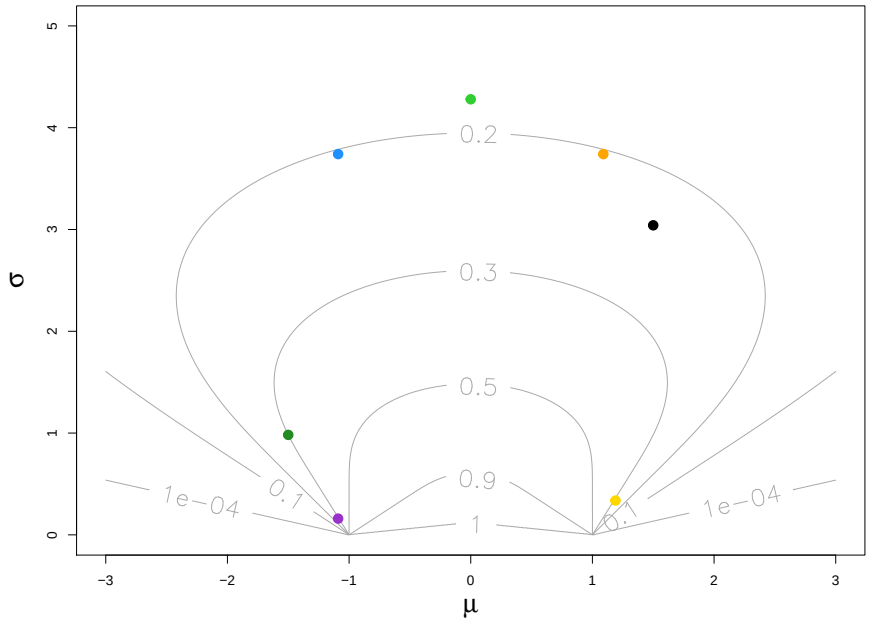

Figure 7: Contourlines of the probability score for normal distribution parameterized by mean μ and standard deviation σ.

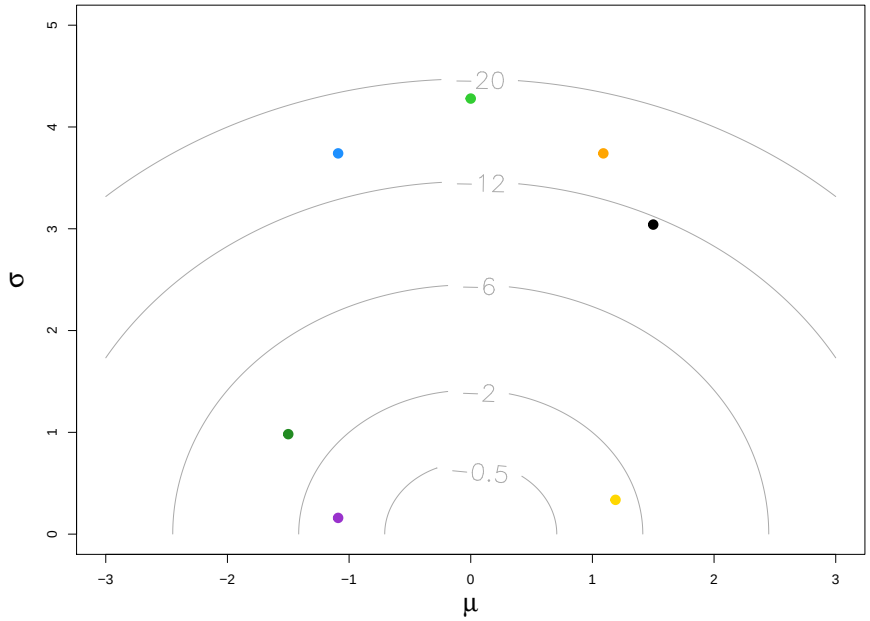

Figure 8: Contourlines of the PMCC score for normal distribution parameterized by mean μ and standard deviation σ.