A new R package for evaluation of probabilistic forecasts based on locally scale invariant proper scoring rules

Abdulaziz Alghanem[1], David Bolin[2], Alexandre Simas[2]

1. Mathematics Department, King Fahd University of Petroleum and Minerals

2. CEMSE, King Abdullah University of Science and Technology

Abstract

It is common to make predictions in the form of probabilistic forecasts in areas such as weather forecasting. Proper scoring rules are the standard way to evaluate these predictions. We implement a newly proposed proper scoring rule, which has some desirable properties, in an R package. We then use it to compare different fitted models with real data.

Introduction

- It is of interest to make predictions about the future based on observed data.
- Probabilistic forecasts are preferred over point forecasts since they capture the uncertainty in the prediction.
- Scoring rules are functions that assign a score based on the predictive distribution and the observed value.
- Proper scoring rules are scoring rules whose expected value is maximized when the observations come from the predictive distribution.
- The logarithmic score and the continuous ranked probability score (CRPS) are commonly used proper scoring rules.
- $\mathrm{CRPS}(\mathbb{P}, y) = \frac{1}{2}\mathbb{E}_{\mathbb{P},\mathbb{P}}|X - X'| - \mathbb{E}_{\mathbb{P}}|X - y|$.
- $\mathrm{LogS}(\mathbb{P}, y) = \log f(x)$, where f is the PDF of the predictive distribution $\mathbb{P}$.

The new score

- The scaled CRPS (SCRPS) is a newly proposed proper scoring rule[1].
- $\mathrm{SCRPS}(\mathbb{P}, y) = -\frac{\mathbb{E}_{\mathbb{P}}|X-y|}{\mathbb{E}_{\mathbb{P},\mathbb{P}}|X-X'|} - \frac{1}{2}\log\left(\mathbb{E}_{\mathbb{P},\mathbb{P}}|X - X'|\right).$ $\quad (\star)$

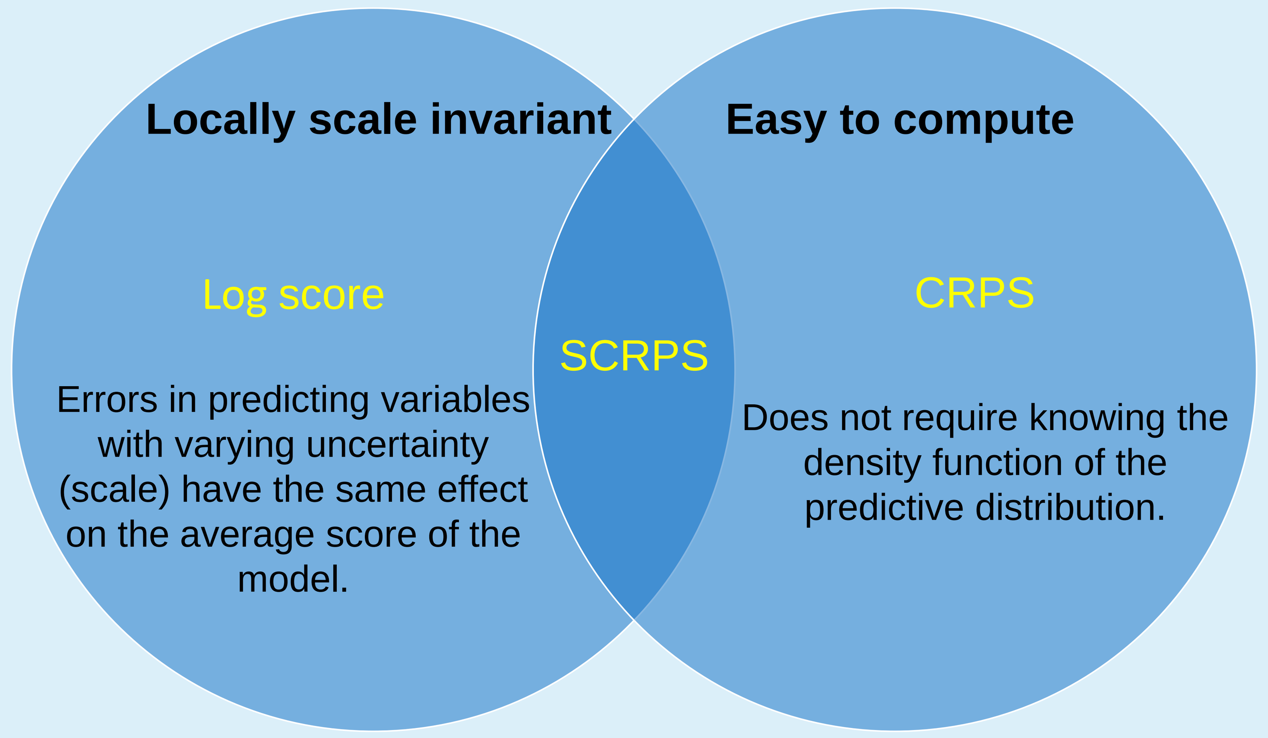

Figure 1 Venn diagram showing how the SCRPS combines properties of known proper scoring rules.

Local scale invariance vs scale dependence

Let $Y_1 \sim \mathrm{N}(0, \sigma_1{}^2)$ and $Y_2 \sim \mathrm{N}(0, \sigma_2{}^2)$, where $\sigma_1 = 0.1$ and $\sigma_2 = 1$. Suppose we have a model with predictive distributions $\mathbb{P}_1 = \mathrm{N}(0, \hat{\sigma}_1{}^2)$ and $\mathbb{P}_2 = \mathrm{N}(0, \hat{\sigma}_2{}^2)$ for Y_1 and Y_2, respectively. The average score for the model given observations y_1, y_2 of Y_1, Y_2, respectively, and using the scoring rule S is given by $\frac{1}{2}\mathrm{S}(\mathbb{P}_1, y_1) + \frac{1}{2}\mathrm{S}(\mathbb{P}_2, y_2)$. Now consider the expected value of the average score and denote it by $\mathrm{S}(\mathbb{P}, Y)$. **Figure 2** shows the $\mathrm{CRPS}(\mathbb{P}, Y)$ and $\mathrm{SCRPS}(\mathbb{P}, Y)$ when $\hat{\sigma}_1$and $\hat{\sigma}_2$ vary. Notice how the CRPS is more sensitive to changes in $\hat{\sigma}_2$ than in $\hat{\sigma}_1$. This means that by using the CRPS, a model that predicts Y_2 very well and makes a big error in predicting Y_1 will be preferred over a model that predicts Y_1 accurately and makes a slight error in predicting Y_2. On the other hand, the SCRPS gives equal importance to errors in predicting both variables which gives a more intuitive ranking.

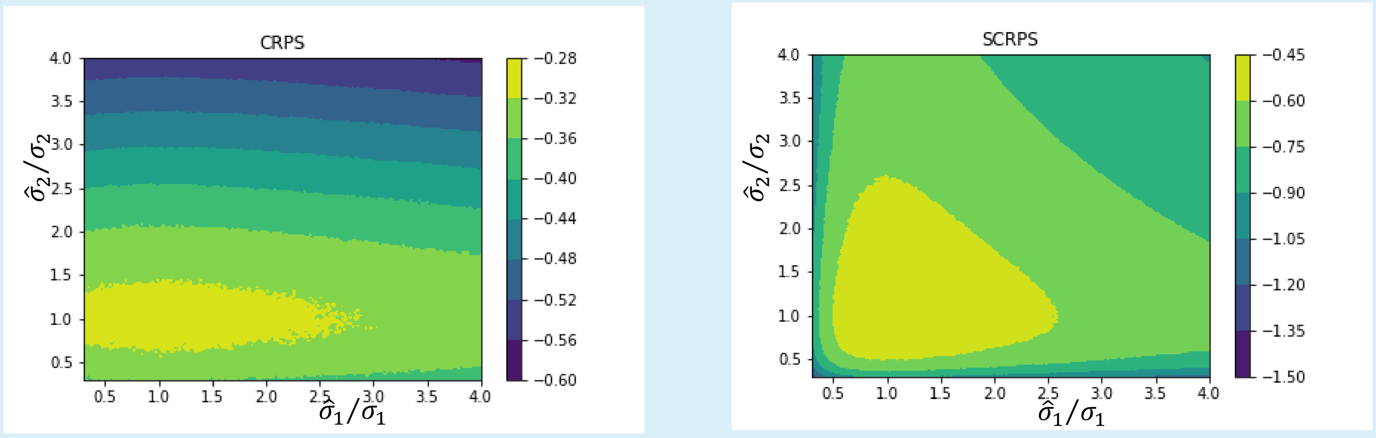

Figure 2 Average expected CRPS and SCRPS as $\hat{\sigma}_1$ and $\hat{\sigma}_2$ vary.

The R package

- We implemented two main functionalities in a new R package that we called SCRPS(Scan QR code for package page).
- The first computes the SCRPS for many known distributions. This is implemented through analytic expressions derived using $(\star)$.
- The supported distributions include the Laplace, the logistic, the normal, the Student's t, the exponential, the gamma, the log-normal, and many other distributions.
- The other functionality allows the computation of the average score for models fitted with INLA. This extends the use of the package significantly.
- Here are two simple examples showing how to use the package.

```r
library(SCRPS) #SCRPS_d(y, parameters) gives SCRPS for distribution d and observation y
y <- rgamma(100000, shape = 2.6, scale = 1.2) #sample from gamma distribution
mean(SCRPS_gamma(y,shape=2.5,scale=1.5)) #compute average score
```

```
## [1] -1.395545
```

```r
x <- rnorm(1000, mean = 0, sd = 1)
y <- rgamma(1000, shape = 3, scale = 2)
df = data.frame(x,y) #generate random data and fir simple inla model
result_y <- inla(y ~ 1 + x, data = df, control.compute=list(return.marginals.predictor=TRUE))
library(SCRPS) #SCRPS_inla(result) gives the SCRPS for the inla fitted model stored in result
SCRPS_inla(result_y)# get the SCRPS
```

```
## [1] -1.662863
```

Implementation of the SCRPS for INLA models

- Consider a model fitted by INLA where $y_i | \mu_i, \theta \sim f(\mu_i; \theta)$ for $i = 1, 2, \ldots, n$. Then the average SCRPS for the model is $\frac{1}{n}\sum_{i=1}^{n} \mathrm{SCRPS}\,(f(\mu_i; \theta),\ y_i)$.
- To evaluate the $\mathrm{SCRPS}(f(\mu_i; \theta), y_i)$ for each i we start by generating m samples of each of μ_i and θ denoted by μ_i^j and θ^j for $j = 1, 2, \ldots m$. Then we approximate $\mathrm{SCRPS}(f(\mu_i; \theta), y_i)$ by $\frac{1}{m}\sum_{j=1}^{m} \mathrm{SCRPS}(f(\mu_i^j; \theta^j), y_i)$. Finally, We use the analytic expression (for likelihoods that are supported) to evaluate the term inside the last sum.
- The supported likelihoods include the gamma, normal, logistic, lognormal, Poisson, binomial, and negative binomial distributions.
- In addition, the package is continuously updated to support more likelihoods.
- We have also added a similar function that computes the CRPS using the same method.

Precipitation in Paraná: data set and model fitting

- The data set contains the amount of daily rain precipitation from January of year 2011 at 616 locations in Paraná state in Brazil (see **Figure 3** and **Figure 4**).
- It seems like locations near the Atlantic Ocean tend to get higher levels of precipitation and hence the distance to the ocean is considered as a new covariate.

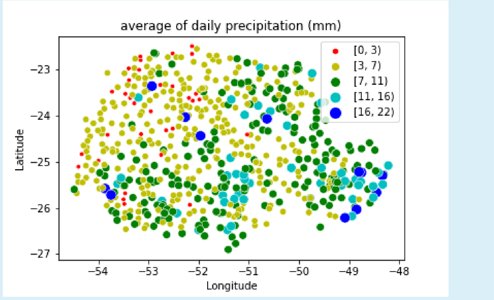

Figure 3 Average of daily accumulated precipitation (mm) during January 2011 in 616 locations in Paraná state.

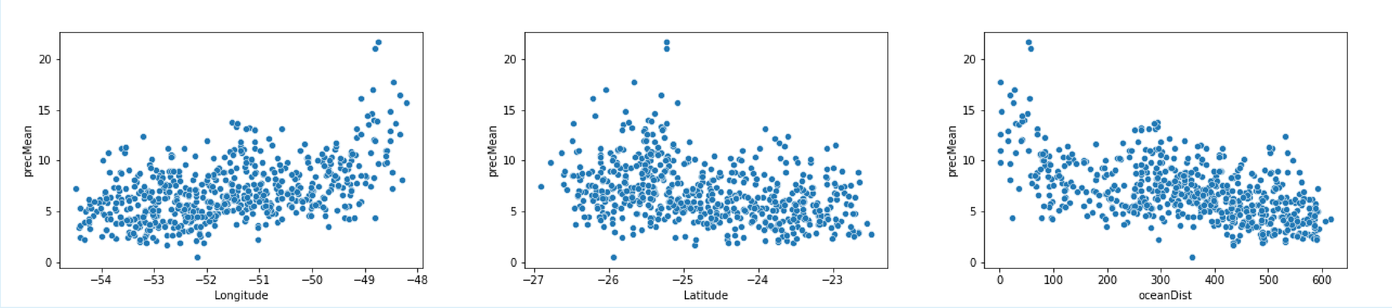

Figure 4 Scatter plots for the average daily accumulated precipitation (mm) against each of the longitude, the latitude, and the distance to the Atlantic Ocean.

- Looking at **Figure 4**, it looks like the average precipitation has a relationship with both the longitude and the distance to the ocean but not with the latitude.
- Using INLA we fit four models with a gamma likelihood and a spatial random effect.
- The first three models ($\mathrm{model}_{\mathrm{Long}}$, $\mathrm{model}_{\mathrm{Lat}}$, $\mathrm{model}_{\mathrm{OD}}$) use a random walk of order 1 (rw1) model on the longitude, the latitude and the distance to the ocean, respectively.
- The fourth model ($\mathrm{model}_{\mathrm{ODL}}$) takes the distance to the ocean as a covariate with linear effect.
- For more details about the models that were used, look at the second reference.

Precipitation in Paraná: evaluation of models

- Both scoring rules give the same ranking to the models as shown in **Table 1**.
- The model that takes the latitude as a covariate got the worst score as expected.
- The models that took the distance to the ocean as a covariate got slightly better scores than the one with longitude as a covariate. This might suggest that the relation between the precipitation and the longitude is just the result of the coast being on the east side of Paraná.
- The best model was the one that took the distance to the ocean as a linear effect. This is reasonable as figure 4 shows that the relation between the average precipitation and the distance to the ocean is almost linear.
- More models can be fitted on this data set and compared to these models. For instance, a model that takes the altitude as a covariate or perhaps a model that takes multiple covariates at once.
- Cross-validation can be used to give a more accurate assessment of the predictive performance of these models.

Model	Average CRPS	Average SCRPS	Rank
$\mathrm{model}_{\mathrm{Long}}$	-0.9798	-1.3196	3
$\mathrm{model}_{\mathrm{Lat}}$	-1.0058	-1.3328	4
$\mathrm{model}_{\mathrm{OD}}$	-0.9656	-1.3119	2
$\mathrm{model}_{\mathrm{ODL}}$	-0.9584	-1.3075	1

Table 1 Average CRPS and SCRPS for the four models. Scores are positively oriented (i.e., higher score is better).

References

1. Bolin, D., & Wallin, J. (2019). Local scale invariance and robustness of proper scoring rules. *Statistical Science* (in press).
2. E. T. Krainski et al. (2019). Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA. Chapman & Hall/CRC Press. Boca Raton, FL.
3. Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. J. Amer. Statist. Assoc. 102 (477), 359–378.
4. Jordan A, Krüger F, Lerch S: Evaluating probabilistic forecasts with scoringRules. *J Stat Softw.* 2019; **90**(12).
5. Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, *71*(2), 319-392.