School of Mathematical
and Physical Sciences

# Department of Mathematics and Statistics

Preprint MPS-2011-15

01 November 2011

# Contrasting Probabilistic Scoring Rules

by

Reason L. Machete

# Contrasting Probabilistic Scoring Rules

Reason L. Machete[*]

Dept. of Mathematics and Statistics, P. O. Box 220, Reading, RG6 6AX, UK

Dated: November 01, 2011

**Abstract**

There are several scoring rules that one can choose from in order to score probabilistic forecasting models or estimate model parameters. Whilst it is generally agreed that proper scoring rules are preferable, there is no clear criterion for preferring one proper scoring rule above another. This manuscript contrasts properties of some commonly used proper scoring rules and provides incremental guidance on scoring rule selection. In particular, it is shown that the *logarithmic scoring rule* prefers erring on the side of caution, but the *Continuous Ranked Probability Score* tends to prefer over-confident forecasts.

## 1 Introduction

Issuing probabilistic forecasts is meant to express uncertainty about the future evolution of a dynamical system. The quality of probabilistic forecasts may be undermined by model mis-specification. This makes it necessary to assess forecast-quality. Forecast-quality can be assessed using either probabi.4.44234(a55985(sa)5(k)32.7)4.44k4ab bteoral  the bng

This paper presents a novel theoretical analysis of strictly proper scoring rules. It focuses upon those scoring rules that are commonly used in the forecasting literature, including econometrics and meteorology. It unravels and contrasts how the di erent scoring rules would rank competing forecasts of specified departures from ideal forecasts and provides incremental guidance on scoring

## 2.2 Logarithmic scoring rule

The logarithmic scoring rule was proposed by Good (1952). It was later termed Ignorance by Roulston & Smith (2002) when they introduced it to the meteorological community. Given a probabilistic forecast $\boldsymbol{f} = ( {}_1 \quad {}_2 \qquad {}_m)$, the Ignorance score is given by $(\boldsymbol{f} \quad ) = -\log \; {}_{j_!} \; = ($ i-5.44495(gd[47[

**Proposition 2** *Given two forecasts* $\boldsymbol{f}_i = (p + \ \ _i\ q - \ \ _i)$, $\ = 1\ 2$ *with* $0\ \ \ _1\ \ \ _2\ \ \ q$ *and* $p\ \ \ q$, *the logarithmic scoring rule prefers* $\boldsymbol{f}_1$ *over* $\boldsymbol{f}_2$, *in agreement with the Brier score.*

**Proof**: In order to prove this proposition, it is su cient to consider the expected logarithmic score of the forecast $\boldsymbol{f} = (p + \ \ q - \ \ )$, which is given by equation (1). Di erentiating the equation with respect to yields

$$\frac{\mathrm{d}}{\mathrm{d}\ }\mathbb{E}[\ \ ] = \frac{}{(p + \ )(q - \ )} \tag{5}$$

Equation (5) implies that, if $q\ \ \ \ \ 0$, $\mathbb{E}[\ \ ]$ is an increasing function of . Hence, the logarithmic scoring rule prefers the forecast $\boldsymbol{f}_1$, in agreement with the Brier score.

On the other hand, if $\ \ 0$ with $|\ |\ \ p$, then equation (5) implies that $\mathbb{E}[\ \ ]$ is a decreasing function of . It then follow that, given $\ _2\ \ \ _1\ \ 0$ with $|\ _2|\ \ p$, the logarithmic scoring rule will prefer the forecast $\boldsymbol{f}_1$, again in agreement with the Brier score.

Finally, let

**Proposition 4** *For positive $\epsilon_1$ and $\epsilon_2$ such that $\epsilon_1 < q < p$ and $\epsilon_2 < p$, the entropy of the forecast $\boldsymbol{f}_1 = (p + \epsilon_1, q - \epsilon_1)$ is lower than that of the forecast $\boldsymbol{f}_2 = (p - \epsilon_2, q + \epsilon_2)$ whenever $\epsilon_2 \leq (p - q)/2$.*

A consequence of this proposition is that the forecast corresponding to $\epsilon_1 = \epsilon^*$ is more informative than $\boldsymbol{f}_2$ provided $\epsilon_2 \leq (p - q)/2$. Otherwise, either forecast could be more informative than the other. We now give the proof of this proposition.

**Proof**: To prove the above proposition, we consider the derivative of equation (4):

$$\frac{\mathrm{d}H}{\mathrm{d}\epsilon} = -\log\left(\frac{p + \epsilon}{q - \epsilon}\right)$$

We then note that $\mathrm{d}H/\mathrm{d}\epsilon \leq 0$ provided that $(p - q) \geq -2\epsilon$. If $\epsilon \geq 0$, this inequality is trivially satisfied. On the other hand, if $\epsilon < 0$, then the inequality is satisfied provided $|\epsilon| \leq (p - q)/2$. If $\epsilon_2 \leq (p - q)/2$, then $H(\epsilon)$ is a strictly decreasing function for all $\epsilon \in [-\epsilon_2, \epsilon_2]$, which implies that $H(\epsilon_1) < H(\epsilon_2)$. If $\epsilon_2 > (p - q)/2$, then $H(\epsilon)$ is an increasing function for all $\epsilon \in (-\epsilon_2, -(p - q)/2)$ (provided $p < 3q$) and strictly decreasing function in $(-(p - q)/2, \epsilon_1)$, which implies that $H(-(p - q)/2) > \max\{H(\epsilon_1), H(-\epsilon_2)\}$. Hence, in this case, we cannot determine which of $H(\epsilon_1)$ and $H(-\epsilon_2)$ is lower.

## 3 Density Forecasts

This section considers scoring rules for for forecasts of continuous variables. It is in some sense a generalisation of the previous section. As before, we consider how each scoring rule would rank two competing predictive distributions of fairly good quality. In the case of the logarithmic scoring rule and the Continuous Ranked Probability Score, we consider errors of each predictive distribution, $f(x)$, from the target distribution, $p(x)$, that are odd functions, i.e. $\delta(x) = f(x) - p(x)$ with $\delta(-x) = -\delta(x)$.

### 3.1 The Quadratic scoring rule

A continuous counterpart of the Brier score is the quadratic score (Gneiting & Raftery, 2007), given by

$$S(f, X) = \|f\|_2^2 - 2f(X)$$

where $X$ is a random variable. Taking the expectation yields

$$\mathbb{E}[S(f, X)] = \|f - p\|_2^2 - \|p\|_2^2 \tag{9}$$

We can now write $f(x) = p(x) + \delta(x)$, where $\int \delta(x)\mathrm{d}x = 0$, and substitute it into (9) to obtain

$$\mathbb{E}[S(f, X)] = \|\delta\|_2^2 - \|p\|_2^2 \tag{10}$$

As was the case with the Brier score, the functions $\pm\delta(x)$ yield the same quadratic score. For any two forecasts, $f_i(x) = p(x) + \delta_i(x)$, $i = 1, 2$ with $\|\delta_1\|_2 < \|\delta_2\|_2$, the quadratic scoring rule would prefer $f_1(x)$.

## 3.2 The Logarithmic scoring rule

The expectation of the logarithmic (or Ignorance) score for this forecast is

$$\mathbb{E}[\;(\;X))] = -\int p(x)\log(p(x) + \;(x))\mathrm{d}x$$

Further more,

$$\mathbb{E}[\;]_{\pm} = \int p(x)\log\left(\frac{p(x) - \;(x)}{p(x) + \;(x)}\right)\mathrm{d}x \tag{11}$$

It is necessary that $|\;(x)| \leq p(x)$ for (11) to be well defined. Consider the case when $p(x) = p(-x)$. If, in addition, $\;(x)$ is an odd function, i.e. $\;(-x) = -\;(x)$, then equation (11) yields $\mathbb{E}[\;]_{\pm} = 0$

We now turn to the general case where $p(x)$ is not necessarily even and $\;(x)$ is not necessarily odd. If we let $\;(x)$ be a test function, then the functional derivative of $\mathbb{E}[\;]_{\pm}$, denoted by $\mathbb{E}[\;]_{\pm}$, satisfies the relation

$$\int \text{—}\mathbb{E}[\;]_{\pm} \;=\; \frac{\mathrm{d}}{\mathrm{d}}\int p(x)\log\left(\frac{p(x) - \;(x) - \;(x)}{p(x) + \;(x) + \;(x)}\right)\mathrm{d}x\Bigg|_{\varepsilon=0}$$

from which it follows that

$$\text{—}\mathbb{E}[\;]_{\pm} = \frac{-2p^2(x)}{p^2(x) - \;^2(x)} \tag{12}$$

Hence, the functional derivative is negative for all $\;(x)$ such that $|\;(x)|\;p(x)$, which implies that $\mathbb{E}[\;]_{\pm}$ is a decreasing functional of $\;(x\;$ x(

where we used $p(|x|) \leq p(x)$ to obtain the last inequality. To justify the use of this inequality, we need to show that the function

Plugging (15) into (15) yields

$$
\begin{aligned}
(\ ) &= -\int_{-\infty}^{\infty} [\log(p(x) + \ (x)) + 1] \ (x)\mathrm{d}x \\
&= -\int_{-\infty}^{0} [\log(p(x) + \ (x)) + 1] \ (x)\mathrm{d}x - \int_{0}^{\infty} [\log(p(x) + \ (x)) + 1] \ (x)\mathrm{d}x \\
&= -\int_{-\infty}^{0} [\log(p(x) + \ (x)) + 1] \ (x)\mathrm{d}x + \int_{0}^{-\infty} [\log(p(-x) + \ (-x)) + 1] \ (-x)\mathrm{d}x \\
&= -\int_{-\infty}^{0} [\log(p(x) + \ (x)) + 1] \ (x)\mathrm{d}x - \int_{-\infty}^{0} [\log(p(-x) + \ (-x)) + 1] \ (-x)\mathrm{d}x \\
&= -\int_{-\infty}^{0} [\log(p(x) + \ (x)) + 1] \ (x)\mathrm{d}x + \int_{-\infty}^{0} [\log(p(-x) - \ (x)) + 1] \ (x)\mathrm{d}x \\
&= -\int_{-\infty}^{0} \log\left(\frac{p(x) + \ (x)}{p(-x) - \ (x)}\right) \ (x)\mathrm{d}x
\end{aligned}
$$

where we have applied a change of variable $x \to -x$ in the second integral of the third line and assumed $(-x) = -\ (x)$ in the fifth line. In particular,

$$
(\ )|_{\gamma=0} = -\int_{-\infty}^{0} \log\left(\frac{p(x)}{p(-x)}\right) \ (x)\mathrm{d}x
$$

Using the assumption that $p(x) \geq p(-x)$ whenever $x$ 0, we consequently obtain

$$
(\ )|_{\gamma=0} \leq 0 \tag{16}
$$

if $(x)$ 0 for all $x$ 0. In e ect, we have just proved the following proposition:

**Proposition 6** *Given that* $(-x) = -\ (x)$, $\int (x)\mathrm{d}x = 0$, $(|x|) \leq 0$, $p(|x|) \leq p(x)$ *and* $|\ (x)|$ $p(x)$, *then the entropy of the forecast density* $_{+}(x) = p(x) + \ (x)$ *is lower than that of the fx)*

Using this result, we obtain the first variation of $\mathbb{E}[\;\;]$ as

$$
\begin{aligned}
\mathbb{E}[\;\;] &= \int_{-\infty}^{\infty} \frac{\mathbb{E}[\;\;]}{\;(x)} \;(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{-p(x)}{p(x) + \;(x)} \;(x)\mathrm{d}x \\
&= \int_{-\infty}^{0} \frac{-p(x)}{p(x) + \;(x)} \;(x)\mathrm{d}x + \int_{0}^{\infty} \frac{-p(x)}{p(x) + \;(x)} \;(x)\mathrm{d}x \\
&= \int_{-\infty}^{0} \frac{-p(x)}{p(x) + \;(x)} \;(x)\mathrm{d}x + \int_{0}^{-\infty} \frac{p(-x)}{p(-x) + \;(-x)} \;(-x)\mathrm{d}x \\
&= \int_{-\infty}^{0} \frac{-p(x)}{p(x) + \;(x)} \;(x)\mathrm{d}x + \int_{-\infty}^{0} \frac{p(-x)}{p(-x) - \;(x)} \;(x)\mathrm{d}x \\
&= \int_{-\infty}^{0} \left( \frac{p(-x)}{p(-x) - \;(x)} - \frac{p(x)}{p(x) + \;(x)} \right) \;(x)\mathrm{d}x \\
&\geq \int_{-\infty}^{0} \left( \frac{p(x)}{p(x) - \;(x)} - \frac{p(x)}{p(x) + \;(x)} \right) \;(x)\mathrm{d}x \\
&= \int_{-\infty}^{0} \frac{2p(x)\;(x)}{p^2(x) - \;^2(x)} \;(x)\mathrm{d}x \\
&\geq 0
\end{aligned}
$$

provided $\;(x) \;\; 0$ for $x \;\; 0$ and $\;(-x) = -\;(x)$, and this completes the proof.

We shall now consider two forecasts, $\;_1(x) = p(x) + \;_1(x)$ and $\;_2(x) = p(x) - \;_2(x)$ with $|\;_1(x)| \leq |\;_2(x)| \leq p(x)$. In this case, the quadratics scoring rule would prefer $\;_1(x)$ over $\;_2(x)$. In

Thinking of ${}_1(x)$ as fixed, the first variation of $\mathbf{H}(\cdot\ {}_2)$ with respect to ${}_2(x)$ is given by

$$
\begin{aligned}
\mathbf{H}(\cdot\ {}_2) &= \int_{-\infty}^{\infty} \frac{\mathbf{H}(\cdot\ {}_2)}{{}_2(x)}\ {}_2(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{-p(x)}{p(x) - {}_2(x)}\ {}_2(x)\mathrm{d}x \\
&= \int_{-\infty}^{0} \frac{-p(x)}{p(x) - {}_2(x)}\ {}_2(x)\mathrm{d}x + \int_{0}^{\infty} \frac{-p(x)}{p(x) - {}_2(x)}\ {}_2(x)\mathrm{d}x \\
&= \int_{-\infty}^{0} \frac{-p(x)}{p(x) - {}_2(x)}\ {}_2(x)\mathrm{d}x - \int_{0}^{-\infty} \frac{p(-x)}{p(-x) + {}_2(x)}\ {}_2(x)\mathrm{d}x \\
&= \int_{-\infty}^{0} \frac{-p(x)}{p(x) - {}_2(x)}\ {}_2(x)\mathrm{d}x + \int_{-\infty}^{0} \frac{p(-x)}{p(-x) + {}_2(x)}\ {}_2(x)\mathrm{d}x \\
&= \int_{-\infty}^{0} \left( \frac{p(-x)}{p(-x) + {}_2(x)} - \frac{p(x)}{p(x) - {}_2(x)} \right) {}_2(x)\mathrm{d}x \\
&\leq \int_{-\infty}^{0} \left( \frac{p(x)}{p(x) + {}_2(x)} - \frac{p(x)}{p(x) - {}_2(x)} \right) {}_2(x)\mathrm{d}x \\
&= \int_{-\infty}^{0} \frac{-2p(x)\ {}_2(x)}{p^2(x) - {}_2^2(x)}\ {}_2(x)\mathrm{d}x
\end{aligned}
$$

Hence, $\mathbf{H}(\cdot\ {}_2) \leq 0$ provided ${}_2$ $0$. It follows that $\mathbf{H}(\cdot\ {}_2)$ has a maximum when ${}_2 = 0$, i.e. $\mathbf{H}(\cdot\ {}_2) \leq \mathbf{H}(\cdot\ 0)$. In particular, $\mathbf{H}(0\ {}_2) \leq \mathbf{H}(0\ 0) = 0$. For ${}_2 \boxminus 0$, we have the strict inequality, $\mathbf{H}(0\ {}_2)$ $0$. But, $\mathbf{H}(\ {}_2\ {}_2)$ $0$. Therefore, continuity implies that $\mathbf{H}(\ {}_1\ {}_2) = 0$ for some ${}_1(x) = {}^*(x)$ such that $|\ {}^*|\ \ |\ {}_2|$, and this completes the proof.

### 3.3 Continuous Ranked Probability Score

Finally, we consider the *Continuous Ranked Probability Score* (CRPS) of the density forecast $(x)$ whose cumulative distribution is $F(x)$. The CRPS is a function of $F$ and the verification $X$ and is defined by (Gneiting & Raftery, 2007)

$$\mathrm{CRPS}(F\ X) = \int_{-\infty}^{\infty} (F(\ ) - \mathbb{I}\{\ \geq X\})^2 \mathrm{d} \tag{18}$$

Its associated entropy function is (Gneiting & Raftery, 2007)

$$(F) = \int_{-\infty}^{\infty} F(\ )(1 - F(\ ))\mathrm{d} \tag{19}$$

and its divergence function is

$$(\ F) = -\int_{-\infty}^{\infty} (\ (\ ) - F(\ ))^2 \mathrm{d} \tag{20}$$

where is the true cumulative distribution function, i.e. $(x) = \int_{-\infty}^{x} p(\ )\mathrm{d}$ . Again, we consider the case when $p(|x|) \leq p(x)$. Since $\mathbb{E}[\mathrm{CRPS}(F\ X)] = (F) - (\ F)$, it follows from (19) and (20) that

$$\mathbb{E}[\mathrm{CRPS}(F\ X)] = \int_{-\infty}^{\infty} F(\ )(1 - F(\ ))\mathrm{d} + \int_{-\infty}^{\infty} (\ (\ ) - F(\ ))^2 \mathrm{d} \tag{21}$$

Define $F_{\pm}(x) = (x) \pm \Gamma(x)$, where $\Gamma(x) = \int_{-\infty}^{x} (\ )\mathrm{d}$ . It can be shown that $(-x) = -(x)$ implies that $\Gamma(-x) = \Gamma(x)$. If we now define $\Delta(F_{\pm}) = (F_{+}) - (F_{-})$, then the following proposition holds:

**Proposition 9** *If* $(x) \geq 0$ *for every* $x$ $0$ *and* $(-x) = -$ $(x)$*, then the first variation of* $\Delta(F_{\pm})$ *is non-negative, i.e.*

$$\{\Delta(F_{\pm})\} \leq 0$$

**Proof**: The functional derivative of $\Delta(F_{\pm})$ with respect to $\Gamma(x)$ is

$$\frac{\{\Delta(F_{\pm})\}}{\Gamma(x)} = 2 - 4 \quad (x) \qquad = 2 \qquad \mathrm{P}(x)$$

)5.25024]3 10.9091 Tf 5.4 0 Td [(P)4]TJ /R63.88733]2

$= 2$

)5.25024]3 10.9091 Tf 5.4 0 Td [(P)4]TJ /R63.88733]2 8.51992 0 Td [(()3.88733]TJ /R33 10.9091 Tf 4.2 0 Td [(x)

rule prefers the higher entropy forecast. Preferring the higher entropy forecast may be thought of as taking a more cautious stance of less confidence. The logarithmic scoring rule selects a lower entropy forecast only if it is nearer to the target distribution in the sense of the $^2$ norm.

We extended the investigation from the binary forecasts to the continuous case, where we considered the Quadratic score, Logarithmic score and the Continuous Ranked Probability Score (CRPS). Just like the Brier score in the binary case, the Quadratic Score does not distinguish between forecasts with equal $^2$ norms of their error from the target distribution. Given two density forecasts whose errors from the target forecast di er by a sign, the logarithmic scoring rule prefers the distribution with higher entropy. On the other hand, the CRPS prefers the forecast distribution with lower entropy; bear in mind that lower entropy corresponds to more confidence (Shannon, 1948).

Our findings indicate that the logarithmic scoring rule encourages a more cautious decision when forecasts depart from the ideal forecast. This is in agreement with the idiom that we should "err on the side of caution." We consider this to be an advantage over the CRPS which ecourages erring on the side of risk. In an investment scenario, erring on the side of risk can result in substantial losses. Some have critised the logarithmic scoring rule for placing a heavy penalty on assigning zero probability to events that materialise (e.g. Boero et al., 2011; Gneiting & Raftery, 2007); but assigning zero probability to events that are possible is also discouraged by *Laplace's rule of succession* (Jaynes, 2003). The logarithmic scoring rule is good at highlighting misplaced confidence of forecasts. Such forecasts may have to be dealt with appropriately. One way of dealing with over-confident forecasts is to apply shrinkage estimators discussed in Casella (1985); Efron & Morris (1977).


## Acknowledgements

This work was supported by the RCUK Digital Economy Programme via EPSRC grant EP/G065802/1 The Horizon Digital Economy Hub.


## References


BOERO, G., SMITH, J. & WALLIS, K. F. (2011). Scoring rules and survey density forecasts. *International Journal of Forecasting* **27**, 379–393.

BRIER, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **78**, 1–3.

BROCKER, J. & SMITH, L. A. (2007). Scoring Probabilistic Forecasts: The importance of Being proper. *Weather and Forecasting* **22**, 382–388.

CASELLA, G. (1985). An Introduction to Empirical Bayes Data Analysis. *The American Statistician* **39**, 83–87.

CORRADI, V. & SWANSON, N. R. (2006). Predictive density evaluation. In *Handbook for Econometric Forecasting*, G. Elliott, C. W. J. Granger & A. Timmermann, eds., vol. 1. North-Holland, pp. 197–284.

DAWID, A. P. (1984). Present position and potential developments: Some Personal Views: Statistical Theory: The Prequencial Approach. *J. R. Statist. Soc. A* **147**, 278–292.

DIEBOLD, F. X., GUNTHER, T. A. & TAY, A. S. (1998). Evaluating density forecasts with application to financial risk management. *International Economics Review* **39**, 863–883.

EFRON, B. & MORRIS, C. (1977). Stein's Paradox in Statistics. [illegible]