Thorpex 2004

Scoring Probabilistic Forecasts

Jochen Bröcker

Devin Kilminster, Liam Clarke and Leonard Smith

Centre for the Analysis of Time Series
London School of Economics
Houghton Street
London WC2A 2AE

United Kingdom

cats.lse.ac.uk

- Why are probabilistic forecasts important?
- What is a skill score?
- What is a ***proper*** skill score?
- The importance of being proper
- Examples

Problem: We want to forecast an observable $T_{\mathbf{n}}$ (e.g. temperature), where n is the time.

- We issue probabilistic forecasts: $p_{\mathbf{n}}(T)$ = {Pobability of $T_{\mathbf{n}}$ = T}

- Usually $p_{\mathbf{n}}$ is built upon some related side information (past observations, weather model simulations)

- Does ***not*** mean $p_{\mathbf{n}}(t)$ the probability of $T_{\mathbf{n}}$ ***given*** that side information

Motivation

- End users don't want to ***know*** $T_{\mathbf{n}}$, they want to base ***decisions*** on $T_{\mathbf{n}}$ (and none of them care about models' 500mB height)

- To take reasonable action, the ***risk*** of taking that action must be factored into that decision

- To do that, information about the ***uncertainty*** of $T_{\mathbf{n}}$ must be known

We need a ***general skill score***, that takes into account the probabilistic character of the forecasts and that is relevant to many different users (incl. model developers, meteorologists).

- A skill score is a function $\mathcal{S}(p\ t)$

- The empirical skill is a sample mean:

$$S = \frac{1}{N}\sum_{n} \mathcal{S}(\underbrace{p_n}_{\text{Our forecast}}\ \underbrace{T_n}_{\text{Reality}})$$

What should skill scores actually measure?

Probabilistic forecasts

A good probabilistic forecast should have:

- Reliability – Looking at those days where a probability $p_n = r$ of rain is forecasted, a fraction r of them should have rain

- Sharpness – High probability is issued to events that acctually happen to occur

Skill scores should take this into account, since we ***believe*** that a probabilistic forecast having these properties is good for a ***multitude*** of speci c problems.

op co

Propriety is the key property for a skill score

- Assume

op co

Propriety is a property of the Skill Score alone, what the actual truth is doesn't matter

op co and a pn p y

Two statements for proper scores:

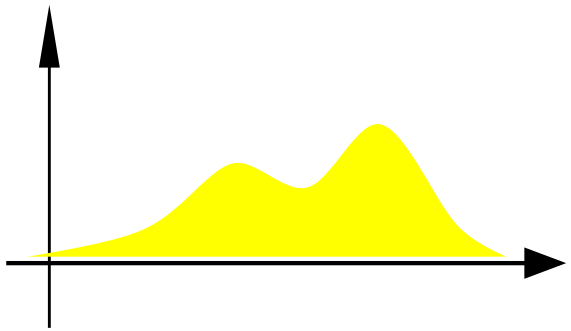
If, given the available information, a reliable forecasts exists, it would yield a maximum score

Of two equally reliable forecasts, the sharper one would score higher

An xa p a o

Bet on temperature at London Heathrow. Objective: Maximize the expected return rate

- Strategy: Distribute your wealth



An xa p a o

- Actual wealth grow rate: S = $\frac{1}{N}\sum_n \log p(T_n)$ + something that depends on the odds only

Can this be used as a skill score?

no anc

- The ***Ignorance Skill Score*** is

$$I(p) = -\log(p_{\mathbf{n}}(T_{\mathbf{n}}))$$

- The Ignorance is proper, local and smooth

- The Ignorance is the ***only*** proper, local and smooth score for continuous forecasts (Good 1952, Gneiting & Raftery 2004)

- The ***Linear Score*** or p ***Score*** $\mathcal{S}(p_{\mathbf{n}}\ t) = p_{\mathbf{n}}(t)$ is ***improper***

- The RMS error depends only on some moments and therefore is ***not*** strictly proper

- Many proper ***nonlocal*** Scores have been suggested and used (see talk by Zoltan Toth about CRPS)

Conclusion

- End users need probabilistic forecasts to make better decisions
- We need skill scores that measure desirable properties of probabilistic forecasts
- We need to use proper scores, since improper scores give misleading answers – we would reject even the optimal forecast
- There are only a handful of essentially different proper skill scores (see www.dime.lse.ac.uk)