Asymmetric prediction intervals using half moment of distribution

Ivan Svetunkov & Nikolaos Kourentzes

Presentation at ISIR 2016, Budapest

23^{rd} August 2016

Lancaster University Management School
Lancaster Centre for Forecasting

Motivation

- Defining safety stock level is important in inventory control.
- The safety stock calculation is connected to the calculation of Prediction Intervals (PI):
  - one vs. two-sided α;
  - cumulative vs. per-period.
- Typically we assume symmetric error distributions $\rightarrow$ often inappropriate.
- Develop (relatively) simple ways to produce assymetric PIs.

LCF

How are PIs typically constructed?

- We calculate PIs as:

$$\mu_{t+h|t} - z_{\alpha/2}\sigma_{t+h|t} < y_{t+h} < \mu_{t+h|t} + z_{1-\alpha/2}\sigma_{t+h|t}, \qquad (1)$$

  - $\mu_{t+h|t}$ is the conditional mean,
  - $\sigma_{t+h|t}$ is the conditional variance,
  - $z_{1-\alpha/2}$ is the z-statistic value for probability α.

- Assuming normality $z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$

- Eq. (1) is also a good approximation for cases of not normal, but **symmetric** distributions.

LCF

What should we do when the distribution is not symmetric?

- We want to use information about asymmetry, in a relatively simple way $\rightarrow$ easy to transfer to practice.
- Idea:
  - use different estimation of lower and upper variance differently
  - use different statistics $\rightarrow$ standard deviation makes sense when the distribution is symmetric and $\bar{Y}$ describes well the central tendency of the error distribution.
- We introduce a statistic that does both, the **half moment**.

Half moment and its properties

- Half moment measures density of distribution on left and right sides from some constant C, which is a measure of central tendency:

$$\text{HM} = \sum_{t=1}^{T} \sqrt{y_t - C},$$

  - y_t is variable of interest.

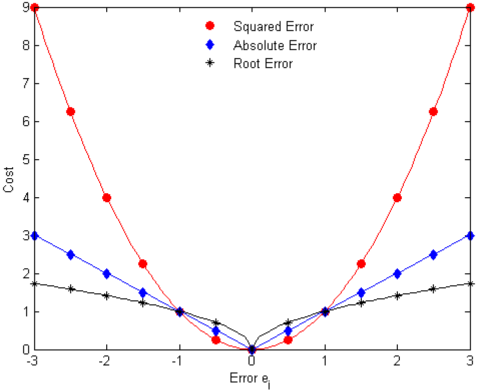
- HM is in general a complex number:

$$\text{HM} = \Re(\text{HM}) + i\Im(\text{HM})$$

.

- i is imaginary unit that satisfies: $i^2 = -1$.

HM is robust to extreme values



- Real part $\Re(\text{HM})$ determines density of right-hand side (from C) of distribution $\rightarrow$ that would be errors above the centre.
- Imaginary part $\Im(\text{HM})$ shows density of left-hand side of distribution $\rightarrow$ that would be errors below the centre.
- The higher values of $\Re(\text{HM})$ or $\Im(\text{HM})$ are, the longer corresponding tail of distribution is.
- Note that $\Re(\text{HM})$ and $\Im(\text{HM})$ do not have to be equal.
- If the size of the real and imaginary parts is the focus then the **Half Absolute Momement** (HAM) is connected to HM:

$$\text{HAM} = \sum_{t=1}^{T} \sqrt{|y_t - C|} = \Re(\text{HM}) + \Im(\text{HM}).$$

LCF

- HM for standard normal distribution is:

$$\text{HM}_N = (1+i)\Gamma(0.75)\pi^{-0.5}2^{-0.75} \approx (1+i)0.411,$$

  - $\Gamma(\cdot)$ is Gamma function.

- Bounds can be constructed using this information:

$$\begin{cases} \mu_{t+h|t} + z_{\alpha/2}\Im(\text{HM}_{t+h|t})^2/\Im(\text{HM}_N)^2 \\ \mu_{t+h|t} + z_{1-\alpha/2}\Re(\text{HM}_{t+h|t})^2/\Re(\text{HM}_N)^2 \end{cases},$$

- so $\Im(\text{HM})^2/\Im(\text{HM}_N)^2$ is estimate of σ_l for left-hand side, while $\Re(\text{HM})^2/\Re(\text{HM}_N)^2$ is estimate of σ_r for right-hand side.

- For standard deviation C is $\bar{Y}$
- The question is how to estimate C for HM (or HAM). This can be:
  - Mean of y_t $\rightarrow$ assumes symmetry;
  - Median of y_t $\rightarrow$ robust to extremes, but still enjoys symmetry;
  - Mode of y_t $\rightarrow$ does not assume symmetry, but needs to be estimated;
  - Optimal value based on minimum of HAM:

$$C = argmin_{c \in \mathbb{R}} \sum \sqrt{|y_t - c|}$$

A standard deviation based alternative

- Another way of constructing asymmetric PI $\rightarrow$ estimate σ_l and σ_r separately:

$$\sigma_l = \frac{1}{T_l} \sum_{y_t < \mu} (y_t - \mu)^2$$
$$\sigma_r = \frac{1}{T_r} \sum_{y_t > \mu} (y_t - \mu)^2,$$

  - T_l is number of observations to the left of μ;
  - T_r is number of observation to the right of μ.

- Update the calculation of PIs:

$$\mu_{t+h|t} + z_{\alpha/2}\sigma_{l,t+h|t} < y_{t+h} < \mu_{t+h|t} + z_{1-\alpha/2}\sigma_{r,t+h|t}. \quad (2)$$

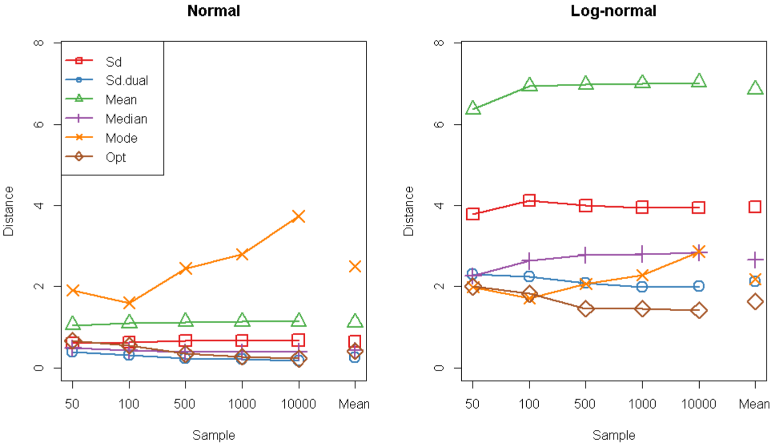
Simulations

- In order to compare all the methods, we do simulations $\rightarrow$ control distributions.
- 1000 samples from **Normal** and **Log-normal** distributions with sizes:
  - 50,
  - 100,
  - 500,
  - 1000,
  - 10000.
- We expect to see the proposed methods to make a difference for the Log-normal case.

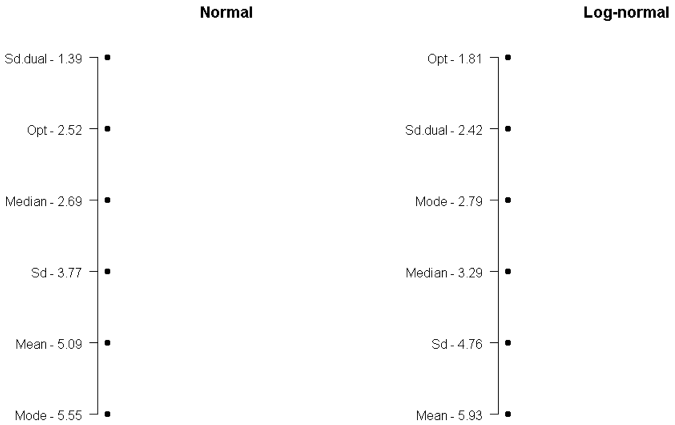
LCF

- Several PI construction methods:
  - Standard (Sd) – *benchmark*;
  - Two standard deviations, method (2) (Sd.dual);
  - HM with $C = \bar{y}$ (Mean);
  - HM with $C = \text{Md}(y)$ (Median);
  - HM with $C = \text{Mo}(y)$ (Mode);
  - HM with optimised C (Opt);
- Typical metric of performance is coverage. We do not use it as it is one-sided (does not evaluate how much more you cover!)
- Instead we will use the absolute *distance* of the PIs from the empirical realised quantiles
  $\rightarrow$ penalises both under- and over-coverage.

Results


Normal
Log-normal
Sd
Sd.dual
Mean
Median
Mode
Opt
Distance
Sample
0
2
4
6
8
50
100
500
1000
10000
Mean

Nemenyi post-hoc test for significant differences



Of course one should keep in mind that we can increase the number of distributions until we get significance!

Normal distribution

	Sd	Sd.dual	Mean	Median	Mode	Opt
50	0.61	**0.39**	1.05	0.49	1.91	0.67
100	0.63	**0.30**	1.09	0.44	1.60	0.55
500	0.66	**0.23**	1.13	0.39	2.45	0.34
1000	0.67	**0.21**	1.14	0.39	2.80	0.28
10000	0.68	**0.18**	1.14	0.40	3.74	0.23
Mean	0.65	**0.26**	1.11	0.42	2.50	0.41

Table: Overall distances for different methods. Normal distribution

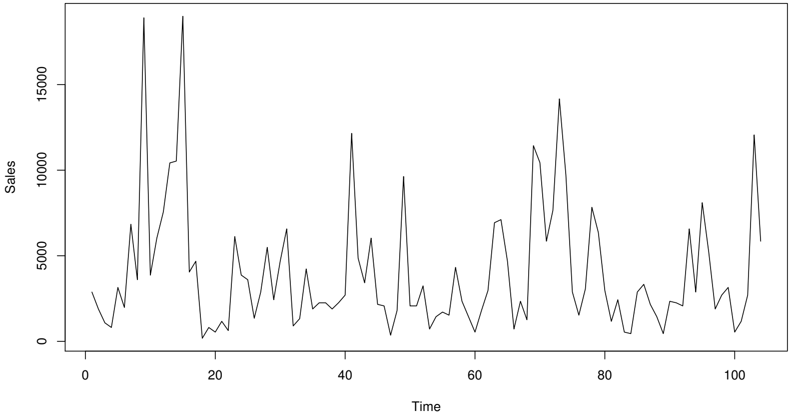
Log-normal distribution

	Sd	Sd.dual	Mean	Median	Mode	Opt
50	3.78	2.30	6.36	2.26	**1.99**	2.00
100	4.12	2.24	6.93	2.63	**1.71**	1.82
500	3.99	2.09	6.98	2.78	2.07	**1.45**
1000	3.95	2.00	7.00	2.81	2.28	**1.45**
10000	3.95	2.01	7.03	2.84	2.86	**1.42**
Mean	3.96	2.13	6.86	2.66	2.18	**1.63**

Table: Overall distances for different methods. Log-normal distribution

Real data experiment

- Use 12 heavily promoted time series
  $n = 103$ weeks, use 52 as test set.
- Perform rolling origin evaluation.
- Evaluate PI distance for one-step ahead predictions.
- Non-seasonal exponential smoothing (allow for any trend or none) with MAE as cost function $\rightarrow$ heavily promoted.


Sales
0
5000
10000
15000
0
20
40
60
80
100
Time

Results

	Dist.Lower	Dist.Upper	Dist.Total
Sd	1.93	1.93	3.85
Sd.dual	1.11	2.66	3.78
Mean	2.02	1.81	3.83
Median	1.12	1.75	2.87
Mode	**1.00**	1.79	**2.79**
Opt	1.47	**1.72**	3.19

Table: Overall distances for different methods.

Conclusions

- HM produces robust asymmetric intervals;
  - is robustness desirable?
  - ... case of baseline + judgemental adjustments
  - ... shocks in the supply chain

- OK for symmetric distributions;

- It performs well error distribution is asymmetric;

- Left/right standard deviations is an interesting alternative.

LCF

Thank you for your attention!

Nikolaos Kourentzes
n.kourentzes@lancaster.ac.uk

Lancaster University Management School
Lancaster Centre for Forecasting