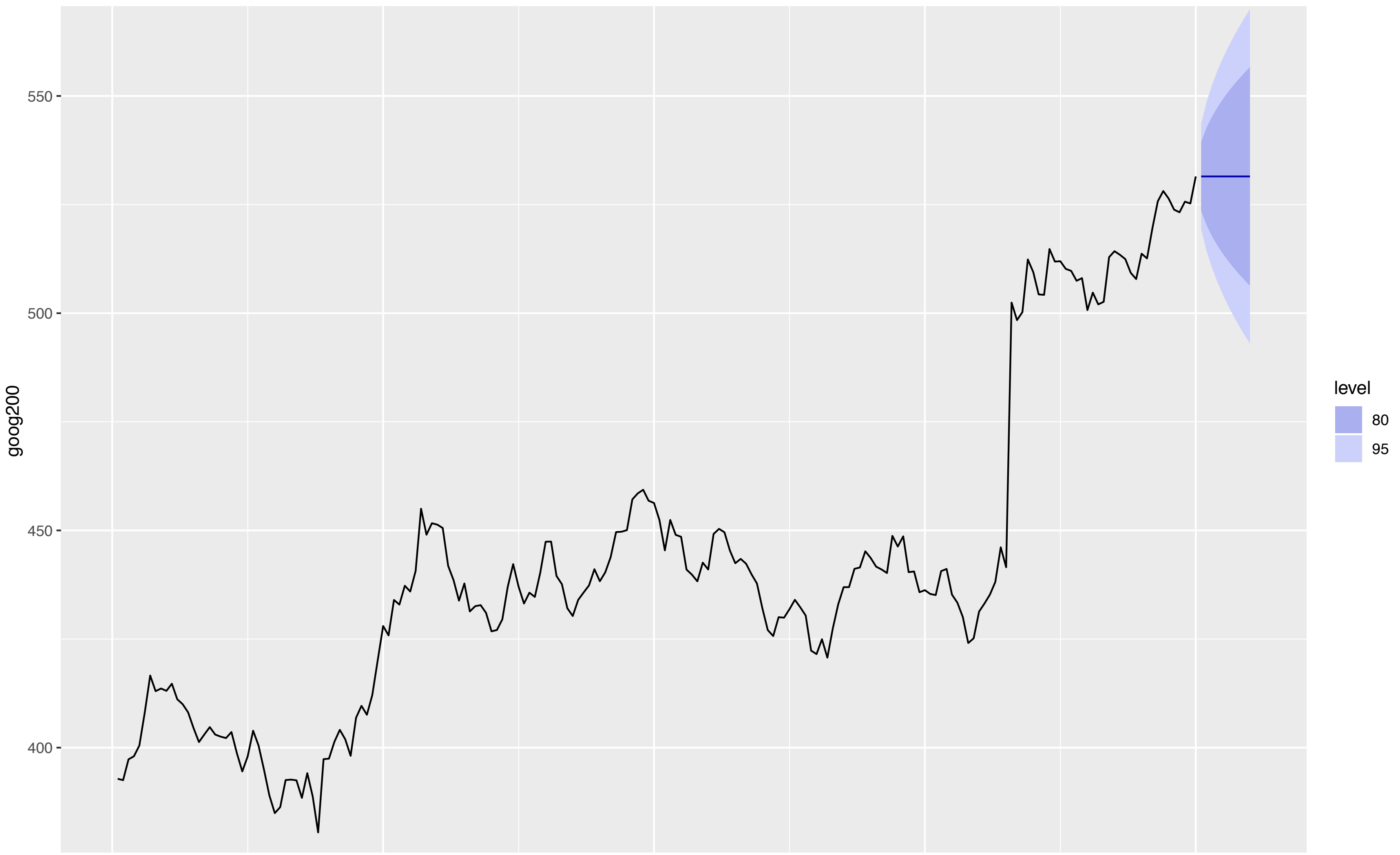
Deriving the prediction intervals for the naïve method

Giorgio Corani (giorgio@idsia.ch)

Naive prediction



Naive prediction

- The naive prediction is characterized by:
  - Next values in future are all assumed to be equal to the last observation:

$$\hat{y}_{t+1} = y_t$$
$$\hat{y}_{t+2} = y_t$$
$$\ldots$$
$$\hat{y}_{t+h} = y_t$$

  - However the uncertainty of the forecast increases with the forecast horizon.

Predicting h steps ahead with the naïve model

The naive model assumes

$y_{t+1} = y_t + \epsilon_{t+1}$ with $\epsilon_{t+1} \sim N(0,\sigma^2)$

Hence:

$y_{t+2} = y_{t+1} + \epsilon_{t+2} = y_t + \epsilon_{t+1} + \epsilon_{t+2}$

$y_{t+3} = y_{t+2} + \epsilon_{t+3} = y_t + \epsilon_{t+1} + \epsilon_{t+2} + \epsilon_{t+3}$

and in general:

$y_{t+h} = \mathrm{y}_{t+h-1} + \epsilon_{\{t+h\}} = y_t + \epsilon_{t+1} + \ldots + \epsilon_{t+h}$

We assume the noise values $(\epsilon_t, \epsilon_{t+1}, \ldots, \epsilon_{t+h})$ to be independent and each noise value $\epsilon_{\{\bullet\}}$ to be drawn from N(0 , σ^2).

Sum of independent Gaussian variables

$X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$

Consider the sum variable: $Y = X_1 + X_2$

For the property of Gaussian: $\mathrm{Y} \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

The sum $\mathrm{Y} = \sum_{i=1}^{h} X_i$ of h independent normal variables $\mathrm{X}_1, \ldots, X_h$

is distributed: $\mathrm{Y} \sim N(\sum_{i=1}^{h} \mu_i, \sum_{i=1}^{h} \sigma_i^2)$

Sum of independent Gaussian variables

If the h independent normal variables have

the same mean 0 ($\mu_1 = \mu_2 = \ldots = \mu_h = 0$)

the same variance $\sigma^2{}_1 = \sigma^2{}_2 = \ldots = \sigma^2{}_h = \sigma^2$

we get:

$$Y \sim N(0,\ h\sigma^2)$$

(standard deviation is hence $\sigma_y = \sigma\sqrt{h}$)

Noise properties to be checked

- We estimate the properties of the noise from the residual of the models
- We need to check the residuals to be unbiased, in order to assume the mean of the noise to be 0.
- We need to check that the residuals are uncorrelated in order to use the summation of the variance of the noise at various time steps.
- We estimate the variance of the noise to be equal to the variance of the residuals.
- Once checked the assumptions about the residuals, we can compute the prediction intervals.

Predicting h steps ahead with the naïve model

$$y_{t+h} = y_t + \epsilon_{t+1} + \ldots + \epsilon_{t+h}$$

Under the previous assumptions:

$$\epsilon_{t+1} + \ldots + \epsilon_{t+h} \sim N(0, \sum_{i=1}^{h} \sigma^2) = N(0, h\sigma^2)$$

Hence

$$y_{t+h} = y_t + N(0, h\sigma^2)$$

The expected value of y_{t+h} is the same (y_t) for any value of h, but its variance grows with h.

Point prediction

Time				t	t+1	t+2
Y	4	3.5	4	4	?	?
Naïve prediction		4	3.5	4		
Residual		3.5 - 4 = -0.5	4 – 3.5 = 0.5	4 – 4 = 0		

The point prediction for any forecast horizon is always the last observed value (at time t).

$$\hat{y}_{t+1} = y_t = 4$$

$$\hat{y}_{t+2} = y_t = 4$$

$$\hat{y}_{t+3} = y_t = 4$$

…

Toy example

Time				t	t+1	t+2
Y	4	3.5	4	4	?	?
Naïve prediction		4	3.5	4		
Residual		3.5 - 4 = -0.5	4 – 3.5 = 0.5	4 – 4 = 0		

- The mean of the residuals is
  - (-0.5 + 0.5 + 0) / 3 = 0
  - (in real applications their mean will not be a sharp 0)

- The variance of the residuals is :
  - σ^2 =[(-0.5-0)^2 + (0.5 -0)^2 + (0-0)^2] /2 = .25

- The squares are computed with respect to the mean value of the residual (0 in our example, usually its value will be slightly different from zero).

- Notice that given T residuals, the variance is estimated by dividing the sum of squares by (T-1).

Prediction intervals

- The standard deviation of the residuals increases with the forecast horizon h.

In this example, $\sigma = \sqrt{0.25} = 0.5.$

The subscript of σ denotes the forecast horizon.

In general $\sigma_h = \sigma\sqrt{h}$:

$h = 1 \to \sigma_1 = \sigma = 0.5$

$h = 2 \to \sigma_2 = \sigma\sqrt{2} = 0.5\sqrt{2}$

$h = 3 \to \sigma_3 = \sigma\sqrt{3} = 0.5\sqrt{3}$

Example: 95% prediction interval, two steps ahead

$$\sigma_2 = \sigma\sqrt{2} = 0.5\sqrt{2} = 0.7$$

A symmetric 95% prediction intervals requires the quantiles 0.025 and 0.975 of the distribution $N(0, \sigma^2{}_2)$.

The easy way (use R!):

$$\text{qnorm}\left(0.025,\ \text{sd} = 0.7,\ \text{mean} = 0\right) = 1.38$$

$$\text{qnorm}\left(0.975,\ \text{sd} = 0.7,\ \text{mean} = 0\right) = -1.38$$

The 95% prediction interval for $\hat{y}_{t+2}$ is :

$$[y_t \pm 1.38] = [2.62, 5.38]$$

Table of the normal density

Look up in table which value of the normal standard (z-value) has cumulative probability equal to the percentile of interest. For instance, percentile 0.975 yields $z = 1.96$.

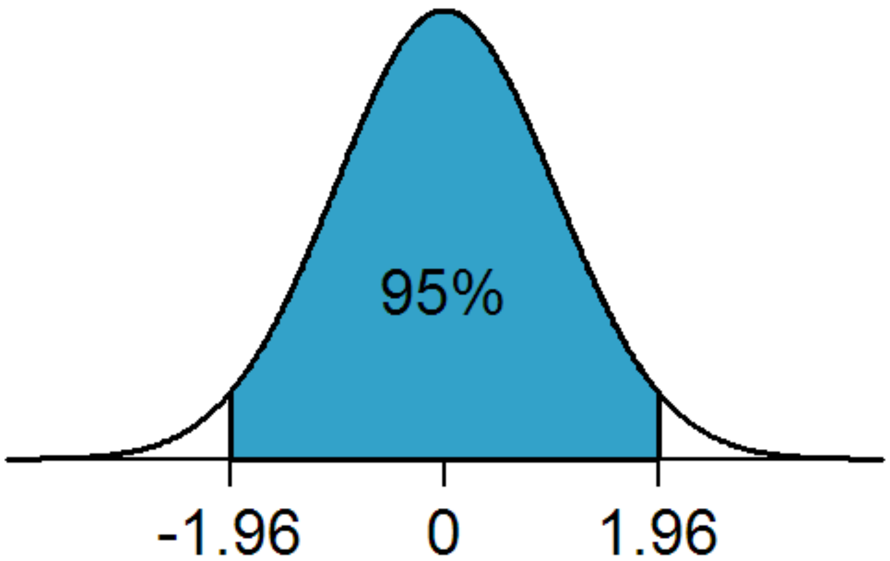
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857

Table of the normal density

The 0.975 percentile of the normal standard is 1.96.

The 0.025 percentile of the normal standard is -1.96.

A random draw from N(0,1) falls within (-1.96, 1.96) with probability 95%.



Prediction interval

For a generic normal distribution $N(\mu, \sigma^2)$, the 95% of the values fall within the interval $(\mu \pm 1.96\sigma)$

Given $\sigma = 0.5$, the 95% prediction interval for the naive prediction two-steps ahead is:

$$y_t \pm 1.96\sigma_2 = 4 \pm 1.96 \cdot (0.5\sqrt{2}) = 4 \pm 1.38 = [2.62, 5.38]$$

Your turn: compute the 80% prediction interval for the 3-steps ahead prediction

Basic relations:

$$\sigma_3 = \sigma\sqrt{3} = 3 \bullet 0.25$$

The quantiles 0.10 and 0.90 of the normal distribution are ± 1.28

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.0156	.0138	.1020	.1003	.0985

Your turn: compute the 80% prediction interval for the 3-steps ahead prediction

The 80% prediction interval is:

$$y_t \pm 1.28\sigma_3 = y_t \pm 1.28\sqrt{3 \bullet 0.25} = [5.1, 2.89]$$

Summing up, the width of the prediction intervals increases:

–with the level of confidence

–with the forecast horizon

Recap: Important percentiles of the normal distribution

- Percentiles .005 and .995: ± 2.58
  - Used for confidence level 99%
- Percentiles .025 and .975: ± 1.96
  - Used for confidence level 95%
- Percentiles .05 and .95: ± 1.64
  - Used for confidence level 90%
- Percentiles .10 and .90: ± 1.28
  - Used for confidence level 80%