Prediction Intervals – Estimating the chance of where future samples will fall

What is a prediction interval?

Source: https://en.wikipedia.org/wiki/Prediction_interval

In statistical inference, specifically predictive inference, a prediction interval is an estimate of an interval in which future observations will fall, with a certain probability, given what has already been observed.

Estimating the range of actual data by random sampling

When actual data samples can be observed, its handy to know how likely it is that you have discovered the range of likely values. This is useful in understanding how likely there is a lower or higher sample yet to be discovered. Like all random sampling, there is absolutely no guarantee that you have discovered any amount of the range, but prediction intervals give you the probability on average.

Probability the next sample is within the previously seen range after "n" samples = $\frac{(n-1)}{(n+1)}$ X 100

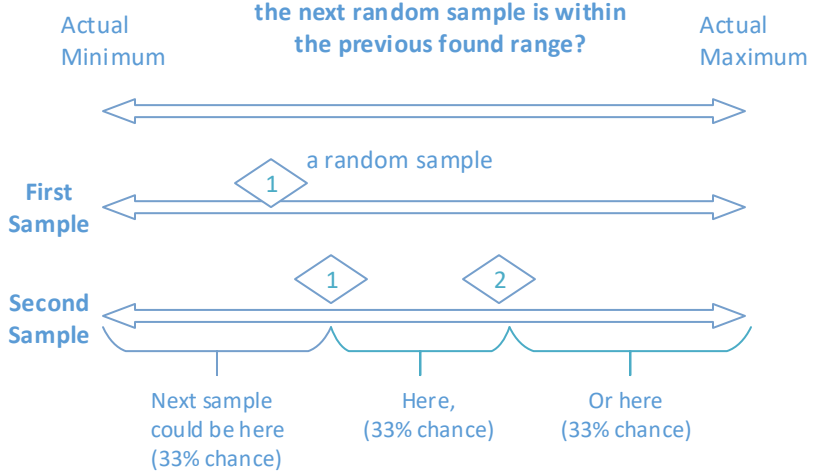
Probability the next sample is lower than the lowest sample so far after "n" = $\frac{1}{(n+1)}$ X 100

Probability the next sample is higher than the highest sample so far after "n" = $\frac{1}{(n+1)}$ X 100

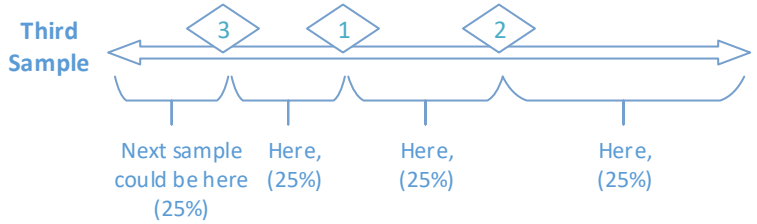
Samples so far (n)	Probability for each interval	Probability next sample in range
1	50.00%	0.00%
2	33.33%	33.33%
3	25.00%	50.00%
4	20.00%	60.00%
5	16.67%	66.67%
6	14.29%	71.43%
7	12.50%	75.00%
8	11.11%	77.78%
9	10.00%	80.00%
10	9.09%	81.82%
11	8.33%	83.33%
12	7.69%	84.62%
13	7.14%	85.71%
14	6.67%	86.67%
15	6.25%	87.50%

Samples so far (n)	Probability for each interval	Probability next sample in range
16	5.88%	88.24%
17	5.56%	88.89%
18	5.26%	89.47%
19	5.00%	90.00%
20	4.76%	90.48%
21	4.55%	90.91%
22	4.35%	91.30%
23	4.17%	91.67%
24	4.00%	92.00%
25	3.85%	92.31%
26	3.70%	92.59%
27	3.57%	92.86%
28	3.45%	93.10%
29	3.33%	93.33%
30	3.23%	93.55%

Q. How can we estimate the chance the next random sample is within the previous found range?



After two samples, there are three spots the next sample could be. Equally splitting the chances, there is a 33.33% chance the next sample is between the previous samples (1) and (2).



After three samples, there are four spots the next sample could be. Equally splitting the chances, there is a 50% chance the next sample is between the lowest (3) and highest (2) so far.

Important assumptions (that are rarely perfectly true)

- The samples are taken at random. Convenient isn't random!
- The distribution is uniform – all values have equal chance.
- The probability is on average. It's when it is more likely than not (~50%)

These are rarely always true in the real world. Milage will vary depending mainly on the underlying distribution. If the distribution is skewed, it can take hundreds of samples to get the lower probability end of the range.

For more probabilistic forecasting resources: FocusedObjective.com Email: troy.magennis@focusedobjective.com