# How to Gauge the Quality of a Testing Method When Ground Truth Is Known with Uncertainty

**N. Gray**[1], **S. Ferson**[1], and **V. Kreinovich**[2]

[1]Institute for Risk and Uncertainty, University of Liverpool
Liverpool L69 7ZF, UK, Nicholas.Gray@liverpool.ac.uk, sandp8@gmail.com

[2]Department of Computer Science, University of Texas at El Paso
El Paso, TX 79968, USA, vladik@utep.edu

***Keywords:*** *sensitivity; specificity; precision; uncertain ground truth.*

## Abstract

The quality of a method for testing a property (e.g., a disease) can be accessed by the following three characteristics.

*Sensitivity* (also called *recall* or *True Positive Rate*, TPR for short) is the proportion of objects with this property that were correctly classified by the test (e.g., the proportion of sick people for which the test recognized the disease).

*Specificity* (also called *True Negative Rate*, TNR, for short), the proportion of objects that do not have the tested property that were correctly classified by the test (e.g., the proportion of healthy people that this test classified as healthy).

*Precision* (also known as Positive Predictive Value, PPV, for short) is the proportion of objects who have the tested property among those that the test classified as having it (e.g., proportion of sick people among those that the test classified as sick).

Computing these values requires that we know the "ground truth", i.e., that we know exactly which objects have the tested property and which do not – in the above example, which patients have the tested disease. In practice, however, this information often comes from experts – e.g., from medical doctors – and experts are often not 100% sure about their statements and their diagnoses. A reasonable way to describe an expert's uncertainty in each statement $S$ is by subjective probability $p(S)$. In this paper, we show how to take this expert uncertainty into account when gauging the quality of a test.

Namely, under this uncertainty. we only have probabilistic information about the ground truth. As a result, the values TPR, TNR, and PPV (which are estimated based on ground truth) become random variables even when the data – and the results of applying the testing method to this data – are fixed. In this paper, we use the Central Limit Theorem (see, e.g., Sheskin (2011)) to show that these variables are normally distributed, and we provide formulas for their means and standard deviations.

## References

Sheskin, D.-J., *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.