Statistical Analysis of Image-Fidelity Ratings

Michael H. Brill (SARNOFF LABS)

This document summarizes a statistical analysis that was undertaken in January, 1996 to determine which of two models better accounts for subjective rating comparisons between distorted and undistorted versions of an image sequence. The study was constrained by time and data volume; therefore, the analysis reported here reflects that constraint. However, the Appendix to this document enumerates some other relevant statistical questions and tools, for future reference.

Data to be Analyzed.

The data consist of three arrays:

x_{1i} = responses of model M1 to image-sequence i (where i = 1,2,...I);
x_{2i} = responses of model M2 to image-sequence i;
y_i = observer-averaged subjective responses to image-sequence i

Question to be Answered.

Which model, M1 or M2, can be said to correlate better with the subjective ratings, and at what level of significance can each such a statement be made?

Our approach to deciding the above question was to make two plots, incorporating all the data: y_i versus x_{1i} for the first plot, and y_i versus x_{2i} for the second plot. By comparing the correlation obtained from plot 1 with the correlation obtained from plot 2 , we hoped to decide which model is a better fit to the data.

Statistical Tools.

The following tools were used to analyze the above question:

1. Linear (Pearson) correlation. This is a measure of how well the data fit a straight line. The correlation, r, is given by

$$r = \frac{\sum_i (x_i - E(x))\,(y_i - E(y))}{[\sum_i (x_i - E(x))^2]^{0.5}\,[\sum_i (y_i - E(y))^2]^{0.5}},$$

where E(x) is the mean of x_i, etc. Note that r = 1 when there is a perfect fit to any line of positive slope, and that r = -1 for a perfect fit to any line of negative slope.

In the presence of random Gaussian noise, the statistic $r\,[(N-2)/(1 - r^2)]^{0.5}$ is distributed approximately as Student's distribution with N-2 degrees of freedom. Therefore, a value of r computed from data is associated with a probability that random noise could have met or exceeded the data-derived value r. This probability is termed the *significance level*. The *Fisher z-transform* of r, $z = 0.5 \ln[(1+r)/(1-r)]$, is distributed approximately normally, and the significance level of a difference between two measured correlation coefficients r_1 and r_2 is given by

$$\mathrm{erfc}\ \{\ |z_1 - z_2|\ [2/(N_1 - 3) + 2/(N_2 - 3)]^{-0.5}\ \}.$$

Here, N_1 and N_2 are the numbers of data points in data sets 1 and 2. The values of z_1 and z_2, as well as level of significance of $|z_1 - z_2|$, will help to answer the question at hand.

2. Spearman rank-order correlation coefficient. This is a nonparametric statistic that is independent of any monotonic function applied to the model output. In fact, the statistic depends only on the rank-ordering of the data and model values. Because of this dependence, the Spearman coefficient reveals relationships between model and data that are hidden by a straight-line fit, while giving a metric of how effectively the model rank-orders image fidelity. If R_i is the rank of data point x_i from data array x, and S_i is the rank of data point y_i from data array y, then the Spearman r_s coefficient between sets x and y is given by

$$r_s = \frac{\sum_i (R_i - E(R))\,(S_i - E(S))}{[\sum_i (R_i - E(R))^2]^{0.5}\,[\sum_i (S_i - E(S))^2]^{0.5}}$$

In the presence of random Gaussian noise, the statistic $r_s\,[(N-2)/(1 - r_s^2)]^{0.5}$ is distributed approximately as Student's distribution with N-2 degrees of freedom. Therefore, a value of r_s computed from data is associated with a probability (significance level) that random noise could have met or exceeded the data-based value r_s. As in (2) above, the statistical significance of a difference in two derived r_s statistics can be obtained from the difference of the Fisher z-transforms of the two evaluations of r_s, though the function erfc. The analysis proceeds as follows:

In the presence of random Gaussian noise, the statistic $r_s\,[(N-2)/(1 - r_s^2)]^{0.5}$ is distributed approximately as Student's distribution with N-2 degrees of freedom. Therefore, a value of r_s computed from data is associated with a probability (significance level) that random noise could have met or exceeded the data-based value r_s. As in (2) above, the Fisher transform of r_s, $z_s = 0.5 \ln[(1+r_s)/(1-r_s)]$, is distributed approximately normally, and the significance level of a difference between two measured Spearman coefficients r_{s1} and r_{s2} is given by

$$\mathrm{erfc}\ \{\ |z_{s1} - z_{s2}|\ [2/(N_1 - 3) + 2/(N_2 - 3)]^{-0.5}\ \}.$$

Here, N_1 and N_2 are the numbers of data points in data sets 1 and 2; subscripts 1 and 2 on r_S and z_S also refer to the data sets 1 and 2, respectively. The values of z_{S1} and z_{S2}, as well as level of significance of $|z_{S1} - z_{S2}|$, will help to answer the question at hand.