<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Mutual Information and a Maximum Entropy Estimate of a Wine Rating's Distribution, Information about context and cross section data can improve analyses of wine ratings.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I want to submit an abstract for:</strong></td>
<td>Conference Presentation</td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
<td>Jeff Bodington</td>
</tr>
<tr>
<td><strong>E-Mail Corresponding Author</strong></td>
<td><a href="mailto:jcb@bodingtonandcompany.com">jcb@bodingtonandcompany.com</a></td>
</tr>
<tr>
<td><strong>Affiliation</strong></td>
<td>Bodington &amp; Company</td>
</tr>
<tr>
<td><strong>Keywords</strong></td>
<td>wine, ratings, statistics</td>
</tr>
<tr>
<td><strong>Research Question</strong></td>
<td>How can the distribution of an uncertain wine tasting be estimated?</td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td>statistics, maximum entropy, mutual information</td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td>Improve accuracy (decrease sum of squared error) of estimate by 30% to over 50%</td>
</tr>
</tbody>
</table>
| **Abstract** | I Introduction  

A difficulty in wine-ratings-related research, and in calculating consensus among judges, is that each rating is one observation drawn from a latent probability distribution that is judge- and wine-specific. One draw from a distribution is a tiny sample size so expert and consumer interpretations of wine ratings, analyses of ratings, predictions of ratings, and calculations of consensus among judges are infused with uncertainty that is difficult to quantify. Bodington (2022a) proposed to quantify that uncertainty using a maximum information entropy estimate of the latent distribution of a wine rating. Using 1,599 blind triplicates rated by judges at the California State Fair Commercial Wine Competition (CSF), and 30 blind triplicates rated by judges at a tasting in Stellenbosch reported by Cicchetti (2014), he showed that the result is much more accurate than ignoring the uncertainty about a rating. This article shows that adding information about the context for a judge’s rating, or adding mutual information from within cross section ratings data, improves the accuracy of the estimated maximum entropy distribution by another 30% to 50%. A check shows that the means and standard deviations of the resulting distributions are similar to those implied by blind triplicate ratings. And an application to the famous 1976 Judgement of Paris shows that the results are generally consistent with sums of scores but more informative and, for some of the wines, imply a different ranking. The ratings data and MATLAB code concerning results reported in this article, including a Nelder-Mead minimizer, are available on request. |
II. Information

We do have information in addition to the observed value of a rating \( x_{ij}^o \) and the categories in the rating system. That information includes the context of a tasting, and it may include cross section data when panels of judges assess flights of wines. Context and cross section data contain information that may enable a more precise, lower entropy, estimate of the distribution of \( u \) in Bodington (2022a). From here forward, that more-informed distribution will be denoted as \( u' \). Ignoring uncertainty is equivalent to asserting that \( u' \) is a degenerate or one-hot distribution.

III.1 Context Information

Information about the context for a rating may include the name and reputation of a wine critic or a sponsoring organization and its judges. For example, the CSF judges are trained and tested, and historical CSF data provide information about the past dispersions of judges’ ratings. The means and standard deviations (SDs) of the ratings that judges assigned to CSF blind triplicates appear in Table 1. For example, the pool of 557 triplicates that included a Silver medal has a mean rank of 6.0 and an SD of 2.1 ranks.

Table 1.
Distributions of CSF Triplicates that Include a Specified Ordinal Category

Making the assumption that historical data are a reasonable guide to expectations, the data in Table 1 yield an estimate of \( u' \). If for example a judge assigns a rating of Gold to a wine, an estimate of the latent discrete and bounded distribution of that rating has a mean and SD of 5.0 and 3.2 rank categories. Using \( (x_{ij}^o=\text{Gold}) \sim u' \) (5.0,3.2) yields a solution to Equation (2) that is consistent with historical data.

Other examples of contextual information could include data about a particular judge’s expertise, serial position carry over, and expectations. Whatever that information may be, \( u' \) is an explicit and flexible description of its analytical impact on the distribution of a rating.

III.2 Cross Section Data & Mutual Information

When one or more panels of judges assess flights of wines, cross section data may be available. Both the Stellenbosch and CSF tastings cited above yielded cross section data. For every rating \( x_{ij}^o \), \( x_i^0 \) is the vector of all ratings assigned to that wine by all judges, \( x_j^0 \) is the vector of all ratings assigned to all wines by that judge, and \( x^0 \) is the union of ratings assigned by all judges to all wines.

Consider a panel of judges that is assessing and rating a flight of wines. All the judges see and taste the same wines, have the same macro-scale tasting-related neurology, and are asked to assign ratings according to the same scheme. The effect of those and other commonalities on the distributions of the ratings the judges assign is an example of what Shannon’s colleague Robert Fano (2001, p. 9) defined as mutual information. Learned-Miller (2013, p. 4) describes it as “how much information is communicated, on average, in one random variable about another.” Rioul (2018, p. 64-67) describes it as an intersection of uncertainty sets where each decreases the entropy of the other. Analytically, for an ordered pair of vectors (A, B), their mutual information is the entropy of A, plus the entropy of B, minus the joint entropy of AB.

Using the notion that the ratings assigned in cross section data by judges \( k \neq j \) may contain information about the distribution of ratings assigned by judge j. Equation (2) in Bodington (2022a) is re-stated as (3) below. Lagrangian maximization of mutual information yields the term within the square brackets […] in Equation (3A) that expresses the weighted average distribution of the other judges’ ratings \( ( \tilde{d}[x] \_ik^0 ) \), where the weights are the shares of mutual information (M) that each judge \( k \neq j \) has about judge j. Equation (3B) then employs that calculation of \( u_{ij}' \) to estimate \( \hat{p}_{ij} \).
\[ u_{ij}^\prime = (1/(1+n_{ij}) \cdot u + (n_{ij}/(1+n_{ij})) \cdot \left( \sum_{k \neq j} \frac{M(x_j^o x_k^o)}{\sum_{k \neq j} M(x_j^o x_k^o)} \right) \cdot d|x_{ij}^0 \right) \] (3A)

\[ \text{arg}[p_{ij}^\prime] = \text{argmin} \left[ (1/(1+n_{ij})) \cdot I(u_{ij}^\prime, p_{ij}) + (n_{ij}/(1+n_{ij})) \cdot I(d|x_{ij}^0, p_{ij}) \right] \] (3B)

Equation (3) has several useful properties. Following Jaynes (1957, p. 623) and Golan Judge & Miller (1996, p. 115-116), it retains consistency with observed data but tempers that consistency with mutual information, the judge sample size \((n_j)\), and the replicate sample size \((n_{ij})\). Judges \(k \neq j\) that share no mutual information with judge \(j\), including a judge whose ratings are indistinguishable from random assignments, have no effect on the estimate of \(p_{ij}^\prime\). And Equation (3) has the essential asymptotic properties that \(u_{ij}^\prime = u\) for sample size \(n_j = 0\) and that \(p_{ij}^\prime \to d|x_{ij}^0\) as \(n_{ij} \to \infty\).

While Equation (3) has the appeal of grounding in Shannon's information theory, it implies a simpler approach that may be practical to implement for wine competition officials using spreadsheet software. That approach uses positive correlation coefficients \((r)\) in place of mutual information, in Equation (3A) and replaces (3A) with a weighted average rather than cross entropy solution.

VI. Test & Example

This section begins with a test of the relative accuracy of the solutions above using the Stellenbosch data, and then it presents an example using the white wines assessed at the 1976 Judgement of Paris.

Chicchetti (2014) published the scores assigned to two flights of eight wines each at a tasting in Stellenbosch, South Africa. Each flight was assessed by 15 judges, and each flight contained a set of blind triplicates. Although it probably understates true variance, assume here that the distribution of the triplicate results for each judge describes the "true" distribution of that judge’s ratings on a triplicate wine. But suppose that wine did not appear in triplicate and that only one of the scores was observed \((x_{ij}^0)\) where \(n_{ij} = 1\). In that case, ignoring the uncertainty about a rating is equivalent to assuming that \(x_{ij}^0\) has a degenerate or one-hot

Consent

✔ I agree to the privacy policy.

You find the link to our privacy policy at the bottom of the page. In the privacy policy you find a link for exporting and/or erasing your personal data stored in our database.