



# Methodologic procedure of beer competitions eliminating the problematic practices and based on the probabilistic approach

Tomáš Vrzal\* and Jana Olšovská

Research Institute of Brewing and Malting, Lipová 511/15,  
120 00 Prague 2, Czech Republic

\*corresponding author: [tomas.vrzal@beerresearch.cz](mailto:tomas.vrzal@beerresearch.cz)

## Abstract

In beer competitions, there are some common problematic practices which ruin the final results, i.e. inappropriate distribution of samples into individual groups, inappropriate scoring of samples, comparison of results obtained from different sub-groups selecting better samples, inappropriate method of handling tied scores, and inappropriate numerical evaluation. To avoid these, the presented study aims to propose a methodologic procedure which eliminates these problematic practices and is based on the probabilistic approach. Further, it wishes to test and evaluate it by the Monte Carlo simulation. The procedure is based on a sensory evaluation of beer samples by assessors through ranking tests. Beer samples and assessors are randomly divided into groups in the lowest round of the competition. Data evaluation, which determines advancing samples or winners, is based on the application of the probability theory, namely the Bayesian theorem, so that the best-evaluated samples could be identified. This procedure achieves a higher probability of accurately recognised best-evaluated samples in comparison to the procedures involving the aforementioned problematic practices.

**Keywords:** sensory analysis; beer evaluation; Bayesian statistics; Bayesian theorem; competition fairness.

## 1 Introduction

Beer competitions have been organized worldwide in various forms for more than a hundred years. A demand to hold a competition is mainly due to two reasons. The first one is the usage in marketing. A prize won in a prestigious competition is a sign of quality, breweries display diplomas in their visitor centres, publish their victories on their websites and in press releases, and often mention medals on product labels. The second reason is the nature of humans because competitiveness is a natural characteristic of the *Homo sapiens* species. Beating the competition is tempting, even if it does not impact the company's economic parameters (Olšovská, 2017).

Beer competitions have a varied organization structure and different evaluation methods; however, several basic attributes unite them all. Firstly, a sensory

evaluation of beer samples is performed where attention is mainly focused on odor and taste including an evaluation of off-flavours and assessment of the beer style conformity. Secondly, the samples are assessed anonymously. In some cases, a certain specification of properties is attached to competing samples, which "declassifies" a sample to a certain extent, however, this is balanced by information without which some beers could not be seriously judged (for example beer style or specific ingredients). Third, beers are divided into categories according to their type, original extract, or finer distinctions. This division is essential, especially for a relative evaluation, and can also cause problems to the organizers if there is only a small number of given types of beer in the competition (Frantík et al., 2005).

In principle, two basic models of competitive beer evaluation exist. The absolute evaluation separates individual samples according to the prescribed criteria without regard to others. The relative one involves a comparison of samples with each other usually according to a ranking test. Both models have their advantages and weaknesses. There are also systems combining both models (Frantík, 2001).

The main criteria of the competition should always be objectivity, fairness, and anonymity. Moreover, it is necessary to highlight that a group of assessors differs from a professional sensory panel which is regularly trained and calibrated according to international methodologies, for example, *Analytica EBC 13.4, 13.11, 13.13, ISO 5492:1992, ISO 6658:2005, ISO 8586-1:1993*, etc. Conversely, a group of assessors in a competition is composed of people with different levels of sensory ability, which are not calibrated. Hence, the whole competition procedure and data evaluation should be designed to obtain reliable results.

Despite efforts of organizers to shift the whole process and results to the highest possible level, several problematic practices, which should be avoided, are used in some competitions. Namely:

- (a) inappropriate distribution of samples into groups in a competition,
- (b) inappropriate scoring of samples,
- (c) comparison of results obtained from different sub-groups to select better samples,
- (d) method for handling tied scores,
- (e) inappropriate numerical evaluation (see the next paragraphs for a detailed discussion regarding these practices).

Such practices can decrease the quality of obtained data or bring some confusion into data evaluation. Hence, it can result in disadvantaging some competition samples or inappropriate mining of the raw data obtained from the assessors (the highest-rated samples might not win a medal).

#### **Description of the problematic practices:**

**Ad (a):** In some types of competitions, the beer samples are distributed into different sub-groups for their evaluation by the assessors in a given competition category (mainly due to the high amount of samples) such that there would not be any beer samples from the same brewery in one sub-group. However, such a practice is problematic as it brings selective handling with samples before the competition. Next, such samples are not compared in the competition (or in lower parts of the competition), and it also advantages breweries with a higher number of beer samples in the competition since there is a higher probability for their advance to the next round (or higher part of the competition).

**Ad (b):** If the competition is based on scoring the beer samples according to the assessor's evaluation, logical scoring should be done to respect the assessor's evaluation. Inappropriate scoring could bring some unwanted interference to the results.

**Ad (c):** Results from relative evaluations (for example in ranking test) performed in different sub-groups are usually not comparable as there are different beer samples in each sub-group. It is not possible to say that one beer sample from the first sub-group has a better or worse evaluation than a beer sample from the second sub-group (even if a sample from the first sub-group has a higher/lower sum of ranks across assessors than a sample from the second sub-group). Such ranking tests give us information about rank only, not about a quantitative difference among samples – in sports terminology “*we do not have a stopwatch, we only know the final rank*”.

**Ad (d):** Tied scores/evaluations could be problematic to handle when they are not acceptable in competition (e.g. when a decision about advancing samples has to be made). Such a decision should be made with extreme care. Several methodologies make such a decision according to the number of better evaluations between two samples with tied scores. Such an approach could be problematic as the decision is only made from a selective part of the available data (better evaluations; unacceptable in the field of data processing), however, all of the data should be inspected and used for the decision.

**Ad (e):** The numerical/mathematical approach for processing/evaluation of data obtained by the assessor's sensory evaluation should be also selected with care and the properties of the data should be considered. A lot of sensory data evaluations, especially in relative evaluations, rely only on calculating mean values or sum of rankings. Despite the easiness of such calculations, it assumes few properties of the input data which are not met. Such properties are:

- (i) data are in the form of continuous variables;
- (ii) the same quantitative difference among individual rankings;
- (iii) data comes from the deterministic process;
- (iv) evaluations of individual beer samples are independent.

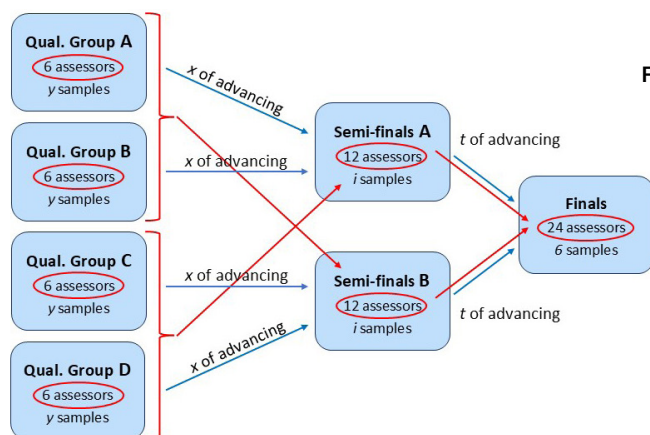
If these properties are met by the data, the calculation of mean values or sums is appropriate. However, the data obtained by ranking test have different properties:

- (i) data in the form of discrete variables;
- (ii) quantitative measure among individual rankings is not possible to define;
- (iii) data comes from the stochastic process;
- (iv) evaluations of individual beer samples are dependent (relative to the other beer samples).

And, in such cases, the calculation of means or sums of rankings leads to a lower quality of the results since this approach does not guarantee that the beer sample with a lower/higher mean or sum value has a lower/higher probability of better/worse evaluation of the beer sample.

All these problematic practices could ruin the evaluation (by bringing some systematic errors) when they occur alone and especially in combination – such a situation was modelled by Monte Carlo simulation and the resulted probabilities are plotted in Figure 1. In such case, the probability that the three best beer samples would win medals is limited to around 80%. This is true only in the case of a sufficiently large difference in the beer samples in comparison to the rest of samples in a competition – such a situation is quite improbable today as most of beer samples in competitions are of high quality. Therefore, the probability falls to lower values and the whole evaluation process might begin random in extreme cases or average samples would win.

From the perspective of the above-mentioned aspects, the aim of this paper is to propose a methodology for beer competitions based on the elimination of the problematic practices described above in order to obtain reliable results based on assessor's evaluations. The development of the methodology for beer competitions was carried out by designing particular steps in the competition and testing it with Monte Carlo simulations. The Monte Carlo simulations were performed by in-house programmed simulations. The principle of the data evaluation is based on the application of probability theory, namely the Bayesian theorem (Bertsekas and Tsitskilis, 2008).

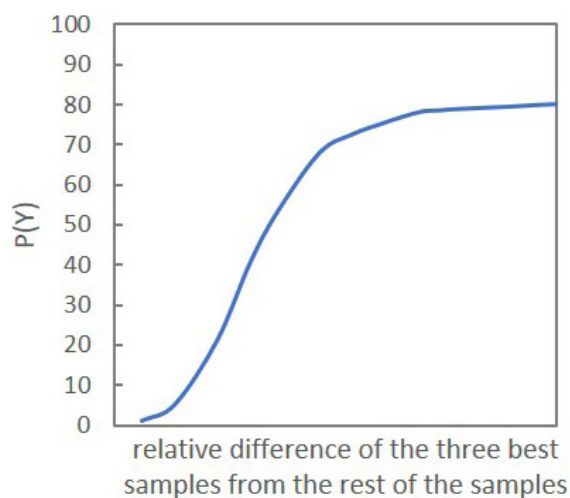


**Figure 2** Scheme of the beer competition – illustration for 24 beer samples. The blue arrows indicate advancing samples, the red arrows indicate distribution of assessors among individual groups in different rounds. Numbers of beer samples in each group of qualification and semifinal are indicated in Table 1 ( $y, i$  values). The number of beer samples in the final round is constantly set to six.

## 2 Principle and procedure of the methodology

### 2.1 Principle

The principle of the proposed competition methodology is based on the sensory evaluation of beer samples by 24 assessors in the qualification, semifinal and final rounds (depending on the total number of samples in the particular category in the competition, see Table 1 and Figure 2). The assessors are randomly divided into groups, and so is the beer sample sequence. After the sensory evaluation of beer samples by the assessors, they rank the samples from the best to the worst one (according to their sensory quality). Such obtained data are evaluated by Bayesian statistics to identify the best-assessed samples across the assessors – evaluation of the conditional probability that the samples are the best-assessed ones given the obtained data from the group of assessors. Samples with the highest probability advance to the next round of the competition (for instance from qualification to the semifinal), and also total probability is used in the final round to identify winners.



**Figure 1** Dependence of probability that the three best beer samples in a competition would win a medal (gold, silver, bronze) on the relative difference of the three samples from the rest of the samples in a competition (when the problematic practices occur – based on Monte Carlo simulation).

**Table 1** Values for organizing the individual rounds according to the total number of beer samples

Total number of samples	Number of samples in qualification (y values)	Number of assessors in qualification groups	Number of advancing samples (x values)	Number of samples in the semi-final (i values)	Number of advancing samples (t values)	Number of samples in the final
6	-	-	-	-	-	6
7	-	-	-	4; 3	3; 2*	6
8	-	-	-	4; 4	3; 3	6
9	-	-	-	5; 4	3; 3	6
10	-	-	-	5; 5	3; 3	6
11	-	-	-	6; 5	3; 3	6
12	-	-	-	6; 6	3; 3	6
13	6; 7	12; 12	5; 6*	6; 6	3; 3	6
14	7; 7	12; 12	6; 6	6; 6	3; 3	6
15	5; 5; 5	8; 8; 8	4; 4 <sup>a</sup> ; 4	6; 6	3; 3	6
16	4; 4; 4; 4	6; 6; 6; 6	3; 3; 3; 3	6; 6	3; 3	6
17	4; 4; 4; 5	6; 6; 6; 6	3; 3; 3; 3	6; 6	3; 3	6
18	4; 4; 5; 5	6; 6; 6; 6	3; 3; 3; 3	6; 6	3; 3	6
19	4; 5; 5; 5	6; 6; 6; 6	3; 3; 3; 3	6; 6	3; 3	6
20	5; 5; 5; 5	6; 6; 6; 6	3; 3; 3; 3	6; 6	3; 3	6
21	5; 5; 5; 6	6; 6; 6; 6	3; 3; 3; 3	6; 6	3; 3	6
22	5; 5; 6; 6	6; 6; 6; 6	3; 3; 3; 3	6; 6	3; 3	6
23	5; 6; 6; 6	6; 6; 6; 6	3; 3; 3; 3	6; 6	3; 3	6
24	6; 6; 6; 6	6; 6; 6; 6	3; 3; 3; 3	6; 6	3; 3	6

\* The lowest ranked samples from both qualifying groups are evaluated among each other and the higher ranked sample will advance to the semi-finals;  
a The first two samples advance to semi-final group A, the third and fourth samples advance to semi-final group B.

## 2.2 Procedure

Before the beginning of the competition, the 24 assessors and all beer samples are randomly distributed into groups in the lowest round in a given category (two, three, or four groups given in Table 1). This distribution can be done by a random draw which also determines the sequence of the beer samples in each group. The sensory evaluation performed by the assessors is based on a ranking test with a focus on the total sensory quality of beer samples. This principle of evaluation applies to all rounds in the competition. Data from each group in a given round are processed by the Bayesian approach (described in sections 2.3 and 2.4) and the posterior conditional probabilities are calculated. These probabilities determine samples advancing to the next round. Such samples are then sensorially evaluated in the next round – samples and assessors are assigned to groups according to the blue and red arrows in Figure 2, respectively. When the total number of samples is equal to 15, three groups in the qualification round are assembled and the distribution of advancing samples to the semifinal groups is also indicated in Table 1.

In case of the total number of samples is 7 or 13, an additional round is organized (see Table 1, indicated by asterisks) to preserve the maximal number of samples in the semifinal (6 in each group) or final (6) round, respectively.

In case of identical results (posterior conditional probabilities) in the qualification of the semifinal, the given samples are additionally sensorially evaluated by the 24 assessors and the better one is determined in order to select which one will advance to the next round.

The final results of the competitions are based only on data obtained from the sensory evaluation in the final round. The final ranking is not determined for samples that did not advance to the final round. The identical results (probabilities) for some beer samples in the final round mean identical final ranking.

## 2.3 Data evaluation – theory

Data evaluation is based on the application of the probability theory, namely the Bayesian theorem (Bertsekas and Tsitskilis, 2008; Oijen, 2020), as

stated in the previous sections. The principle could be demonstrated by a simple toy example. Imagine two bags with red and blue balls (see Figure 3). The bag number 1 contains three blue balls and three red balls. The bag number 2 contains one blue ball and five red balls. A man (let’s call him Thomas – after Thomas Bayes, the father of Bayesian statistics) randomly selects one of the bags. Now, he would like to know which of the bags he selected (without directly seeing the contents of the bag). Therefore, he made a random draw of three balls – three red balls were drawn. By using the Bayesian theorem in Figure 3, Thomas can calculate the posterior conditional probabilities that he selected bag 1 or bag 2 given his observation from the random draw. This probability directly tests the hypotheses (hypothesis 1: bag 1, hypothesis 2: bag 2) according to our data, it also does not assume unreal hypothetical situations on the side of data as it is common in frequentistic statistics. In Figure 3, it can be seen that the conditional probabilities are 0.09 and 0.91 for bag 1 and bag 2, respectively. Hence, there is quite a high probability (91%) that the selected bag was bag 2.

the best beer sample in the competition with gold, silver, and bronze medals (three best beers), hypotheses that a given trio of samples are better and, therefore, evaluated as better than the rest of samples, can be postulated. Such hypotheses are specified for all possible combinations. The hypotheses indicate that given samples in a given combination are twice likely to be evaluated as the best samples in a ranking test in comparison to the rest of the samples in a group. The number of combinations (hypotheses) is calculated by equation 1.

$$N = \frac{n!}{k!(n-k)!} \tag{1}$$

Where:  $N$  ... is a total number of combinations,  
 $n$  ... is a total number of samples in a given group,  
 $k$  ... is a number of advancing samples from a given group (specified in Table 1).

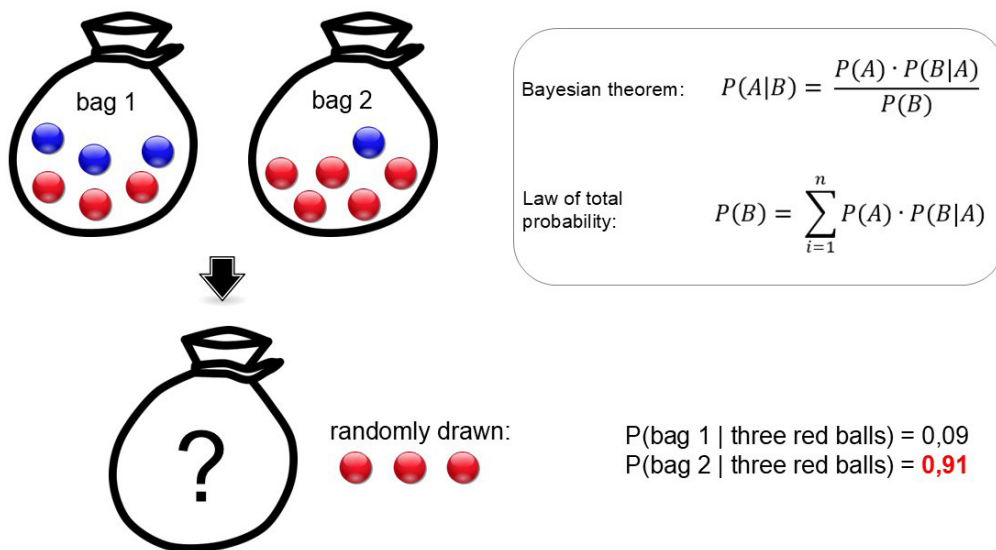


Figure 3 A toy example of Bayesian theorem usage for a basic explanation of the data evaluation in the proposed beer competition.

The same principle can be also used for the evaluation of data in a beer competition. However, in such case, we do not have any information similar to the knowledge of the ball distribution in the bags like in the toy example in Figure 3. Therefore, we need to specify some hypotheses which would act as a basis for the evaluation by the Bayesian theorem. As there is a usual purpose to award

In some specific cases, such hypotheses can also be postulated for groups of any other number of samples. Conditional probability of the hypotheses is calculated according to the Bayesian theorem, equation 2.

$$P(H_x|D) = \frac{P(H_x) \cdot P(D|H_x)}{P(D)} \tag{2}$$

Where:  $P(H_x|D)$  ... is posterior conditional probability of hypothesis  $x$  ( $x=1,2, \dots, N$ ) given the obtained data,  
 $P(H_x)$  ... is aprior probability of hypothesis  $x$ ,  
 $P(D|H_x)$  ... is conditional probability of obtained data given the hypothesis  $x$ ,  
 $P(D)$  ... is total probability of obtained data.

The total probability of obtained data is calculated according to law of total probability, equation 3.

$$P(D) = \sum_{i=1}^n P(H_x) \cdot P(D|H_x) \tag{3}$$

The conditional probability of obtained data given in the hypothesis  $x$ ,  $P(D|H_x)$ , is calculated according to equation 4 and values from table of hypotheses (an example of the table is shown in section 2.4, Table 4).

$$P(D|H_x) = \prod_{h=1}^{As} \frac{I_{x1h}}{Q} \cdot \frac{I_{x2h}}{Q-I_{x1h}} \cdot \dots \cdot \frac{I_{x(n-1)h}}{Q-\sum_{s=1}^{n-1} I_{xsh}} \tag{4}$$

Where:  $As$  ... is a total number of assessors in a given group,  
 $h$  ... is an assessor's number,  
 $s$  ... is a ranking of the sample in the ranking test by assessor  $h$ ,  
 $I_{x1h}$  ... is a value from table of hypotheses (for hypothesis  $x$  and a sample ranked by the assessor  $h$  at the first place).

Equivalently,

$I_{x2h}$  ... is a value for hypothesis  $x$  and a sample ranked by the assessor  $h$  at the second place.

Hence,  $I_{x(n-1)h}$  ... is a value for hypothesis  $x$  and a sample ranked by the assessor  $h$  at  $n-1$  place.

$Q$  ... is sum of values in the table of hypothesis across all samples.

The aprior probability of hypothesis  $x$ ,  $P(H_x)$ , is calculated by uniform distribution (all of the hypotheses have the same aprior probability), equation 5.

$$P(H_x) = \frac{1}{N} \tag{5}$$

By combination of equations 2, 3, and 4, posterior conditional probability  $P(H_x|D)$  can be calculated. Hypotesis with the highest  $P(H_x|D)$  determines samples advancing to next round of the competition. The final results of the competition is calculated by equation 3 from  $P(H_x|D)$  obtained from data in final round.

### 2.4 Data evaluation - illustrative example

In order to illustrate usage of the theoretical part of the evaluation described in the previous section in practice, an example of a fictitious competition was defined – 4 beer samples (A, B, C, D), 6 assessors, and 3 advancing samples.

The ranking of the beer samples by individual assessors is in Table 2.

**Table 2** Ranking of beer samples by individual assessors in the illustrative example

	1	2	3	4
Assessor 1	A	B	C	D
Assessor 2	A	B	C	D
Assessor 3	C	D	B	A
Assessor 4	C	D	B	A
Assessor 5	B	A	D	C
Assessor 6	D	C	A	B

The number of hypotheses can be calculated by equation 1:

$$N = \frac{4!}{3!(4-3)!} = 4$$

Hence, the total number of combinations of advancing samples is 4, the combinations are listed in Table 3.

**Table 3** Hypotheses and their combinations of advancing samples

$H_x$	combination
$H_1$	ABC
$H_2$	ABD
$H_3$	ACD
$H_4$	BCD

According to these combinations, the table of hypotheses with values for equation 4 can be constructed in a way so that the values for advancing samples in a given hypothesis are 2, and values of 1 for the rest of the samples, Table 4.

**Table 4** Values for  $P(D|Hx)$  calculation by equation 4 - illustrative example

	A	B	C	D	sum
H1	2	2	2	1	7
H2	2	2	1	2	7
H3	2	1	2	2	7
H4	1	2	2	2	7
index of sample	1	2	3	4	-



Aprior probabilities  $P(H_x)$  are 0.25 for all the hypotheses in this example (calculated according to the equation 5). The calculation of the conditional probabilities  $P(D|H_x)$  according to equation 4 and values from Table 4:

$$P(D|H_1) = \frac{2}{7} \cdot \frac{2}{5} \cdot \frac{2}{3} \cdot \frac{2}{7} \cdot \frac{2}{5} \cdot \frac{2}{3} \cdot \frac{2}{7} \cdot \frac{1}{5} \cdot \frac{2}{4} \cdot \frac{2}{7} \cdot \frac{1}{5} \cdot \frac{2}{4} \cdot \frac{2}{7} \cdot \frac{2}{5} \cdot \frac{1}{3} \cdot \frac{1}{7} \cdot \frac{2}{6} \cdot \frac{2}{4} = 4.2982 \cdot 10^{-9}$$

And equivalently,

$$P(D|H_2) = 1.7909 \cdot 10^{-9}$$

$$P(D|H_3) = 2.1491 \cdot 10^{-9}$$

$$P(D|H_4) = 3.5818 \cdot 10^{-9}$$

Calculation of the total probability  $P(D)$  according to equation 3:

$$P(D) = 0,25 \cdot 4.2982 \cdot 10^{-9} + 0,25 \cdot 1.7909 \cdot 10^{-9} + 0,25 \cdot 2.1491 \cdot 10^{-9} + 0,25 \cdot 3.5818 \cdot 10^{-9} = 2.9550 \cdot 10^{-9}$$

Obtained probabilities  $P(H_x)$ ,  $P(D|H_x)$ , and  $P(D)$  are put into equation 2 and the posterior probabilities of each hypothesis with respect to given data are obtained:

$$P(H_1|D) = \frac{0,25 \cdot 4.2982 \cdot 10^{-9}}{2.9550 \cdot 10^{-9}} = 0.3636$$

$$P(H_2|D) = \frac{0,25 \cdot 1.7909 \cdot 10^{-9}}{2.9550 \cdot 10^{-9}} = 0.1515$$

$$P(H_3|D) = \frac{0,25 \cdot 2.1491 \cdot 10^{-9}}{2.9550 \cdot 10^{-9}} = 0.1818$$

$$P(H_4|D) = \frac{0,25 \cdot 3.5818 \cdot 10^{-9}}{2.9550 \cdot 10^{-9}} = 0.3030$$

According to these probabilities, hypothesis  $H_1$  is most probable and, therefore, samples A, B, and C advance to the next round.

In case the illustrative example would represent the final round, the calculation of the final results would be based on total probability  $P(D)$ , equation 3. Where  $P(H_x)$  is equal to posterior conditional probability  $P(D|H_x)$  calculated above, and conditional probability  $P(D|H_x)$  is calculated from values in Table 4 as a ratio of a value for a given sample and a given hypothesis to the sum of values for a given hypothesis.

$$P(A) = 0.3636 \cdot \frac{2}{7} + 0.1515 \cdot \frac{2}{7} + 0.1818 \cdot \frac{2}{7} + 0.3030 \cdot \frac{1}{7} = 0.2424$$

$$P(B) = 0.3636 \cdot \frac{2}{7} + 0.1515 \cdot \frac{2}{7} + 0.1818 \cdot \frac{1}{7} + 0.3030 \cdot \frac{2}{7} = 0.2597$$

$$P(C) = 0.3636 \cdot \frac{2}{7} + 0.1515 \cdot \frac{1}{7} + 0.1818 \cdot \frac{2}{7} + 0.3030 \cdot 2 = 0.2641$$

$$P(D) = 0.3636 \cdot \frac{1}{7} + 0.1515 \cdot \frac{2}{7} + 0.1818 \cdot \frac{2}{7} + 0.3030 \cdot \frac{2}{7} = 0.2338$$

According to the final calculation, sample C won a gold medal, sample B won a silver medal, and sample A won a bronze medal.

## 2.5 Evaluation of the proposed methodology by Monte Carlo simulation

The procedure described in the previous sections was computationally simulated to evaluate its properties. The Monte Carlo simulations were performed by in-house programmed simulations in RStudio version 1.1.456 (R version 4.0.1) with 100 000 iterations. The whole competition procedure, 24 samples, and 24 assessors were incorporated. In the simulation, the best three samples were defined (simulated by their higher probability of selection by the assessors – this probability was randomly modified to evaluate the effect of the relative difference of these samples from

the rest of the samples), and after the simulation, the probability that these samples would win one of medals (gold, silver or bronze) was computed.

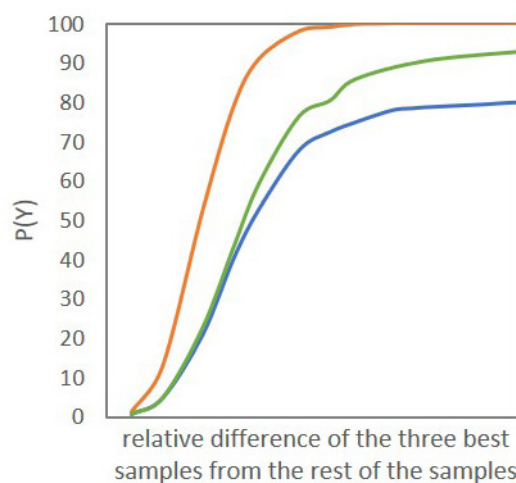
In the same way, the second set of simulations was performed for a different variation of the competition procedure including the problematic practices mentioned in the Introduction section. In brief, this procedure was designed as follows: samples were randomly distributed into four groups in the first round of the competition; assessors evaluated the samples in a rank test (simulated by the probabilities of selection mentioned in the previous paragraph); the obtained data were processed by the sum of rankings for each sample across the assessors; three samples with the lowest sums in each group advanced to the second round of the competition (tied scores were assessed by the problematic practice described in the Introduction section); in the second round, samples and assessors were divided into two groups and the samples were evaluated in the same way as in the first round; the sums of rankings were compared between the two groups in the second round – the lower the sum, the better ranking in the second round; and finally, the final results were obtained by summation of scores for each sample (based on the ranking in the first and the second round) – see [Table 5](#).

**Table 5** Scoring of the samples in the variation of the competition with problematic practices – for simulation only

total ranking	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. round	30	20	15	10	5	3	–	–	–	–	–	–
2. round	40	35	30	26	23	20	17	14	12	10	5	0

The scoring shown in [Table 5](#) is an example of the inappropriate way mentioned in the Introduction section since it gives an advantage to samples with a combination of total ranking in the first and the second round at 1. and 2. place, respectively, over samples with a combination of total ranking in the first and the second round at 2. and 1. place. The third set of simulations was performed in the same way as the previous one, however, the scoring was removed and the final results were only calculated from the total rankings in the second round.

The results of the simulations are in [Figure 4](#) in the form of dependency of probability that the best three samples in a competition would win a medal on the relative difference of the three samples from the rest of the samples in a competition. These results clearly show that the problematic practices included in the second set of simulations (blue curve) are able to ruin the results of



**Figure 4** Dependence of probability that the three best beer samples in a competition would win a medal (gold, silver, bronze) on the relative difference of the three samples from the rest of the samples in a competition (computed by Monte Carlo simulations). Blue – procedure with the problematic practices; green – the procedure with the problematic practices but without scoring; orange – the methodology proposed in this paper.

the competition (as already shown and described in the Introduction section and [Figure 1](#)). Removing the scoring system (described in the previous paragraph and [Table 5](#)) led to a slight increase in the maximal probability in [Figure 4](#) (green curve) – the other problematic practices are still negatively influencing the results. On the other hand, the probability in [Figure 4](#) is limited to 100% for the procedure proposed in this paper (orange curve). Also, the curve is steeper in comparison to the blue and green ones – the procedure is more sensitive. It means that the proposed probabilistic procedure (without the problematic practices) is able to identify the best samples in a competition correctly (when their relative difference from the rest of the samples is high enough). Further, it can identify them with higher probabilities when the difference is lower; in a real-world situation, the differences among beer samples in beer competitions are usually lower.



### 3 Conclusion

The procedure of beer competition based on a probabilistic approach and elimination of the problematic procedures described in this paper is able to provide meaningful results and award beer samples which were evaluated by the assessors as the best samples (since it does not bring unnecessary interference or confusion to the data). The procedure is specific for its ruggedness, correctness from the point of view of the data properties, and also it eliminates tied results. The basic principle of the procedure can also be used for competitions with different numbers of assessors and samples other than specified in section 2 Principle and procedure of the methodology, as well as, in competitions with different designs. However, the authors of this paper suggest not exceeding the ratio of samples to assessors more than 1.2 since probability in Figure 4 would decrease.

### 4 Acknowledgement

The study was supported by the Ministry of Agriculture of the Czech Republic within the Institutional Support MZE-R01923.

### 5 References

- Olšovská, J., Čejka, P., Štěrba K., Slabý, M., Frantík, F. (2017). Senzorická analýza piva. Výzkumný ústav pivovarský a sladařský, a. s., Prague. ISBN 978-80-86576-74-9
- Frantík, F., Čejka, P., Čapková, V. (2005). Beer Competition in the Czech Republic. *Kvasny prumysl*, 51(10), 11–12. <https://doi.org/10.18832/kp2005018>
- Frantík, F., Čejka, P. (2001). Způsoby hodnocení piv při soutěžích v České republice. *Kvasny prumysl*, 47(2), 34–38. <https://doi.org/10.18832/kp2001004>
- ISO 5492:1992. Sensory analysis – Vocabulary.
- ISO 6658:2005. Sensory analysis – Methodology – General guidance.
- ISO 8586-1:1993. Sensory analysis – General guidance for the selection, training and monitoring of assessors – Part 1: Selected assessors.
- ISO 8586-2:1994. Sensory analysis – General guidance for the selection, training and monitoring of assessors – Part 2: Experts.
- Analytica EBC 13.4. Sensory Analysis: Selection and Training of Assessors (1997). Available online: <https://brewup.eu/ebc-analytica/sensory/sensory-analysis-selection-and-training-of-assessors-im-am/13.4> [accessed on 25 July 2024]
- Analytica EBC 13.11. Sensory Analysis: Ranking Test (1997). Available online: <https://brewup.eu/ebc-analytica/sensory/sensory-analysis-ranking-test-im/13.11> [accessed on 25 July 2024]
- Analytica EBC 13.13. Sensory Analysis: Routine Descriptive Test Guideline (2004). Available online: <https://brewup.eu/ebc-analytica/sensory/sensory-analysis-routine-descriptive-test-guideline/13.13> [accessed on 25 July 2024]
- Bertsekas, D.P., Tsitskilis, J.N. (2008). Introduction to Probability. 2nd Edition, Athena Scientific, p.528. ISBN 978-1-886529-23-6
- van Oijen, M. (2020). Bayesian Compendium. 1st Edition, Springer Cham, p. 204. ISBN 978-3-030-55896-3 <https://doi.org/10.1007/978-3-030-55897-0>