Statistical analysis of the conveyor milking model

T Borodina¹*, V Zorin¹, S Strebulyaev², O Tareeva³, N Shkilev⁴ and S Kazakov³

¹Department “Software Engineering”, Lobachevsky State University of Nizhny Novgorod, 23 Gagarina Prospekt, 603950, Nizhny Novgorod, Russian Federation
²Department “Differential Equations, Mathematical and Numerical Analysis”, Lobachevsky State University of Nizhny Novgorod, 23 Gagarina Prospekt, 603950, Nizhny Novgorod, Russian Federation
³Department “Technical and Biological Systems”, Nizhny Novgorod State Engineering and Economic University, 22 а Oktyabrskaya Street, 606340, Knyaginino, Russian Federation
⁴Center for Support of Sustainable Development of Rural Areas, Nizhny Novgorod State Engineering and Economic University, 22 а Oktyabrskaya Street, 606340, Knyaginino, Russian Federation

*E-mail: borodina@itmm.unn.ru

Abstract. The purpose of the study is a computer analysis of the milking process mathematical model as well as the influence of statistical characteristics allowing developing adaptive algorithms to control the process of rotary conveyor milking on the “Carousel” lines. The methods of probability theory and mathematical statistics are used in the research. When conducting a computational experiment, the Maple analytical computing system was used. Using of Pearson’s chi-squared test sets the gamma distribution of the individual milking time and speed for each cow with acceptable reliability. One of the Morgenstern-Gumbel models is used as a first approximation of the joint distribution of time and speed of the rotary conveyor milking. It was found that this model does not contradict the sample data. The authors believe that it is necessary to conduct further research using statistical sequential analysis to reduce the amount of sample cows from the herd and to refine the estimates of the model parameters. It will lead to improving the algorithm for adaptive control of the “Carousel” rotation speed and will increase the productivity of the milking process in the rural areas of Russia and abroad.

1. Introduction

For many years, scientists and designers have tried to bring the milking process closer to the technological lines of industrial enterprises. This problem is partially solved via the development and operation of conveyor ring milking machines. There is a developed deterministic algorithm for adaptive control of the platform rate angular rotation depending on the each cow milking time. It takes into account the conditions for compensation of abnormal milking cycles of individual animals in order to optimize the milking time, the number of milking places, as well as to exclude the conveyor and animals’ downtime.

One of the features of the milking process on carousel milking machines (figure 1a-d) is a continuous and uniform flow of cows that is created from the storage unit to the outlet. It makes possible achieving the highest productivity, on average 5.5 heads/h per cattle place in comparison to other milking systems.
Figure 1. Schemes of milking units with movable machines: (a) – conveyor with a radial arrangement of cows; (b) – Morsotto conveyor; (c) – conveyor with a sequential arrangement of cows; (d) – conveyor with an oblique arrangement of cows.

The process of servicing each cow during milking includes the following parameters:

\[ t_c = t_{in} + t_{po} + t_m + t_{st} + t_{fo} + t_{ex} \]  \hspace{1cm} (1)

- \( t_c \) – service cycle time of each cow, min;
- \( t_{in} \) – time of admission of the animal to the platform, min;
- \( t_{po} \) – time of preparatory operations, min;
- \( t_m \) – milking time, min;
- \( t_{st} \) – downtime for various reasons, min;
- \( t_{fo} \) – time of final operations, min;
- \( t_{ex} \) – time of release of a cow from the platform, min.

The milking time \( (t_m) \) of each cow depends on its physiological characteristics, position in the life cycle and a number of other parameters and affects the performance of milking carousels [1]. A number of authors have been engaged in research in the field of optimization and statistical analysis of the described process [2-8].

However, to create an adequate algorithm for milking machine controlling, it is necessary to take into account that the parameters of the milking time distribution, as well as the amount of samples to which they relate, always change during the milking process. It represents the scientific innovation of this research. Based on the physics of the problem, hypotheses about the gamma distribution, normal distribution, and logarithmically normal distributions were formulated separately for each of the obtained arrays of milking time and speed. It is established that the gamma distribution is the most preferable for the milking characteristics (speed and time separately).

It seems natural to us to fix two characteristics simultaneously for each cow at once: the speed and time of milking, so it is necessary to consider the joint distribution of the speed and time of milking in
3D space. In this paper, one of the Morgenstern-Gumbel models was used for the joint distribution of the time and speed of the rotary conveyor milking process. An element of novelty is also the use of a 3D model allowing description of the correlation between the characteristics under consideration. Also, it allowed obtaining new statistical data for each cow, which will be taken into account in the adaptive algorithm for controlling the speed of rotation of the “Carousel”, that is the purpose of this work.

The conducted research and methodology makes it possible to introduce a unified approach to the study of rotary conveyor milking for more successful development of rural areas both in Russia and abroad.

2. Methodology
The data on the time and speed of milking in the morning and afternoon for several days, obtained in LLC “Zhdanovsky” of the Nizhny Novgorod region, were used as the materials. The data on the time and speed of morning milking for one of the days which are used in this work are shown in figure 2 and figure 3, respectively.

Based on the above graphic images, it is quite difficult to make an assumption about the types of distributions that adequately describe the process under study. In this regard, hypotheses about the gamma distribution, normal distribution, and logarithmically normal distribution were formulated separately for each of the obtained arrays of milking time and speed. Afterwards, histograms and density graphs were plotted for milking characteristics (figures 4 and 5). When performing this procedure on a computer, specially developed software was used.

From the analysis of these graphs, it can be concluded that the most suitable graphs are the density graphs of the gamma distribution, which is confirmed of statistical values of Pearson’s chi-squared test
\( \chi^2 \) in the table 1 below.

![Figure 4](image1.png) **Figure 4.** Different types of distribution densities for the milking time.

![Figure 5](image2.png) **Figure 5.** Different types of distribution densities for the milking speed.

The statistical values \( \chi^2 \) were compared with the critical values \( \chi^2_{cr} \) for the number of degrees of freedom equal to three. Thus, the application of the Pearson’s chi-squared test establishes with acceptable reliability the gamma distribution of separate milking time and separate milking speed for each cow.

**Table 1.** The statistics values of the Pearson’s chi-squared test for various hypotheses.

| Types of distributions | Milking time, \( T \), s | Milking speed, \( V \), g/s | Milking time, \( T \), s | Milking speed, \( V \), g/s |
|------------------------|--------------------------|-----------------------------|--------------------------|-----------------------------|
| gamma                  | 4.65                     | 1.78                        | 10.80                    | 7.12                        |
| normal                 | 15.06                    | 2.44                        | 20.99                    | 1.78                        |
| logarithmically normal | 82.83                    | 77.70                       | 80.89                    | 80.23                       |

Next, we fix two characteristics for each cow at once: milking speed and time. First, it is well known that the joint distribution function contains information that is not deduced from the particular distribution laws, and second, we assume that there is a statistical relationship between the variables (milking speed \( - V \), g/s and milking time \( - T \), s), which it needs to be studied.

The papers of McKay A, Dussauchoy A and Berland R, Cherian K, Arnold B and Strauss D, Becker P and Roux J, Smith O and Adelfang S, Nadarajah S and other scientists are devoted to the multidimensional gamma distribution [9-11]. The joint probability distribution densities \( p(x,y) \) of some multidimensional gamma distributions are given below, namely:

1. McKay’s joint distribution density:

   \[
   p(x,y) = \frac{b^{pq}}{\Gamma(p)\Gamma(q)} (y-x)^{p-1} x^{q-1} \exp(-by) 
   \]  
   \[ (2) \]

   for \( y > x > 0, b > 0, p > 0, q > 0 \).

2. Dussauchoy and Berland’s joint distribution density:
\[ p(x, y) = \frac{\rho a^q_{2}}{\Gamma(p) \Gamma(q-p)} (\rho x)^{p-1} (y-\rho x)^{q-p-1} \exp(-a x) \exp \left( -\frac{a}{\rho} (y-\rho x) \right) \times \]

\[ \times H \left( p; q-p; \left( \frac{a_1}{\rho} - a_2 \right) (y-\rho x) \right) \] (3)

for \( x > 0, y > 0, 0 \leq \rho \leq 1, 0 < a_2 \leq a_1/\rho, 0 < p < q \).

3. Cherian’s joint distribution density:

\[ p(x, y) = M \exp(-x) \exp(-y) \min(x, y) \int_{0}^{\min(x, y)} \exp(t)(x-t)^{\theta_1-1}(y-t)^{\theta_2-1}I_{\theta_3-1}dt \] (4)

for \( x > 0, y > 0, \theta_1 > 0, \theta_2 > 0, \theta_3 > 0 \).

Special functions are used in (2)-(4):

\[ \Gamma(x) = \int_{0}^{\infty} t^{x-1} e^{-t} dt \] (5)

\( \Gamma(x) \) is gamma function;

\[ H(\alpha; \beta; x) = \sum_{k=0}^{\infty} \frac{(\alpha)_k x^k}{(\beta)_k k!} \] (6)

\( H(\alpha; \beta; x) \) is confluent hypergeometric function and \( M \) is normalizing constant given by

\[ \frac{1}{M} = \Gamma(\theta_1) \Gamma(\theta_2) \Gamma(\theta_3) \] (7)

In this article, we consider the following general problem formulated by Frechet M: find a class of two-dimensional distribution functions \( \{P(x, y)\} \), having given partial (marginal) one-dimensional distributions \( F(x) \) and \( F_1(y) \).

Gumbel suggested the following example of functions included in the Frechet class \( K\{F(x), F_1(y)\} \) for any given \( F(x) \) and \( F_1(y) \):

\[ P(x, y) = F(x) F_1(y) \left[ 1 + \lambda (1-F(x))(1-F_1(y)) \right], \ -1 < \lambda < 1 \] (8)

In the case of existing densities denoted by the corresponding lowercase letters, we have

\[ p(x, y) = f(x) f_1(y) \left[ 1 + \lambda (2F(x)-1)(2F_1(y)-1) \right] \] (9)

For the normal case, using the notation \( \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-t^2/2} dt \), we get

\[ p(x, y) = \frac{\exp \left[ -\frac{x^2 + y^2}{2} \right]}{2\pi} \left[ 1 + 4\lambda \Phi(x) \Phi(y) \right] \] (10)
It is known that if the densities \( p_k(x, y) \) belong to a certain Frechet class, a linear combination of their form \( \sum_k \varepsilon_k p_k(x, y) \), where \( \varepsilon_k > 0 \) and \( \sum_k \varepsilon_k = 1 \) also belong to the same class. As an example

\[
p(x, y) = f(x)f_1(y) \left[ 1 + \sum_{k=1}^{\infty} \varepsilon_k \cos 2k\pi \left( F(x) - 1/2 \right) \cos 2k\pi \left( F_1(y) - 1/2 \right) \right]
\]

(11)

where \( \sum_{k=1}^{\infty} |\varepsilon_k| \leq 1 \), and the corresponding family of distribution functions is

\[
P(x, y) = F(x)F_1(y) + \frac{1}{4\pi^2} \sum_{k=1}^{\infty} \varepsilon_k \sin 2k\pi \left( F(x) - 1/2 \right) \sin 2k\pi \left( F_1(y) - 1/2 \right)
\]

(12)

In this paper, the Morgenstern-Gumbel model of the form (8)-(9) is used for the joint distribution of the milking speed \((V)\) and time \((T)\), where the random variables \(V\) and \(T\) have a gamma distribution, that is, their densities \(f(x)\) and \(f_1(y)\) are represented as

\[
f(x) = f_1(y) = \begin{cases} 
0, & at \ x \leq 0 \\
\frac{\lambda^a x^{a-1}}{\Gamma(a)} e^{-\lambda x}, & at \ x > 0
\end{cases}
\]

(13)

where \( \lambda > 0 \), \( \alpha > 0 \) and \( \Gamma(a) \) given in (5).

Note that the parameters \( \lambda \) and \( \alpha \) are different for the variables \(V\) and \(T\), and their estimates are found by the method of moments.

It should also be noted that in classical mathematical statistics, it is established that maximum likelihood estimates (MLE) are usually better (in terms of efficiency) than estimates based on the method of moments. However, the maximum likelihood method in practice is often connected to calculation difficulties. To find the MLE for samples from the gamma distribution, it is necessary to apply numerical methods, so in this paper we use the method of moments, with which the estimates of the parameters in (13) are easily calculated. Generally speaking, the choice of the “best” estimates of the model parameters (8)-(9) requires additional research. In the future, the authors plan to use the method of sequential analysis to solve this problem, and use the method of moments as an initial approximation.

To check whether the Morgenstern-Gumbel model corresponds to real data, the Pearson chi-squared test is used.

If observations are quite expensive, then the criteria for testing a statistical hypothesis may be advantageous when the hypothesis can be accepted from a small number of observations, and rejected from a larger sample. Such is the r-fold chi-square criterion, which generalizes the usual Pearson chi-squared test criterion. Its essence is as follows.

Let there be independent tests with \(m\) results. According to the main hypothesis \(H_0\), we will assume that the probability of \(j\) result equals \(p_j\), \(j=1,.., m\). Let \(n_1 < n_2 < ... < n_r\) are increasing volumes of consecutive, nested samples. We introduce random variables \(\mu_j^{(k)}\), assuming that \(\mu_j^{(k)} = 1\), if at \(k\) testing \(j\) result occurred, and \(\mu_j^{(k)} = 0\) in all other cases. Denote by \(v_j = \sum_{k=1}^{n_j} \mu_j^{(k)}\) the number of \(j\) results in \(n_i\) initial tests.

By volume selection \(n_i\) make the statistics:
\[
\chi_i^2 = \sum_{j=1}^{n} \left( \frac{(v_j - n_j p_j)^2}{n_j p_j} \right)
\tag{14}
\]

The suggested r-fold criterion \(\chi^2\) is plotted using vector statistics \(\chi^2 = (\chi_1^2, \chi_2^2, \ldots, \chi_r^2)\) as follows.

Hypothesis \(H_0\) is accepted if one of the following events occurs, \(A_k, k = 1, 2, \ldots, r\): \[A_k = \{\chi_1^2 > \chi_{1,cr}^2, \chi_2^2 > \chi_{2,cr}^2, \ldots, \chi_k^2 > \chi_{(k-1),cr}^2, \chi_k^2 \leq \chi_{cr}^2\}.\] If \(\chi_k^2 > \chi_{cr}^2\) at all \(k = 1, 2, \ldots, r\), then hypothesis \(H_0\) is rejected.

Thus, the error of the first kind \(\alpha\), that is, the probability of the main hypothesis \(H_0\) rejecting, when it is true, equals to \(\alpha = P\{\chi_1^2 > \chi_{1,cr}^2, \chi_2^2 > \chi_{2,cr}^2, \ldots, \chi_r^2 > \chi_{cr}^2 | H_0\}\). The error of the second kind \(\beta\), the probability of the main hypothesis \(H_0\) accepting, when some competing hypothesis \(H_1\) is true equals to \(\beta = 1 - P\{\chi_1^2 > \chi_{1,cr}^2, \chi_2^2 > \chi_{2,cr}^2, \ldots, \chi_r^2 > \chi_{cr}^2 | H_1\}\).

3. Results and discussion

We focus in more detail on the main stages of developing and verifying the adequacy of the Morgenstern-Gumbel model \((8)-(9)\), applied to the distribution of time and speed of the rotary conveyor milking process. The steps of the algorithm for checking the Morgenstern-Gumbel model for adequacy are set below:

- **Step 1.** Formulating the initial hypothesis \(H_0\). To test the hypothesis \(H_0\) that the system of random variables \((V, T)\) obeys a certain distribution law given by the distribution function \(F_0(x, y)\) of the type \((8)\), which is \(H_0: F(x, y) = P(x, y)\) against the alternative hypothesis \(H_1: F(x, y) \neq F_0(x, y) = P(x, y)\). Pearson’s chi-squared test is applied.

- **Step 2.** Estimation of model parameters based on sample data. Based on the assumptions made above about the separate distribution of the milking speed and time, we first found parameters estimates of the corresponding gamma distributions based on sample data obtained at “Zhdanovsky” LLC in the Nizhny Novgorod region. According to the method of moments, a system of algebraic equations was solved:

\[
\begin{align*}
\frac{a}{\lambda} &= \bar{X} \\
\frac{a}{\lambda^2} &= S^2
\end{align*}
\tag{15}
\]

where \(\lambda, a\) – parameters of the gamma distribution at \((13)\), \(\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i\) and \(S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2\), and \(X\) – variable that characterizes the observed value \((V\ or\ T)\). It is received that \(V \in \Gamma\left(8.006; 0.3593\right)\), and \(T \in \Gamma\left(7.448; 0.0174\right)\). A graphical 3D representation of the joint distribution of milking characteristics is shown in figure 6.
Figure 6. Program fragment: a graph of the joint distribution of milking characteristics.

- Step 3. Splitting the range of variable values into a system of rectangles. At this stage, the entire range of variable values \( (V, T) \) was divided into a system of rectangles (figure 7).

Figure 7. Values of variables \( (V, T) \) and the system of rectangles.

In figure 7, the dots mark the boundaries of the rectangular areas, and the asterisks mark the cows according to their speed and milking time.

- Step 4. Calculation of the probabilities of falling into rectangular areas (provided that the initial hypothesis is valid). The calculated probabilities of falling rectangular areas (figure 7) are given in the matrix (figure 8).
• Step 5. Calculation of theoretical frequencies. At the fifth step, the theoretical frequencies of falls in each of the rectangular areas were calculated and compared with the observed frequencies.
• Step 6. Calculating chi-square statistics. The value of the chi-square criterion was calculated from the sample data. It is obtained that the value of $\chi^2 = 45.95$.
• Step 7. Finding the critical value of the chi-square criterion. Critical value of the criterion $\chi^2_{cr} = 46.98$ for the significance level $\alpha = 0.025$ and the number of degrees of freedom, equal to 30, obtained from special tables of mathematical statistics.
• Step 8. Comparison of the observed chi-square criterion value with the critical value. Since the calculated value $\chi^2 = 45.95$ is less than the critical value $\chi^2_{cr} = 46.98$, then the hypothesis $H_0$ does not contradict the sample data, and we can use the Morgenstern-Gumbel model for the joint distribution of the of the rotary conveyor milking characteristics.

The research shows the possibility of using the Morgenstern-Gumbel model to describe the joint distribution of two variables – the speed and the time of milking. The peculiarity of this model is that both of the above variables are taken into account simultaneously, and this allows us to find a correlation between them. Further studies are needed to reduce the amount of sample cows from the herd and to refine the parameters estimates of the of the rotary conveyor milking model. The authors plan to apply the method of statistical sequential analysis, which will let improve the algorithm for the adaptive control of the “carousel” rotation speed.

4. Conclusion
It is possible to more accurately characterize the specifics of the dairy herd by choosing the correct hypothesis of the distribution of the studied processes. This stage of research allows the development of an adaptive algorithm for controlling the platform rotation speed, which would increase the productivity of the milking process, taking into account the number of cows and the capabilities of the “Carousel”.

The research results can be used by developers to create more productive rotary conveyor lines for milking cows and other animals. At the same time, it will be necessary to equip the rotary conveyor lines with more modern computing equipment providing sufficient information about the cows, including for zootechnical and veterinary purposes. At this stage of digitalization, it is also necessary to automate the milking process using robotics. All this will contribute to the development of rural areas both in Russia and abroad.

References
[1] Kirsanov V V, Izmaylov A Y, Lobachevsky Y P, Tareeva O A, Strebulyayev S N and Filonov R F 2019 Models and algorithms of adaptive animal flow control in rotary milking parlours. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis 67 1465 doi: 10.11118/actaun201967061465
[2] Kucheruk V Yu, Palamarchuk E A and Kulakov P I 2014 The development of statistical models
of milking duration on conveyor milking units. *East.-Eur. J. Enterp. Technol.* 4(2) 72 doi: 10.15587/1729-4061.2014.28951

[3] Klymenko I, Holovko O, Hilliaka M and Mytsio Y 2016 The development of means of definition of the optimum ratio of computational algorithm and the reconfigurable structure. *East.-Eur. J. Enterp. Technol.* 3(2) 81 doi: 10.15587/1729-4061.2016.71460

[4] Jago J G, Davis K L, Copeman P J and Woolford M M 2006 The effect of pre-milking teat-brushing on milk processing time in an automated milking system. *J. Dairy Res.* 73(2) 187 doi: 10.1017/S002202990500155X

[5] Zagidullin L R, Khisamov R R, Kayumov R R, Lomakin I V and Kanalina N M 2020 Lactive activities and the process of milking of first-calf cows using robotic milking. *BIO Web Conf.* 17 00038 doi: 10.1051/bioconf/20201700038

[6] Hogeveen H, Ouweltjes W, Koning C and Stelwagen K 2001 Milking interval, milk production and milk flow-rate in an automatic milking system. *Livest. Prod. Sci.* 72(1-2) 157 doi: 10.1016/S0301-6226(01)00276-7

[7] Nitzan R, Bruckental I, Bar Shira Z, Maltz E and Halachmi I 2006 Stochastic models for simulating parallel, rotary, and side-opening milking parlors. *J. Dairy Sci.* 89(11) 4462 doi: 10.3168/jds.S0022-0302 (06) 72495-X

[8] Stal M, Pinzke S, Hansson G A and Kolstrup C 2003 Highly repetitive work operations in a modern milking system. A case study of wrist positions and movements in a rotary system. *Ann. Agric. Environ. Med.* 10(1) 67

[9] Nadarajah S and Kotz S 2004 The beta gumbel distribution. *Math. Probl. Eng.* 2004 4 doi: 10.1155/S1024123X04403068

[10] Nadarajah S 2005 Reliability for some bivariate beta distributions. *Math. Probl. Eng.* 2005 1 doi: 10.1155/MPE.2005.101

[11] Nadarajah S 2005 Reliability for some bivariate gamma distributions. *Math. Probl. Eng.* 2005 2 doi: 10.1155/MPE.2005.151