Autocorrelation function of the poly-L-glycine

LIJUBICA VILENDECIC-GRKINIC
Gymnasium Nova Gradiska, Serbian Republic, BOSNIA AND HERCEGOVINA

Abstract: Fundamental concept of autocorrelation function was studied. Autocorrelation parameters of the poly-L-glycine formation was derived. Poly-L-glycine formation involves interaction between inner angles of rotation $\phi$ and $\psi$. Variation coefficient, mean values and autocovariation of the $\alpha$ poly-L-glycine were determined. The range of possible outcomes was determined on the basis of past experience. Dynamic series was made history data in time.

Key-Words: Autocorrelation, model, dynamics, parameters, variation, protein.

1 Introduction
Present day statistics may be defined as a tool for decision making in the light of uncertainty. Uncertainty does not imply ignorance, merely that an exact determination of the outcome of an experiment is not possible. A range of possible outcomes can often be determined on the basis of past experience or from observed sample data. Assuming the range or pattern or variability to be expected, a decision can be made as to whether or not observed data could reasonably have been taken from a population which has this assumed pattern[1]-[2]. Since random variation is present in all measurements, the real variation in an experiment must be detected in the presence of this random variation or measurement error. This is may also be defined as decision making in the light of random variation.

Statistical inference refers to the process of inferring something about a population from a sample drawn from that population. The population consist of all possible values of some random variable $Y$. The response $Y$ may represent tensile strength, weight, score, reaction time, or whatever criterion is being used to evaluate the experimental results. Characteristics of the population of this random variable are called parameters. The average or expected value of the random variable is designated as $E(Y) = \mu$. If the probability function defining the random variable is known,

$$e(Y) = \sum Y_i p(Y_i)$$ (1)

where $p(Y_i)$ is a discrete probability function, and

$$e(Y) = \int Y f(Y) dY$$ (2)

where $f(Y)$ is a continuous probability density function.

The long–range average of squared deviations from the mean of a population is called the population variance $\sigma^2_Y$.

$$e[(Y - \mu)^2] = \sigma^2_Y$$ (3)

The squares root of this parameter is called population standard deviation $\sigma_Y$. Quantities computed from the sample values drawn from the population are called sample statistics. Examples include the sample mean:

$$Y_{\text{avg}} = \frac{1}{n} \sum Y_i$$ (4)

where $n$ is the number of observations in the sample, and the sample variance:

$$s^2 = \frac{1}{n-1} \sum (Y_i - Y_{\text{avg}})^2$$ (5)

Most statistical theory is based on the assumption that samples drawn are random samples, that is each member of the population has an equal chance of being included in the sample and that the pattern of variation in the population is not changed by this deletion of the $n$ members for the sample.

The notion of statistical inference may be divided into two parts estimation and tests of hypotheses.

The objective of statistical estimation is to make an estimate of a population parameter based on a sample statistic drawn from this population. Two types of estimates are usually needed, point estimates and interval estimates.

A point estimate is a single statistic used to estimate a parameter. For example, the sample mean $Y_{\text{avg}}$ is a point estimate of the population mean $\mu$. Point estimates are usually expected to have certain desirable characteristics. They should be unbiased, consistent, and have minimum variance.

An unbiased statistic is one whose expected or average value taken over an infinite number of similar
samples equals the population parameter being estimated.

2 Autocorrelation function formulation

Autocorrelation function of probability variable $Y(t)$ characterize dependence a value in any time from a value in other time.

$$r_{yy}(t_1, t_2) = \varepsilon [Y(t_1)Y(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 p(y_1, y_2, t_1, t_2) \, dy_1 \, dy_2$$  \hspace{1cm} (6)

Autocorrelation function is defined as mathematical expectation probability system variable $Y(t)$ in time $t_1, t_2$. Autocorrelation function carries information about how system variable changing [3],[4].

According to autocorrelation property

$$r_{yy}(t_1, t_2) = \varepsilon [Y'(t_1)Y'(t_2)]$$  \hspace{1cm} (7)

the following expression for derivation can be derived.

$$r_{yy}(t_1, t_2) = \frac{\partial r_{yy}(t_1, t_2)}{\partial t_1}$$  \hspace{1cm} (8)

and

$$r_{yy}(t_1, t_2) = \frac{\partial r_{yy}(t_1, t_2)}{\partial t_2}$$  \hspace{1cm} (9)

Note that $r_{yy}(t_1, t_2)$ is not probability function.

Autocorrelation for steady-state conditions depends only for time difference $\tau = t_2 - t_1$.

$$r_{yy}(t_1, t_2) = r_{yy}(\tau) = \varepsilon [Y(t_1)Y(t + \tau)] = r_{yy}(-\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 p(y_1, y_2, \tau) \, dy_1 \, dy_2$$  \hspace{1cm} (11)

and derivations, too.

$$r_{yy}(\tau) = \frac{dr_{yy}(\tau)}{d\tau}$$  \hspace{1cm} (12)

Also, $r_{yy}(\tau) = r_{yy}(-\tau)$.

If autocorrelation function is stisfied the following conditions

$$\varepsilon[Y(t)] = \mu_Y$$

$$\varepsilon[Y(t)Y(t + \tau)] = r_{yy}(\tau)$$

Then it is weak stationary autocorrelation function.

3 Covariation formulation

Scientists very often need to estimate is there any connection between two quantity, qualitative and if possible quantitative. If common probability distribution for two quantities is known then can be calculated some measure linear dependence between them, which called correlation coefficient. At this it is the same for dependent and independent variables. Covariation function of two quantities $X(t)$ i $Y(t)$ are called a function:

$$\sigma_{XY}(t_1, t_2) = \varepsilon[(X(t_1) - \mu_X(t_1))(Y(t_2) - \mu_Y(t_2))] = \iint_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y) \cdot p(x,y;t_1,t_2) \, dx \, dy$$  \hspace{1cm} (14)

For steady state assembly covariation:

$$\sigma_{XY}(\tau) = r_{xy}(\tau) - \mu_X \mu_Y$$  \hspace{1cm} (15)

Since, covariation quantity depends from unities of $X$ and $Y$ two undimensional quantities can be introduced.

$$\frac{X - \mu_X}{\sigma_X(0)} \quad i \quad \frac{Y - \mu_Y}{\sigma_Y(0)}$$  \hspace{1cm} (16)

where argument (0) denotes $t=0$. Correlation coefficients for steady state assembly is called covariation of two undimensional quantities:

$$\gamma_{XY}(\tau) = \frac{\sigma_{XY}(\tau)}{\sigma_X(0) \cdot \sigma_Y(0)}$$  \hspace{1cm} (17)

For non-correlated $X$ and $Y$ covariation and correlation coefficient are equal zero. In oposite statement is not accurate. For example, two stochastic quantities which have Gauss normal distribution probability law can be uncorrelated but dependent from each other [1]-[2].

Correlation coefficients allow to estimates measure between two quantities by one number. Positive correlation means that $\sigma_{XY}$, mean square deviation always positive, and negative value $\sigma_{XY}$ means that

$$r_{yy}(\tau) = \frac{dr_{yy}(\tau)}{d\tau} = \frac{d^2 r_{yy}(\tau)}{d\tau^2}$$  \hspace{1cm} (13)
large value of one quantity has connected with little value of the other.

4 Protein α-poly-L-glycine
In this paper polypeptide structures and their angles of inner rotation were examined. Every amino acid residual in the main polypeptide strands formed three covalent branches, because of peptide branch σ CO-NH has character double branch. All atoms of the main polypeptide strand between two neighbors α-carbon atoms are on the same surface. R-group represent branch strands i.e. that are amino acids residuals. If these groups are equal that is homopolymers, and if these groups different then they copolymers, proteins. Proteins are biopolymers whose polymer strands consist from α-L-amino acids residuals connected interrelation by peptide branches.

Long strands some amino acids such as glycine, valyne, leucyne, and iso-leucyne have freedom driving rotation. Glucacon consists of 29 amino acids residuals and as well as insulin which consists 51 amino acids counting in peptide hormone. Practically, all molecules of this hormone represents one helix. In this case conformation of the polypeptide by the values of rotation angles ϕ and ψ are determined. If values ϕ and ψ of the amino acids residuals are equal then regular structures are formed, while in opposite case non-regular structures be stated. For example, regular structures are α-helix, β-structure and non-regular structure is coil.

Secondary protein structure is α-helix, β-structure and β-rotation. Which one structure there is in protein depends on hydrogen bonds in the molecule of protein, and their depends from inner angles ϕ and ψ distance between atoms.

For studying protein properties use graphical data about connection between angles ϕ and ψ in each peptide bell. If cups of angles ϕ and ψ have values -119° and 113° than protein formed structure β. If angles ϕ and ψ have values -57° and -48° than protein conformation correspond structure α-helix.

Conformation map of angles ϕ and ψ has made for some protein by rendgen structural analysis shows how much is fraction in the protein α-helix and β-structure.

5 Autocorrelation function for poly-L-glycine
At the poly-L-glycine formation there are interaction between inner angles of rotation ϕ and ψ. Autocorrelation function for inner angle of rotation ϕ k-th amino acid:

\[ r_{ϕϕ}(k) = \varepsilon[ϕ_o(k)ϕ_k(k)] = \int_{-∞}^{∞} \int_{-∞}^{∞} ϕ_o(k)ϕ_k(k)θ(ϕ_o,k,ϕ_k)dϕ_odϕ_k \]

or

\[ r_{ϕϕ}(k) = \varepsilon[ϕ_o(k)ϕ_k(k)] = \int_{-∞}^{∞} \int_{-∞}^{∞} ϕ_o(k)θ(ϕ_o,k,ϕ_k)dϕ_odϕ_k \]

Mean values or mathematical expectation for assembly \( ϕ_o(k) \) is:

\[ \mu{ϕ_o(k)} = \int_{-∞}^{∞} ϕ_o(k)θ(ϕ_o,k)dϕ_o \]

This is steady state function and equation (21) becomes.

\[ r_{ϕϕ}(k) = \mu{ϕ_o(k)} \mu_ϕ(k) \]

\[ r_{ϕϕ}(k) = -r_{ϕϕ}(k) \]

and autocorrelation function achieved maximum for k=0. The values of autocorrelation function for k=0 for inner angle ϕ at the poly-L-glycine are shown in Table 1. These values are represented average squares value, power of the process. Also, dispersion and mean values have shown for different k.

Autocovariation for ϕ for steady state assembly:

\[ σ_{ϕϕ}(k) = r_{ϕϕ}(k) - \mu_{ϕϕ}^2 \] (23)

According to equation (22) and (23) is obtained average value of the inner angle ϕ as a function of the number of the amino acids residual \( ϕ_o(k) \).

\[ ϕ_o(k) = \frac{σ_{ϕϕ}(k)}{μ_ϕ} + μ_{ϕ_o} \]

Including mean value for \( ϕ = -51.9° \) from experimental data Go et.al.[5] is obtained:
Coefficient of variation is calculated according to following equation.

\[ \gamma_{\varphi}(k) = \frac{\sigma_{\varphi}(k)}{\mu_{\varphi}} \]  

(25)

The obtained values from equations (22)-(25) have shown in Table 1.

The value of autocorrelation function for \( k=0 \) for inner angle \( \psi \) for poly-L-glycine is obtained:

\[ r_{\psi}(k) = \mu_{\psi}(k)\mu_{\psi}(k) \]  

(26)

Autocovariation probability quantity of \( \psi \) for steady state assembly is:

\[ \sigma_{\psi}(k) = r_{\psi}(k) - \mu_{\psi}^2 \]  

(27)

and combining with equation (26) the following equation is obtained:

\[ \mu_{\psi}(k) = \frac{\sigma_{\psi}(k)}{\mu_{\psi}^0} + \mu_{\psi}^0 \]  

(28)

This equation defines mean value inner angle \( \psi \) as a function of the number of the amino acids residual \( \mu_{\psi}(k) \).

The obtained values from equations (26)-(28) have shown in Table 2.

Table 1. Correlation parameters values for \( \varphi \) calculated by autocorrelation function

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \sigma_{\varphi}(k) )</th>
<th>( \gamma_{\varphi}(k) )</th>
<th>( \mu_{\varphi}(k) )</th>
<th>( r_{\varphi}(k) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>80</td>
<td>-1.54140</td>
<td>-53.440</td>
<td>2773.536</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>-0.15414</td>
<td>-52.054</td>
<td>2701.603</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>-0.48160</td>
<td>-52.380</td>
<td>2718.522</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>-51.900</td>
<td>2693.610</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>-0.19260</td>
<td>-52.092</td>
<td>2703.575</td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>0.01926</td>
<td>-51.880</td>
<td>2692.608</td>
</tr>
<tr>
<td>6</td>
<td>+1</td>
<td>-0.01926</td>
<td>-51.919</td>
<td>2694.596</td>
</tr>
</tbody>
</table>

Table 2. Correlation parameters values for \( \varphi \) calculated by autocorrelation function

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \sigma_{\psi}(k) )</th>
<th>( \gamma_{\psi}(k) )</th>
<th>( \mu_{\psi}(k) )</th>
<th>( r_{\psi}(k) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>110</td>
<td>-2.09125</td>
<td>-53.440</td>
<td>2773.536</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-52.054</td>
<td>2701.603</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>-0.66539</td>
<td>-52.380</td>
<td>2718.522</td>
</tr>
<tr>
<td>3</td>
<td>-5</td>
<td>0</td>
<td>-51.900</td>
<td>2693.610</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>-0.19260</td>
<td>-52.092</td>
<td>2703.575</td>
</tr>
<tr>
<td>5</td>
<td>-5</td>
<td>0.01926</td>
<td>-51.880</td>
<td>2692.608</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>-0.01926</td>
<td>-51.919</td>
<td>2694.596</td>
</tr>
</tbody>
</table>

6 Results and discussion

The obtained calculated results are shown in Fig. 1 and Fig.2. Fig. 1 shows autocovariation for \( \varphi \) as a function of the number of amino acids residual. Fig.2 shows shows autocovariation for \( \psi \) as a function of the number of amino acids residuals.

7 Conclusion

The autocovariation as a function of the amino acids residuals was derived. Poly-L-glycine are interactions between inner angles of rotation \( \varphi \) and \( \psi \) were studied.

Response of the autocovariation with varying assembly \( \varphi_0(k) \) and \( \psi_0(k) \) was examined.

Notation

- \( r_{\varphi}(k) \) – autocorrelation function variable \( \varphi \)
- \( r_{\psi}(k) \) – autocorrelation function of variable \( \psi \)

Greek symbol

- \( \gamma(k) \) – variation coefficient or relative deviation \( \varphi(k) \)
- \( \mu \) – mean value
- \( \sigma \) – dispersion
- \( \varphi \) – inner rotation angle in the protein main strand
- \( \psi \) – inner angle rotation in the protein main strand
Fig. 1  Autocovariation of $\phi$ for poly-L-glycine vs. $k$

Fig. 2  Autocovariation of $\psi$ for poly-L-glycine vs. $k$
References:


