KURAM VE UYGULAMADA EĞİTİM BİLİMLERİ EDUCATIONAL SCIENCES: THEORY & PRACTICE

Received: August 25, 2017
Revision received: April 2, 2018
Accepted: April 3, 2018

www.estp.com.tr
DOI 10.12738/estp.2018.6.179 ⬩ December 2018 ⬩ 18(6) ⬩ 2782-2793

Research Article

Application of Statistical Inference in Education and Teaching

Zhaoxia Huang[1]
Ankang University

Abstract

Statistical inference is a statistical analysis method and the basic method of contemporary statistics. It has strong practicability and can be widely used in many fields. This paper introduces the concept and style of statistical inference in detail, and makes a brief data analysis by SPSS software. Finally, it combines statistical inference with education and teaching, and makes an example analysis.

Keywords

Statistical Inference • Education and Teaching • SPSS Software

[1]**Correspondence to:** Zhaoxia Huang (PhD), School of Mathematics and Statistics, Ankang University, Ankang 725000, China; Institute of Mathematics and Applied Mathematics, Ankang University, Ankang 725000, China, Email: 2007990011@aku.edu.cn

Citation: Huang, Z. X., Application of Statistical Inference in Education and Teaching. *Educational Sciences: Theory & Practice, 18(6)*, 2782-2793. http://dx.doi.org/10.12738/estp.2018.6.179

In addition to the three common methods of least squares estimation, Bayesian estimation and maximum likelihood estimation, the number of estimation methods is increasing with the increasing requirements of modern engineering practice, such as moment estimation, maximum and minimum probability estimation, minimum contrast estimation and shortest distance estimation. Method and so on. Next, we mainly introduce the development of maximum likelihood estimation, Bayesian estimation and least square estimation in recent years. In the early nineteen twenties, Gauss's estimation method was mainly used to find the parameters of the correlation probability density function of a sample set. This method was first named by the British statistician, sir Fisher, and named as maximum likelihood estimation. Based on the application and extension of this method in research, a series of research results have been produced: Taraski, based on the application of maximum likelihood estimation, realizes the analysis of stochastic differential equations to clarify the problem of parameter estimation, and discovers the asymptotic normality of the results in the proof process (Taraskin, 1974); Ishra applied this method in the analysis of general linear time-varying homogeneous stochastic differential equations, and estimated the parameters of the equation. At the same time, based on the research and analysis, Berry-Esseed's boundedness theorem (Mishra & Rao, 1985); KticWer are also used to analyze the homogeneous exponential form (Kiichler & Soensen, 1994). In this process, not only the method is applied, but also the semi-martingale theory is applied to realize the study of parameter estimation, and the demonstration is carried out. The results can be compared with local ones. The mixed asymptotic normality coincide. Compared with linear stochastic systems, experts and scholars pay more attention to non-linearity and get more achievements. Rubin and others used the maximum likelihood estimation method to analyze the problem of parameter estimation in the equation, and found that the conclusion has strong consistency and asymptotic normality (Rao & Rubin, 1981; Ike, 2018). Yoshida analyzed the system with unknown parameters, determined the approximate likelihood rate function under the discrete observation state, and estimated the unknown parameters of the system by using the maximum likelihood estimation method, and discussed the weak convergence and asymptotic property of the random domain of likelihood rate (Yoshida, 1990).

Bayesian estimation is also a common estimation method. The formation of this method is based on the Bayesian school. This method first appeared in 1763 and was proposed by Bayes. It was applied to biomedical statistics and played a certain role. In fact, Bayesian theory is under the condition of incomplete information, based on subjective probability, estimating the unknown state, and based on the application of Bayesian formula, so as to realize the revision of occurrence probability, based on the revision probability, expectation value, and so on, so as to determine the optimal decision. Classical school pointed out that statistical inference needs two kinds of information, one is sample information and the other is general information. But for Bayesian school, not only the two kinds of information are highly valued, but also the third kind of information is extremely concerned, which is the collection, mining and processing of prior information. Wait. In this case, it is helpful to quantify this type of information, and to get a prior distribution, which can be incorporated into statistical inference, so that the efficiency of inference can be effectively improved. In the actual application process, if we fail to correctly understand the role of prior information, it may affect the conclusions, and thus get unreasonable conclusions. For this reason, after the emergence of Bayesian estimation, people have paid great attention to this method, and applied it to practice, so as to obtain relevant results. Bauer carried out research work based on this method, solved the problem of estimating system parameters with unknown drift terms, and validated the relevant characteristics of conclusions, such as validity and gradualness (Bauer, 1980). Kutoyants studied the parameter estimation of nonlinear time-varying homogeneity, based on the application of Bayesian method (Stoyan, 1986). The estimation results are effectively demonstrated.

The least square method occupies an important position in the estimation method. The proposed method can be traced back to the end of the eighteenth century to the beginning of the nineteenth century. Gauss et al. obtained the method in the research process, and Markov continued to study in this field. The least square estimation method has been well developed. From the point of view of this method, when solving location data, this method can often be simplified, which makes the solution more convenient, and then makes the sum of squares of errors corresponding to the obtained data and actual data to be controlled in the minimum range. It is precisely because of the advantages of least squares method, for this reason, it is obtained in the linear statistical model. Full application, so as to achieve the problem of parameter estimation, so that such problems can be effectively solved. When estimating unknown parameters of the system, the least square method is often used, and more research results have been achieved in this field. Kailath (1968) makes an in-depth study of the estimation methods used in reference (Bode & Shannon, 1950), so as to realize the estimation of systems with unknown parameters. Bai (1994) studies linear equations, and in the process of parameter estimation, the least square method is also used. From the above literature, the least squares method has certain requirements for drift terms, which need to be small enough to obtain the asymptotic distribution of the variable point estimator, and to realize the understanding of its related characteristics, such as compatibility. In addition, some researchers also use the least square method to estimate the parameters in the process of studying the non-linear equation, and get a series of results. Wu compares the results obtained by the least square method in the process of studying linear stochastic systems, so as to study the nonlinear systems with unknown variables, so as to clarify the corresponding conditions of the estimators (Wu, 1981). At the same time, in order to explore its performance in convergence, the estimation problem caused by parameter space limitation is analyzed in detail. Bishwal has also carried out research work on this issue, which is also aimed at non-linear, but belongs to time-varying homogeneous stochastic system, thus realizing the exploration of parameter estimation (Kristensen, Madsen & Jorgenseii, 2004). In order to realize the understanding of its asymptotic normality, the corresponding stochastic regions are established. Kan and others have studied the linear model (Kan & Shu, 2012). With the help of Bayesian theory, they have realized the research of parameter estimation. Based on the research, the results have excellent performance in asymptotic normality and other aspects.

Basic methods of statistical inference

1. Parameter estimation: Parameter estimation is a method of estimating global parameters by constructing sample statistics, which includes point estimation and interval estimation.

(1) point estimation. Point estimation is a direct estimation of the values of group parameters, and a statistic of each sample data is used as the estimation of group parameters. For example, when estimating the average of a normal whole, the average of sample data is taken as the estimator of the overall average. In point

estimation, sample statistics are required to be unbiased, that is to say, the average number of distributions produced by such sample statistics is the estimated parameter in countless repeated sampling. It is also required that the variance of this sample distribution is less than the variance of other unbiased estimators.

If the distribution function $F(x;\theta_1,\ldots\theta_m)$ of the whole $X1,\ldots,Xn$, $\theta_1,\ldots\theta_m$. X is known and the parameter $\theta_1,\ldots\theta_m$ is unknown, then the sample can be extracted from the whole X, and the appropriate statistic $\hat{\theta}_i(X_1,\ldots,X_n)$ can be constructed as the estimator of the parameter θ_i according to the characteristics of the parameters to be estimated. Then the observed value $x_1,\ldots x_n$ of the estimator can be calculated from the observed value $\hat{\theta}_i(X_1,\ldots,X_n)$ of the sample extracted, which is the unknown parameter θ_i. Estimation (i=1,2,…m) is called point estimation of parameters.

The commonly used methods for constructing point estimation are moment estimation, least squares estimation, maximum likelihood estimation and Bayesian estimation. The moment estimation method estimates the global moment by the sample moment, and estimates the corresponding function of the global moment by the function of the sample moment, thus estimating the global parameters. Maximum likelihood estimation (MLE), also known as maximum likelihood estimation (MLE), is the maximum likelihood estimation of parameters by using sample distribution density to form a likelihood function. Least squares estimation is mainly used for parameter estimation in linear statistical models. Bias estimates that Beth's theorem, new evidence and prior transcendental probability are combined to get a new probability.

(2) interval estimation

Interval estimation is the estimation of the true value of unknown parameters or functions of parameters in group distribution by composing an appropriate interval according to the requirements of fixed accuracy and accuracy of the sample extracted. For example, if a normal distribution is known, its average is μ and its variance is δ. Samples with the number of data n are sampled repeatedly. According to the central limit theorem, the average x of these samples constitutes a normal distribution with the group average μ as the average and the variance δ/n as the variance. According to the nature of normal distribution, for the average x of any sample, there is a probability of $\mu - \frac{1.96^{\delta}}{\sqrt{n}} < \bar{x} < \mu + \frac{1.96^{\delta}}{\sqrt{n}}$, that is, the probability of 0.95. The total average is unknown, and the value of x can be obtained from a sample we extracted. From the above formula, we can infer that the total average will fall in the interval $\left\{\bar{x} - \frac{1.96^{\delta}}{\sqrt{n}}, \bar{x} + \frac{1.96^{\delta}}{\sqrt{n}}\right\}$ at the probability level of 0.95, which is the confidence interval of the total average. In the upper and lower bounds of this interval, the total variance δ is generally unknown. We still need to use sample data for point estimation. In the actual construction of the confidence interval, we do not necessarily use normal distribution, but use t distribution and other distributions to make the inference more reliable.

2. Hypothesis testing: Hypothesis testing is to make judgments about ordinary things, often to make assumptions about concerns or speculations. Whether these hypotheses are correct or not is unknown. The content of the description is about the distribution characteristics of the population or its parameters, which is called statistical hypothesis. Under normal circumstances, we first need to propose two mutually exclusive

assumptions: the original hypothesis and the alternative hypothesis. In most cases, we propose a statistical hypothesis only to reject this assumption.

The basic principle of hypothesis testing is "small probability principle". The so-called "small probability principle" means that in a particular experiment, the probability of a random event is extremely small, and the event will not occur in this experiment. At a certain saliency level, a sample is randomly selected from the population, and if the original hypothesis holds, it is found that the observed value of the sample taken into the statistics is a small probability event, that is to say, a small probability event occurs, which violates the principle of small probability and shows that the original hypothesis is wrong. They have every reason to refuse it; on the contrary, they are forced to accept it. However, due to a series of reasons such as the randomness of samples, we may have two kinds of errors in judging whether the original hypothesis is correct or not: discarding the true and taking the false. When the sample size is fixed, one type of error will be reduced and the other one will increase. In order to eliminate these two kinds of errors at the same time, we can only solve this problem by increasing sample size, which is unrealistic.

The basic steps of hypothesis testing are as follows:

(1) Establishing hypothesis: the symbol of invalid hypothesis H_0; the symbol of alternative hypothesis H_1.

H_0: The difference between sample and population or between sample and sample is caused by sampling error.

H_1: There are essential differences between samples and population or between samples and samples.

(2) Determine test level ∂, usually ∂=0.05 or ∂=0.01.

(3) To calculate the test statistics, different methods (*Z* test, *T* test, *F* test, chi-square test, etc.) and specific formulas are used.

T -test means that in the case of small samples (n<30), the mathematical expectation of random variables is not different from that of a given value.

Z-test is an example of large sample average difference test (n>30). It uses the theory of standard normal distribution to estimate the probability of difference, so as to check whether the difference between the two averages is significant.

F test is used in the T test of two samples. In order to compare the two samples, it is necessary to determine whether the variance of the two samples is the same, that is, the homogeneity of variance. F test was used to determine whether the two global variances were the same. Chi-square test refers to the statistical verification of probability and critical value obtained by Chi-square distribution, and it is a widely used hypothesis test method for counting data.

Determine a value of *P* according to the statistics and the corresponding threshold table: If $P>\partial$, accept H_0. Considering that the level taken according to ∂ is not significant, the difference is mostly caused by sampling errors, which is not statistically valid; if $P\leq\partial$, H_0 is rejected. It is believed that the difference may be caused only by sampling errors and most probably by differences in experimental factors, so it is statistically valid.

3. The Internal Relation between Parameter Estimation and Hypothesis Testing We consider general issues:

Example: The length of steel produced by a factory obeys normal distribution. Its average value is 240 centimetre, a product is randomly extracted from the factory. The results are as follows:

239.7, 239.6, 239, 240, 239.2

Try to determine whether the length of the steel meets the set requirements? This is a two-sided hypothesis test for normal mean.

(1) First, two hypotheses are proposed. H_0:μ=240, Original hypothesis， Alternative hypothesis H_1:μ≠240.

(2) In the case of the original false H_0:μ=240, the test statistics are $T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{\partial(n-1)}$.

(3) From a given ∂=0.05， According to the t distribution table, the critical value of rejection is obtained as follows: $t_{\frac{\partial}{2}}(n-1) = t_{0.025}(4) = 2.776$.

The rejection field |T| ≥ $t_{\frac{\partial}{2}}(n-1)$ which has this rejection field H_0 is (-∞, -2.776)∪(2.776, +∞).

(4) Computation of the observed value of T from the sample

$|T| = \frac{|239.5-240|}{\frac{0.4}{\sqrt{5}}} = 2.795 \subseteq (-\infty, -2.776) \cup (2.776, +\infty)$

So refuse H_0, Therefore, it is considered that the length of the steel does not meet the requirements.

Parametric interval estimation and hypothesis test are the same way of thinking. Their basic principle is to use sample information to infer the nature and characteristics of the whole. Both of them apply the estimation method of mathematical statistics theory, take sampling distribution as the theoretical basis, and based on probability reasoning theory, select a statistic, according to which it is packed by an interval. The probability of inclusion is the result of inference. The two can transform each other and form duality.

For example, Set total $N(\mu,\sigma^2)$ σ^2 unknown. Selection statistics for $T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{\partial(n-1)}$

If the confidence level is 1-∂, The probability is $P\left\{\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \le t_{\frac{\partial}{2}}(n-1)\right\}$

The confidence interval for μ to be 1-∂ is $\left(\bar{X} - \frac{S}{\sqrt{n}} t_{\frac{\partial}{2}}(n-1), \bar{X} + \frac{S}{\sqrt{n}} t_{\frac{\partial}{2}}(n-1)\right)$

If hypothesis testing is considered H_0:μ=μ_0; H_0:μ≠μ_0 Statistics for $T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{\partial(n-1)}$

For a given saliency level ∂, Identify a small probability event $\left\{\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} > t_{\frac{\partial}{2}}(n-1)\right\}$ The probability of its occurrence is ∂.

Bring in sample observations，See if $|\bar{x}-\mu_0|\frac{\frac{1}{S}}{\sqrt{n}} > t_{\frac{\partial}{2}}(n-1)$ is established，Decide whether to reject the original hypothesis，Domain of rejection is $W=\left\{|\bar{x}-\mu_0| > \frac{S}{\sqrt{n}}t_{\frac{\partial}{2}}(n-1)\right\}$

Acceptance domain is $W=\left\{|\bar{x}-\mu_0| \le \frac{S}{\sqrt{n}}t_{\frac{\partial}{2}}(n-1)\right\}$，It can be rewritten as $\overline{W}=\left\{\bar{x}-\frac{S}{\sqrt{n}}t_{\frac{\partial}{2}}(n-1) \le \mu_0 \le \bar{x}+\frac{S}{\sqrt{n}}t_{\frac{\partial}{2}}(n-1)\right\}$

μ_0 has no restrictions.，If μ_0 is given a value in (-∞,+∞), the confidence interval of 1-∂ of μ can be obtained as follows: $\left(\bar{x}-\frac{S}{\sqrt{n}}t_{\frac{\partial}{2}}(n-1), \bar{x}+\frac{S}{\sqrt{n}}t_{\frac{\partial}{2}}(n-1)\right)$

Simple data analysis using SPSS

Data selection

Table 1
Grade 2 in English *Unit: Division*

Serial number	Gender	Achievement	Serial number	Gender	Achievement
1	female	73	11	female	65
2	male	68	12	male	83
3	female	82	13	female	70
4	female	82	14	female	79
5	female	92	15	male	69
6	male	73	16	male	64
7	male	71	17	female	81
8	male	88	18	female	78
9	female	90	19	male	72
10	female	95	20	female	79
Serial number	Gender	achievement	Serial number	Gender	achievement
21	male	66	31	female	67
22	female	56	32	female	75
23	female	76	33	male	69
24	male	58	34	female	65
25	female	78	35	female	63
26	male	45	36	male	64
27	female	89	37	male	67
28	female	78	38	female	69
29	female	67	39	male	69
30	female	67	40	female	77
Serial number	Gender	achievement	Serial number	Gender	achievement
41	female	81	42	male	80
43	male	71	44	female	79
45	female	60	46	female	82
47	female	83	48	female	69
49	male	63	50	male	65

The school has a total of two students, including boys and girls. After a monthly exam, the school leaders wanted to know the English level of the students in our school, so they decided to make a statistical analysis of the English scores of the second grade in junior middle school, so as to infer the English learning situation of all the students in our school, so as to formulate the next teaching plan and take pertinent measures to improve our school's English achievements. A random sample was drawn from all the second grade English test papers, as shown in Table 1.

The number of points intervals is shown in table 2.:

Table 2
Number of Points Interval *Unit: Human*

Fractional interval	Number of people
(0-10]	0
(10-20]	0
(20-30]	0
(30-40]	0
(40-50]	1
(50-60]	3
(60-70]	19
(70-80]	15
(80-90]	10
(90-100]	2

Data analysis

The boys and girls were divided into two groups for data analysis and the following results were obtained. It can be seen from the scatter plot that the scores of male and female students are basically linear, showing a linear correlation.

Using SPSS, the results of male and female students are analyzed as follows:

The results of boys' analysis are shown in table 3:

Table 3
Male Descriptive Statistics

		Statistic	Standard error	Bootstrap deviation	Bootstrap Standard error	95% confidence interval lower limit	95% confidence interval Upper limit
Gender	N	19		0	0	19	19
	Full distance	0					
	Minimum value	1					
	maximum value	1					
	mean value	1.00	.000	.00	.00	1.00	1.00
	Standard deviation	.000		.000	.000	.000	.000
	variance	.000		.000	.000	.000	.000
	skewness	.	.	.[b]	.[b]	.[b,c]	.[b,c]
	kurtosis	.	.	.[b]	.[b]	.[b,c]	.[b,c]
achievement	N	19		0	0	19	19
	Full distance	43					
	Minimum value	45					
	maximum value	88					
	mean value	68.68	2.110	.09	2.09	64.37	72.74
	Standard deviation	9.196		-.423	1.935	4.823	12.668
	variance	84.561		-3.865	34.037	23.262	160.474
	skewness	-.259	.524	.229	.879	-1.699	1.734
	kurtosis	2.158	1.014	-.525	1.728	-.970	5.937
Effective N (List state)	N	19		0	0	19	19

Table 4
Descriptive Statistics for Girls

		Statistic	Standard error	Bootstrap deviation	Bootstrap Standard error	95% confidence interval lower limit	95% confidence interval Upper limit
Gender	N	31		0	0	31	31
	Full distance	0					
	Minimum value	2					
	maximum value	2					
	mean value	2.00	.000	.00	.00	2.00	2.00
	Standard deviation	.000		.000	.000	.000	.000
	variance	.000		.000	.000	.000	.000
	skewness	.	.	.[b]	.[b]	.[b,c]	.[b,c]
	kurtosis	.	.	.[b]	.[b]	.[b,c]	.[b,c]
achievement	N	31		0	0	31	31
	Full distance	39					
	Minimum value	56					
	maximum value	95					
	mean value	75.71	1.698	.00	1.67	72.48	78.97
	Standard deviation	9.452		-.214	1.077	7.027	11.253
	variance	89.346		-2.845	19.880	49.384	126.637
	skewness	-.021	.421	-.019	.312	-.657	.575
	kurtosis	-.377	.821	.003	.505	-1.153	.816
Effective N (List stat)	N	31		0	0	31	31

From Table 3, we can see that there is a total of male students in the selected famous students. The average score of this male student is the highest score and the lowest score. The difference between variance and standard deviation is about two, which shows that the difference between each student's scores is about one-sided; the skewness is negative, and the skewness of this group of variables is different from the skewness of normal distribution. That is to say, the data are more concentrated on the left side of the mean. The kurtosis is positive. We know that these variables are different from the kurtosis of normal distribution and steeper than the peak of normal distribution.

The results of female student achievement analysis are shown in table 4.

Table 4 shows that there are girls among the selected students. The average score of this girl student is score, the highest score is score, and the lowest score is score. The variance and standard deviation are respectively 0 and 0, so it can be seen that each student's score gap is approximately about the same; skewness is negative, which shows that the skewness of this group of variables is different from that of normal distribution, that is, the data are more concentrated on the left side of the mean; kurtosis is negative, which shows that the group of variables is different from that of normal distribution, and is more than normal distribution. The peak is more gentle.

Conclusion and Countermeasures

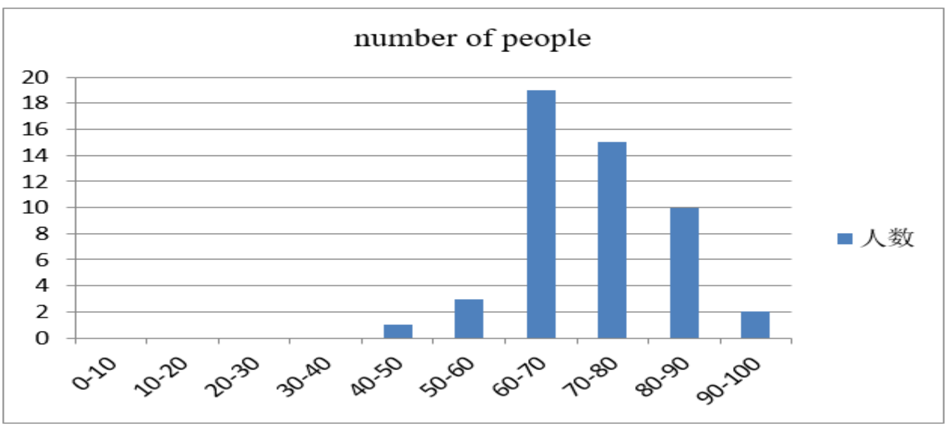

Figure 3. statistical chart of 50 students' English Achievements

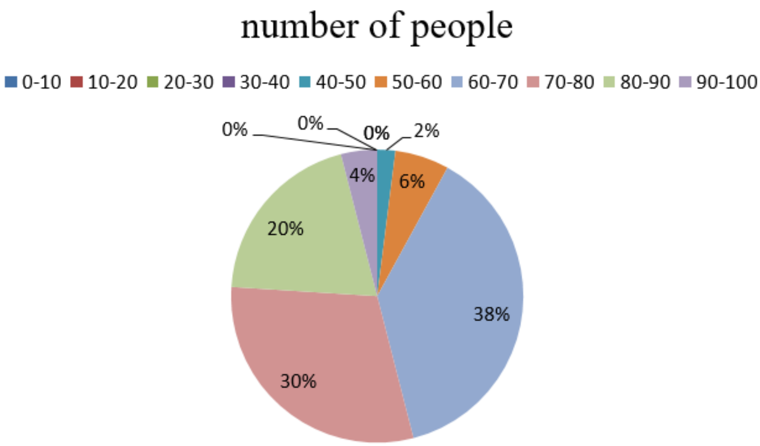

Figure 4. 50 pie chart of students' English achievement

Conclusion: By comparing male and female groups, it can be seen that the English scores of female students are generally better than those of male students, and most of the students' English scores are concentrated between them. In order to draw a general conclusion from the sample more clearly and intuitively, the following bar and pie charts are made for the distribution of the English scores of the famous students.

Countermeasures: schools should focus on English learning for boys. Of course, girls can not relax. It can adopt the mode of guiding boys' poor English achievement with good English achievement of girls, and make common progress with pertinence and goal.

References

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, *15*(5), 453-472. https://dx.doi.org/10.1111/j.1467-9892.1994.tb00204.x

Bauer, K. (1980, January). On asymptotic properties of estimates of the drift coefficient parameter of a diffusion process. In *Theory of Probability and Its Applications 25* (2). 429-431.

Bode, H. W., & Shannon, C. E. (1950). A simplified derivation of linear least square smoothing and prediction theory. *Proceedings of the IRE*, *38*(4), 417-425. https://dx.doi.org/10.1109/JRPROC.1950.231821

Ike, C. C. (2018). Energy formulation for flexural – torsional buckling of thin-walled column with open cross-section, *Mathematical Modelling of Engineering Problems*, *5*(2), 58-66. https://dx.doi.org/10.18280/mmep.050202

Kailath, T. (1968). An innovations approach to least-squares estimation-Part I: Linear filtering in additive white noise. *IEEE Transactions on Automatic Control*, *6*(13), 646-655. https://dx.doi.org/10.1109/TAC.1968.1099025

Kan, X., & Shu, H. (2012). Uniform approximate estimation for nonlinear nonhomogenous stochastic system with unknown parameter. Mathematical Problems in Engineering, *2*(12), 330-342. https://dx.doi.org/10.1155/2012/421754

Kuchler, U., & Sorensen, M. (1994). Exponential-families of stochastic-processes with time-continuous likelihood functions. *Scandinavian Journal of Statistics*, *21*(4), 421-431. https://dx.doi.org/10.1016/0378-3758(94)90206-2

Kristensen, N. R., Madsen, H., & Jørgensen, S. B. (2004). Parameter estimation in stochastic grey-box models. *Automatica*, *40*(2), 225-237. https://dx.doi.org/10.1016/j.automatica.2003.10.001

Mishra, M. N., & Rao, B. P. (1985). On the Berry-Esseen bound for maximum likelihood estimator for linear homogeneous diffusion processes. *Sankhyā: The Indian Journal of Statistics, Series A*, 392-398. https://dx.doi.org/10.1007/BF00146615

Rao, B. P., & Rubin, H. (1981). Asymptotic theory of estimation in nonlinear stochastic differential equations. *Sankhyā: The Indian Journal of Statistics, Series A*, 170-189.

Stoyan, D. (1986). Kutoyants, Yu. A.: Parameter Estimation for Stochastic Processes. Translated and edited by BLS Prakasa Rao. Heldermann Verlag, Berlin—, DM. *Biometrical Journal*, *28*(3), 294-294. https://dx.doi.org/10.1002/bimj.4710280306

Taraskin, A. F. (1974). On the asymptotic normality of vector valued stochastic integrals and estimates of a multidimensional diffusion process. *Theory of Probability and Mathematical Statistics*, *2*, 209-224.

Wu, C. F. (1981). Asymptotic theory of nonlinear least squares estimation. *The Annals of Statistics*, 501-513. https://dx.doi.org/10.1214/aos/1176345455

Yoshida, N. (1990). Asymptotic behavior of M-estimator and related random field for diffusion process. *Annals of the Institute of Statistical Mathematics*, *42*(2), 221-251. https://dx.doi.org/10.1007/BF00050834