الأكاديمية الليبية
The Libyan Academy
Misurata Branch

The Evolution of Shrinkage Estimators between Statistical Theory and Modern Intelligent Applications

Aisha Mohamed Abutartour

Statistics Department, Faculty of Science, University of Tripoli, Tripoli, Libya
A.Abutartur@uot.edu.ly

Article information

Key words

: Artificial Intelligence (AI), Bias and Variance, Model Stability, Model Regularization, Overfitting, Shrinkage Estimator

Received **22 02 2026**, Accepted **11 03 2026,** Available online **12 03 2026**

Abstract

Mathematical statistics underwent a major transformation during the last century with the discovery of shrinkage estimators, which provided an improved version of the classical tradition based on unbiased estimators. Since the discovery of Stein's theory in the early 1960s, accepting a degree of bias has become an effective tool for reducing overall error and producing more stable estimators. This paper aims to provide a comprehensive review of the historical and theoretical development of shrinkage estimators, from the James-Stein model to modern models such as ridge regression, Least Absolute Shrinkage and Selection Operator (LASSO), and elastic net regression. It also focuses on their fundamental statistical basis and their growing role in the development of artificial intelligence (AI) and deep learning models. The paper concludes by presenting a unified vision that links statistical theory and modern intelligent applications by adopting shrinkage estimators as an integrated framework that combines mathematical precision with applied innovation in complex data environments. This unified framework illustrates how shrinkage estimators have evolved to become the cornerstone of intelligent predictive modeling.

I. Introduction

The concept of shrinkage has evolved significantly, moving from traditional statistical uses to modern applications in AI and machine learning. It first appeared in the framework of statistical estimation with the aim of reducing the mean squared error (MSE) by balancing bias and variance through approximating estimates toward a reference value. This balances bias and variance and improves prediction accuracy (James and Stein, 1961). As models became more complex and data dimensions

A.Abutartur@uot.edu.ly

increased, contraction techniques emerged in linear regression, such as Ridge (Hoerl and Kennard, 1970) and Lasso (Tibshirani, 1996). However, these techniques are used to reduce overfitting and enhance generalization. Early research in neural networks also showed that incorporating mechanisms such as weight decay is essentially a form of contraction to improve generalization (Krogh and Hertz, 1991). This paradigm shift demonstrates the transition from a purely statistical concept to intelligent application in complex predictive models. In the modern era, contraction techniques have become an essential part of deep learning model design, where they used within a structured, organized framework to build more efficient and accurate intelligent systems (Goodfellow et al., 2016) .

With the rapid expansion of AI models for analyzing high-dimensional data, the need has grown for statistical tools capable of improving model performance while mitigating the overfitting problem. In this context, contraction estimators play a pivotal role. Ridge regression (Hoerl and Kennard, 1970), LASSO regression (Tibshirani, 1996), and elastic net (Zou and Hastie, 2005) are among the most important tools developed to address the overfitting problem, which can lead to reduced predictive performance when applied to new data. Integrating these methods into the construction of statistical models and machine learning algorithms is essential to ensure models can generalize and produce results that are more reliable. Based on all of the above, and with the development of data science and the emergence of AI, contraction techniques have moved beyond the theoretical realm to become a fundamental component in the design of intelligent models. However, there is still a lack of comprehensive studies that systematically explore how shrinkage estimators have transformed from a purely theoretical discovery in classical statistics into a central tool in modern AI and big data science. From this perspective, this paper aims to provide a comprehensive analytical review of the development of shrinkage estimators from a theoretical and applied perspective, focusing on their role in improving the performance of statistical models and AI in terms of estimation, variance control, and enhancing prediction efficiency in complex environments. To achieve this, the study adopts an analytical literature review approach, focusing on two main criteria: foundational influential contributions that shaped major shrinkage models such as Ridge, Lasso, and Elastic Net, and recent advancements in their modern applications in statistical learning. Particular attention is given to their theoretical properties and their role in addressing statistical challenges such as multicollinearity and overfitting in high-dimensional data.

2. Theoretical Background and Literature Review

2.1 Definition of Shrinkage Estimators

Shrinkage estimators are a class of statistical estimators that aim to minimize the mean square error (MSE) by shrinking the estimated values toward a reference point (usually the mean or zero) rather than relying on classical estimators such as the maximum likelihood estimator (MLE). A simplified form of the shrinkage principle can be written as follows:

$$\hat{\theta} shrink = (1 - \lambda)\hat{\theta}_{MLE} + \theta_0 \lambda \quad (2.1)$$

where $\hat{\theta}\, shrink$ is a shrink estimator, $\hat{\theta}_{MLE}$ is a traditional estimator (such as the maximum likelihood estimator) , θ_0 is the reference point (usually the zero vector) , and $\lambda \in [0\,,1]$ the shrinkage factor that controls the amount of shrinkage toward the reference θ_0. This formula illustrates the fundamental equation between bias and variance, which forms the core of the concept of contraction. As λ increases, the estimator approaches the reference point, reducing variance at the cost of a slight increase in bias, thus achieving a lower overall mean error. The contraction principle forms the theoretical basis for many modern statistical and machine learning models, including the James-Stein estimator (James and Stein, 1961), the Lasso model (Tibshirani, 1996), and the elastic network (Zou and Hastie, 2005). Its elasticity lends particular strength to the processing of high-dimensional or small-sample data, effectively bridging statistical theory with modern applications (Efron and Morris, 1977; Tibshirani, 1996).

2.2 The Theoretical Basis of Shrinkage Estimators

Interest in shrinkage estimators began in the mid-twentieth century when Stein (1956), in his famous study "The Unacceptability of the Usual Estimator of the Mean of a Multivariable Normal Distribution," showed that the conventional estimator of the sample mean (i.e., maximum likelihood estimator) becomes statistically unacceptable when the number of dimensions p ≥ 3. This concept was later extended to include more complex contexts, such as minimum and maximum (minimax) risks and wave shrinkage (Donoho and Johnstone, 1994) . Earlier, at the beginning of the 20th century, Fisher (1922) established the theoretical framework for the concept of the maximum error, particularly the maximum likelihood method (MLE), which became a cornerstone of classical statistical estimation theory, as discussed by Aldrich (1997) and Lehmann &

Casella, (1998) in his historical analysis of the development of the maximum likelihood methodology. This estimation is generally expressed as follows:

$$\arg\max_{\theta} L(\theta; x) = \hat{\theta}_{MLE} , \quad (2.2)$$

where $\arg\max_{\theta}$ is the value of the parameter θ that makes the probability function as large as possible, $MLE\hat{\theta}$ is the maximum probability estimator, L(θ;x) is the value that maximizes the probability function based on the observed data. In the case of a normal distribution with known variance, the sample mean $\bar{X}$ is estimated as the maximum probability of the population mean μ. Although the maximum likelihood estimator possesses consistency and asymptotic efficiency, Stein (1956) demonstrated that it becomes unstable when the number of variables is large or the sample size is small. This led to the need for new, more efficient estimators based on the concept of introducing a simple shrinkage toward a specific reference value, as opposed to a slight bias, which achieves a significant reduction in variance and mean squared error (MSE). Later, James and Stein (1961) developed this idea in their research "Estimation with Quadratic Loss", and proposed the famous James-Stein Estimator, which is expressed by the relation:

$$\hat{\theta}_{JS} = \bar{X} - \lambda(\bar{X} - a), \quad \text{where} \quad \lambda = \frac{\sigma^2(p-2)}{\|\bar{X}-a\|^2} , \quad (2.3)$$

where $\hat{\theta}_{JS}$ is James-Stein estimator of the mean vector , $\bar{X}$ is a sample mean vector , a is the shrinkage target, usually a=0 , p is a number of dimensions (variables)($p \geq 3$), σ^2 is a population variance , $\|\bar{X} - a\|^2$ is a square of the Euclidean distance between the mean and the target a and λ is the shrinkage factor . This estimator represents a paradigm shift in estimation theory because it introduces a systematic bias toward the reference point a (often zero) in exchange for a large reduction in variance. With the increasing number of independent variables and the emergence of the problem of multicollinearity, Hoerl and Kennard (1970) introduced the Ridge Regression model, which is an extension of the idea of contraction to the analysis of multicollinearity:

$$\hat{\beta}_{ridge} = [X'X + \lambda I]^{-1}X'Y \quad , \quad (2.4)$$

where $\hat{\beta}_{ridge}$ is the ridge estimator for the regression coefficients, X is a matrix of independent variables ($n \times p$), X' is the matrix transposition of X , p is the number of dimensions , λ is the shrinkage parameter that controls the amount of shrinkage ($\lambda \geq 0$) , I is the identity matrix , Y is the dependent variable vector of size(n × 1) , The Ridge model adds a quadratic penalty to the regression coefficients, reducing their magnitude relative to zero and lowering the variance in estimates. By introducing a controllable bias via λ , if λ=0 the model reverts to normal linear regression (OLS). As λ increases the bias increases and the regression coefficients decrease, while the variance decreases, thus improving the stability of estimates, particularly in cases of high linear correlation or small sample sizes. In the 1990s, Tibshirani (1996) developed this concept and introduced the Lasso (Least Absolute Shrinkage and Selection Operator) estimator, which replaced the squared term with a penalty based on the absolute value, The standard mathematical formulation for this estimator (Hastie et al., 2009) is given by the following equation:

$$\widehat{\beta}_{lasso} = \arg min_{\beta} \{\tfrac{1}{2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} X_{ij}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\},$$

(2.5)

where $\hat{\beta}_{lasso}$ represents the lasso estimates of the regression coefficients, $\arg min_{\beta}$ is the value of β that makes the function inside the brackets as small as possible, y_i is the observed value of the dependent variable for observation i , β_0 is the intercept of the model, X_{ij} represents the value of the $j-th$ independent variable for the $i-th$ observation, β is the vector of the regression coefficients to be estimated, β_j is the value of the regression coefficient for the variable j. n is the total number of observations, p is the number of independent variables. The factor $\frac{1}{2}$ is a mathematical convenience for derivation. λ is the contraction factor that determines the degree of added penalty; the higher its value, the greater the contraction. The lasso method is unique in that the penalty limit $|\beta_j|$ completely zeros out certain small coefficients, enabling it to perform variable selection in addition to its role in contraction and variance reduction.

After both the Ridge Regression model and the Lasso estimator demonstrated their effectiveness in addressing multicollinearity and overfitting problems, Zou and Hastie (2005) noted that each model has distinct advantages and disadvantages. While Ridge effectively limits variance when there are strong correlations between independent variables, it does not completely zero out small regression coefficients and therefore does not perform variable selection. Lasso, on the other hand, introduces a penalty limit of type L1 that zeros out some coefficients, but it may fail to select the correct variable when there are high correlations between explanatory variables.

Based on this, Zou and Hastie (2005) proposed the Elastic Net estimator, which combines the features of both Ridge and Lasso by integrating the penalties L1 and L2 simultaneously, as in the following equation:

$$\hat{\beta}_{EN} = \arg min_{\beta} \{\tfrac{1}{2}\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} X_{ij}\beta_j\right)^2 + \lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p} {\beta_j}^2\},$$

(2.6)

where $\hat{\beta}_{EN}$ represents the Elastic Net estimators for the regression coefficients, $\arg min_{\beta}$ is the value of β that makes the function inside the brackets as small as possible, y_i is the true value of the dependent variable at observation i , β_0 is the intercept of the model, X_{ij} represents the value of the $j-th$ independent variable for the $i-th$ observation, β is the vector of the regression coefficients to be estimated, β_j is the value of the regression coefficient for the variable j. n is the total number of observations, p is the number of independent variables. The factor $\frac{1}{2}$ is a mathematical convenience for derivation. λ_1 and λ_2 are positive coefficients for controlling the degree of penalization; λ_1 represents the penalty term of type L_1 (used in Lasso), while λ_2 represents the penalty term of type L_2 (used in Ridge). The Elastic Net model is a hybrid formula capable of balancing shrinkage and variable selection. It is more stable than the Lasso estimator when strong correlations exist between explanatory variables, making it one of the most widely used regularized regression models in modern high-dimensional applications. After presenting the theoretical basis for the characteristics of the different shrinkage estimators and in order to clarify the key differences between the most widely used shrinkage regression models (Hastie et al., 2009), Table (2.1) provides a brief comparison between the Ridge, Lasso and Elastic Net models in terms of the type of penalty, the possibility of feature selection, and their ability to address the problem of multicollinearity between variables (Hoerl & Kennard, 1970; Tibshirani, 1996; Zou & Hastie, 2005).

Table 2.1: Comparison of Ridge, Lasso, and Elastic Net Estimators

Estimators	Penalty Type	Feature Selection	Handling Multicollinearity
Ridge Regression	L_2 penalty (sum of squared coefficients)	No (Shrinks coefficients towards zero but does not set them exactly to zero)	Very effective
Lasso Regression	L_1 penalty (sum of absolute values of coefficients)	Yes (Can shrink some coefficients to exactly zero, effectively dropping them from the model)	It tends to select only one variable from a group of highly correlated predictors
Elastic Net	Combination of L_1 and L_2 penalties	Yes (Retains the variable selection property of Lasso)	Highly effective, especially with highly correlated predictors

Table (2.1) illustrates the key differences among the three shrinkage estimators: Ridge, Lasso, and Elastic Net. The Ridge model reduces the variance of regression coefficients by applying an L_2 penalty without removing variables from the model. In contrast, the Lasso model applies an L_1 penalty, allowing for simultaneous shrinkage and variable selection by setting some regression coefficients to zero. The Elastic Net model combines both L_1 and L_2 penalties, providing a balance between variance reduction and variable selection, resulting in a more stable and accurate model, particularly when strong correlations exist among explanatory variables.

This relationship is expressed as follows (Casella & Berger, 2002):

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + \left[Bias(\hat{\theta})\right]^2 , \quad (2.7)$$

Based on the preceding theoretical foundations, it becomes clear that shrinkage estimators practically embody the principle of equilibrium between bias and variance (the trade-off between bias and variance). This equilibrium is not merely a mathematical formula, but rather the cornerstone for constructing stable and generalizable statistical models across various fields. To illustrate the relationship described in Equation (2.7), Figure (2.1) shows the trade-off between bias and variance as a function of the shrinkage parameter. As the value of the shrinkage parameter increases (moving along the horizontal axis), the variance of the model decreases steadily, while the square of the bias increases simultaneously. The total mean squared error, represented by the U-shaped curve, initially decreases to reach an optimal minimum point and then rises again as bias begins to dominate the model. This minimum point on the mean squared error curve represents the ideal equilibrium at which the shrinkage estimator achieves the highest possible predictive accuracy. In light of this, the research now moves to reviewing the practical applications of these estimators in the fields of medicine, economics, education, engineering, and artificial intelligence, where controlling variance and improving accuracy are among the most prominent statistical challenges.

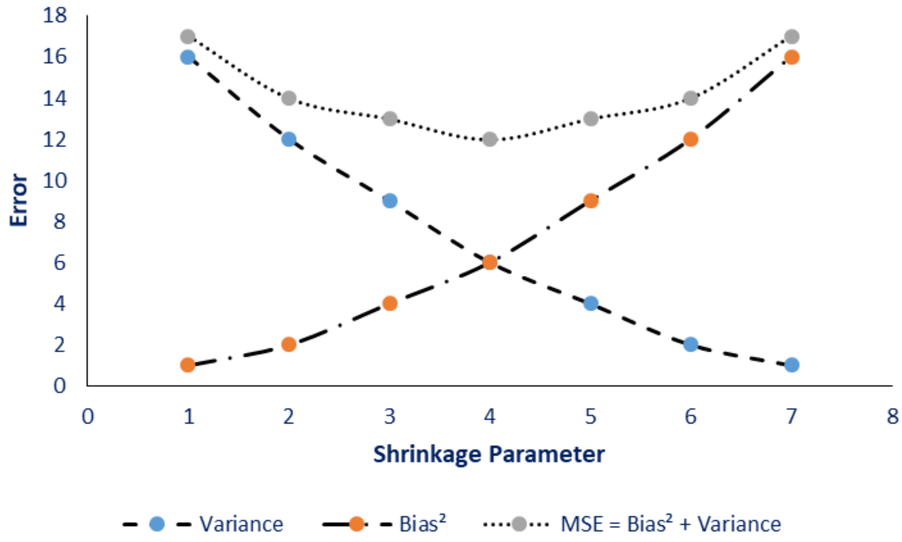

Figure 2.1: The bias-variance tradeoff as a function of the shrinkage parameter.

3. Applications of Shrinkage Estimators

The importance of shrinkage estimators is not limited to being a mathematical innovation; their true value lies in their practical applications, which have proven effective in a variety of fields. This section will discuss the most prominent applications in which shrinkage estimators have proven their ability to provide accurate and effective solutions to estimation problems in medicine, economics, finance, and other fields.

3.1 Medicine

Shrinkage estimators are essential tools in high-dimensional medical analyses with a large number of variables and small sample sizes. Schäfer and Strimmer (2005) demonstrated that incorporating shrinkage into principal components analysis improves the stability of estimates and reflects the true biological structure of the data, while Lasso (Tibshirani, 1996) is an essential technique for selecting genes affecting complex diseases. Group-Lasso and overlapping/hierarchical group penalty methods have been developed to integrate multi-omics data and exploit the ensemble structure of features (Jacob et al., 2009), as highlighted in recent reviews within the Balance and Integration literature. Penalized versions of the Cox model are also used to select biomarkers associated with survival time in high-dimensional medical contexts. For instance, these methods have been applied to the Boston Lung Cancer Survival Cohort study, one of the largest cancer epidemiology cohorts, investigating the complex mechanisms of lung cancer. (Gui and Li, 2005; Salerno et al., 2023). In addition, recent studies in the field of radiomics have demonstrated the potential of combining imaging features with omics data to build more stable and accurate functional prognostic models (You et al., 2024). Finally, shrinkage estimators have been applied to improve polygenic risk scores via shrinkage of effect sizes correction techniques to reduce noise and increase generalizability (Vilhjálmsson et al., 2015).

3.2 Finance and Economics

Shrinkage estimators are essential statistical tools in financial and macroeconomic analysis, as financial models typically face the problem of high dimensionality when the number of assets or variables exceeds the number of observations. Ledoit and Wolf (2004) introduced an improved variance-covariance matrix estimator, known as the Ledoit–Wolf shrinkage estimator, which produces more stable matrices and reduces biases resulting from small samples, which has positive implications for portfolio management. Subsequent studies, such as Zou and Hastie (2005), demonstrated the effectiveness of Ridge and Elastic Net techniques in financial predictive models, by reducing overfitting and improving performance in predicting returns and market volatility. In recent years, the application of shrinkage has expanded to include Bayesian shrinkage and adaptive Lasso shrinkage methods in macroeconomic modeling. For example, Korobilis (2022) reviews Bayesian shrinkage patterns in high-dimensional models, noting that these methods enhance forecasting accuracy for dense models when the sample is limited. These applications highlight the vital role that shrinkage estimates play in enhancing the stability of investment decisions and developing more reliable tools for risk management and financial planning in complex global markets.

3.3 Education and Social Sciences

Shrinkage estimators are widely used in survey data analysis and educational predictive models, especially when multiple behavioral or psychological variables are involved. Foundational texts in statistical learning Hastie, Tibshirani, and Friedman (2009) establish that methods such as LASSO and ridge regression improve prediction accuracy and variable selection. in the social sciences, this capability is crucial for identifying the most important factors influencing outcomes like academic performance or job satisfaction. In recent years, new applications of shrinkage have emerged, including the analysis of emotional intelligence and intrinsic motivation data using Elastic Net to extract psychological patterns affecting academic achievement (Wettstein et al., 2023). The logistic LASSO technique has also been used to predict teacher burnout and analyze factors associated with job satisfaction (Wettstein et al., 2023). These trends indicate that shrinkage estimators have become a central tool in building data-driven educational and social models, enhancing the accuracy of educational decisions and policies based on advanced statistical analysis.

3.4 Industry and Engineering

Shrinkage estimators are used in the context of quality control and industrial process control, particularly for multivariate or limited-sample data, enabling the construction of more accurate and stable predictive models. For example, Ridge Regression and Lasso shrinkage estimation are traditional methods used to improve performance predictions of industrial systems and reduce variance resulting from correlations between variables or partial measurements. In recent years, shrinkage applications have expanded to include predictive maintenance in Industrial Internet of Things (IIoT) systems, where big sensor data is combined with shrinkage methods to predict component failure and reduce operation costs. For example, Çınar et al. (2020) reviewed how machine learning frameworks, incorporating shrinkage and variable selection techniques contribute to improving equipment maintenance efficiency. Likewise, researchers have also applied Elastic Net technology in multivariate contexts within production lines to detect anomalies early; Yu and Zhao (2019) demonstrated that combining a denoising network with Elastic Net produces more stable performance indicators for robust monitoring and fault isolation in complex production systems.

These applications demonstrate that shrinkage estimators have become an essential part of the Fourth Industrial Revolution, contributing to enhanced operational forecasting, intelligent quality control, and operational risk management through statistical models adapted to industrial environments with high-dimensional data or multiple sensors.

3.5 Data Science and Machine Learning

Shrinkage estimators have become an integral part of machine learning algorithms, particularly in high-dimensional predictive models that seek to improve estimation accuracy and reduce overfitting. For example, the principle of controlled bias introduced by Stein (1961) is fundamental to understanding how shrinkage estimators can minimize the mean square error (MSE). Studies such as Ledoit and Wolf (2004), Schäfer and Strimmer (2005), and Krogh and Hertz (1991) have shown that the concepts of statistical shrinkage and "regularization" have moved beyond classical models and are used across disciplines. In recent decades, applications have expanded to include techniques such as weight decay in deep learning networks. As demonstrated in the study D'Angelo et al. (2023), using shrinkage in weight updating is an effective factor for balancing the bias–variance tradeoff in modern deep networks.

4. Conclusion

This paper explores the concept of shrinkage estimators and their pivotal role in statistical estimation, from their theoretical origins and historical development to their diverse applications in medicine, finance, economics, engineering, and other fields. Studies have shown that using shrinkage techniques improves the accuracy of estimates and reduces the problem of overfitting. They also provide flexible tools for handling high-dimensional data and statistical noise. We hope this review will provide a comprehensive and clear overview of the topic, helping researchers, students, and professionals understand the theoretical foundations of estimation using shrinkage estimators and inspiring potential practical applications in various scientific and applied fields. We also hope this overview will encourage further research exploring improvements to shrinkage methodologies and expanding their applications. We recommend further studies integrating shrinkage principles with generative AI and adaptive machine learning frameworks, given their significant potential for building more transparent, stable, and effective models in highly complex environments. Ultimately, shrinkage estimation represents not only a statistical advance, but also a philosophical shift in how uncertainty is managed in intelligent systems.

References

1. Aldrich, J. (1997). RA Fisher and the making of maximum likelihood 1912-1922. Statistical Science, 12(3), 162-176.
2. Casella, G., & Berger, R. (2024). *Statistical inference*. Chapman and Hall/CRC.
3. Çınar, Z. M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. Sustainability, 12(19), 8211.
4. D'Angelo, F., Andriushchenko, M., Varre, A. V., & Flammarion, N. (2023). Why do we need weight decay in modern deep learning?. Advances in Neural Information Processing Systems, 37, 23191-23223.
5. Donoho, D. L., & Johnstone, I. M. (1994). Minimax risk over lp-balls for lp-error. Probability Theory and Related Fields, 99(2), 277-303.
6. Efron, B., & Morris, C. (1977). Stein's paradox in statistics. Scientific American, 236(5), 119-127.
7. Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character, 222(594-604), 309-368.
8. Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2). Cambridge: MIT Press.
9. Gui, J., & Li, H. (2005). Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. Bioinformatics, 21(13), 3001-3008.
10. Hastie, T. (2009). The elements of statistical learning: data mining, inference, and prediction.
11. Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical learning with sparsity. Monographs on statistics and applied probability, 143(143), 8.
12. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.
13. Jacob, L., Obozinski, G., & Vert, J. P. (2009, June). Group lasso with overlap and graph lasso. In Proceedings of the 26th annual international conference on machine learning (pp. 433-440).
14. James, W., & Stein, C. (1961, June). Estimation with quadratic loss. In Proceedings of the fourth Berkeley symposium on mathematical statistics and probability (Vol. 1, No. 1961, pp. 361-379).
15. Korobilis, D., & Shimizu, K. (2022). Bayesian approaches to shrinkage and sparse estimation. Foundations and Trends® in Econometrics, 11(4), 230-354.
16. Krogh, A., & Hertz, J. (1991). A simple weight decay can improve generalization. Advances in neural information processing systems, 4.
17. Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. Journal of multivariate analysis, 88(2), 365-411.
18. Lehmann, E. L., & Casella, G. (1998). Theory of point estimation. New York, NY: Springer New York.

19. Salerno, S., & Li, Y. (2023). High-dimensional survival analysis: Methods and applications. Annual review of statistics and its application, 10(1), 25-49.
20. Schäfer, J., & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Statistical applications in genetics and molecular biology, 4(1).
21. Stein, C. (1956, January). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In Proceedings of the third Berkeley symposium on mathematical statistics and probability, volume 1: Contributions to the theory of statistics (Vol. 3, pp. 197-207). University of California Press.
22. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267-288.
23. Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., ... & Marsal, S. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. The american journal of human genetics, 97(4), 576-592.
24. Wettstein, A., Jenni, G., Schneider, I., & Kühne, F. (2023). grosse Holtforth, M.; La Marca, R. Predictors of Psychological Strain and Allostatic Load in Teachers: Examining the Long-Term Effects of Biopsychosocial Risk and Protective Factors Using a LASSO Regression Approach. Int. J. Environ. Res. Public Health, 20, 5760.
25. You, C., Su, G. H., Zhang, X., Xiao, Y., Zheng, R. C., Sun, S. Y., ... & Gu, Y. J. (2024). Multicenter radio-multiomic analysis for predicting breast cancer outcome and unravelling imaging-biological connection. NPJ Precision Oncology, 8(1), 193.
26. Yu, W., & Zhao, C. (2019). Robust monitoring and fault isolation of nonlinear industrial processes using denoising autoencoder and elastic net. IEEE Transactions on Control Systems Technology, 28(3), 1083-1091.
27. Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology, 67(2), 301-320.

تطور مقدرات الانكماش بين النظرية الإحصائية والتطبيقات الذكية الحديثة

عائشة محمد ابوطرطور
قسم الإحصاء، كلية العلوم، جامعة طرابلس، طرابلس، ليبيا
A.Abutartur@uot.edu.ly

الملخص

شهد علم الإحصاء الرياضي تحولًا جذريًا خلال القرن الماضي مع اكتشاف مُقدِّرات الانكماش، التي قدمت نسخة مُحسَّنة من المنهج الكلاسيكي القائم على المُقدِّرات غير المتحيزة. منذ اكتشاف نظرية شتاين في أوائل الستينيات، أصبح قبول درجة من التحيز أداة فعالة لتقليل الخطأ الإجمالي وإنتاج مُقدِّرات أكثر استقرارًا. تهدف هذه الورقة إلى تقديم مراجعة شاملة للتطور التاريخي والنظري لمُقدِّرات الانكماش، بدءًا من نموذج جيمس-شتاين وصولًا إلى النماذج الحديثة مثل انحدار ريدج (RIDGE REGRESSION)، ومعامل الانكماش والاختيار المطلق الأدنى (LASSO)، وانحدار الشبكة المرنة (ELASTIC NET REGRESSION). كما تُركز على أساسها الإحصائي ودورها المتنامي في تطوير نماذج الذكاء الاصطناعي والتعلم العميق. وتختتم الورقة بعرض رؤية موحدة تربط النظرية الإحصائية بالتطبيقات الذكية الحديثة من خلال تبني مُقدِّرات الانكماش كإطار عمل متكامل يجمع بين الدقة الرياضية والابتكار التطبيقي في بيئات البيانات المعقدة. يوضح هذا الإطار الموحد كيف تطورت مقدرات الانكماش لتصبح حجر الزاوية في النمذجة التنبؤية الذكية.

استلمت الورقة بتاريخ 2026/02/22، وقبلت بتاريخ 2026/03/11، ونشرت بتاريخ 2026/03/12

الكلمات المفتاحية: الذكاء الاصطناعي (AI)، التحيز والتباين، استقرار النموذج، تنظيم النموذج، التجاوز، مقدر الانكماش، التقدير الإحصائي.