The Anderson-Darling Test

Study of the Authors' Methodology

Reda Mdair

2024

1 Introduction

Introduced in 1954, the Anderson-Darling test is a non-parametric approach used to assess whether a sample follows a given probability distribution. Its most popular application is normality testing.

Initially, this test assumed that the parameters of the sample were known. A few years later, statistician Michael A. Stephens refined this approach by incorporating parameter estimation for unknown parameters. With these advancements, the Anderson-Darling test has become a comprehensive and powerful tool for goodness-of-fit testing.

Sharing the same principles as the Kolmogorov-Smirnov test, it is often regarded as a variant of the latter. The main difference lies in the weighting of the test statistic. The Kolmogorov-Smirnov statistic is more sensitive to deviations near the median, whereas the Anderson-Darling statistic is more influenced by extreme values.

The objective of this project is to analyze the two papers published by the authors of this test. We will interpret the key findings to address the central question: what are the advantages and limitations of the Anderson-Darling test compared to other classical approaches?

2 Test Construction

Let $X = (X_1, \dots, X_n)$ be a sample of n i.i.d. random variables from an unknown distribution P. The goal is to test whether this distribution follows a specified continuous distribution P^1. Given an observation $x = (x_1, \dots, x_n)$, we formulate the hypotheses:

$$(H_0) : P = P^1 \qquad\qquad (H_1) : P \neq P^1$$

Denoting F and F^1 as the cumulative distribution functions (CDFs) associated with P and P^1, respectively, Anderson and Darling proposed the following two test statistics:

$$W_n^2(X) = n \int_{\mathbb{R}} [F_n(x) - F^1(x)]^2 \psi[F^1(x)] \, dF^1 \tag{1.0}$$

$$K_n(X) = \sup_{x \in \mathbb{R}} \sqrt{n} |F_n(x) - F^1(x)| \sqrt{\psi[F^1(x)]} \tag{2.0}$$

where $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i \leq x}$ is the empirical cumulative distribution function (ECDF) of the sample, and $\psi(t)_{0 \leq t \leq 1}$ is a predefined positive weighting function. This function allows emphasis to be placed on specific regions of the distribution.

The objective is then to determine the asymptotic distributions of W_n^2 and K_n. This will enable precise computation of their respective critical values, z_1 and z_2, particularly for large sample sizes.
A key aspect lies in the choice of the weighting function $\psi(t)_{0\leq t\leq 1}$. This function is left to the discretion of the statistician to weight discrepancies according to the importance assigned to different parts of the distribution. For instance, when $\psi(t) = 1$, W_n^2 corresponds to the Cramér-von Mises criterion, while K_n corresponds to the Kolmogorov-Smirnov criterion.

By ordering the observations, the test statistics simplify. Let $(X_{(1)}, \ldots, X_{(n)})$ be the order statistic associated with the sample X. Since $F^1(x)$ is assumed to be continuous, the variable transformation $u = F^1(x)$ is well-defined, leading to the transformed observations $u_i = F^1(x_i), i \in \{1, \ldots, n\}$. Under hypothesis (H_0), these can be considered as realizations from the uniform distribution on $[0, 1]$. Denoting $G_n(u)$ as the empirical cumulative distribution function (ECDF) of $(u_i)_{1\leq i\leq n}$, we obtain the new formulations:

$$W_n^2(X) = n \int_0^1 [G_n(u) - u]^2 \psi(u)\, du \tag{1.1}$$

$$K_n(X) = \sup_{0\leq u\leq 1} \sqrt{n}|G_n(u) - u|\sqrt{\psi[u]} \tag{2.1}$$

To study the asymptotic properties of the test statistics, Anderson and Darling define:

$$Y_n(u) = \sqrt{n}[G_n(u) - u], u \in [0, 1].$$

The advantage of this transformation is that the joint distribution of $Y_n(u_1), \ldots, Y_n(u_k)$ (for fixed $u_1, \ldots, u_k$) approaches a multivariate normal distribution asymptotically:

$$(Y_n(u_1), \ldots, Y_n(u_k)) \xrightarrow[n\to\infty]{\mathcal{L}} (y(u_1), \ldots, y(u_k))$$

where $y(u_i)$ are normally distributed centered variables with covariance:

$$Cov(y(u), y(v)) = E[y(u)y(v)] = \min(u, v) - uv$$

This step is crucial in their research towards the asymptotic distribution, as the normal distribution allows the application of numerous theorems in stochastic processes.

Following this approach, they define the two quantities:

$$A_n(z) = \mathbb{P}(W_n^2 \leq z) = \mathbb{P}\left(\int_0^1 Y_n^2(u)\psi(u)\, du \leq z\right)$$

$$B_n(z) = \mathbb{P}(K_n \leq z) = \mathbb{P}\left(\sup_{0\leq u\leq 1} |Y_n(u)|\sqrt{\psi[u]} \leq z\right)$$

The objective is then to determine:

$$A_n(z) = a(z) = \mathbb{P}\left(\int_0^1 y^2(u)\psi(u)\, du \leq z\right)$$

$$B_n(z) = b(z) = \mathbb{P}\left(\sup_{0\leq u\leq 1} |y(u)|\sqrt{\psi[u]} \leq z\right)$$

in order to establish a solid foundation for computing the limiting distributions of W_n^2 and K_n.

Donsker's theorem allows them to conclude that $A_n(z) = a(z)$. This result is built upon the assumption that the function $\psi(u)$ is bounded. The two authors further generalize this result, under specific conditions, to more common weighting functions $\psi(u)$.

They then proceed to test different choices for the function ψ through examples. For the test statistic W_n^2, they consider:

$$\psi_1(u) \equiv 1 \qquad\qquad \psi_2(u) = \frac{1}{u(1-u)}$$

The second choice proves to be more interesting, as it emphasizes deviations in the tails of the distribution. In other words, the function $\psi_2(u) = \frac{1}{u(1-u)}$ is more effective at detecting deviations near 0 and 1. The first choice, on the other hand, does not favor any particular part of the distribution and assigns uniform weight across the entire interval $[0, 1]$.

For the test statistic K_n, the considered weighting functions are more technical:

$$\psi_3(u) = q_k \geq 0 \qquad \forall u_k \leq u \leq u_{k+1}, \quad k \in \{1, \dots, n\}, \quad u_0 = 0, u_{n+1} = 1$$

$$\psi_4(u) = \frac{1}{u(1-u)} \mathbb{1}_{0<a\leq u\leq b<1}$$

Although these choices appear promising, interpreting the results proves to be quite challenging. Moreover, certain theoretical uncertainties reduce the reliability of the asymptotic distribution associated with K_n, leading to its exclusion from further consideration.

Ultimately, the test statistic W_n^2 is the one retained in Anderson and Darling's second paper, using the weighting function $\psi(u) = \frac{1}{u(1-u)}$. The formula given in (1.0) is then simplified as follows:

$$\boxed{W_n^2(X) = -n - \tfrac{1}{n}\textstyle\sum_{i=1}^{n}(2i-1)\left[\ln\left(F^1(X_{(i)})\right) + \ln\left(F^1(X_{(n-i+1)})\right)\right]}$$

with its characteristic function:

$$\phi(t) = \sqrt{\frac{-2\pi i t}{\cos(\frac{\pi}{2}\sqrt{1+8it})}}$$

These findings enabled the authors to construct the statistical table that now bears their names.

Remark: In 1979, statistician Michael A. Stephens proposed, for the case where the parameters are unknown, the estimators $\mu = \bar{X}$ and $\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2$ for the normality test. He then defined the modified test statistic:

$$W^* = W_n^2 \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right)$$

3 Discussions

Let us recall the formula for the test statistic associated with the Kolmogorov-Smirnov approach:

$$T_{KS}(X) = \sqrt{n} \max_{1 \le i \le n} \left\{ \max \left\{ \left| F^1(X_{(i)}) - \frac{i}{n} \right|, \left| F^1(X_{(i)}) - \frac{i-1}{n} \right| \right\} \right\}$$

Its interpretation is straightforward: the distance is based on the largest discrepancy between the empirical and hypothetical cumulative distribution functions.

This raises an important question: is the information utilized by this statistic sufficient? To answer this, it is useful to introduce the Cramér-von Mises test (1928), which is based on the following distance measure:

$$T_{CVM}(X) = n \int_{\mathbb{R}} \left[F_n(x) - F^1(x)\right]^2 \, dF^1$$

This function appears to be more qualitative than that of Kolmogorov-Smirnov. Indeed, rather than measuring only the maximum vertical deviation, it quantifies the overall goodness-of-fit across the entire distribution. Therefore, it may seem more reliable for goodness-of-fit testing.

However, the Cramér-von Mises approach presents a slight imprecision. Specifically, it does not account for the fact that the empirical cumulative distribution function (ECDF) exhibits greater variability near the median than at the extremes. This reasoning helps to explain the motivation behind Anderson and Darling's approach. The function $\psi(u) = \frac{1}{u(1-u)}$ (the inverse of the variance of the empirical CDF) is introduced to weight the statistic $T_{CVM}(X)$, assigning greater importance to deviations at the extremes.
It is often preferable to prioritize accuracy in low-density regions (extremes) rather than in high-density regions (the median).

The Anderson-Darling technique is thus best suited for tests involving heavy-tailed distributions. Examples include the Pareto distribution, the Cauchy distribution, and the Lévy distribution. Furthermore, many problems require special attention to extreme regions (such as rare events), and the A-D approach is particularly effective in addressing these cases.

However, when testing for conformity to a uniform distribution, the Kolmogorov-Smirnov test is clearly advantageous. Since the hypothetical CDF increases linearly, measuring cumulative deviations uniformly across the entire interval is appropriate. In this case, the use of the Anderson-Darling test is unnecessary, and its computed distance may be excessively sensitive.

4 Summary of Results

The Anderson-Darling test is a highly powerful tool that has expanded the range of options available for goodness-of-fit testing. Its formula measures discrepancies across the entire interval of the empirical and theoretical distributions, assigning greater weight to the tails. As a result, it is particularly well-suited for heavy-tailed distributions. Additionally, it is useful for analyzing rare events, which require careful consideration of observations that deviate significantly from the median.

However, it remains important to alternate between different tests, such as K-S and C-V-M. The study of uniformity testing highlights the limitations of the Anderson-Darling approach. In this context, the test may produce an excessively high statistic by overemphasizing weight in non-critical regions. This could make it overly conservative, leading to an increased likelihood of rejecting the null hypothesis (H_0) of conformity.

The two articles analyzed in this study illustrate the trade-offs that statisticians face when designing hypothesis tests. Decisions must be made regarding which regions of the distribution should be measured, how to weight each of these regions appropriately, and how to minimize computational costs. This makes the research process highly complex.

References

[1] Anderson, T. W., and Darling, D. A. *Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes.* Annals of Mathematical Statistics, 23 (1952), 193-212.

[2] Anderson, T. W., and Darling, D. A. *A Test of Goodness of Fit.* Journal of the American Statistical Association, Vol. 49, No. 268. (Dec., 1954), pp. 765-769.

[3] Stephens, M. A. *The Anderson-Darling Statistic.* Technical report no. 39 - October 31, 1979.
Available online : `https://apps.dtic.mil/sti/pdfs/ADA079807.pdf`