January 2023

Uniformity test based on the empirical Bernstein distribution

Ran Sun
sun16j@uwindsor.ca

Uniformity Test based on the empirical Bernstein distribution

by

Ran Sun

A Major Research Paper
Submitted to the Faculty of Graduate Studies
through the Department of Mathematics and Statistics
in Partial Fulfillment of the Requirements for
the Degree of Master of Science at the
University of Windsor

Windsor, Ontario, Canada

Uniformity Test based on the empirical Bernstein distribution

by

Ran Sun

APPROVED BY:

A. Hussein
Department of Mathematics and Statistics

M. Belalia, Co-Advisor
Department of Mathematics and Statistics

S. Nkurunziza, Co-Advisor
Department of Mathematics and Statistics

December 19, 2022

Author's Declaration of Originality

I hereby certify that I am the sole author of this paper and that no part of this paper has been published or submitted for publication.

I certify that, to the best of my knowledge, my paper does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my paper, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my paper and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my paper, including any final revisions, as approved by my paper committee and the Graduate Studies office, and that this paper has not been submitted for a higher degree to any other University or Institution.

Abstract

In this paper, we firstly review the origin of Bernstein polynomial and the various application of it. Then we review the importance of goodness-of-fit test, especially the uniformity test, and we examine lots of different test statistics proposed by far. After that we suggest two new statistics for testing the uniformity. These two statistics are based on Komogorov-Smirnov test type and Cramér-Von Mises test type, respectively. Also we embed Bernstein polynomial into those test type and take advantage of great approximation performance of this polynomial. Finally, we run a Monte-Carlo simulation to compare the performance of our statistics to those without embedding the Bernstein polynomials. We compare their performance in term of powers and inefficiencies. We found that by choosing suitable value for parameter, our statistics can perform better than the original form in most of the cases. The suggestion of choosing optimal value will be given.

To my loving grandparents

Qinggui Li and Tianhui Zhang

Acknowledgments

I would like to thank my supervisors, Dr. Mohamed Belalia and Dr. Sévérien Nkurunziza for their constant support and guidance during my graduate study since I started my M.Sc. program, especially in every stage of this paper. During my program, they never hesitated to provide me suggestions whenever I had difficulties in study and research. These support and suggestions made me finally succeed in this work.

I would also like to thank Dr. Hussein for being my department reader. He also taught me course of statistical analysis, which is very useful for practical applications of theoretical results.

In addition, I would like to thank all faculty and staff members, and the graduate students in the Department of Mathematics and Statistics who helped me in many different ways during my study.

Moreover, my profound gratitude goes to my grandparents. They have given me constant love, encouragement and support since my birth, and have guided me to face my life positively and go ahead in the right direction. Finally, I would like to thank everyone who directly or indirectly helped me in the completion of this paper.

Contents

List of Tables

Chapter 1

Introduction

Goodness of fit test is a crucial tool in statistical analysis that has developed very broadly by statisticians in the past several decades. The interests of it come from the widely-used nature of this method. Statisticians develop theoretical results for more effective testing, and scientists in other fields use this testing method to model real problems. These fields include economics (see Chandra and Singpurwalla (1981) for discussion of Lorentz Curve and Gini Index), finance (see Dang et al. (2018)), medical research (see Boyle et al. (1997)) and biochemistry (see Kornegay et al. (1993) for goodness-of-fit test of arcelin segregation curves), etc. In this paper, we will mainly consider uniformity test. Indeed, the hypothesis testing of some samples follow other distribution can be transferred to an uniformity test problem. For example, if we want to test whether some samples $X_i's$ are taken from a cumulative distribution function (c.d.f.) F, we can instead do a uniformity test for $F(X_i)'s$, as $X_i's$ have c.d.f. of F is equivalent to $F(X_i)'s$ are distributed uniformly.

In Chapter 2 entitled **"Literature Review"**, we will review the background and recent research results of Bernstein polynomial. Bernstein polynomial was firstly served as a constructional proof of Stone–Weierstrass theorem. The quoted theorem states that every continuous bounded real-valued function f on a closed interval $[a, b]$ can be uniformly approximated by a sequence of polynomials with real-valued coefficients. For the construction of this polynomial and the proof of uniformly convergence, see the original paper Bernshtein (1952). After that, Bernstein polynomial received spotlight in the following years because of its excellent performance in approximate continuous functions. Bernstein polynomial also plays an important role in statistics. As we all know, probability density function (p.d.f.) and cumulative density function (c.d.f.) are both bounded and continuous functions in most of cases. As a result, they can be approximated by Bernstein polynomials. The behavior of such approximation was studied deeply in the last several decades and the effectiveness of that approximation was compared to other classical estimators, in terms of bias and mean squared error (MSE), etc. Further, when Bernstein polynomial is applied in multivariate and conditional density distribution case, it also works well. Recently, Bernstein polynomial was applied in copula analysis to testing the independence of two random variables. The proposed method is also based on this famous polynomial.

By far, lots of statistics for uniformity test are developed, as we will review some of them in Section 2.3 entitled **"An overview of uniformity tests"**. However, no test statistic is the uniformly most powerful. To evaluate the efficiency of test statistics, we use the power function for a given significance level. Some of the statistics have satisfactory power when dealing with some certain alternative distributions, but works bad when dealing with other ones. We will also see some statistics which are good when dealing with special cases. For example, one statistic was proposed to deal with data which have significant covariance between them. Some statistics have high power almost universal for all variation of distributions, but they have complicated forms and so require huge load of calculation. The most useful statistic is required to have considerably high power against most of variation of distributions, and simultaneously does not need too heavy computation load. In this paper, we consider to improve two famous statistics, C_n and S_n (see Marhuenda et al. (2005)). We will show in the simulation study that these two new statistics perform better than C_n and S_n when dealing with two families of common-seen alternative distributions. Further, as these two improved statistics basically keep the original expression, the calculation load will not be increased too much.

In Chapter 3 entitled **"Uniformity tests based on Bernstein polynomials"**, we give theoretical result of those two new statistics, including the explicit expression and their asymptotic distributions. We use the generalized continuous mapping theorem (GCMT) to prove their asymptotic distributions. The empirical process can be expressed as follows:

$$\mathbb{F}_n(x) = \sqrt{n}(F_n(x) - F(x)), \qquad 0 \le x \le 1$$

where F_n stands for the empirical cumulative distribution function (e.c.d.f.) and F is the c.d.f. Van der Vaart (2000) shows that this $\mathbb{F}_n$ converges in distribution to a Gaussian process as n goes to infinity. We show that after embedding the Bernstein polynomial, the resulting Bernstein empirical process also converges in distribution to the same limit. Then, a simulation study follows to show their good performance compared to the original form. For the convenience of readers, recall the definition of the Gaussian process:

Definition 1.1. *(Gaussian process)*

A continuous time stochastic process $\{X_t; t \in T\}$ *is Gaussian if and only if for every finite set of indices* $t_1, \ldots, t_k$ *in the index set* T, $\mathbf{X}_{t_1,\ldots,t_k} = (X_{t_1}, \ldots, X_{t_k})$ *is a multivariate Gaussian random vector. Or in other words, every linear combination of* $(X_{t_1}, \ldots, X_{t_k})$ *has a univariate normal (or Gaussian) distribution.*

As the explicit expressions of the statistics are very complicated, we take advantage of strong law of large number (SLLN) to compute the value of them. We use two classical families of distributions to cover most type of alternative distributions, for a thoroughly overview of how our statistics performs. Following that, we give suggestions on the best situation of when to use each statistics and the optimal parameter to choose. Finally, we suggest some future research directions.

Chapter 2

Literature Review

Bernstein polynomial is currently widely used in scientific research. In fields of Mathematics and Engineering, this famous polynomial is employed to smooth and approximate functions. For example, in field of control engineering, Nataraj and Arounassalame (2007) use Bernistein polynomial approach to improve an algorithm for unconstrained global optimization. In this chapter, we introduce Bernstein polynomial from its origin. As we can see, it is closely related to a theorem in topology.

2.1 Weierstrass Theorem and Bernstein Polynomials

Some basic concepts in topology are needed for this section, which can be found in Appendix A. We first introduce the following theorem:

Theorem 2.1. *(Stone–Weierstrass theorem) Let K be a compact Hausdorff space, let $\mathcal{A}$ be a subalgebra of $C(K, \mathbb{R})$ with the following properties:*

- $1 \in \mathcal{A}$
- $\exists f \in \mathcal{A}$ *such that* $f(x) \neq f(y) \quad$ *for* $\forall x \neq y \in K$

Then, we have: $\mathcal{A}$ *is dense in* $C(K, \mathbb{R})$*. Namely, every element in* $C(K, \mathbb{R})$ *is either in* $\mathcal{A}$ *or a limit point of* $\mathcal{A}$*.*

This theorem in topology was presented and a proof was given in Stone (1937), and then Stone gave a simplified proof in Stone (1948). There are lots of generalized and applied form of this theorem. Among those, one application of this theorem, namely, the Weierstrass approximation theorem, is closely related to our topic, and it is stated as follows:

Theorem 2.2. *(Weierstrass approximation theorem) Let* $a < b \in \mathbb{R}$ *and* f *be a continuous function mapping on the closed interval* $[a, b]$ *to* $\mathbb{R}$*. Then, there exists a sequence of polynomials with coefficients in* $\mathbb{R}$ *that converges to* f *uniformly on* $[a, b]$*.*

There are only theoretical proof available for the above theorem until 1940s. Then, Bernstein gave a constructional proof of the above theorem by proposing the following Bernstein polynomials.

Definition 2.1. *The Bernstein polynomial* $\widehat{f}_m(x)$ *of a continuous function* $f(x)$ *mapping from a closed interval* $[a, b]$ *to* $\mathbb{R}$ *is defined as follows:*

$$\widehat{f}_m(x) = \sum_{k=0}^{m} f\left(\frac{k}{m}\right) P_{m,k}(x), \qquad k = 0, ..., m$$

where $P_{m,k}(x) = \binom{m}{k} x^k (1-x)^{m-k}, \quad k = 0, ..., m$, stands for the binomial probability mass function and m is a bandwidth parameter.

It can be shown that $\widehat{f}_m(x)$ is a continuous function which converges to $f(x)$ uniformly on $[a, b]$, as m tends to infinity. The proof follows from the following theorem which is given in Feller et al. (1965).

Theorem 2.3. *(Feller et al. (1965)) If $f(x)$ is a continuous bounded function mapping from $[0, 1]$ to $\mathbb{R}$, then $\widehat{f}_m(x) = \sum_{k=0}^{m} f\left(\frac{k}{m}\right) P_{m,k}(x)$ converges uniformly to $f(x)$ on $[0, 1]$, as m goes to infinity.*

More properties of Bernstein polynomial are thoroughly discussed in Lorentz (1986). An illustration on Bernstein approximation is given below:

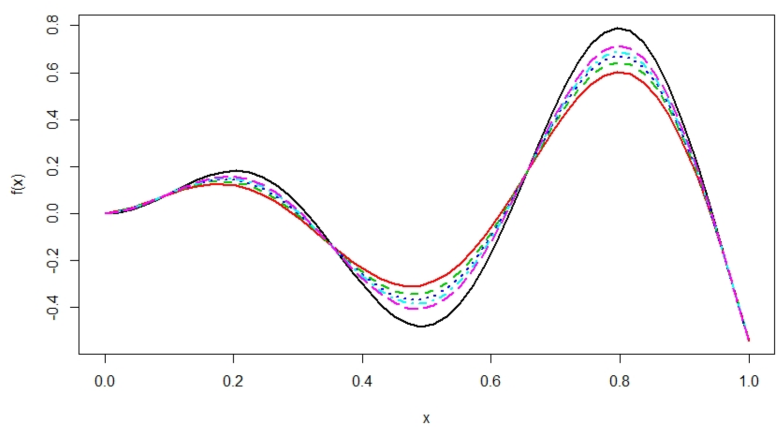

Figure 2.1: Bernstein Approximation of a Function $f(x) = x \sin(10x)$

Figure 2.1 gives graph of original function ($f(x) = x \sin(10x)$, in black line) and its Bernstein polynomial approximation for $k = 30, 40, 50, 60, 80$ respectively. We can see from Figure 2.1 that as the value of k increases, the Bernstein polynomial gets closer to the approximated function.

2.2 Applications of Bernstein Polynomials

2.2.1 Estimation of Unknown Density Function

Started from the 1970s, Bernstein polynomial was more deeply studied, and its usage was no longer confined to be a proof of theorem. Vitale (1975) developed a Bernstein polynomial density function estimator for estimating an unknown p.d.f. Let $\mathbb{I}_A$ denote the indicator function of the event A. Let $X_1, \ldots, X_N$ be i.i.d. random variables from an unknown density function f, which has positive density only on interval $[0, 1]$, then a Bernstein polynomial density function estimator proposed by Vitale (1975) is constructed as follows: for positive integer n, firstly define

$$m_{jn}^N = \sum_{i=1}^{N} \mathbb{I}\left(X_i \in \left[\frac{j}{n+1}, \frac{j+1}{n+1}\right]\right), \qquad j = 0, 1, \ldots, n. \tag{2.1}$$

Then, the estimator introduced by Vitale (1975) is as follows:

$$\widehat{f}_{nN}(x) = \frac{n+1}{N} \sum_{j=0}^{N} m_{jn}^N \binom{n}{j} x^j (1-x)^{n-j}.$$

Vitale (1975) argued that this density function estimator has comparable order of convergence as its competitors, such as kernel estimators. However, this new estimator has significant lower load of computation which made it better than competitors in practical usage. After that, Tenbusch (1994) extended this estimation of unknown marginal p.d.f. to the case of joint p.d.f. To estimate the joint p.d.f. $f(x, y)$, he used a similar method of Bernstein polynomial as follows: Suppose that $(X_1, Y_1), \ldots, (X_N, Y_N)$ are i.i.d. vectors from an unknown joint p.d.f. $f(x, y)$, which has positive density in unit square region, and let n be a positive integer which depends on the sample size N. We firstly define

$$A_{jin} = \left[(x, y) : \frac{j}{n+1} < x \leq \frac{j+1}{n+1}, \frac{i}{n+1} < y \leq \frac{i+1}{n+1}\right]$$

and then,

$$m_{jin}^{(N)} = \sum_{i=1}^{N} \mathbb{I}((X_i, Y_i) \in A_{jin}).$$

Then, the Bernstein polynomial density estimator proposed by Tenbusch (1994) of $f(x, y)$ is given as follows:

$$\widehat{f}_{nN}(x, y) = \frac{(n+1)^2}{N} \sum_{j=0}^{n} \sum_{i=0}^{n} m_{jin}^{(N)} p_{jn}(x) p_{in}(y),$$

where $p_{jn}(x) = \binom{n}{j} x^j (1-x)^{n-j}$.

Tenbusch (1994) argued that this estimator has bias at non-boundary region comparable in terms of order to classical kernel estimator, which is shown in Theorem B.1 given in Appendix B. Further, this estimator is unbiased at boundary, whereas kernel estimator is biased at that points. Also, on the border, or in other words, when either x or y is equal to 0 or 1, and the other takes value between them, the bias of this new estimator will outperform classical one, as shown in Theorem B.2. Taking the lower computation load and lower bias at boundary and border into consideration, this new estimator outperforms classical kernel one in similar way as the univariate case.

2.2.2 Approximation of p.d.f. and c.d.f

Recently, Bernstein polynomial was used to approximate p.d.f. and c.d.f. using an empirical distribution function approach. We point out that the empirical cumulative density function (ecdf) is a step function. Nevertheless, as it was proposed in Babu et al. (2002), the ecdf can be smoothed by using Bernstein polynomials. Specifically, the resulting empirical Bernstein distribution estimator is given by:

$$\widehat{F}_{n,m}(x) = \sum_{k=0}^{m} F_n\left(\frac{k}{m}\right) P_{m,k}(x), \qquad k = 0, ..., m, \tag{2.2}$$

Also, some properties of the difference between ecdf and its Bernstein approximation were studied in Babu et al. (2002). The result is stated in the following theorem.

Theorem 2.4. *(Babu et al. (2002)) Let F be continuous and differentiable on the interval $[0,1]$ with density f. If f is Lipschitz of order 1, then for $n^{2/3} \leq m \leq (n/\log n)^2$, we have a.s. as $n \longrightarrow \infty$,*

$$\sup_{0\leq x\leq 1} |\widehat{F}_{m,n}(x) - F_n(x)| = O((n^{-1}\log n)^{1/2}(m^{-1}\log m)^{1/4}).$$

Therefore, for $m = n$, we will have:

$$\sup_{0\leq x\leq 1} |\widehat{F}_{n,n}(x) - F_n(x)| = O(n^{-3/4}(\log n)^{3/4}) \qquad a.s.$$

The Bernstein probability density function estimator $\widehat{f}_{m,n}(x)$ is the derivative of $\widehat{F}_{m,n}(x)$ with respect to x, which is expressed as:

$$\widehat{f}_{m,n}(x) = \sum_{k=0}^{m-1} \left[F_n\left(\frac{k+1}{m}\right) - F_n\left(\frac{k}{m}\right) \right] \beta_{k+1,m-k}(x)$$

where $\beta_{a,b}(x)$ stands for the beta density of parameters a and b :

$$\beta_{a,b}(x) = \frac{\gamma(a+b)}{\gamma(a)\gamma(b)} x^{a-1}(1-x)^{b-1} \qquad for \qquad x \in [0,1]$$

and $\beta_{a,b}(x) = 0$ otherwise.

Theorem B.3 gives the rate at which the Bernstein estimator of p.d.f. $(\widehat{f}_{m,n}(x))$ converges to f.

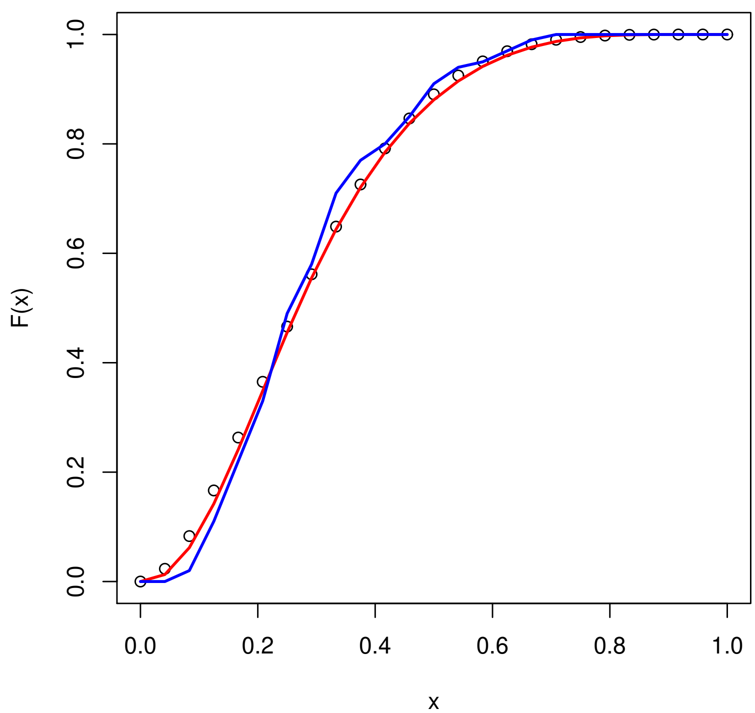

Figure 2.2: Distribution function estimation based on Bernstein polynomials

Figure 2.2 shows the Bernstein approximation of c.d.f. of a $Beta(2,5)$ distribution, with sample size $n = 100$ and optimal m value which is 25.
Further, the asymptotic normality was studied in Babu et al. (2002). For this paper to be self-contained, we recall this result in Theorem B.4 in Appendix.
Moreover, Leblanc (2010) studied the bias and mean squared error (MSE) of $\widehat{F}_{m,n}(x)$ and $\widehat{f}_{n,m}(x)$. See Theorem B.5 by Leblanc (2010) for details. The significance of this theorem is that the Bernstein estimator of p.d.f has uniform bias over the whole interval $[0,1]$. Therefore, by using this estimator, one can get rid of any boundary bias. Following that, Leblanc (2012) derived a theorem which gives the bias for Bernstein estimator of c.d.f. For this paper to be self-contained, we recall this result in the Appendix (see Theorem B.6)
Also, the MSE of this Bernstein estimator of c.d.f. is given by Theorem B.7.
As a result, by the fact that

$$\text{MSE}[\widehat{F}_n(x)] = n^{-1}F(x)[1 - F(x)].$$

Leblanc (2012) argued that by artificially choosing good value of m, the estimator $\widehat{F}_{m,n}(x)$ dominates $\widehat{F}_n(x)$ over the whole interval $[0,1]$ in terms of MSE.
Finally, in Leblanc (2012), the author establishes the asymptotic normality of $\widehat{F}_{m,n}(x)$. For this paper to be self-contained, we recall this result in the Appendix B (see Theorem B.8)
As we see from the above, the theoretical results are pretty firmly built for Bernstein estimator. Further, lots of articles use simulation to approximate the real distribution function using Bernstein polynomial. For example, in Babu et al. (2002), different

combinations of m and n are selected to generate Bernstein approximation for c.d.f. and p.d.f. of Beta, exponential and normal distributions. It is found in Babu et al. (2002) that among all combinations, $m = n/\log n$ works best for large n. Further, it outperforms the kernel estimator by a huge amount. We can clearly see this trend in the figures given in the Appendix.

After that, Leblanc (2012) gave more emphasis on the boundary bias of estimators on his graphic illustrations. He uses two real datasets and applies several different estimators to approximate the density function. As we can see from graphs in Leblanc (2012), Bernstein estimator outperforms others at boundary. It coincides with the real density very well near points of 0 and 1, whereas other estimators are kinds of out of control at these points. Finally, the performance of Bernstein estimator in terms of mean-integrated squared error (MISE) was also included in Leblanc (2012). For the convenience of the reader, recall that MISE of an estimator $\widehat{F}$ of a c.d.f F is defined as

$$\text{MISE}(\widehat{F}) = \mathbb{E}\left[\int_0^1 \left(\widehat{F}(x) - F(x)\right)^2 dx\right].$$

Leblanc (2012) compare MISE of Bernstein estimator, Empirical c.d.f. (F_n) and Kernel estimator for a $Beta(2, 1)$ distribution. He fixed the sample size n and plot MISE for different values of m. In his graph, we can see clearly that if we choose m deliberately according to the value of n, the MISE of Bernstein estimator can be lower to about half of MISE of other two estimators.

2.2.3 Estimation of Conditional Density Function

Bernstein polynomials are recently developed to have other applications. For example, given i.i.d. $(X_1, Y_1), \ldots, (X_n, Y_n)$ with joint continuous c.d.f., they are used for estimating the conditional c.d.f., which is defined as:

$$F_y(x) = \mathbb{P}[Y < y | X = x]. \tag{2.3}$$

A classical approach is given by Nadaraya (1964), which is given as follows:

$$\widehat{F}_{x,h}(y) = \frac{\sum_{i=1}^{n} W_h(x - X_i)\mathbb{I}(Y_i < y)}{\sum_{j=1}^{n} W_h(x - X_j)} = \sum_{i=1}^{n} w_{i,h}(x)\mathbb{I}(Y_i < y) \tag{2.4}$$

where the weights

$$w_{i,h}(x) = \frac{W_h(x - X_i)}{\sum_{j=1}^{n} W_h(x - X_j)}$$

where W_h is the kernel function and h is bandwidth parameter. However, a well-known problem of this estimator is that it has a boundary bias which is larger than the bias in other region by a full magnitude h. Belalia et al. (2017) suggested to use Bernstein polynomial to smooth the conditional c.d.f. and the estimator is given by:

$$\widehat{F}^{B(2)}_{x,mh}(y) = \sum_{k=0}^{m} \widehat{F}^{(2)}_{x,h}(k/m) P_{m,k}(y) \tag{2.5}$$

where $P_{m,k}(y)$ is defined as previous and $\widehat{F}^{(2)}_{x,h}$ is given by:

$$\widehat{F}^{(2)}_{x,h}(y) = \frac{\sum_{i=1}^{n} W_i^*(x,h)\mathbb{I}(Y_i < y)}{\sum_{j=1}^{n} W_j^*(x,h)} \tag{2.6}$$

where $W_i^*(x,h)$ satisfy:

$$W_i^*(x,h) = \max[0, [S_2(x) - (x - X_i)S_1(x)]W((x - X_i)/h)]$$

and where $S_l(x) = \sum_{i=1}^{n}(x - X_i)^l W((x - X_i)/h)$ for $l = 1, 2$. Belalia et al. (2017) argued that this Bernstein estimator has no variance and no bias at the end point, namely, at $y = 0, 1$. Moreover, Belalia et al. (2017) proved that the estimator in (2.6) maintains the bias in order h^2 at boundary regions. This is a significant improvement from the traditional estimator which has order h bias at boundary regions.

Further, the Bernstein estimator of the joint c.d.f F in order m, denoted as $\widehat{F}_{m,n}(x,y)$, is defined as stated in Babu and Chaubey (2006):

$$\widehat{F}_{m,n}(x,y) = \sum_{k=0}^{m}\sum_{l=0}^{m} F_n\left(\frac{k}{m}, \frac{l}{m}\right). \tag{2.7}$$

Belalia et al. (2019) mentioned that we can have a smooth estimator for the joint p.d.f. $\widehat{f}_{m,n}(x,y)$ by taking the second derivative of $\widehat{F}_{m,n}(x,y)$:

$$\widehat{f}_{m,n}(x,y) = \frac{d^2}{dxdy}\widehat{F}_{m,n}(x,y) = m^2 \sum_{k=0}^{m-1}\sum_{l=0}^{m-1} B_{k,l,m} P_{k,m-1}(x) P_{l,m-1}(y), \tag{2.8}$$

where

$$B_{k,l,m} = F_n\left(\frac{k+1}{m}, \frac{l+1}{m}\right) - F_n\left(\frac{k+1}{m}, \frac{l}{m}\right) - F_n\left(\frac{k}{m}, \frac{l+1}{m}\right) + F_n\left(\frac{k}{m}, \frac{l}{m}\right),$$

where F_n denotes the bivariate empirical distribution. Belalia et al. (2019) proposed a new estimator for conditional p.d.f.$\widehat{f}_{x,m,n}(y)$, which is as follows:

$$\widehat{f}_{x,m,n}(y) = \frac{\widehat{f}_{m,n}(x,y)}{\widehat{g}_{m,n}(x)}, \tag{2.9}$$

where

$$\widehat{g}_{m,n}(x) = \frac{m}{n}\sum_{k=0}^{m-1} M_{k,m}P_{k,m-1}(x)$$

and $M_{k,m}$ denotes the number of observations in the interval $(\frac{k}{m}, \frac{k+1}{m}]$.

2.2.4 Estimation of Regression Function

Bernstein polynomial can also help to estimate regression function, which naturally follows from the previous results. The regression function $r(x)$ is defined as:

$$r(x) = \mathbb{E}(Y|X = x) = \int_0^1 y f_x(y)dy. \tag{2.10}$$

Belalia et al. (2019) considered to estimate $r(x)$ by:

$$\widehat{r}_{m,n}(x) = \int_0^1 y\widehat{f}_{x,m,n}(y)dy = \sum_{l=0}^{m-1} \frac{l+1}{m+1} W_{x,l,m} \tag{2.11}$$

where

$$W_{x,l,m} = \frac{\sum_{k=0}^{m-1} M_{k,l,m}P_{k,m-1}(x)}{\sum_{k=0}^{m-1} M_{k,m}P_{k,m-1}(x)}.$$

Lots of simulation studies and real-data applications can be found in Belalia et al. (2019).

2.2.5 Estimation of Copula and Testing of Independence

The hypothesis testing problem of two sets of samples $(X_1, Y_1), \ldots, (X_n, Y_n)$ also plays an important role in statistics. Belalia et al. (2017) took advantage of Bernstein polynomial to approach this kind of testing by using empirical Bernstein copula. Firstly, as proposed by Deheuvels (1979), the empirical copula $C_n(u)$ can be defined as:

$$C_n(u) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\mathbb{I}(V_{i;j} \leq u_j), u = (u_1, \ldots, u_d) \in [0,1]^d \tag{2.12}$$

where $V_{i;j} = F_{j,n}(X_{i,j})$ and $F_{j,n}$is the empirical cumulative distribution function of the component $X_{i,j}$. From empirical cumulative distribution, Belalia et al. (2017) define Bernstein copula estimator as:

$$C_{k,n}(u) = \sum_{v_1=0}^{k}\cdots\sum_{v_d=0}^{k} C_n\left(\frac{v_1}{k}, \ldots, \frac{v_d}{k}\right)\prod_{j=1}^{d} P_{v_j,k}(u_j) \tag{2.13}$$

where $P_{v_j,k}$ is the binomial probability function defined in Definition 2.1. Further, Belalia et al. (2017) defined empirical Bernstein copula process as:

$$\mathbb{B}_{k,n}(u) = n^{1/2}(C_{k,n}(u) - C_\pi(u)) \tag{2.14}$$

where $C_\pi(u) = \prod_{j=1}^{d} u_j$, for $u = (u_1, \ldots, u_d) \in [0,1]^d$. Finally, Belalia et al. (2017) proposed a new statistics T_n to test the independence based on this empirical Bernstein process, which is:

$$T_n = \int_{[0,1]^d} \mathbb{B}_{k,n}(u)du. \tag{2.15}$$

In Belalia et al. (2017), the explicit expression of T_n is shown and it is also shown that under the null hypothesis, this statistics converges in distribution to integral of a Gaussian process:

$$T_n \xrightarrow{d} \int_{[0,1]^d} \mathbb{C}_\pi^2(u)du. \tag{2.16}$$

Belalia et al. (2017) also ran a Monte-Carlo simulation to calculate the critical value of this statistic, and a comparison in terms of power with other statistics is also available in Belalia et al. (2017).

2.3 An overview of uniformity tests

In this chapter, we review the methodology of uniformity test. In fact, the problem of testing whether a sample follows any continuous distribution can be transformed to a problem of uniformity test. Then, the widely-used nature will be briefly examined and others result about the statistics of this test will be reviewed.

2.3.1 Uniformity Test and Existing Studies

Basically, a test of uniformity is performed as follows: Let $X_1, \ldots, X_n$ be i.i.d. random variables drawn from a cumulative distribution function F, and consider the problem of testing the null hypothesis $H_0 : F = F_0$ against $H_a : F \neq F_0$. In this paper, we typically consider the case where F_0 is the cdf of uniform distribution on $[0, 1]$. This particular case can be generalized to the problem of testing the simple null hypothesis H_0 that $X_1, \ldots, X_n$ are i.i.d. from any fixed continuous c.d.f. F on the real line. In fact, define $Y_i = F(X_i)$, $i = 1, 2, ..., n$. We can easily show that the $Y_i's$ are i.i.d. as a uniform random variable on $[0, 1]$ under H_0; then, testing the hypothesis that $Y_1, \ldots, Y_n$ are i.i.d. as a uniform random variable on $[0, 1]$ is equivalent to testing the hypothesis that $X_1, \ldots, X_n$ are from the c.d.f F.

As the universal nature of uniformity test, it is useful in lots of fields. Basically, different statistics are chosen for different purposes to test whether the samples are coming from a certain distribution. For example, Dang et al. (2018) used this goodness-of-fit test in the field of corporate finance to evaluate the nuance effects of different measurements of the "firm size effect". Also, Boyle et al. (1997) used this test in medical science to argue that the rare events in epidemiological area fit better in a Poisson model instead of a more popular chi-square model.

Many test statistics have been proposed in literature. For instance, Marhuenda et al. (2005) compared the most commonly used statistics for uniformity test. These statistics can be divided to two major classes. One is the supremum test statistics, which include the Kolmogorov-Smirnov test statistics and related versions. The other is the Cramér-Von Mises type statistics, which take integral form. The above mentioned test statistics are based on the empirical cumulative distribution function (ecdf). Namely, given a sample $(X_1, X_2, ..., X_n)$, F_n is defined by

$$F_n(x) = n^{-1} \sum_{i=1}^{n} \mathbb{I}\{X_i \leq x\}. \tag{2.17}$$

Some well-known and relatively powerful test statistics are as follows: Cramér-Von Mises family type statistics given by

$$Q_n^2 = n \int_0^1 \left(F_n(x) - F_0(x)\right)^2 \phi(x) dF_0(x). \tag{2.18}$$

Kolmogorov-Smirnov test statistics:

$$D_n = \sup_{0 \leq x \leq 1} \left| \widehat{F}_n(x) - F_0(x) \right|. \tag{2.19}$$

D_n can be expressed as the maximum of two values, D_n^+ and D_n^-, which are as follows:

$$D_n^+ = \max_{1 \leq i \leq n} \left(\frac{i}{n} - U_{(i)} \right), \tag{2.20}$$

and

$$D_n^- = \max_{1 \leq i \leq n} \left(U_{(i)} - \frac{i-1}{n} \right). \tag{2.21}$$

where $U_{(i)}$ stands for the i^{th} largest observations i.e. $U_{(1)} < U_{(2)} < ... < U_{(n)}$. Kuiper (1960) suggested a closely related statistics V_n to be used when sample are distributed among a circle, as this statistics does not depend on the origin choice. V_n is defined as:

$$V_n = D_n^+ + D_n^-. \tag{2.22}$$

The closed form and the asymptotic distribution of those statistics are well-studied. See Kolmonorgov (1933) for a detailed discussion of D_n. Also, the Cramér-Von Mises type statistics have some variations. When $\phi(x) = 1$, the Q_n is the Cramér-Von Mises statistics W_n and when $\phi(x) = \frac{F(x)}{1-F(x)}$ it is called Anderson-Darling's A_n. Anderson and Darling (1954) showed that the closed form of these statistics can be expressed as:

$$W_n^2 = \sum_{i=1}^{n} \left(U_i - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n}. \tag{2.23}$$

and

$$A_n^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln U_{(i)} + \ln(1-U_{(n+1-i)})]. \quad (2.24)$$

Noticeably, as no simulation software was available that time, the critical value was gained theoretically in Anderson and Darling (1954).

Also, some small modifications to those well-studied statistics exist. For example, Watson's U_n^2 is a modification of W_n^2 and is defined as:

$$U_n^2 = W_n^2 - n(\bar{U} - 0.5)^2 \quad (2.25)$$

where $\bar{U}$ stands for the average of all observed sample.

Besides that, there exist other statistics, such as statistics based on spacing. We can defined the spacing K_i between order statistics as:

$$K_i = U_{(i)} - U_{(i-1)} \qquad \forall i = 2, \ldots, n \qquad \mathit{and} \qquad K_1 = U_{(1)}, K_{n+1} = 1 - U_{(n)}. \quad (2.26)$$

Then, we have statistics based on D_i. The simplest one is from Greenwood (1946) and it is defined as:

$$G_n = \sum_{i=1}^{n+1} K_i^2.$$

It can also be as complicated as the one introduced by Read and Cressie (2012), which is defined as:

$$2nI^{\lambda}(D, E(D)) = \frac{2n}{\lambda(\lambda+1)}\sum_{i=1}^{n+1} D_i[((n+1)D_i)^{\lambda} - 1].$$

Further, some statistics are introduced for special use, for example, Quesenberry and Jr. (1977) propose the statistic:

$$Q = \sum_{i=1}^{n+1} D_i^2 + \sum_{i=1}^{n} D_i D_{i+1}. \tag{2.27}$$

As the notation suggests, this statistic will be sensitive when dealing with autocorrelation between samples. In Quesenberry and Jr. (1977), the critical value of the statistics Q was calculated by Monte Carlo simulation. More details of those statistics are discussed in the above quoted papers.

An improvement of statistics based on spacing is that based on higher order spacings. We follow the approach in Deken (1981) and define higher order spacings as:

$$G_i^{(m)} = U_{(i+m)} - U_{(i)} \qquad \forall i = 0, \ldots, n+1-m \qquad \textit{and} \qquad U_0 = 0, U_{n+1} = 1. \tag{2.28}$$

In Cressie (1978), the author found a statistic, based on this higher order spacing, which is defined as:

$$L_n^{(m)} = \sum_{i=0}^{n+1-m} \ln G_i^{(m)}. \tag{2.29}$$

Cressie (1978) showed that this statistic is asymptotically normal, and it is better for some particular alternatives. After that, Cressie (1979) studied another statistics which have a better asymptotic relative efficiency. This statistic is defined as:

$$S_n^{(m)} = \sum_{i=0}^{n+1-m} n {G_i^{(m)}}^2. \tag{2.30}$$

To present some others statistics, we recall first the definition of Legendre polynomial.

For more details about this polynomial, we refer to Weisstein (2015)

Definition 2.2.

$$\frac{1}{(1-2xt+t^2)^{1/2}} = \sum_{n=0}^{\infty} P_n(x)t^n$$

The solution $P_n(x)$ satisfy the above equation when $t=1$ is the normalized Legendre polynomial.

It have the following explicit expressions:

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n. \tag{2.31}$$

Or it can be expressed as a more convenient form for calculation:

$$P_n(x) = \sum_{k=0}^{n} \binom{n}{k}\binom{n+k}{k}\left(\frac{x-1}{2}\right)^k. \tag{2.32}$$

Further, Neyman (1937) used approach of Legendre polynomials to introduce a powerful statistics for uniformity test, which is defined as:

$$N_h = \frac{1}{n}\sum_{j=1}^{h}(\sum_{i=1}^{n} P_n(U_i))^2, \tag{2.33}$$

where $P_n(x)$ stands for the normalized Legendre polynomials.

As Neyman (1937) mentioned, the problem of this statistics consists in the choice of h. Although sometimes one can find a value of h which makes N_h performs very well, it does not work for any alternative distributions. For certain alternatives, to find an optimal h is very time-consuming because of the complexity of the expression of N_h. As time goes, in the beginning of the $21st$ century, Zhang (2002) developed a novel approach based on the statistics

$$Z_a = -\sum_{i=1}^{n} \left(\frac{\ln(U_{(i)})}{n-i+0.5} + \frac{\ln\left(1-U_{(i)}\right)}{i-0.5} \right), \tag{2.34}$$

and

$$Z_c = \sum_{i=1}^{n} \left(\ln \left(\frac{U_{(i)}^{-1} - 1}{(n-0.5)/(i-0.75) - 1} \right) \right)^2. \tag{2.35}$$

It is shown that these statistics behave better in term of power than most of classical ones.

The reason for developing so many statistics is that different kinds of statistics have higher sensitivity on different alternative distributions F_a. For example, as mentioned in Marhuenda et al. (2005), Z_a is more powerful when testing against alternative distribution which has more density on both end (0 and 1) than uniform distribution. Conversely, the Kolmogorov-type statistics D_n, which is mentioned in (2.19), performs better when dealing with alternative distribution which has more density around 0.5. For goodness-of-fit test problem, there is infinite variation of alternative distributions. Because of that, the uniformly most powerful test (UMP test) seems impossible. However, the pursuit of more generally suitable or more specific fitted test statistics will never be ceased.

Chapter 3

Uniformity tests based on Bernstein polynomials

In this chapter, we propose two test statistics based on the empirical Bernstein process. The closed form and asymptotic distribution are theoretically proved. Further, we present a simulation study showing the advantages of the proposed test statistics.

3.1 Introduction

Let $X_1, \ldots, X_n$ be i.i.d. real valued observations drawn from a cumulative distribution function F, and consider the problem of testing the null hypothesis $\mathcal{H}_0 : F = F_0$ against $\mathcal{H}_a : F \neq F_0$. In this paper, we typically consider the case where F_0 is the cdf of uniform distribution on $[0, 1]$. This particular case can be generalized to the problem of testing the simple null hypothesis $\mathcal{H}_0$ that $X_1, \ldots, X_n$ are i.i.d. from any fixed continuous c.d.f. F on the real line. In fact, define $Y_i = F(X_i)$, so that the Y_i are i.i.d. $U(0, 1)$ under $\mathcal{H}_0$; then, test the hypothesis that $Y_1, \ldots, Y_n$ are i.i.d. uniform on $[0, 1]$. This test ignited interest for a long time. Dating back to 1950s, for example, Anderson and Darling (1954) proposed a statistic A_n for goodness-of-fit test, which was given in Equation 2.24, and the significance point are theoretically calculated.

Many test statistics have been proposed in literature, and efficiencies of them are often compared in terms of power. For instance, Quesenberry and Jr. (1977) compared most of available statistics for testing uniformity. He picked lots of widely used one, like Kolmogorov's D_n, Cramer-Von Mises W_n and Anderson-darling A_n. We reviewed them in previous chapter. Also, see Liang et al. (2001) for a comparison of statistics for testing multivariate uniformity. Later, Marhuenda et al. (2005) did a more complete comparison between most commonly used statistics for the uniformity test. These statistics can be divided into two classes, the supremum test statistics, which include the Kolmogorov-Smirnov test statistics and related versions, and the Cramér-Von Mises type statistics. The above mentioned test statistics are based on the empirical cumulative distribution function (ecdf), namely, given a sample $(X_1, X_2, ..., X_n)$, F_n is defined by

$$F_n(x) = n^{-1} \sum_{i=1}^{n} \mathbb{I}\{X_i \leq x\}. \tag{3.1}$$

There are two statistics that we are going to improve, namely, C_n and S_n, and they are defined as follows respectively:

- Komogorov-Smirnov test type

$$S_n = \sup_{0 \leq x \leq 1} \left| \sqrt{n} \left(\widehat{F}_n(x) - F(x) \right) \right|. \tag{3.2}$$

- Cramér-Von Mises test type

$$C_n = n \int_0^1 (F_n(x) - F(x))^2 \, dF(x). \tag{3.3}$$

We point out that the ecdf is a step function. However, as proposed in Babu et al. (2002), it can be smoothed by using Bernstein polynomials. Specifically, the empirical Bernstein distribution estimator is given by

$$\widehat{F}_{n,m}(x) = \sum_{k=0}^{m} F_n\left(\frac{k}{m}\right) P_{m,k}(x), \qquad k = 0, ..., m, \tag{3.4}$$

where $P_{m,k}(x) = \binom{m}{k} x^k (1-x)^{m-k}, \quad k = 0, ..., m$, stands for the binomial probability mass function and m is a bandwidth parameter which increases to infinity as n tends to infinity. A multivariate extension of this estimator was proposed in Babu and Chaubey (2006) and studied in Belalia (2016). The estimator (3.4) was deeply studied by Leblanc (2009) and Leblanc (2012). It was shown in the quoted papers that the empirical Bernstein estimator $\widehat{F}_{m,n}$ outperforms the ecdf one in the sense of the integrated mean square error. This finding motivated us to propose an alternative goodness-of-fit test based on the empirical Bernstein distribution estimator instead of the ecdf one. Specifically, the proposed test statistics are as follow:

- Bernstein Komogorov-Smirnov test type

$$S_{m,n} = \sup_{0 \le x \le 1} \left| \sqrt{n} \left(\widehat{F}_{m,n}(x) - F(x) \right) \right|. \tag{3.5}$$

- Bernstein Cramér-Von Mises test type

$$C_{m,n} = n \int_0^1 \left(F_{n,m}(x) - F(x)\right)^2 dF(x). \tag{3.6}$$

The remainder of this chapter is organized as follow. In Section 3.2, an explicit expression of $C_{m,n}$ is provided and the asymptotic distributions of $C_{m,n}$ and $S_{m,n}$ are established under the null hypothesis $\mathcal{H}_0$ and also alternative hypotheses. In Section 3.3, a simulation study is carried out to show the powers and efficiencies of the proposed test statistics compared to the ones based on the ecdf.

3.2 Asymptotic distribution of $S_{m,n}$ and $C_{m,n}$

We start this section by providing a proposition which gives an explicit expression of $C_{m,n}$ under the null hypothesis $\mathcal{H}_0$.

Proposition 3.1. *Under $\mathcal{H}_0$, we have*

$$\begin{aligned} C_{m,n} = n &\sum_{k,\ell=0}^{m} F_n\left(\frac{k}{m}\right) F_n\left(\frac{\ell}{m}\right) \binom{m}{k}\binom{m}{\ell} \beta(k+\ell+1, 2m-k-\ell+1) \\ &- 2n \sum_{k=0}^{m} F_n\left(\frac{k}{m}\right) \binom{m}{k} \beta(k+2, m-k+1) + \frac{n}{3}, \end{aligned}$$

where $\beta(a,b)$ stands for the beta function for two positive integers a and b, which is defined as:

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx.$$

Proof. Note that under the null hypothesis $\mathcal{H}_0$, $F_0(x) = x$, then, $C_{m,n}$ can be decomposed as follow:

$$\begin{aligned} C_{m,n} &= n\int_0^1 \left(\widehat{F}_{n,m}(x) - x\right)^2 dx \\ &= n\int_0^1 \left(\widehat{F}_{n,m}(x)\right)^2 dx - 2n\int_0^1 x\left(\widehat{F}_{n,m}(x)\right) dx + n\int_0^1 x^2 dx. \end{aligned}$$

Then,

$$C_{m,n} = n(A - 2B + C),$$

where

$$A = \int_0^1 \left(\widehat{F}_{n,m}(x)\right)^2 dx, \qquad B = \int_0^1 x\left(\widehat{F}_{n,m}(x)\right) dx, \qquad \textit{and} \qquad C = \int_0^1 x^2 dx$$

Furthermore, we have

$$\begin{aligned} A &= \int_0^1 \sum_{k,\ell=0}^{m} F_n\left(\frac{k}{m}\right) F_n\left(\frac{\ell}{m}\right) P_{m,k}(x) P_{m,\ell}(x) \\ &= \sum_{k,\ell=0}^{m} \binom{m}{k}\binom{m}{\ell} F_n\left(\frac{k}{m}\right) F_n\left(\frac{\ell}{m}\right) \int_0^1 x^{k+\ell}(1-x)^{2m-k-\ell} dx. \end{aligned}$$

This gives

$$A = \sum_{k,\ell=0}^{m} F_n\left(\frac{k}{m}\right) F_n\left(\frac{\ell}{m}\right) \binom{m}{k}\binom{m}{\ell} \beta(k+\ell+1, 2m-k-\ell+1).$$

Similarly, one can show that

$$B = \sum_{k=0}^{m} F_n\left(\frac{k}{m}\right)\binom{m}{k}\beta(k+2, m-k+1).$$

Finally, it is trivial that $C = \frac{1}{3}$. Which concludes the proof. □

3.2.1 Asymptotic properties of $C_{m,n}$ and $S_{m,n}$

This subsection is devoted to the derivation of the distribution limit under the null hypothesis $\mathcal{H}_0$ and all legitimate alternative distributions $\mathcal{H}_A$ of the test statistics $C_{m,n}$ and $S_{m,n}$. The following lemma from Chapter 19 of Van der Vaart (2000) will be needed in establishing these limiting distributions. Let $\mathbb{F}_n$ denote the empirical process, which is expressed as:

$$\mathbb{F}_n(x) = \sqrt{n}(F_n(x) - F(x)), \qquad 0 \leq x \leq 1. \tag{3.7}$$

Let $D[0,1]$ denote the space of all functions which are right-continuous with left limits on $[0,1]$. We use the notation $\mathbb{B}_n \rightsquigarrow \mathbb{B}$ to stand for the process $\mathbb{B}_n(.)$ converges weakly to the process $\mathbb{B}(.)$, with respect to Skorokhod topology on $D[0,1]$.

Lemma 3.1. *(convergence of $\mathbb{F}_n$)*

Let $\mathbb{F}_n(x)$ be as defined in equation 3.7, let $X_1, \ldots, X_n$ be i.i.d. random samples and let $F(x)$ be the c.d.f. for $X_i's$ and $F(x)$ is a continuous function in $[0,1]$.

Then, we have:

$$\mathbb{F}_n \rightsquigarrow \mathbb{F} \qquad as \qquad n \longrightarrow \infty,$$

where $\mathbb{F}$ is a Gaussian process which satisfies:

i. $\mathbb{E}\left(\mathbb{F}(t)\right) = 0 \qquad \forall 0 \leq t \leq 1$;

ii. $\mathbb{C}\mathrm{ov}\left(\mathbb{F}(s), \mathbb{F}(t)\right) = F(s)(1 - F(t)) \qquad \forall 0 \leq s \leq t \leq 1$.

The proof of this lemma can be found in chapter 19 of Wellner et al. (2013). Further, for this paper to be self-contained, a similar proof is outlined in the Appendix B.

For deriving the asymptotic properties of $C_{m,n}$, we firstly define the Bernstein operator

$$B_m : (D[0,1], ||\cdot||_\infty) \longrightarrow (C[0,1], ||\cdot||_\infty)$$

as:

$$B_m(F)(u) = \sum_{k=0}^{m} F(k/m) P_{m,k}(u),$$

where $C[0,1]$ stands for the space of all continuous functions over $[0,1]$ and $D[0,1]$ stands for the space of all functions over $[0,1]$ with left limit and right continuous. Further, the norm $||\cdot||_\infty$ is defined as $||f||_\infty = \sup|f(x)|$ where the supremum is taken over domain of f (therefore $[0,1]$ in our space).

Then, the empirical Bernstein process is defined as

$$\mathbb{B}_n(x) = \sqrt{n}\left(\widehat{F}_{m,n}(x) - F(x)\right).$$

Further, the following result will be of interest, also known as the generalized continuous mapping theorem, which is given in (Whitt, 2002, Theorem 3.4.4).

Theorem 3.1 (Generalized Continuous Mapping Theorem)**.** *Let g and g_n, $n \geq 1$, be measurable functions mapping (S,d) into (S',d'). Let the range (S',d') be separable. Let E be the set of x in S such that $g_n(x_n) \longrightarrow g(x)$ fails for some sequence*

$\{x_n, n \geq 1, n \in \mathbb{N}\}$ with $x_n \longrightarrow x$ in S. If $X_n \xrightarrow{d} X$, and $P[X \in E] = 0$, then $g_n(X_n) \xrightarrow{d} g(X)$, in (S', d').

The proof of this theorem is given in Whitt (2002). For the convenience of the reader, a similar proof is outlined in the Appendix. By using Theorem 3.1, the core result of this paper is established in the following theorem.

Theorem 3.2. *Assume that the derivative of F exists and is continuous on $[0, 1]$. Further, if m satisfies $n^{1/2}m^{-1/2} \longrightarrow 0$ and $n \longrightarrow \infty$, then*

$$\mathbb{B}_n(x) = n^{1/2} \left(B_m \left(F_n\right)(x) - F(x)\right) \rightsquigarrow \mathbb{F}(x)$$

where $\mathbb{F}$ is given in Lemma 3.1.

Proof of Theorem 3.2. First, the empirical Bernstein process can be rewritten as:

$$\begin{aligned} n^{1/2}\left[B_m(F_n)(x) - F(x)\right] &= n^{1/2}\left[B_m(F_n)(x) - B_m(F)(x) + B_m(F)(x) - F(x)\right] \\ &= n^{1/2} B_m(F_n - F)(x) + n^{1/2}\left(B_m(F)(x) - F(x)\right) \\ &= T_{mn,1} + T_{mn,2} \end{aligned}$$

Within the second term, $(B_m(F)(x) - F(x))$ converges uniformly to 0 by original paper of Bernstein (1912).

This result is revised a little bit, and together with the assumption that $n^{1/2}m^{-1/2} \longrightarrow 0$, it will be proven that the term $T_{mn,2}$ converges to 0 uniformly under our condition on F. To this end, the second term can be expressed as

$$\begin{aligned} T_{mn,2} = \sqrt{n}\left|B_m(F)(x) - F(x)\right| &= \sqrt{n}\left|\sum_{k=0}^{m} F\left(\frac{k}{m}\right) p_{m,k}(x) - F(x)\right| \\ &\leq \sqrt{n}\sum_{k=0}^{m} p_{m,k}(x)\left|F\left(\frac{k}{m}\right) - F(x)\right|. \end{aligned}$$

Then, a Taylor expansion for F at point x leads to

$$\begin{aligned} T_{mn,2} &\leq \sqrt{n}\sum_{k=0}^{m} p_{m,k}(x)\left|f(x)\left(\frac{k}{m} - x\right) + O\left(\frac{k}{m} - x\right)^2\right| \\ &\leq \sqrt{n}\sum_{k=0}^{m} p_{m,k}(x)f(x)\left|\frac{k}{m} - x\right| + \sqrt{n}\sum_{k=0}^{m}\left|O\left(\frac{k}{m} - x\right)^2\right| p_{m,k}(x) \\ &\leq \sqrt{n}\sum_{k=0}^{m} p_{m,k}(x)f(x)\left|\frac{k}{m} - x\right| + O\left(\left|\sqrt{n}m^{-2}\sum_{k=0}^{m}(k - mx)^2 p_{m,k}(x)\right|\right) \\ &\leq \sqrt{n}\sum_{k=0}^{m} p_{m,k}(x)f(x)\left|\frac{k}{m} - x\right| + O\left(\left|\sqrt{n}m^{-2}mx(1 - x)\right|\right) \\ &= \sqrt{n}\sum_{k=0}^{m} p_{m,k}(x)f(x)\left|\frac{k}{m} - x\right| + O(n^{1/2}m^{-1}). \end{aligned}$$

Noticed that the previous inequality comes from the triangular inequality and also from the fact that $f(x)$ is non-negative. Also the summation inside big-O notation is the variance of a random variable Y which follows binomial distribution of parameters m and x. Then we use the fact that $x(1 - x)$ is bounded by 1.

Then we apply Cauchy-Schwartz inequality to first term and use again the variance of a binomial distribution to get

$$\begin{aligned}
\sqrt{n}\sum_{k=0}^{m} p_{m,k}(x)f(x)\left|\frac{k}{m}-x\right| &= n^{1/2}m^{-1}f(x)\sum_{k=0}^{m}|k-mx|\,p_{m,k}(x) \\
&\leq n^{1/2}m^{-1}f(x)\left(\sum_{k=0}^{m}(k-mx)^2 p_{m,k}(x)\right)^{1/2} \\
&= \ n^{1/2}m^{-1}f(x)\left(mx(1-x)\right)^{1/2} \\
&\leq f(x)n^{1/2}m^{-1/2} = O\left(n^{1/2}m^{-1/2}\right).
\end{aligned}$$

We have the big O notation as $f(x)$ is a continuous function on a compact support $[0,1]$, so it achieve it's maximum in the support. Therefore, we finally conclude that $\sup_{x\in[0,1]}\sqrt{n}\,|B_m(F)(x)-F(x)| = O\left(n^{1/2}m^{-1/2}\right)$. Hence $T_{mn,2}$ converges to 0 uniformly.

It remain to use Theorem 3.1 to prove that the first term $T_{mn,1}$ converges to $\mathbb{F}$. to achieve this goal a similar method followed by Neumann et al. (2019) will be employed. Let (S,d) be $(D[0,1],||\cdot||_\infty)$ and (S',d') be $(C[0,1],||\cdot||_\infty)$. Also, let g_n be defined as $g_n = B_m$ and g be the identity function on S' and take arbitrary value on $S\backslash S'$.

Then we check 3 conditions:

i. S' has a countable dense subset, namely, the set of all rational coefficient polynomials. Therefore, S' is separable.

ii. Let E be the set of $f \in S$ such that $g_n(f_n) \longrightarrow g(f)$ fails for some sequence f_n with $f_n \longrightarrow f$ in S. Then if $f \in E$ and $f \in S$, we will have $f \notin S'$. Since if $f \in S'$, we can choose $f_n = f$ for all n and get uniform convergence of $g_n(f_n)$ to $g(f)$ by Bernstein's Theorems. Therefore we claim that $E \subset S\backslash S'$. Therefore we will have $P(F \in E) \leq P(F \in (S\backslash S')) = 0$. As $P(F \in E) \geq 0$, we have

$P(F \in E) = 0$.

iii. Also, the convergence of $\mathbb{F}_n$ to $\mathbb{F}$ is showed in Van der Vaart (2000).

Then by Theorem 3.1, with all conditions meet, we have

$$B_m(\mathbb{F}_n) \rightsquigarrow \mathbb{F} \qquad as \qquad n \longrightarrow \infty.$$

Theorem Therefore, 3.2 is proved.

□

For further discussion about the process $\mathbb{F}_n$, see Chapter 2 of Van der Vaart (2000). Based on Theorem 3.2 and the generalized continuous mapping theorem, we establish the asymptotic distribution of $C_{m,n}$ and $S_{m,n}$, as stated in the following corollary. Noticing that this corollary holds for any F that satisfies the conditions of Theorem 3.2. The established result serves as a linkage to get the limiting distribution of $C_{m,n}$ and $S_{m,n}$ under $\mathcal{H}_0$ and $\mathcal{H}_A$.

Remark 3.1. *It is pointed out that if the condition is strengthened to let F admits two continuous and bounded derivatives, Leblanc (2012) proved that the term $(B_m(F)(x) - F(x))$ is dominated by m^{-1}. The formula is as follows:*

$$B_m(F)(x) = F(x) + m^{-1}b(x) + o(m^{-1}),$$

where $b(x) = x(1-x)f'(x)/2$.

Corollary 3.1. *Under the same conditions as Theorem 3.2, we have*

$$C_{m,n} \xrightarrow[n \longrightarrow \infty]{d} \int_0^1 \mathbb{F}^2(x)dF(x)$$

and

$$S_{m,n} \xrightarrow[n\longrightarrow\infty]{d} \sup_{0\le x\le 1} |\mathbb{F}(x)|,$$

where $\mathbb{F}$ *is a Gaussian process with properties given in Lemma 3.1.*

Proof. Corollary 3.1 follows directly from Theorem 3.2 and the continuous mapping theorem. Indeed, note that $h(f) = \int_0^1 f^2(x)dF(x)$ and $k(f) = \sup_{0\le x\le 1} |f(x)|$ are continuous functionals. □

3.3 Simulation study

To assess and compare the efficiencies of those statistics, we firstly need to calculate the critical value for each statistic. Let F_0 be the c.d.f. of uniform distribution on $[0, 1]$, then for any given statistics T, the critical value c_T is defined as follows:

$$\mathbb{P}\left[T \ge c_T \mid F = F_0\right] = \alpha$$

where $\alpha = 0.05$ is the significance level of the test, F is the c.d.f. from which the sample was drawn, and T is a function of sample data only.

To calculate the statistics $C_{m,n}$, we take advantage of the strong law of large numbers, which is given below.

Theorem 3.3. *(Strong Law of Large Numbers) Let* $X_1, \ldots, X_n$ *be i.i.d. as a r.v.* X *with* $\mathbb{E}(X) = \mu$ *and* $\mathbb{E}(|X^4| < \infty)$*. Let* $\bar{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$*, then*

$$\bar{X_n} \xrightarrow{a.s.} \mu.$$

Noticed that the general version of SLLN only requires that the means of $X_i's$ exist. However, as we shown below, we only apply theorem 3.3 to random variables which are continuous on a compact base (which is $[0,1]$) in $\mathbb{R}$, therefore the fourth moments always exist. Moreover, a proof of this theorem can be found in Etemadi (1981). For this paper to be self-contained, a similar proof is outlined in the Appendix B.

We denote $(F_{n,m}(x) - F_0(x))^2$ as $g(x)$. Therefore, we generate a sequence of 10000 copies of $\mathsf{Uniform}(0,1)$ sample and calculate the value of $g(x)$ for each of them and then take the average. Finally, we multiply this value by n and we get the value of $C_{m,n}$. To see why this algorithm works, notice that when $x_i's$ are distributed as $\mathsf{Uniform}(0,1)$, we have $C_{m,n} = n\int_0^1 g(x)dx = n\mathbb{E}[g(x)]$. The generated copies of $g(x)'s$ give a good approximation of $\mathbb{E}[g(x)]$ in respect of the strong law of large numbers.

Moreover, although the asymptotic distribution of those statistics can be obtained, it is hard to use the limiting distribution to calculate the critical value directly. Therefore, we use **R** software to obtain them by simulations. For a given statistics T_0, we first generate n copies of random variables from uniform $[0,1]$, and then the corresponding test statistics T_0 is calculated. In this paper, we use $n = 100$. Finally, we repeat this process 10000 times, and get the 95% percentile, therefore we obtain the critical value c_{T_0} for T_0.

Further, the power of those test statistics when testing against different alternative distributions is calculated. For a given alternative distribution F_a, the power of a test statistic T_0 is denoted as $\beta(T_0, F_a)$, and is defined as:

$$\beta(T_0, F_a) = \mathbb{P}\left[T_0 > c_{T_0} \mid F = F_a\right].$$

To calculate this, we generate $n = 100$ copies of sample according to F_a, then calculate T_0. Then, we repeat this process 10000 times and record the proportion of T_0's which are greater than the critical value.

For the alternative distributions, similar distributions as in Marhuenda et al. (2005) and Stephens (1974) are used. Namely, the following families of distributions:

$$A_k : F_{1,k}(x) = 1 - (1-x)^k, \qquad 0 \leq x \leq 1$$

$$B_k : F_{2,k}(x) = \begin{cases} 2^{k-1}x^k, & 0 \leq x \leq 0.5 \\ 1 - 2^{k-1}(1-x)^k, & 0.5 \leq x \leq 1 \end{cases}$$

where is a positive real number. It should be noticed that when $k = 1$, both A_k and B_k reduced to the uniform distribution. When $k < 1$, as k decreasing to zero, A_k gives more densities near 1, and B_k gives more densities near 0.5. Conversely, when $k > 1$, as k increases, A_k gives more densities near 0 and B_k gives more densities near 0 and 1.

We compute the power and inefficiencies for $k = 0.20, 0.40, 0.60, 0.80, 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.75, 3.00$, in order to compare with the findings in Marhuenda et al. (2005). We use the same definition of inefficiency as in Marhuenda et al. (2005), which is as follows:

First, we calculate the maximum power of test statistics over each fixed family with fixed k_0:

$$\beta_{\max}(f, k_0) = \max_T (\beta(T, F_a))$$

where f denotes families of A_k or B_k, and F_a is that family with fixed k_0.

Then, the inefficiency of statistic T_0 of the family f with parameter k_0 is the difference between power of T_0 and the maximum power of this same alternative distribution:

$$i_{T_0}(f, k_0) = \beta_{\max}(f, k_0) - \beta(T_0, F_a).$$

Finally, for the family f, the maximum inefficiency of the statistic T_0 is the maximum inefficiency over all possible k:

$$i_{\max T_0}(f) = \max_{k_0}(i_{T_0}(f, k_0)).$$

As we proposed, $S_{m,n}$ and $C_{m,n}$ are served as the improvements for S_n and C_n respectively. As a result, we made an additional comparison between each pair of them. Firstly, we use large scale m value, namely, we take $m = 15, 30, 50$ respectively, and the critical value and power are summarized in Table 3.1. The results in Table 3.2 show that $m = 30$ works the best. Then, we take the values of m around 30 and compare them. Here we compare the value of m from 25 to 35.

Table 3.1: Table of Critical Value

Statistics	Critical Value
C_n	0.462437
$C_{m,n}, m = 15$	0.3623181
$C_{m,n}, m = 30$	0.3969091
$C_{m,n}, m = 50$	0.4029511
S_n	0.1295154
$S_{m,n}, m = 25$	0.09705955
$S_{m,n}, m = 30$	0.0990622
$S_{m,n}, m = 35$	0.09984095

Table 3.2: Powers of C_n and $C_{m,n}$ for family A_k

	k									
Statistic	0.2	0.4	0.6	0.8	1	1.25	1.5	1.75	2	2.25
C_n	1	1	0.988	0.4684	0.05	0.4726	0.951	0.998	1	1
$C_{m,n}, m = 15$	1	1	0.9905	0.431	0.05	0.4345	0.9355	0.998	1	1
$C_{m,n}, m = 30$	1	1	0.994	0.4864	0.05	0.4815	0.9545	1	1	1
$C_{m,n}, m = 50$	1	1	0.9905	0.4835	0.05	0.471	0.9485	0.999	1	1

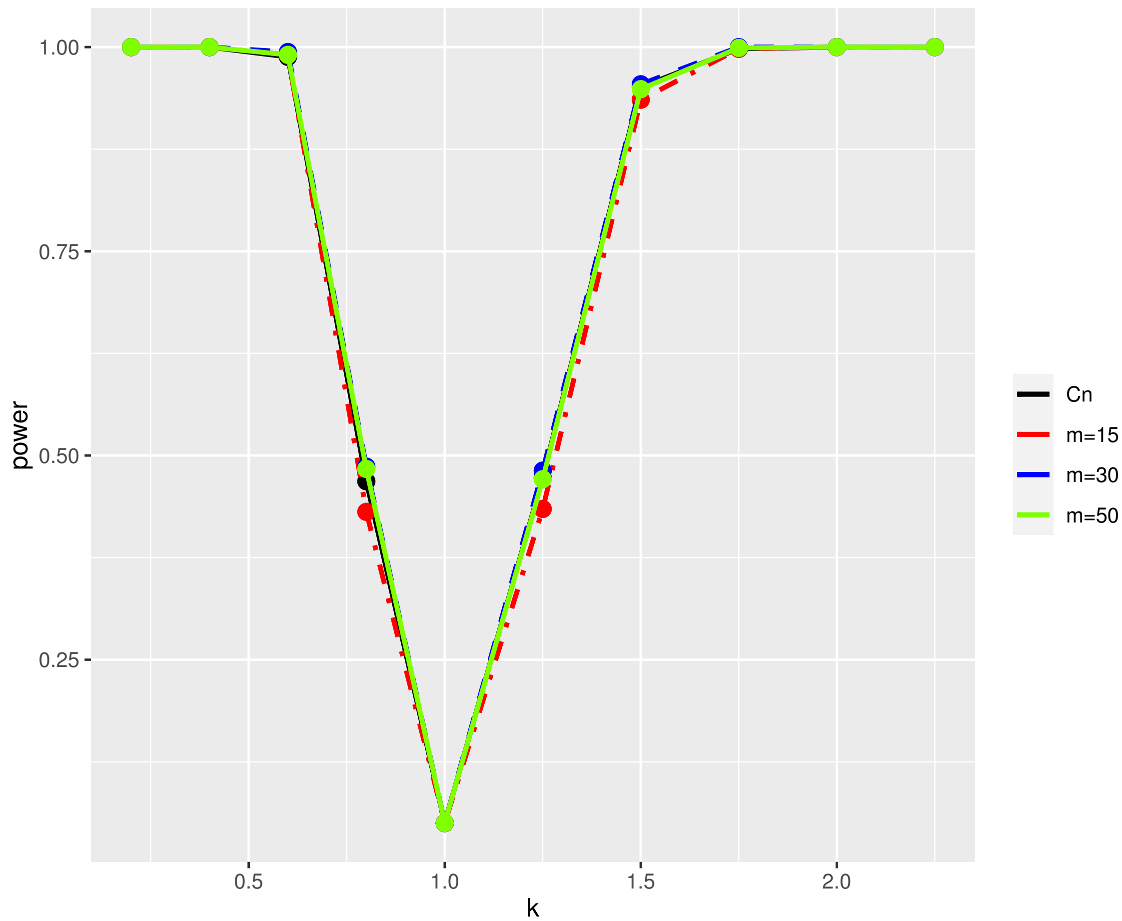

Figure 3.1: Powers of C_n and $C_{m,n}$ for family A_k

Table 3.2 and Figure 3.1 show the power of our statistics $C_{m,n}$ with different value of m compared to the power of C_n. All the data here use 10000 times repetitions. From the table, we can see that for sample size $n = 100$, small numbers of m do not give good results. Both $m = 30$ and $m = 50$ improve somewhat compared to the original C_n, and $m = 30$ works better, as its power is always greater than that of C_n for all value of k. Therefore, we carry out the simulations for small-scale variation around $m = 30$ and see which value works best. Table 3.3 shows the critical value for

$m = 25, 35, 40$. Further, in Table 3.4, we also report the power of the tests for these values of m, and a graph for comparison is also provided.

Table 3.3: Critical Value of $C_{m,n}$

	Critical Value
$C_{m,n}, m = 25$	0.4142262
$C_{m,n}, m = 35$	0.3894583
$C_{m,n}, m = 40$	0.4063481

Table 3.4: Powers of C_n and $C_{m,n}$ for family A_k

	k									
Statistic	0.2	0.4	0.6	0.8	1	1.25	1.5	1.75	2	2.25
C_n	1	1	0.988	0.4684	0.05	0.4726	0.951	0.998	1	1
$C_{m,n}, m = 25$	1	1	0.9856	0.4514	0.05	0.4544	0.9402	0.999	1	1
$C_{m,n}, m = 30$	1	1	0.994	0.4864	0.05	0.4815	0.9545	1	1	1
$C_{m,n}, m = 35$	1	1	0.9908	0.5024	0.05	0.4936	0.9512	0.999	1	1
$C_{m,n}, m = 40$	1	1	0.9882	0.4832	0.05	0.483	0.9528	0.999	1	1

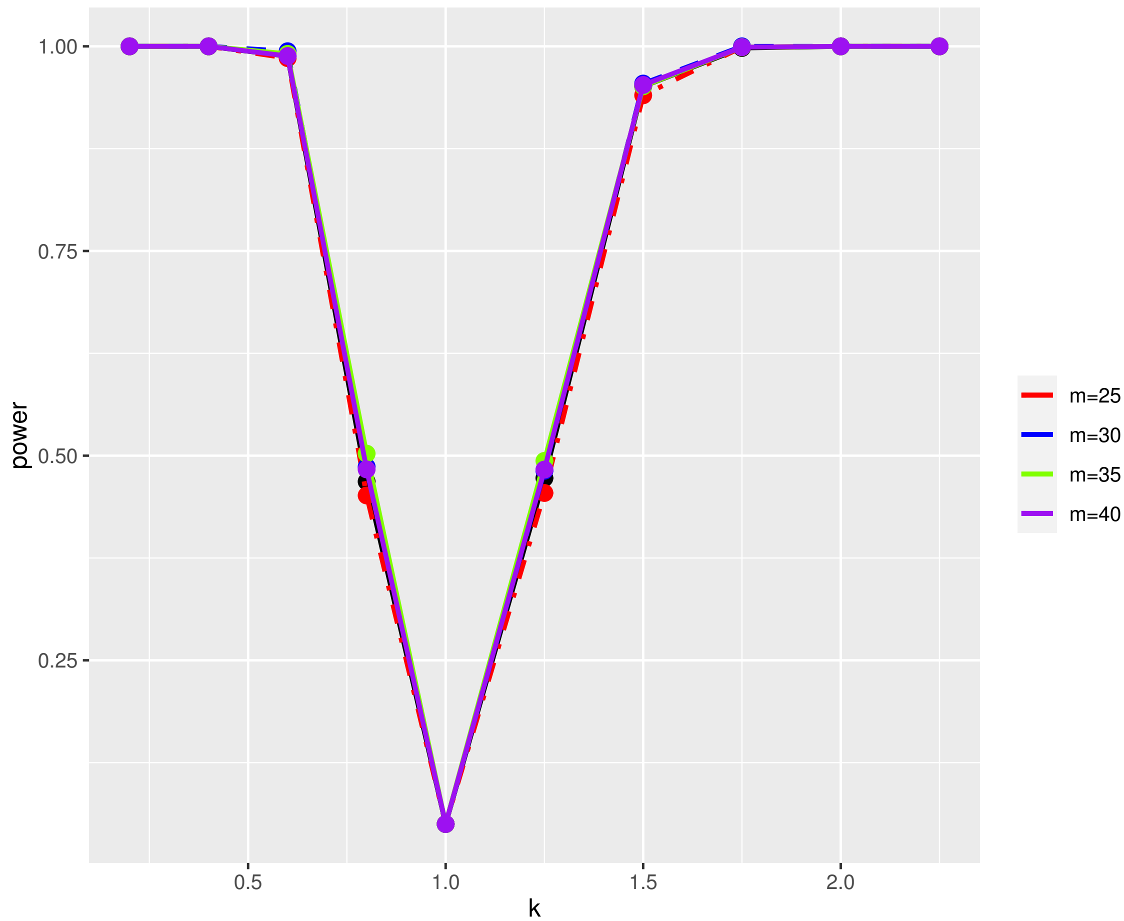

Figure 3.2: Powers of $C_{m,n}$ for family A_k

We keep the row of C_n and $C_{m,n}$ for $m = 30$ for comparison. We can see from Table 3.4 that when the alternative distribution is behaved far from uniform, both statistics perform well. As a result, the competition happens when k takes the value of $0.6, 0.8, 1.25$ and 1.5. When $m = 30$, $C_{m,n}$ has the maximum power dealing with $k = 0.6, 1.5$ and 1.75. However, when k takes the value of 0.8 and 1.25, which means the alternative distribution is close to uniform, taking $m = 35$ makes $C_{m,n}$ performs better. It outperforms $m = 30$ by approximately 0.015 and even more for C_n.

Noticed that for $k = 0.8$ and 1.25, $C_{m,n}$ with $m = 35$ has power around 0.5. It is not really satisfactory. To see the reason, we refer to a graph which shows the pdf of family $A_k, k = 0.4, 0.8$ and 1.25 together with the pdf of uniform distribution.

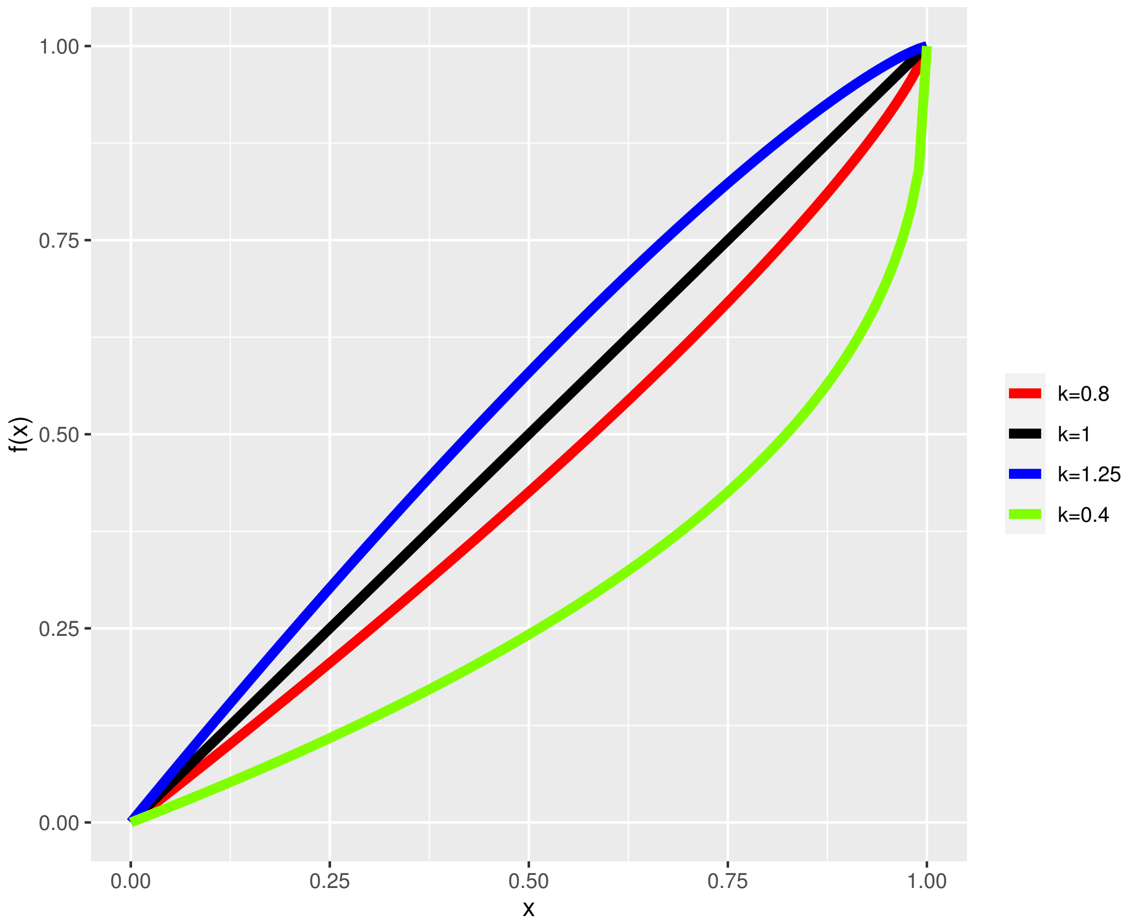

Figure 3.3: Density functions for family A_k when $k = 0.4, 0.8$ and 1.25

Comparing to $k = 0.4$, the distributions of $F_{A,k}$ are close to uniform when they take values of 0.8 and 1.25. As a result, it is much harder for a statistic to distinguish the difference between these alternative distributions from uniform.

Further, we also report in Table 3.5 and Table 3.6 the numerical results which show the power of C_n and $C_{m,n}$ against alternatives family B_k.

Table 3.5: Powers of C_n and $C_{m,n}$ for family B_k

	k			
	0.4	0.6	0.8	1
C_n	0.9994	0.7388	0.1371	0.05
$C_{m,n}, m = 15$	0.998	0.534	0.071	0.05
$C_{m,n}, m = 25$	1	0.6175	0.1075	0.05
$C_{m,n}, m = 30$	1	0.7	0.12	0.05
$C_{m,n}, m = 35$	1	0.716	0.1405	0.05
$C_{m,n}, m = 40$	1	0.701	0.132	0.05
$C_{m,n}, m = 50$	1	0.723	0.1425	0.05

Table 3.6: Powers of C_n and $C_{m,n}$ for family B_k

	k			
	1.25	1.5	1.75	2
C_n	0.0911	0.3925	0.8433	0.9884
$C_{m,n}, m = 15$	0.05615	0.23	0.657	0.945
$C_{m,n}, m = 25$	0.0732	0.25	0.69	0.963
$C_{m,n}, m = 30$	0.0708	0.2965	0.7585	0.972
$C_{m,n}, m = 35$	0.076	0.353	0.83	0.9835
$C_{m,n}, m = 40$	0.07905	0.3265	0.7915	0.979
$C_{m,n}, m = 50$	0.0815	0.3725	0.8295	0.9785

From Table 3.5 and Table 3.6, although bigger values of m make statistics $C_{m,n}$

perform better, it is not satisfactory compared to C_n.

Then, we compare the power of S_n and $S_{m,n}$.

Table 3.7: Powers of S_n and $S_{m,n}$ for family A_k

	k									
	0.2	0.4	0.6	0.8	1	1.25	1.5	1.75	2	2.25
S_n	1	1	0.9533	0.4412	0.05	0.4315	0.9433	0.998	1	1
$S_{m,n} m = 25$	1	1	0.9842	0.4316	0.05	0.4252	0.9553	0.999	1	1
$S_{m,n} m = 30$	1	1	0.9861	0.4349	0.05	0.4274	0.9502	0.999	1	1
$S_{m,n} m = 35$	1	1	0.9885	0.4374	0.05	0.428	0.9578	0.999	1	1

Table 3.7 shows that, although the improvement of $S_{m,n}$ is not satisfactory for dealing with alternatives distribution close to uniform, the power is increased to a level comparable to that of $C_{m,n}$ when $k = 0.6$ and 1.25. The calculation load of $S_{m,n}$ is much lower than $C_{m,n}$. Taking this into consideration, we point out that it is competitive when the alternative distribution is different from uniform in a moderate extent. For that purpose, we can see from Table 3.7 that $m = 35$ works the best.

For family B_k, the performances of $S_{m,n}$ compared to that of S_n will be much better.

Table 3.8: Powers of S_n and $S_{m,n}$ for family B_k

	k								
	0.4	0.6	0.8	1	1.25	1.5	1.75	2	2.25
S_n	0.9984	0.6558	0.1434	0.05	0.1048	0.3498	0.7104	0.9374	0.9932
$S_{m,n}, m = 25$	0.999	0.7189	0.1641	0.05	0.1069	0.3551	0.7406	0.9559	0.9972
$S_{m,n}, m = 30$	1	0.7166	0.1672	0.05	0.1059	0.3554	0.7512	0.9572	0.9985
$S_{m,n}, m = 35$	1	0.7278	0.1656	0.05	0.1064	0.3712	0.756	0.9621	0.9968

We can see from Table 3.13 that for all value of m in comparison, the power of $S_{m,n}$ dominates that of S_n almost for every k, and when $m = 35$ the statistics $S_{m,m}$ performs the best.

We point out that all of test statistics tested behave bad when dealing with family B_k, $k = 0.8$ and 1.25 in case of fixed sample size $n = 100$. The graph below shows how these two alternatives are close to an Uniform $(0, 1)$. We can see that the cumulative distribution function of family B_k coincides with Uniform $(0, 1)$ at point $x = 0.5$ which makes them behave more closely to Uniform $(0, 1)$ than family A_k does.

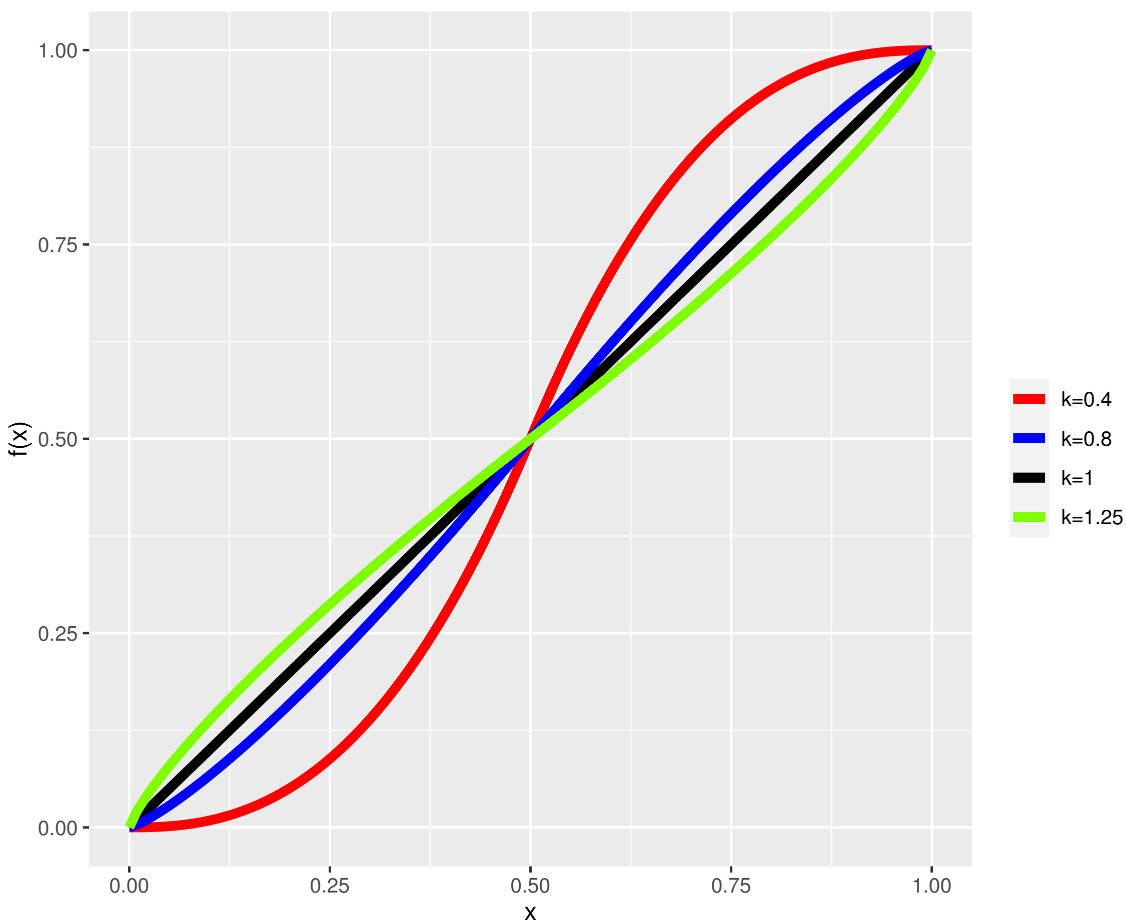

Figure 3.4: Density functions for family B_k when $k = 0.4, 0.8$ and 1.25

We give the table of inefficiencies of all test statistics. Noticed that this table summarizes the statistics in the context where the sample size $n = 100$.

Table 3.9: Maximum Power and Inefficiencies of All Statistics for family A_k

	k								
	0.4	0.6	0.8	1	1.25	1.5	1.75	2	$\sum(i_T)$
$\beta_{\max}$	1	0.994	0.5024	0.05	0.4936	0.9578	1	1	N/A
$i_{C_n}(A_k)$	0	0.006	0.034	0	0.021	0.0068	0.001	0	0.0688
$i_{C_{m,n}}(A_k), m = 15$	0	0.0035	0.0714	0	0.0591	0.0223	0.001	0	0.1573
$i_{C_{m,n}}(A_k), m = 25$	0	0.0084	0.0051	0	0.0392	0.0176	0.001	0	0.0713
$i_{C_{m,n}}(A_k), m = 30$	0	0	0.016	0	0.021	0.0033	0	0	0.0403
$i_{C_{m,n}}(A_k), m = 35$	0	0.0032	0	0	0	0.0066	0.001	0	0.0108
$i_{C_{m,n}}(A_k), m = 40$	0	0.0058	0.0092	0	0.0106	0.005	0.001	0	0.0316
$i_{C_{m,n}}(A_k), m = 50$	0	0.0035	0.0191	0	0.0226	0.0093	0.001	0	0.0555
$i_{Sn}(A_k)$	0	0.0407	0.0612	0	0.0621	0.0145	0.002	0	0.1805
$i_{S_{m,n}}(A_k), m = 25$	0	0.0092	0.0708	0	0.0684	0.0025	0.001	0	0.1519
$i_{S_{m,n}}(A_k), m = 30$	0	0.0073	0.0675	0	0.0662	0.0076	0.001	0	0.1496
$i_{S_{m,n}}(A_k), m = 35$	0	0.0055	0.065	0	0.0656	0	0.001	0	0.1371

We use the summation of all powers for a given statistics instead of maximum of them, as this will give a holistic overview of how it perform. From Table 3.9, we can see that inefficiencies of $S_{m,n}$ is higher than that of $C_{m,n}$. However, embedding Bernstein Polynomial into both of ecdf statistics (C_n and S_n) will improve their performances. Further, $S_{m,n}$ works more efficiently when dealing with special alternative distributions, as we mentioned above.

At this point, one problem left is no matter whether our Bernstein ecdf statistics can improve the power or not, they behave bad overall in the case that the real sample are coming from a distribution close to uniform $(0, 1)$. One natural solution to deal with this is to improve the sample size n. The following table shows the power of $S_{m,n}$ in testing against family B_k with $k = 0.8$, 0.9, 1.125 and 1.25, with a large sample size $n = 500$. We will not include the power for family A_k as all statistics have satisfactory power (> 0.98) and the differences are pretty small. Also, a graph is provided to make visually comparison easily. Noticed that we change the scale of value m because the optimal m should increase as n increases.

Table 3.10: Powers of S_n and $S_{m,n}$ for family B_k, $n = 500$

	Critical Value	$B_{0.8}$	$B_{0.9}$	B_1	$B_{1.125}$	$B_{1.25}$
S_n	0.05945	0.622	0.141	0.050	0.119	0.598
$S_{m,n}, m = 30$	0.47	0.677	0.156	0.051	0.132	0.632
$S_{m,n}, m = 40$	0.04498	0.712	0.172	0.049	0.167	0.692
$S_{m,n}, m = 50$	0.04652	0.682	0.172	0.051	0.161	0.640
$S_{m,n}, m = 60$	0.04672	0.687	0.175	0.050	0.174	0.661
$S_{m,n}, m = 70$	0.04773	0.682	0.171	0.053	0.156	0.6373

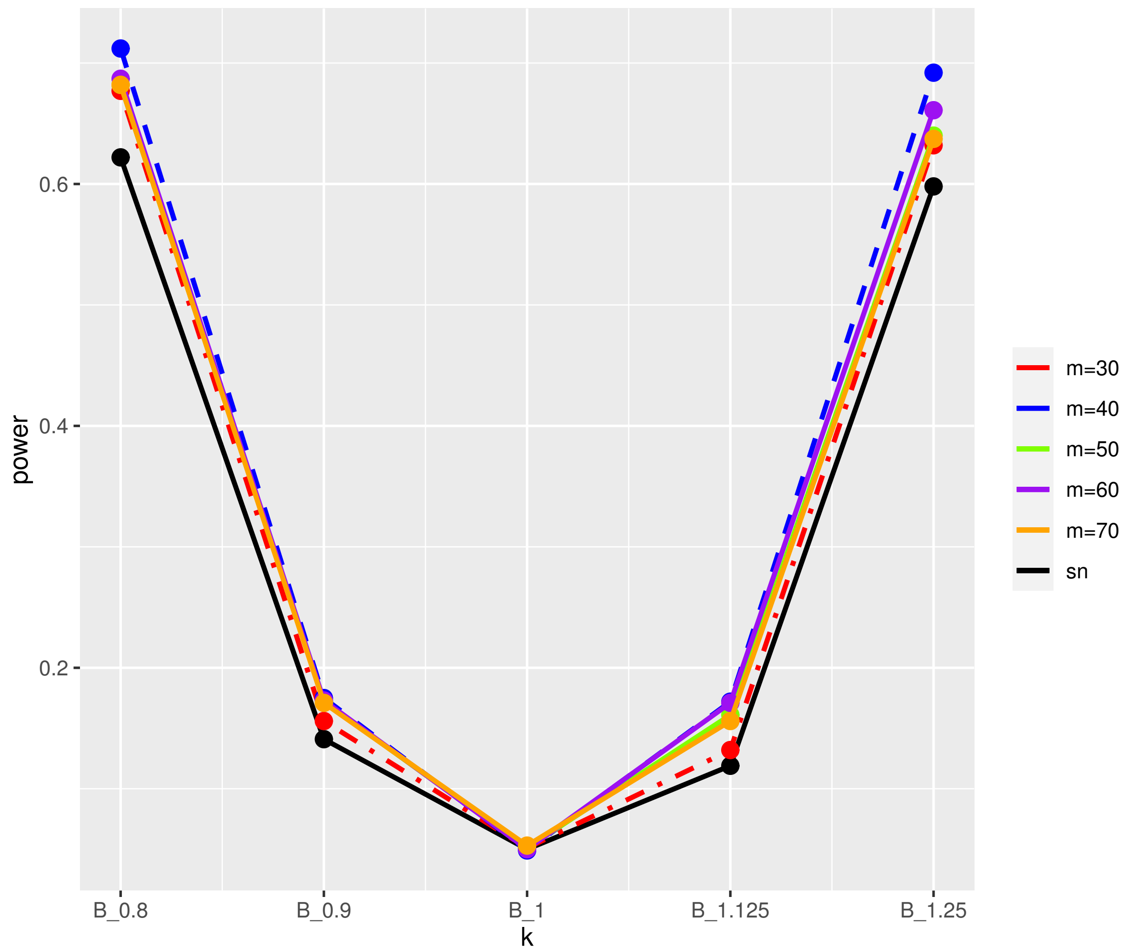

Figure 3.5: Powers of S_n and $S_{m,n}$ for family B_k, $n = 500$

Noticed that we include the power for even more tough case of $k = 0.9$ and 1.25 for comparison. If we want to detect that nuance difference from null, we are inevitable to increase the sample size even more. We see from the table and graph above that almost all choices of m improve the power greatly compared to S_n. When the underlying alternatives are $B_{0.8}$ and $B_{1.25}$, our statistics improve the power for about 0.05 to 0.08 and reach a somewhat satisfactory level. Further, noticed that $m = 40$ works the best, we then test m in smaller scale to find an optimal m value. We keep the rows for S_n and $S_{40,n}$ for comparison.

Table 3.11: Powers of S_n and $S_{m,n}$ for family B_k, small scale

	Critical Value	$B_{0.8}$	$B_{0.9}$	B_1	$B_{1.125}$	$B_{1.25}$
S_n	0.05945	0.622	0.141	0.050	0.119	0.598
$S_{m,n}, m = 38$	0.04516	0.690	0.171	0.049	0.163	0.650
$S_{m,n}, m = 39$	0.04504	0.709	0.181	0.052	0.177	0.688
$S_{m,n}, m = 40$	0.04498	0.712	0.172	0.049	0.167	0.692
$S_{m,n}, m = 41$	0.04560	0.696	0.165	0.051	0.167	0.648
$S_{m,n}, m = 42$	0.04553	0.691	0.166	0.049	0.163	0.651

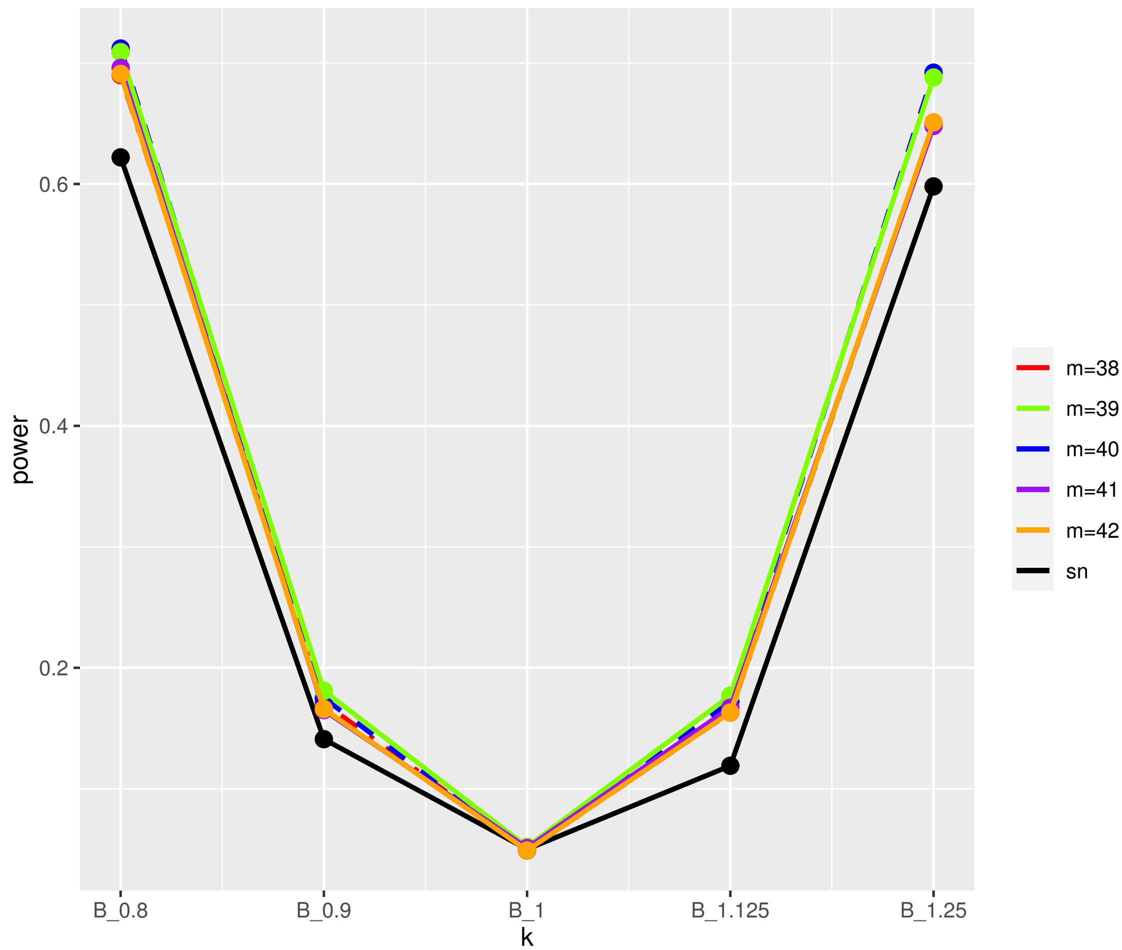

Figure 3.6: Powers of S_n and $S_{m,n}$ for family B_k, $n = 500$

We found that both $m = 39$ and $m = 40$ are the most competitive. Additionally, the following table shows the power of statistics $C_{m,n}$ testing against those four alternative distributions.

Table 3.12: Powers of C_n and $C_{m,n}$ for family A_k and B_k, $n = 500$

	Critical Value	$B_{0.8}$	$B_{0.9}$	B_1	$B_{1.125}$	$B_{1.25}$
C_n	0.47633	0.690	0.133	0.053	0.123	0.687
$C_{m,n}, m = 35$	0.4035633	0.676	0.133	0.051	0.125	0.650
$C_{m,n}, m = 45$	0.4112255	0.690	0.131	0.047	0.123	0.666
$C_{m,n}, m = 55$	0.4168085	0.698	0.136	0.046	0.116	0.671
$C_{m,n}, m = 65$	0.4227846	0.703	0.140	0.051	0.143	0.681
$C_{m,n}, m = 75$	0.426895	0.718	0.147	0.051	0.141	0.691
$C_{m,n}, m = 85$	0.4208105	0.707	0.141	0.053	0.151	0.698

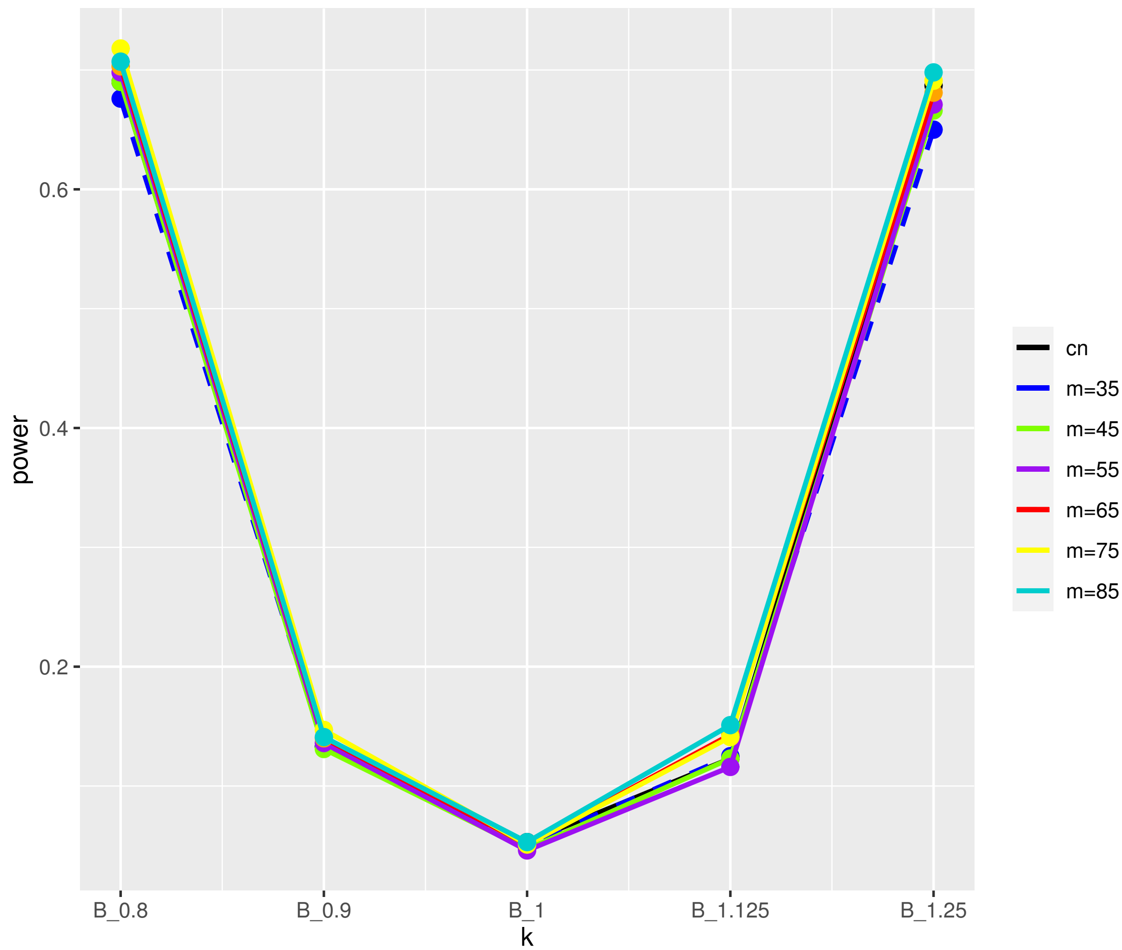

Figure 3.7: Powers of C_n and $C_{m,n}$ for family B_k, $n = 500$

We see from the table and graph that C_n itself has relatively high power for the large sample size. However, its power against B_k is a little lower than that of $S_{m,n}$ and its power against A_k is outperformed by $C_{m,n}$. Further, we see that $m = 75$ and $m = 85$ perform better than others in all of these four alternatives. However, their performances differ with each other. So, we do a small scale test to find the optimum value of m.

Table 3.13: Powers of C_n and $C_{m,n}$ for family A_k and B_k, $n = 500$

	Critical Value	$B_{0.8}$	$B_{0.9}$	B_1	$B_{1.125}$	$B_{1.25}$
$C_{m,n}, m = 71$	0.421671	0.707	0.121	0.052	0.125	0.690
$C_{m,n}, m = 73$	0.456934	0.684	0.117	0.056	0.128	0.663
$C_{m,n}, m = 75$	0.42698	0.718	0.147	0.049	0.143	0.691
$C_{m,n}, m = 77$	0.424712	0.709	0.131	0.044	0.141	0.696
$C_{m,n}, m = 79$	0.43433	0.693	0.130	0.053	0.140	0.676
$C_{m,n}, m = 81$	0.426542	0.717	0.143	0.051	0.131	0.704
$C_{m,n}, m = 83$	0.43975	0.690	0.129	0.053	0.129	0.667

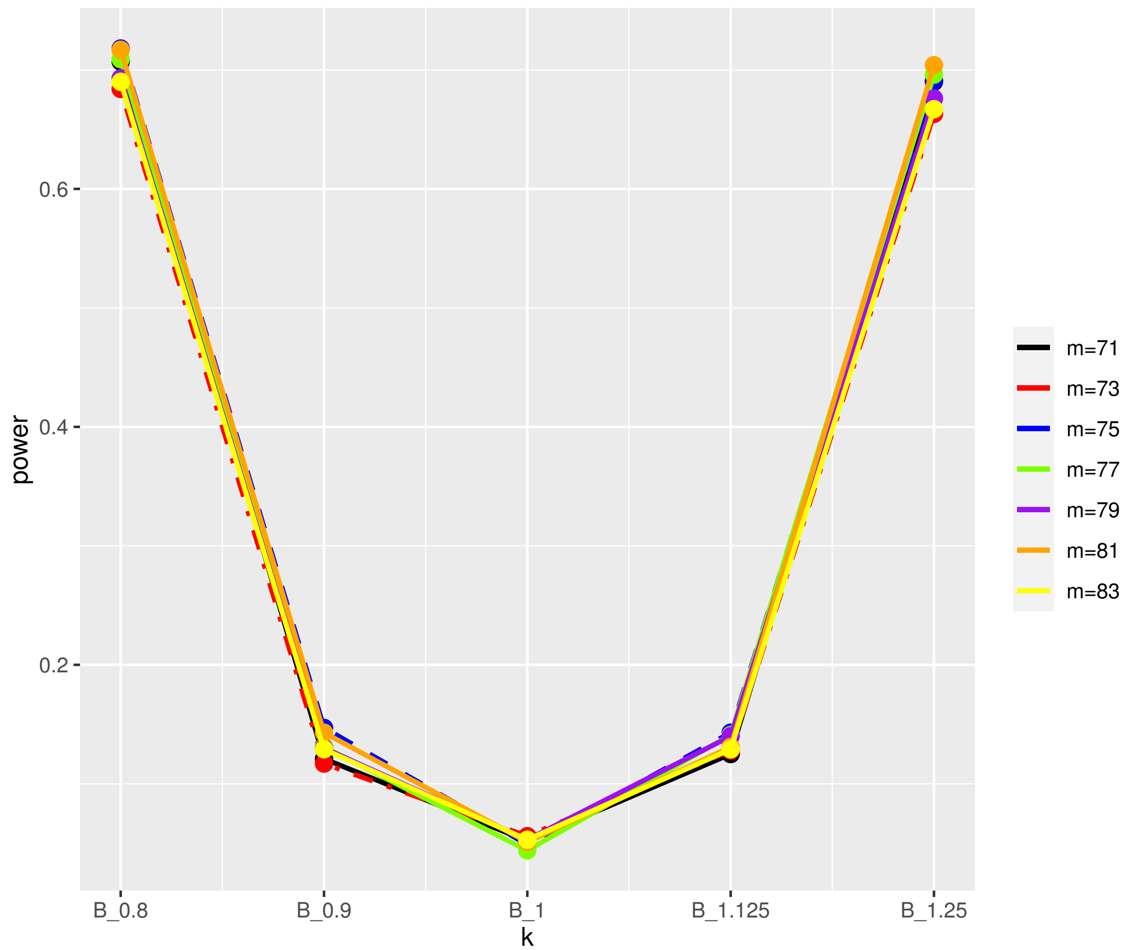

Figure 3.8: Powers of $C_{m,n}$ for family B_k, $n = 500$

From the table and graph, we see basically $m = 75, 77, 81, 85$ are competitive. Overall $m = 81$ has the highest power against family B_k when $k = 1.25$, and it is the only one who has the power larger than 0.7.

In conclusion, Bernstein ecdf statistics $C_{m,n}$ and $S_{m,n}$ have improved the performance overall in terms of the power for a given significance level. When the suspected underlying alternative distribution of sample are deviated from Uniform to a moderate extent, take sample size $n = 100$ will lead to satisfactory result. We suggest to use

$C_{m,n}$ with $m = 35$ in this case. However, if the alternative distribution behaves close to Uniform, like the families with $k = 0.8$ and 1.25 in previous examples, we suggest to increase the sample size for better detecting the nuance. If the density of data reaches maximum around tails, which is similar to the A_k family, we suggest to use $C_{m,n}$ with $m = 75$ to get the biggest power. Conversely, if the density reaches the highest around 0.5, using $C_{m,n}$ with $m = 81$ yield the best result. However, if the shape of underlying distribution is hard to judge, but only the closeness to uniformity is supposed, then one can try $S_{m,n}$ with $m = 39$ to reach the overall higher efficiency.

Conclusion

In this major paper, we add two new members, $C_{m,n}$ and $S_{m,n}$, to the huge family of uniformity testing statistics. They perform better than C_n and S_n respectively, especially in dealing with alternative distributions which have more density at end points 0 and 1. We also have discussed the optimal choice of the parameter m when n is fixed, and interestingly it does not coincide with optimal choice $m = \frac{n}{\log n}$ in function approximation. However, the purposed statistics can still be improved for those alternative distributions which are very close to uniform distribution. We think the future research works could be in the following two directions.

i. As mentioned in this work, the statistics $C_{m,n}$ and $S_{m,n}$ are asymptotically converge to some linear functional imposed on a Gaussian process $\mathbb{F}(x)$. Although it seems mathematically challenging, it maybe possible to derive the concrete distribution of these two limit through some advanced mathematical techniques.

ii. A statistic V_n is closely related to the Kolmogorov-Smirnov test type, and it is calculated in similar way as that of D_n. We talked about D_n^+ and D_n^- by Kuiper (1960) in this major paper, and V_n is the summation of D_n^+ and D_n^-. It is showed in Marhuenda et al. (2005) that it has much less inefficiencies than D_n and Cramér-Von Mises type C_n. As a result, we suggest that embedding

Bernstein polynomial into this statistic V_n might have even better results. We may denote this new statistic as $V_{m,n}$. The expression of V_n is more complicated than that of D_n. As a result, it will be more difficult to derive the properties of $V_{m,n}$. However, $V_{m,n}$ has the potential to approach the lowest inefficiency among all e.c.d.f. estimators.

Appendix A

Some Topological Background

Definition A.1. *(Topological Space)*

Let $X(\neq \emptyset)$ be a set and let $\mathcal{P}(X)$ stands for the power set of X. A topology on X is a collection $\tau \subset \mathcal{P}(X)$ satisfying:

i. $\emptyset, X \in \tau$.

ii. *If $U_i \in \tau$ for $i \in I$, then $\bigcup_{i \in I} U_i \in \tau$.*

iii. *If $U_1, U_2 \in \tau$, then $U_1 \cap U_2 \in \tau$.*

A set X together with a topology τ is called a topological space.

We also need the definitions of compact space and Hausdorff space:

Definition A.2. *(Compactness)*

Let (X, τ) be a topological space. X is said to be compact if every open cover of X ($X \subset \bigcup_{\alpha \in A} G_\alpha$ with $G_\alpha \in \tau$) has a finite subcover, namely, there exist $\alpha_1, \ldots, \alpha_n$ such that $X \subset \bigcup_{i=1}^{n} G_{\alpha_i}$.

Definition A.3. *(Hausdorff Space)*

Let (X, τ) be a topological space. X is said to be Hausdorff if:

$\forall x \neq y (\in X)$, $\exists x \in U, y \in V$ *such that* $U, V \in \tau$ *and* $U \cap V = \emptyset$.

The uniform convergence is a type of convergences of a series of functions to a limit function. It is defined as follows:

Definition A.4. *(Uniform Convergence)*

Suppose E is a set, and $(f_n)_{n\in\mathbb{N}}$ be a sequence of real-valued functions on E. We say that the sequence $(f_n)_{n\in\mathbb{N}}$ uniformly converge to $f : E \longrightarrow \mathbb{R}$ on E if for arbitrary $\epsilon > 0$, there exist a $N \in \mathbb{N}$ such that for all $n > N$ and $x \in E$, we will have:

$$|f_n(x) - f(x)| < \epsilon.$$

With the same definition of E and $(f_n)_{n\in\mathbb{N}}$, it is equivalent to write the definition of $(f_n)_{n\in\mathbb{N}}$ uniformly converge to f as:

$$\lim_{n\longrightarrow\infty} \sup_{x\in E} |f_n(x) - f(x)| = 0.$$

Appendix B

Theorems in Literature

Theorem B.1. *(Tenbusch (1994)) Assume that $f(x,y)$ is bounded on the square and all partial derivatives of order two exist and continuous in a neighborhood of (x,y). Further, assume that $f(x,y) \neq 0$. Let n tend to infinity such that $KnN^{-1/3}$ converges to 1 for some constant K. Then*

$$\mathbb{E}[(\widehat{f}_{nN}(x,y) - f(x,y))^2] = \frac{K^2\alpha^2(x,y) + K^{-1}B(x,y)}{N^{2/3}} \qquad if \qquad 0 < x, y < 1$$

where

$$\alpha(x,y) = \frac{1}{2}\left(\frac{\partial f(x,y)}{\partial x}(1-2x) + \frac{\partial f(x,y)}{\partial y}(1-2y) + \frac{\partial^2 f(x,y)}{\partial x^2}x(1-x) + \frac{\partial^2 f(x,y)}{\partial y^2}y(1-y)\right)$$

and

$$B(x,y) = \frac{f(x,y)}{4\pi(x(1-x)y(1-y))^{1/2}}.$$

Theorem B.2. *(Tenbusch (1994)) Assume that $f(x,y)$ is bounded on the square and all partial derivatives of order two exist and continuous in a neighborhood of (x,y).*

Further, assume that $f(x,y) \neq 0$. Let n tend to infinity such that $KnN^{-1/3}$ converges to 1 for some constant K. Then

$$\mathbb{E}[(\widehat{f}_{nN}(x,y) - f(x,y))^2] = O(N^{-4/7})$$

If $x = 0$ or $x = 1$ and $0 < y < 1$ or if $y = 0$ or $y = 1$ and $0 < x < 1$.

Theorem B.3. *(Babu et al. (2002)) For $2 \leq m \leq (n/\log n)$, we have a.s. as $n \longrightarrow \infty$,*

$$\sup_{0\leq x\leq 1} |\widehat{f}_{m,n}(x) - f(x)| = O((m^{1/2}(n^{-1}\log n)^{1/2})) + O(\sup_{0\leq x\leq 1} |F'_m(x) - f(x)|)$$

where F'_m denotes the derivative of F^*_m, and

$$F'_m(x) = \sum_{k=0}^{m} F\left(\frac{k}{m}\right) P_{m,k}(x)$$

.

Theorem B.4. *(Babu et al. (2002)) If $f(x) > 0$, then*

$$n^{1/2}m^{1/4}\left(\widehat{f}_{m,n}(x) - f(x)\right) \xrightarrow{d} \mathcal{N}\left(0, f(x)(4\pi x(1-x))^{-1/2}\right)$$

as $m, n \longrightarrow \infty$ such that $2 \leq m \leq (n/\log n)$ and $n^{2/3}/m \longrightarrow 0$.

Theorem B.5. *(Leblanc (2010))*

If f is continuous (and bounded) and admits two continuous and bounded derivatives on $[0,1]$, then:

$\mathbb{B}ias[\widehat{f}_{m,n}(x)] = \mathbb{E}[f_{m,n}(x)] - f(x) = \frac{1}{2m}(1-2x)f^{'}(x) + x(1-x)f^{''}(x)] + o(m^{-1}) \qquad \forall x \in [0,1]$

Theorem B.6. *(Leblanc (2012))*

Under same conditions as assumed in Theorem B.5, we have:

$\mathbb{B}ias[\widehat{F}_{m,n}(x)] = \mathbb{E}[F_{m,n}(x)] - F(x) = m^{-1}b(x) + o(m^{-1}) \qquad \forall x \in [0,1]$ *where* $b(x) = x(1-x)f^{'}(x)/2$

Theorem B.7. *(Leblanc (2012))*

Under same conditions as assumed in Theorem B.5, we have:

$\mathrm{MSE}[\widehat{F}_{m,n}(x)] = n^{-1}F(x)[1-F(x)] - m^{-1/2}n^{-1}f(x)[2x(1-x)/\pi]^{1/2} + m^{-2}b^2(x) + o(m^{-2}) + o(m^{-1/2}n^{-1})$

Theorem B.8. *(Leblanc (2012))*

Assuming $F(x)$ *is continuous and holds two continuous and bounded derivatives on* $[0,1]$*, we have:*

$$n^{1/2}(\widehat{F}_{m,n}(x) - \widehat{F}_m(x)) \xrightarrow{d} \mathcal{N}(0, F(x)[1-F(x)]).$$

Proof of Theorem 3.1. To show $g_n(X_n) \Rightarrow g(X)$, it's suffice to show

$$P(g(X) \in G) \leq \lim \inf_{n \longrightarrow \infty} P(g_n(X_n) \in G)$$

For any open G in (S', d'). Suppose that $g(x) \in G$ for some open set $G \in S'$, as g is continuous, there exist k and δ such that $g_i(y) \in G$ for $i \geq k$ if $m(x,y) < \delta$. Then $x \in T_k$ where $T_k = \bigcap_{i \geq k} g_i^{-1}(G)$, as $x \in g_i^{-1}(G)$ for every $i \geq k$.

Then, as $T_k \subset T_{k+1} \subset T_{k+2} \subset \ldots$, so we have $g^{-1}(G) \subset E \bigcup (\bigcup_{k=1}^{\infty} T_k)$ Then, as $T_k \subset T_{k+1}$ and $P(x \in E) = 0$, for $\forall \epsilon > 0$, there exist a k such that

$$P(X \in g^{-1}(G)) \leq P(X \in \bigcup_k T_k) \leq P(X \in T_k) + \epsilon.$$

The second inequality hold as the set T_k is increasing, and $lim_{k \longrightarrow \infty} T_k = \bigcup_k T_k$. As a result, we can always find a large enough k for any small ϵ to let the inequality holds.

Additionally, since $X_n \Rightarrow X$ and $T_k \subset g_n^{-1}(G)$ for $n \geq k$, we have

$$P(X \in T_k) \leq \lim \sup_{n \longrightarrow \infty} P(X_n \in T_k) \leq \lim_{n \longrightarrow \infty} P(X_n \in g_n^{-1} G).$$

Finally, as ϵ is arbitrary, we combine these two inequalities to get the desired result.

□

Proof of Theorem 3.3. We assume that $E\left(|X_i^4|\right) < \infty$ for all i's, as stated in chapter 3.3. Then we denote $Var(X_i) = \sigma^2$ for all i's. Further, without the loss of generality, we assume X_i's are i.i.d. with zero mean and going to prove the sample mean $\bar{X_n}$ converge to 0 almost surely. Indeed, if we have $E(X_i) = \mu$, we can let $Y_i = X_i - \mu$ and prove $\bar{Y_n}$ converge to 0 almost surely. Let $S_n = \sum_{i=1}^{n} X_i$. Then, we apply Chebyshev's Inequality to S_n at the power of 4. For any $\epsilon > 0$, We have

$$P\left(|S_n| > n\epsilon\right) \leq \frac{1}{(n\epsilon)^4} \mathbb{E}\left(|S_n|^4\right).$$

For $|S_n|^4$, we have

$$\mathbb{E}\left(\left(\sum_{i=1}^{n} X_i\right)^4\right) = n\mathbb{E}(X_1^4) + 3n(n-1)\mathbb{E}(X_1^2)\mathbb{E}(X_2^2)$$

Then,

$$\mathbb{E}\left(|S_n|^4\right) = n\mathbb{E}(X_1^4) + 3n(n-1)\sigma^4.$$

As all other term written as $\mathbb{E}(X_iX_jX_hX_k)$ and $\mathbb{E}(X_iX_j^3)$ will be 0 under our assumption of zero mean and independence. Then, as $\sigma^2 < \infty$ and $E\left(|X_i^4|\right) < \infty$ by our assumption, this term is bounded by n^2 times some constant, which written as

$$\mathbb{E}\left(|S_n|^4\right) \leq n^2 C.$$

Then, we have

$$P\left(S_n > n\epsilon\right) \leq P\left(|S_n| > n\epsilon\right) \leq \frac{1}{(n\epsilon)^4}\mathbb{E}\left(|S_n|^4\right) \leq \frac{n^2 C}{n^4\epsilon^4} = \frac{C}{n^2\epsilon^4}.$$

Therefore, we conclude $\sum_{n=1}^{\infty} P\left(S_n > n\epsilon\right) < \infty$. By Borel-Cantelli Lemma, we have the probability of event $(S_n > n\epsilon)$ happen infinitely often (i.o.) is 0. So that $P\left\{\lim_{n\to\infty} \bar{X}_n = 0\right\} = 1$. Therefore, we have the almost surely convergence for $\bar{X}_n$ to the population mean.

□

Proof of Lemma 3.1. We take partition $t_1, t_2, \ldots, t_k$ over $[0,1]$ as $0 \leq t_1 < t_2 < \cdots < t_k \leq 1$. Noticing that

$$\mathbb{E}(F_n(x)) = \mathbb{E}(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(X_i < t)) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{P}(X_i < t) = F(x),$$

we apply Central Limit theorem for the multidimensional case and get

$$\sqrt{n}\left(F_n(t_1) - F(t_1), \ldots, F_n(t_k) - F(t_k)\right) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where the covariance matrix Σ is given by

$$Cov(F_n(t_i) - F(t_i), F_n(t_j) - F(t_j)) = F(\max(t_i, t_j)) - F(t_i)F(t_j).$$

Further, in our cases, the expectation matrix of the limiting distribution is all zero, and the covariance matrix Σ is bounded because $F(\max(t_i, t_j))$ and $F(t_i)F(t_j)$ are both bounded by 1. Also the empirical process is asymptotically continuous when n goes to infinity. As this limiting random variable is bounded in a compact space (namely continuous bounded function space $\mathbb{C}_B(0, 1)$), it is a tight measure and therefore the process is Gaussian. We can denote it as $\mathbb{F}$.

Then the Lemma follows by taking $t_i = s$ and $t_j = t$ with $0 \leq s \leq t \leq 1$.

□

Bibliography

Anderson, T. W. and D. A. Darling (1954). A test of goodness of fit. *Journal of the American statistical association 49*(268), 765–769.

Babu, G. J., A. J. Canty, and Y. P. Chaubey (2002). Application of Bernstein polynomials for smooth estimation of a distribution and density function. *Journal of Statistical Planing and Inference. 105*, 377–392.

Babu, G. J. and Y. P. Chaubey (2006). Smooth estimation of a distribution and density function on a hyper-cube using Bernstein polynomials for dependent random vectors. *Statistics and Probability Letters 76*, 959–969.

Belalia, M. (2016). On the asymptotic properties of the bernstein estimator of the multivariate distribution function. *Statistics & Probability Letters 110*(C), 249–256.

Belalia, M., T. Bouezmarni, and A. Leblanc (2017). Smooth conditional distribution estimators using bernstein polynomials. *Computational Statistics and Data Analysis 111*, 166 – 182.

Belalia, M., T. Bouezmarni, and A. Leblanc (2019). Bernstein conditional density estimation with application to conditional distribution and regression functions. *Journal of the Korean Statistical Society 48*(3), 356–383.

Belalia, M., T. Bouezmarni, F. Lemyre, and A. Taamouti (2017). Testing independence based on bernstein empirical copula and copula density. *Journal of Nonparametric Statistics 29*(2), 346–380.

Bernshtein, S. (1952). On the best approximation of continuous functions by polynomials of given degree. *Collected works 1*, 11–104.

Bernstein, S. (1912). Démonstration du théorème de weierstrass fondée sur le calcul des probabilités. *13*(1), 1–2.

Boyle, P., R. Flowerdew, and A. Williams (1997). Evaluating the goodness of fit in models of sparse medical data: a simulation approach. *International journal of epidemiology 26*(3), 651–656.

Chandra, M. and N. D. Singpurwalla (1981). Relationships between some notions which are common to reliability theory and economics. *Mathematics of Operations Research 6*(1), 113–121.

Cressie, N. (1978). Power results for tests based on high-order gaps. *Biometrika 65*(1), 214–218.

Cressie, N. (1979). An optimal statistic based on higher order gaps. *Biometrika 66*(3), 619–627.

Dang, C., Z. F. Li, and C. Yang (2018). Measuring firm size in empirical corporate finance. *Journal of Banking & Finance 86*, 159–176.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. académie royale de belgique. *Bulletin de la Classe des Sciences 65*(5), 274–292.

Deken, J. G. (1981). Exact distributions for gaps and stretches. Technical report, STANFORD UNIV CA DEPT OF STATISTICS.

Etemadi, N. (1981). An elementary proof of the strong law of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 55*(1), 119–122.

Feller, A., H. Kaupp, and J. Digiacomo (1965). Crosstalk and reflections in high-speed digital systems. In *Proceedings of the November 30–December 1, 1965, fall joint computer conference, part I*, pp. 511–525.

Greenwood, M. (1946). The statistical study of infectious diseases. *Journal of the Royal Statistical Society 109*(2), 85–110.

Kolmonorgov, K. (1933). Sulla determinazione empirica di una legge di distibuzione, g.

Kornegay, J., C. Cardona, and C. E. Posso (1993). Inheritance of resistance to mexican bean weevil in common bean, determined by bioassay and biochemical tests. *Crop Science 33*(3), 589–594.

Kuiper, N. H. (1960). Tests concerning random points on a circle. In *Nederl. Akad. Wetensch. Proc. Ser. A*, Volume 63, pp. 38–47.

Leblanc, A. (2009). Chung-Smirnov property for Bernstein estimators of distribution functions. *Journal of Nonparametric Statistics 22*(2), 459–475.

Leblanc, A. (2010). A bias-reduced approach to density estimation using Bernstein polynomials. *Journal of Nonparametric Statistics 22*, 459–475.

Leblanc, A. (2012). On estimating distribution functions using Bernstein polynomials. *Annals of the Institute of Statistical Mathematics 64*, 919–943.

Liang, J.-J., K.-T. Fang, F. Hickernell, and R. Li (2001). Testing multivariate uniformity and its applications. *Mathematics of Computation 70*(233), 337–355.

Lorentz, G. (1986). *Bernstein Polynomials* (2nd ed.). New York: Chelsea Publishing.

Marhuenda, Y., D. Morales, and M. C. Pardo (2005). A comparison of uniformity tests. *Statistics 39*(4), 315–327.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications 9*(1), 141–142.

Nataraj, P. S. and M. Arounassalame (2007). A new subdivision algorithm for the bernstein polynomial approach to global optimization. *International journal of automation and computing 4*(4), 342–352.

Neumann, A., T. Bodnar, D. Pfeifer, and T. Dickhaus (2019). Multivariate multiple test procedures based on nonparametric copula estimation. *Biometrical Journal 61*(1), 40–61.

Neyman, J. (1937). » smooth test» for goodness of fit. *Scandinavian Actuarial Journal 1937*(3-4), 149–199.

Quesenberry, C. and F. M. Jr. (1977). Power studies of some tests for uniformity. *Journal of Statistical Computation and Simulation 5*(3), 169–191.

Read, T. R. and N. A. Cressie (2012). *Goodness-of-fit statistics for discrete multivariate data*. Springer Science & Business Media.

Stephens, M. A. (1974). Edf statistics for goodness of fit and some comparisons. *Journal of the American statistical Association 69*(347), 730–737.

Stone, M. H. (1937). Applications of the theory of boolean rings to general topology. *Transactions of the American Mathematical Society 41*(3), 375–481.

Stone, M. H. (1948). The generalized weierstrass approximation theorem. *Mathematics Magazine 21*(5), 237–254.

Tenbusch, A. (1994). Two-dimensional Bernstein polynomial density estimation. *Metrika 41*, 233–253.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.

Vitale, R. (1975). A Bernstein polynomial approach to density estimation. In M. L. Puri (Ed.), *Statistical Inference and Related Topics*, Volume 2, New York, pp. 87–99. Academic Press.

Weisstein, E. W. (2015). Legendre polynomial. *Mathematical Methods 43*.

Wellner, J. et al. (2013). *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media.

Whitt, W. (2002). *Stochastic-process limits: an introduction to stochastic-process limits and their application to queues*. Springer Science & Business Media.

Zhang, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(2), 281–294.

Vita Auctoris

Mr. Ran Sun was born in 1994, Beijing, China. He graduated from Beijing High School Four in 2010. From there he went to University of California, Santa Barbara, where he obtained a B.S degree with a major in Actuarial Science and a minor in Chemistry. He is currently a candidate for the Master of Science degree in Statistics at the University of Windsor. He hopes to graduate in December 2022.