Goodness of fit via phi-divergences: a new family of test statistics

Jon A. Wellner

Joint work with Leah Jager

University of Washington

- joint work with Leah Jager,
  University of Washington

- Talk at Princeton University,
  Workshop on the Frontiers of Statistics,
  in honor of Peter J. Bickel

- *Email: jaw@stat.washington.edu*
  http: //www.stat.washington.edu/jaw/jaw.research.html

Outline

- Semiparametric models and Peter Bickel

Outline

- Semiparametric models and Peter Bickel
- Introduction: some history

Outline

- Semiparametric models and Peter Bickel
- Introduction: some history
- A new family of statistics via phi-divergences

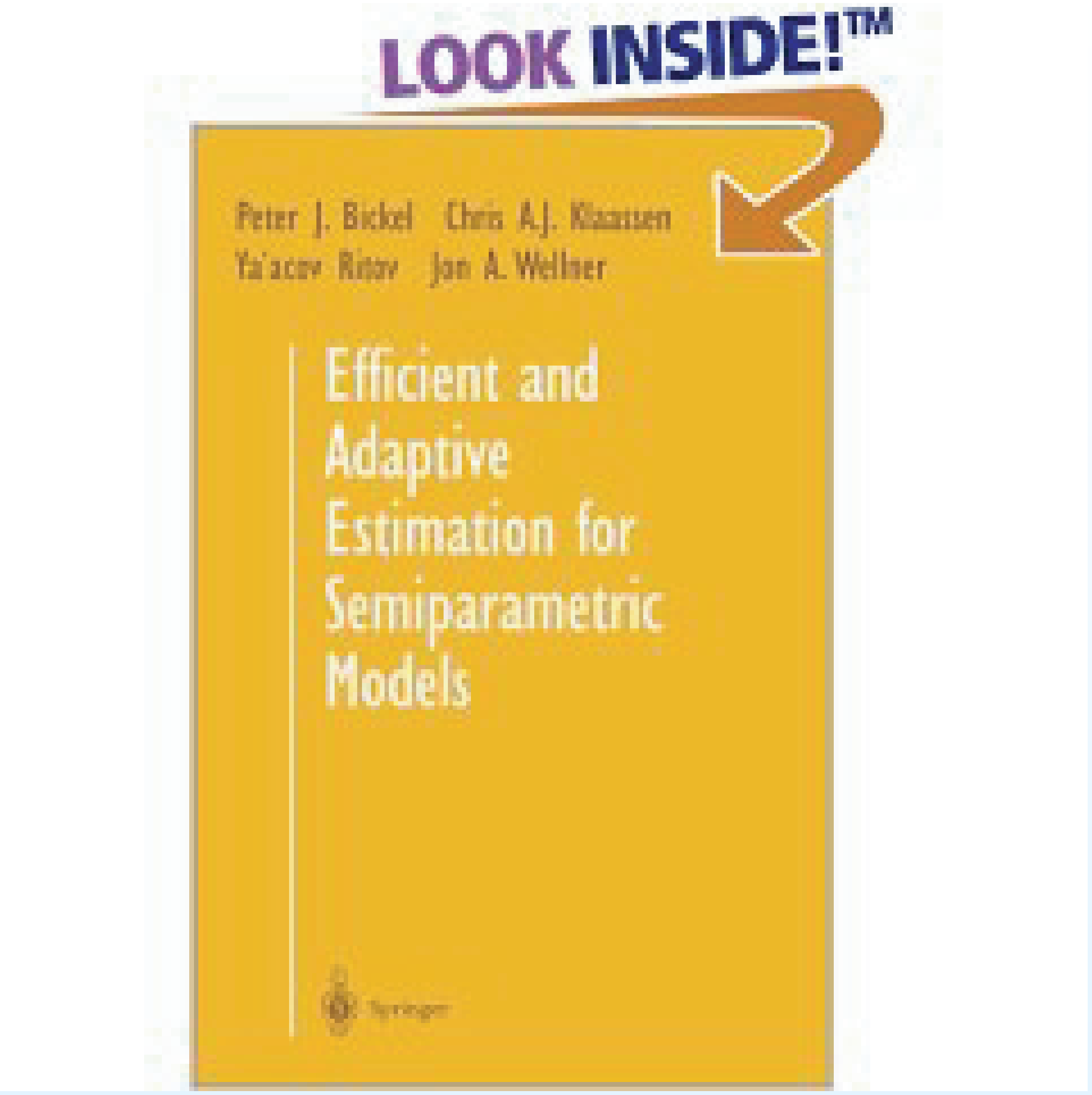
Outline

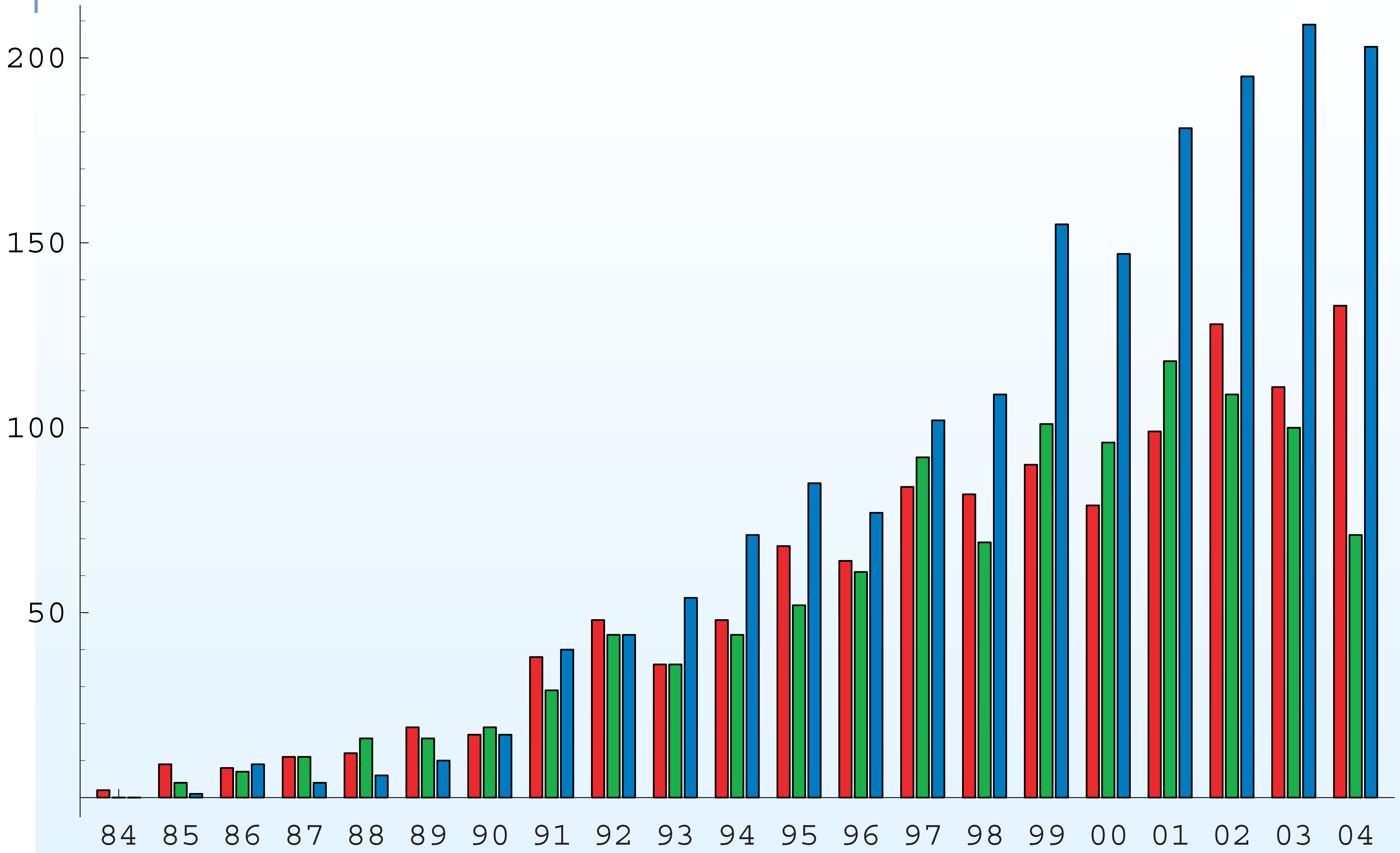
- Semiparametric models and Peter Bickel
- Introduction: some history
- A new family of statistics via phi-divergences
- Null hypothesis distribution theory: finite - sample and limit theory

Outline

- Semiparametric models and Peter Bickel
- Introduction: some history
- A new family of statistics via phi-divergences
- Null hypothesis distribution theory: finite - sample and limit theory
- Limit theory under alternatives and power

Robert Serfling, Peter Bickel, and Alan Karr
Johns Hopkins University, June 1983


LOOK INSIDE!™
Peter J. Bickel Chris A.J. Klaassen
Ya'acov Ritov Jon A. Wellner
Efficient and Adaptive Estimation for Semiparametric Models
Springer


200
150
100
50
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
00
01
02
03
04

- Missing data models
- Testing and profile likelihood theory
- Semiparametric mixture model theory
- Rates of convergence via empirical process methods
- Bayes methods in semiparametric models
- Model selection methods
- Empirical likelihood
- Transformation and frailty models
- Semiparametric regression models
- Extensions to non-i.i.d. data
- Critiques and possible alternative theories

1. Introduction: some history

- **Setting: classical “goodness - of - fit”**

1. Introduction: some history

- **Setting: classical "goodness - of - fit"**
- $X_1, \ldots, X_n$ i.i.d. with distribution function F

1. Introduction: some history

- **Setting: classical “goodness - of - fit”**
- $X_1, \ldots, X_n$ i.i.d. with distribution function F
- $\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{[X_i \leq x]}$

1. Introduction: some history

- **Setting: classical "goodness - of - fit"**
- $X_1, \ldots, X_n$ i.i.d. with distribution function F
- $\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{[X_i \leq x]}$
- Test $H : F = F_0$ versus $K : F \neq F_0$, F_0 is continuous

1. Introduction: some history

- **Setting: classical "goodness - of - fit"**
- $X_1, \ldots, X_n$ i.i.d. with distribution function F
- $\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{[X_i \leq x]}$
- Test $H : F = F_0$ versus $K : F \neq F_0$, F_0 is continuous
- Without loss of generality $F_0(x) = x$, the $U(0, 1)$ distribution

1. Introduction: some history

- **Setting: classical "goodness - of - fit"**
- $X_1, \dots, X_n$ i.i.d. with distribution function F
- $\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{[X_i \le x]}$
- Test $H : F = F_0$ versus $K : F \neq F_0$, F_0 is continuous
- Without loss of generality $F_0(x) = x$, the $U(0,1)$ distribution

- Break hypotheses down into family of pointwise hypotheses: $H_x : F(x) = F_0(x)$ versus $K_x : F(x) \neq F_0(x)$

1. Introduction: some history

- **Setting: classical "goodness - of - fit"**
- $X_1, \ldots, X_n$ i.i.d. with distribution function F
- $\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{[X_i \le x]}$
- Test $H : F = F_0$ versus $K : F \neq F_0$, F_0 is continuous
- Without loss of generality $F_0(x) = x$, the $U(0,1)$ distribution

- Break hypotheses down into family of pointwise hypotheses: $H_x : F(x) = F_0(x)$ versus $K_x : F(x) \neq F_0(x)$
- $H = \cap_x H_x$, $K = \cup_x K_x$

- Likelihood ratio statistic for testing H_x versus K_x:

$$\begin{aligned}\lambda_n(x) &= \frac{\sup_{F(x)} L_n(F(x))}{L_n(F_0(x))} = \frac{L_n(\mathbb{F}_n(x))}{L_n(F_0(x))} \\ &= \frac{\mathbb{F}_n(x)^{n\mathbb{F}_n(x)}(1-\mathbb{F}_n(x))^{n(1-\mathbb{F}_n(x))}}{F_0(x)^{n\mathbb{F}_n(x)}(1-F_0(x))^{n(1-\mathbb{F}_n(x))}} \\ &= \left(\frac{\mathbb{F}_n(x)}{F_0(x)}\right)^{n\mathbb{F}_n(x)} \left(\frac{1-\mathbb{F}_n(x)}{1-F_0(x)}\right)^{n(1-\mathbb{F}_n(x))}\end{aligned}$$

- Thus

$$\log \lambda_n(x) = n\mathbb{F}_n(x) \log\left(\frac{\mathbb{F}_n(x)}{F_0(x)}\right) + n(1 - \mathbb{F}_n(x)) \log\left(\frac{1 - \mathbb{F}_n(x)}{1 - F_0(x)}\right)$$

$$= nK(\mathbb{F}_n(x), F_0(x))$$

- Thus

$$\begin{aligned}\log \lambda_n(x) &= n\mathbb{F}_n(x) \log \left(\frac{\mathbb{F}_n(x)}{F_0(x)} \right) \\ &\quad + n(1 - \mathbb{F}_n(x)) \log \left(\frac{1 - \mathbb{F}_n(x)}{1 - F_0(x)} \right) \\ &= nK(\mathbb{F}_n(x), F_0(x))\end{aligned}$$

- $K(u, v) \equiv u \log \left(\frac{u}{v} \right) + (1 - u) \log \left(\frac{1-u}{1-v} \right)$,
  Kullback - Leibler "distance"
  Bernoulli(u), Bernoulli(v)

- Thus

$$\begin{aligned} \log \lambda_n(x) &= n\mathbb{F}_n(x) \log\left(\frac{\mathbb{F}_n(x)}{F_0(x)}\right) \\ &\quad + n(1-\mathbb{F}_n(x)) \log\left(\frac{1-\mathbb{F}_n(x)}{1-F_0(x)}\right) \\ &= nK(\mathbb{F}_n(x), F_0(x)) \end{aligned}$$

- $K(u,v) \equiv u \log\left(\frac{u}{v}\right) + (1-u)\log\left(\frac{1-u}{1-v}\right)$,
  Kullback - Leibler "distance"
  Bernoulli(u), Bernoulli(v)
- Berk-Jones (1979) test statistic:

$$R_n \equiv \sup_x n^{-1} \log \lambda_n(x) = \sup_x K(\mathbb{F}_n(x), F_0(x)) \,.$$

- **History:**

- **History:**
  - Berk and Jones (1979)

- **History:**
  - Berk and Jones (1979)
  - Groeneboom and Shorack (1981)

- **History:**
  - Berk and Jones (1979)
  - Groeneboom and Shorack (1981)
  - Shorack and Wellner (1986, p. 786)

- **History:**
  - Berk and Jones (1979)
  - Groeneboom and Shorack (1981)
  - Shorack and Wellner (1986, p. 786)
  - Owen (1995): inversion of R_n to get confidence bands finite - sample distribution via Noé's recursion

- **History:**
  - Berk and Jones (1979)
  - Groeneboom and Shorack (1981)
  - Shorack and Wellner (1986, p. 786)
  - Owen (1995): inversion of R_n to get confidence bands finite - sample distribution via Noé's recursion
  - Einmahl and McKeague (2002): integral version of R_n

- **History:**
  - Berk and Jones (1979)
  - Groeneboom and Shorack (1981)
  - Shorack and Wellner (1986, p. 786)
  - Owen (1995): inversion of R_n to get confidence bands finite - sample distribution via Noé's recursion
  - Einmahl and McKeague (2002): integral version of R_n
  - Donoho and Jin (2002): supremum version of Anderson-Darling statistic
    with comparison to Berk - Jones statistic R_n

2. A new family of statistics via phi-divergences

- For $s \in \mathbb{R}$, $x \geq 0$ define

$$\phi_s(x) = \begin{cases} \frac{1-s+sx-x^s}{s(1-s)}, & s \neq 0, 1 \\ x \log x - x + 1, & s = 1 \\ -\log x + x - 1, & s = 0\,. \end{cases}$$

2. A new family of statistics via phi-divergences

- For $s \in \mathbb{R}$, $x \geq 0$ define

$$\phi_s(x) = \begin{cases} \frac{1-s+sx-x^s}{s(1-s)}, & s \neq 0, 1 \\ x \log x - x + 1, & s = 1 \\ -\log x + x - 1, & s = 0\,. \end{cases}$$

- Then define

$$K_s(u, v) = v\phi_s(u/v) + (1-v)\phi_s((1-u)/(1-v))\,.$$

- Special cases:

$$
\begin{aligned}
K_1(u,v) &= K(u,v) \\
&= u\log(u/v) + (1-u)\log((1-u)/(1-v)) \\
K_0(u,v) &= K(v,u) \\
K_2(u,v) &= \frac{1}{2}\frac{(u-v)^2}{v(1-v)} \\
K_{-1}(u,v) &= K_2(v,u) = \frac{1}{2}\frac{(u-v)^2}{u(1-u)} \\
K_{1/2}(u,v) &= 2\{(\sqrt{u}-\sqrt{v})^2 + (\sqrt{1-u}-\sqrt{1-v})^2\} \\
&= 4\{1-\sqrt{uv}-\sqrt{(1-u)(1-v)}\}.
\end{aligned}
$$

- The new family of statistics:

$$S_n(s) = \begin{cases} \sup_{x\in\mathbb{R}} K_s(\mathbb{F}_n(x), F_0(x)), & s \geq 1 \\ \sup_{x\in[X_{(1)},X_{(n)})} K_s(\mathbb{F}_n(x), F_0(x)), & s < 1\,, \end{cases}$$

- The new family of statistics:

$$S_n(s) = \begin{cases} \sup_{x\in\mathbb{R}} K_s(\mathbb{F}_n(x), F_0(x)), & s \geq 1 \\ \sup_{x\in[X_{(1)},X_{(n)})} K_s(\mathbb{F}_n(x), F_0(x)), & s < 1\,, \end{cases}$$

- Thus, with $F_0(x) = x$,

$$S_n(1) = R_n, \qquad S_n(0) = \text{“reversed” Berk-Jones} \equiv \widetilde{R}_n$$

$$S_n(2) = \frac{1}{2}\sup_{x\in\mathbb{R}} \frac{(\mathbb{F}_n(x) - x)^2}{x(1-x)}, \qquad S_n(-1) = \frac{1}{2}\sup_{x\in[X_{(1)},X_{(n)})} \frac{(\mathbb{F}_n(x)}{\mathbb{F}_n(x)(1 -}$$

$$S_n(1/2)$$

$$= 4 \sup_{x\in[X_{(1)},X_{(n)})} \{1 - \sqrt{\mathbb{F}_n(x)x} - \sqrt{(1-\mathbb{F}_n(x))(1-x)}\}$$

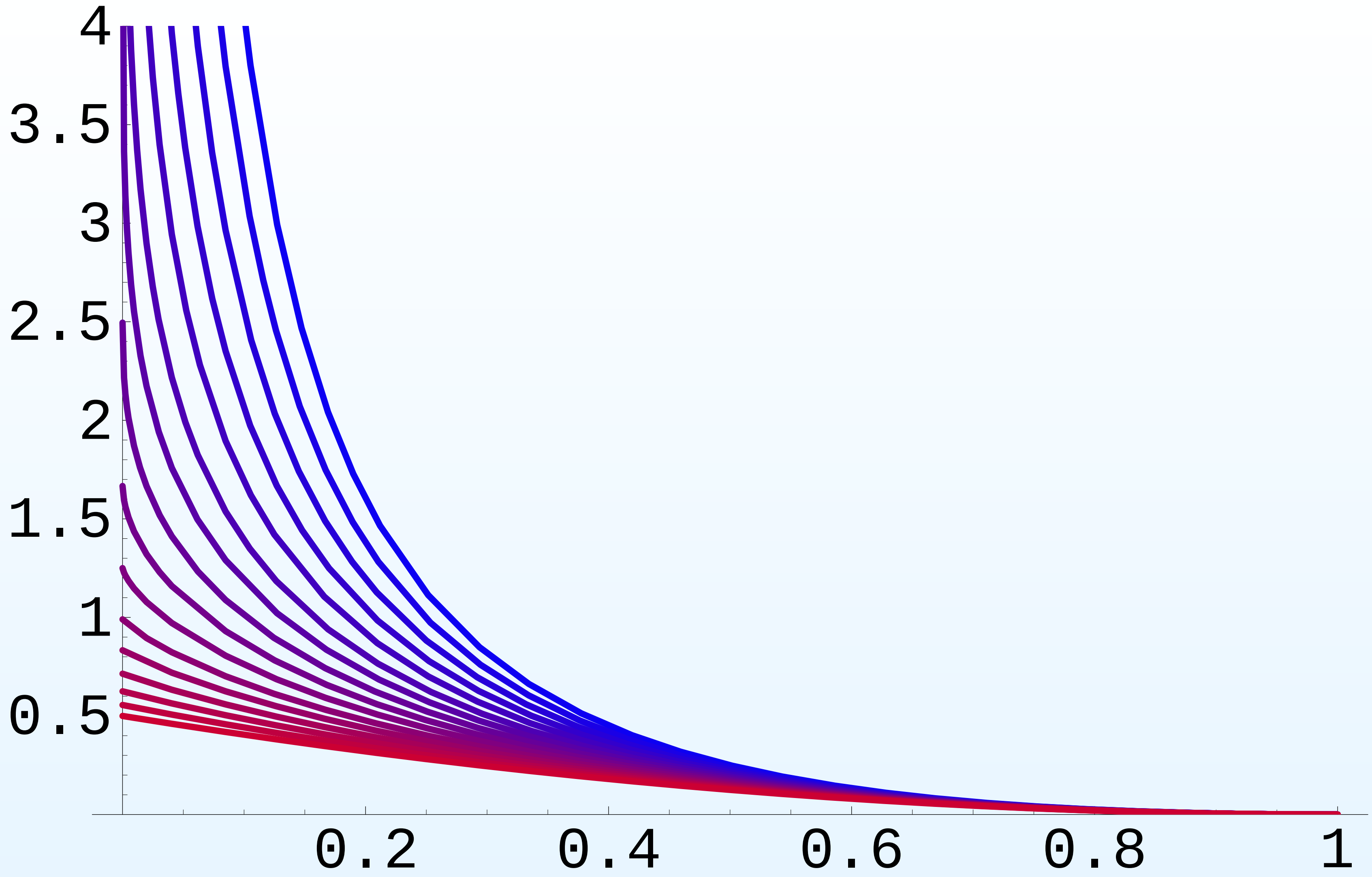


Fig. 1: $\phi_s(x)$, $s \in \{-1, -0.8, -0.6, -0.4, -0.2, 0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.8, 2.0\}$

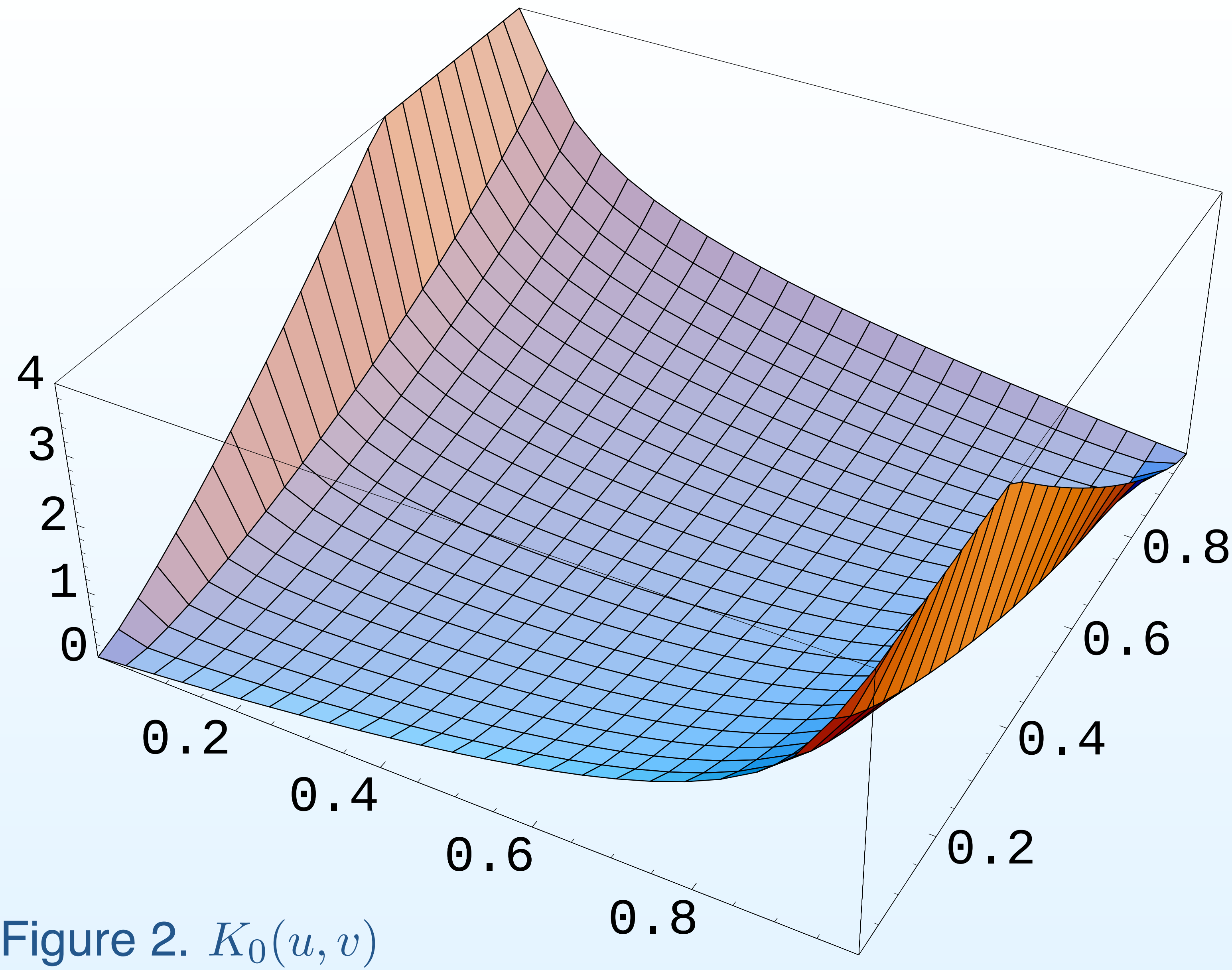


Figure 2. $K_0(u, v)$

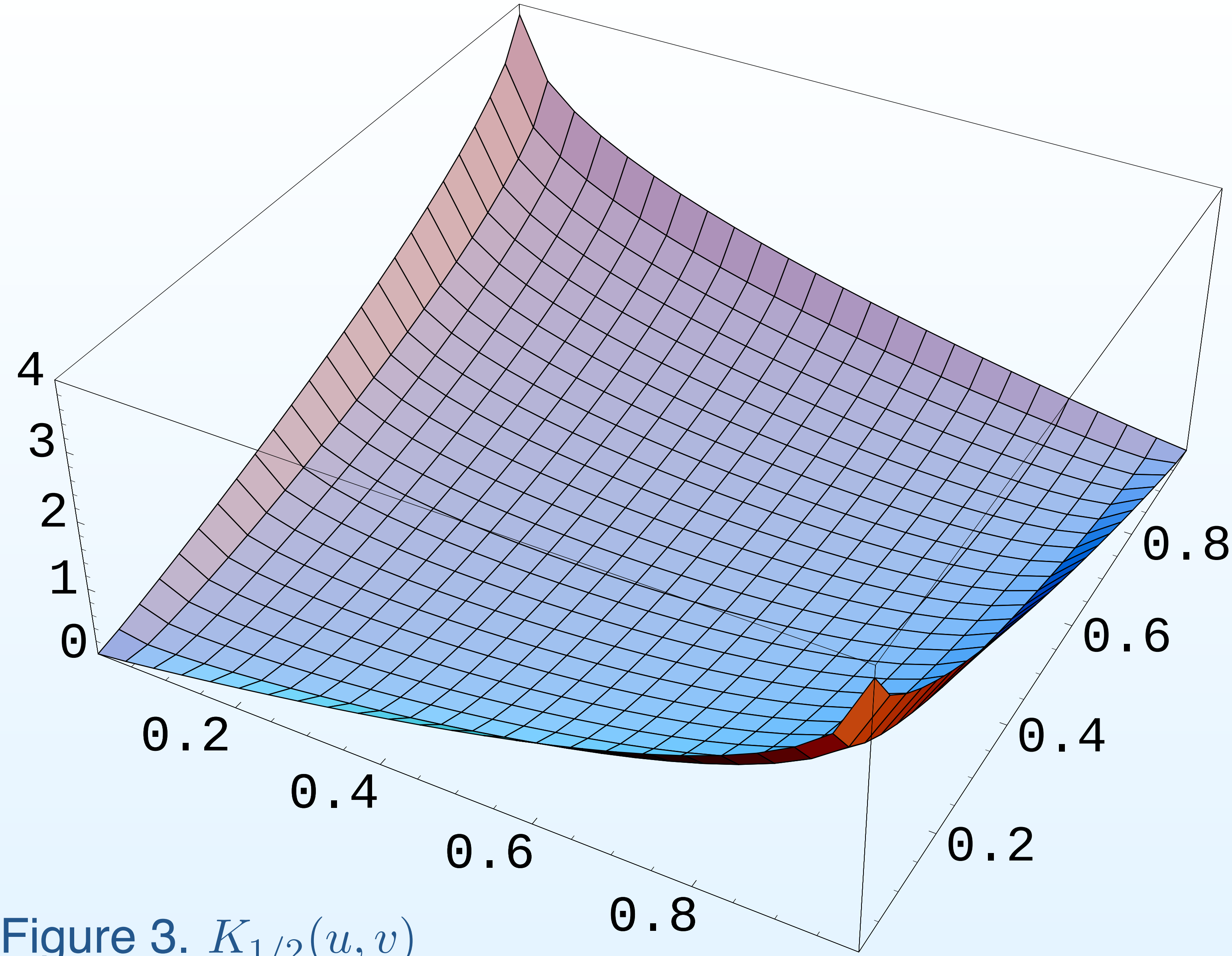


Figure 3. $K_{1/2}(u, v)$

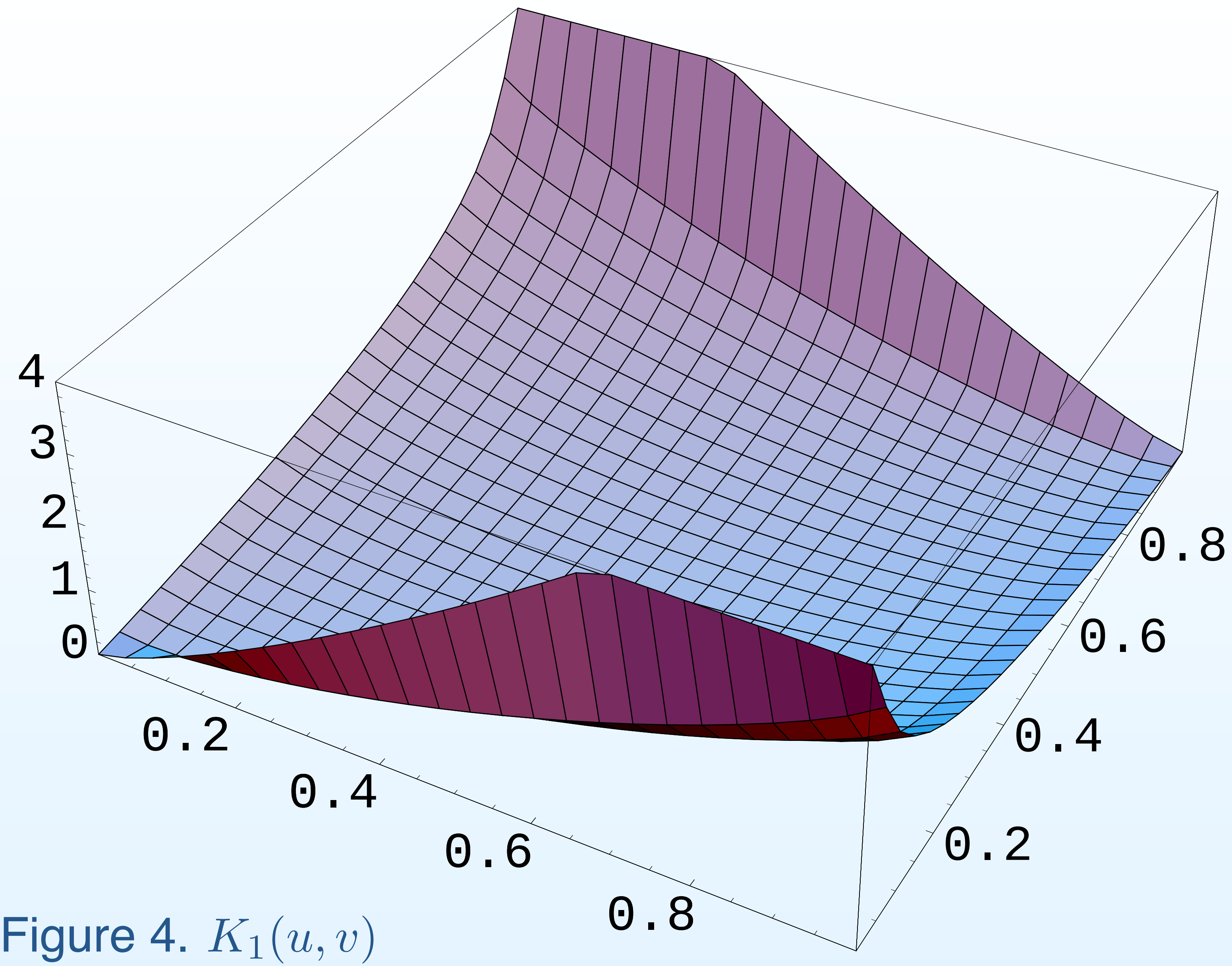


Figure 4. $K_1(u,v)$

3. Null hypothesis distribution theory

- Owen (1995) and Jager (2006):
  finite sample critical points via Noé's recursion for $n \leq 3000$

3. Null hypothesis distribution theory

- Owen (1995) and Jager (2006):
  finite sample critical points via Noé's recursion for $n \leq 3000$
- For $n \geq 3000$, asymptotic theory via Jaeschke (1979) and Eicker (1979) (cf. SW p. 597 - 615), together with

$$K_s(u, v) \approx 2^{-1} \frac{(u - v)^2}{v(1 - v)}$$

3. Null hypothesis distribution theory

- Owen (1995) and Jager (2006):
  finite sample critical points via Noé's recursion for $n \leq 3000$
- For $n \geq 3000$, asymptotic theory via Jaeschke (1979) and Eicker (1979) (cf. SW p. 597 - 615), together with

$$K_s(u,v) \approx 2^{-1} \frac{(u-v)^2}{v(1-v)}$$

- Let $r_n \equiv \log_2 n + (1/2) \log_3 n - (1/2) \log(4\pi)$.

3. Null hypothesis distribution theory

- Owen (1995) and Jager (2006):
  finite sample critical points via Noé's recursion for $n \leq 3000$
- For $n \geq 3000$, asymptotic theory via Jaeschke (1979) and Eicker (1979) (cf. SW p. 597 - 615), together with

$$K_s(u,v) \approx 2^{-1} \frac{(u-v)^2}{v(1-v)}$$

- Let $r_n \equiv \log_2 n + (1/2) \log_3 n - (1/2) \log(4\pi)$.
- **Theorem.** If $F = F_0$, the uniform distribution on $[0,1]$, then for $-1 \leq s \leq 2$

$$nS_n(s) - r_n \to_d Y_4$$

  where $P(Y_4 \leq x) = \exp(-4 \exp(-x))$.

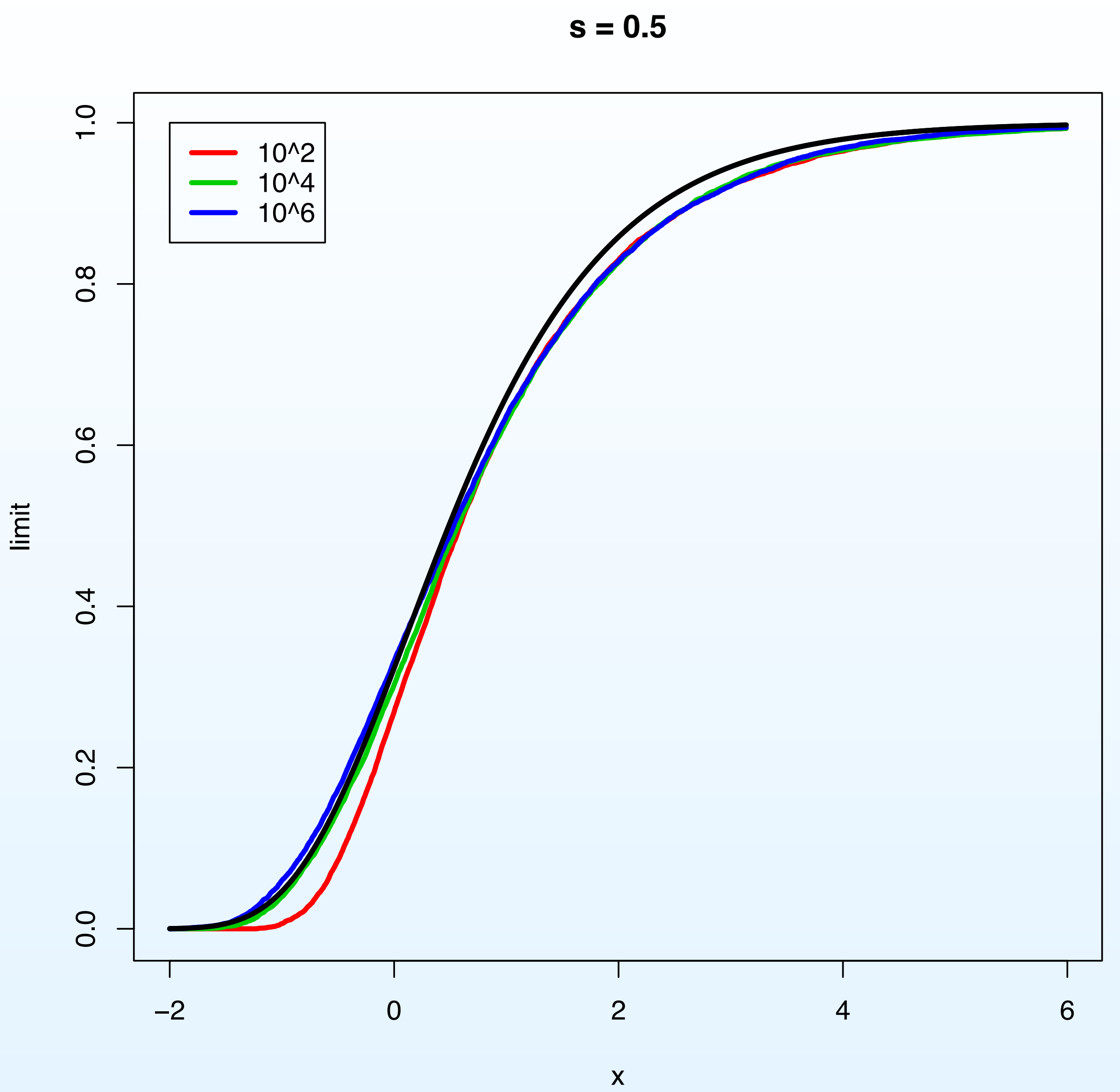


Figure 5. $P(nS_n(1/2) - r_n \leq x)$ for $n = 10^2, \; 10^4, \; 10^6$ and $P(Y_4 \leq x)$

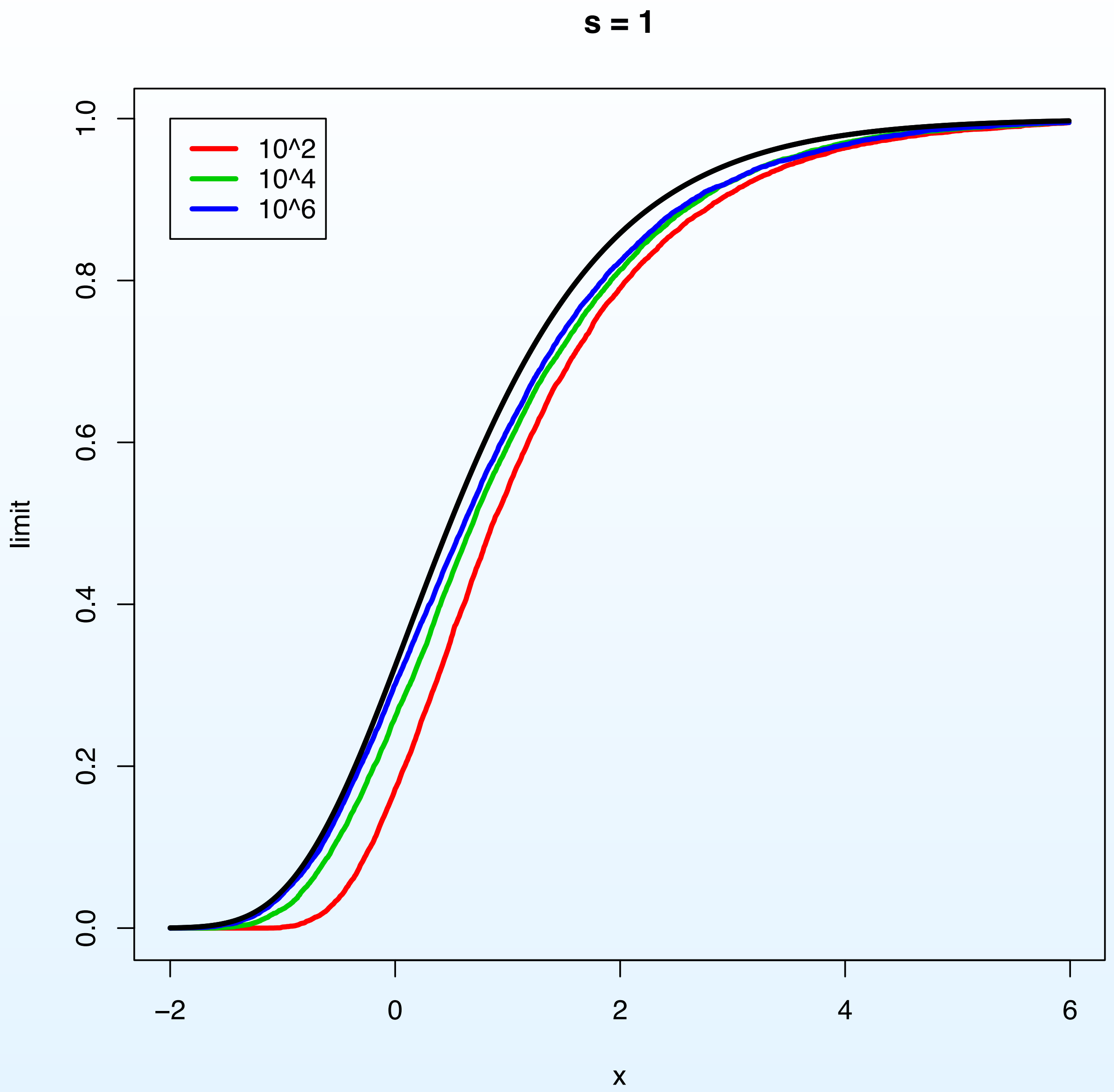


Figure 6. $P(nS_n(1) - r_n \leq x)$ for $n = 10^2$, 10^4, 10^6 and $P(Y_4 \leq x)$

- $b_n \equiv \sqrt{2 \log_2 n}, \quad c_n \equiv b_n^2 + (1/2)\{\log_3 n - \log(4\pi)\}$

- $b_n \equiv \sqrt{2 \log_2 n}, \quad c_n \equiv b_n^2 + (1/2)\{\log_3 n - \log(4\pi)\}$
- $q_n^{(1)}(\alpha) = y_{4,\alpha} + r_n$

- $b_n \equiv \sqrt{2 \log_2 n}, \quad c_n \equiv b_n^2 + (1/2)\{\log_3 n - \log(4\pi)\}$
- $q_n^{(1)}(\alpha) = y_{4,\alpha} + r_n$
- $q_n^{(2)}(\alpha) = y_{4,\alpha} + c_n^2/(2b_n^2),$

- $b_n \equiv \sqrt{2\log_2 n}, \quad c_n \equiv b_n^2 + (1/2)\{\log_3 n - \log(4\pi)\}$
- $q_n^{(1)}(\alpha) = y_{4,\alpha} + r_n$
- $q_n^{(2)}(\alpha) = y_{4,\alpha} + c_n^2/(2b_n^2),$
- $q_n^{(3)}(\alpha) = x_{\alpha,n} + c_n^2/(2b_n^2)$

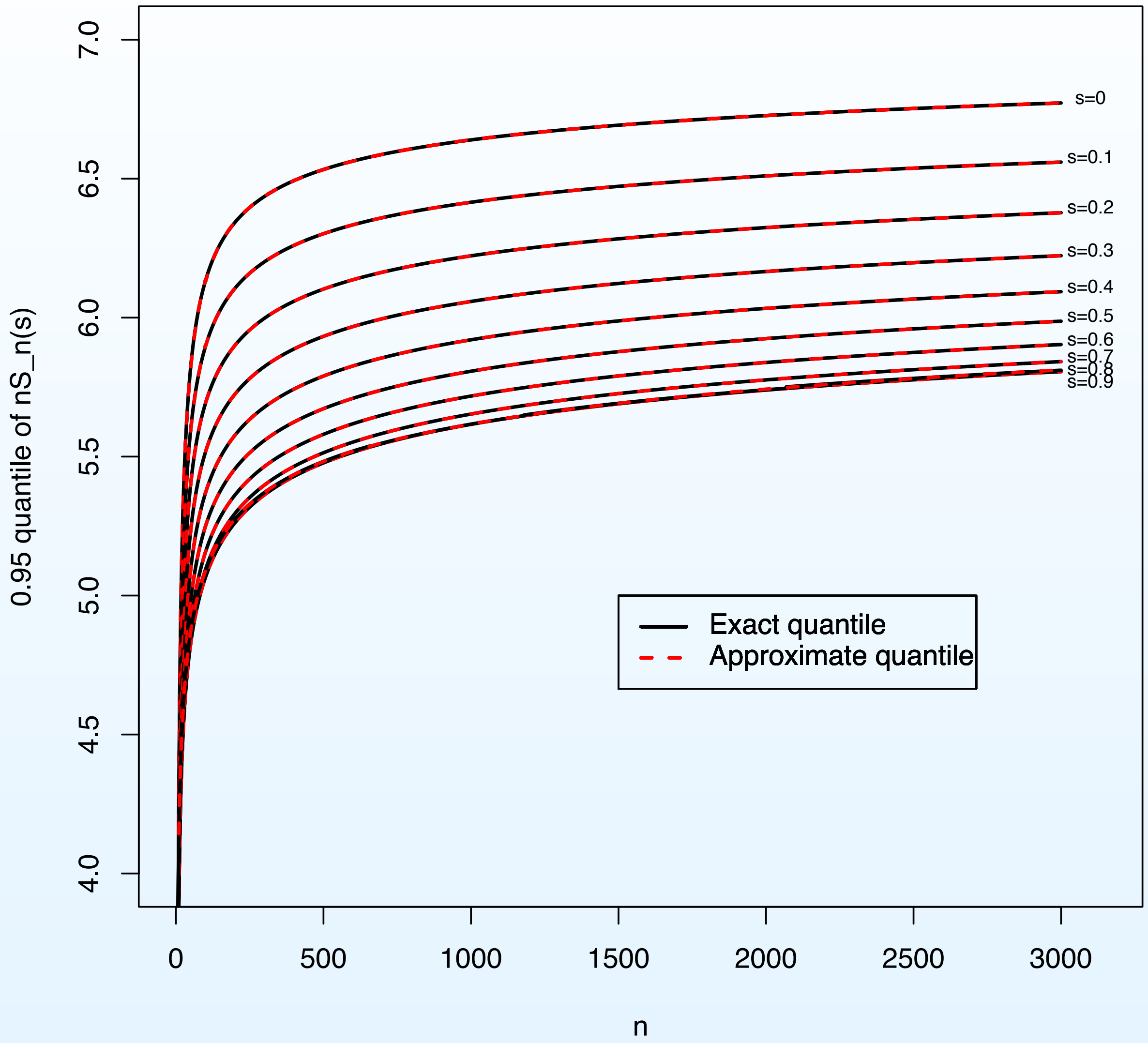


Figure 7. Exact and approximate $.95$ quantiles of $nS_n(s)$, $10 \leq n \leq 3000$.

4. Limit theory under alternatives and power

- **Theorem 1.** If $X_1, \ldots, X_n$ are i.i.d. $F \in K$ and $0 < s < 1$, then

$$S_n(s) \rightarrow_{a.s.} \sup_{0<x<1} K_s(F(x), x) \equiv S_\infty(s, F). \qquad (1)$$

- **Theorem 2.** If $X_1, \ldots, X_n$ are i.i.d. $F \in K$ and $s > 1$, then (1) holds if and only if

$$\int_0^1 \{F^{-1}(u)(1 - F^{-1}(u))\}^{(1-s)/s} du < \infty.$$

- Poisson boundary distributions

- Poisson boundary distributions
- **Theorem: (Berk & Jones, 1979).** If $F(x) = 1/(1 + \log(1/x))$, and $X_1, \ldots, X_n$ are i.i.d. F, then

$$R_n = S_n(1) \rightarrow_d 1/U \stackrel{d}{=} \sup_{t>0} \frac{\mathbb{N}(t)}{t}$$

where $U \sim U(0,1)$, $\mathbb{N}$ is a standard Poisson process.

- Poisson boundary distributions
- **Theorem: (Berk & Jones, 1979).** If $F(x) = 1/(1+\log(1/x))$, and $X_1, \ldots, X_n$ are i.i.d. F, then

$$R_n = S_n(1) \rightarrow_d 1/U \stackrel{d}{=} \sup_{t>0} \frac{\mathbb{N}(t)}{t}$$

where $U \sim U(0,1)$, $\mathbb{N}$ is a standard Poisson process.
- Generalization: let

$$F_s(x) = \begin{cases} (1 + \frac{x^{1-s}-1}{s-1})^{-1/s}, & 1 < s < \infty, \\ (1 + \log(1/x))^{-1}, & s = 1, \\ (1 - s(x^{s-1} - 1))^{1/s}, & s < 0. \end{cases}$$

- **Theorem.** (Poisson boundaries for $s \geq 1$ and $s < 0$).
  (i) Fix $s \geq 1$ and suppose that $X_1, \ldots, X_n$ are i.i.d. F_s. Then

$$S_n(s) \to_d \frac{1}{s}\left(\sup_{t>0} \frac{\mathbb{N}(t)}{t}\right)^s \stackrel{d}{=} \frac{1}{sU^s}$$

  (ii) Fix $s < 0$ and suppose that $X_1, \ldots, X_n$ are i.i.d. F_s. Then

$$S_n(s) \to_d \frac{1}{1-s}\left(\sup_{t \geq S_1} \frac{t}{\mathbb{N}(t)}\right)^{-s}$$

  where $S_1 = E_1$ is the first jump point of $\mathbb{N}$.

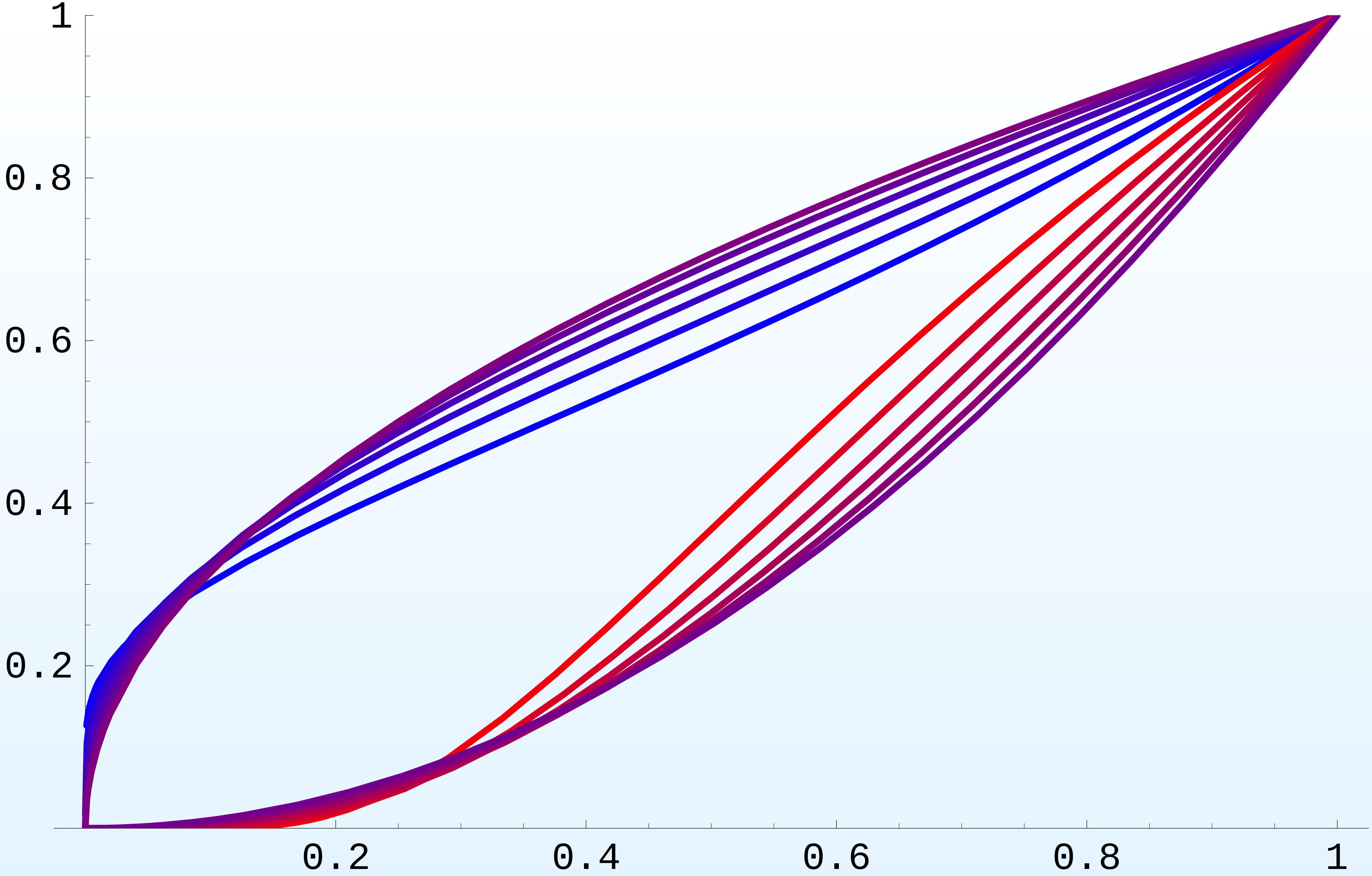


Figure 7. Poisson boundary distribution functions,
$s \in \{1, 1.2, 1.4, 1.6, 1.8, 2\} \cup \{-1, -0.8, -0.6, -0.4, -0.2, 0.0\}$.

- Ingster - Donoho - Jin testing problem

- Ingster - Donoho - Jin testing problem
- Suppose $Y_1, \ldots, Y_n$ i.i.d. G on $\mathbb{R}$

- Ingster - Donoho - Jin testing problem
- Suppose $Y_1, \ldots, Y_n$ i.i.d. G on $\mathbb{R}$
- test $H : \; G = \Phi$, the standard $N(0,1)$ d.f. versus $H_1 : G = (1-\epsilon)\Phi + \epsilon\Phi(\cdot - \mu)$, and, in particular, against

$$H_1^{(n)} : G = (1-\epsilon_n)\Phi + \epsilon_n\Phi(\cdot - \mu_n)$$

for $\epsilon_n = n^{-\beta}$, $\quad \mu_n = \sqrt{2r\log n}$
$1/2 < \beta < 1$, $0 < r < 1$.

- transform to $X_i \equiv 1 - \Phi(Y_i) \in [0, 1]$ i.i.d.

$$F = 1 - G(\Phi^{-1}(1 - \cdot)).$$

- transform to $X_i \equiv 1 - \Phi(Y_i) \in [0, 1]$ i.i.d.

$$F = 1 - G(\Phi^{-1}(1 - \cdot)).$$

- Then the testing problem becomes: test

$$H_0 : F = F_0 = U(0, 1) \qquad \text{versus}$$
$$H_1^{(n)} : F(u) = u + \epsilon_n\{(1 - u) - \Phi(\Phi^{-1}(1 - u) - \mu_n)\}$$

- transform to $X_i \equiv 1 - \Phi(Y_i) \in [0,1]$ i.i.d.

$$F = 1 - G(\Phi^{-1}(1 - \cdot)).$$

- Then the testing problem becomes: test

$$H_0 : F = F_0 = U(0,1) \qquad \text{versus}$$
$$H_1^{(n)} : F(u) = u + \epsilon_n\{(1-u) - \Phi(\Phi^{-1}(1-u) - \mu_n)\}$$

- Test statistics: Donoho-Jin: Berk-Jones $R_n = S_n(1)$ and

$$HC_n^* \equiv \sup_{X_{(1)} \le x < X_{([n/2])}} \frac{\sqrt{n}(\mathbb{F}_n(x) - x)}{\sqrt{x(1-x)}}$$
$$\equiv \text{Tukey's “higher criticism statistic”}$$

- Define the optimal detection boundary $\rho^*(\beta)$ by

$$\rho^*(\beta) = \begin{cases} \beta - 1/2, & 1/2 < \beta \leq 3/4 \\ (1 - \sqrt{1-\beta})^2, & 3/4 < \beta < 1 \end{cases}$$

- Define the optimal detection boundary $\rho^*(\beta)$ by

$$\rho^*(\beta) = \begin{cases} \beta - 1/2, & 1/2 < \beta \leq 3/4 \\ (1 - \sqrt{1-\beta})^2, & 3/4 < \beta < 1 \end{cases}$$

- **Theorem:** (Donoho - Jin, 2004). For $r > \rho^*(\beta)$ the tests based on HC_n^* and $R_n = S_n(1)$ are size and power consistent for testing H_0 versus $H_1^{(n)}$.

- Define the optimal detection boundary $\rho^*(\beta)$ by

$$\rho^*(\beta) = \begin{cases} \beta - 1/2, & 1/2 < \beta \le 3/4 \\ (1 - \sqrt{1-\beta})^2, & 3/4 < \beta < 1 \end{cases}$$

- **Theorem:** (Donoho - Jin, 2004). For $r > \rho^*(\beta)$ the tests based on HC_n^* and $R_n = S_n(1)$ are size and power consistent for testing H_0 versus $H_1^{(n)}$.
- **Theorem:** (Jager - Wellner, 2006). For $r > \rho^*(\beta)$ the tests based on $S_n(s)$ with $-1 \le s \le 2$ are size and power consistent for testing H_0 versus $H_1^{(n)}$.

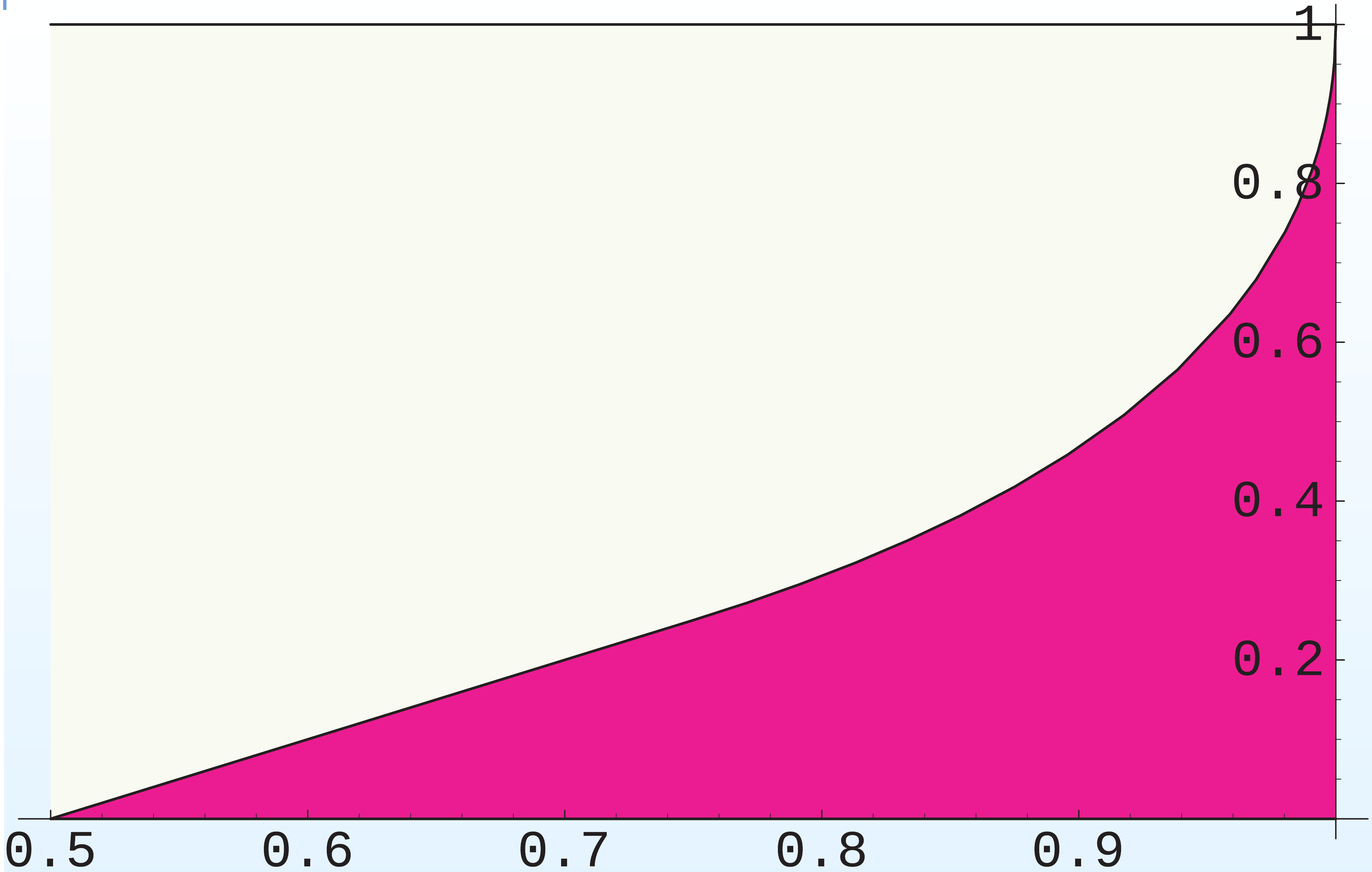


Figure 8. Detection boundary: $r > \rho^*(\beta)$ detectable

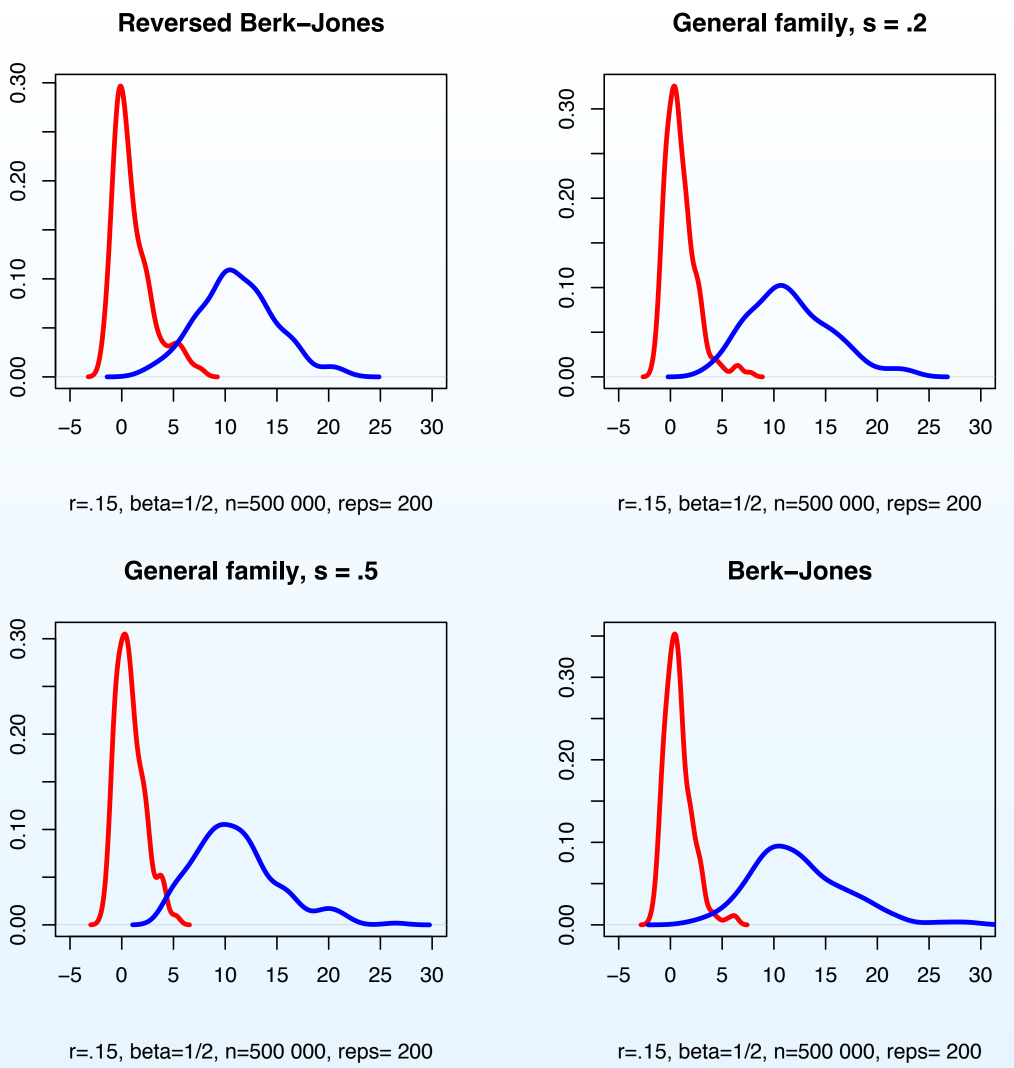


Figure 9. Separation plots: $n = 5 \times 10^5$, $r = .15$, $\beta = 1/2$
Smoothed histograms of reps $= 200$ of the statistics under the null hypothesis and the the alternative hypothesis

- $$d_n = \frac{(\log n)^5}{n} < 1/2 \qquad \text{if } \; n > 1010388 \approx 10^6$$

- $$d_n = \frac{(\log n)^5}{n} < 1/2 \qquad \text{if } \; n > 1010388 \approx 10^6$$
- Number theory; Littlewood $Li(x) - \pi(x)$ changes sign infinitely often for x large.

- $$d_n = \frac{(\log n)^5}{n} < 1/2 \qquad \text{if} \;\; n > 1010388 \approx 10^6$$
- Number theory; Littlewood $Li(x) - \pi(x)$ changes sign infinitely often for x large.
- Skewes (1933): first sign change of $Li(x) - \pi(x)$ before

$$10^{10^{10^{79}}}$$

- $$d_n = \frac{(\log n)^5}{n} < 1/2 \qquad \text{if } \; n > 1010388 \approx 10^6$$
- Number theory; Littlewood $Li(x) - \pi(x)$ changes sign infinitely often for x large.
- Skewes (1933): first sign change of $Li(x) - \pi(x)$ before

$$10^{10^{10^{79}}}$$

- Current estimate: first sign change of $Li(x) - \pi(x)$ before 10^{316}.