Testing for sparse normal means: Is there a signal?

Jon A. Wellner

University of Washington

- joint work with Leah Jager,
  U. S. Naval Academy

- Talk in the program on
  Statistical Theory and Methods for Complex,
  High-Dimensional Data

  Isaac Newton Institute for Mathematical Sciences,
  Cambridge 30 April, 2008

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

Outline

- Testing problems for normal means

Outline

- Testing problems for normal means
- Detection boundaries and Tukey's "higher criticism" statistic

Outline

- Testing problems for normal means

- Detection boundaries and Tukey's "higher criticism" statistic

- A new family of statistics via phi-divergences

Outline

- Testing problems for normal means

- Detection boundaries and Tukey's "higher criticism" statistic

- A new family of statistics via phi-divergences

- Beyond normality: generalized Gaussian distributions

Outline

- Testing problems for normal means

- Detection boundaries and Tukey's "higher criticism" statistic

- A new family of statistics via phi-divergences

- Beyond normality: generalized Gaussian distributions

- Estimating the proportion of false null hypotheses

Outline

- Testing problems for normal means

- Detection boundaries and Tukey's "higher criticism" statistic

- A new family of statistics via phi-divergences

- Beyond normality: generalized Gaussian distributions

- Estimating the proportion of false null hypotheses

- Application: the WMAP data

1. Testing problems for sparse normal means

- Initial setting: multiple testing of normal means
  For $i = 1, \ldots, n$ consider testing

$$H_{0,i} : X_i \sim N(0, 1)$$

versus

$$H_{1,i} : \ X_i \sim N(\mu_i, 1) \ \text{ with } \ \mu_i > 0.$$

1. Testing problems for sparse normal means

- Initial setting: multiple testing of normal means
  For $i = 1, \ldots, n$ consider testing

$$H_{0,i} : X_i \sim N(0, 1)$$

versus

$$H_{1,i} : \ X_i \sim N(\mu_i, 1) \ \text{ with } \ \mu_i > 0.$$

- Sparsity: proportion $\epsilon_n \equiv n^{-1}\#\{i \leq n : \ \mu_i > 0\}$ is small; $\epsilon_n \sim n^{-\beta}$ with $0 < \beta < 1$.

1. Testing problems for sparse normal means

- Initial setting: multiple testing of normal means
  For $i = 1, \ldots, n$ consider testing

$$H_{0,i} : X_i \sim N(0, 1)$$

versus

$$H_{1,i} : \ X_i \sim N(\mu_i, 1) \ \text{ with } \ \mu_i > 0.$$

- Sparsity: proportion $\epsilon_n \equiv n^{-1} \#\{i \leq n : \ \mu_i > 0\}$ is small; $\epsilon_n \sim n^{-\beta}$ with $0 < \beta < 1$.

- Three questions (in increasing order of difficulty):

1. Testing problems for sparse normal means

- Initial setting: multiple testing of normal means
  For $i = 1, \ldots, n$ consider testing

$$H_{0,i} : X_i \sim N(0, 1)$$

versus

$$H_{1,i} : \; X_i \sim N(\mu_i, 1) \; \text{ with } \; \mu_i > 0.$$

- Sparsity: proportion $\epsilon_n \equiv n^{-1} \# \{ i \le n : \; \mu_i > 0 \}$ is small; $\epsilon_n \sim n^{-\beta}$ with $0 < \beta < 1$.

- Three questions (in increasing order of difficulty):
  - Q1: Can we tell if at least one null hypothesis is false?

1. Testing problems for sparse normal means

- Initial setting: multiple testing of normal means
  For $i = 1, \ldots, n$ consider testing

$$H_{0,i} : X_i \sim N(0, 1)$$

versus

$$H_{1,i} : \ X_i \sim N(\mu_i, 1) \ \text{ with } \ \mu_i > 0.$$

- Sparsity: proportion $\epsilon_n \equiv n^{-1} \#\{i \leq n : \ \mu_i > 0\}$ is small;
  $\epsilon_n \sim n^{-\beta}$ with $0 < \beta < 1$.

- Three questions (in increasing order of difficulty):
  - Q1: Can we tell if at least one null hypothesis is false?
  - Q2: What is the proportion of false null hypotheses?

1. Testing problems for sparse normal means

- Initial setting: multiple testing of normal means
  For $i = 1, \ldots, n$ consider testing

$$H_{0,i} : X_i \sim N(0, 1)$$

versus

$$H_{1,i} : \ X_i \sim N(\mu_i, 1) \ \text{ with } \ \mu_i > 0.$$

- Sparsity: proportion $\epsilon_n \equiv n^{-1} \#\{i \leq n : \ \mu_i > 0\}$ is small; $\epsilon_n \sim n^{-\beta}$ with $0 < \beta < 1$.

- Three questions (in increasing order of difficulty):
  - Q1: Can we tell if at least one null hypothesis is false?
  - Q2: What is the proportion of false null hypotheses?
  - Q3: Which null hypotheses are false?

1. Testing problems for sparse normal means

- Initial setting: multiple testing of normal means
  For $i = 1, \ldots, n$ consider testing

$$H_{0,i} : X_i \sim N(0,1)$$

versus

$$H_{1,i} : \ X_i \sim N(\mu_i, 1) \ \text{with} \ \mu_i > 0.$$

- Sparsity: proportion $\epsilon_n \equiv n^{-1} \#\{i \leq n : \ \mu_i > 0\}$ is small;
  $\epsilon_n \sim n^{-\beta}$ with $0 < \beta < 1$.

- Three questions (in increasing order of difficulty):
  - Q1: Can we tell if at least one null hypothesis is false?
  - Q2: What is the proportion of false null hypotheses?
  - Q3: Which null hypotheses are false?

- Main focus here: Q1; partial review of work on Q2.

- Previous work: Q1: is there any signal?
    - Ingster (1997, 1999)
    - Jin (2004)
    - Donoho and Jin (2004)
    - Jager and Wellner (2007)

- Previous work: Q1: is there any signal?
  - Ingster (1997, 1999)
  - Jin (2004)
  - Donoho and Jin (2004)
  - Jager and Wellner (2007)
- Previous work: Q2: What is the proportion of non-null hypotheses?
  - Swanepoel (1999)
  - Efron, Tibshirani, Storey, and Tusher (2001)
  - Meinshausen and Rice (2006)
  - Jin and Cai (2007)

- Previous work: Q1: is there any signal?
  - Ingster (1997, 1999)
  - Jin (2004)
  - Donoho and Jin (2004)
  - Jager and Wellner (2007)
- Previous work: Q2: What is the proportion of non-null hypotheses?
  - Swanepoel (1999)
  - Efron, Tibshirani, Storey, and Tusher (2001)
  - Meinshausen and Rice (2006)
  - Jin and Cai (2007)
- Previous work: Q3: Where is the signal and how big is it?
  - Benjamini and Hochberg (1995)
  - Efron, Tibshirani, Storey, and Tusher (2001)
  - Storey, Dai, and Leek (2005)
  - Abramovich, Benjamini, Donoho, and Johnstone (2006).
  - Donoho and Jin (2006)

2. Detection boundaries and
Tukey's "higher criticism statistic

- Change of setting: Ingster - Donoho - Jin testing problem

2. Detection boundaries and

Tukey's "higher criticism statistic

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

2. Detection boundaries and

Tukey's "higher criticism statistic

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

$$H_1^{(n)} : G = (1 - \epsilon_n)N(0,1) + \epsilon_n N(\mu_n, 1).$$

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

2. Detection boundaries and

Tukey's "higher criticism statistic

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

$$H_1^{(n)} : G = (1 - \epsilon_n)N(0,1) + \epsilon_n N(\mu_n, 1).$$

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

- Let $\Phi(z) \equiv P(Z \leq z) = \int_{-\infty}^{z} (2\pi)^{-1/2} \exp(-x^2/2)dx$,
  $Z \sim N(0,1)$.

- 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

$$H_0 : F = F_0 = U(0, 1) \qquad \text{versus}$$

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

- 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

$$H_0 : F = F_0 = U(0, 1) \qquad \text{versus}$$

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

$$= (1 - \epsilon_n)u + \epsilon_n\{1 - \Phi(\Phi^{-1}(1 - u) - \mu_n)\}$$

- Test statistics: Donoho-Jin

$$HC_n^* \equiv \sup_{X_{(1)} \leq u < X_{([n/2])}} \frac{\sqrt{n}(\mathbb{F}_n(u) - u)}{\sqrt{u(1 - u)}}$$

$$\equiv \text{Tukey's "higher criticism statistic"}$$

where $\mathbb{F}_n(u) \equiv n^{-1}\sum_{i=1}^{n} 1_{[0,u]}(X_i) =$ empirical distribution function of the X_i's.

- Optimal detection boundary $\rho^*(\beta)$ defined by:

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

- Optimal detection boundary $\rho^*(\beta)$ defined 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 1:** (Donoho - Jin, 2004). For $r > \rho^*(\beta)$ the test based on HC_n^* is size and power consistent for testing H_0 versus $H_1^{(n)}$.

- Optimal detection boundary $\rho^*(\beta)$ defined 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 1:** (Donoho - Jin, 2004). For $r > \rho^*(\beta)$ the test based on HC_n^* is size and power consistent for testing H_0 versus $H_1^{(n)}$.

- With $h_n(\alpha_n) = \sqrt{2 \log \log(n)}(1 + o(1))$

$$P_{H_0}(HC_n^* > h_n(\alpha_n)) = \alpha_n \to 0, \qquad \text{and}$$
$$P_{H_1^{(n)}}(HC_n^* > h_n(\alpha_n)) \to 1, \qquad \text{as } n \to \infty.$$
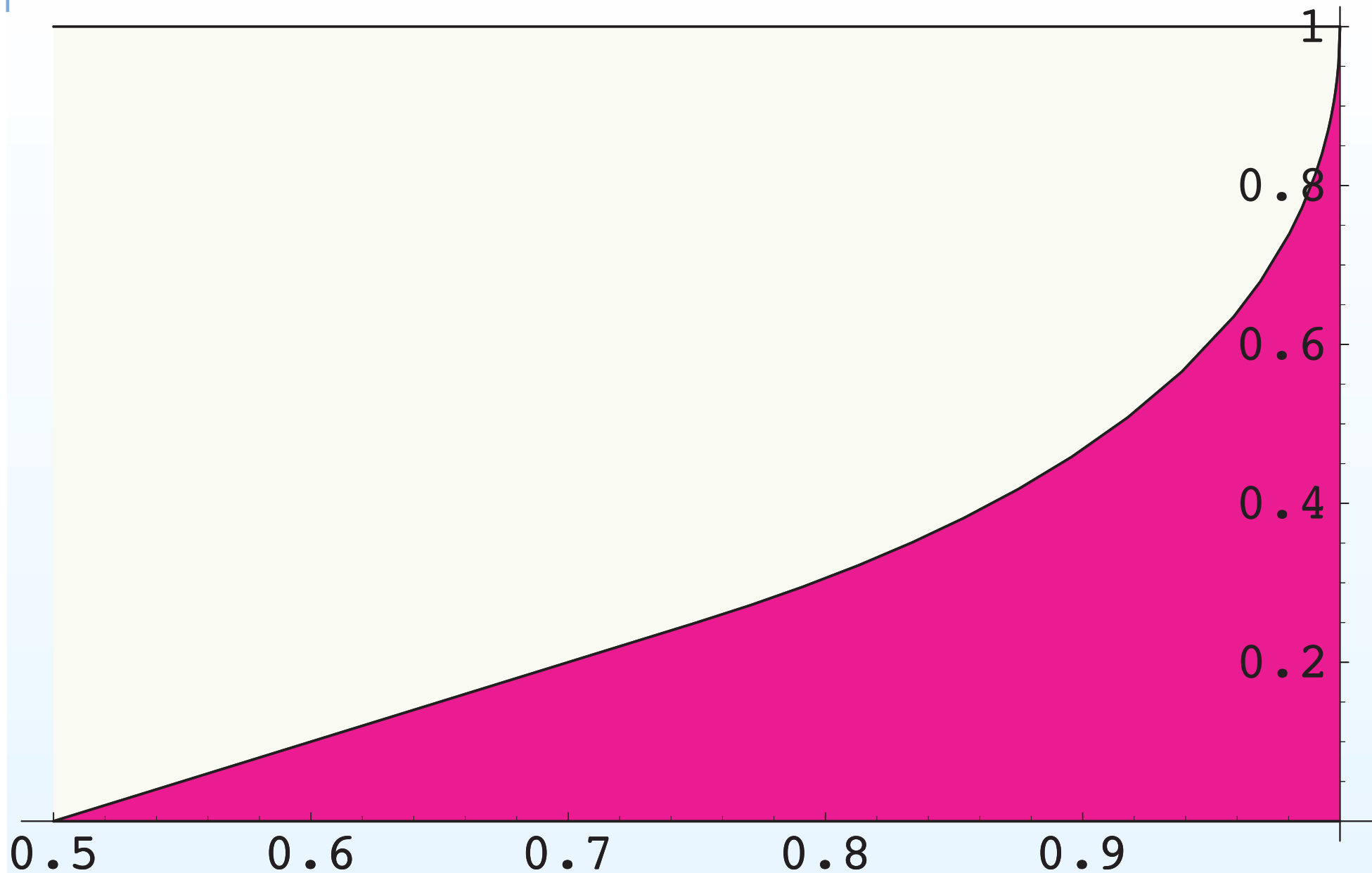
Figure 1. Detection boundary: $r > \rho^*(\beta)$ detectable

Some alternative statistics:

- Berk-Jones (1979) test statistic:

$$R_n \equiv \sup_x \log \lambda_n(x) = \sup_x K(\mathbb{F}_n(x), F_0(x)) \text{ with}$$

$$K(u, v) \equiv u \log \left(\frac{u}{v} \right) + (1 - u) \log \left(\frac{1 - u}{1 - v} \right)$$

Some alternative statistics:

- Berk-Jones (1979) test statistic:

$$R_n \equiv \sup_x \log \lambda_n(x) = \sup_x K(\mathbb{F}_n(x), F_0(x)) \ \text{ with}$$

$$K(u, v) \ \equiv \ u \log \left(\frac{u}{v} \right) + (1 - u) \log \left(\frac{1 - u}{1 - v} \right)$$

- Adaptation to one-sided $p-$value setting:

$$BJ_n^+ \equiv n \sup_{X_{(1)} \le u \le 1/2} K^+(\mathbb{F}_n(u), u)$$

where

$$K^+(u, v) \equiv \begin{cases} K(u, v), & \text{if } 0 < v < u < 1, \\ 0, & \text{if } 0 \le u \le v \le 1, \\ +\infty, & \text{otherwise}. \end{cases}$$

- **Theorem 2:** (Donoho - Jin, 2004). For $r > \rho^*(\beta)$ the test based on BJ_n^+ is size and power consistent for testing H_0 versus $H_1^{(n)}$; i.e. with $h_n(\alpha_n) = \sqrt{2 \log \log(n)}(1 + o(1))$

$$P_{H_0}(BJ_n^+ > h_n(\alpha_n)) = \alpha_n \to 0, \qquad \text{and}$$

$$P_{H_1^{(n)}}(BJ_n^+ > h_n(\alpha_n)) \to 1, \qquad \text{as } n \to \infty.$$

3. A new family of statistics via phi-divergences

A family of test statistics connecting "Higher criticism" and Berk-Jones:

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

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

3. A new family of statistics via phi-divergences

A family of test statistics connecting "Higher criticism" and Berk-Jones:

- 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:

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

s=-1.0

s=-0.5

s= 0.0

s= 0.5

s= 1.0

s= 1.5

s= 2.0
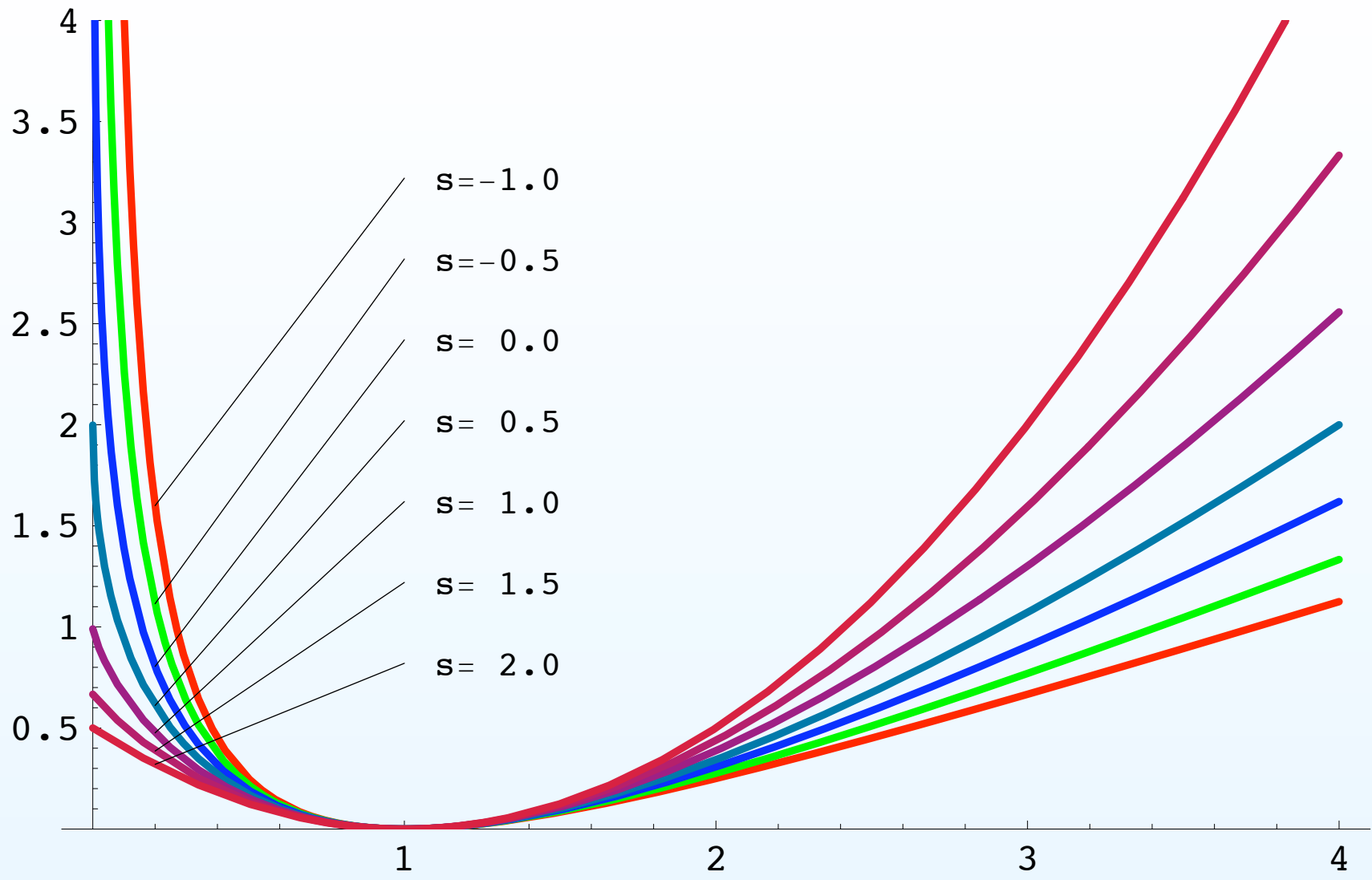
- The new family of statistics:

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

- 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 $F_0(x) = x$,

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

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

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

$$S_n(1/2)$$

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

- Version of the statistics for one-sided $p-$value setting:

$$S_n^+ \equiv n \sup_{X_{(1)} \leq u \leq 1/2} K_s^+(\mathbb{F}_n(u), u)$$

where

$$K_s^+(u, v) \equiv \begin{cases} K_s(u, v), & \text{if } 0 < v < u < 1, \\ 0, & \text{if } 0 \leq u \leq v \leq 1, \\ +\infty, & \text{otherwise.} \end{cases}$$

- Version of the statistics for one-sided $p-$value setting:

$$S_n^+ \equiv n \sup_{X_{(1)} \leq u \leq 1/2} K_s^+(\mathbb{F}_n(u), u)$$

where

$$K_s^+(u, v) \equiv \begin{cases} K_s(u, v), & \text{if } 0 < v < u < 1, \\ 0, & \text{if } 0 \leq u \leq v \leq 1, \\ +\infty, & \text{otherwise.} \end{cases}$$

- **Theorem:** (Jager - Wellner, 2007). For $r > \rho^*(\beta)$ the tests based on $S_n^+(s)$ with $-1 \leq s \leq 2$ are size and power consistent for testing H_0 versus $H_1^{(n)}$; i.e. with $s_n(\alpha_n) = \log\log(n)(1 + o(1))$

$$P_{H_0}(S_n^+ > s_n(\alpha_n)) = \alpha_n \to 0, \quad \text{and}$$
$$P_{H_1^{(n)}}(S_n^+ > s_n(\alpha_n)) \to 1, \quad \text{as } n \to \infty.$$
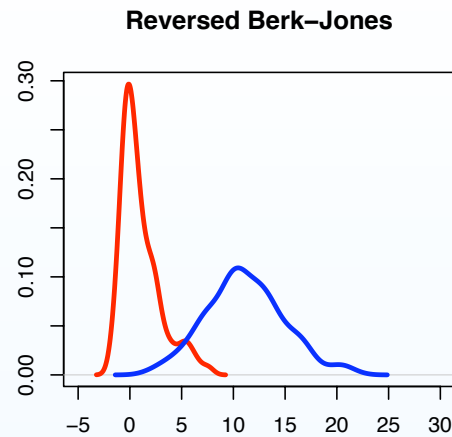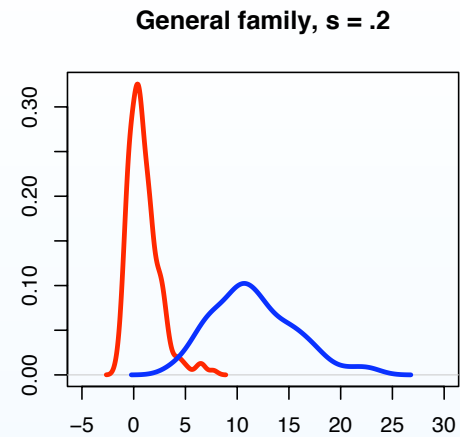
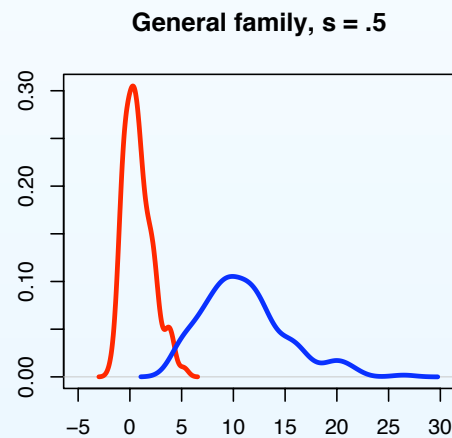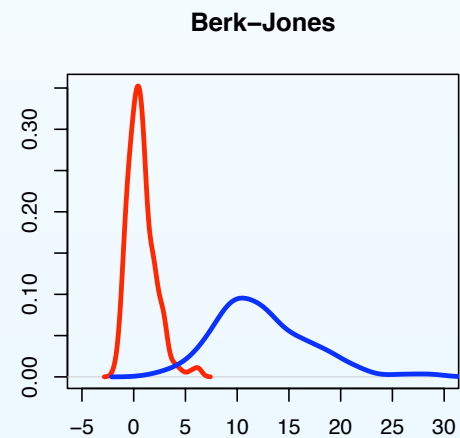Figure 2. Separation plots: $n = 5 \times 10^5$, $r = .15$, $\beta = 1/2$
Smoothed histograms of reps $= 200$ of the statistics under the
null hypothesis and the the alternative hypothesis

4. Beyond normality:

generalized Gaussian distributions, ...

- Donoho and Jin (2004) also computed detection boundaries for sparse mixtures of "Generalized Gaussian" or Subbotin distributions: $X \sim GN_\gamma(\mu)$ has density function

$$f_{\gamma,\mu}(x) = \frac{1}{C_\gamma} \exp\left(-\frac{|x - \mu|^\gamma}{\gamma}\right), \quad C_\gamma = 2\Gamma(1/\gamma)\gamma^{1/\gamma - 1}.$$

4. Beyond normality:

generalized Gaussian distributions, ...

- Donoho and Jin (2004) also computed detection boundaries for sparse mixtures of "Generalized Gaussian" or Subbotin distributions: $X \sim GN_\gamma(\mu)$ has density function

$$f_{\gamma,\mu}(x) = \frac{1}{C_\gamma} \exp\left(-\frac{|x-\mu|^\gamma}{\gamma}\right), \quad C_\gamma = 2\Gamma(1/\gamma)\gamma^{1/\gamma - 1}.$$

- Suppose $Y_1, \ldots, Y_n$ i.i.d. G on $\mathbb{R}$.

4. Beyond normality:

generalized Gaussian distributions, ...

- Donoho and Jin (2004) also computed detection boundaries for sparse mixtures of "Generalized Gaussian" or Subbotin distributions: $X \sim GN_\gamma(\mu)$ has density function

$$f_{\gamma,\mu}(x) = \frac{1}{C_\gamma} \exp\left(-\frac{|x-\mu|^\gamma}{\gamma} \right), \quad C_\gamma = 2\Gamma(1/\gamma)\gamma^{1/\gamma - 1}.$$

- Suppose $Y_1, \ldots, Y_n$ i.i.d. G on $\mathbb{R}$.
- Test $H_0 : G = GN_\gamma(0)$ versus
$H_1^{(n)} : G = (1 - \epsilon_n)GN_\gamma(0) + \epsilon_n GN_\gamma(\mu_n)$ where

$$\epsilon_n = n^{-\beta}, \qquad \mu_{\gamma,n} = (\gamma r \log n)^{1/\gamma},$$

where $1/2 < \beta < 1, \ \ 0 < r < 1$.

- Detection boundary for $1 < \gamma \leq 2$:

$$\rho_\gamma^*(\beta) = \begin{cases} (2^{1/(\gamma-1)} - 1)^{\gamma-1}(\beta - 1/2), & 1/2 < \beta \leq 1 - 2^{-\gamma/(\gamma-1)}, \\ (1 - (1 - \beta)^{1/\gamma})^\gamma, & 1 - 2^{-\gamma/(\gamma-1)} \leq \beta < 1. \end{cases}$$

- Detection boundary for $1 < \gamma \leq 2$:

$$\rho_\gamma^*(\beta) = \begin{cases} (2^{1/(\gamma-1)} - 1)^{\gamma-1}(\beta - 1/2), & 1/2 < \beta \leq 1 - 2^{-\gamma/(\gamma-1)}, \\ (1 - (1-\beta)^{1/\gamma})^\gamma, & 1 - 2^{-\gamma/(\gamma-1)} \leq \beta < 1. \end{cases}$$

- Detection boundary for $0 < \gamma \leq 1$:

$$\rho_\gamma^*(\beta) = 2(\beta - 1/2), \qquad 1/2 < \beta < 1.$$

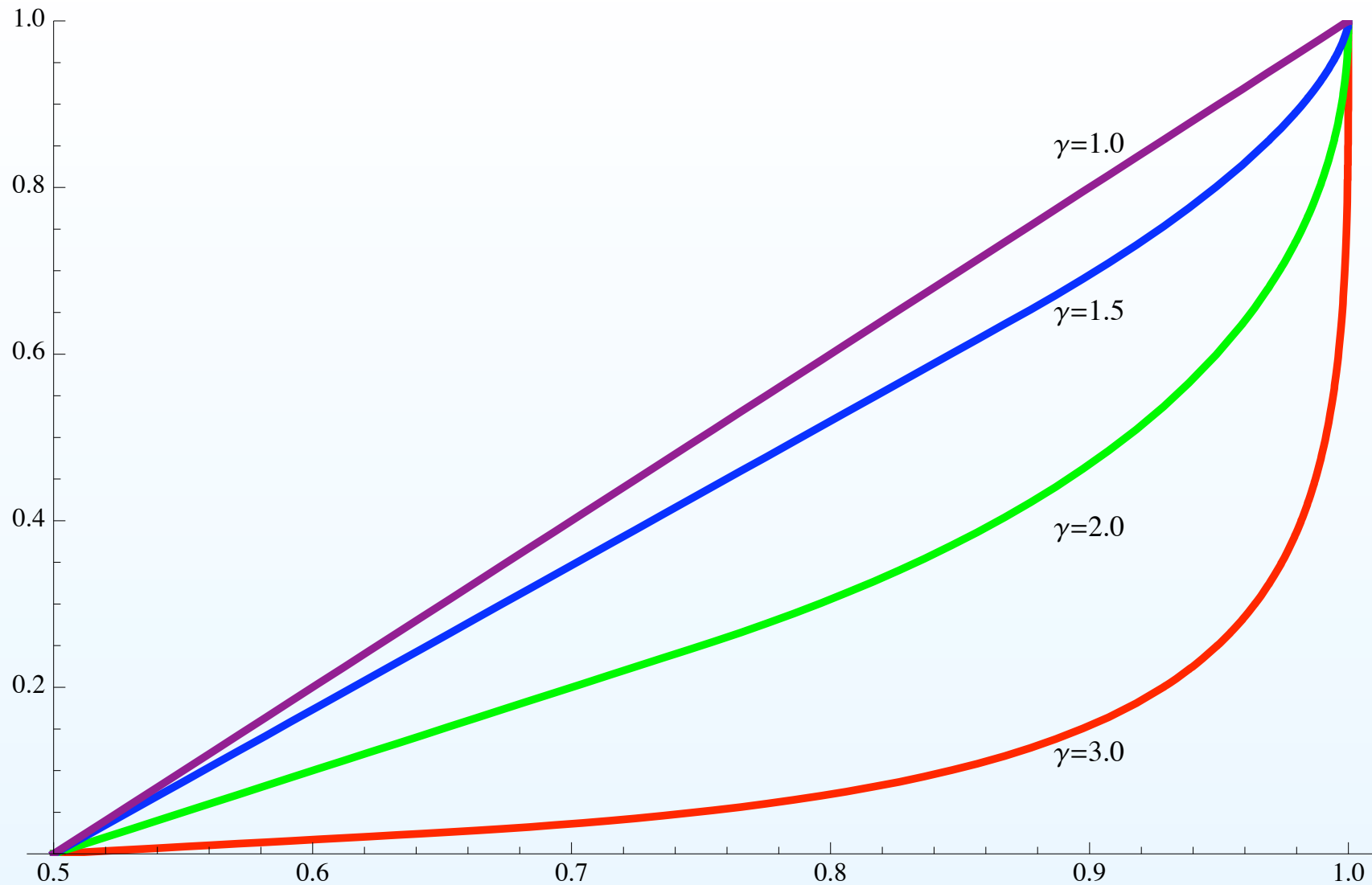Note: The detection boundary is the same for all for $0 < \gamma \leq 1$!

Figure 3. Detection boundaries for GN testing problem, $\gamma \in \{1, 1.5, 2, 3\}$.

- **Theorem:** (Donoho - Jin, 2004). For the higher criticism test statistic applied to the p-values $p_i \equiv P(GN_\gamma(0) > Y_i)$, $i = 1, \ldots, n$. Then the detection boundary $\rho_{HC,\gamma}$ for this procedure is the same as the efficient detection boundary:

$$\rho_{HC,\gamma}(\beta) = \rho_\gamma^*(\beta), \qquad 1/2 < \beta < 1.$$

- **Theorem:** (Donoho - Jin, 2004). For the higher criticism test statistic applied to the p-values $p_i \equiv P(GN_\gamma(0) > Y_i)$, $i = 1, \ldots, n$. Then the detection boundary $\rho_{HC,\gamma}$ for this procedure is the same as the efficient detection boundary:

$$\rho_{HC,\gamma}(\beta) = \rho_\gamma^*(\beta), \qquad 1/2 < \beta < 1.$$

- Similar theorem for χ_ν^2 mixtures.

5. Estimating the proportion of false null hypotheses

- Meinshausen and Rice (2006): Assume
  $Y_i \sim (1 - \epsilon_n)N(0,1) + \epsilon_n F$, F arbitrary.
  $\epsilon_n = n^{-\beta}$, $1/2 < \beta < 1$.
  M & R construct $\hat{\epsilon}_n^{MR}$ such that

$$P_{\epsilon_n, F}(\epsilon_n \geq \hat{\epsilon}_n^{MR}) \geq 1 - \alpha$$

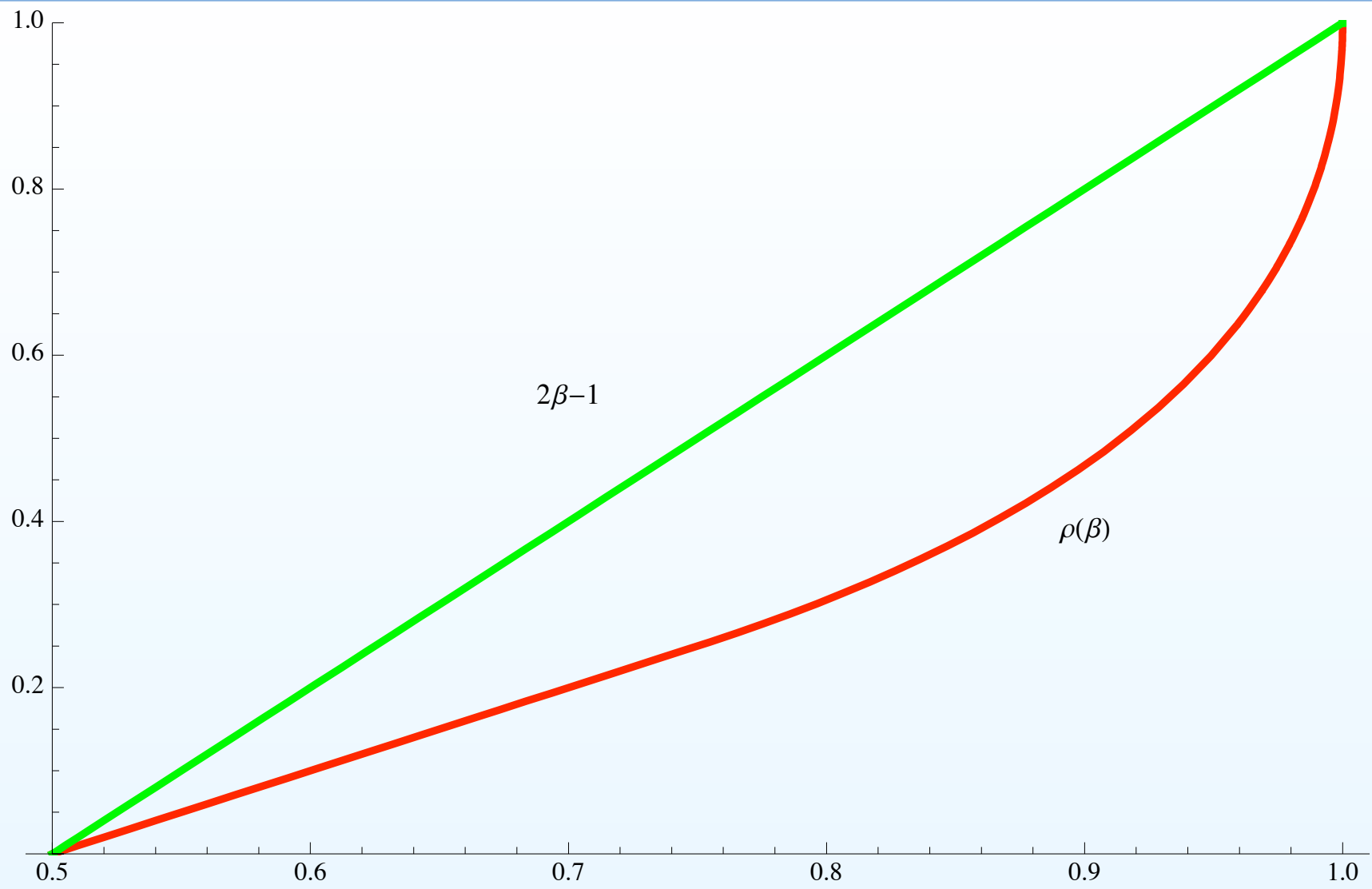# 5. Estimating the proportion of false null hypotheses

- Meinshausen and Rice (2006): Assume
  $Y_i \sim (1 - \epsilon_n)N(0, 1) + \epsilon_n F$, F arbitrary.
  $\epsilon_n = n^{-\beta}$, $1/2 < \beta < 1$.
  M & R construct $\hat{\epsilon}_n^{MR}$ such that

$$P_{\epsilon_n, F}(\epsilon_n \geq \hat{\epsilon}_n^{MR}) \geq 1 - \alpha$$

- When $F = N(\mu_n, 1)$, $\mu_n = \sqrt{2r \log n}$, then if $r > 2\beta - 1$,

$$P_{\epsilon_n, \mu_n} \left(\left| \frac{\hat{\epsilon}_n^{MR}}{\epsilon_n} - 1 \right| > \delta \right) \to 0$$

  for every $\delta > 0$.

$2\beta-1$

$\rho(\beta)$

- Cai, Jin, and Low (2007): Assume
  $Y_i \sim (1 - \epsilon_n)N(0, 1) + \epsilon_n N(\mu_n, 1)$,
  $\epsilon_n = n^{-\beta}$, $\mu_n = \sqrt{2r \log n}$, with $1/2 < \beta < 1$, $0 < r < 1$.
  Cai, Jin, and Low construct $\hat{\epsilon}_n^{CJL}$ such that

$$P_{\epsilon_n, \mu_n}(\epsilon_n \geq \hat{\epsilon}_n^{CJL}) \geq 1 - \alpha$$

- Cai, Jin, and Low (2007): Assume
$Y_i \sim (1 - \epsilon_n)N(0,1) + \epsilon_n N(\mu_n, 1)$,
$\epsilon_n = n^{-\beta}$, $\mu_n = \sqrt{2r \log n}$, with $1/2 < \beta < 1$, $0 < r < 1$.
Cai, Jin, and Low construct $\hat{\epsilon}_n^{CJL}$ such that

$$P_{\epsilon_n, \mu_n}(\epsilon_n \geq \hat{\epsilon}_n^{CJL}) \geq 1 - \alpha$$

- For any closed set Ω in the interior of $\{(\beta, r): \ r > \rho^*(\beta)\}$,

$$\sup_{(\beta, r) \in \Omega} P_{\epsilon_n, \mu_n} \left(\left| \frac{\hat{\epsilon}_n^{CJL}}{\epsilon_n} - 1 \right| > \delta \right) \to 0$$

for every $\delta > 0$.

- Cai, Jin, and Low (2007): Assume
$Y_i \sim (1 - \epsilon_n)N(0,1) + \epsilon_n N(\mu_n, 1)$,
$\epsilon_n = n^{-\beta}$, $\mu_n = \sqrt{2r \log n}$, with $1/2 < \beta < 1$, $0 < r < 1$.
Cai, Jin, and Low construct $\hat{\epsilon}_n^{CJL}$ such that

$$P_{\epsilon_n, \mu_n}(\epsilon_n \geq \hat{\epsilon}_n^{CJL}) \geq 1 - \alpha$$

- For any closed set Ω in the interior of $\{(\beta, r) : r > \rho^*(\beta)\}$,

$$\sup_{(\beta, r) \in \Omega} P_{\epsilon_n, \mu_n}\left(\left|\frac{\hat{\epsilon}_n^{CJL}}{\epsilon_n} - 1\right| > \delta\right) \to 0$$

for every $\delta > 0$.
- Moreover ...

-
$$E_{\epsilon_n,\mu_n} \left(\frac{\hat{\epsilon}_n^{CJL}}{\epsilon_n} - 1 \right)^2 \leq C_n(\beta, r)$$

for $C_n(\beta, r) \to 0$ at the optimal rate up to powers of $\log n$.

6. Application: non-Gaussian-ness in the WMAP data?

The Cosmic Microwave Background radiation, CMB
also CMBR, or CBR, or MBR
or Relic Radiation

- Predicted by George Gamow, Ralph Alpher and Robert Herman in 1948.

6. Application: non-Gaussian-ness in the WMAP data?

The **Cosmic Microwave Background radiation**, **CMB**
also CMBR, or CBR, or MBR
or Relic Radiation

- Predicted by George Gamow, Ralph Alpher and Robert Herman in 1948.

- CMB detectability suggested by Doroshkevich and Novikov, 1964.

6. Application: non-Gaussian-ness in the WMAP data?

The Cosmic Microwave Background radiation, CMB
also CMBR, or CBR, or MBR
or Relic Radiation

- Predicted by George Gamow, Ralph Alpher and Robert Herman in 1948.

- CMB detectability suggested by Doroshkevich and Novikov, 1964.

- CMB observed by Penzias and Wilson, 1964-1965 (scooping D. T. Wilkinson and P. Roll).

6. Application: non-Gaussian-ness in the WMAP data?

The Cosmic Microwave Background radiation, CMB
also CMBR, or CBR, or MBR
or Relic Radiation

- Predicted by George Gamow, Ralph Alpher and Robert Herman in 1948.

- CMB detectability suggested by Doroshkevich and Novikov, 1964.

- CMB observed by Penzias and Wilson, 1964-1965 (scooping D. T. Wilkinson and P. Roll).

- 1983: RELIKT-1 Soviet CMB anistropy experiment launched.

6. Application: non-Gaussian-ness in the WMAP data?

The Cosmic Microwave Background radiation, CMB
also CMBR, or CBR, or MBR
or Relic Radiation

- Predicted by George Gamow, Ralph Alpher and Robert Herman in 1948.

- CMB detectability suggested by Doroshkevich and Novikov, 1964.

- CMB observed by Penzias and Wilson, 1964-1965 (scooping D. T. Wilkinson and P. Roll).

- 1983: RELIKT-1 Soviet CMB anistropy experiment launched.

- 1989: COBE NASA Satellite launched: 7° angular resolution 1992: first anisotropies detected by COBE

- 1995: Wilkinson Microwave Anisotropy probe proposed.

- 1995: Wilkinson Microwave Anisotropy probe proposed.

- 2001: WMAP satellite launched: $.2^o \approx 15'$ angular resolution

- 1995: Wilkinson Microwave Anisotropy probe proposed.
- 2001: WMAP satellite launched: $.2^o \approx 15'$ angular resolution
- 2003: second acoustic peak confirmed

- 1995: Wilkinson Microwave Anisotropy probe proposed.

- 2001: WMAP satellite launched: $.2^o \approx 15'$ angular resolution

- 2003: second acoustic peak confirmed

- 2008: 5 year report issued on 5 March; third peak more precisely located.

- 1995: Wilkinson Microwave Anisotropy probe proposed.

- 2001: WMAP satellite launched: $.2^o \approx 15'$ angular resolution

- 2003: second acoustic peak confirmed

- 2008: 5 year report issued on 5 March; third peak more precisely located.

- 2008: European Space Agency Planck CMB probe to be launched on 31 October: $5'$ angular resolution.
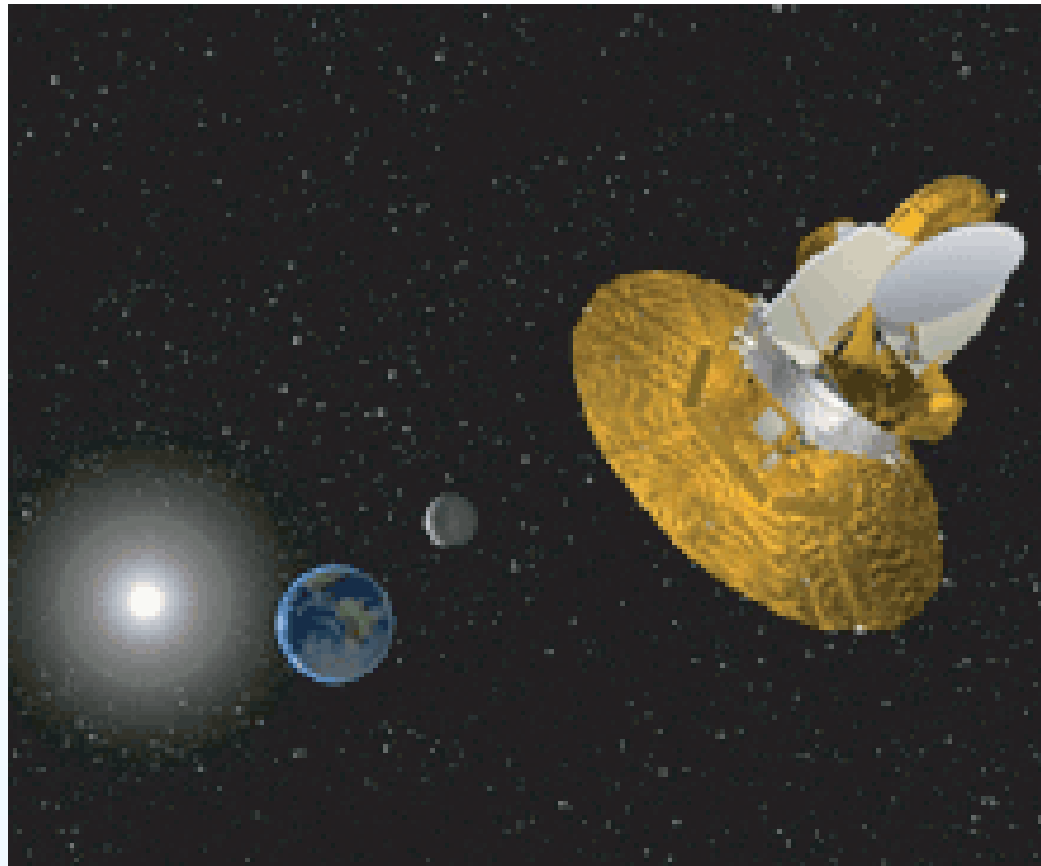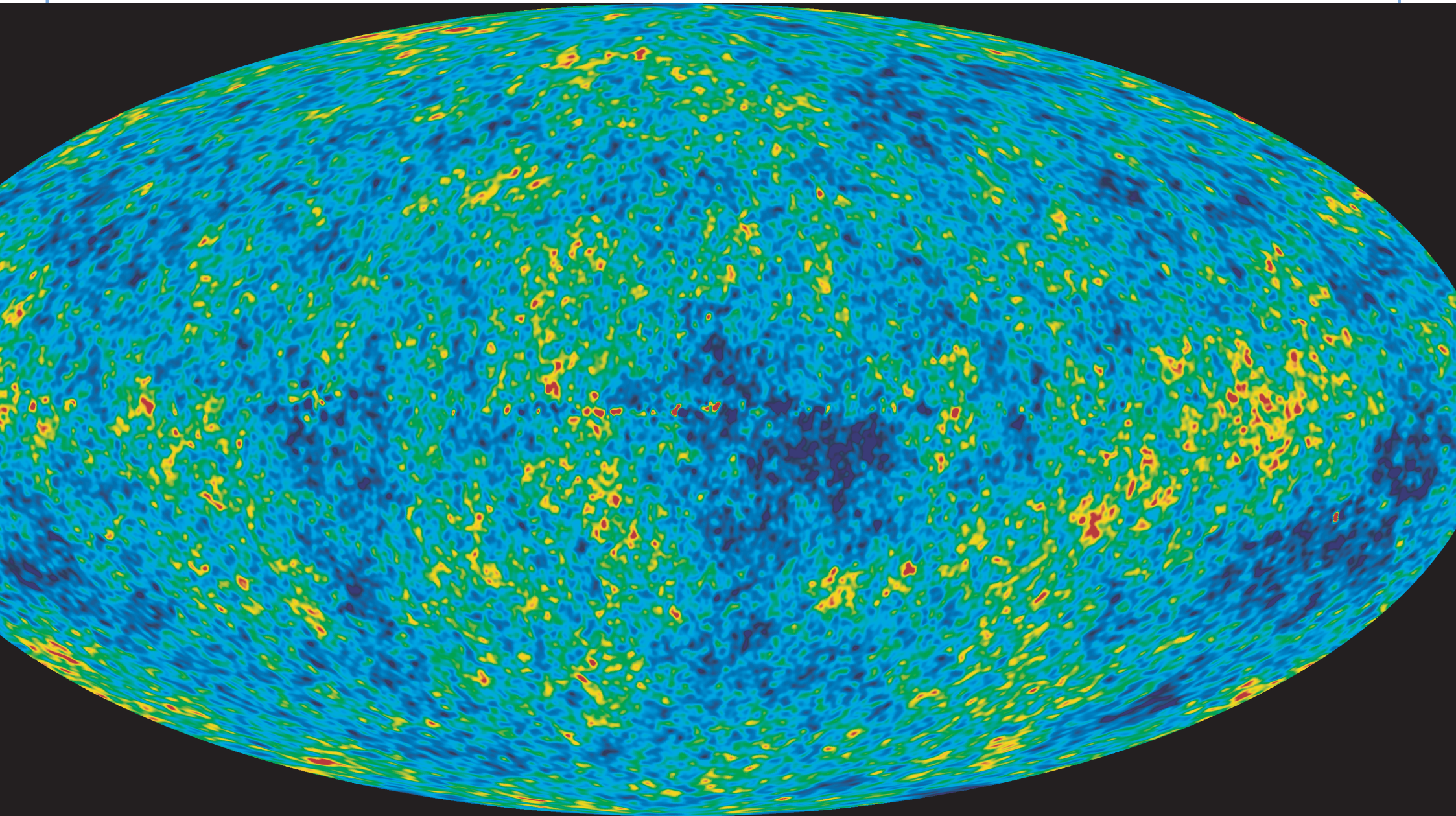
Conclusions from WMAP data:

- Average temperature of CMB: 2.275 Kelvin (-270 C or -455 F): red; warmer by $\approx 0.0002^o$; blue; colder by $\approx 0.0002^o$ K.

Conclusions from WMAP data:

- Average temperature of CMB: 2.275 Kelvin (-270 C or -455 F): red; warmer by $\approx 0.0002^o$; blue; colder by $\approx 0.0002^o$ K.

- Age of universe is: 13.73 ± 0.12 billion years old.

Conclusions from WMAP data:

- Average temperature of CMB: 2.275 Kelvin (-270 C or -455 F): red; warmer by $\approx 0.0002^o$; blue; colder by $\approx 0.0002^o$ K.

- Age of universe is: 13.73 ± 0.12 billion years old.

- Composition of universe:
  - 4.6% ordinary (baryonic) matter.
  - 23% unknown type of dark matter (does not emit or absorb light)
 - 72% mysterious dark energy (acts to accelerate expansion)
 - $< 1\%$ neutrinos.

Conclusions from WMAP data:

- Average temperature of CMB: 2.275 Kelvin (-270 C or -455 F): red; warmer by $\approx 0.0002^o$; blue; colder by $\approx 0.0002^o$ K.

- Age of universe is: 13.73 ± 0.12 billion years old.

- Composition of universe:
  - 4.6% ordinary (baryonic) matter.
  - 23% unknown type of dark matter (does not emit or absorb light)
  - 72% mysterious dark energy (acts to accelerate expansion)
  - $< 1\%$ neutrinos.

- Geometry: "flat"

Conclusions from WMAP data:

- Average temperature of CMB: 2.275 Kelvin (-270 C or -455 F): red; warmer by $\approx 0.0002^o$; blue; colder by $\approx 0.0002^o$ K.

- Age of universe is: 13.73 ± 0.12 billion years old.

- Composition of universe:
  - 4.6% ordinary (baryonic) matter.
  - 23% unknown type of dark matter (does not emit or absorb light)
  - 72% mysterious dark energy (acts to accelerate expansion)
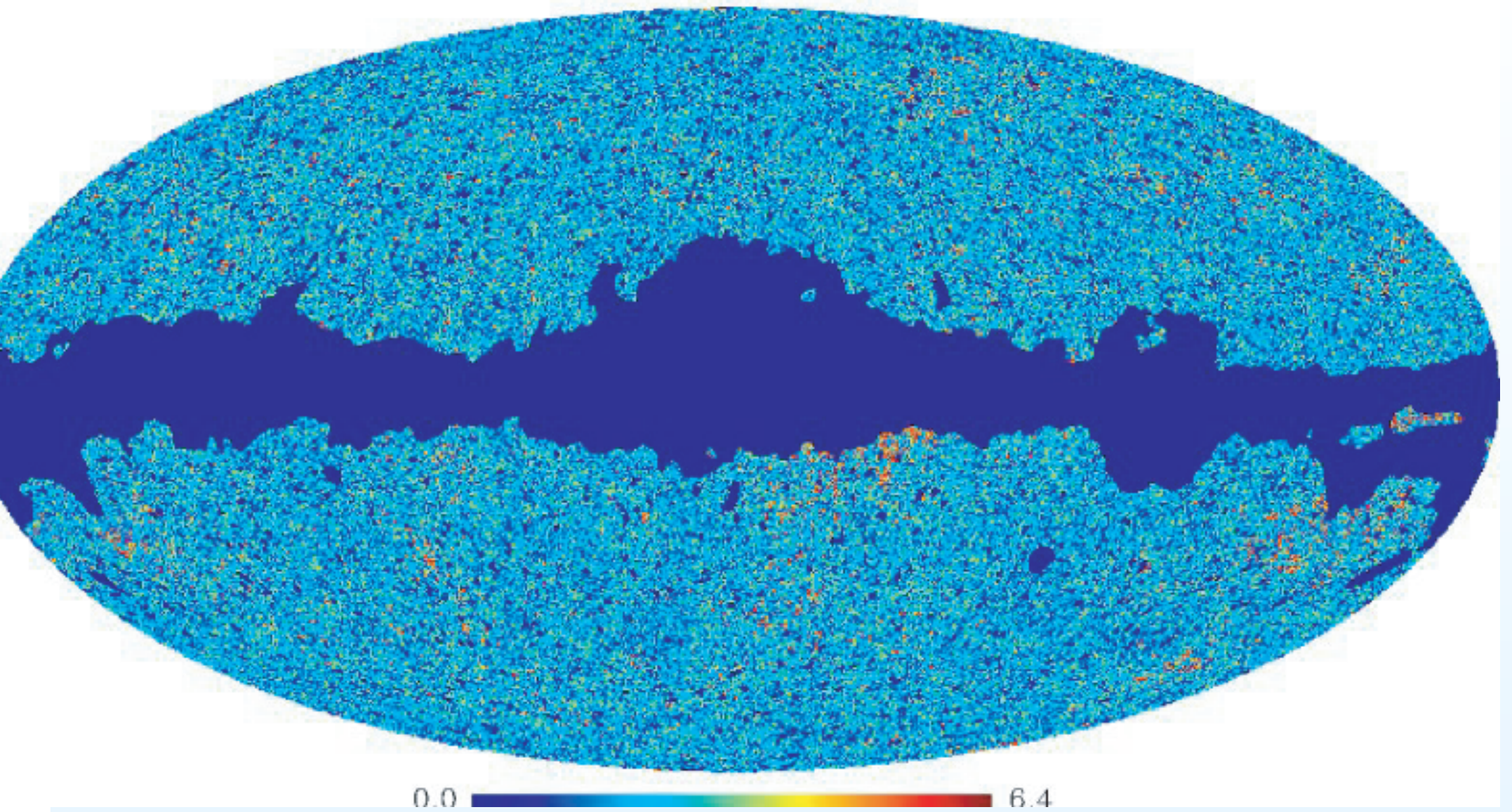 - $< 1\%$ neutrinos.

- Geometry: "flat"
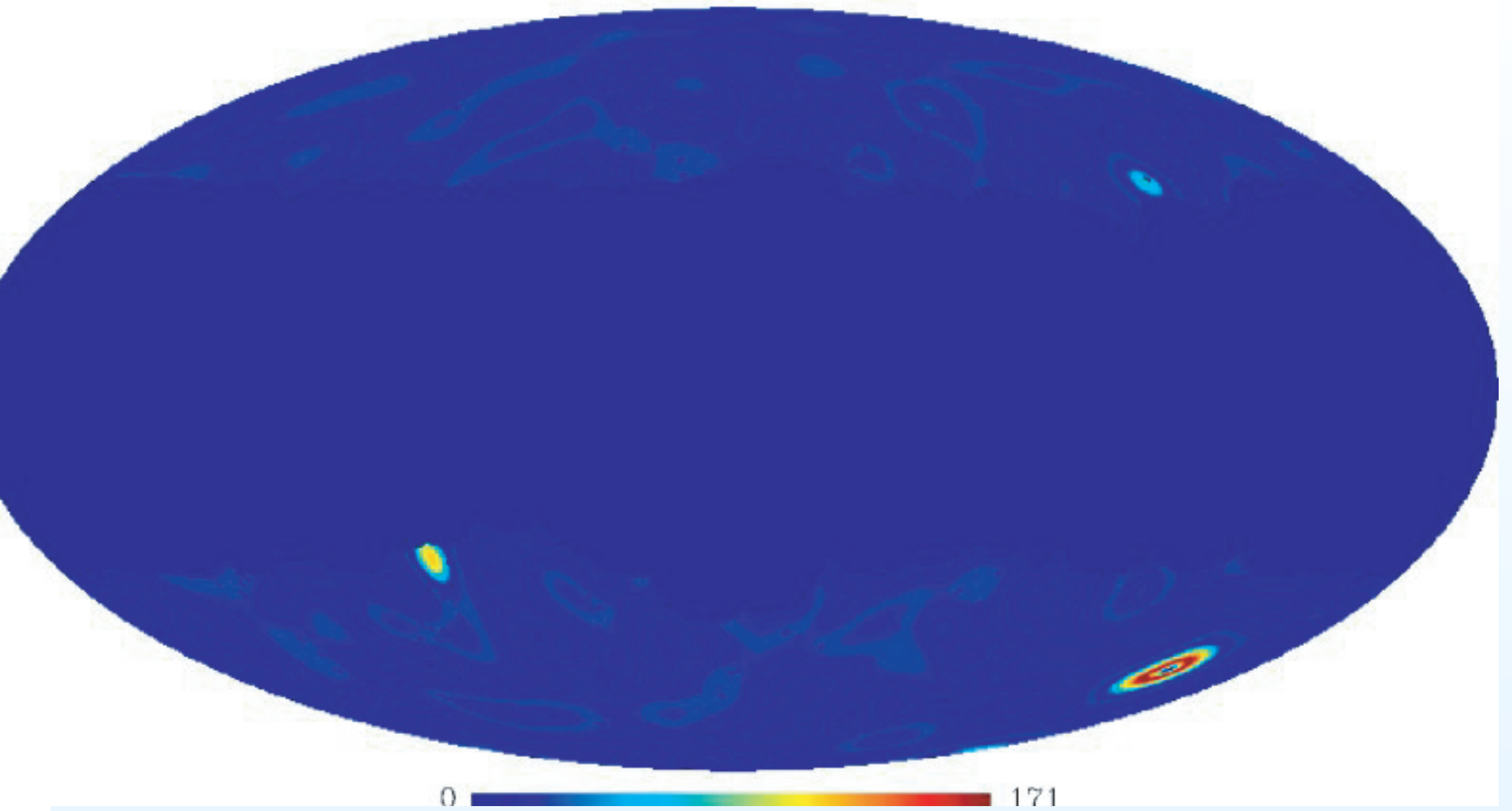
- Results support the "Lambda -Cosmic Dark Matter" model

HC WMAP

0.0 ▭ 6.4
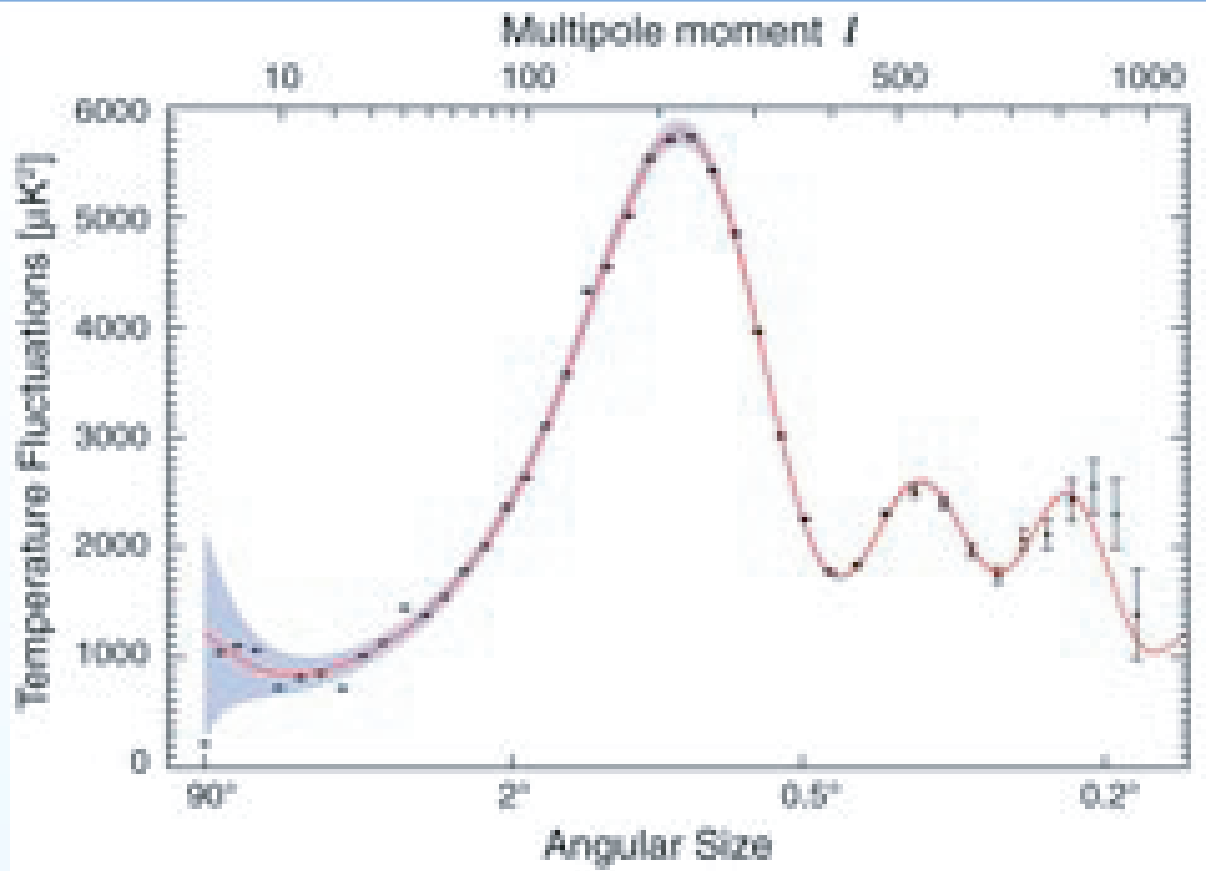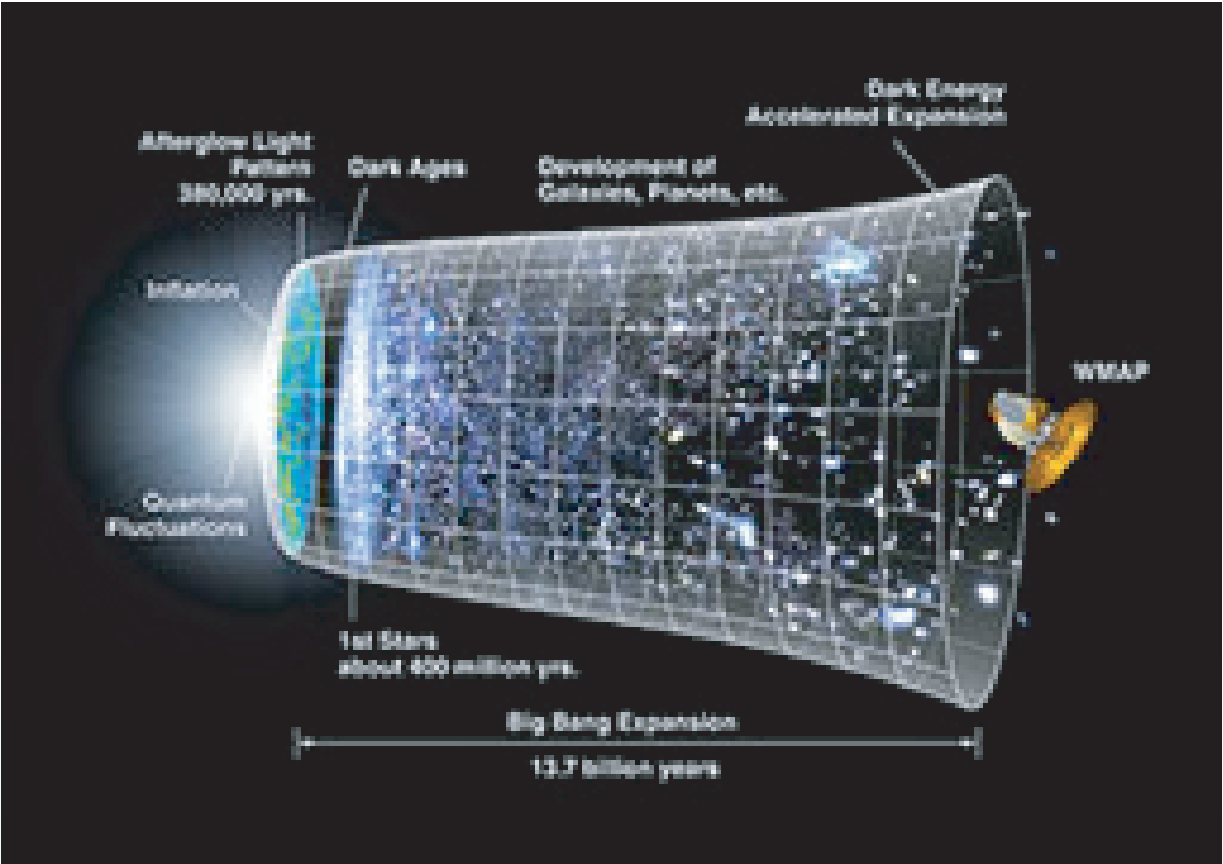
HC WMAP sc=300 arcmin

0 ▬▬▬▬▬▬▬▬▬▬▬▬▬▬ 171

From Spergel et al., 2007:

"The detection of primordial non-Gaussian fluctuations in the CMB would have a profound impact of the physics of the early universe. While the simplest inflationary models predict only mild non-Gaussianities that should be undetectable in the WMAP data, there are a wide range of plausible mechanisms for generating significant and detectable non-Gaussian fluctuations … "

7. Further problems and challenges

- Exact contiguity results for HC_n and the phi-diverence statistics under (some refinement of) the exact optimal boundary

7. Further problems and challenges

- Exact contiguity results for HC_n and the phi-diverence statistics under (some refinement of) the exact optimal boundary

- Do all the phi-divergence statistics achieve the optimal detection regions for the Generalized Gaussian (Subbotin) sparse mixture model?

7. Further problems and challenges

- Exact contiguity results for HC_n and the phi-diverence statistics under (some refinement of) the exact optimal boundary

- Do all the phi-divergence statistics achieve the optimal detection regions for the Generalized Gaussian (Subbotin) sparse mixture model?

- Can the phi-divergence test statistics be used to estimate ϵ_n (efficiently)?

7. Further problems and challenges

- Exact contiguity results for HC_n and the phi-diverence statistics under (some refinement of) the exact optimal boundary

- Do all the phi-divergence statistics achieve the optimal detection regions for the Generalized Gaussian (Subbotin) sparse mixture model?

- Can the phi-divergence test statistics be used to estimate ϵ_n (efficiently)?

- Estimation of the null hypothesis (Efron, Jin and Cai (2007))? Alternatives to the empirical characteristic function methods of Jin and Cai?

7. Further problems and challenges

- Exact contiguity results for HC_n and the phi-diverence statistics under (some refinement of) the exact optimal boundary

- Do all the phi-divergence statistics achieve the optimal detection regions for the Generalized Gaussian (Subbotin) sparse mixture model?

- Can the phi-divergence test statistics be used to estimate ϵ_n (efficiently)?

- Estimation of the null hypothesis (Efron, Jin and Cai (2007))? Alternatives to the empirical characteristic function methods of Jin and Cai?

- Dependence? Strong dependence studied by Hall and Jin (2007).
  Higher criticism improved on by maximum statistic.

7. Further problems and challenges

- Exact contiguity results for HC_n and the phi-diverence statistics under (some refinement of) the exact optimal boundary

- Do all the phi-divergence statistics achieve the optimal detection regions for the Generalized Gaussian (Subbotin) sparse mixture model?

- Can the phi-divergence test statistics be used to estimate ϵ_n (efficiently)?

- Estimation of the null hypothesis (Efron, Jin and Cai (2007))? Alternatives to the empirical characteristic function methods of Jin and Cai?

- Dependence? Strong dependence studied by Hall and Jin (2007).
  Higher criticism improved on by maximum statistic.

- Weak dependence models?

8. Some references for WMAP

- Cayón, L., Jin, J., and Treaster, A. (2005). Higher Criticism statistic: detecting and identifying non-gaussianity in the WMAP first year data. **Mon. Not. R. Astron. Soc. 362**, 826 - 832.

- Cayón, L, Banday, A. J., Jaffe, T., Eriksen, H.K.K., Hansen, F. K., Gorski, K.M., and Jin, J. (2006). No higher criticism of the Bianchi corrected WMAP data. **Mon. Not. R. Astron. Soc. 369**, 598-602.

- Cruz, M., Cayón, L., E. Martinez-Gonzalez, Vielva, P., and Jin, J. (2006). The non-Gaussian cold spot in the 3-year WMAP data. **The Astrophysical Journal 655**, 11.

- Spergel, D. N. et al. (2007). Wilkinson microwave anisotropy probe (WMAP) three year observations: implications for cosmology. **Astrophysical Journal Supplement Series 170**, 377.

- Dunkley, J. et al. (2008). Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) observations: likelihoods and parameters from the WMAP data. Submitted 5 March 2008 to the **Astrophysicial Journal Supplement Series.** See: *http://lambda.gsfc.nasa.gov/product/map/current/map_bibliography.cfm*

- Planck: The Scientific Program ("the Bluebook"). European Space Agency. Available at *http://www.rssd.esa.int/Planck*.

9. Some references for Higher Criticism and …

- Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32**, 962-994.

- Hall, P. and Jin, J. (2008). Properties of higher criticism under strong dependence. *Ann. Statist.* **36**, 381-402.

- Jager, L. and Wellner, J. (2007). Goodness-of-fit tests via phi-divergences. *Ann. Statist.* **35**, 2018-2053.