

Criticality und Zusammenhang zu Higher Critism

Patrick Wilhelm

January 31, 2011

Inhaltsverzeichnis

- 1 Einführung
- 2 Voraussetzungen und Modellannahmen
 - p-Werte
 - FDR und pFDR
 - Mixture-Model
- 3 BH95-Prozedur und Criticality
 - Benjamini-Hochberg-Prozedur
 - Criticality
 - Beispiele
- 4 Higher-Critism
 - Higher-Critism
 - Sparse-Mixture-Model
 - Beispiele

Einführung

Wenn man eine große Anzahl von Hypothesen testen will, dann ergeben sich drei zentrale Problemstellungen:

- **Detection Problem:** Existieren wahre Alternativen?
- **Estimation Problem:** Wieviele Hypothesen sind wahre Alternativen?

Einführung

Wenn man eine große Anzahl von Hypothesen testen will, dann ergeben sich drei zentrale Problemstellungen:

- **Detection Problem:** Existieren wahre Alternativen?
- **Estimation Problem:** Wieviele Hypothesen sind wahre Alternativen?
- **Selection Problem:** Welche Hypothesen sind wahre Alternativen?

Einführung

Wenn man eine große Anzahl von Hypothesen testen will, dann ergeben sich drei zentrale Problemstellungen:

- **Detection Problem:** Existieren wahre Alternativen?
- **Estimation Problem:** Wieviele Hypothesen sind wahre Alternativen?
- **Selection Problem:** Welche Hypothesen sind wahre Alternativen?

p-Werte

Proposition 1: Einseitiger p-Wert

Der einseitige p-Wert für eine Beobachtung $x \in \mathbb{R}$ ist $p(x) = 1 - F_0(x)$. Die zugehörige Verteilungsfunktion G_1 und die Dichte g_1 unter der Alternative H_1 sind gegeben durch:

$$G_1(u) = 1 - F_1(F_0^{-1}(1 - u))$$

$$g_1(u) = \frac{f_1}{f_0}(F_0^{-1}(1 - u))$$

für alle $u \in [0, 1]$

p-Werte

Proposition 2: zweiseitiger p-Wert

Sei o.B.d.A. F_0 symmetrisch, also $\forall x \in \mathbb{R} : F_0(x) + F_0(-x) = 1$ bzw. $\forall x \in \mathbb{R} : f_0(x) = f_0(-x)$, dann kann der zweiseitige p-Wert für ein $x \in \mathbb{R}$ geschrieben werden als $p(x) = 2(1 - F_0(|x|))$. Dann ist die zugehörige Verteilungsfunktion G_1 und deren Dichte g_1 unter der Alternative H_1 gegeben durch

$$G_1(u) = 1 - F_1(F_0^{-1}(1 - \frac{u}{2})) + F_1(F_0^{-1}(\frac{u}{2}))$$

$$g_1(u) = \frac{1}{2}(\frac{f_1}{f_0}(F_0^{-1}(1 - \frac{u}{2})) + \frac{f_1}{f_0}(F_0^{-1}(\frac{u}{2})))$$

für alle $u \in [0, 1]$

FDR und pFDR

Angenommen wir führen m simultane Hypothesentests durch. Dann ist die False Discovery Rate zum Schwellenwert t definiert durch:

$$FDR(t) = E[FDP(t)] := E\left[\frac{V(t)}{R(t) \vee 1}\right],$$

Die positive False Discovery Rate zum Schwellenwert t ist definiert durch:

$$pFDR(t) := E\left[\frac{V(t)}{R(t)} \mid R(t) > 0\right]$$

Die FDR ist eng verknüpft mit der pFDR durch

$$FDR(t) = pFDR(t)P(R(t) > 0).$$

FDR und pFDR

Als Funktion in Abhängigkeit von einem Schwellenwert t ergibt sich die FDP im Zusammenhang mit p-Werten also durch:

$$FDP(t) = \frac{V(t)}{R(t) \vee 1} = \frac{\sum_{i=1}^m 1_{\{P_i \leq t\}} (1 - H_i)}{\sum_{i=1}^m 1_{\{P_i \leq t\}} + 1_{\{\text{alle } P_i > t\}}}$$

Mixture-Model

Setting:

m Tests, $H_1, \dots, H_m$ Hypothesen-Indikatorfunktionen für
 $i \in \{1, \dots, m\}$, X_i sei die zugehörige Statistik.

$\Rightarrow (X_i, H_i)_{1 \leq i \leq m}$ iid, $H_i \sim \text{Bernoulli}(\pi)$,
(π = Anteil **falscher** Nullhypothesen zu allen Hypothesen i.S.v. :
 $P(H_i = 1) = \pi$).

Die bedingte Verteilung von $X_i | H_i = 1$ sei F_1 und von $X_i | H_i = 0$
sei F_0 .

$\Rightarrow$ Randverteilung der X_i :

$$F = (1 - \pi)F_0 + \pi F_1,$$

Mixture-Model

Da $P_i|H_1 = 0 \sim U(0, 1)$, erhalten wir mit $P_i|H_i = 1 \sim G_1$ insgesamt als Randverteilung der p-Werte:

$$G = (1 - \pi)U + \pi G_1$$

bzw.

$$G(u) = (1 - \pi)u + \pi G_1(u),$$

mit Dichte:

$$g(u) = 1 - \pi + \pi g_1(u)$$

wobei G_1 typischerweise in einer Klasse von (Verteilungs-) Funktionen ist derart, dass

$$\Gamma_C := \{G_1 : G_1 \text{ ist konkav, } G_1 \in C([0, 1])\}$$

Benjamini-Hochberg-Prozedur

Wir wollen die FDR zum Level α steuern. Seien $P_{(1)} \leq \dots \leq P_{(m)}$ geordnete P-Werte zu den m Tests und sei t_{BH} gemäß Benjamini-Hochberg-Prozedur definiert durch

$$t_{BH} = \max\{i : P_{(i)} \leq \alpha \frac{i}{m}, 0 \leq i \leq m\}.$$

Verteilungsfunktion der p-Werte

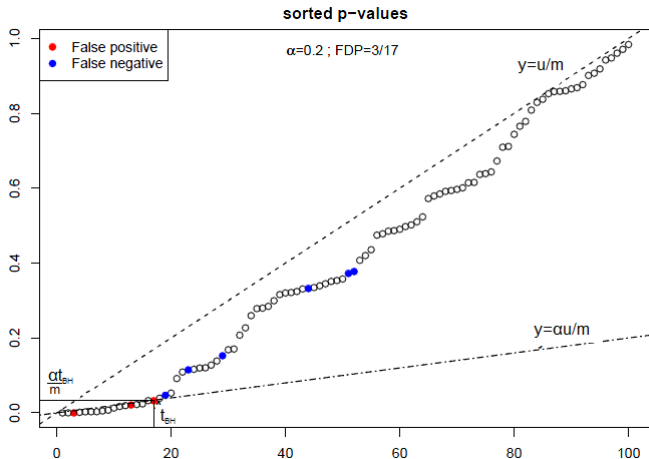


Figure: Benjamini-Hochberg Prozedur für geordnete p-Werte

Benjamini-Hochberg-Prozedur

Sei t fix, dann gilt:

$$\begin{aligned}
 FDR(t) &= E\left[\frac{\sum_{i=1}^m 1_{\{P_i \leq t\}}(1 - H_i)}{\sum_{i=1}^m 1_{\{P_i \leq t\}} + 1_{\{\text{alle } P_i > t\}}}\right] \\
 &\approx \frac{E\left[\frac{1}{m} \sum_{i=1}^m 1_{\{P_i \leq t\}}(1 - H_i)\right]}{E\left[\frac{1}{m} \sum_{i=1}^m 1_{\{P_i \leq t\}} + \frac{1}{m} P\{\text{alle } P_i > t\}\right]} \\
 &= \frac{(1 - \pi)t}{G(t) + \frac{1}{m}(1 - G(t))^m} \approx \frac{(1 - \pi)t}{G(t)}
 \end{aligned}$$

Benjamini-Hochberg-Prozedur

Sei nun $G_m(t) = \frac{1}{m} \sum_{i=1}^m 1_{\{P_i \leq t\}}$ die empirische Verteilungsfunktion der p-Werte. Dann gilt im stetigen Fall mit $G_m(P_i) \approx \frac{i}{m}$:

$$\begin{aligned} t_{BH}(P^m) &= \sup\{t : t \leq \alpha G_m(t)\} \\ &= \sup\{t : G_m(t) \leq \frac{t}{\alpha}\} \\ &= \sup\{t : \frac{t}{G_m(t)} \leq \alpha\} \end{aligned}$$

Benjamini-Hochberg-Prozedur

Zwei Interpretationen:

Schätzer der FDR

$$FDR(t) = \frac{t}{G_m(t)}$$

t_{BH} als Plug-In-Schätzer

$$u_*(\pi, G) = \max\{t : G(t) \leq \frac{t}{\alpha}\} \leq \max\{t : \frac{t}{G(t)} \leq \alpha\},$$

$$\text{denn } \lim_{m \rightarrow \infty} G_m(t) = G(t)$$

Daraus folgt $E[FDR(t_{BH})] = FDR(t_{BH}) \leq (1 - \pi)\alpha$ und
 $pFDR = \frac{(1-\pi)t}{G(t)}$

Criticality

Verschiedene Situationen für die BH95-Prozedur:

$\alpha < \alpha_*$, $\alpha > \alpha_*$ und $\alpha = \alpha_*$

Insbesondere entsteht so eine Situation für $\alpha < \alpha_*$ in der FDR und pFDR asymptotisch nicht mehr äquivalent sind.

Criticality

Kritische Werte für die BH95-Prozedur

$$\alpha_* := \inf_{u>0} \frac{u}{G(u)} \leq 1, \beta_* = (1 - \pi)\alpha_*$$

Da $G \in \Gamma_C$ ist G konkav, $G(0) = 0$ ist $u \mapsto \frac{u}{G(u)}$ ist nicht wachsend auf $[0, 1]$.

$$\implies \alpha_* := \inf_{u>0} \frac{u}{G(u)} = \lim_{u \rightarrow 0} \frac{u}{G(u)} = \lim_{u \rightarrow 0} \frac{1}{g(u)},$$

sowie

$$\beta_* = \frac{1 - \pi}{g(u)} \implies \alpha_* = \frac{\beta_*}{1 - \pi} \implies \alpha_* = 0 \iff \beta_* = 0$$

Bemerkung: Man kann also die Betrachtungen auf den Fall $\alpha_* = 0$ oder nicht reduzieren.

Criticality

Für die BH95-Prozedur hatten wir im asymptotischen Fall:

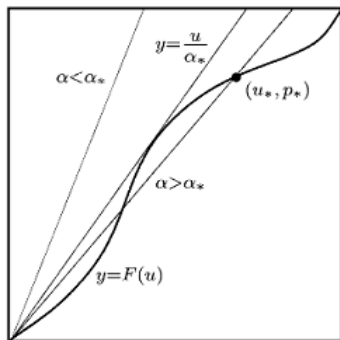
$$u_*(\pi, G) = \max\left\{t : \frac{t}{\alpha} \leq G(t)\right\}$$

als Grenzwert des größten abgelehnten p-Wert und somit

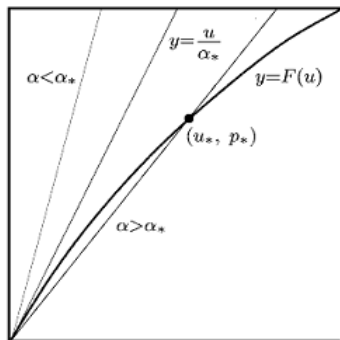
$$p_* = G(u_*) = \frac{u_*}{\alpha}$$

als Grenzwert des Anteils der p-Werte

Verteilungsfunktion der p-Werte



(a)



(b)

Figure: links: kritischer Wert α_* für allgemeines G , rechts: kritischer Wert α_* für konkaves G

Satz 1: $\alpha < \alpha_*$

Sei $\alpha < \alpha_*$. Wenn die Menge m der Hypothesen gegen unendlich strebt dann ist Anzahl der abgelehnten Alternativen t_{BH} beschränkt im Sinne von:

$$t_{BH} \xrightarrow{d} \tau$$

Mit anderen Worten: bezeichne $p_*(\alpha) = \frac{t_{BH}}{m}$ der Anteil der Ablehnungen durch die Benajmini-Hochberg-Prozedur, dann konvergiert $p_*(\alpha)$ gegen 0 für $m \rightarrow +\infty$. Es gilt außerdem:

$$\text{pFDR} \rightarrow (1 - \pi)\alpha_*$$

Interpretation von Satz 1

Es kann also keine Prozedur ein pFDR kleiner als $\beta_* = \inf_{t>0} pFDR(t) = \inf_{t>0} \frac{(1-\pi)t}{G(t)}$ erreichen.

Criticality:

FDR und pFDR asymptotisch nicht mehr äquivalent, denn es folgt für ein multiples Testproblem mit $\beta_* > 0$, dass jede Prozedur zum Level $\alpha \leq \beta_*$:

$$P(R(t) = 0) = 1 - \frac{FDR(t)}{pFDR(t)} \geq 1 - \frac{\alpha}{\beta_*} > 0$$

Bemerkung: Dieser Wert hängt jeweils von der betrachteten Prozedur ab

Satz 2: $\alpha > \alpha_*$

Sei $\alpha > \alpha_*$ und $\delta := 1 - \alpha g(u_*) > 0$ und sei $q_* = 1 - p_*$. Dann gilt:

$$\limsup_m + \frac{t_{BH} - mp_*}{\sqrt{m \log(\log(m))}} = \frac{\sqrt{2p_*q_*}}{\delta} := p_\infty, \text{ f.s.}$$

und $p_*(\alpha)$ ist asymptotisch proportional zur Macht:

$$M_m = \frac{t_{BH}}{m} \left(\frac{1 - \alpha}{\pi} + \alpha \right) + o_p(1) \rightarrow G_1(u_*)$$

Mit anderen Worten heißt das, dass in diesem Fall $p_*(\alpha)$ konvergiert gegen einen positiven Wert p_∞ für $m \rightarrow \infty$.

Lemma 1: Criticality und Likelihoodquotienten

Sei G_1 konkav und die Verteilungsfunktion der Teststatistik symmetrisch. Dann gilt:

- 1 Wenn $\frac{f_1}{f_0}(t)$ beschränkt ist für $t \rightarrow \infty$, dann hat die Dichte g_1 der p-Werte unter der Alternative einen endlichen Grenzwert bei 0 (Notation: $g_1(0)$). In diesem Fall entsteht *Criticality*. Der kritische Wert ist gegeben durch:

$$\alpha_* = \frac{1}{1 + \pi + \pi g_1(0)}.$$

- 2 Wenn $\frac{f_1}{f_0}(t) = +\infty$ für $t \rightarrow \infty$, dann gilt $\lim_{u \rightarrow 0} \frac{G(u)}{u} = +\infty$ und $\alpha_* = 0$. Dann entsteht keine *Criticality* und alle gewünschten FDR-Level sind erreichbar.

(Dieses Lemma gilt sowohl für einseitige, als auch zweiseitige p-Werte)

Gauss-Test

Seien m Statistiken gemäß $N(0, 1)$ verteilt unter der Nullhypothese und gemäß $N(\theta, 1)$ mit $\theta \neq 0$ unter der Alternative. Dann ist der Likelihoodquotient gegeben durch:

$$\frac{f_1}{f_0} = \exp\left(-\frac{1}{2}(t - \theta)^2 + \frac{1}{2}t^2\right) = \exp\left(-\frac{\theta^2}{2} + \theta t\right)$$

Da der Likelihood-Quotient nicht beschränkt ist für $t \rightarrow \infty$, folgt aus dem Lemma, dass keine *Criticality* entsteht, also $\alpha_* = 0$.

Gauss-Test

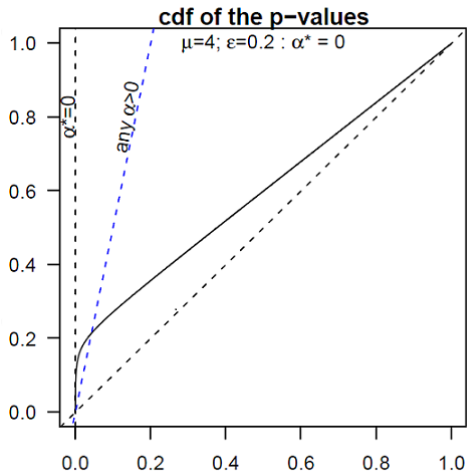


Figure: Gauss-Tes

Laplace-Test

Sei die Dichte der Teststatistik gegeben durch $f_0 : t \rightarrow \frac{1}{2}e^{-|t|}$ unter der Nullhypothese und $f_1 : t \rightarrow \frac{1}{2}e^{-|t-\theta|}$ unter der Alternative mit $\theta > 0$. Dann ist der Likelihood-Quotient gegeben durch

$$\frac{f_1}{f_0} = e^{2t-\theta}, \text{ falls } t \leq \theta$$

$$\frac{f_1}{f_0} = e^{\theta}, \text{ falls } t > \theta$$

Criticality

$$\alpha_* = \frac{1}{\pi e^{\theta} + (1 - \theta)}$$

BH95 hat asymptotisch Macht Null für $\alpha < \alpha_*$

Laplace-Test

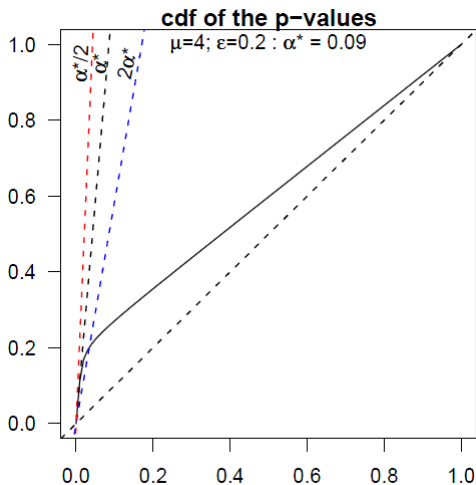


Figure: Laplace-Test

Higher-Critism

Im Rahmen des Detection Problems, also der Möglichkeit zu entscheiden ob, ein Signal *Sparse-Data* (seltene Signale) enthält oder nicht untersuchen wir die:

Detection Boundary

Asymptotische Grenze über der wir Signale entdecken und unter der es unmöglich ist Signale auszumachen

Beispiel - Verstecktes Kommunikationssignal

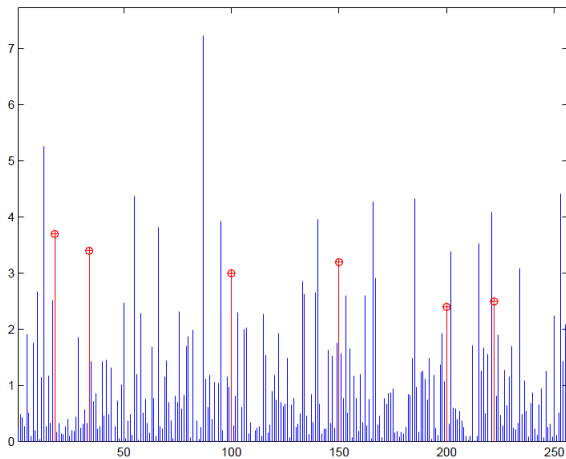


Figure: Versteckte Kommunikationssignale

Sparse-Mixture-Model

Wir betrachten das Gauss-Modell. Seien m unabhängige Beobachtungen $X_i \sim N(\mu_i, 1)$ gegeben und nur ein kleiner Anteil der μ 's ist merklich von Null verschieden.

Nach Definition ist der Vektor $\mu = \{\mu_1, \dots, \mu_m\}$ sparse.

Sparsity

$$\frac{\#\{i : \mu_i \neq 0\}}{m} \leq \pi \approx 0$$

Sparse-Mixture-Model

Betrachtet man in diesem Zusammenhang die l_p -Norm mit

$$\frac{1}{m} \sum_{i=1}^m |\mu_i|^p \leq \nu^p, \nu \text{ klein}, 0 < p < 2$$

folgt für $p \rightarrow 0$

Sparsity

$$\sum_{i=1}^m |\mu_i|^p \xrightarrow{p \rightarrow 0} \#\{i : \mu_i \neq 0\}$$

Sparse-Mixture-Model

Im Gegensatz zum Mixture-Model von oben und der Frage: welche X_i kommen von F_1

Simultanes Testen von m Hypothesen

$$X_i | H_i = 0 \sim F_0$$

$$X_i | H_i = 1 \sim F_1$$

untersuchen wir hier die Frage ob π gleich Null ist?

Binäres-Test-Problem

$$X_i | H = 0 \sim F_0, 1 \leq i \leq m$$

$$X_i | H^n = 1 \sim (1 - \pi)F_0 + \pi F_1, 1 \leq i \leq m$$

Detection-Boundary

Man setzt

$$\pi = \pi_m = m^{-\beta}, \mu = \mu_m = \sqrt{2r \log(m)}, \frac{1}{2} < \beta < 1, 0 < r < 1$$

und definiert dann die

Detection-Boundary

$$\rho_*(\beta) = \beta - \frac{1}{2}, \text{ für } \frac{1}{2} < \beta \leq \frac{3}{4}$$

$$\rho_*(\beta) = (1 - \sqrt{1 - \beta})^2, \text{ für } \frac{3}{4} < \beta < 1$$

Satz 3: Detection Boundary

- ① Sei $r > \rho_*(\beta)$ und sei

$$LR_m = \sum_{i=1}^m \log(1 - \pi + \pi e^{\mu_m x_i - \mu_m^2/2}) > 0$$

der Likelihood-Quotiententest, der H_0 ablehnt, dann:

$$P_{H_0}(\{\text{lehne } H_0 \text{ ab}\}) + P_{H_1^m}(\{\text{akzeptiere } H_0\}) \xrightarrow{m \rightarrow \infty} 0$$

- ② Sei $r < \rho_*(\beta)$, dann:

$$P_{H_0}(\{\text{lehne } H_0 \text{ ab}\}) + P_{H_1^m}(\{\text{akzeptiere } H_0\}) \xrightarrow{m \rightarrow \infty} 1$$

Detection Boundary

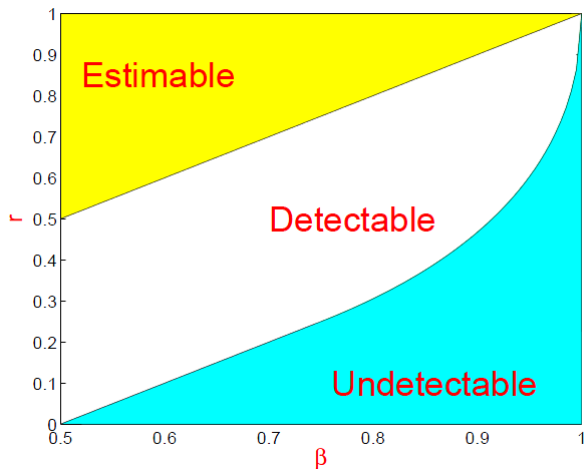


Figure: Detection Boundary spaltet die *entdeckbare Region* von der *unentdeckbaren Region*

Gauss-Test

Gaussian-Mixtures



$$X_i | H = 0 \sim N(0, 1) \text{ iid}$$

$$X_i | H^n = 1 \sim (1 - \pi)N(0, 1) + \pi N(\mu_m, 1) \text{ iid}$$

- **Sparsity**

$$\pi = \pi_m = m^{-\beta}, \mu = \mu_m = \sqrt{2r \log(m)}, \frac{1}{2} < \beta < 1, 0 < r < 1$$

Gauss-Test

Gaussian Detection Boundaries

- **Optimale Boundary**

$$\rho_*(\beta) = \beta - \frac{1}{2}, \text{ für } \frac{1}{2} < \beta \leq \frac{3}{4}$$

$$\rho_*(\beta) = (1 - \sqrt{1 - \beta})^2, \text{ für } \frac{3}{4} < \beta < 1$$

- **Benjamini-Hochberg Boundary**

$$\rho_{BH}(\beta) = (1 - \sqrt{1 - \beta})^2, \text{ für } \frac{1}{2} < \beta < 1$$

Gauss-Test: Detection Boundary

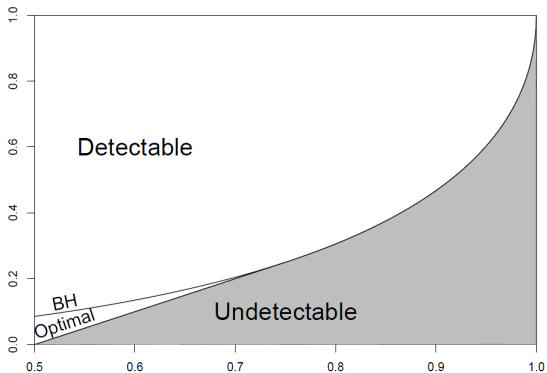


Figure: Detection Boundary - Gauss-Test

Laplace-Test: Detection Boundary

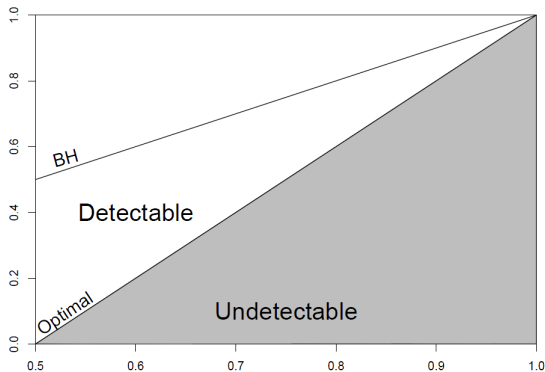


Figure: Detection Boundary - Laplace-Test

Danke

Vielen Dank für Ihre Aufmerksamkeit