4.2 Methoden um Tests zu finden: Likelihood Quotienten Tests (LRT)

Falls $X_1, \ldots, X_n$ iid aus $f(x|\theta)$, so gilt für die Likelihood Funktion

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)\,.$$

Falls $L(\theta'|\mathbf{x}) > L(\theta''|\mathbf{x})$, für $\theta', \theta'' \in \Theta$, ist es plausibler, dass θ' und nicht θ'' der wahre Wert des Parameters θ ist, gegeben die vorliegenden Beobachtungen $\mathbf{x}$.

Definition 4.2.1: Die **(LRT) Statistik** für den Test

$$H_0 : \theta \in \Theta_0 \quad \text{gegen} \quad H_1 : \theta \in \Theta_0^c\,,$$

mit einer Untermenge Θ_0 des Parameterraums Θ und ihrem Komplement Θ_0^c, ist

$$\lambda(\mathbf{x}) = \frac{\sup\limits_{\theta \in \Theta_0} L(\theta|\mathbf{x})}{\sup\limits_{\theta \in \Theta} L(\theta|\mathbf{x})} \qquad \in [0, 1]\,.$$

Ein LRT ist ein Test mit Verwerfungsbereich R der Form

$$R = \{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$$

für eine beliebige Zahl $0 \leq c \leq 1$. Ist $\lambda(\mathbf{x})$ klein, so ist dies ein Hinweis **gegen** H_0. Ist hingegen $\lambda(\mathbf{x})$ groß, dann spricht dies **für** H_0.

Sei $\hat{\theta} = \hat{\theta}(\mathbf{x})$ der MLE von θ und sei $\hat{\theta}_0 = \hat{\theta}_0(\mathbf{x})$ der MLE von θ unter dem eingeschränkten Parameterbereich Θ_0. Damit ist die LRT Statistik gleich

$$\lambda(\mathbf{x}) = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}\,.$$

Beispiel 4.2.1: (Normal LRT) Seien $X_1, \ldots, X_n$ iid Normal$(\theta, 1)$. Betrachte den **zweiseitigen Test**

$$H_0 : \theta = \theta_0 \quad \text{gegen} \quad H_1 : \theta \neq \theta_0 \,.$$

Da unter H_0 nur 1 Wert spezifiziert ist, gilt $\Theta_0 = \{\theta_0\}$ und

$$\sup_{\theta \in \Theta_0} L(\theta|\mathbf{x}) = L(\theta_0|\mathbf{x}) \,.$$

Der uneingeschränkte MLE von θ ist $\overline{X}$ (bereits gezeigt), womit folgt, dass

$$\sup_{\theta \in \Theta} L(\theta|\mathbf{x}) = L(\overline{x}|\mathbf{x}) \,.$$

Somit resultiert

$$\lambda(\mathbf{x}) = \frac{L(\theta_0|\mathbf{x})}{L(\overline{x}|\mathbf{x})} \,.$$

Die LRT Statistik vereinfacht sich zu

$$\begin{aligned}
\lambda(\mathbf{x}) &= \frac{(2\pi)^{-n/2}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta_0)^2\right]}{(2\pi)^{-n/2}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right]} \\
&= \exp\left\{\frac{1}{2}\left[-\sum_{i=1}^{n}(x_i-\theta_0)^2+\sum_{i=1}^{n}(x_i-\overline{x})^2\right]\right\} \\
&= \exp\left\{-\frac{n}{2}(\overline{x}-\theta_0)^2\right\},
\end{aligned}$$

denn

$$\sum_{i=1}^{n}(x_i-\theta_0)^2=\sum_{i=1}^{n}(x_i-\overline{x}+\overline{x}-\theta_0)^2=\sum_{i=1}^{n}(x_i-\overline{x})^2+n(\overline{x}-\theta_0)^2\,.$$

Was nun?

Vorgehen:

(1) Bestimme zuerst $\lambda(\mathbf{x})$.

(2) Vereinfache die Beschreibung des kritischen Bereichs (falls möglich) auf einen Ausdruck der auf einer einfacheren Statistik beruht.

In unserem Fall ist $\lambda(\mathbf{x})$ klein (und wir verwerfen H_0) falls

$$\begin{aligned} \lambda(\mathbf{x}) = \exp\left\{-\frac{n}{2}(\overline{x}-\theta_0)^2\right\} &\leq c \\ (\overline{x}-\theta_0)^2 &\geq -\frac{2}{n}\log(c) \\ |\overline{x}-\theta_0| &\geq \sqrt{-\frac{2}{n}\log(c)} =: c^* \qquad 0 \leq c^*\,, \end{aligned}$$

also wenn $|\overline{x}-\theta_0|$ groß ist.

Der Verwerfungsbereich $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$ ist somit äquivalent mit

$$\{\mathbf{x} : |\overline{x} - \theta_0| \geq c^*\}\,.$$

Wie wählt man c^*? Nimm c^* derart, dass beispielsweise gilt

$$P(H_0 \text{ verwerfen}|H_0 \text{ richtig}) = \alpha = 0.05$$

(Wahrscheinlichkeit für **Fehler erster Art**, type I error probability, α).

Allgemein gilt

$$Z = \frac{\overline{X} - \theta}{1/\sqrt{n}} \sim \text{Normal}(0, 1)$$

aber wahrer Wert von θ ist unbekannt.

Unter H_0 jedoch folgt dafür

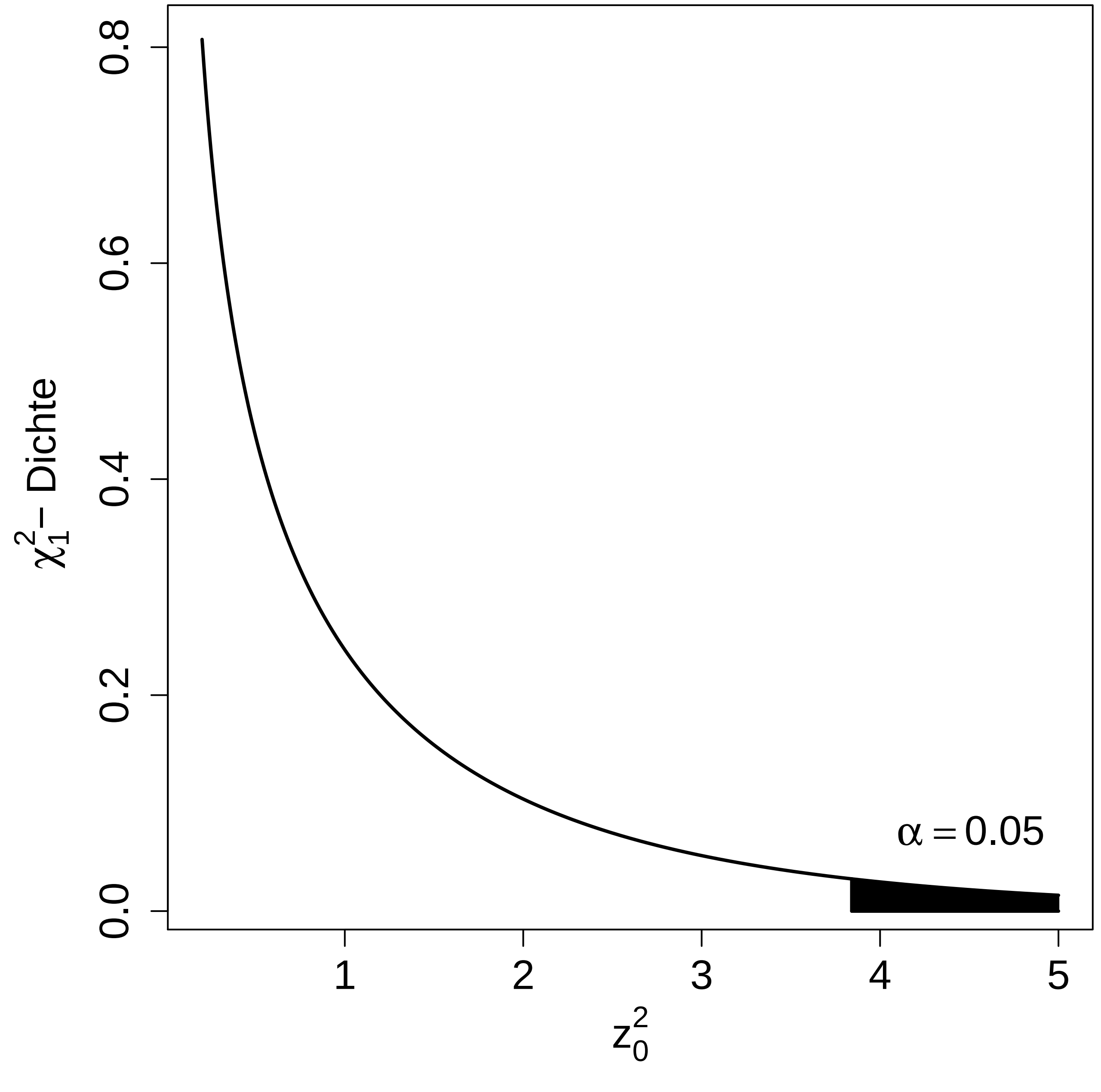
$$Z_0 = \frac{\overline{X} - \theta_0}{1/\sqrt{n}} \sim \text{Normal}(0, 1)$$

und damit ist

$$\lambda(\mathbf{x}) = \exp\left\{-\frac{n}{2}(\overline{x} - \theta_0)^2\right\} = \exp\left(-\frac{1}{2}z_0^2\right) .$$

Die Entscheidung ist äquivalent mit der Entscheidung bzgl.

$$-2\log\lambda(\mathbf{X}) = Z_0^2 \sim \chi_1^2 .$$



0.0
0.2
0.4
0.6
0.8
χ_1^2– Dichte
$\alpha = 0.05$
1
2
3
4
5
z_0^2

Beispiel 4.2.2 Seien $X_1, \ldots, X_n$ iid Exponential$(1, \theta)$, d.h. aus einer Population mit Dichte

$$f(x|\theta) = \exp(-(x-\theta)) I_{[\theta,\infty)}(x)\,, \qquad \theta \in \mathbb{R}\,.$$

Betrachte den **einseitigen Test**

$$H_0 : \theta \le \theta_0 \quad \text{gegen} \quad H_1 : \theta > \theta_0\,,$$

wobei θ_0 frei gewählt ist.

Als Likelihood Funktion resultiert mit $x_{(1)} := \min_i(x_i)$

$$L(\theta|\mathbf{x}) = \begin{cases} \exp\left(-n(\overline{x} - \theta)\right) & \text{für } \theta \le x_{(1)} \\ 0 & \text{sonst.} \end{cases}$$

$L(\theta|\mathbf{x})$ ist monoton wachsend in θ auf dem Bereich $\theta \le x_{(1)}$.

$\Rightarrow$ uneingeschränkte Maximum von $L(\theta|\mathbf{x})$ ist in $\theta = x_{(1)}$ mit

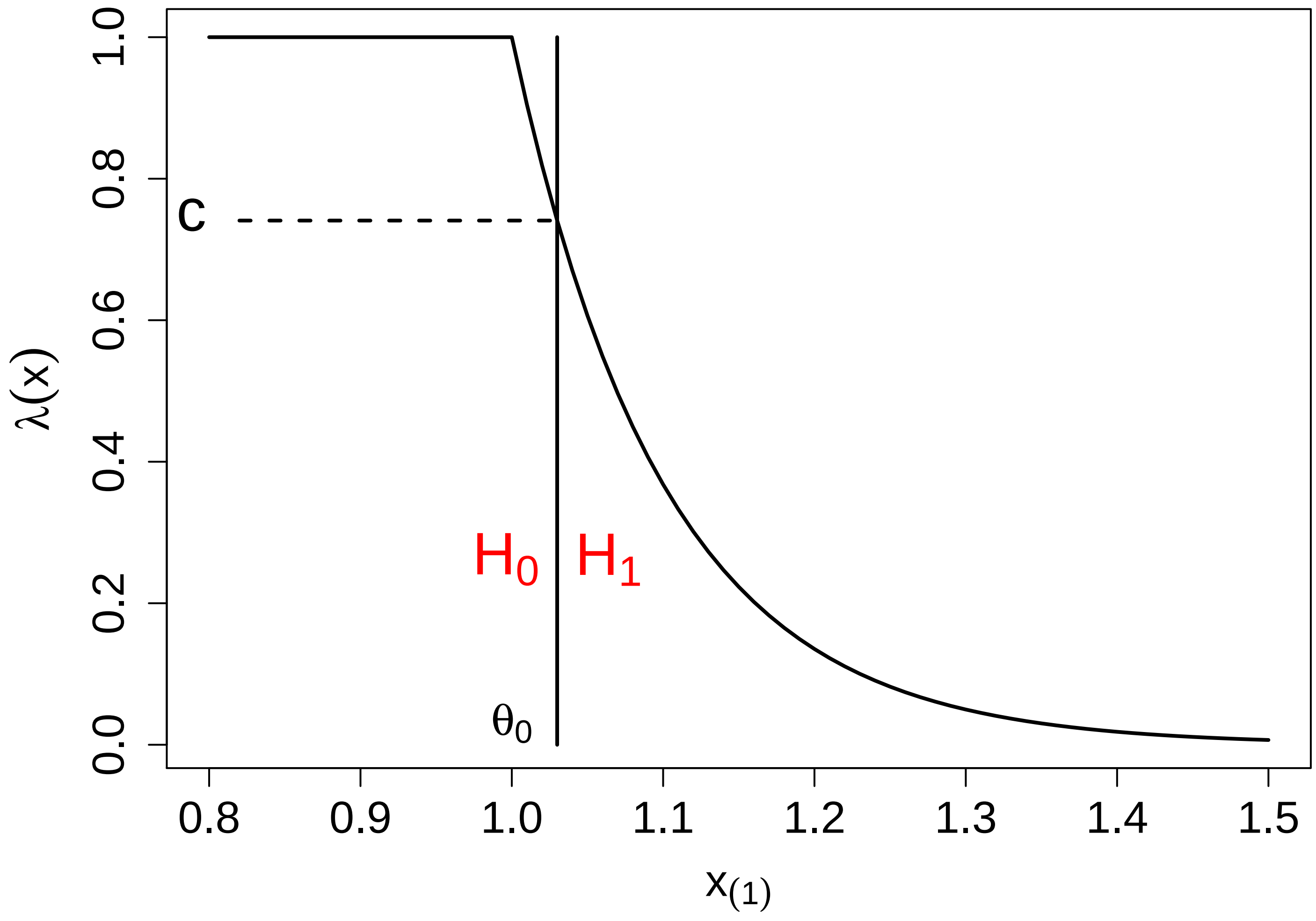
$$L(x_{(1)}|\mathbf{x}) = \exp\left(-n(\overline{x} - x_{(1)})\right)$$

$\Rightarrow$ eingeschränkte Maximum von $L(\theta|x)$ unter $H_0 : \theta \leq \theta_0$ ist

$$\max_{\theta\leq\theta_0} L(\theta|\mathbf{x}) = \begin{cases} L(\theta_0|\mathbf{x}) & \text{falls } \theta_0 < x_{(1)} \\ L(x_{(1)}|\mathbf{x}) & \text{falls } x_{(1)} \leq \theta_0. \end{cases}$$

Als LRT Statistik folgt

$$\lambda(\mathbf{x}) = \frac{L(\min(x_{(1)}, \theta_0)|\mathbf{x})}{L(x_{(1)}|\mathbf{x})} = \begin{cases} 1 & \text{für } x_{(1)} \leq \theta_0 \\ \exp\left(-n(x_{(1)} - \theta_0)\right) & \text{für } x_{(1)} > \theta_0 \end{cases}$$



1.0
0.8
0.6
0.4
0.2
0.0
λ(x)
c
H0
H1
θ0
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
x(1)

Für $x_{(1)} > \theta_0$ verwerfen wir H_0 falls $\lambda(\mathbf{x}) \leq c$, d.h. falls

$$\exp\left(-n(x_{(1)} - \theta_0)\right) \leq c\,.$$

Der entsprechende Verwerfungsbereich

$$\left\{\mathbf{x} : x_{(1)} \geq \theta_0 - \frac{1}{n}\log(c)\right\}$$

hängt von der Stichprobe nur über die suffiziente Statistik $X_{(1)}$ ab.

In beiden Beispielen hatten wir

$$X_i \overset{iid}{\sim} \text{Normal}(\theta, 1) \qquad\qquad \lambda(\mathbf{x}) = \lambda^*(\overline{x})$$

$$X_i \overset{iid}{\sim} \text{Exponential}(1, \theta) \qquad\qquad \lambda(\mathbf{x}) = \lambda^*(x_{(1)})\,,$$

d.h. die LRT Statistiken hingen nur von den suffizienten Statistiken ab.

Gilt dies allgemein?

Satz 4.2.1: Ist $T(\mathbf{X})$ eine suffiziente Statistik für θ, und sind $\lambda^*(t)$ sowie $\lambda(\mathbf{x})$ die LRT Statistiken basierend auf T und auf $\mathbf{X}$, dann gilt für jedes $\mathbf{x}$ aus dem Stichprobenraum

$$\lambda^*(T(\mathbf{x})) = \lambda(\mathbf{x})\,.$$

ad (2) "Vereinfachte Beschreibung des kritischen Bereichs": das vereinfachte $\lambda(\mathbf{x})$ sollte nur noch von $T(\mathbf{x})$ abhängen.

Beispiel 4.2.3: Seien $X_1, \ldots, X_n$ iid Normal(μ, σ^2) mit σ^2 unbekannt: Teste

$$H_0 : \mu \leq \mu_0 \qquad \text{gegen} \qquad H_1 : \mu > \mu_0 \,.$$

In diesem Zusammenhang nennt man σ^2 einen **nuisance parameter** (ist lästig und bringt nur Ärger, Störparameter).

Wir brauchen

$$\lambda(\mathbf{x}) = \frac{\sup\limits_{\mu \leq \mu_0, \sigma^2} L(\mu, \sigma^2 | \mathbf{x})}{\sup\limits_{\mu \in \mathbb{R}, \sigma^2} L(\mu, \sigma^2 | \mathbf{x})} \,.$$

Wie bereits gezeigt ergeben sich ohne Restriktion $\hat{\mu} = \overline{x}$ und $\hat{\sigma}^2 = \frac{1}{n} \sum_i (x_i - \overline{x})^2$.

Im eingeschränkten Parameterbereich $\mu \leq \mu_0$ ($\sigma^2 > 0$) maximieren wir

$$L(\mu, \sigma^2 | \mathbf{x}) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}.$$

Nun ist

$$\frac{d}{d\mu}\sum_{i=1}^{n}(x_i - \mu)^2 = -2n(\overline{x} - \mu) = \begin{cases} < 0 & \text{für } \overline{x} > \mu \\ = 0 & \text{für } \overline{x} = \mu \\ > 0 & \text{für } \overline{x} < \mu \end{cases}$$

Die Minimalstelle von $\sum_{i=1}^{n}(x_i - \hat{\mu}_0)^2$ über $\mu \leq \mu_0$ ist somit

$$\hat{\mu}_0 = \min(\overline{x}, \mu_0).$$

Es kann nun gezeigt werden, dass $L(\hat{\mu}_0, \sigma^2|\mathbf{x})$ maximal ist in

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_0)^2 .$$

Zusammen ergibt sich

$$\left(\hat{\mu}_0, \hat{\sigma}_0^2\right) = \begin{cases} \left(\overline{x}, \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right) & \overline{x} \leq \mu_0 \\ \left(\mu_0, \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2\right) & \overline{x} > \mu_0 . \end{cases}$$

Das führt zu

$$\lambda(\mathbf{x}) = \frac{L(\hat{\mu}_0, \hat{\sigma}_0^2|\mathbf{x})}{L(\overline{x}, \hat{\sigma}^2|\mathbf{x})} = \begin{cases} 1 & \text{für } \overline{x} \leq \mu_0 \text{ (spricht für } H_0) \\ \dfrac{L\left(\mu_0, \frac{1}{n}\sum_{i}(x_i - \mu_0)^2 \Big| \mathbf{x}\right)}{L(\overline{x}, \hat{\sigma}^2|\mathbf{x})} & \text{für } \overline{x} > \mu_0. \end{cases}$$

Für den Fall $\overline{x} > \mu_0$ ist dies

$$\begin{aligned}
\lambda(\mathbf{x}) &= \frac{(2\pi)^{-n/2}\left[\frac{1}{n}\sum_i (x_i-\mu_0)^2\right]^{-n/2}\exp\left\{-\frac{1}{2}\frac{\sum_i (x_i-\mu_0)^2}{\frac{1}{n}\sum_i (x_i-\mu_0)^2}\right\}}{(2\pi)^{-n/2}\left[\frac{1}{n}\sum_i (x_i-\overline{x})^2\right]^{-n/2}\exp\left\{-\frac{1}{2}\frac{\sum_i (x_i-\overline{x})^2}{\frac{1}{n}\sum_i (x_i-\overline{x})^2}\right\}} \\
&= \left[\frac{\sum_i (x_i-\mu_0)^2}{\sum_i (x_i-\overline{x})^2}\right]^{-n/2} = \left[\frac{\sum_i (x_i-\overline{x})^2 + n(\overline{x}-\mu_0)^2}{\sum_i (x_i-\overline{x})^2}\right]^{-n/2} \\
&= \left[1 + \frac{n(\overline{x}-\mu_0)^2}{\sum_i (x_i-\overline{x})^2}\right]^{-n/2} = \left[1 + (n-1)^{-1}\left(\frac{\overline{x}-\mu_0}{s/\sqrt{n}}\right)^2\right]^{-n/2} .
\end{aligned}$$

Der LRT verwirft H_0 falls

$$\lambda(\mathbf{x}) \le c \;\Leftrightarrow\; \left(\frac{\overline{x}-\mu_0}{s/\sqrt{n}}\right)^2 \ge (n-1)(c^{-2/n}-1) \;\Leftrightarrow\; \frac{\overline{x}-\mu_0}{s/\sqrt{n}} \ge c^* > 0\,.$$