Balanced adjusted empirical likelihood

Art B. Owen
Stanford University

Sarah Emerson
Oregon State University

Empirical likelihood

Observations $x_i \in \mathbb{R}^d$, $x_i \overset{\text{iid}}{\sim} F_0$

$$L(F) = \prod_{i=1}^{n} F\{X_i\}$$

Nonparametric MLE is at $F = F_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$ ECDF Kiefer & Wolfowitz (1956)

Some other NPMLE results:

Survival analysis	Kaplan & Meier (1958)
Survey sampling	Hartley & Rao (1967)
Interval censoring	Peto (1973)
Censoring & truncation	Lynden-Bell (1971)
Monotone density	Grenander (1956)

etc., see monograph O (2001)

Nonparametric likelihood ratios

$$R(F) = \frac{L(F)}{L(F_n)} = \prod_{i=1}^{n} nw_i \quad \text{where} \quad w_i = F\{X_i\}.$$

Profile likelihood

Statistic $T = T(F)$, true value $\tau_0 = T(F_0)$, NPMLE $\widehat{T} = T(F_n)$

$$\mathcal{R}(\tau) = \max\{R(F) \mid T(F) = \tau\}$$

When do we get $-2\log(\mathcal{R}(\tau_0)) \overset{d}{\approx} \chi^2$?

IE, a Wilks like result without assuming a parametric form.

Then $C = \{\tau \mid \mathcal{R}(\tau) \geqslant \eta\}$ is an approximate confidence region.

For the mean

$$T(F) = \int x \, \mathrm{d}F(x), \quad \text{ie} \quad \mu = \mathbb{E}(X)$$

Degeneracy

$$R\big((1-\epsilon)F_n + \epsilon 1_x\big) \to 1 \quad \text{as } \epsilon \to 0 \quad \text{any } x$$

For small enough $\epsilon > 0$,

$$(1-\epsilon)\bar{X} + \epsilon x \in C \equiv \{\mu \mid \mathcal{R}(\mu) \geqslant \eta\}$$

Letting $\|x\| \to \infty$ we get $C = \mathbb{R}^d$

Non-degeneracy

For the mean, restrict F to have support in a bounded set $B \subset \mathbb{R}^d$.

It is enough to take $B = B_n \equiv \mathrm{chull}(x_1, \dots, x_n)$

B_n grows with n

Upshot

$$\mathcal{R}(\mu) = \sup\left\{ \prod_i (nw_i) \,\middle|\, \sum_{i=1}^n w_i x_i = \mu, w_i > 0, \sum_i w_i = 1, \right\}$$

Then $-2\log(\mathcal{R}(\mu_0)) \xrightarrow{\mathrm{d}} \chi^2_{(d)}$ under moment conditions. O (1990)

Some good things about EL

1) (correct) data driven shape for confidence sets Hall
2) power optimality of tests Kitamura
3) allows side constraints O (1991), Qin & Lawless (1993)
4) Bartlett correctable DiCiccio, Hall & Romano (1991)
5) extends for
   a) censoring
   b) truncation
   c) biased sampling,
6) methods for
   a) time series Kitamura
   b) survey sampling Qin, Chen, Sitter, . . .

Many more extensions S.-X. Chen; Hjort, McKeague & van Keilegom; Lahiri

Drawbacks

1) Region for mean bounded by convex hull of B_n

2) Profiling the likelihood can be hard

Hull

Coverage of EL is at most coverage of hull of B_n.

This is a problem for small n and/or large d.

Profiling

Profiling is also hard for parametric likelihoods.

Empirical likelihood is usually easy to compute for a fixed parameter vector.

Convexity

Profile empirical likelihood

$$\mathcal{R}(\mu) = \sup\left\{ \prod_{i=1}^{n} (nw_i) \;\middle|\; w_i \geqslant 0,\ \sum_i w_i = 1,\ \sum_i w_i(x_i - \mu) = 0 \right\}$$

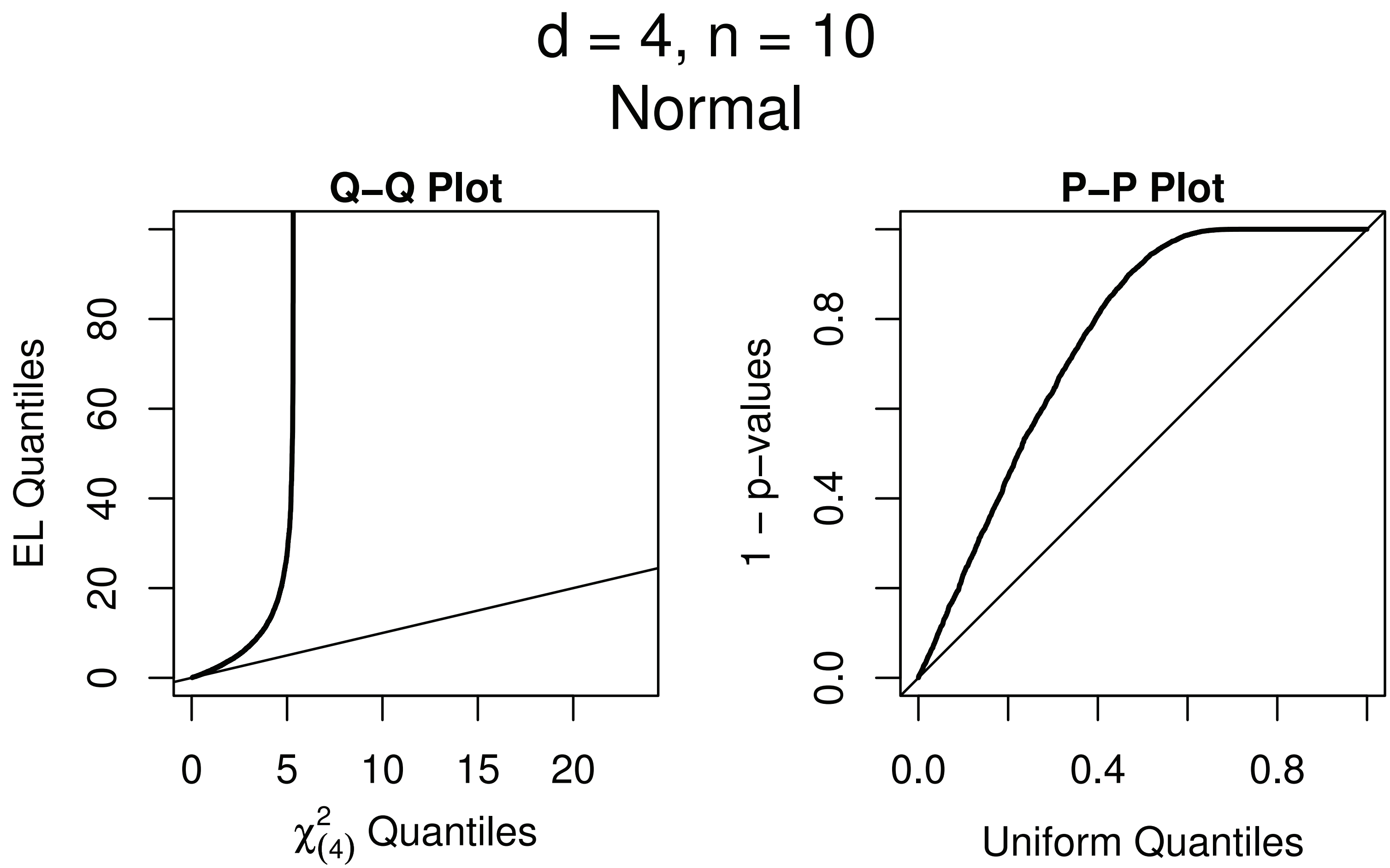
Confidence region for a mean

$$C(\epsilon) = \{\mu \in \mathbb{R}^d \mid \mathcal{R}(\mu) > \epsilon\}$$

Nested inside convex hull

$$C(\epsilon) \subset C(0) = \text{convex hull}(x_1, \ldots, x_n), \quad \forall \epsilon > 0$$

Adjusted EL coverage (extreme case)



Emerson & O (2009)

Vertical asymptote from atom at $+\infty$ for $-2\log\mathcal{R}(\mu_0)$.

Escape from the hull

Idea: extend B_n to ensure that $\mu \in B_n$

Add an artificial point (undata) x_{n+1}. Now,

$$T(F) = \sum_{i=1}^{n+1} w_i x_i, \quad \text{and,}$$

$$L(F) = \prod_{i=1}^{n} w_i, \quad \text{or,}$$

$$L(F) = \prod_{i=1}^{n+1} w_i.$$

The second version is easier computationally and asymptotically the same (if $\|x_{n+1}\|$ reasonable).

Chen, Variyath & Abraham (2008) originate this approach.

Adjusted empirical likelihood

Chen, Variyath & Abraham (2008) use

$$x_{n+1} = \mu - a_n(\bar{x} - \mu), \quad a_n = \log(n)/2$$

$$a_n = o_p(n^{2/3}) \quad \text{preserves 1st order asymptotics}$$

Note: new point x_{n+1} depends on μ

Now μ is between $\bar{x}$ and x_{n+1}:

$$\mu = \frac{x_{n+1} + a_n \bar{x}}{1 + a_n}$$

Hull of $x_1, \ldots, x_{n_1}$ contains μ

Not all is well yet

Let $\mathcal{R}^*$ be adjusted profile empirical likelihood. Then:

$$-2\log\mathcal{R}^*(\mu) \leqslant -2\left[n\log\left(\frac{(n+1)a_n}{n(a_n+1)}\right)+\log\left(\frac{n+1}{a_n+1}\right)\right]$$

which is bounded, even if $\|\mu\| \to \infty$.

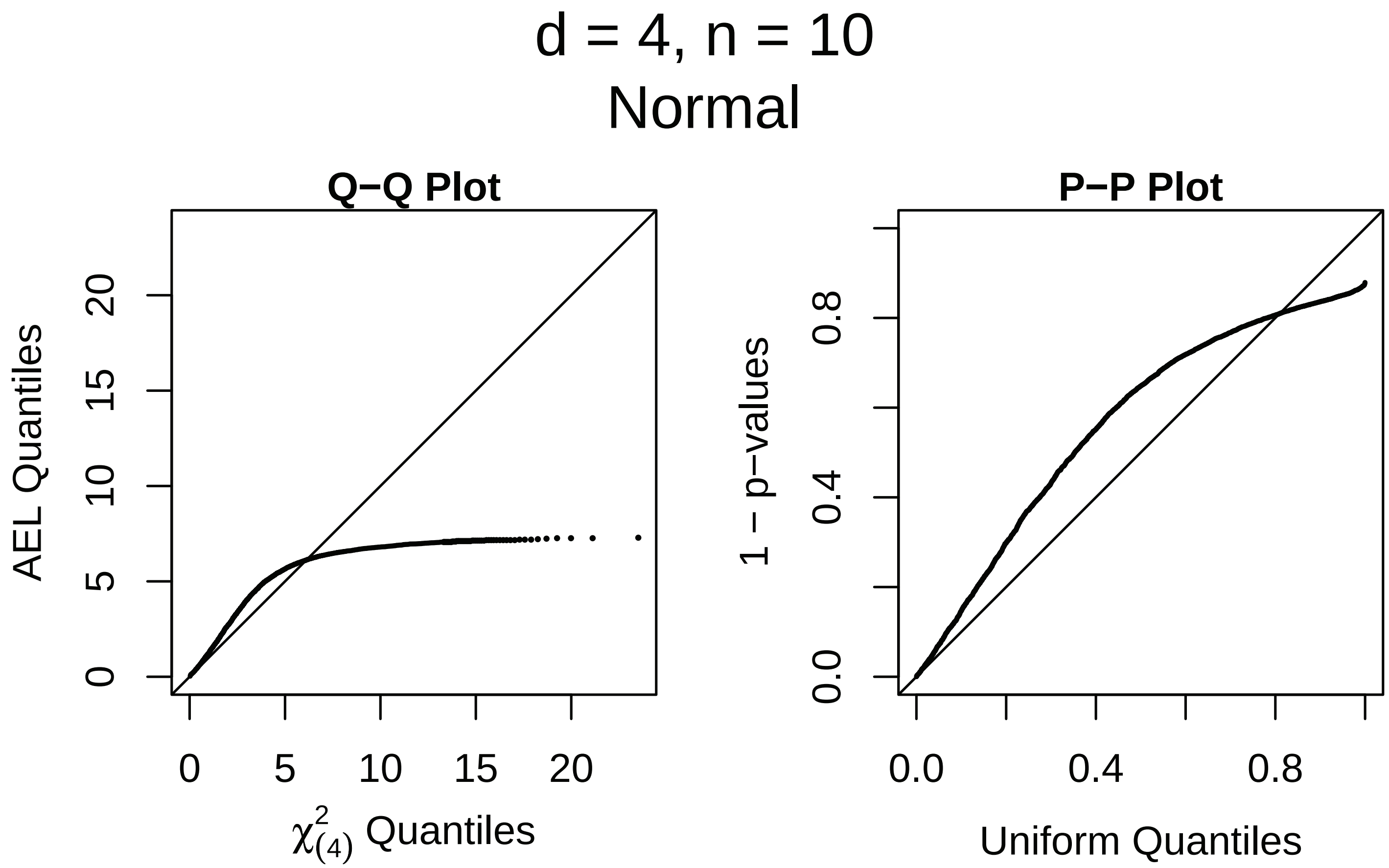
Opposite problem from $\log\mathcal{R}(\mu)$ which diverged at finite $\|\mu\|$.

Instead of a bounded 100% region we can get all of $\mathbb{R}^d$ at less than 100% confidence.

Extreme example ctd.

$n = 10$, $d = 4$, 88.1% region is $\mathbb{R}^4$.

Coverage (extreme case)



Emerson & O (2009)

Balanced adjusted empirical likelihood

Dissertation: Emerson (2009)

1) Add 2 points x_{n+1} and x_{n+2}
2) $(x_{n+1} + x_{n+2})/2 = \bar{x}$ (preserving sample mean)
3) farther new points if $\mu - \bar{x}$ is a direction where the sample varies a lot

Add points

$$x_{n+1} = \mu - s c_{u^*} u^*$$

$$x_{n+2} = 2\bar{x} - \mu + s c_{u^*} u^*$$

where

$$u^* = \frac{\bar{x} - \mu}{\|\bar{x} - \mu\|} \qquad c_{u^*} = ({u^*}^{\mathsf{T}} S^{-1} u^*)^{-1/2}$$

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^{\mathsf{T}} \qquad s \approx 1.9$$

Choice of s is based on empirical work. The best s depends (weakly) on d.

Related

Independently Liu & Chen (2009) also added 2 points.

Their 2 points were designed to improve Bartlett correction.

Ours were tuned to give good small sample coverage in high dimensions.

Invariance

Let $A \in \mathbb{R}^{d\times d}$ be non-singular.

Set $\widetilde{x}_i = Ax_i$ and $\widetilde{\mu} = A\mu$.

Let C be the balanced adjusted empirical likelihood region for μ_0 based on $x_1, \dots, x_n$.

Let $\widetilde{C}$ be the balanced adjusted empirical likelihood region for $\widetilde{\mu}_0 = A\mu_0$ based on $\widetilde{x}_1, \dots, \widetilde{x}_n$.

Then $\mu \in C \iff \widetilde{\mu} \in \widetilde{C}$.

Emerson & O (2009) Proposition 4.1.

Hotelling's T^2 and the original EL are also invariant this way.

Avoiding the boundedness

Recall $-2\log\mathcal{R}^*$ was bounded.

The new criterion $-2\log\mathcal{R}^{**}$ is unbounded.

The ultimate cause is that

$\|x_{n+1}-\mu\|$ is proportional to $\|\bar{x}-\mu\|$ in AEL but is of constant order in BAEL

The larger $\|x_{n+1}-\mu\|$ in AEL means that less weight needs to go there.
Less weight there $\cdots$ allows more weight on the other n points and a higher likelihood.

Connection to T^2

Recall

$$x_{n+1} = \mu - s c_{u^*} u^* \qquad x_{n+2} = 2\bar{x} - \mu + s c_{u^*} u^*, \quad \text{where}$$

$$u^* = \frac{\bar{x} - \mu}{\|\bar{x} - \mu\|} \quad \text{and} \quad c_{u^*} = ({u^*}^{\mathsf{T}} S^{-1} u^*)^{-1/2}.$$

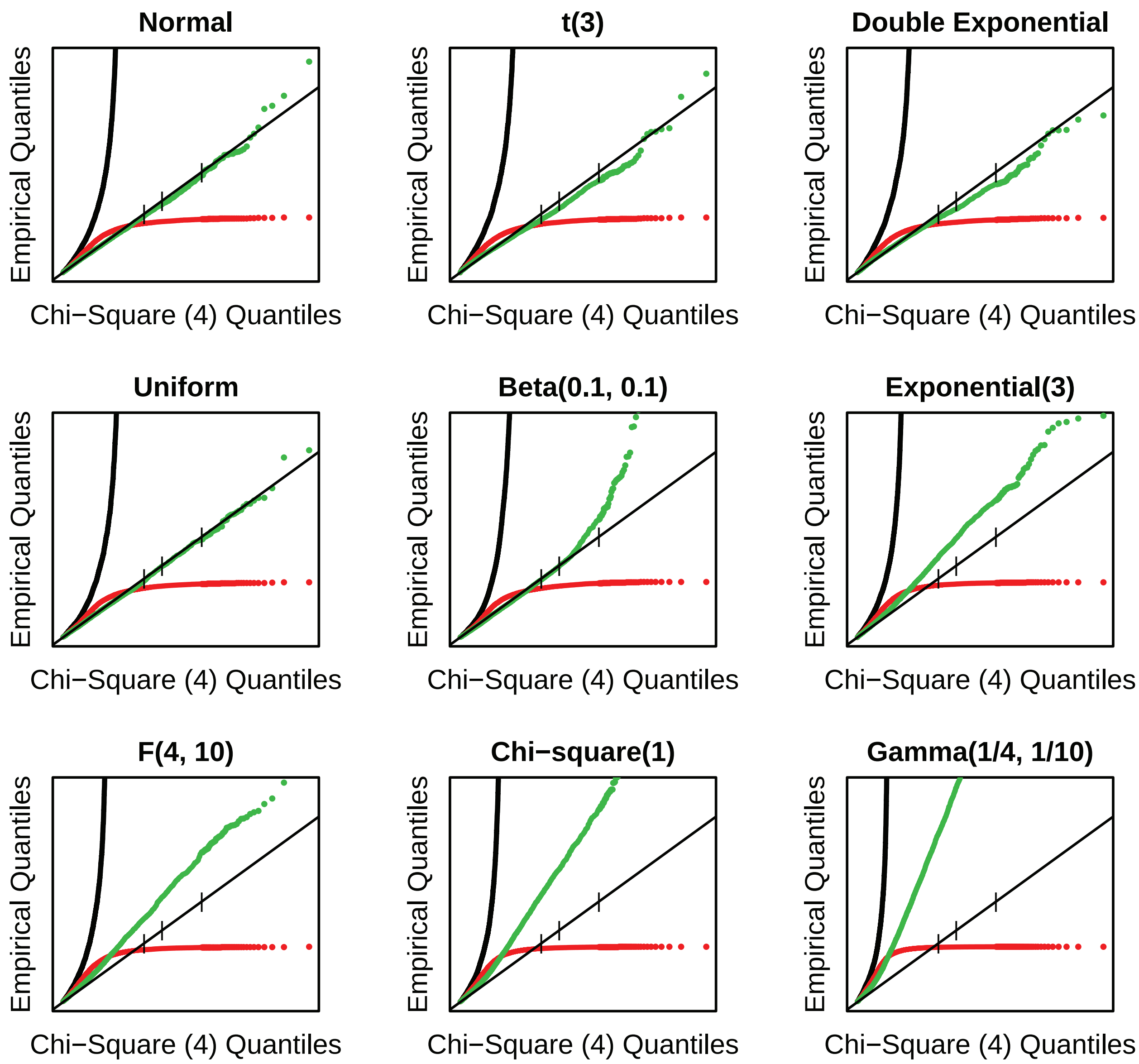
Theorem 4.2

$$\lim_{s \to \infty} \frac{2ns^2}{(n+2)^2} \Big(-2 \log \mathcal{R}^{**}(\mu) \Big) = T^2(\mu)$$

Emerson & O (2009)

Quantile−Quantile Plots

d = 4, n = 10



Emerson & O (2009)

Comments

1) More examples in the article
2) Good calibration for distributions with shorter tails
3) High kurtosis is harder
4) Even there the calibration is almost linear so a Bartlett correction could help a lot
5) Exact nonparametric CI.s for the mean are unobtainable Bahadur & Savage (1956)

Now, regression

$$y \sim x^{\mathsf{T}}\beta, \quad x \in \mathbb{R}^d \quad y \in \mathbb{R}$$

Estimating equations*

$$\mathbb{E}\big(x(y - x^{\mathsf{T}}\beta)\big) = 0$$

Normal equations

$$\sum_{i=1}^{n} x_i(y_i - x_i^{\mathsf{T}}\beta) = 0 \in \mathbb{R}^d$$

In principle we let $z_i = z_i(\beta) \equiv x_i(y_i - x_i^{\mathsf{T}}\beta) \in \mathbb{R}^d$, adjoin z_{n+1} and z_{n+2}, and carry on.

*residuals $\epsilon = (y - x^{\mathsf{T}}\beta)$ are uncorrelated with x.

They have mean zero too, when as usual, x contains a constant.

Regression hull condition

$$\mathcal{R}(\beta) = \sup\left\{ \prod_{i=1}^{n}(nw_i) \,\Big|\, w_i \geqslant 0,\ \sum_{i=1}^{n} w_i = 1,\ \sum_{i=1}^{n} w_i x_i (y_i - x_i^{\mathsf{T}}\beta) = 0 \right\}$$

$$\mathcal{P} = \mathcal{P}(\beta) = \{x_i \mid y_i - x_i^{\mathsf{T}}\beta > 0\}$$ x with pos resid

$$\mathcal{N} = \mathcal{N}(\beta) = \{x_i \mid y_i - x_i^{\mathsf{T}}\beta < 0\}$$ x with neg resid

Convex hull condition O (2000)

$$\mathrm{chull}(\mathcal{P}) \bigcap \mathrm{chull}(\mathcal{N}) \neq \varnothing \implies \beta \in C(0)$$

For $x_i = (1, t_i)^{\mathsf{T}} \in \mathbb{R}^2$ $\mathcal{P}$ and $\mathcal{N}$ are intervals in $\{1\} \times \mathbb{R}$.

Converse

Suppose that $\tau \notin \{t_1, \ldots, t_n\}$ and

$$\text{Sign}(y_i - \beta_0 - \beta_1 t_i) = \begin{cases} 1 & t_i > \tau \\ -1 & t_i < \tau \end{cases}$$

Suppose also that

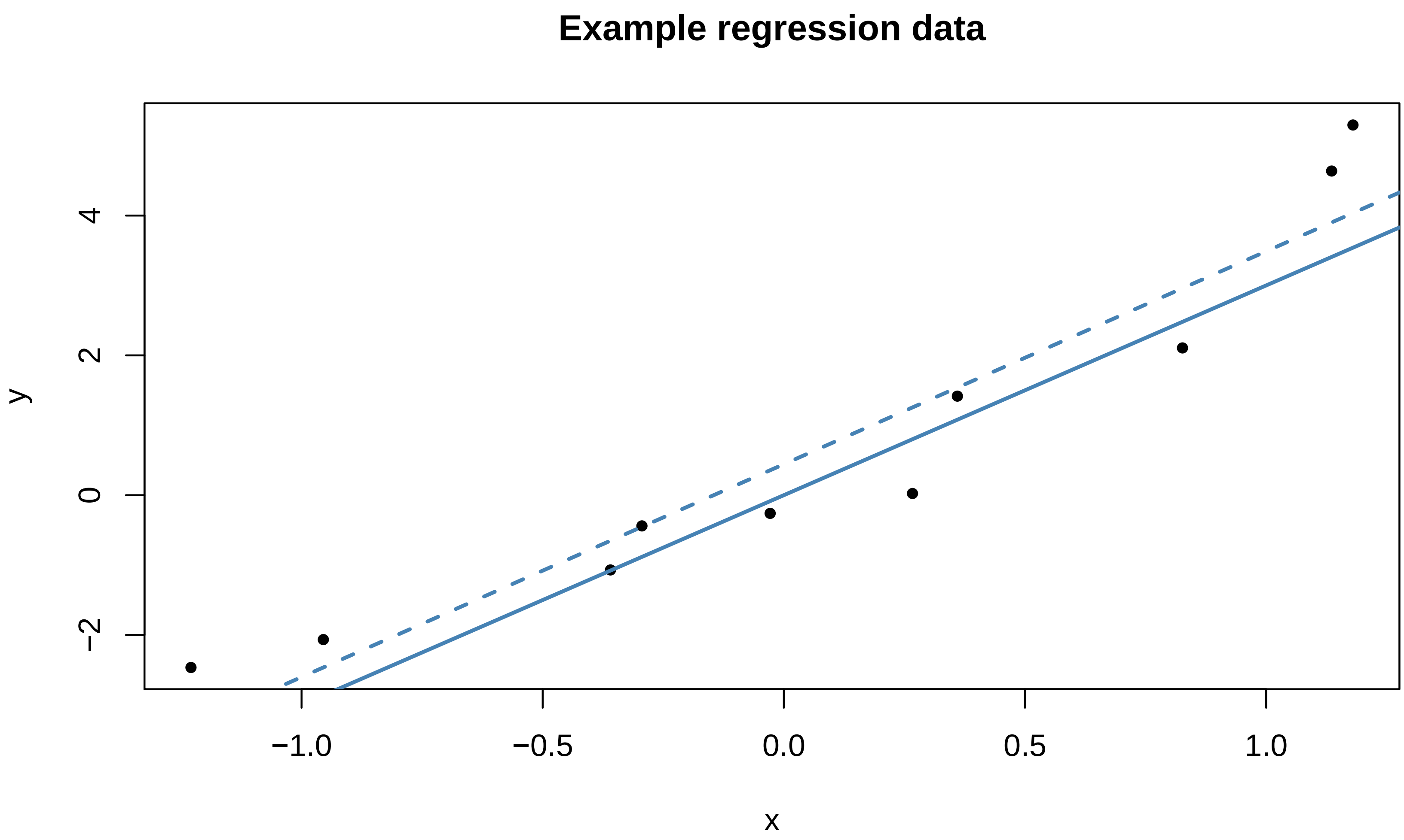
$$\sum_i w_i \begin{pmatrix} 1 \\ t_i \end{pmatrix} (y_i - \beta_0 - \beta_1 t_i) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Then

$$\sum_i w_i (y_i - \beta_0 - \beta_1 t_i)(t_i - \tau) = 0$$

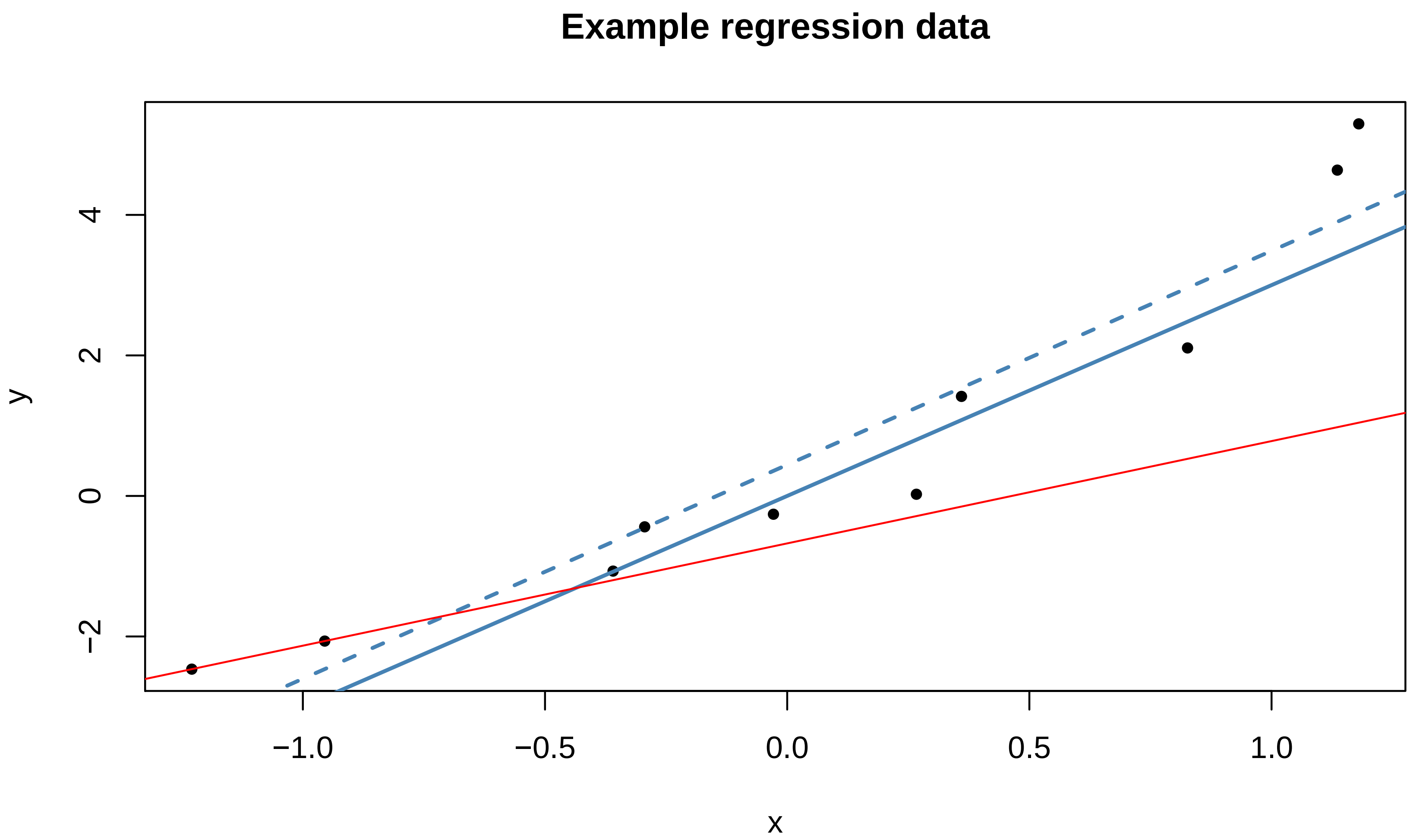
But $(y_i - \beta_0 - \beta_1 t_i)(t_i - \tau) > 0 \; \forall i$

Therefore the hull condition is necessary.

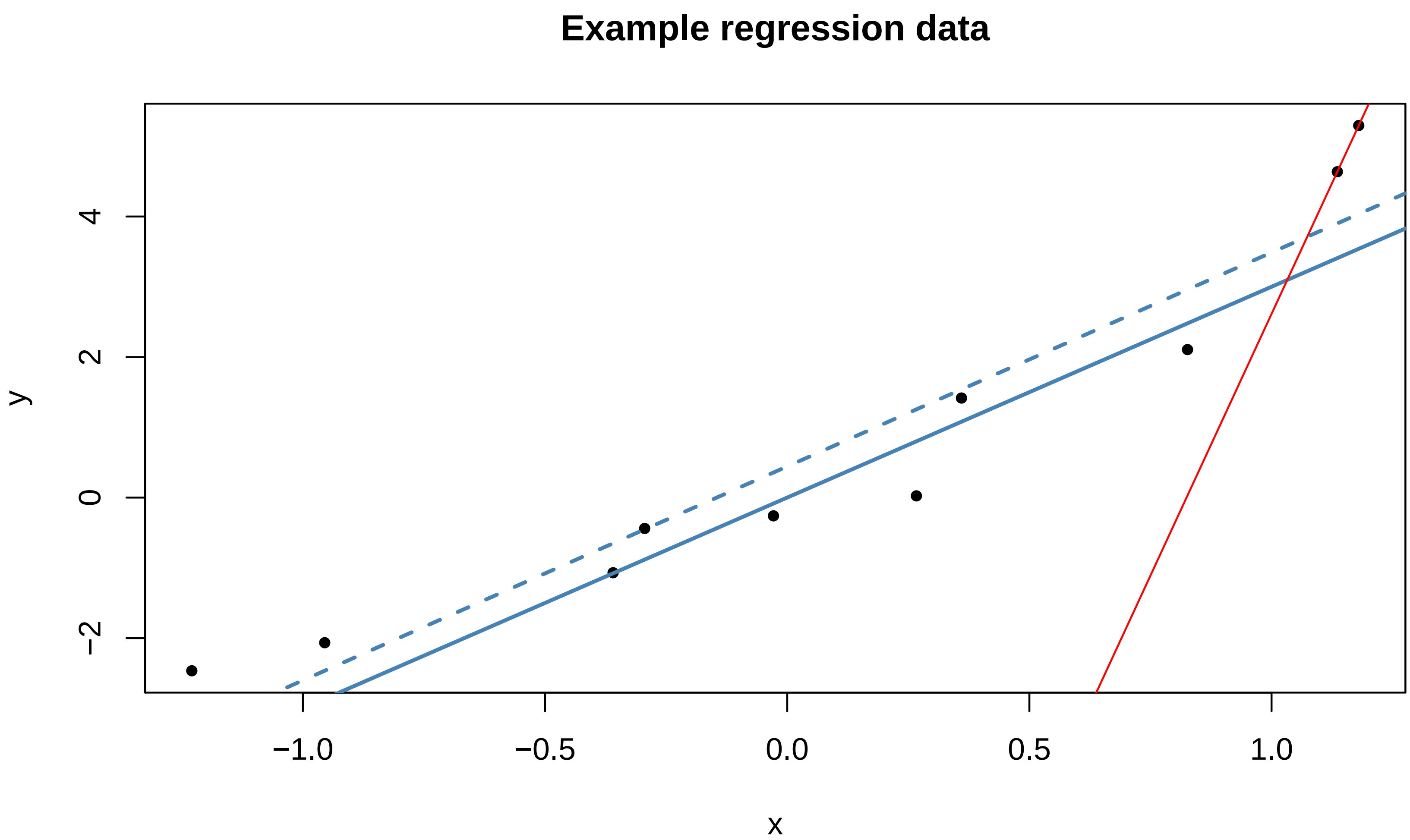


$Y = \beta_0 + \beta_1 X + \sigma\epsilon \quad \beta = (0,3)^\mathsf{T}, \sigma = 1$

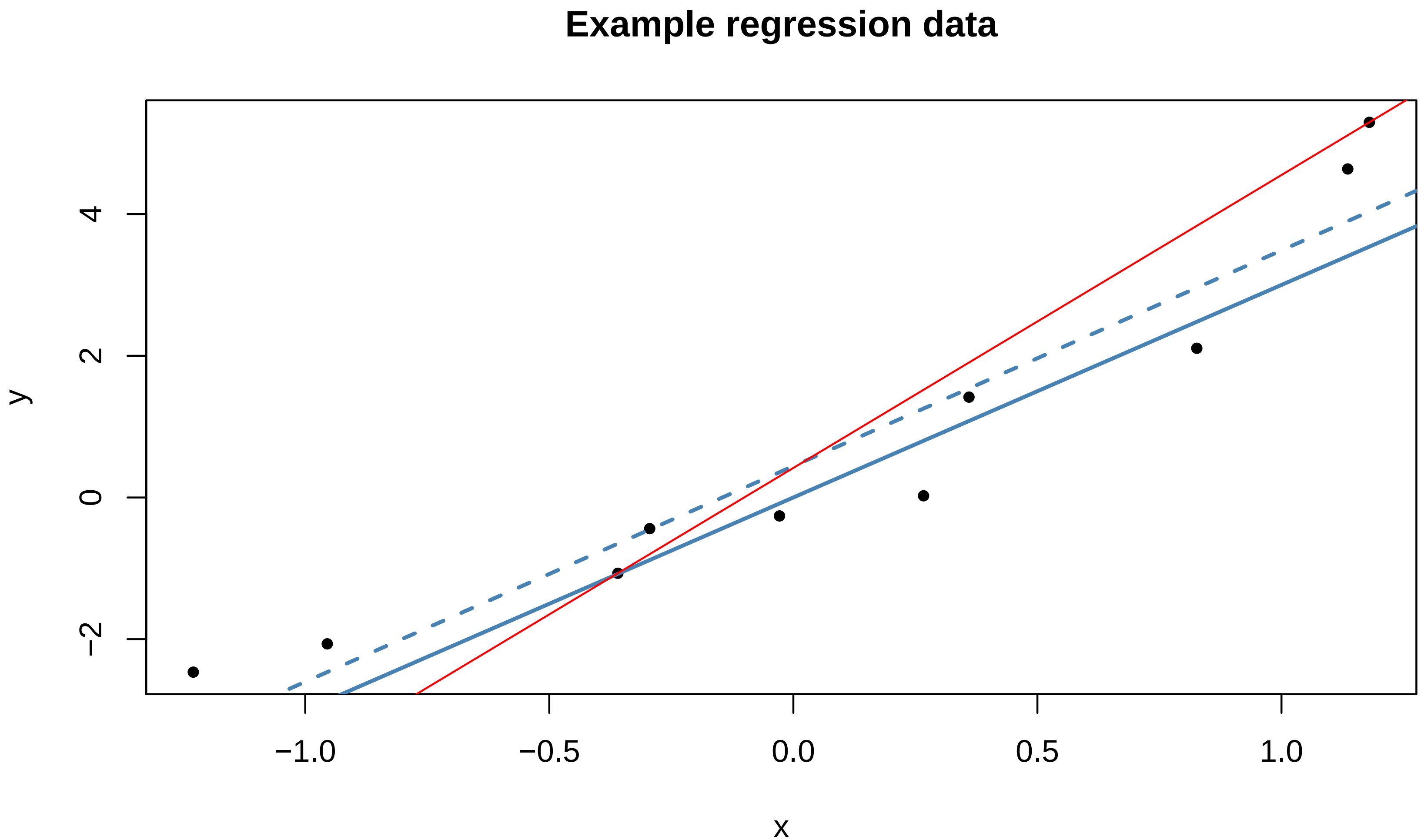
β solid $\quad \hat{\beta}$ dashed



Red line is on boundary of set of (β_0, β_1) with positive empirical likelihood



Another boundary line.

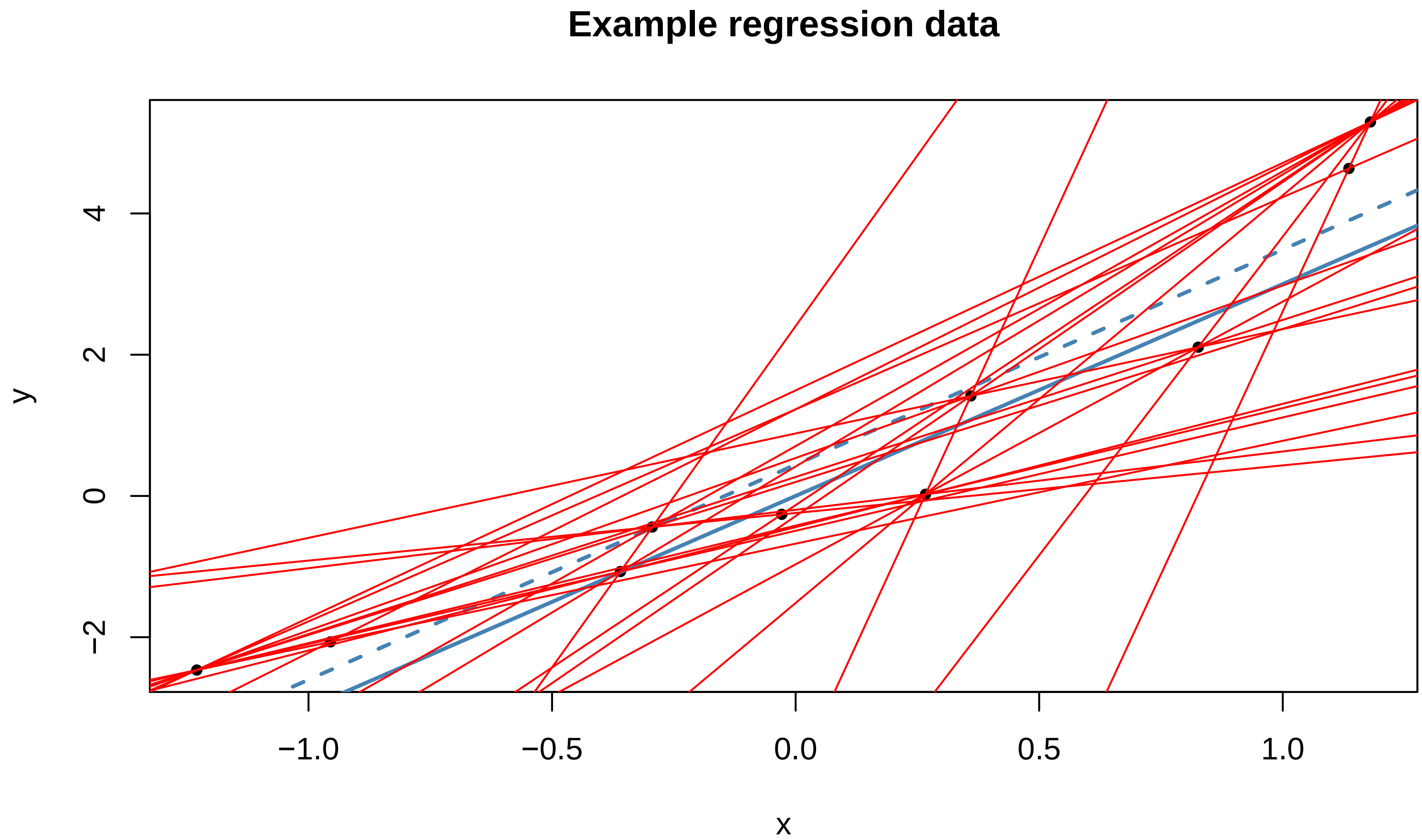


Yet another boundary line.

Left side has positive residuals; right side negative.

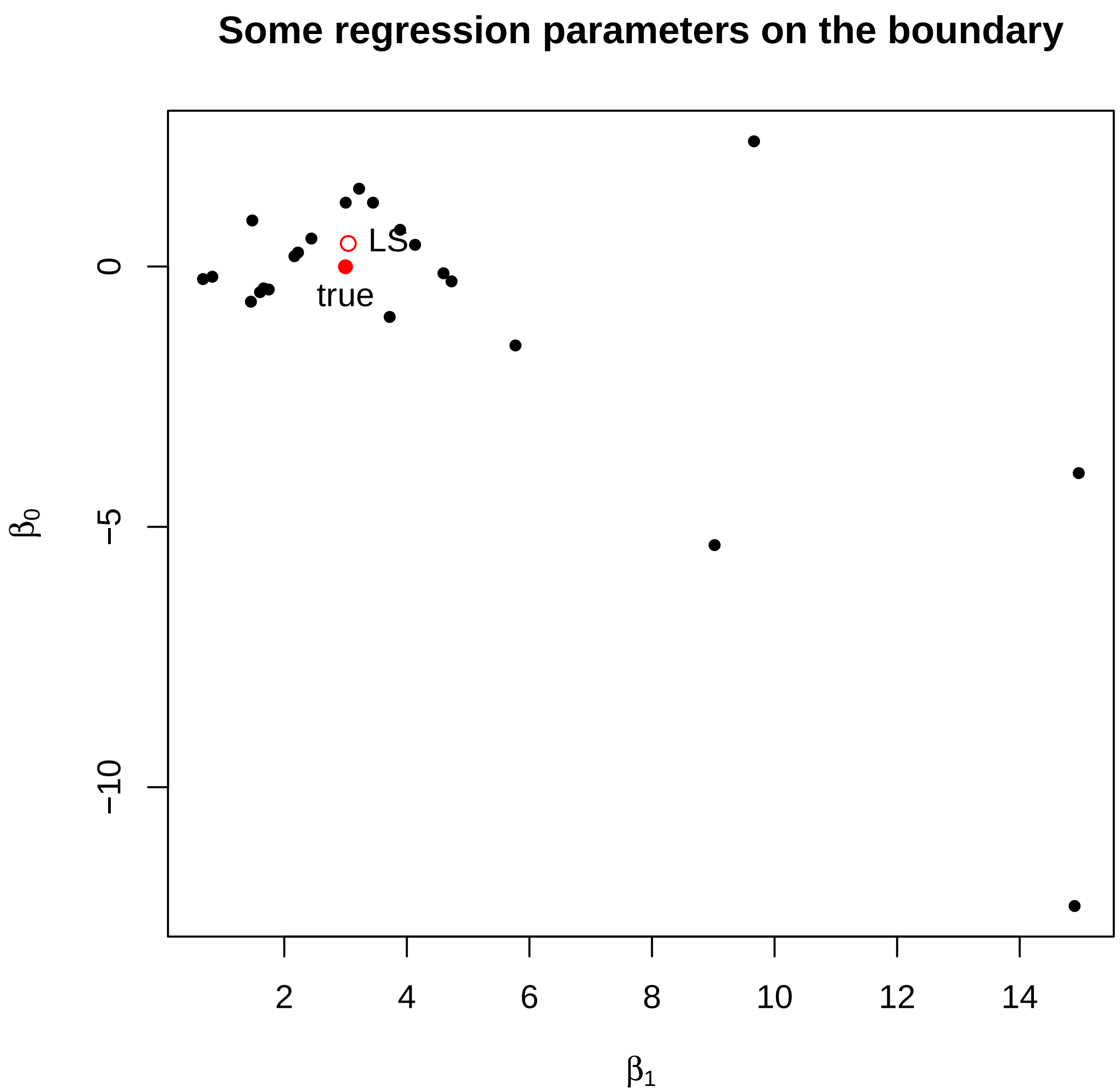
Wiggle it up and point 3 gets a negative residual $\Longrightarrow$ ok.

Wiggle down $\Longrightarrow$ NOT ok.



All the boundary lines that interpolate two data points.

They are a subset of the boundary.



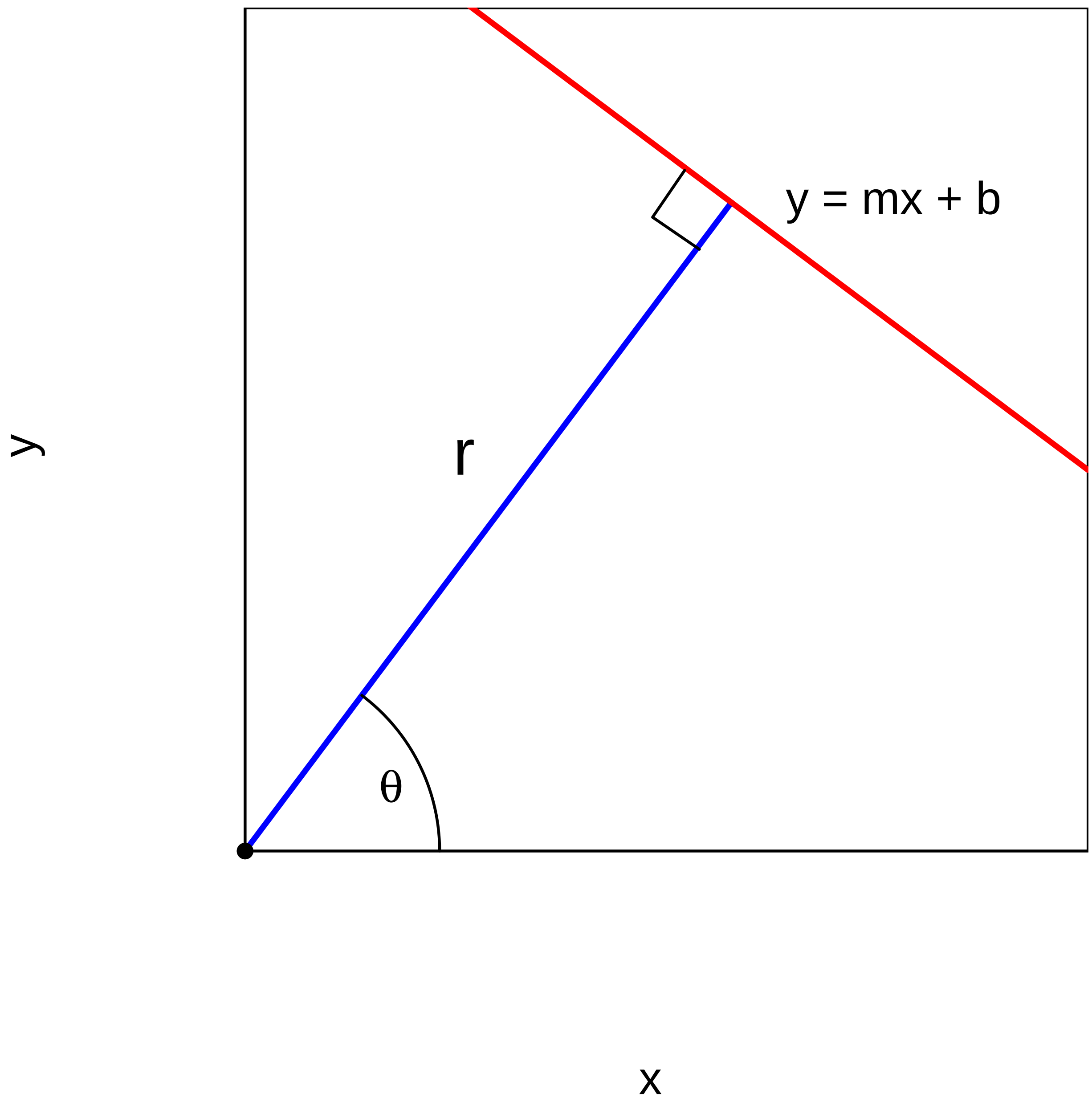
Boundary points (β_0, β_1). Region is not convex.

It **is** convex in β_0 (vertical) for fixed β_1 (horizontal).

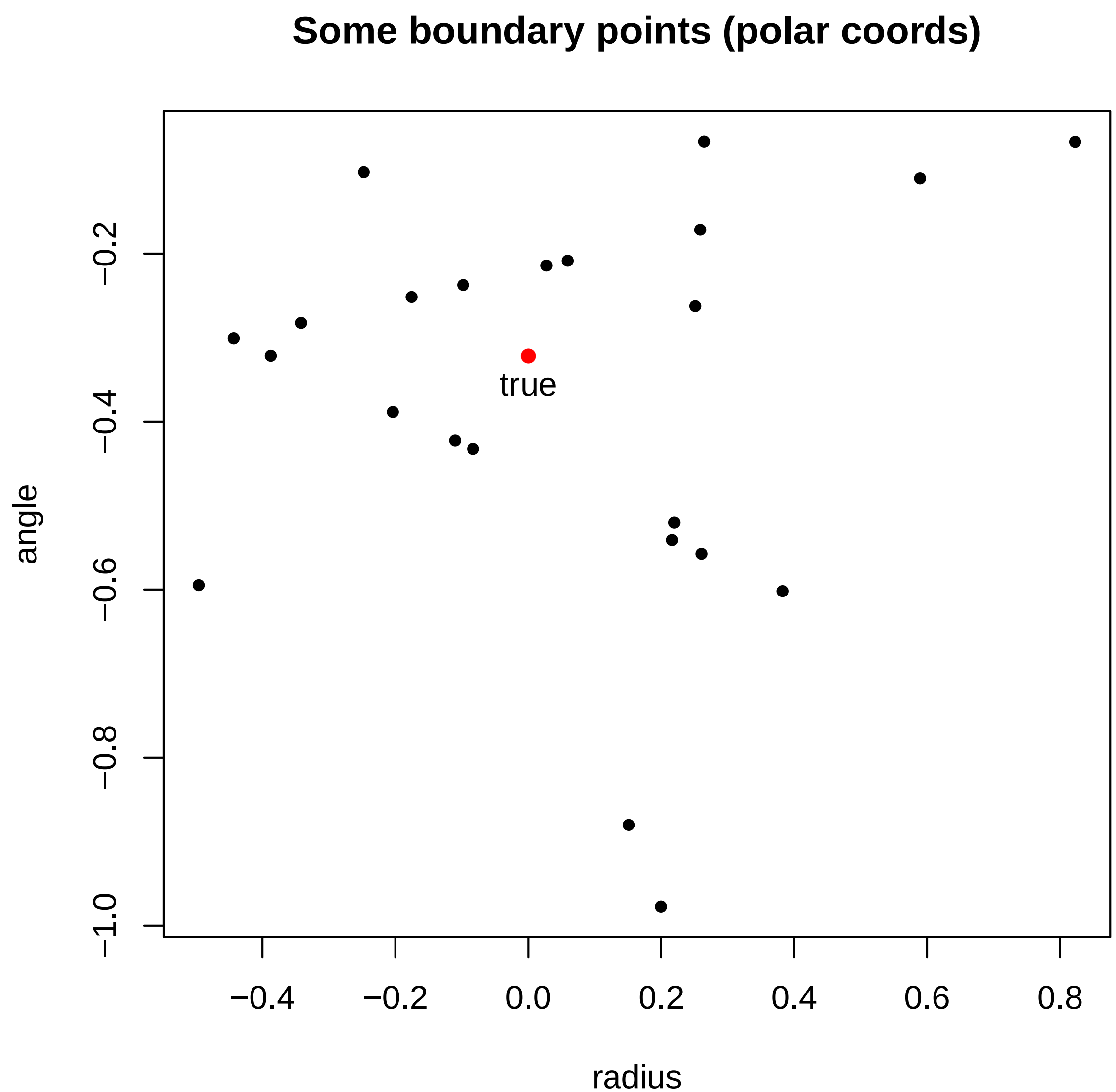
What is a convex set of lines?

- convex set of (β_0, β_1)?
- convex set of (ρ, θ)? (polar coordinates)
- convex set of (a, b) ($ax + by = 1$)?

Polar coordinates of a line



Boundary pts in polar coords



Not convex here either.

Intrinsic convexity

There is a geometrically intrinsic notion for a convex set of linear flats.

J. E. Goodman (1998) “When is a set of lines in space convex?”

Maybe . . . that can support some computation.

Dual definition

The set of flats that intersects a convex set $C \subset \mathbb{R}^d$ is a convex set of flats.

So is the set of flats that intersect **all of** $C_1, \ldots, C_k \subset \mathbb{R}^d$ for convex C_j.

Thanks

- Sarah Emerson
- Invitation: Heping Zhang
- Contacts: Mingxiu Hu, Tianxi Cao, Hongliang Shi & Naitee Ting
- NSF DMS-0906056