Should Bayesians and frequentists calibrate their parameters?

D Fraser
Statistics Dept
U of Toronto

December 8 2005
U of Western Ontario

www.utstat.toronto.edu/dfraser
- - - - - - - - - - - - /documents/xxx

Calibrate?

Calibrate parameters? B | (μ, σ) $\frac{d\mu d\sigma}{\sigma}$

f | σ σ^2 $\log \sigma^2$

Conceptualize parameter? what is real interest parameter?

Measure of departure of
data from parameter value?

Common issues... B & f ?

Claim: If Bayesians & frequentists calibrate
then they'll agree.... same conclusions

xxx = westerndec5

Does it matter?

Does it make a difference?

- Examples
- There are issues
- Calibration is central
- There are methods

Thanks
Acknowledgements to
Many colleagues —

B <u>Example 1</u>

One Normal

$y_1, \ldots, y_n \quad N \quad \mu \quad \sigma^2 \qquad$ or $\quad \underset{\sim}{y} = X\beta + \sigma \underset{\sim}{z} \qquad$ N 0 1
Student(7)

Interest in ... σ

Calibrate? $\quad \sigma \quad \sigma^2 \quad \log \sigma^2 \qquad$ "same" parameter

$$\hat{\sigma}^2 = n^{-1} \Sigma (y_i - \hat{y}_i)^2$$
$$s_y^2 = (n-p)^{-1} \Sigma (y_i - \bar{y})^2$$

N case

Unbiasedness for σ^2... for σ ?

II Examples

1) One Normal

$y_1, \dots, y_n$ N μ σ^2 or $\underset{\sim}{y} = X\beta + \sigma \underset{\sim}{z}$ N 0 1

Student (7)

Interest in ... σ

f Calibrate? σ σ^2 $\log \sigma^2$ "same" parameter

$\hat{\sigma}^2 = n^{-1} \Sigma (y_i - \hat{y}_i)^2$ | N case

$s_y^2 = (n-p)^{-1} \Sigma (y_i - \bar{y})^2$ | Unbiasedness for σ^2 ... for σ ?

Bayes? $\sigma^{-2} d\mu d\sigma$ $\sigma^{-1} d\mu d\sigma$ (Left inv., Right inv.)

II Examples

1) One Normal

$y_1, \dots, y_n \quad N \; \mu \; \sigma^2$ or $\underset{\sim}{y} = X\beta + \sigma \underset{\sim}{z}$ $\quad$ $N \; 0 \; 1$
Student(7)

Interest in ... σ

Calibrate? $\quad \sigma \quad \sigma^2 \quad \log \sigma^2$ "same" parameter

$$\hat{\sigma}^2 = n^{-1} \Sigma (y_i - \hat{y}_i)^2$$
$$s_y^2 = (n-p)^{-1} \Sigma (y_i - \bar{y})^2$$

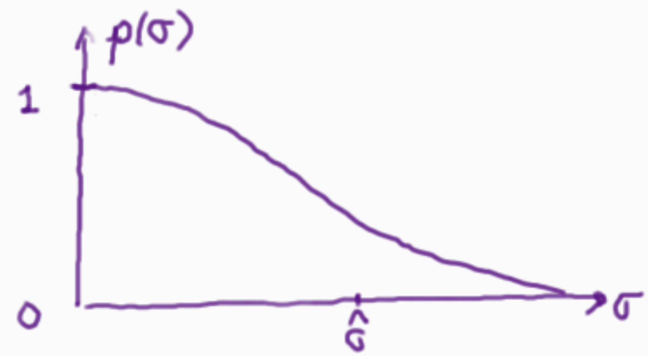
| N case
| Unbiasedness for σ^2 ... for σ ?

Bayes? $\quad \sigma^{-2} d\mu d\sigma \quad \sigma^{-1} d\mu d\sigma$ (Left inv, Right inv.)

p-value

$$p(\sigma) = P\{ \text{pivotal} \le \text{obs. pivotal} \}$$
$$= P\left\{ \frac{\Sigma(y_i - \hat{y}_i)^2}{\sigma^2} \le \frac{\Sigma(y_i^0 - \hat{y}_i^0)}{\sigma^2} \right\}$$

$\uparrow \chi^2_{n-p}$ in N case



Measure of departure ... built into pivotal | Intrinsic (continuity)

p-value = Confidence distribution function

C Example 2

Two normals

$y_{11}, \dots, y_{1n}$ $N\ \mu_1\ \sigma_1^2$ Interest in $\delta = \mu_1 - \mu_2$

$y_{21}, \dots, y_{1m}$ $N\ \mu_2\ \sigma_2^2$

Calibration? Difference in means ... ! Obvious!

f

$$z = \frac{\bar{y}_1 - \bar{y}_2}{\left\{ \frac{\Sigma(y_{1i} - \hat{y}_{1i})^2}{n(n-1)} + \frac{\Sigma(y_{2i} - \hat{y}_2)^2}{m(m-1)} \right\}^{1/2}}$$

cf "?" (Long standing issue!)

Behrens 1929 Fisher 1934

II Examples

2) Two normals

$y_{11}, \ldots, y_{1n}$ $N\ \mu_1\ \sigma_1^2$ Interest in $\delta = \mu_1 - \mu_2$

$y_{21}, \ldots, y_{2m}$ $N\ \mu_2\ \sigma_2^2$

Calibration? Difference in means ... ! Obvious !

f $$z = \frac{\bar{y}_1 - \bar{y}_2}{\left\{ \frac{\Sigma(y_{1i} - \hat{y}_{1i})^2}{n(n-1)} + \frac{\Sigma(y_{2i} - \hat{y}_2)^2}{m(m-1)} \right\}^{1/2}}$$ cf ? Long standing issue

B $$\pi(\theta) = d\mu_1 d\mu_2 \frac{d\sigma_1\, d\sigma_2}{\sigma_1^2 \sigma_2^2} \quad \text{or} \quad d\mu_1 d\mu_2 \frac{d\sigma_1\, d\sigma_2}{\sigma_1 \sigma_2}$$

Left inv. / Jeffreys

Right inv. / modified Jeffreys / Fisher

Behrens 1929 Fisher 1934 Jeffreys 1939 1946

II Examples

2) Two normals

$y_{11}, \ldots, y_{1n}$ $N\ \mu_1\ \sigma_1^2$

$y_{21}, \ldots, y_{2m}$ $N\ \mu_2\ \sigma_2^2$

Interest in $\delta = \mu_1 - \mu_2$

Calibration ? Difference in means ... ! Obvious !

f

$$z = \frac{\bar{y}_1 - \bar{y}_2}{\left\{ \frac{\Sigma (y_{1i} - \hat{y}_{1i})^2}{n(n-1)} + \frac{\Sigma (y_{2i} - \hat{y}_2)^2}{m(m-1)} \right\}^{1/2}}$$

cf ? Long standing issue

B

$$\pi(\theta) = d\mu_1 d\mu_2 \frac{d\sigma_1\, d\sigma_2}{\sigma_1^2 \sigma_2^2} \quad \text{or} \quad d\mu_1 d\mu_2 \frac{d\sigma_1\, d\sigma_2}{\sigma_1 \sigma_2}$$

Left inv. / Jeffreys

Right inv. / modified Jeffreys / Fisher

$$\pi(\theta) = d\mu_1 d\mu_2 \frac{d\sigma_1\, d\sigma_2}{\sigma_1^3 \sigma_2^3 \left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right)^{-1}}$$

Ghosh Kim (2001)

Behrens 1929 Fisher 1934 Jeffreys 1939 1946 Ghosh Kim 2001 (CJS)

Example Behrens–Fisher 1929 1934 Ghosh Kim 2001

$y_{11}, \dots, y_{1n}$ N μ_1 σ_1^2 Interest $\psi = \mu_1 - \mu_2$

$y_{21}, \dots, y_{2n}$ N μ_2 σ_2^2

GK: $\pi = \sigma_1^{-3} \sigma_2^{-3} \left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right)$ vs Jeffreys(right): $\sigma_1^{-1} \sigma_2^{-1}$

Some simulation numbers: Case

n	m	σ_1^2	σ_2^2	ψ
2	2	2	1	2

Nominal 5% 95% points by various methods

Example Behrens–Fisher 1929 1934 Ghosh Kim 2001

$y_{11}, \dots, y_{1n}$ N μ_1 σ_1^2

$y_{21}, \dots, y_{2n}$ N μ_2 σ_2^2

Interest $\psi = \mu_1 - \mu_2$

GK: $\pi = \sigma_1^{-3} \sigma_2^{-3} \left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right)$ vs Jeffreys(right): $\sigma_1^{-1} \sigma_2^{-1}$

Some simulation numbers: Case

n	m	σ_1^2	σ_2^2	ψ
2	2	2	1	2

Nominal 5% 95% points by various methods

	Nominal	5%	95%
B	Jeffreys	.7%	99.1%
B	Kim Ghosh	1.7%	97.9%
f	Lik ratio	13.2%	86.9%

Example Behrens–Fisher 1929 1934 Ghosh Kim 2001

$y_{11}, \dots, y_{1n}$ N μ_1 σ_1^2

$y_{21}, \dots, y_{2n}$ N μ_2 σ_2^2

Interest $\psi = \mu_1 - \mu_2$

GK: $\pi = \sigma_1^{-3} \sigma_2^{-3} \left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right)$ vs Jeffreys(right): $\sigma_1^{-1} \sigma_2^{-1}$

Some simulation numbers: Case

n	m	σ_1^2	σ_2^2	ψ
2	2	2	1	2

Nominal 5% 95% points by various methods including flat prior

Nominal	5%	95%
Jeffreys	.7%	99.1%
Kim Ghosh	1.7%	97.9%
Lik ratio	13.2%	86.9%
$p(\psi), s(\psi)$	4.23%	95.8%
Sim 95% limits	(4.86, 5.14)	(94.9, 95.14)

flat prior
data dep.
$O(n^{-3/2})$
$B \equiv f$

$N = 100{,}000$

Augustine Wong York U (2004)

D Data dependent priors

- Sounds like a contradiction?
- Box & Cox 1964 $\quad y^{\lambda} = X\beta + \sigma \underset{\sim}{z} \qquad z_i \; N \; 0 \; 1$

  Used data dependent prior ---- Later...
- Cox: Prior depend on data? Of course!
- Wasserman 2000

but Why?

D Data dependent priors

- Sounds like a contradiction?
- Box & Cox 1964 $\quad y^{\lambda} = X\beta + \sigma \underset{\sim}{z} \quad z_i \; N \; 0 \; 1$

  Used data dependent prior
- Cox: Prior depend on data? Of course!
- Wasserman 2000

Prior depend on data ?

Prior depend on model ? .... default prior (objective ???)

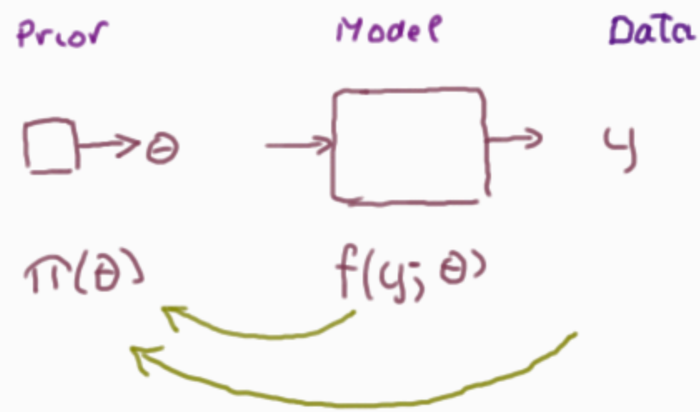
D Data dependent priors

- Sounds like a contradiction?
- Box & Cox 1964 $y^{\lambda} = X\beta + \sigma \underset{\sim}{z}$ z_i N 0 1

  Used data dependent prior
- Cox: Prior depend on data? Of course!
- Wasserman 2000

Prior depend on data?

Prior depend on model? .... default prior (objective ???)



?

D Data dependent priors

Should: Bayesian right tail posterior probabilities
be equal to p-values i.e. have frequentist properties?

Widely addressed: "matching" in B community!

Judith Rousseau ... "important to be as close as possible
to the uniform distribution"

D Matching:

Pure case: Scalar, scalar θ $\qquad f(y;\theta)$

Implications!

"p-value = right tail posterior probability"

$$p^{\circ}(\theta) = F(y^{\circ};\theta) = \int_{\theta}^{\infty} L^{\circ}(\theta)\,\pi(\theta)\,d\theta$$

D Matching:

Pure case: Scalar, scalar θ $\qquad f(y;\theta)$

p-value = right tail posterior probability

$$p^0(\theta) = F(y^0;\theta) = \int_\theta^\infty L^0(\theta)\,\pi(\theta)\,d\theta$$

$$F(y^0;\theta) = \int_\theta^\infty c\,f(y^0;\theta)\,\pi(\theta)\,d\theta$$

Differentiate:

$$F_{;\theta}(y^0;\theta) = -c\,F_y(y^0;\theta)\,\pi(\theta)$$

$$\pi(\theta) = -c\,\frac{F_{;\theta}(y^0;\theta)}{F_y(y^0;\theta)}$$

Needed Data-dependent model-based prior

D Matching:

Pure case: Scalar, scalar θ $\qquad f(y;\theta)$

p-value = right tail posterior probability

$$p^{\circ}(\theta) = F(y^{\circ};\theta) = \int_{\theta}^{\infty} L^{\circ}(\theta)\,\pi(\theta)\,d\theta$$

$$F(y^{\circ};\theta) = \int_{\theta}^{\infty} c\,f(y^{\circ};\theta)\,\pi(\theta)\,d\theta$$

Differentiate

$$F_{;\theta}(y^{\circ};\theta) = -c\,F_y(y^{\circ};\theta)\,\pi(\theta)$$

$$\pi(\theta) = -c\,\frac{F_{;\theta}(y^{\circ};\theta)}{F_y(y^{\circ};\theta)} = \left.\frac{dy}{d\theta}\right|_{\substack{\text{data}\\\text{fixed pivot}}}$$

Need: Data-dependent prior $\Rightarrow$ "matching"

Note - prior is flat if model is location

- $F(y;\theta) \sim U(0,1)$ .... is pivotal
- $\boxed{\text{Prior} = \left.\frac{dy}{d\theta}\right|_{y^{\circ}} \quad \text{(with fixed pivot)}}$

E Inference elements

$f(y;\theta) + y^\circ \Rightarrow$?

Case: Scalar y & θ ... Where things should start!

Obvious: $f^\circ = f(y^\circ;\theta)$ $\quad L^\circ(\theta)$ $\quad$ Prob at

$F^\circ = F(y^\circ;\theta)$ $\quad p^\circ(\theta)$ $\quad$ Prob left

E Inference elements

$f(y;\theta) + y^0 \Rightarrow$?

Case: Scalar y & θ ... Where things should start!

Obvious: $f^0 = f(y^0;\theta)$ $L^0(\theta)$ Prob at

$F^0 = F(y^0;\theta)$ $p^0(\theta)$ Prob left

Q Is there anything more?

E **Inference elements**

$f(y;\theta) + y^0 \Rightarrow$?

Case: Scalar y & θ ... Where things should start!

Obvious: $f^0 = f(y^0;\theta)$ $\quad L^0(\theta) \quad$ Prob at

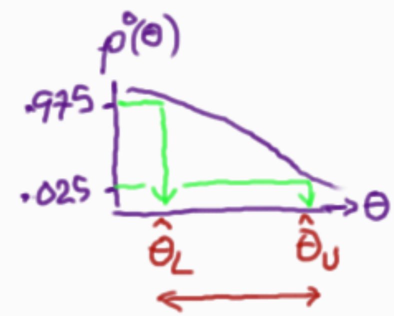
$F^0 = F(y^0;\theta)$ $\quad p^0(\theta) \quad$ Prob left

Q Is there anything more?

$u = F(y;\theta)$ is a pivot

Invert; get all confidence bounds, conf. int's

$95\%\ CI = \{p^{-1}(.975),\ p^{-1}(.025)\}$



A It's all there!

Inference & what more than likelihood

$f(y;\theta) + y^0 \Rightarrow$?

Case: Scalar y & θ ... Where things should start!

Obvious:	$f^0 = f(y^0;\theta)$	$L^0(\theta)$	Prob at
	$F^0 = F(y^0;\theta)$	$p^0(\theta)$	Prob left

Is there anything more?

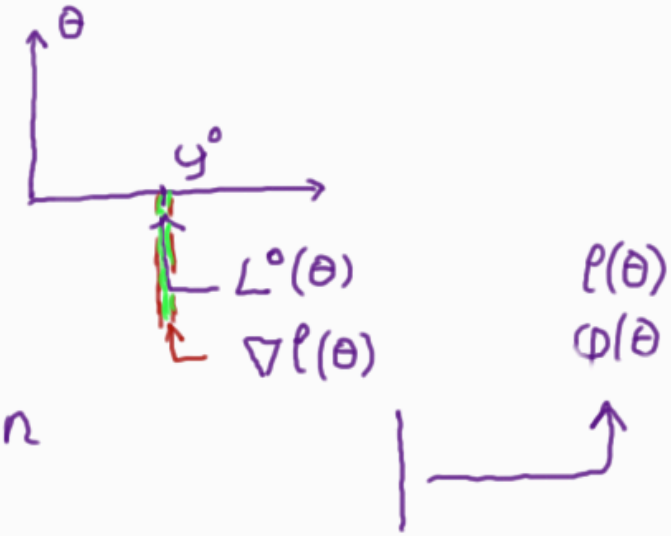
$u = F(y;\theta)$ is a pivot

Invert; get all confidence bounds, conf. int's

$$95\%\ CI = \{p^{-1}(.975),\ p^{-1}(.025)\}$$

Can do in general;

Min. assumptions!



But: you need more than likelihood

HOL: overview Fraser, Reid, Wong 2004 xxx = 231

Inference: Need more than Likelihood !

$f(y;\theta) + y^0 \Rightarrow \ ?$

y in R^n $\quad \theta$ in R^p $\qquad$ moderate regularity

Inference: Need more than Likelihood !

$f(y;\theta) + y^0 \Rightarrow$?

y in R^n θ in R^p moderate regularity

Need: $\ell(\theta; y^0) = a + \log f(y^0; \theta)$

$\varphi(\theta) =$ gradient likelihood in directions $V = (\underset{\sim}{v}_1, \ldots, \underset{\sim}{v}_p)$
tangent to "conditioning"

Inference: Need more than Likelihood !

$f(y;\theta) + y^0 \Rightarrow$?

y in R^n $\quad \theta$ in R^p $\qquad$ moderate regularity

Need: $\quad \ell(\theta; y^0) = a + \log f(y^0;\theta)$

$\varphi(\theta) =$ gradient likelihood in directions $V = (\underset{\sim}{v}_1, \ldots, \underset{\sim}{v}_p)$
tangent to "conditioning"

Get $\quad$ p-value for any scalar $\psi(\theta)$ $\quad$ 3rd

likelihood " " " or vector $\psi(\theta)$ $\quad$ 3rd

p-value for any vector $\psi(\theta)$ $\quad$ 2nd

Uses just mle's, info's & CPU ...

Fraser Reid Wu BKa '99 $\qquad$ $xxx = 196$

F Calibrate: priors

Get: 3rd order inference using

$V = \frac{dy}{d\theta}\Big|_{(y^0, \hat\theta)}$ Differentiate for fixed pivotal

Get: 3rd order matching prior

$V(\theta) = (\underset{\sim}{v}_1(\theta), \ldots, \underset{\sim}{v}_p(\theta)) = \frac{dy}{d\theta}\Big|_{(y^0, \theta)}$

$\text{Prior} = c\,|M' V(\theta)|$

$M = \frac{\partial^2}{\partial y \partial \theta} \ell(\theta; y)\Big|_{\hat\theta^0, y^0}$

$= \ell_{\underset{\sim}{y}, \theta}(\hat\theta^0; y^0)$

F Calibrate: priors

Got: 3rd order inference using

$$V = \left.\frac{dy}{d\theta}\right|_{(y^0, \hat\theta)}$$

Differentiate for fixed pivotal

Get: 3rd order matching prior

$$V(\theta) = (\underset{\sim}{v}_1(\theta), \ldots, \underset{\sim}{v}_p(\theta)) = \left.\frac{dy}{d\theta}\right|_{(y^0, \theta)}$$

$$\text{Prior} = c\,|M' V(\theta)|\, d\theta$$

$$M = \left.\frac{\partial^2}{\partial y\, \partial\theta}\ell(\theta; y)\right|_{\hat\theta^0, y^0} = \ell_{\underset{\sim}{y}, \theta}(\hat\theta^0; y^0)$$

Example: $\underset{\sim}{y} = X\beta + \sigma \underset{\sim}{z}$ z_i Student or Extreme value

$$\text{Prior} = c\,|M' V(\theta)|\, d\theta = d\underset{\sim}{\beta}\,\frac{d\sigma}{\sigma}$$

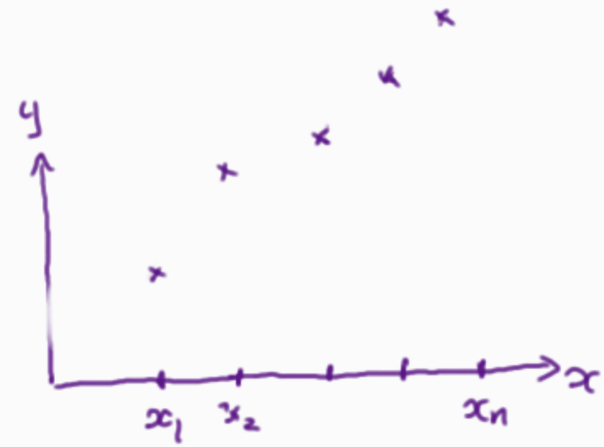
and also gives confidence

Fraser, Reid, Marras (Rome) Yi (Waterloo) 2005

G Example: Box & Cox 1964

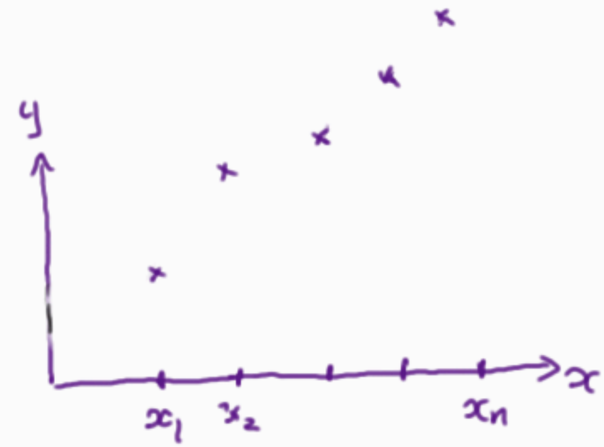
Finding the parameter

<u>Model</u> $y^\lambda = \alpha + \beta x + \sigma z$



G Example: Box & Cox 1964

Finding the parameter



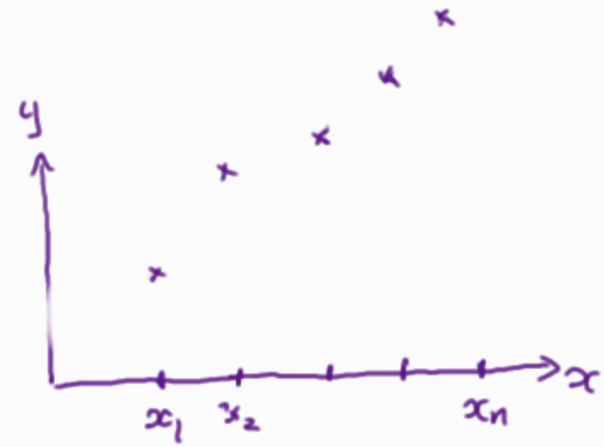
<u>Model</u> $y^\lambda = \alpha + \beta x + \sigma z$

a) If you knew λ
just ordinary regression

Box, Cox 1964 Fraser Wong 2005

G Example: Box & Cox 1964

Finding the parameter



Model $y^{\lambda} = \alpha + \beta x + \sigma z$

a) If you knew λ
just ordinary regression

Psychologist is uncertain as to how to express/calibrate/scale response y

Possibly: y^2 or $y^{1/2}$ or $\log y$ to get: Lin model, Normality, homosced... | $y>0$

try y^{λ}

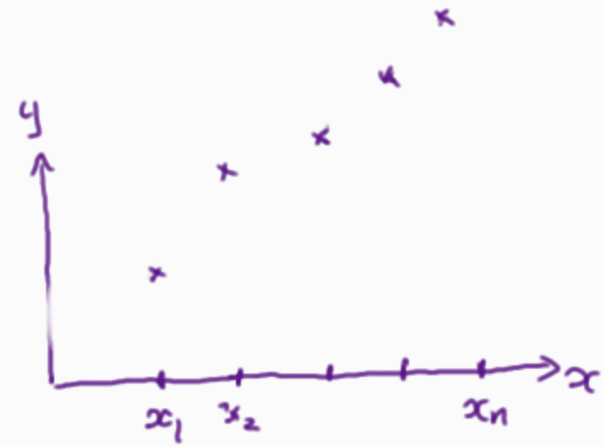
Model: $y = (\alpha + \beta x + \sigma z)^{1/\lambda}$ ... say $z \sim N$

G Example: Box & Cox 1964

Finding the parameter

<u>Model</u> $y^\lambda = \alpha + \beta x + \sigma z$

a) If you knew λ
just ordinary regression



Psychologist is uncertain as to how to express/calibrate/scale response y
Possibly y^2 or $y^{1/2}$ or $\log y$ to get Lin model, Normality, homosced ...

Try y^λ "$y > 0$"

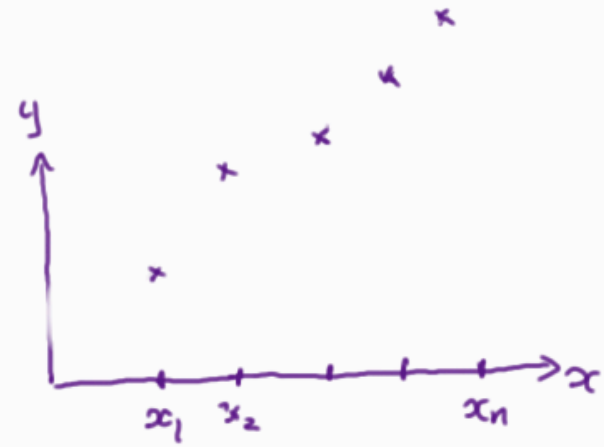
Model: $y = (\alpha + \beta x + \sigma z)^{1/\lambda}$ ... say $z \sim N$

but what does β mean ? Box Cox 1964 Bickel Doksum 1981 Box Cox 1982 !

Box, Cox 1964 Fraser Wong 2005

G Example: Box & Cox 1964

Finding the parameter



<u>Model</u> $\quad y^\lambda = \alpha + \beta x + \sigma z$

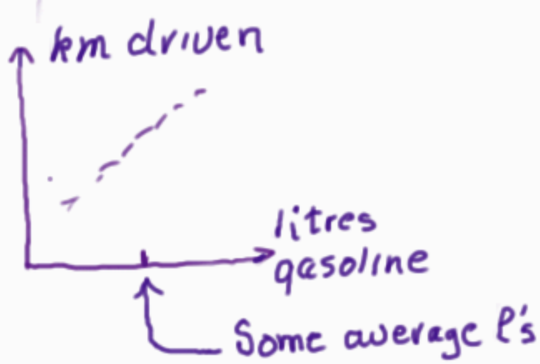
a) If you knew λ
just ordinary regression

Psychologist is uncertain as to how to express/calibrate/scale response y
Possibly y^2 or $y^{1/2}$ or $\log y$ to get Lin model, Normality, homosced...

Try $\quad y^\lambda$ $\qquad$ "$y > 0$"

Model: $\quad y = (\alpha + \beta x + \sigma z)^{1/\lambda}$ $\quad$ ... say $z \sim N$

<u>Example</u>



Interest?
How "km" change with x change

$\psi(\theta) = \frac{d}{dx} (\alpha + \beta x)^{1/\lambda} \Big|_{x_0}$ $\quad$ <u>real</u> parameter

<u>So</u> get p-value.

Chen Lockhart Stephens 2002 $\quad$ F & Wong 2005

A message:

Calibrate parameter

1) Need : likelihood, gradient of likelihood f

get : 3rd order p-value

2) Need : likelihood, pivotal B

get : 3rd order posterior tail probs

3) Bayes & likelihood agree ♡ $(f\ B)$