Gamma distribution:

$Y \sim \Gamma(\alpha, \beta)$ $\quad \left[f(y) = \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} e^{-\beta y} y^{\alpha-1} \right], \; y > 0$

$\alpha > 0, \; \beta > 0$

Properties

$$\left\{ \Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \right\}$$

- $E(Y) = \dfrac{\alpha}{\beta}$, $\quad var(Y) = \dfrac{\alpha}{\beta^2}$

- $\alpha = 1 \Rightarrow \Gamma(\alpha) = 1$, $\quad f(y) = \beta e^{-\beta y}$; $Y \sim exp(\beta)$

- closed under convolution:

$$\text{ind.} \begin{cases} Y_1 \sim \Gamma(\alpha_1, \beta) \\ Y_2 \sim \Gamma(\alpha_2, \beta) \end{cases} \Rightarrow Y_1 + Y_2 \sim \Gamma(\alpha_1 + \alpha_2, \beta)$$

- Eg $\quad T =$ waiting time until event

$$\sim exp(\lambda) = \Gamma(\underset{\alpha}{1}, \underset{\beta}{\lambda}) \qquad \left(E(T) = \frac{1}{\lambda}\right) \;\; \frac{\alpha}{\beta}$$

$Y = T_1 + \dots + T_m$, $\qquad T_i$ independent

$$\Rightarrow Y \sim \Gamma(1 + 1 \dots + 1, \lambda) = \Gamma(\underset{\alpha}{m}, \underset{\beta}{\lambda}) \qquad E(Y) = m \cdot \frac{1}{\lambda}$$

$\alpha \cdot \beta$

□ In framework of GLIM;

$Y_1, \dots, Y_n \underset{ind}{\sim} \Gamma(\alpha, \beta)$.

$\underset{\sim}{x}_{(1)}, \dots, \underset{\sim}{x}_{(n)}$

Q: $Y_i \sim \begin{cases} \Gamma(\alpha_i, \beta) \\ \Gamma(\alpha, \beta_i) \end{cases}$? $\{ Y_i \sim N(\mu_i, \sigma^2) \}$

A: Want $E(Y_i) = \mu_i$

$$var(Y_i) = \frac{\varphi}{w_i} b''(\theta_i)$$

$$= \varphi \cdot v(\mu_i) \quad \text{for some } v \quad (w_i \equiv 1)$$

□ Have $E(Y) = \dfrac{\alpha}{\beta}$

$$var(Y) = \frac{\alpha}{\beta^2}$$

If $Y_i \sim \Gamma(\alpha_i, \beta)$ then

$$\mu_i = \frac{\alpha_i}{\beta} \text{ and } var(Y_i) = \left(\frac{\alpha_i}{\beta}\right) \cdot \frac{1}{\beta}$$

$$= \underbrace{\left(\frac{1}{\beta}\right)}_{\varphi} \cdot \mu_i, \text{ OK}$$

If $Y_i \sim \Gamma(\alpha, \beta_i)$ then

$$\mu_i = \frac{\alpha}{\beta_i} \text{ and } var(Y) = \frac{\alpha}{\beta_i^2} = \frac{1}{\alpha}\left(\frac{\alpha}{\beta_i}\right)^2$$

$$= \underbrace{\left(\frac{1}{\alpha}\right)}_{\varphi} \cdot \mu_i^2, \text{ also OK}$$

(43)

☐ So need to express Γ pdf as $f(y \mid \theta, \phi)$

{ formula page } $\log f = \dfrac{[y\theta_i - b(\theta_i)]}{\phi} + c(y, \phi)$

$$f(y) = \exp\{\alpha \log(\beta) - \log \Gamma(\alpha) - \beta y + (\alpha-1)\log y\}.$$

$$\{\ \} = \underbrace{y\cdot(-\beta) + \alpha\log(\beta)}_{\frac{y\cdot\theta - b(\theta)}{\phi}} \underbrace{- \log\Gamma(\alpha) + (\alpha-1)\log y}$$

ϕ depends on α ?

$$= \alpha\cdot\left(-\frac{\beta}{\alpha}\right)\cdot y + \alpha\log(\beta) + \{\ \}$$

need to

θ θ?

$$= \frac{\left(-\frac{\beta}{\alpha}\right)\cdot y + \log(\beta)}{\left(\frac{1}{\alpha}\right)} + \{\ \}$$

$\log(\beta) - \log(\alpha)$

$$= \frac{\left(-\frac{\beta}{\alpha}\right)y + \log\left(\frac{\beta}{\alpha}\right)}{\frac{1}{\alpha}} + \frac{\log(\alpha)}{\frac{1}{\alpha}} + \{\ \}$$

$$= \frac{\left(-\frac{\beta}{\alpha}\right)y + \log\left(\frac{\beta}{\alpha}\right)}{(1/\alpha)} + \underbrace{\alpha\log(\alpha) - \log\Gamma(\alpha) + (\alpha-1)\log y}_{c(\frac{1}{\alpha}, y)}$$

☐ So: $\theta_i = -\left(\frac{\beta_i}{\alpha}\right)$ ($\{\theta_i < 0\}$

$\phi = \left(\frac{1}{\alpha}\right)$

$$b(\theta_i) = -\log\left(\frac{\beta_i}{\alpha}\right) = -\log(-\theta_i)$$

□ Can check that

$$b'(\theta_i) = \mu_i = \left(\frac{\alpha}{\beta_i}\right)$$

$$\varphi \cdot b''(\theta_i) = var(y_i) = \frac{1}{\alpha}\left(\frac{\alpha_i}{\beta_i}\right)^2$$

□ Link function is:

canonical ; $g(\mu_i) = \theta_i$

$\Rightarrow \quad g\left(\frac{\alpha}{\beta_i}\right) = -\left(\frac{\beta_i}{\alpha}\right)$

so $g(\mu_i) = -\frac{1}{\mu_i}$ (increasing)

□ ~~II~~ Instead, can use $g(\mu) = \frac{1}{\mu}$

{ μ = expected time to failure, $\frac{1}{\mu}$ = failure rate }

($Y \sim \exp(\lambda)$, $E(Y) = \frac{1}{\lambda}$) so model failure rate in terms of $\underset{\sim}{x}$

□ Problem ; $\frac{1}{\mu} = \underset{\sim}{x}^t \underset{\sim}{\beta} > 0$; numerical constraints

Often use $g(\mu) = \log(\mu) = \log$ [expected time to failure]

☐ Estimation of dispersion parameter ϕ to later (estimation of β does not depend on $\hat{\phi}$)

☐ Example: mismatch data gamma.1.pdf

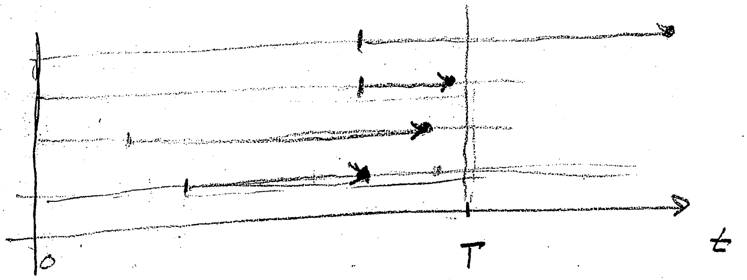
☐ Lin reg (only reject sig)

- $\hat{\beta}$ reject > 0 ??

[reject = 1] ⇒ longer survival time?

Usually reject (eventually); perhaps those who died early died for other reasons, so didn't get chance to reject

- $\hat{\beta}$ mismatch, $\hat{\beta}$ age OK
- $\hat{\beta}$ waiting time - N.S.
- $\hat{\beta}$ cal time < 0;



Data is only for patients who died before end of study. So someone who entered program late will only appear if died early.

☐ Example of bias due to censoring.

☐ Residuals → Γ regression

Comparing linear regression & GLM results

Effect of reject and mis match

	lin. reg y = survival time	Γ: "mortality rate" $\frac{1}{E(y)} = \underset{\sim}{x}^t \underset{\sim}{\beta}$	p-value
reject	1.96	−1.68	0.10
mis-match	−1.48	1.54	0.13

$\underbrace{\text{t-values}}$

- Dispersion parameter, pearson. resids — later
- Convergence problems → "step halving"

For $y \sim \Gamma(\alpha, \beta)$ where $\alpha > 0$, $\beta > 0$,

$$E(y) = \mu = \frac{\alpha}{\beta} > 0.$$

natural parameters: $\theta = -\frac{\beta}{\alpha} = -\frac{1}{\mu} < 0$

$$\phi = \left(\frac{1}{\alpha}\right)$$

- So constraints on sign of θ

But $\underset{\sim}{x}_i^t \underset{\sim}{\beta} = g(\mu_i) = \left\{ \begin{array}{l} -\frac{1}{\mu_i} < 0 \\ \frac{1}{\mu_i} > 0. \end{array} \right\} \Rightarrow$ constraints on $\underset{\sim}{\beta}$.

{If $\underset{\sim}{x}_i^t \underset{\sim}{\beta} = g(\mu_i) = \log(\mu_i)$ then no constraints on $\underset{\sim}{\beta}$}

☐ Question What is "step-halving"?

Why does it lead to convergence?

Consider simplest possible case of one observation $y \sim \Gamma(\alpha, \beta)$ with α known, and no explanatory variables.

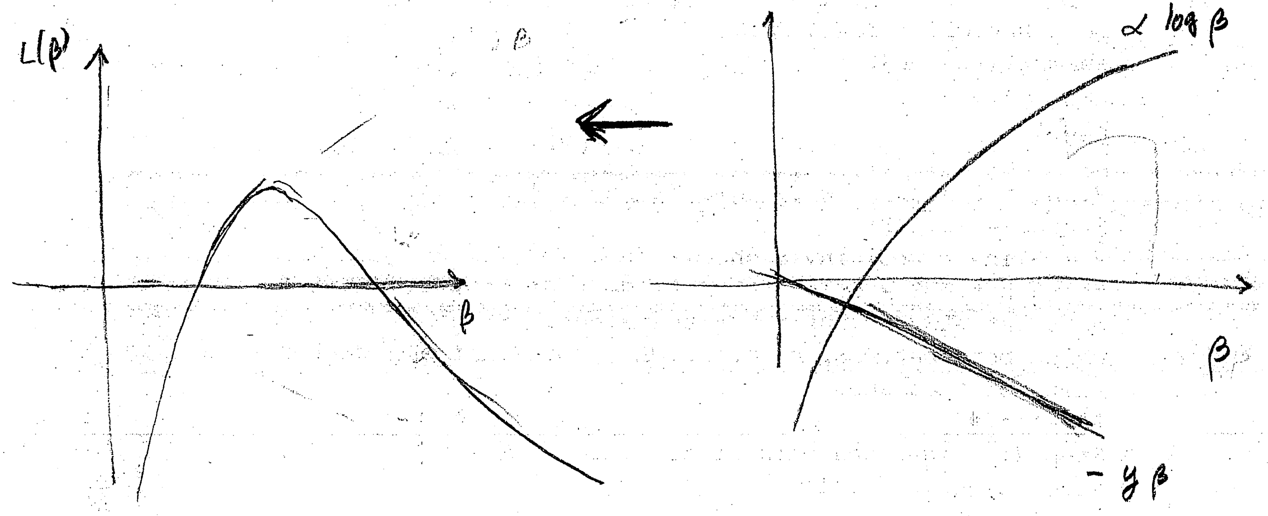
So need to estimate β using y.

☐ Convergence problems "step halvings"

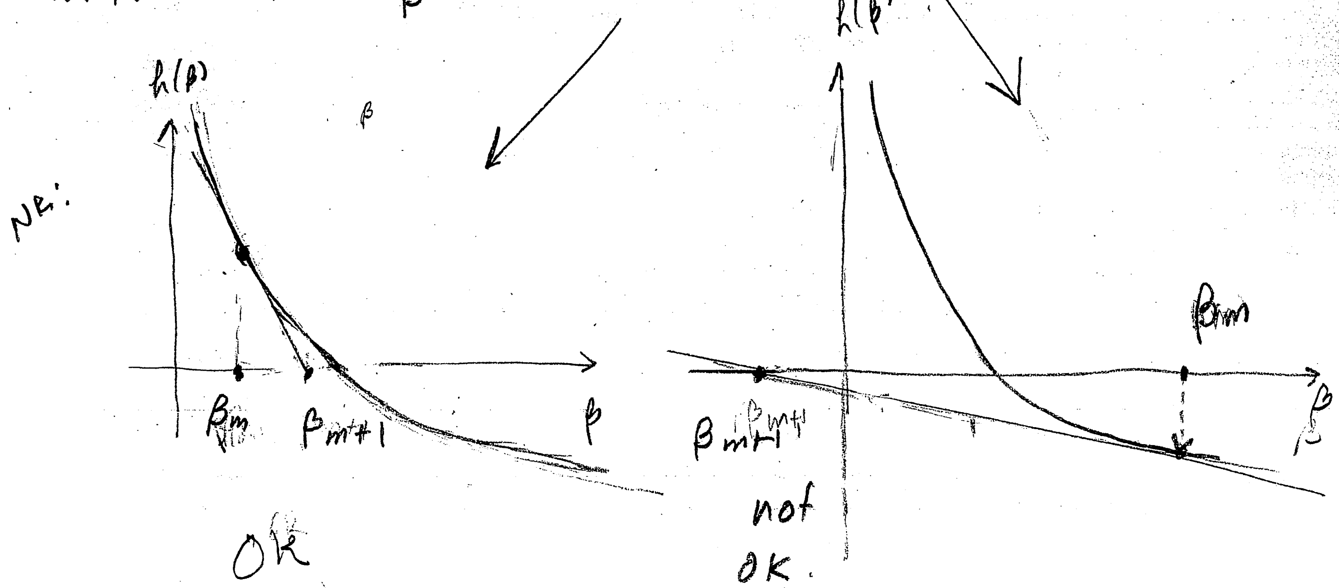
☐ For $Y \sim \Gamma(\alpha, \beta)$,

$$L = \text{log-lik} = \alpha \log \beta - \log \Gamma(\alpha) - \beta y + (\alpha - 1) \log y$$

$$L(\beta) = \alpha \log \beta - y\beta + \text{const.}$$



☐ $h(\beta) = L'(\beta) = \frac{\alpha}{\beta} - y$; to solve $h(\beta) = 0$



So halve step until get

$h(\beta_{m+1}) > 0$

☐ Estimation of ϕ (if at all): {formula page}

L = log like fc'n

$$= \sum_{i=1}^{n} \left\{ \frac{w_i}{\phi} [y_i \theta_i - b(\theta_i)] + c(y_i, \phi) \right\}$$

$$= L(\underset{\sim}{\beta}, \phi)$$

☐ MLE $(\hat{\underset{\sim}{\beta}}, \hat{\phi})$ satisfies:

(1) $\nabla_{\underset{\sim}{\beta}} L(\underset{\sim}{\beta}, \phi) = \underset{\sim}{0}$

(2) $\frac{\partial L}{\partial \phi}(\underset{\sim}{\beta}, \phi) = 0$.

Can see that for any ϕ, obtain same solution $\hat{\underset{\sim}{\beta}}$ for $\underset{\sim}{\beta}$ in $\nabla_{\underset{\sim}{\beta}} L(\underset{\sim}{\beta}, \phi) = \underset{\sim}{0}$, and can compute $\hat{\underset{\sim}{\beta}}$ (using IRWLS) without knowing ϕ. $\{ (A^{-1})^{-1}, A^{-1}$ where $var(y_i) = \phi v_i \}$

☐ Thus, to solve (1) and (2) in $(\underset{\sim}{\beta}, \phi)$:

(a) Compute $\hat{\underset{\sim}{\beta}}$ solving (1), using IRWLS NR

~~Estimate ϕ~~

(b) Solve $\frac{\partial}{\partial \phi} L(\hat{\underset{\sim}{\beta}}, \phi) = 0$ for ϕ.

□ To solve $\frac{\partial}{\partial \varphi} L(\hat{\underset{\sim}{\beta}}, \varphi) = 0$ where

$Y_i \sim \Gamma(\alpha, \beta_i) \Rightarrow$

$$f(y_i) = \exp\left\{ \frac{\left(-\frac{\beta_i}{\alpha}\right) y_i + \log\left(\frac{\beta_i}{\alpha}\right)}{1/\alpha} + \alpha \log \alpha - \log \Gamma(\alpha) + (\alpha - 1) \log y_i \right\}$$

and

$$\theta_i = -\frac{\beta}{\alpha}, \qquad \varphi = \frac{1}{\alpha}.$$

□ Thus $L = \sum_{i=1}^{n} \log f(y_i)$

$$= \frac{\sum_i [y_i \theta_i + \log(-\theta_i)]}{\varphi} + n\left[\frac{1}{\varphi}(-\log\varphi) - \log\Gamma\left(\frac{1}{\varphi}\right)\right] + \left(\frac{1}{\varphi} - 1\right) \sum_i \log y_i$$

$$= S \cdot \varphi^{-1} - n\left[\varphi^{-1} \log \varphi + \log \Gamma(\varphi^{-1})\right] + (\varphi^{-1} - 1) \cdot T$$

$\equiv g(\varphi)$, for $S = \sum_i [y_i \theta_i + \log(-\theta_i)]$ and $T = \sum_i \log y_i$

□ So need to solve

$$g'(\varphi) = 0$$

where θ_i replaced by $\hat{\theta}_i$ (for $\hat{\underset{\sim}{\beta}}$) in S and T.

50

□ $g(\varphi) = S \cdot \varphi^{-1} - n\left[\varphi^{-1}\log\varphi + \log\Gamma(\varphi^{-1})\right] + (\varphi^{-1}-1)\,T$

$\Rightarrow$

$g'(\varphi) = -S\varphi^{-2} - n\left\{\left[-\varphi^{-2}\log(\varphi) + \varphi^{-2}\right] + \left[\frac{d}{dx}\log\Gamma(x)\Big|_{x=\varphi^{-1}} \cdot (-\varphi^{-2})\right]\right\} - \varphi^{-2}\,T$

$= 0 \iff$

$-S + n\left[\log\varphi - 1 + \psi(\varphi^{-1})\right] - T = 0$,

where $\psi(x) = \frac{d}{dx}\log\Gamma(x)$ for $\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}\,du$ is the <u>digamma function</u>

$\iff \underbrace{\left[\log\varphi + \psi(\varphi^{-1})\right]}_{h(\varphi)} = \frac{1}{n}(S+T) + 1 \qquad (*)$

□ So $\hat{\varphi}$ is the solution to $(*)$,

with S and T (which depend on θ_i)

replaced by $\hat{S}$ and $\hat{T}$ (which use $\hat{\theta}_i$).

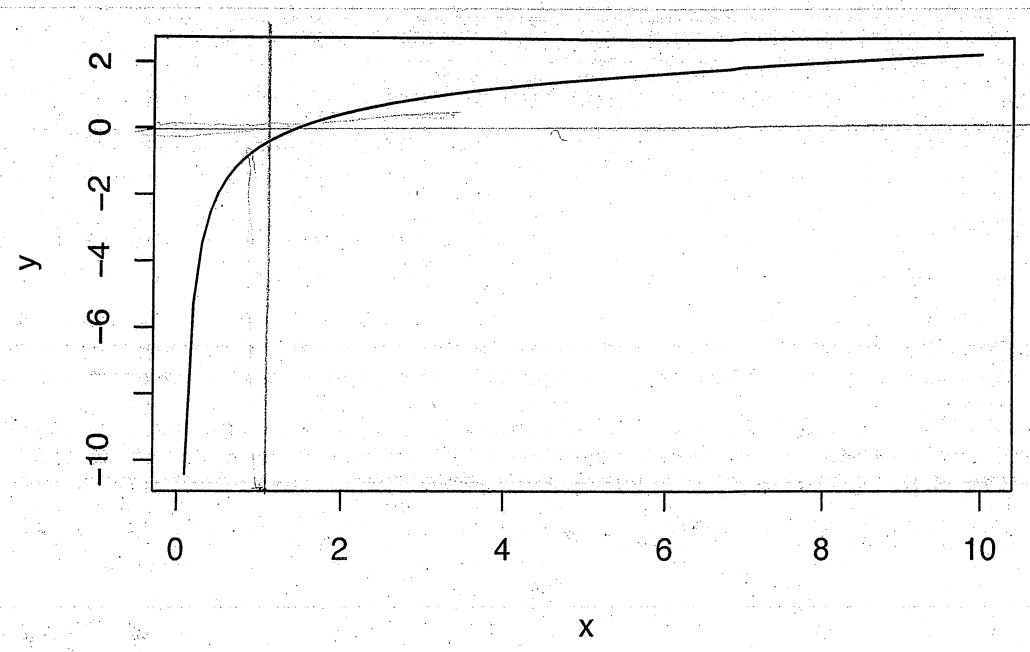
□ How does ψ and $h(\varphi)$ look?

next page

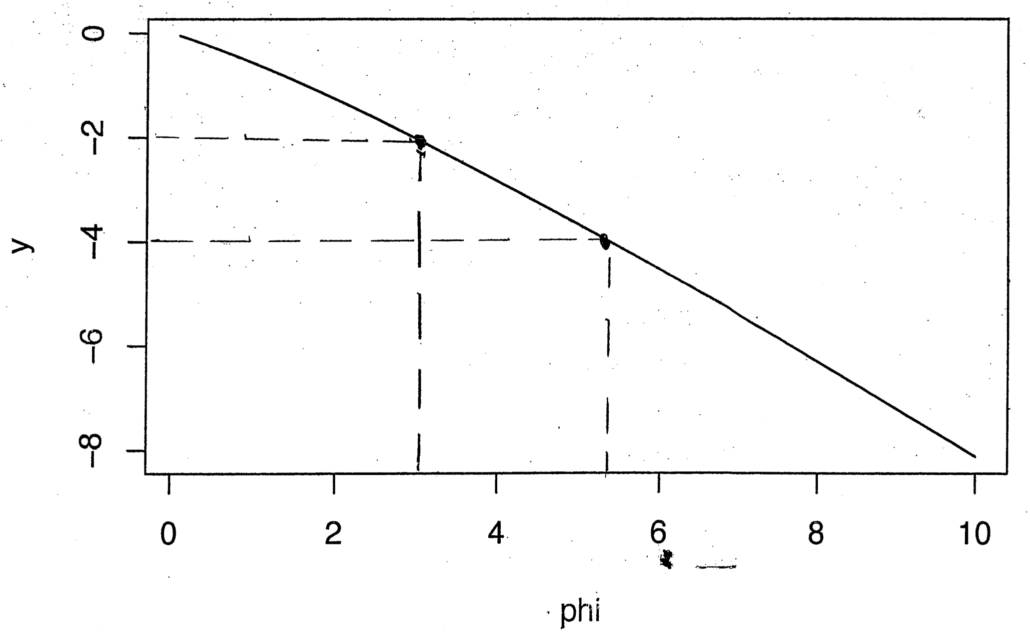
<u>NOTE</u>: $|h(\varphi_1) - h(\varphi_2)| = \Delta \Rightarrow |\varphi_1 - \varphi_2| > \Delta$

50.1

digamma function



$h(\phi) =$ **log(phi) + digamma(1/phi)**



□ Problem: If some $y_i \approx 0$, $\hat{\varphi}$ will be very sensitive to <u>rounding errors</u>:

(1) $\hat{\varphi}$ is solution to $h(\varphi) = \frac{1}{n}(\hat{S} + \hat{T}) + 1$ (*)

(2) $\hat{S} + \hat{T} = \sum_i [y_i \hat{\theta}_i + \log(-\hat{\theta}_i) + \log(y_i)]$, where $\hat{\theta}_i = \frac{-1}{\hat{\mu}_i}$

(3) Consider two sets of observations: I: $(y_1, y_2, \ldots, y_n)$

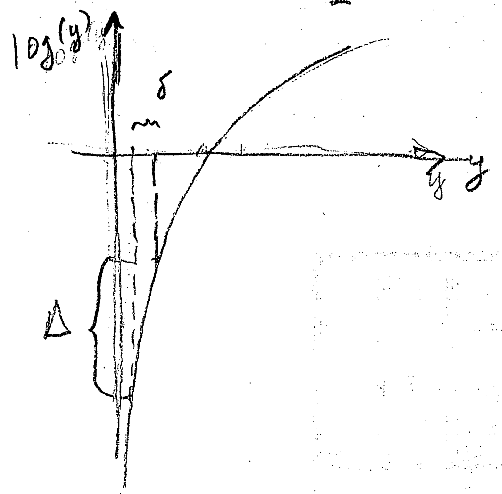
II: $(y_1 + \delta_1, y_2, \ldots, y_n)$

where $y_1 \approx 0$, $\delta_1 \approx 0$.

- will get $\hat{\beta}_I \approx \hat{\beta}_2$ and thus $\hat{\theta}_i^{(I)} \approx \hat{\theta}_i^{(II)}$

(first $y \approx 0$ in both cases)

- Thus $(\hat{S} + \hat{T})_{II} - (\hat{S} + \hat{T})_I \approx \log(y_1 + \delta) - \log(y_1) \equiv \Delta$,



Δ LARGE

(4) So $|h(\hat{\varphi}_I) - h(\hat{\varphi}_2)| \approx \Delta$ and thus

(see graph)

$$|\hat{\varphi}_I - \hat{\varphi}_{II}| > \Delta .$$

☐ Thus have <u>numerical instability</u>

when estimating φ for Γ using MLE, i.e.

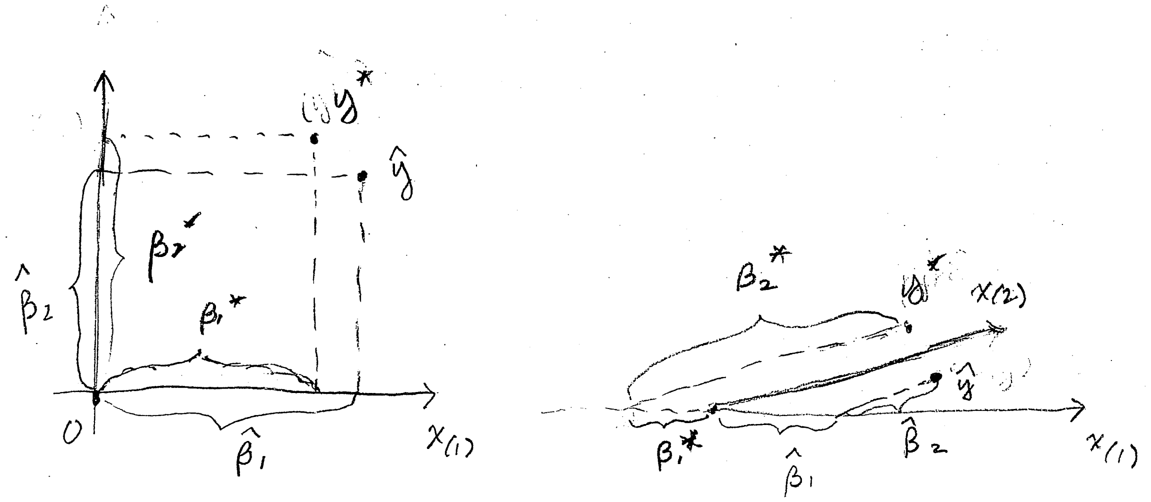
if one $y_i \approx 0$ then using $y_i + \delta_i$ (e.g. because of rounding error) gives large $|\hat{\varphi}_I - \hat{\varphi}_{II}|$

☐ Remark: Same problem occurs in multiple linear regression,

$$\hat{\beta} = (X^t X)^{-1} X^t y$$

when X is (approximately) multicollinear:

<u>EG</u>, let $\hat{y}$ = projection of $y = (y_1, \ldots, y_n)^t$ onto span$\{x_{(1)}, x_{(2)}\}$, y^* = projection of $(y + \delta)$



☐ So how estimate ϕ for Γ dist?

Hint: for linear regression,

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \begin{cases} \text{- MLE assuming } \varepsilon_i \sim N(0,\sigma^2) \\ \text{- Method of moments estimator assuming } E(\varepsilon_i)=0,\ var(\varepsilon_i)=\sigma^2 \end{cases}$$

☐ So use method of moments for GLIM