ESS 212

FALL 2024: *Homework # 4*

Problem 1 The data in the table is believed to obey a model of the form $y_i = ax_i^2 + bx_i + c + e_i$ with e_i assumed independent and identically distributed errors with $e_i \sim N(0, \sigma^2)$. Estimate a, b, and c and assign appropriate errorbars to your estimates. Also estimate the value of $y(t = 12)$ and assign an appropriate errorbar to the estimate.

t	y
1.0	-2.73
2.0	-2.71
3.0	-2.65
4.0	-0.87
5.0	-3.10
6.0	-1.03
7.0	0.63
8.0	1.46
9.0	5.90
10.0	8.38

My notes are messy I know so I also typed out the general procedure.

- Estimate a, b, ad c and assign appropriate error bars:

*Jeffrey's Prior assumed, likelihood fxn given in notes, can be expanded for each observation, but simplifies to (see notes). y = A*beta + e and we can now solve for e and beta (parameter vector [a b c]'). See nots for design matrix A (t_i are predictor variable values). e is the error vector, and e = y – A*beta, so we simplify likelihood function again (in notes, it's the one that starts with 1/ (2*pi)^5 * sigma^10.*

The posterior distribution is product of likelihood and prior normalized by Prob(data) ((evidence)). We simplify the expression by making posterior proportional to product of likelihood and prior but omitting evidence. This is allowed because the evidence (Prob(data)) doesn't depend on estimated parameters and can be treated like a constant when we are doing Bayesian proportionality.

The Jeffrey's Prior doesn't prefer any value of sigma over another value of sigma. Integrate out sigma to get marginal posterior dist. For beta, and simplify the integral. In simplified expression we introduce "neu" , denoted here as ***V : V****(beta) = (A*beta – y)'(A*beta – y) . k: # of parameters. Now we now the post. Dist. For beta is proportionate to 1 / ((e'e)^ (-3/2)). It follows a student t distribution which has heavier tails than normal distribution.*

Now revisit previous noes on posterior computed for this model. (we have to add a 1/ (sigma^(N+1) term to multiply the generalized form due to uncertainty in sigma.

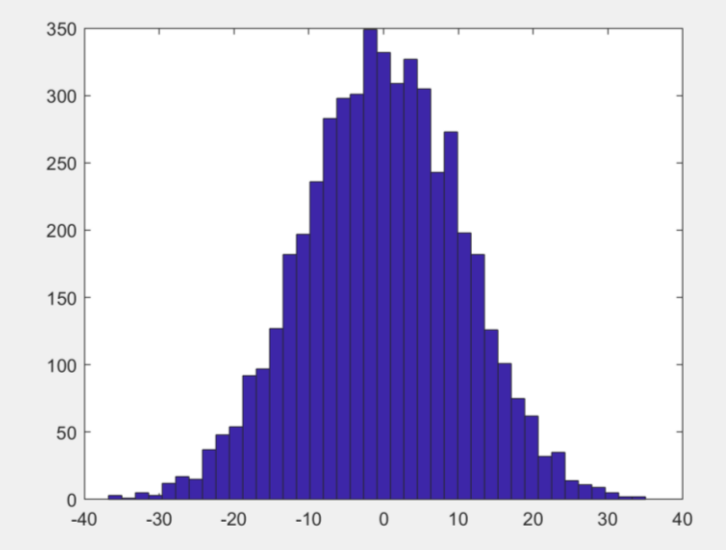
Beta_hat is the most probable values of beta, minimize the sum of squared residuals. To find most probable values for the parameters in beta, we take log of both sides of posterior distribution. Then take deriv. wrt/ beta and set it equal to 0 to find B_hat. Plug beta_hat back into model or evaluating model @ beta_hat gives us the most likely values of the parameters. This involves normal equations and shows that it is consistent with the solution from least squares/linear regression approach.

After we find beta_hat, we find sigma_hat by taking the log of the posterior distribution and differentiating wrt/ sigma, setting that equal to zero, and solving for sigma. See notes for final formula for sigma_hat ^ 2 (variance estimate). We can use the beta_hat that we found previously.

- Estimate y(t=12) and assign error bar:

To estimate y at t=12 (Bayesian inference), we derive the predictive dist. For y at a new point using the posterior dest. Of beta. See notes for formula for Prob(y12|beta) ..It includes a design matrix with a constant, linear and quadratic term for t. This is the probability dist of y12 given beta. But we don't actually know true beta.

So we must consider the posterior predictive dist. Of y at t = 12, incorporating the uncertainty in beta. (see notes). Then we integrate out sigma and get a distribution proportional to (see notes, it includes **H**., which = A'A / sigma^2. Probability of y12 given data is equal to a triple integral from –infinity to infinity (see notes), to address all uncertainty (in all parameters a b and c). A new symbol (del? Squiggle? Dirac delta?) . To solve the triple integral I use matlab to simulate samples from a t-distribution and take into account uncertainty in a b and c.

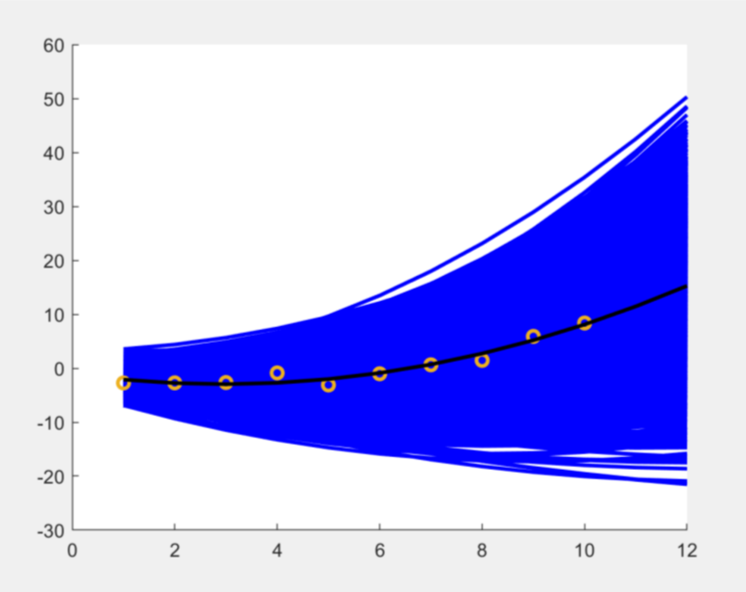


SOLUTION

Command Window

```
a = 2.223864e-01 +/- 5.231020e-02
b = -1.310614e+00 +/- 5.904336e-01
c = -1.025500e+00 +/- 1.413731e+00

error_12 =

   10.5416

>> d12

d12 =

   15.2708
```

- Estimates for parameters a, b, and c with errorbars
- d12(y at t = 12) with error_12 being the associated error



Notes from class below.

HW #4

ASSIGN PRIOR

(1) $\text{prob}(a, b, c, \log\sigma) \propto \text{const.}$

$\sigma) \propto 1/\sigma$

$\text{prob}(\log\sigma)\, d\log\sigma = \text{prob}(\sigma)\, d\sigma$

$\left| \frac{d\log\sigma}{d\sigma} \right| = \frac{1}{\sigma}$ (2) likelihood fxn

$\rightarrow \text{prob}(\text{data} \mid a, b, c, \sigma)$

how does data vary w/ Δ (the aspects of the model) (3) find posterior:

$\text{prob}(\underbrace{a, b, c, \sigma}_{\text{model}} \mid \text{data}) \propto$ conditioned on data

product rule

$\text{prob}(A|B)\,\text{prob}(B) = \text{prob}(\cancel{A|B}\ B|A)\,\text{prob}(A) = \text{prob}(A + B)$

$\rightarrow \propto \dfrac{\text{likelihood} \cdot \text{PRIOR}}{\underbrace{\text{prob}(\text{data})}_{\text{evidence}}} \propto \text{prob}(\text{data} \mid abc\sigma) \cdot \text{prob}(abc\sigma)$

$y_i = a t_i^2 + b t_i + c + e_i$

$e_i \sim N(\emptyset, \sigma^2)$ $\quad \underline{\beta} = \begin{bmatrix} c \\ b \\ a \end{bmatrix}$

rewrite $\underline{y} = \underline{\underline{A}}\, \underline{\beta} + \underline{e}$

$\underline{\underline{A}} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t & \vdots \\ 1 & t & \\ \vdots & \vdots & \\ 1 & t_{10} & t_{10}^2 \end{bmatrix} = [\underline{t}.\hat{\ }[0\ 1\ 2]]$ in MATLAB

(4) write prob distribution fxn:

likelihood

$\text{prob}(\text{data} \mid abc\sigma) \propto \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2} \frac{(y_i - (a t_i^2 + b t_i + c))^2}{\sigma^2}\right.$

$\cdot \frac{1}{\sqrt{2\pi}\,\sigma} \exp\Big(\frac{1}{}$ same as prev but 2, 3, ..., 10

Simplify!

sum of squares of difference between

$= \frac{1}{\sqrt{2\pi}(2\pi)^5 \sigma^{10}} \cdot \exp\Big(-\frac{1}{2}\big(\underline{e}\big)^T \cdot \big(\underline{e}\big)\Big)$

dot product here gives sum of squares

data & model

solve for e

$\underline{e} = (\underline{y} - \underline{A}\,\underline{\beta})$

(data − model)

$\Rightarrow \frac{1}{(2\pi)^5 \sigma^{10}} \exp\Big(-\frac{1}{2\sigma^2}(y - A\beta)^T(y - A\beta)\Big)$

likelihood

posterior

$\text{prob}(\underline{\beta}, \sigma \mid \text{data}) \propto \frac{1}{\sigma^{11}} \Big(\text{but } \hat{\sigma}^2 \Big)$

" bc Jeffrey's prior?

proportionality lets you drop the 2π term

MATLAB

design mat w/ col of 1s

p = [t.^1, t.^2]

lm = fitlm(p, y);   LINEAR MODEL

(col ones implicit)

$\hat{a}$ = lm.coefficients.Estimate (3)   backwards?

b (2)   c (1)

$\hat{\beta} = \underline{\underline{A}} \backslash \underline{y}$

$= (\underline{\underline{A}}'\underline{\underline{A}}) \backslash \underline{\underline{A}}'\underline{y}$

we want to know error bar for $\hat{a}$

$a = \hat{a} \pm \sigma_a$

$b = \hat{b} \pm \sigma_b$ $c = \hat{c} \pm \sigma_c$

assign std dev to a, b, + c

$\sigma_a =$

lin. coefficients · SE(3) $\sqrt{\nu/\nu-2}$

"nu" $\nu = N - K$ ← # params (a, b, c)

$k=3$

degrees of freedom   what N again? 10?

we want to integrate σ explicitly to get new PDF?

$$\text{prob}(\underline{\beta} \mid \text{data}) = \int_{-\infty}^{\infty} d\log\sigma \{\text{prob}(\underline{\beta}, \log\sigma \mid \text{data})\}$$

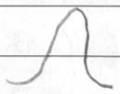
sum rule

weighted avg taking into account we don't know σ

$$\propto V(\beta)^{-k/2} \quad \text{where } V(\beta) = (\underline{A}\underline{\beta} - \underline{y})'(\underline{A}\beta - \underline{y})$$

$k = \text{length}(\underline{\beta})$

$$\frac{1}{(\underline{e}^T\underline{e})^{-3/2}}$$



$\hat{\sigma}$ find where deriv a, b, c $= 0$ TO FIND $\hat{\sigma}$:

deriv of posterior

for finding prob($\underline{\beta}$, σ | data)

$\mathcal{L} = \log(\text{prob}(\beta, \sigma \mid \text{data}))$

$$\frac{\partial \mathcal{L}}{\partial \underline{\beta}} = -\frac{1}{\cancel{2}\sigma^2}(\underline{A}\underline{\beta} - \underline{y})^T \underline{A}$$

gets rid of factor of 2

$$\mathcal{L} = -\frac{1}{2\sigma^2}(\underline{\underline{A}}\underline{\beta} - \underline{y})^T(\underline{\underline{A}}\underline{\beta} - \underline{y}) - (N+1)\log\sigma$$

$$\frac{\partial \mathcal{L}}{\partial \sigma} = +\frac{1}{4\sigma^3}(\underline{\underline{A}}\underline{\beta} - \underline{y})^T(\underline{\underline{A}}\,\underline{\beta} - \underline{y}) - \frac{(N+1)}{\sigma}$$

Now 4 eqns, 4 unknowns

Set $\left.\frac{\partial \mathcal{L}}{\partial \sigma}\right|_{\sigma=\hat{\sigma}} = 0$ + $\left.\frac{\partial \mathcal{L}}{\partial \beta}\right|_{\beta=\hat{\beta}} = 0 \Rightarrow \hat{\beta}$

$\rightarrow = (A^T A)^{-1} \underline{\underline{A}}^T \underline{y}$ $\hat{\sigma} = \frac{1}{4}(A\hat{\beta} - y)^T(A\hat{\beta} - y) -$

ctnd from Monday week 1∅

$B = \underline{\underline{A}} \backslash g = (A'A) \backslash (A'g)$

$a = \hat{a} \pm \sigma_a$ ... b + c too

find in matlab

$$\text{prob}(B \mid \text{data}) = \int_{-\infty}^{\infty} d\log\sigma \, \{ p(\beta \log\sigma \mid \text{data}) \}$$

$$= \int_{-\infty}^{\infty} d\sigma \, \{ p(B\sigma \mid \text{data}) \}$$

$$\propto V(B)^{-K/2} \quad \text{where}$$

$$V(B) = (AB - g)'(AB - g)$$

$K = \text{length}(B)$

...

$$\left. \frac{\partial \mathcal{L}}{\partial B} \right| = \frac{1}{\sigma^3}(AB - g)^T(AB - g) - \frac{N+1}{\sigma} = \varnothing$$

$$\left. \frac{\partial \mathcal{L}}{\partial B} \right|_{\substack{B = \hat{B} \\ \sigma = \hat{\sigma}}} = \frac{1}{\sigma^3}(AB - g)^T(AB - g) - \frac{N+1}{\sigma} = \varnothing$$

$$\Longrightarrow \frac{1}{4}(A\hat{B} - g)^T(A\hat{B} - g) - \hat{\sigma}^2(N+1) = \varnothing$$

hat?

$$\hat{\sigma}^2 = \frac{(AB - g)^T(A\hat{B} - g)}{(N+1)}$$

3/14/24

1) $y_i = a x_i^2 + b x_i + c + e_i$ ← MODEL

e_i assumed independent

$e_i \sim N(\emptyset, \sigma^2)$

- estimate a b c + error bars
- estimate value $y(t=12)$ + error bar

First, ASSIGN PRIOR:

$\text{prob}(a, b, c, \log\sigma) \propto \text{const}$

$\text{prob}(a, b, c, \sigma) \propto 1/\sigma$

$\Rightarrow \text{prob}(\log\sigma)\, d\log\sigma = \text{prob}(\sigma)\, d\sigma$

Second LIKELIHOOD FXN:

$\text{prob}(\text{data} \mid a, b, c, \sigma)$

third find POSTERIOR:

$\text{prob}(\underbrace{a, b, c}_{\text{model params}}, \sigma \mid \text{data}) \propto \text{prob}(A|B)\,\text{prob}(B)$ (apply product rule)

$\rightarrow = \text{prob}(B|A)\,\text{prob}(A) = \text{prob}(A + B)$

$\rightarrow \propto \dfrac{\text{likelihood} \cdot \text{prior}}{\text{prob}(\underbrace{\text{data}}_{\text{evidence}})} \propto \text{prob}(\text{data} \mid a, b, c, \sigma) \cdot \text{prob}(a, b, c, \sigma)$

Switch notation of model xs to ts

Rewrite model:

$\underline{y} = \underline{\underline{A}}\,\underline{\beta} + \underline{e}$

$\underline{\underline{A}} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \end{bmatrix}$ $\quad \underline{\beta} = \begin{bmatrix} c \\ b \\ a \end{bmatrix}$

Rewrite model:

Next, write PROB DISTRIBUTION FXN:

likelihood

$$\text{prob}(\text{data} \mid a\, b\, c\, \sigma) \propto \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(y_i - (at_i^2 + bt_i + c))^2}{\sigma^2}\right)$$

$$\cdot \frac{1}{\sqrt{2\pi}\sigma} \exp(\quad \leftarrow 2, 3, 4 \cdots 10)$$

SIMPLIFY!

$$= \frac{1}{(2\pi)^5 \sigma^{10}} \exp\left(-\frac{1}{2} (\underline{e})^T \cdot (\underline{e})\right)$$

SOLVE FOR e

$\underline{e} = (\underline{y} - \underline{\underline{A}}\, \underline{\beta}) \leftarrow$ data - model

LIKELIHOOD

$$\frac{1}{(2\pi)^5 \sigma^{10}} \exp\left(-\frac{1}{2\sigma^2} (y - A\beta)^T (y - A\beta)\right)$$

POSTERIOR

$$\text{prob}(\underline{\beta}, \sigma \mid \text{data}) \propto \frac{1}{\sigma^{11}} \exp\left(-\frac{1}{2\hat{\sigma}^2} \quad \right)$$

can drop the 2π — Jeffreys prior

FIT MODEL IN MATLAB + find $\hat{a}\ \hat{b}\ \hat{c}$ + their errorbars

$\hat{\beta} = \underline{\underline{A}} \backslash \underline{y} \quad (\underline{\underline{A}}'\underline{\underline{A}}) \backslash \underline{\underline{A}}'\underline{y}$ — ex. $a = \hat{a} \pm \sigma_a$

in matlab, same as

gaussian 3/13/24

$\text{prob}(\beta, \sigma | d) \propto \frac{1}{\sigma^{N+1}} \exp\left(-\frac{1}{2\sigma^2}(A\beta - d)'(A\hat{\beta} - d)\right)$

posterior

comes from prior

$\text{prob}(g(12) | \beta) = $

errors are assume random noise + we know true value of β

or

$\delta([1 \ \ 12 \ \ 12^2]\underline{\beta} - g(12))$

prob. dist. of Ø width

would be no uncertainty

but we don't know β

$\text{prob}(g(12), \beta | \underline{d}) = \text{prob}(g(12) | \beta)\,\text{prob}(\beta | d)$

product rule

integral

$\int_0^\infty d\sigma \ \text{prob}(\underline{\beta}, \sigma | d)$

$\propto [\nu + (\underline{\beta} - \hat{\beta})^T \underline{\underline{H}} (\underline{\beta} - \hat{\underline{\beta}})]^{-N/2}$

K: # params a b c $\Rightarrow 3$

N: # data $\Rightarrow 10$

$\nu = N - K$

$\underline{\underline{H}}$: precision of multivar. t dist

i.e. inv. cov. mat.

$= \frac{\underline{\underline{A}}'\underline{\underline{A}}}{s^2}$

$\underline{\underline{A}} = [\underline{t}^{\wedge}[0,1,2]]$

$s^2 = (\underline{d} - \underline{\underline{A}}\hat{\underline{\beta}})'(\underline{d} - \underline{\underline{A}}\hat{\underline{\beta}})$

makes error vect smallest ν (all over ν)

$\hat{\beta} = (\underline{\underline{A}}'\underline{\underline{A}})^{-1}\underline{\underline{A}}'\underline{d}$ ← least sq est

location of max of posterior

$s^2 = \frac{e'e}{\nu}$

Black lines have integrated out σ

C = off diagonals are covariance

3 x 3 (bc a, b, c)

CM = correlation matrix

1. also 3 x 3

$$\text{prob}(g(12) \mid \underline{\beta}\, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(g(12) - \underline{\underline{A}}\,\underline{\beta})'(g(12) - \underline{\underline{A}}\,\underline{\beta})\right)$$

$$2.\ \text{prob}(\underline{\beta}, \sigma, \underline{d}) \propto \frac{1}{\sigma^{N+1}} \exp\left(-\frac{1}{2\sigma^2}(\underline{\underline{A}}\,\underline{\beta} - \underline{d})'(\underline{\underline{A}}\,\underline{\beta} - \underline{d})\right)$$

combine 1 + 2 to get: / prob of projection

3. prob(g(12), $\underline{\beta}$, σ | $\underline{d}$)

S: sqrt diag elements of matrix

in code

$$S = \begin{bmatrix} \sigma_a & & \emptyset \\ & \sigma_b & \\ \emptyset & & \sigma_c \end{bmatrix} \quad S \backslash C / S \Rightarrow S^{-1} \underline{\underline{C}}\, S^{-1}$$

$\Rightarrow \underline{\underline{CM}} \rightsquigarrow$ correlation matrix (dimensionless) unlike $\underline{\underline{C}}$

(All prob #1, Now prob #2)

ignore last 3 points, ...

use that model as a prior

(on prev page)

$\text{prob}(\varsigma(12), \beta | d) =$

$\text{prob}(\varsigma(12) | \underline{d}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} da\,db\,dc\,(\delta\,\varsigma(12) - \underline{\underline{A}}\,\underline{\beta}))$

$\rightarrow [D + [\beta - \hat{\beta}]\,\underset{=}{H}\,(\beta - \hat{\beta})]^{-N/2}$ all params

then can plot
histogram / probability solve in MATLAB
by drawing random samples

assume prior
 └ a b c all equally uncertain
└ errors in model random + indep

Now assume there's an actual rain
process happening (not random)

need both

$\text{prob}(\varsigma(12), \underline{\beta}, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma^2} \exp\left(-\frac{1}{2}(\varsigma(12) - \underline{\underline{A}}\,\beta)'\right.$
$\left.(\varsigma(12) - \underline{\underline{A}}\,\beta)\right)$

+

$\text{prob}(\varsigma(12), \beta, \sigma | d) =$
$\text{prob}(\varsigma(12) | \beta, \sigma)\,\text{prob}(\beta, \sigma | \underline{d})$

$\rightarrow \propto \frac{1}{\sigma^{N+2}} \exp\left(-\frac{1}{2\sigma^2}(\varsigma(12) - \underline{\underline{A}}\,\beta)'(\varsigma(12) - \underline{\underline{A}}\,\beta)\right.$
$\left. - \frac{1}{2\sigma^2}(\underline{\underline{A}}\,\underline{\beta} - \underline{d})'(\underline{\underline{A}}\,\underline{\beta} - d)\right)$

Now can integrate out
β + σ + get prob $(\varsigma(12))$

Problem 2 The data in the table is believed to obey a model of the form $y_i = mx_i + b + e_i$ where e_i is a gaussian error term. For the first two measurements the error is known and indicated in the table. For the last three measurements the error is not known, but is assumed to have a constant variance.

x	y
0	0.0434 ± 0.1
1	1.0343 ± 0.1
2	$-0.2588 \pm \sigma$
3	$3.68622 \pm \sigma$
4	$4.3188 \pm \sigma$

Estimate m and b and assign appropriate errorbars to your estimates.

m = 1.12027

b = -0.47576

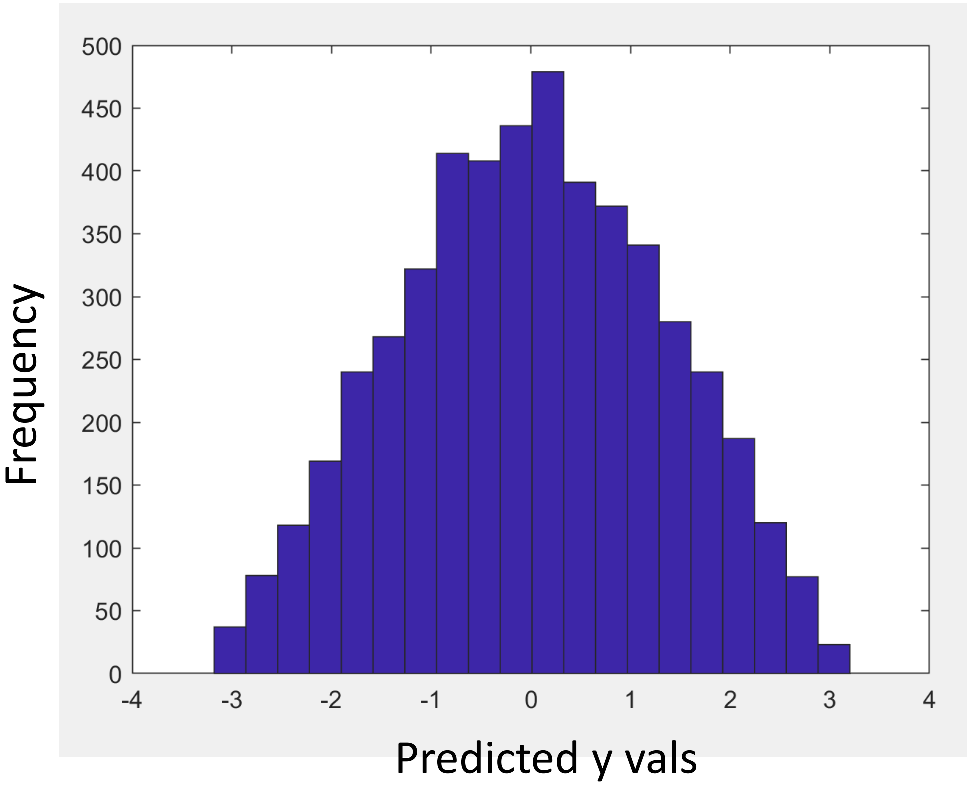
sigma_m = 0.7250

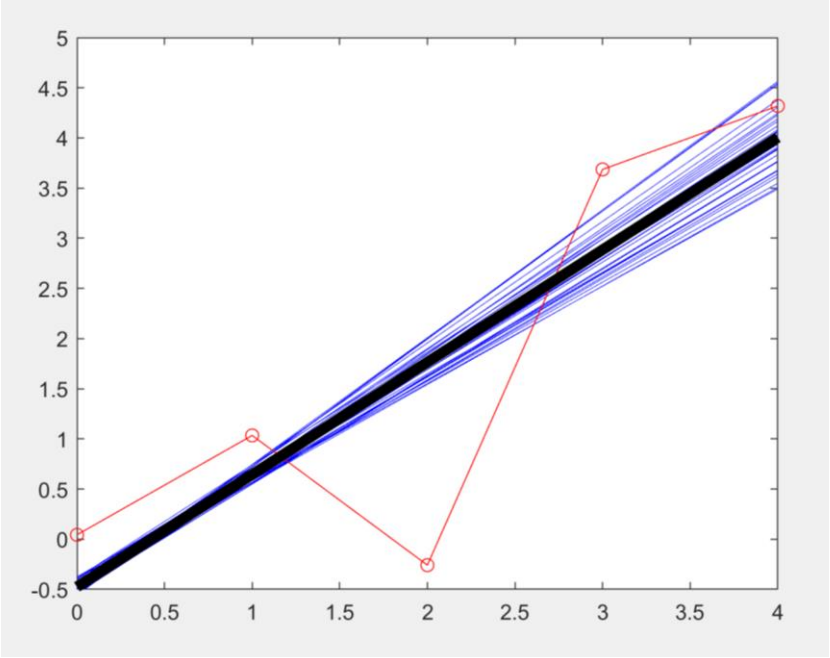
sigma_b = 1.7759

errorbar_x2 = 1.1125

errorbar_x3 = 1.3321

errorbar_x4 = 1.6306





Blue lines represent model error at each x value, red points are the data, and black line is my model.

Problem 3 The data in the table below are assumed to obey a model of the form $y_{i,j} = e_{i,j}\exp(-bt_j)$ with $e_{i,j}$ assumed independent and identically distributed errors with $\log(e_{i,j}) \sim N(\mu, \sigma^2)$. Estimate b, μ, and σ and assign appropriate errorbars to your estimates.

$t_:$	$y_{1,:}$	$y_{2,:}$	$y_{3,:}$	$y_{4,:}$
1.0	3.75	0.36	0.58	2.06
2.0	0.93	0.32	0.67	1.01
3.0	0.38	0.11	0.12	0.60
4.0	0.05	0.15	0.05	0.11
5.0	0.04	0.03	0.08	0.06

3) $y_{ij} = e_{ij}\exp(-bt_j)$

$i = 1, 2, 3, 4, 5$

• est. b, μ, σ & errorbars

$\log(e_{ij}) \sim N(\mu, \sigma^2)$

$\underbrace{\phantom{\log(e_{ij})}} = \mu + \varepsilon_{ij}$ where $\varepsilon_{ij} \sim N(\emptyset, \sigma^2)$

$\underbrace{\log(y_{ij})} = \log(e_{ij}) - bt_j$

$x_{ij} = \mu - bt_j + \varepsilon_{ij}$

$\boxed{\langle \log e_{ij} \rangle = \mu \rightarrow \langle \log e_{ij} - \mu \rangle = \emptyset}$

After taking log, we see is linear in terms of parameter b. We can start building likelihood function where: Beta = [mu; b] prior : $\text{prob}(\mu, b, \sigma) \propto \frac{1}{\sigma}$

likelihood : $\text{prob}(\text{data} \mid \text{model})$

$= \text{prob}(y \mid \mu, b, \sigma) \propto$

$$\frac{1}{\sigma^{10}}\left(\exp\left(-\frac{1}{2\sigma^2}(\varepsilon^T \varepsilon)\right)\right)$$

$\varepsilon = y - A\beta \qquad \beta = \begin{bmatrix} \mu \\ b \end{bmatrix}$

$$\text{prob}(y \mid \mu, b, \sigma) \propto \frac{1}{\sigma^{10}} \exp\left(-\frac{1}{2}(A\beta - y)^T(A\beta - y)\right)$$

$$\text{prob}(\sigma, \mu, b \mid y) \propto \frac{1}{\sigma^{11}} \exp\left(-\frac{1}{2} \quad \right)$$

here

also here

$$\log(\text{posterior}) = \text{const} - 11 \log \sigma - \frac{1}{2\sigma^2}$$

The likelihood function should be represented as Prob (z_i,j (i.e. the data, I just call it y) given both parameters of beta (mu and b) and the standard deviation of the errors, sigma. A is the design matrix incorporating time points t_j and appropriately aligns with the structure of beta. This is assuming the prior (highlighted in blue). Each time point is repeated four times to account for the four different y vals.

$$\text{prob}(y \mid m, b, \sigma) \propto \frac{1}{\sigma^{10}} \exp\left(-\frac{1}{2}(A\beta - y)^T(A\beta - y)\right)$$

here

$$\text{prob}(\sigma, m, b \mid y) \propto \frac{1}{\sigma^{11}} \exp\left(-\frac{1}{2} \quad \right)$$

also here

$$\log(\text{posterior}) = \text{const} - 11 \log \sigma - \frac{1}{2\sigma^2}$$

$$\hat{\beta} = \qquad \log(\text{posterior}) = \mathcal{L}$$

$\wedge$?

$$\left.\frac{\partial \mathcal{L}}{\partial \beta}\right|_{\substack{\beta = \hat{\beta} \\ \sigma = \hat{\sigma}}} = 0 \qquad \left.\frac{\partial \mathcal{L}}{\partial \beta}\right|_{\substack{\beta = \hat{\beta} \\ \sigma = \hat{\sigma}}} = 0 \qquad A^T A \hat{\beta} = A^T y$$

$$\frac{d\mathcal{L}}{d\sigma} = -\frac{11}{\sigma} - \sigma^{-3}\left[(\beta - \hat{\beta})A^T A(\beta - \hat{\beta}) + (y - A\hat{\beta})^T (y - A\hat{\beta})\right.$$

$$0 = \cdots \quad 11\,\hat{\sigma}^2 = -\left[(y - A\hat{\beta})^T(y - A\hat{\beta})\right]$$

The posterior becomes (highlighted in blue).
We then define the log posterior denoted as squiggly L (highlighted in green). We want to find beta_hat, the maximum likelihood estimate for beta. We take derivative of squiggle_L wrt beta and set it equal to 0, then evaluate at beta_hat. We expand this and solve through differentiation to get beta_hat.

I didn't keep up in the notes but derivative wrt beta evaluated at beta = beta_hat is:

d/d_beta [-1/2*sigma^2 (A*beta – y)' (A*beta – y)] = 0

Solve to get:

beta_hat = (A'A)^-1 * A'y

Note: ' implies transpose

Now we want to find the maximum likelihood estimate for sigma. We differentiate squiggle_L wrt/ sigma and set it equal to 0. Set derivative equal to 0 and solve for sigma to get sigma_hat. Multiply through by sigma_hat^3.

$$\| \hat{\sigma}^2 = -[(y - A\hat{\beta})^T(y - A\hat{\beta})]$$

Then divide by residual sum of squares to isolate sigma_hat^2, the maximum likelihood for variance parameter assuming the beta_hat we found.

$$\text{prob}(\beta \mid y) = \int_0^\infty d\sigma \, \text{prob}(\beta, \sigma \mid y)$$

$$\text{prob}(\mu \mid y) = \int_{-\infty}^{\infty} db \int_0^\infty d\sigma \, \text{prob}(\beta, \sigma \mid y)$$

$$\Rightarrow \log(\text{prob}(\mu \mid y)) = \frac{-1}{2(N-K+1)} \log(\nu + $$

$$\rightarrow (\mu - \hat{\mu})^2 / ? \;)$$

$$b \mid y \rightarrow (b - \hat{b})^2 / c(2,2)$$

$$\nu \doteq N - K \qquad \underline{\underline{H}} = s^2 \underline{\underline{A}}^T \underline{\underline{A}} \qquad s^2 = \frac{\underline{e}^T \underline{e}}{\nu}$$

$$\underline{\underline{C}} = \underline{\underline{H}}^{-1} \qquad c(1,1)$$

Given posterior distribution (Beta, sigma | y), we find the distribution of beta independent of sigma. We integrate sigma from the joint posterior. The integration results in a distribution like the student’s t-distribution, which is a better representation of beta uncertainty when sigma isn’t exactly know.. Integrate over sigma and other components of beta that are not mu to find Prob(mu|y) (highlighted in green). Log of marginal posterior of mu (highlighted in orange). Now we can see the log (Prob(b|y)) = -1/2(N – k + 1) * log(***v*** + (the box highlighted in blue)).

$$\text{prob}(\beta \mid y) = \int_0^\infty d\sigma \, \text{prob}(\beta, \sigma \mid y)$$

$$\text{prob}(\mu \mid y) = \int_{-\infty}^{\infty} db \int_0^\infty d\sigma \, \text{prob}(\beta, \sigma \mid y)$$

$$\Rightarrow \log(\text{prob}(\mu \mid y)) = \frac{-1}{2(N-K+1)} \log\left(\nu + (\mu - \hat{\mu})^2 / ? \right)$$

$$b \mid y \rightarrow (b - \hat{b})^2 / c(2,2)$$

$$\nu = N - K \qquad \underline{\underline{H}} = s^2 \underline{\underline{A}}^T A \qquad s^2 = \frac{\underline{e}^T \underline{e}}{\nu}$$

$$\underline{\underline{C}} = \underline{\underline{H}}^{-1} \qquad c(1,1)$$

Given posterior distribution (Beta, sigma | y), we find the distribution of beta independent of sigma. We integrate sigma from the joint posterior. The integration results in a distribution like the student's t-distribution, which is a better representation of beta uncertainty when sigma isn't exactly known.. Integrate over sigma and other components of beta that are not mu to find Prob(mu|y).

Solve in Matlab

ANSWERS

b =

-0.8395

mu =

2.8975

sigma_b =

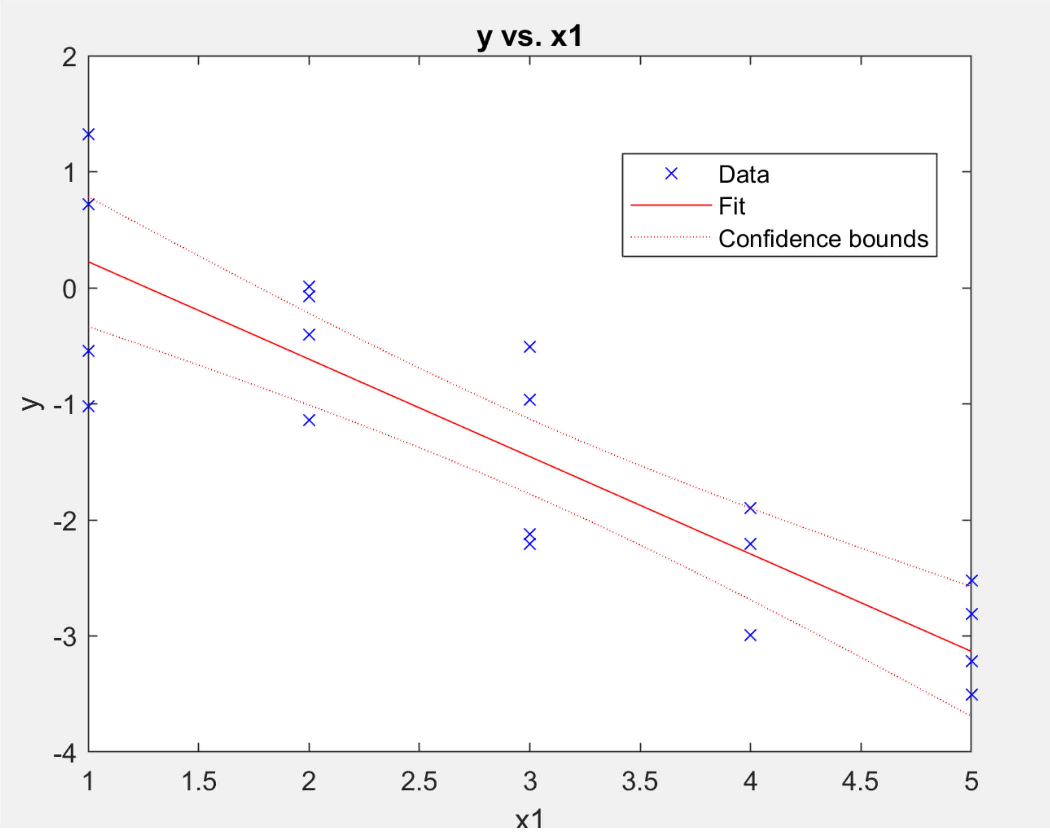
0.1149

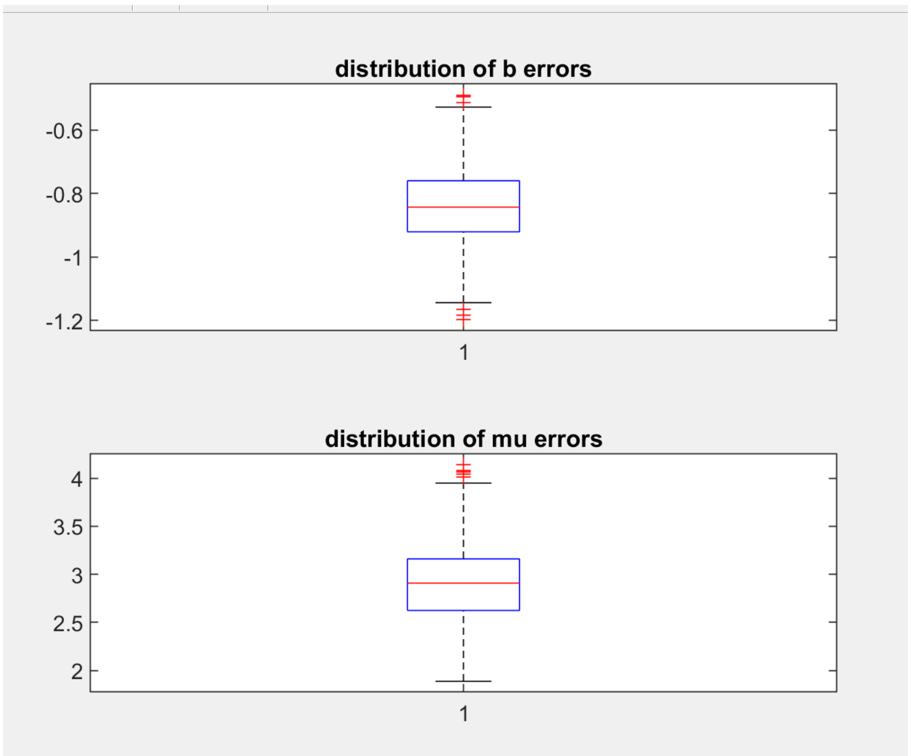
sigma_mu =

0.3812

error =

sigma_error = 0.1300





Boxplots of the individual b and mu errors (sigma) for all x_i : x_N