problem set 3 #1

① extra notes 15 Feb (part 1)

$$\left.\begin{array}{l} \lambda \sim \Gamma^{-1}(\alpha, \beta) \\ (Y_i \mid \lambda) \overset{iid}{\sim} \mathcal{E}(\lambda) \end{array}\right\} \quad (\lambda \mid y) \sim \Gamma^{-1}(\alpha + n, \beta + s)$$

$$s = \sum_{i=1}^{n} y_i$$

$$(Y_i \mid \lambda) \overset{iid}{\sim} \mathcal{E}(\lambda) \leftrightarrow p(y_i \mid \lambda) = \begin{cases} \frac{1}{\lambda} e^{-\frac{y_i}{\lambda}} & y > 0 \\ 0 & \text{else} \end{cases}$$

$$\lambda \sim \Gamma^{-1}(\alpha, \beta) \to p(\lambda) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-(\alpha+1)} e^{-\frac{\beta}{\lambda}} & \lambda > 0 \\ 0 & \text{else} \end{cases}$$

$$\eta = g(\lambda) = \frac{1}{\lambda} = \lambda^{-1}$$

MCMC solution

dead easy but not in closed form

$$\begin{array}{c|c} \lambda & \eta = \frac{1}{\lambda} \\ \hline \vdots & \vdots \end{array}$$

Q: if I know dist.n of λ, $p(\lambda)$, what is density of $g(\lambda) = \eta = \frac{1}{\lambda}$? prior (dens.)

A: work out CDF & diff. to get

density: $F_\eta(t) = P(\eta \le t) = P(\frac{1}{\lambda} \le t)$ ②

$\stackrel{(\lambda>0)}{=} P(\lambda \ge \frac{1}{t}) = 1 - P(\lambda < \frac{1}{t}) \stackrel{\lambda \text{ cont}}{=}$

$= 1 - P(\lambda \le \frac{1}{t}) = 1 - F_\lambda(\frac{1}{t})$, so

$p_\eta(t) = \frac{d}{dt}\left[1 - F_\lambda(\frac{1}{t})\right] = t^{-2} \boxed{f_\lambda(\frac{1}{t})}$

$p_\eta(t) = t^{-2} \frac{\beta^\alpha}{\Gamma(\alpha)} (\frac{1}{t})^{-(\alpha+1)} e^{-\beta/(1/t)}$

$= \frac{\beta^\alpha}{\Gamma(\alpha)} t^{(\alpha-1)} e^{-\beta t}$

i.e. $= \Gamma(\alpha, \beta)$

if $\lambda \sim \Gamma^{-1}(\alpha, \beta)$ then $\frac{1}{\lambda} \sim \boxed{\Gamma(\alpha, \beta)}$

---

② 1:k $\ell(\lambda | y) = c \lambda^{-n} e^{-s/\lambda}$ so

$\ell(\eta | y) = c\, \eta^n e^{-s\eta}$ (direct subst. $\eta \leftrightarrow \frac{1}{\lambda}$)

$= \Gamma(n+1, s)$

③ post $(\eta \mid y) \sim \Gamma(\alpha + n, \beta + s)$ ✓ ③

PS 3 #3 $(D_i \mid \sigma^2) \stackrel{iid}{\sim} N(0, \sigma^2)$

$D = (D_1, \ldots, D_n)$ $d_n = (d_1, \ldots, d_n)$

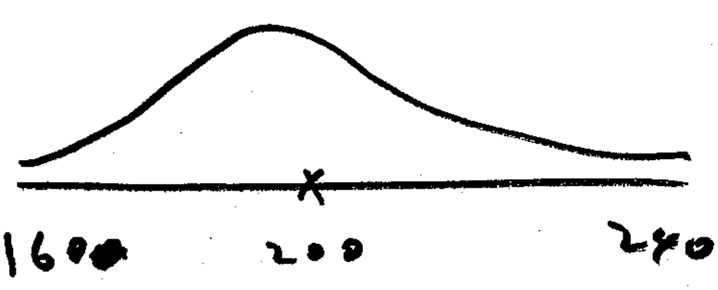
$$\ell(\sigma^2 \mid \cancel{y}\, d) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(d_i - 0)^2}$$

$$= c\,\underbrace{(\sigma^2)}_{\theta}^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\underbrace{\sum_{i=1}^{n} d_i^2}_{s}}$$

(s: suff. stat.)

$$= \cancel{\chi}^{-2}(\underbrace{n-2}_{\text{"sample size"}}, \hat{\sigma}^2_{MLE})$$

$$\ell\ell(\sigma^2 \mid d) = c - \frac{n}{2}\ln\sigma^2 - \frac{s}{2\sigma^2}$$



$$\hat{\sigma}^2_{MLE} = \frac{s}{n} = \frac{1}{n}\sum_{i=1}^{n} d_i^2 = 191.8$$

~~noninformative~~ $\leftrightarrow$ diffuse, flat ④

$\nu_0 \to 0 \qquad p(\sigma^2) = c \, \frac{1}{\sigma^2}$

this is an improper prior (inf. to $+\infty$)

$\boxed{154 \text{ #2}}$

model $\quad \lambda \sim$ diffuse

① $(y_i \mid \lambda) \overset{iid}{\sim}$ Exponential(λ)

$\quad (\alpha, \beta) \sim$ diffuse

② $(y_i \mid \alpha, \beta) \overset{iid}{\sim}$ Gamma(α, β)

model ① is special case of ② with $\alpha = 1$, i.e. ② is expansion of ① in direction suggested by diagnostics (heavy tail?)

here $\alpha > 0, \ \beta > 0$: $\quad \alpha \sim \Gamma(\epsilon, \epsilon)$, $\epsilon = .001$

$\qquad \beta \sim \Gamma(\epsilon, \epsilon)$

model exp. not necessary: (5)

$p(\alpha \mid y) \doteq 1$, $p(y^* \mid y, ①) \doteq p(y^* \mid y, ②)$

alt. "flat" prior:

$\alpha \sim \Gamma(\epsilon, \epsilon)$ | $\alpha \sim U(0, c_\alpha)$ ← 4.0

$\beta \sim \Gamma(\epsilon, \epsilon)$ | $\beta \sim U(0, c_\beta)$ ← .001

c_α, c_β chosen to not truncate $l(\alpha \mid y)$ & $l(\beta \mid y)$ $\boxed{Q}$ 2 diff

"flat" priors have somewhat diff. ans.; what now? $\boxed{A}$ simulation study: gen. many diff data sets with known truth, try 2 or more Bayesian inference methods & see which is best at recovering truth.

this can be done, e.g. in R/winBUGS

simulations tend to favor $U(0, c)$ over $\Gamma(\epsilon, \epsilon)$ for $(\theta > 0)$ (positive parameters), but uniform on what scale? ex.

$(\mu, \sigma^2) \sim$ diffuse

$(Y_i \mid \mu, \sigma^2) \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$

$U(0, c)$ on σ?
$U(0, c)$ on σ^2?
etc.

(research) (can do your own research by simulation pretty easily)

(with "large n" none of ~~the~~ the diffuse specification details matter)

want $p(\theta \mid y)$, $\theta = (\theta_1, \ldots, \theta_k)$

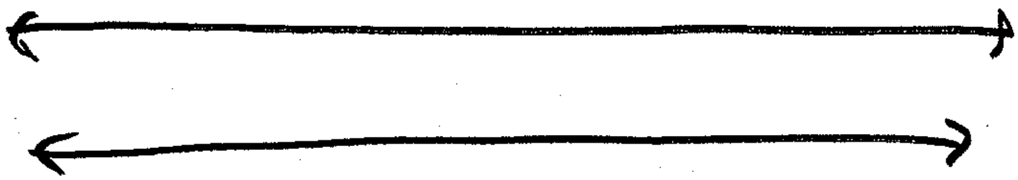
$p(\theta_1 \mid y, \theta_2, \ldots, \theta_k)$

$p(\theta_2 \mid y, \theta_1, \theta_3, \ldots, \theta_k)$

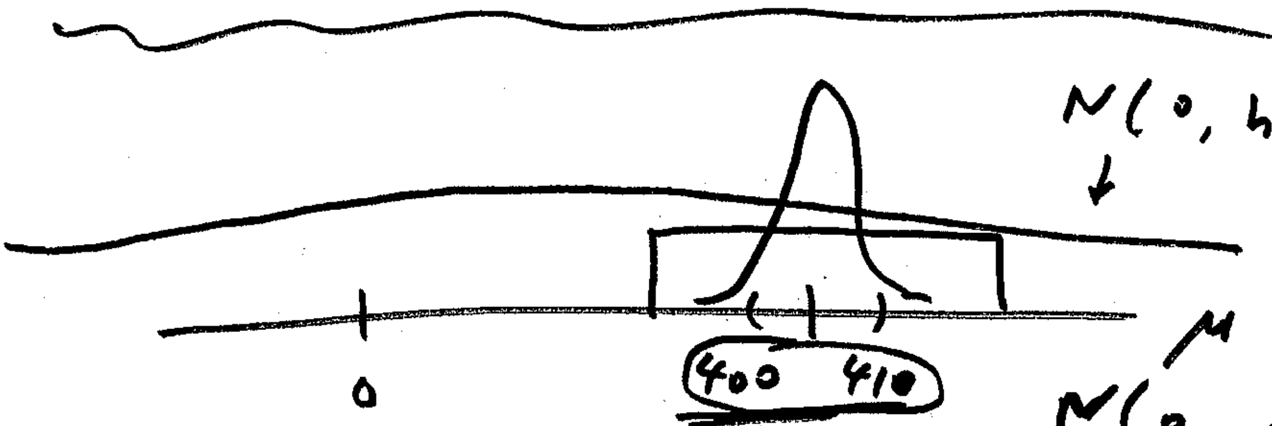
$\vdots$

how draw? from

$U(-\infty, \infty)$ improper



μ



$N(0, \text{huge var})$

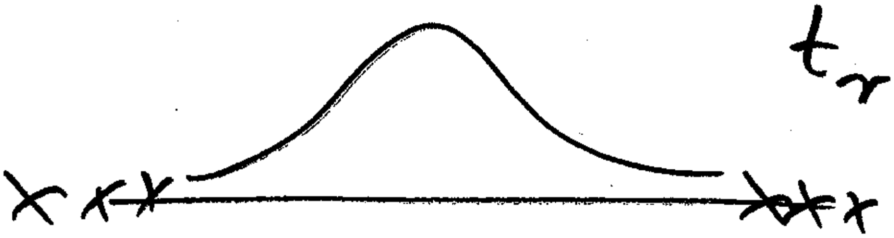
μ

$N(0, \text{tiny precision})$

$\mu \sim N(0, \underbrace{1.0e-6})$

precision $= \frac{1}{1{,}000{,}000}$

prior SD $= 1000$

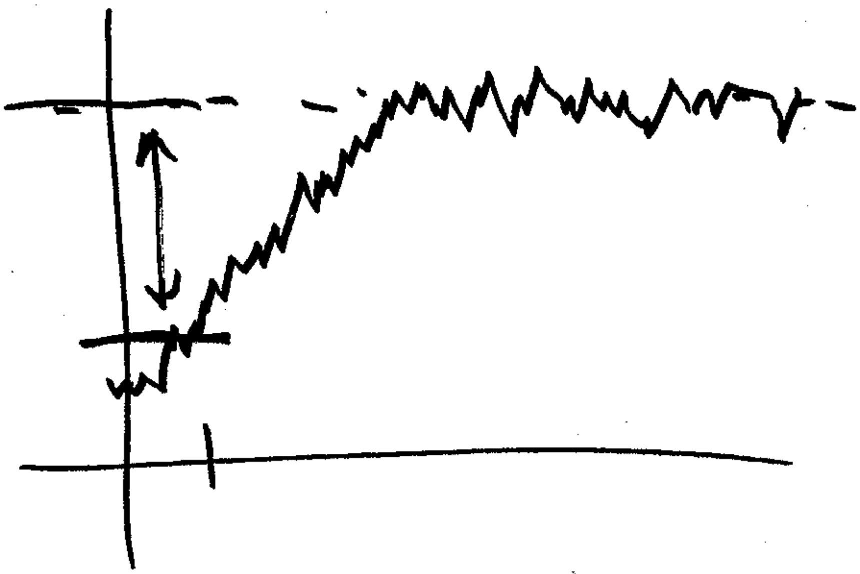


t_r, r small

$r \geq 2$ variance should exist

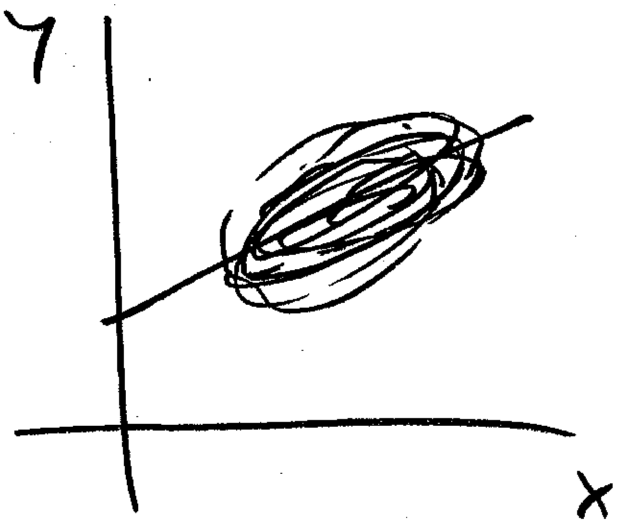
$r \leq$ small value $r \sim U(2, 12)$

not truncate lik. fn



Geweke

Gelman-Rubin



$$\left(\begin{array}{c|c} y_1 & x_1 \\ \vdots & \vdots \\ y_n & x_n \end{array}\right) \quad \text{(OLS)}$$

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

simple linear regression model

(y regressed on x)

(9)

$$\hat{SE}(\bar{\theta}^*) = \frac{\hat{\sigma}_\theta}{\sqrt{m}} \sqrt{\frac{1+\hat{\rho}_1}{1-\hat{\rho}_1}} \leq T$$

1.171   +.86

$\theta = \gamma$

$$m \geq \frac{\hat{\sigma}_\theta^2}{T^2}\left(\frac{1+\hat{\rho}_1}{1-\hat{\rho}_1}\right) \doteq 729\text{k}$$

1.171^2 +.86 $(.005)^2$

model	posterior for μ: mean	SD
→ Gaussian	404.6	0.65
→ t	404.3	0.47

<u>more</u> model uncertainty but <u>less</u> inferential uncertainty

among all symmetric unimodal dens. on $(-\infty, +\infty)$, Gaussian has <u>minimal</u> Fisher <u>info</u> for location (maxent)

$$(C_i \mid \mu_C, \sigma_C^2) \overset{iid}{\sim} N(\mu_C, \sigma_C^2)$$
$$i = 1, \ldots, n_C$$

$$(E_j \mid \mu_E, \sigma_E^2) \overset{iid}{\sim} N(\mu_E, \sigma_E^2)$$

$\bar{C}$ is a good est of μ_C

$\bar{E}$ ———————— μ_E

q of i: $\Delta = (\mu_E - \mu_C)$

$(\bar{E} - \bar{C}) = \hat{\Delta} =$ a good est of $(\mu_E - \mu_C)$

$\hat{SE}(\bar{E} - \bar{C}) = ?$ $\quad$ $\hat{SE}(\bar{E}) = ?$

$\hat{SE}(\bar{C}) = ?$

$$SE(\bar{C}) = \frac{\sigma_C}{\sqrt{n_C}}$$

$\hat{SE}(\bar{C}) = \dfrac{s_C}{\sqrt{n_C}}$, sim. $\hat{SE}(\bar{E}) = \dfrac{s_E}{\sqrt{n_E}}$

$$\hat{SE}(\bar{E}-\bar{C}) = ?$$

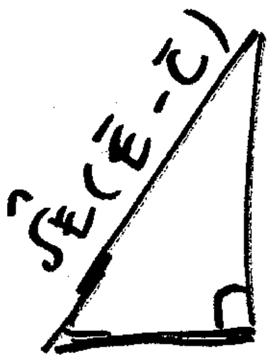
$$SE^2(\bar{E}-\bar{C}) = V(\bar{E}-\bar{C}) \quad \text{(in r.s.)}$$

$$= V(\bar{E}) + V(-\bar{C}) + \underbrace{2\,C(\bar{E}, -\bar{C})}_{0}$$

$$V(cY) = c^2 V(Y) \quad \therefore$$

$$\hat{SE}(\bar{E}-\bar{C}) = \sqrt{\hat{V}(\bar{E}) + \hat{V}(\bar{C})}$$

$$= \sqrt{\left(\hat{SE}(\bar{E})\right)^2 + \left(\hat{SE}(\bar{C})\right)^2}$$

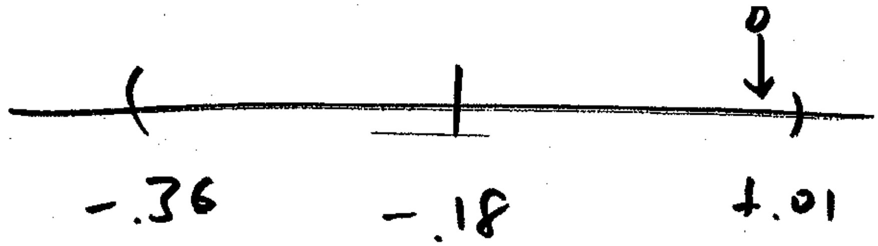


$$\hat{SE}(\bar{E}) = \frac{s_E}{\sqrt{n_E}}$$

$$\hat{SE}(\bar{C}) = \frac{s_C}{\sqrt{n_C}}$$

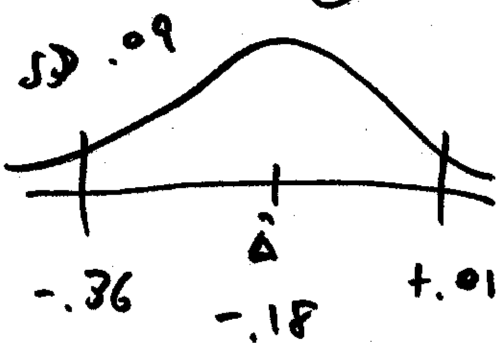
2 indep samples

$$\boxed{\hat{SE}(\bar{E}-\bar{C}) = \sqrt{\frac{s_E^2}{n_E} + \frac{s_C^2}{n_C}}}$$



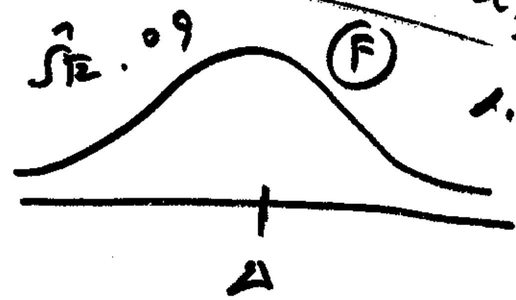
insufficient evidence to reject $H_0 : \mu_C = \mu_E$ (not stat. sig.) (at .05 level)

B



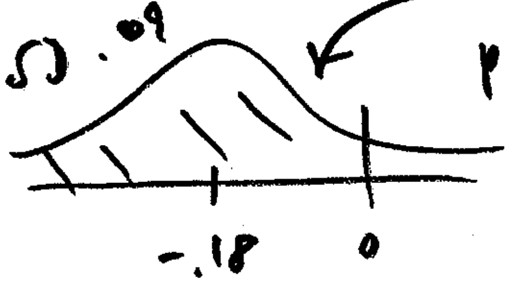
post. dist. for Δ given y

F



dist of $\hat{\Delta}$

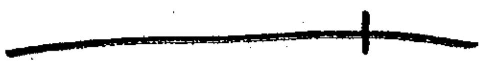
$p($treatment effective$\,|\, y) = p(\Delta < 0 \,|\, y)$



$p(\Delta | y) \doteq 97\%$ post. odds

raw of almost 30

su to 1 that

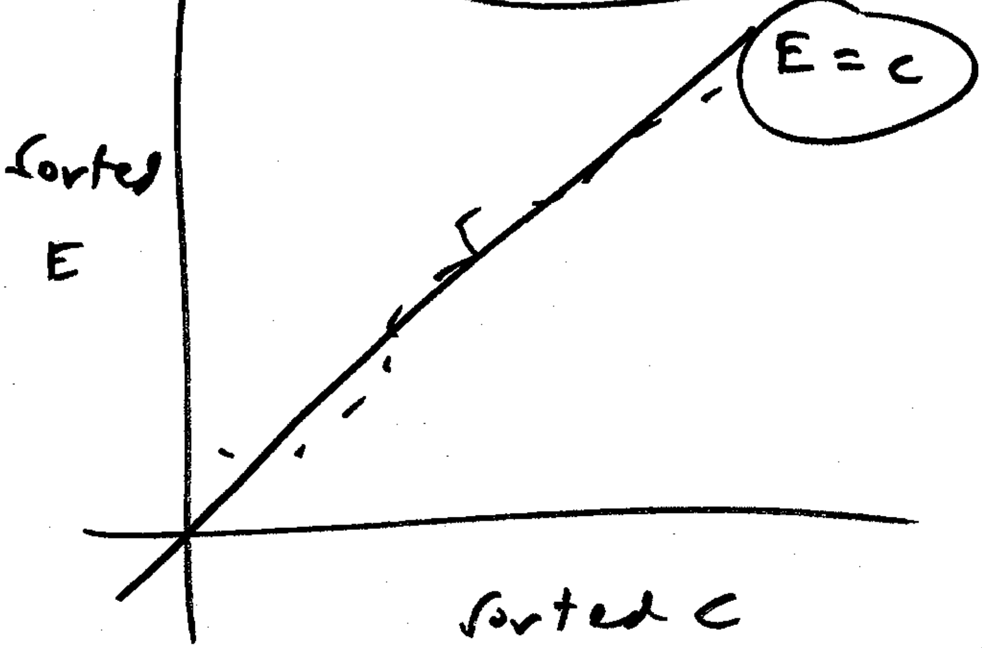


$$\frac{0 - (-.18)}{.09} =$$

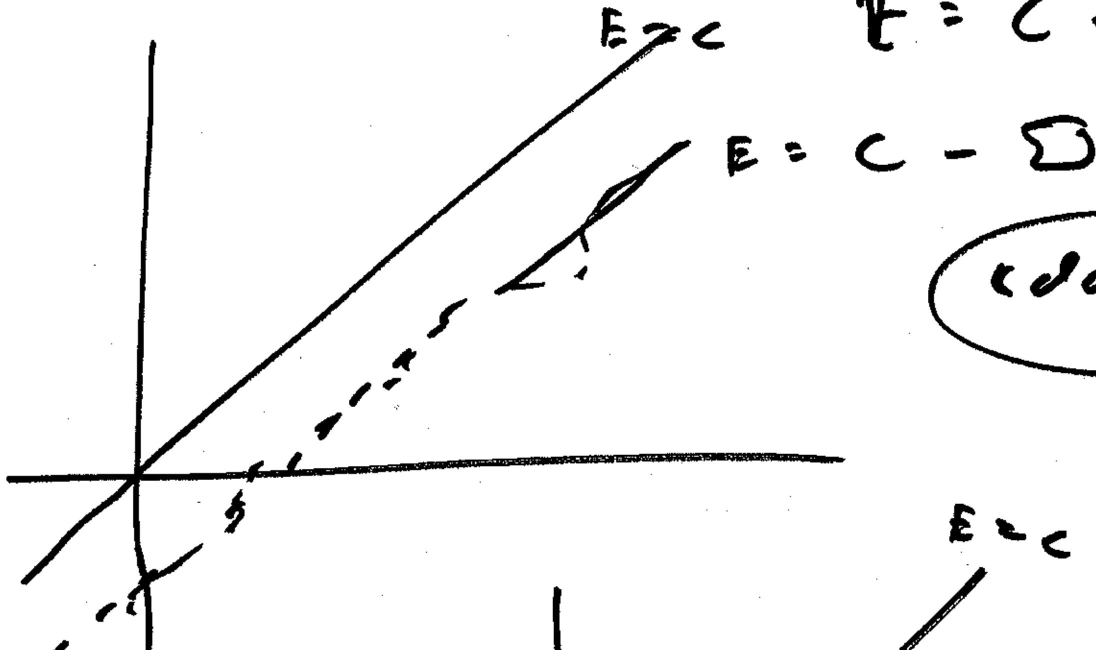
treatment is helpful

no T effect

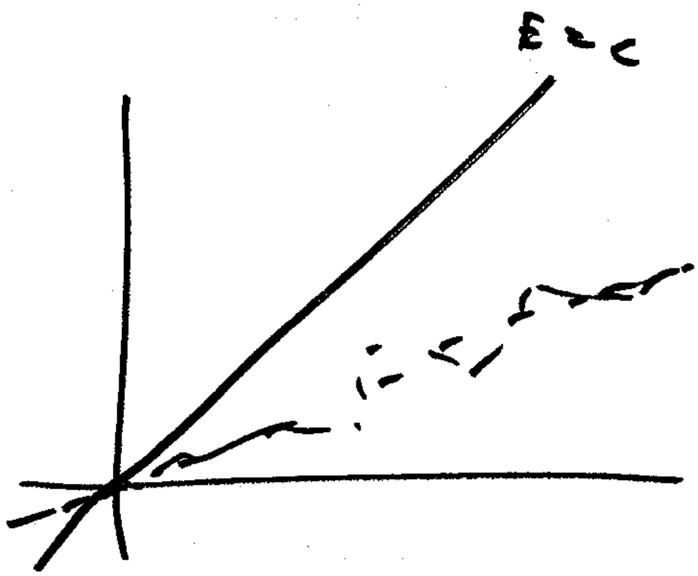
suppose $n_C = n_E$



$E \stackrel{st}{=} C + \Delta$



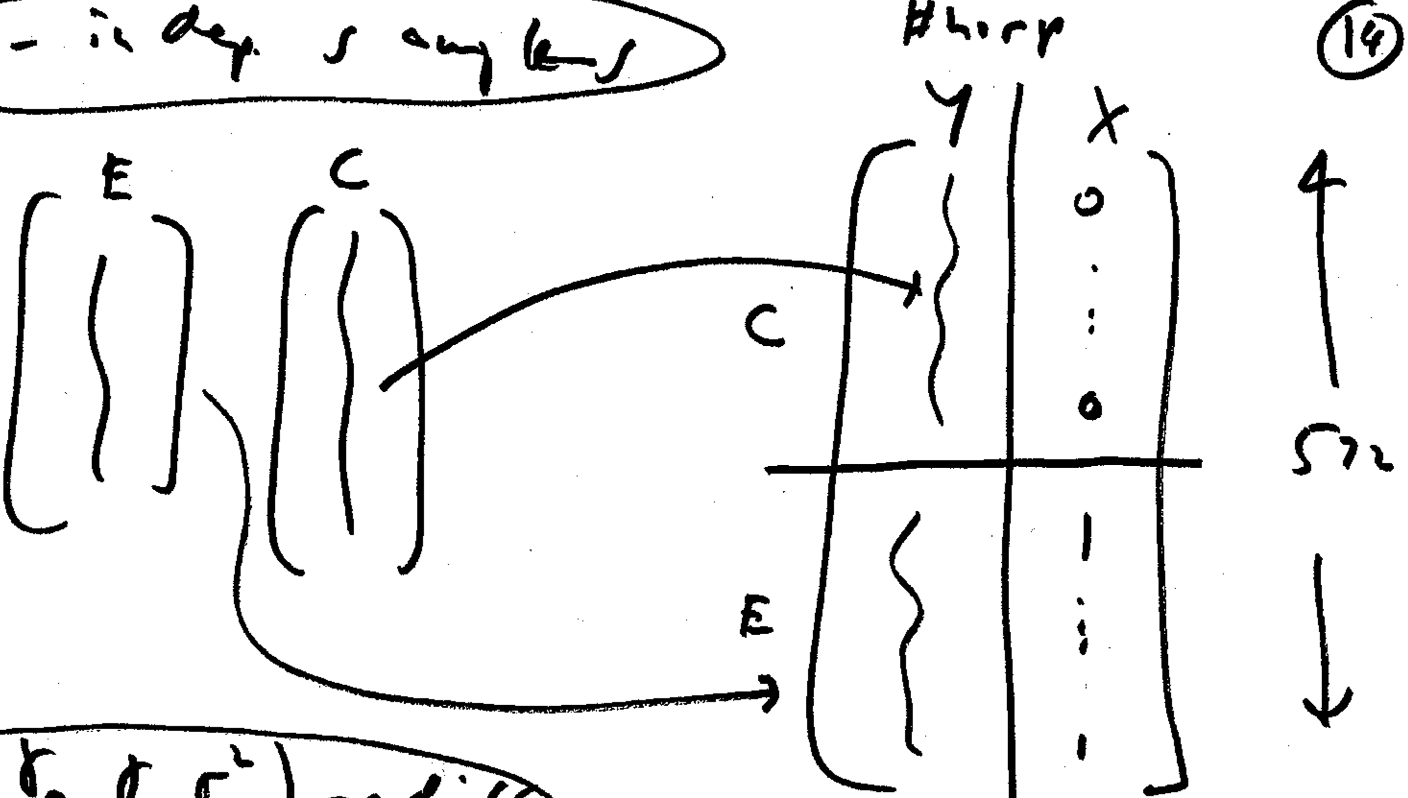
additive



mult.

$E \stackrel{st}{=} (1+\Delta)c$

2 - indep samples



regression

$(\gamma_0, \gamma_1, \sigma_e^2) \sim$ diffuse

$(Y_i \mid \lambda_i) \overset{\text{indep}}{\sim} p(\lambda_i)$

$\left\{ \begin{matrix} 1 & \text{if E} \\ 0 & \text{if C} \end{matrix} \right\}$ baseline health status

$$\log(\lambda_i) = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2}$$

age, sex

unknown $+ \ldots + \gamma_k x_{ik}$

$+ \; e_i$ ⟸ latent

↖ random effects

$e_i \overset{\text{iid}}{\sim} N(0, \sigma_e^2)$