(extra notes 1 Feb) ① part 1

$$\begin{array}{cl} \left[\begin{array}{c} 1 \\ \vdots \\ 1 \end{array}\right] & \alpha \\ \left[\begin{array}{c} 0 \\ \vdots \\ 0 \end{array}\right] & \beta \end{array}$$

mean $\dfrac{\alpha}{\alpha+\beta}$

$(\underline{\alpha+\beta}) = 30$ "prior sample size"

"prior data set"

$$\begin{array}{cl} \left[\begin{array}{c} 1 \\ \vdots \\ 1 \end{array}\right] & s \\ \left[\begin{array}{c} 0 \\ \vdots \\ 0 \end{array}\right] & n-s \end{array}$$

mean $\dfrac{s}{n} = \bar{y}$

$n = 400$ data sample size

always works with conjugate priors

sample data set

---

merge these 2 data sets: ① use likelihood methods on merged data; ② use Bayes methods on sample data alone with Beta(α, β) prior →

if $\theta \sim \text{Beta}(\alpha, \beta)$ then

$$V(\theta) = \frac{\alpha \beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

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

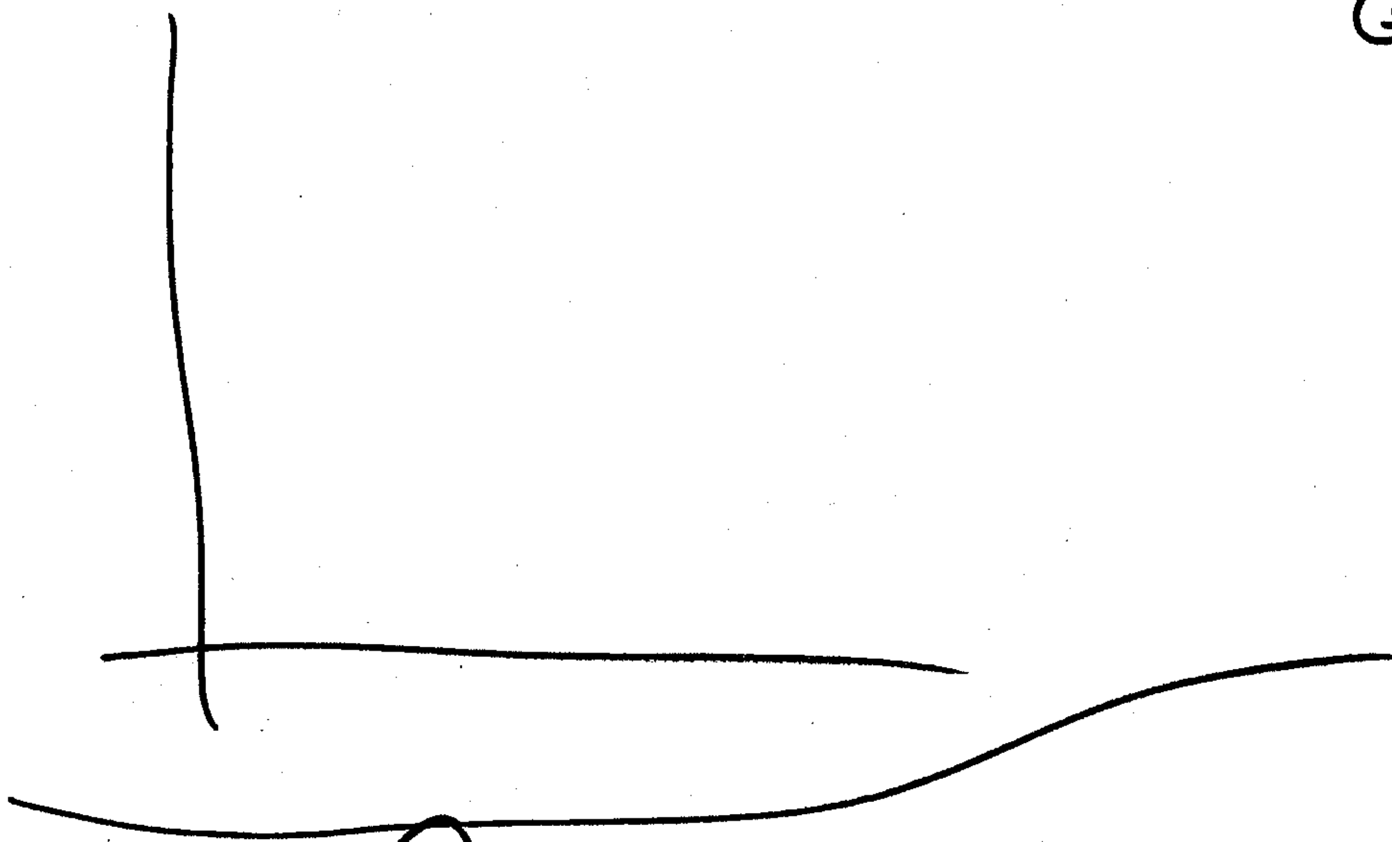
$$(\theta \mid y) \sim \text{Beta}(\alpha^*, \beta^*)$$

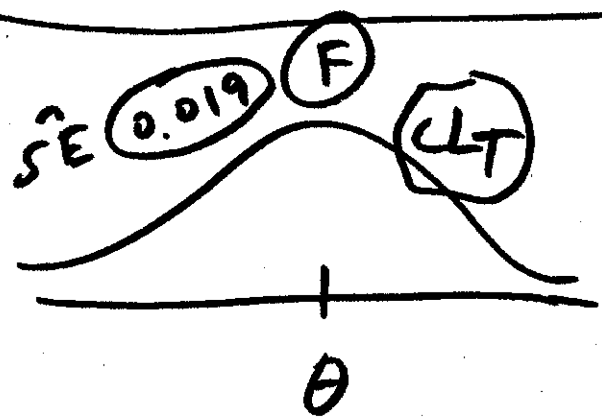
where $\alpha^* = \alpha + s$, $\beta^* = \beta + n - s$

$$+ \; V(\theta \mid y) \doteq \left(\frac{\alpha^*}{\alpha^*+\beta^*}\right)\left(\frac{\beta^*}{\alpha^*+\beta^*}\right)\left(\frac{1}{\alpha^*+\beta^* \cancel{+1}}\right)$$

$$= \frac{\hat{\theta}_B \, (1 - \hat{\theta}_B)}{n_B}$$

post. sample size $(\alpha+\beta+n)$

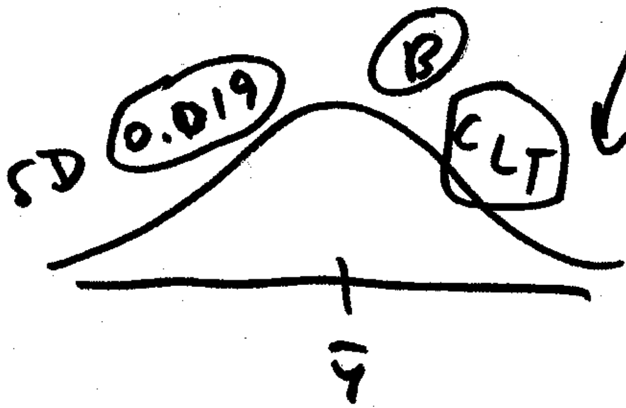




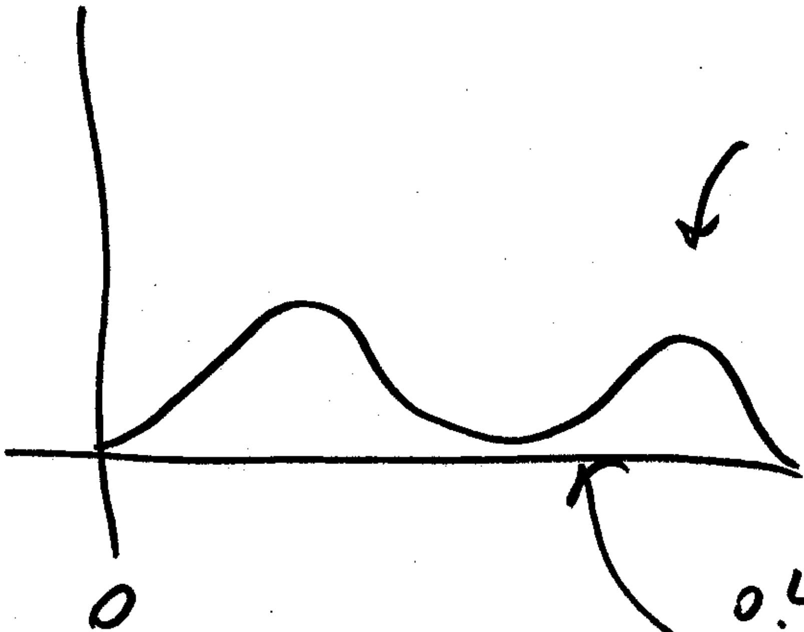
dist. of $\bar{y}$ in repeated sampling

n large

(post.) dist. of θ given $\bar{y}$ & diffuse prior info



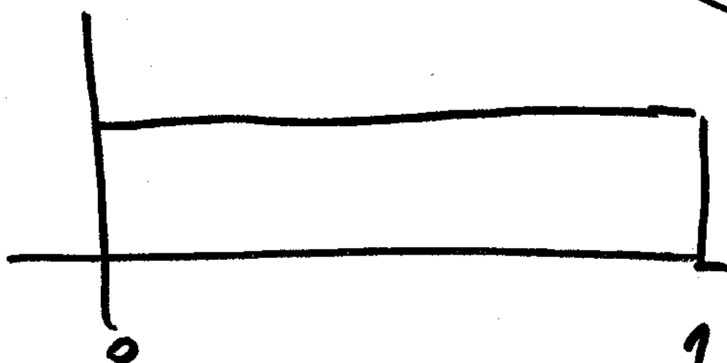
$$c_1 \, e^{-\frac{1}{2c_2}(\theta - \bar{y})^2}$$



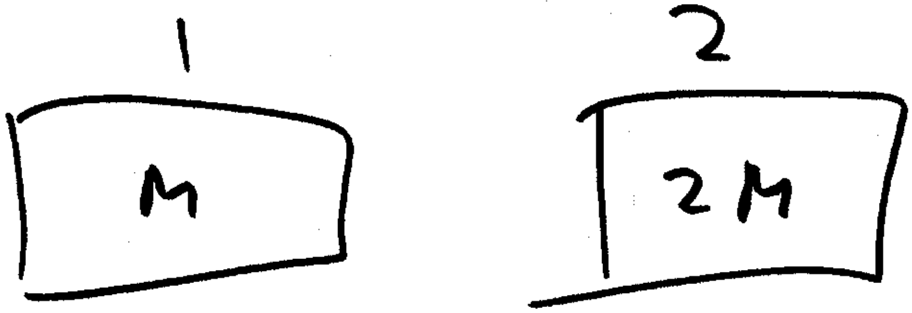
impossible in Beta family

(latent var)

$$p_1 \,\text{Beta}(\alpha_1, \beta_1) + (1-p_1)\,\text{Beta}(\alpha_2, \beta_2)$$



mixture modeling



X = amount of money in my env.

Y = ——————— opp's env.

$$p(X = \underline{x} \mid M = m) = \begin{cases} m & \text{w.p. } \frac{1}{2} \\ 2m & \frac{1}{2} \end{cases}$$

$$= p(Y = y \mid M = m)$$

$a_1 = \{\text{offer to trade}\}$ $\quad$ $a_2 = \{\text{don't}\}$

$U(a_1) = Y$ $\quad$ $U(a_2) = X$

$$E[U(a_1)] = E(Y) = \underbrace{E_M[E(Y \mid M)]}_{= \circledast}$$

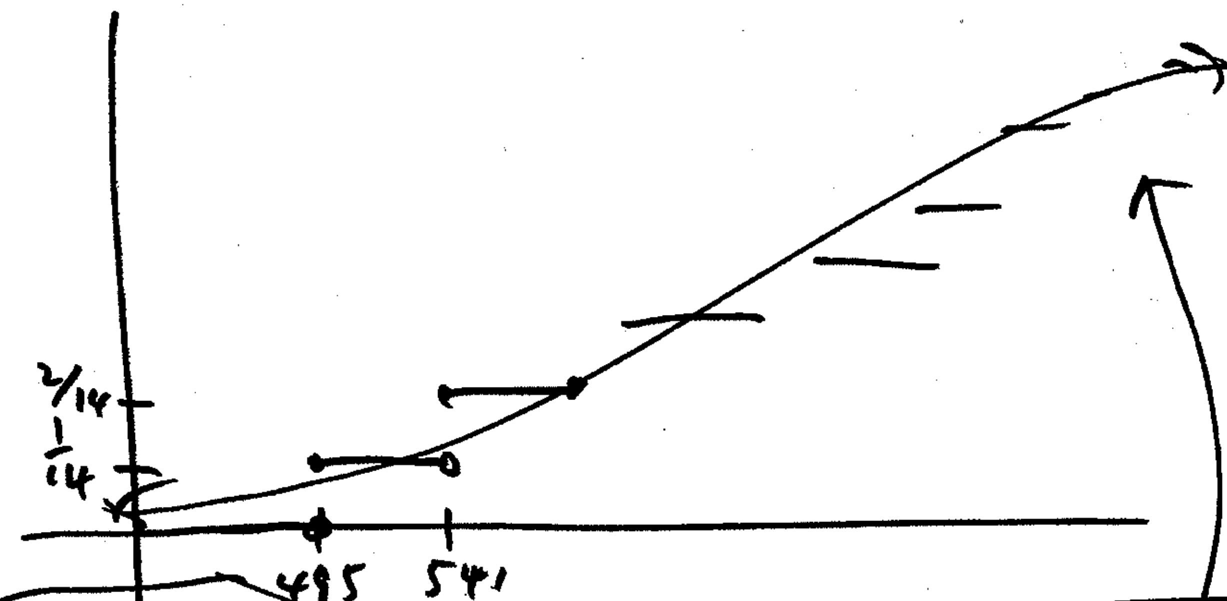
---

part 1 of double-expectation theorem	$E(Y) = E_X[E(Y \mid X)]$

$$E(Q \mid M = m) = \frac{1}{2}(m) + \frac{1}{2}(2m) = \frac{3m}{2}$$

$$E[u(a_1)] = E_M\left[\frac{3M}{2}\right] = \frac{3}{2}E(M)$$

$E[u(a_2)] = \frac{3}{2}E(M)$ so no advantage to trading



Could plot empirical CDF vs. theoretical CDF

$\frac{1}{n}$ mass at y_i

$$Y \sim \mathcal{E}(\lambda) \qquad p_Y(y \mid \lambda) = \begin{cases} \frac{1}{\lambda} e^{-\frac{y}{\lambda}} & \text{for } y > 0 \\ 0 & \text{else} \end{cases}$$

$$F_Y(y) = P(Y \le y) = \int_0^y p_Y(t \mid \lambda)\, dt$$

$$= 1 - e^{-\frac{y}{\lambda}}$$

MLE of λ:

1. write down $\ell(\lambda | y)$
2. take logs to get $\ell\ell(\lambda | y)$
3. diff wrt λ, set to 0 & solve

$$\ell(\lambda | y) = c \, P(Y_1 = y_1, \dots, Y_n = y_n)$$

$$= c \prod_{i=1}^{n} p_Y(y_i)$$

$$= c \prod_{i=1}^{n} \left(\frac{1}{\lambda} e^{-\frac{y_i}{\lambda}} \right)$$

$$= c \, \lambda^{-n} e^{-\frac{1}{\lambda} \sum_{i=1}^{n} y_i} \quad \swarrow s$$

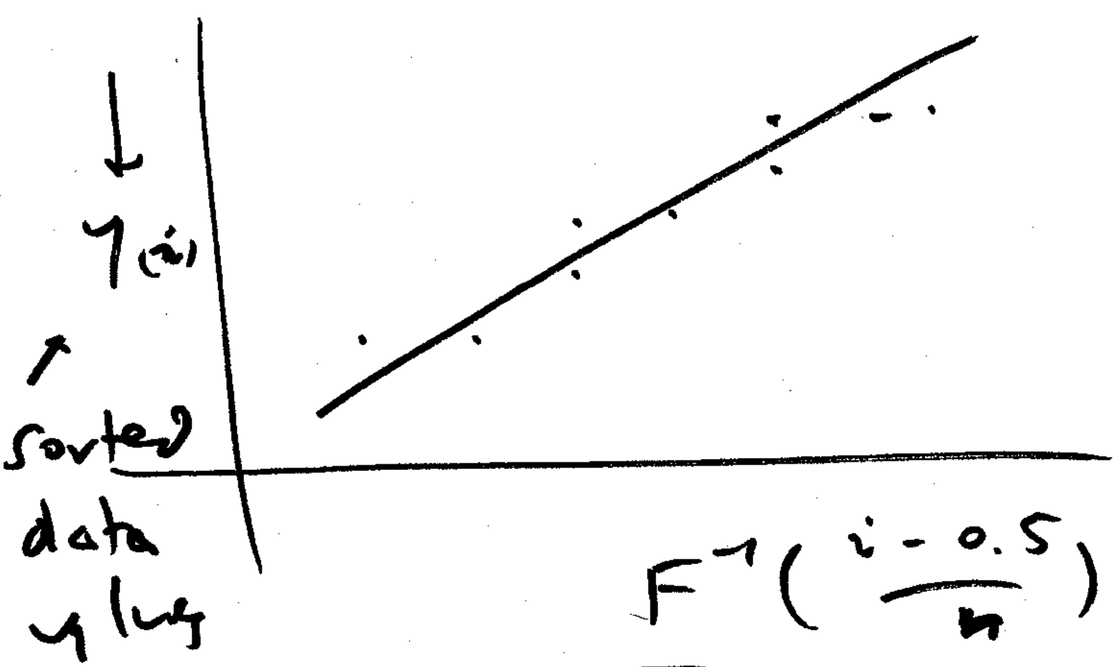
$$\ell(\lambda | y) = c \, \lambda^{-n} e^{-\frac{s}{\lambda}}$$

$\boxed{1}$ here $s = 70612$ & $n = 14$

$\boxed{2}$ $\ell\ell(\lambda \mid y) = -n \ln \lambda - \frac{S}{\lambda}$

$\boxed{3}$ $\frac{\partial}{\partial \lambda} \ell\ell(\lambda \mid y) = -\frac{n}{\lambda} + \frac{S}{\lambda^2} = 0$

iff $\lambda = \hat{\lambda}_{MLE} = \frac{S}{n} = \bar{y}$



QQ plot

$$F_Y(y) = 1 - e^{-\frac{y}{\lambda}} = p$$

$$1 - p = e^{-\frac{y}{\lambda}}$$

$$\ln(1-p) = -\frac{y}{\lambda}$$

$$F^{-1}_{Y}(p) = -\lambda \ln(1-p)$$

---

steps in conjugate Bayesian analysis

① work out $\ell(\lambda | y) = c\, \lambda^{-n} e^{-\frac{s}{\lambda}}$

② find conj. prior, by inspection: mult. ℓ by another f'n of same form get same f'n:

$$\text{conj pr.} = c\, \lambda^{-\alpha}\, e^{-\frac{\beta}{\lambda}}$$

from Gelman $\quad p(\lambda) = \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{-(\alpha+1)} e^{-\frac{\beta}{\lambda}}$

for $\alpha, \beta > 0$

$$\ell(\lambda | y) = c\, \lambda^{-n} e^{-\frac{s}{\lambda}}$$

$$= \Gamma^{-1}(\lambda \mid \underline{n-1}, \overset{n\bar{y}}{s})$$

(11)

$$p(\lambda \mid y) = \left[c\,\lambda^{-(\alpha+1)} e^{-\frac{\beta}{\lambda}}\right] \cdot \left[c\,\lambda^{-n} e^{-\frac{s}{\lambda}}\right]$$

$$= c\,\lambda^{-([\alpha+n]+1)} e^{-\frac{(\beta+s)}{\lambda}}$$

$$= \boxed{\Gamma^{-1}(\lambda \mid \alpha+n, \beta+s)}$$

$n\bar{y}$

$$\lambda \sim \Gamma^{-1}(\alpha, \beta) \rightarrow E(\lambda) = \frac{\beta}{\alpha-1}$$

$$E(\lambda \mid y) = \frac{\beta_0+s}{\alpha_0+n-1} \qquad \lambda \sim \Gamma^{-1}(\alpha_0, \beta_0)$$

$$= \left(\frac{\beta_0}{\alpha_0-1}\right)\left(\frac{\alpha_0-1}{\alpha_0+n-1}\right) + \left(\frac{s}{n}\right)\left(\frac{n}{\alpha_0+n-1}\right)$$

$\therefore$ prior sample size is $n_0 = \alpha_0 - 1$

$$E(\lambda) = \boxed{\frac{\beta_0}{\alpha_0 - 1} = \mu_0} \approx 4500$$

$$\sqrt{V(\lambda)} = \sqrt{\frac{\beta_0^2}{(\alpha_0-1)^2(\alpha_0-2)}} = \boxed{\frac{\beta_0}{(\alpha_0-1)\sqrt{\alpha_0-2}} = \overset{1800}{\sigma}}$$

	Prior	Lik Max	Lik Int	Post
Mean/Est	4500	5044	5884	4858
SD/SE	1800	1348	1774	1080

$$\frac{\partial}{\partial\lambda} ll(\lambda|y) = -\frac{n}{\lambda} + \frac{s}{\lambda^2}$$

$$\frac{d^2}{d\lambda^2} ll(\lambda|y) = \frac{n}{\lambda^2} - \frac{2s}{\lambda^3}$$

$$\left(\qquad \right)\Big|_{\lambda=\frac{s}{n}} = \frac{n}{(s/n)^2} - \frac{2s}{(s/n)^3}$$

$$\hat{I}(\hat{\lambda}_{MLE}) = -\frac{d^2}{d\lambda^2} \ell\ell(\lambda \mid y)\Big|_{\lambda = \hat{\lambda}_{MLE}}$$

$$= \frac{n}{\bar{y}^2}$$

$$\hat{SE}(\hat{\lambda}_{MLE}) = \frac{\bar{y}}{\sqrt{n}} = \frac{\hat{\lambda}_{MLE}}{\sqrt{n}}$$

$$\doteq 13.48$$

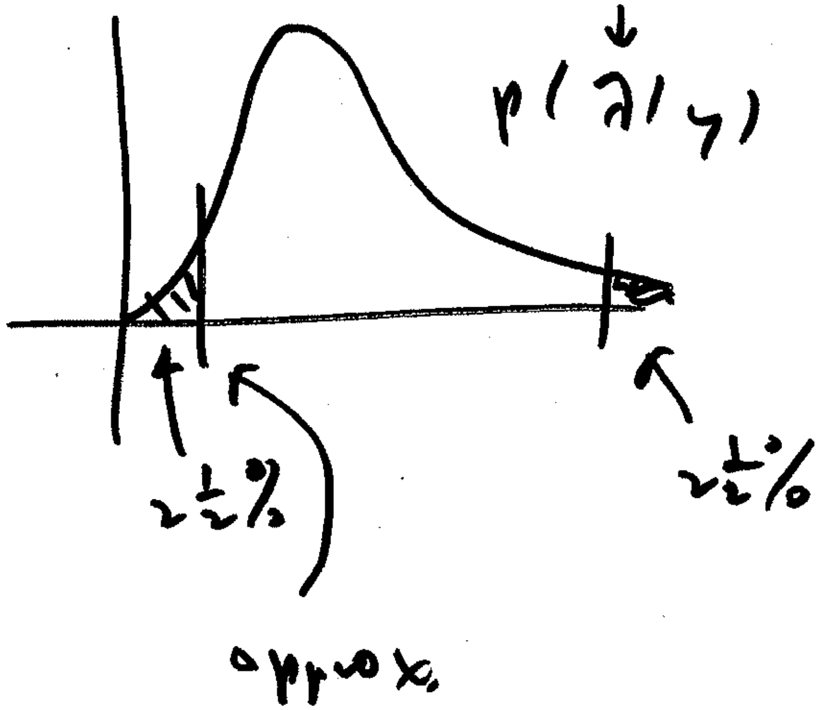
$$\lambda \sim \Gamma^{-1}(\alpha, \beta) \rightarrow$$

$$E(\lambda) = \frac{\beta}{\alpha - 1}, \quad V(\lambda) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$$

lik: $\alpha = n - 1, \quad \beta = s$

$$E(\lambda)_{lik} = \frac{s}{n-2}, \quad \sqrt{V(y)_{lik}} = \frac{s}{n-2\sqrt{n-3}}$$

95% central
post.
int.



MLE: 95% int $\hat{\lambda}_{MLE} \pm 1.96\,\hat{SE}$

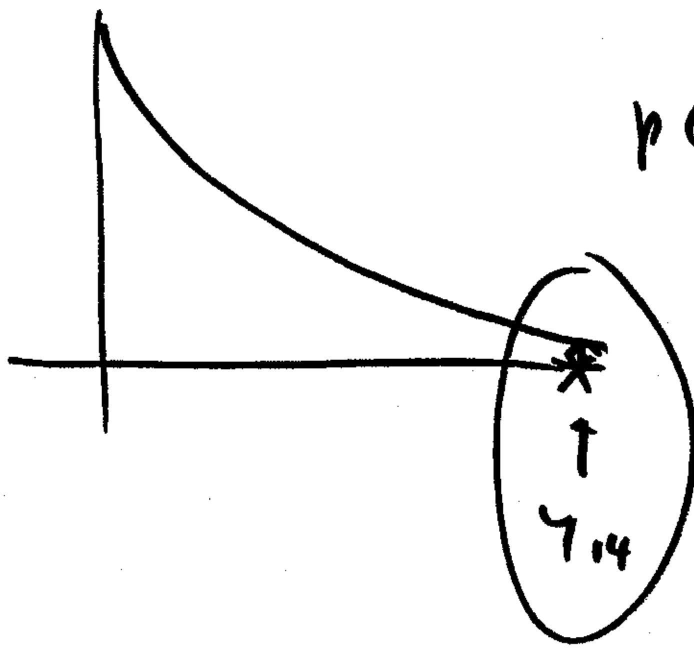
$5044 \pm 1.96\,(1348)$

$$p(y_{n+1}|y) = \int_0^\infty p(y_{n+1}, \lambda|y)\,d\lambda$$

$$= \int_0^\infty \boxed{p(\underline{\underline{y_{n+1}}}|\lambda, \cancel{y})}\, p(\lambda|y)\,d\lambda$$

past, future cond. indep. given truth

$$= \int_0^\infty \frac{1}{\lambda} e^{-\frac{y_{n+1}}{\lambda}} \underbrace{\Gamma^{-1}(\lambda|\alpha^*, \beta^*)}\,d\lambda$$



$p(y_{14} \mid y_1, \dots, y_{13})$

biggest obs

unusual?

---

1. if Bayesian processing then behavior X
2. we observe behavior X

---

3. $\therefore$ Bayesian processing

$A \to B \,;\, B \,;\, \therefore A$ really bad

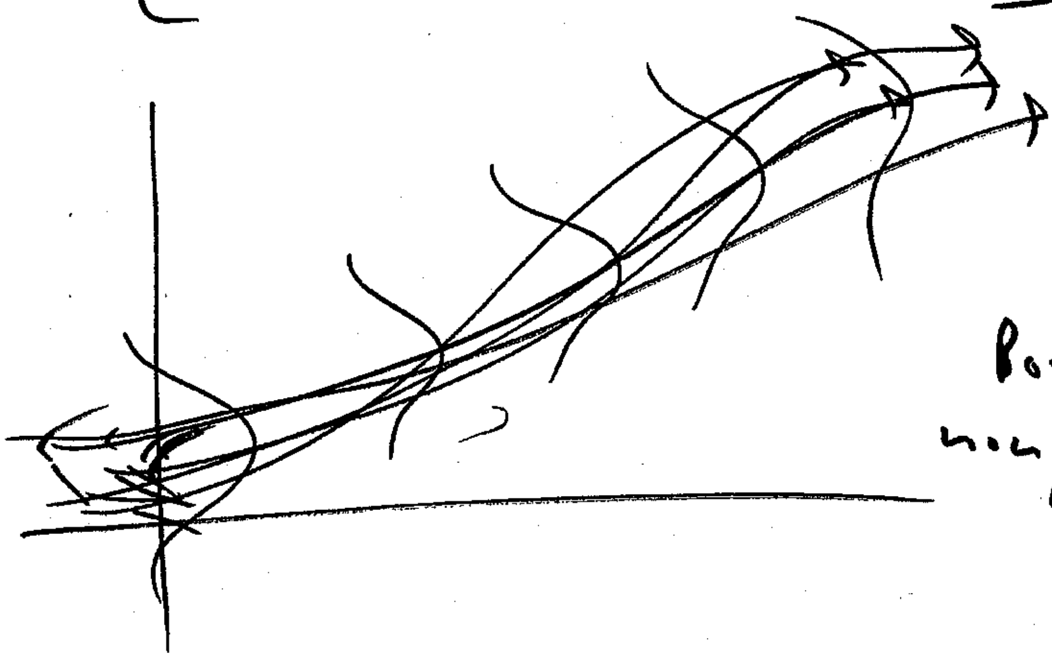
$A_1 \to B \,;\, B \,;\, \therefore A_1$

$A_2 \to B$

Y_i: real-valued; (inf) exch

→ think about CDF $F = F_Y$ in infinite collective

$$\left[\begin{array}{l} F \sim p(F) \\ (Y_i \mid F) \overset{iid}{\sim} F \end{array}\right]$$



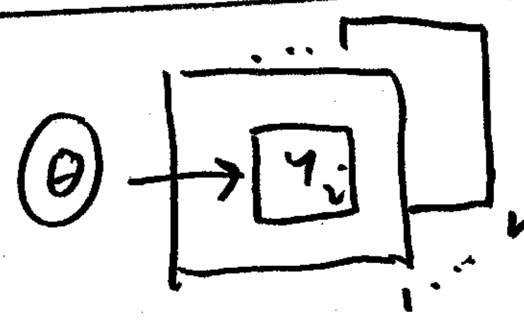
Bayesian nonparametrics

graphical representation of models of the form

$$\left\{ \begin{array}{c} \theta \sim p(\theta) \\ (Y_i \mid \underline{\theta}) \overset{iid}{\sim} F(\theta) \\ i = 1, \ldots, n \end{array} \right\}$$

modus ponens	if A then B; A; therefore B ✓
modus tollens	if A then B; not B; therefore not A ✓
? = names	if A then B; B; therefore A ✗

(post hoc ergo propter hoc) or (the fallacy of affirming the alternative)



□ knowns

○ unknowns

→ probabilistic causation

$\theta \to \boxed{Y}$: knowing θ allows you to simulate from Y