~~to~~ divide by Z ESS to get Z Monte carlo error

∴ Z Estimate is Z expectation, is Z Monte Carlo approx to whatever you asked. gives 95% bounds on a particular prediction

all of these questions are Monte Carlo questions, makes sense just to write down Z integral hes asking for, see Z distri is Z posterior predictive, get samples from that with >>predict(...)

>>cyc_preds %>% for summary=FALSE gives ~390 rows, ~30 columns ∴ 30pts predicted ∴ each column giving all 4000 samples, ∵ 4000 samples in fourth model, ∴ for each column getting all Z samples for Z particular prediction ∴ can average these samples in Z right way to do any monte carlo

inference worth more than convergence (convergence easiest worth 10-20%)

inference more than 20%.

Modelling, writting model down, talking about Z priors, & model checking, & interplay betw them worth more

Week 11 + 12: Decision theory

1.] A set of possible decisions $d \in D$

2.] A set of possible outcomes/uncertain states $\theta \in \Theta$ $(\theta(d), \theta)$

3.] Consequences (numerical) associated with {decision & outcome pairs} (d, θ) pairs

4.] $p(\theta)$ (probab over Z outcomes

Ex 5.1 1.) D • d take measures
• $\tilde{d}$ dont take measures

2.) outcomes Θ • θ is severe weather • $\tilde{\theta}$

3.) (payoff table) consequences:

	θ	$\tilde{\theta}$
d	-300	-100
$\tilde{d}$	-1000	0

in £10^3

4.] $p(\theta) = p \in [0,1]$

for now Z best decision maximises Expected Money Value (EMV) ω)

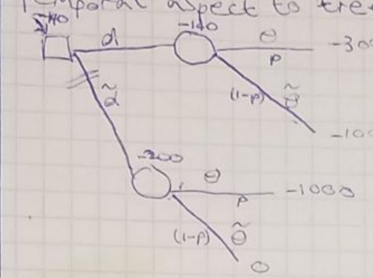
$EMV(d) = \sum_{\Theta} M(d,\theta) P(\Theta = \theta) = -300p - 100(1-p) = -100(1+2p)$

$EMV(\tilde{d}) = -1000p$ ∴ is $EMV(d) > EMV(\tilde{d})$ choose d eg let $p = 0.2$:

$\therefore EMV(d) = -140$, $EMV(\tilde{d}) = -200$ $\therefore$ choose d

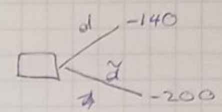
A decision tree has 2 types of node. decision nodes (square) 2 chance nodes (circles) 2 shows all paths to all possible outcomes

Temporal aspect to tree

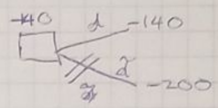


if $p = 0.2$

$\therefore$ step ①:



then for ②:



Roll back procedure: Start at 2 right hand side of 2 tree

① As you reach a chance node, mark it with 2 expected val of 2 consequence to 2 right (then remove 2 tree to 2 right)

② As each decision node is reached, mark it with 2 expected val of 2 branch that is highest to its right, then remove 2 tree to 2 right

repeat until 2 start node is reached

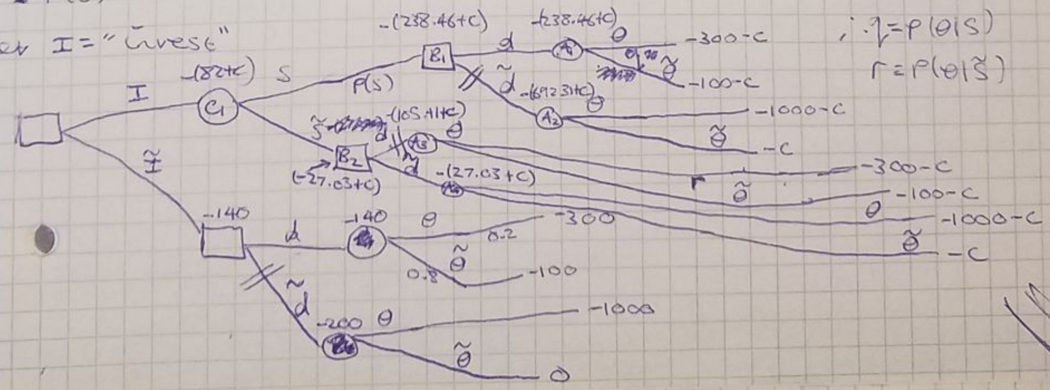
Ex 5.2 / $d, \tilde{d}, \theta, \tilde{\theta}$. Suppose we can, before choosing $d/\tilde{d}$, invest £C in a new forecast system that is "90% accurate"

let S be event that new system gives 'severe forecasts'. then '90% accuracy' will mean $P(S|\theta) = 0.9 = P(\tilde{S}|\tilde{\theta})$ $\left\{ P(\tilde{S}|\tilde{\theta}) = \frac{P(\tilde{\theta}|\tilde{S})P(\tilde{S})}{P(\tilde{\theta})} = \frac{P(\tilde{\theta}|\tilde{S})(1-P(S))}{1-P(\theta)} \quad P(S|\theta) = \frac{P(\theta|S)P(S)}{P(\theta)} \right\}$

let I = "invest"

$\therefore q = P(\theta|S)$

$r = P(\theta|\tilde{S})$



to divide by Z Ess to get Z Monte ...

$P(\theta|S)$, $P(\theta|\tilde{S})$ $\{P(\tilde{\theta}|S) = 1 - P(\theta|S)\}$, $P(S)$ $P(\theta) = 0.2$

$\therefore P(\theta|S) = \dfrac{P(S|\theta)P(\theta)}{P(S)}$ by Bayes $P(S) = P(S|\theta)P(\theta) + P(S|\tilde{\theta})P(\tilde{\theta})$ by LOTP

$\{\because \theta \,\&\, \tilde{\theta}$ form a partition$\}$ $= 0.9 \times 0.2 + 0.1 \times 0.8 = \frac{13}{50} \Longrightarrow$

$P(\tilde{S}) = \frac{37}{50}$ $\therefore$ $P(\theta|S) = \dfrac{(9/10)(2/10)}{13/50} = \frac{9}{13}$ $\therefore$

$P(\tilde{\theta}|S) = \frac{4}{13}$

$P(\theta|\tilde{S}) = \dfrac{P(\tilde{S}|\theta)P(\theta)}{P(\tilde{S})} = \dfrac{0.1 \times 0.2}{37/50} = \frac{1}{37}$ $\therefore$ $P(\tilde{\theta}|\tilde{S}) = \frac{36}{37}$

at A1.] $EMV(d|I,S) = -(300+c)P(\theta|S) - (100+c)P(\tilde{\theta}|S) = -(238.46+c)$

$\{$A2: $EMV(\tilde{d}|I,S) = -(1000+c)P(\theta|S) - cP(\tilde{\theta}|S) = -(1000+c)\frac{9}{13} - c\frac{4}{13} =$

$-(\frac{9000}{13} + c) \approx -(692.31+c)$

A3: $EMV(d|I,\tilde{S}) = -(300+c)P(\theta|\tilde{S}) - (100+c)P(\tilde{\theta}|\tilde{S}) = -(300+c)\frac{1}{37} - (100+c)\frac{36}{37}$

$= -(\frac{3900}{37} + c) \approx -(105.41+c)$

A4: $EMV(\tilde{d}|I,\tilde{S}) = -(1000+c)P(\theta|\tilde{S}) - cP(\tilde{\theta}|\tilde{S}) = -(1000+c)\frac{1}{37} - c\frac{36}{37} =$

$-(\frac{1000}{37} + c) \approx -(27.03+c)$ $\therefore$

$A_4 > A_1, A_2, A_3$

$A_1 > A_2$ $\therefore$ B_1 is $-(238.46+c)$

$A_4 > A_3$ $\therefore$ B_2 is $-(27.03+c)$

C1:] $EMV(I) = -(238.46+c)P(S) - (27.03+c)P(\tilde{S}) = -(238.46+c)\frac{13}{50} - (27.03+c)\frac{37}{50}$

$= -(82+c)$

$\therefore C_1 > -140$ is: $-(82+c) > -140$ $\therefore -82-c > -140$ $\therefore 58 > c$

$\therefore$ choose I is $c < 58$ $\{$this does make physical sense (Sanity check)$\}$

A2.] $EMV(\tilde{d}|I,S) = -(1000+c)P(\theta|S) - cP(\tilde{\theta}|S) = -(692.31+c)$

A3: $-(105.41+c)$ A4: $-(27.03+c)$

C1: $EMV(I) = -(238.46+c)P(S) - (27.03+c)P(\tilde{S}) = -(82.00+c)$

Choose I is $-(82+c) > -140 \iff c < 58$

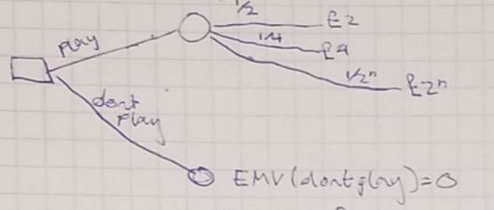
if a severe forecast: take measures, else dont take measures

Consider Z game: Toss a fair coin until Z first tail is thrown. You recieve £2^n as a prize, where n is Z number of tosses you had. how much would you pay to play Z game?

{I would pay ? note: $P(\text{Tail}) = 0.5$ $\therefore$ $\frac{1}{0.5} = 2 = E(n)$ $n \geq 1$ $\therefore$

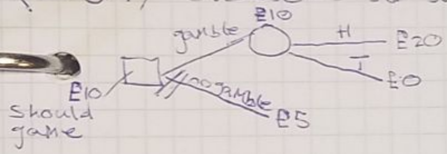
$2^2 = 4$ $\therefore$ Should bet any amount less than £4, I would:

$2^1 = 2$ $\therefore$ only pay any amount less than £2}

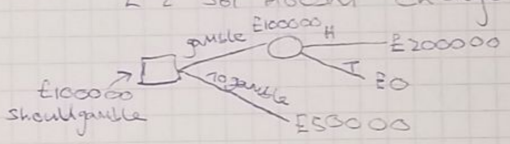


$$EMV(\text{play}) = \sum_{i=1}^{\infty} P_i x_i = \sum_{i=1}^{\infty} 2^i \frac{1}{2^i} = \sum_{i=1}^{\infty} 1 = \infty$$

\Ex 5.4/ choose £5 for sure or £20 probab 0.5, 0 otherwise



but could add more zeros to Z question & Z Sol doesnt change:



\Ex 5.5/ Decision 'get breakdown cover' d,

outcomes: 'you break down' (θ_1)

Z rewards are $r_{11}, r_{12}, r_{21}, r_{22}$

$d_2 = \tilde{d}_1$ $\theta_2 = \tilde{\theta}_1$

	θ_1	θ_2
d_1	r_{11}	r_{12}
d_2	r_{21}	r_{22}

Let r_1, r_2 be rewards then:

\Def 5.1/ $r_1 \overset{*}{>} r_2$ means ~~means~~ you prefer r_1 to r_2

\Def 5.2/ $r_1 \overset{*}{\geq} r_2$ means you view r_1 at least as good as r_2

\Def 5.3/ $r_1 \sim^* r_2$ means you are indifferent betw r_1 & r_2

$r_{22} \overset{*}{>} r_{12} \overset{*}{>} r_{11} \overset{*}{>} r_{21}$

$d_1 \equiv \begin{cases} r_{11} & \text{probability } P(\theta_1) \\ r_{12} & \text{probability } P(\theta_2) \end{cases}$ $d_2 \equiv \begin{cases} r_{12} & \text{probability } P(\theta_1) \\ r_{22} & \text{probability } P(\theta_2) \end{cases}$

\Def 5.4/ for rewards r_1, r_2 & $p \in [0,1]$ Z 'gamble' where you recieve r_1 with probab p & r_2 otherwise is written $pr_1 +_g (1-p) r_2$ this extends to any finite reward set $r_1, \dots, r_k$ with

~~to divide~~ by Z Ess ~~to get~~ Z Monte ...

probabs $p_1, \dots, p_k$ so Z gamble is $p_1 r_1 +_g p_2 r_2 +_g \dots +_g p_k r_k$

$0 \le p_i \le 1 \quad \sum p_i = 1$

$\therefore d_1 \equiv p r_{11} +_g (1-p) r_{12} = G_1 \quad P(\theta_1) = p \quad \{\therefore d_1 \equiv G_1\}$

$d_2 \equiv p r_{21} +_g (1-p) r_{22} = G_2$

do we prefer G_1 or G_2? eg $G_1 \overset{*}{>} G_2$

\Def 5.5 Utility:/ A utility function, U on gambles $g = \{p_1 r_1 +_g \dots +_g p_k r_k\}$ over rewards $\{r_1, \dots, r_k\}$ assigns a real number, $U(g)$ to each gamble g, subject to Z following 2 conditions:

(i) if $g_1 \overset{*}{<} g_2$ then $U(g_1) < U(g_2)$

if $g_1 \overset{*}{\sim} g_2$ then $U(g_1) = U(g_2)$

(ii) for any two rewards r, s & $p \in [0,1]$

$U(pr +_g (1-p)s) = pU(r) + (1-p)U(s)$

<u>Comment 1</u>: (i) says utility respects preferences

(ii) says utility of a gamble is equal to its expected utility

<u>Comment 2</u>: (i)+(ii) imply Z optimal decision betw gambles is to choose Z gamble with Z highest expected utility

~~Nut~~ $\overset{*}{>}$ ~~Normal~~ $\overset{*}{>}$ Dark $\overset{*}{>}$ White

procedure for defining utilities

(i) choose Z best reward give $U(\text{Nut}) = 1$

(ii) choose Z worst reward set $U(\text{White}) = 0$

(iii) for another reward (eg choose normal) consider choice betw normal for sure & Z gamble (for Z best reward) Nut with probab p & white (worst reward) probab $1-p$

$\therefore G_1 = \text{Normal} \quad G_2 = p \times \text{Nut} +_g (1-p) \times \text{white}$

let q be Z smallest p st $G_2 \overset{*}{>} G_1$

then $U(\text{Normal}) = q$

(iv) repeat step iii for all rewards (all chocolates) $\therefore$

to show this procedure defines a utility, we need to show it satisfies (Def 5.5) {(i) it respects Z preferences}

(i) is automatic (ii) requires more assumptions

Assumptions (a) Rewards are <u>coherently</u> comparable

implications: (1) for any pair of rewards r_i, r_j exactly one of Z following holds either: $r_i <^* r_j$ $r_i \,{}^*\!\!> r_j$ $r_i \sim^* r_j$

(2) $r_i <^* r_j$, $r_j <^* r_k$ then $r_i <^* r_k$

(b) Gambles can be coherently compared two parts: (1) if $r_1 \le^* r_2$ & $p < q$ ($p = 1-q$) then $pr_2 +_g qr_1 <^* qr_2 +_g pr_1$ (Monotonicity, M property)

(2) if $s_1 \le^* s_2$ then $ps_1 +_g qr \le^* ps_2 +_g qr$ (Substitutionability, S property)

(c) we can find reward R at least as good as all reward, & r at least as bad as all rewards. [boundedness] $s \in S$

$r \le^* s \le^* R$ $r, R \in S$

Suppose general reward set S. $r \le^* s \le^* R$ $s \in S$ (by C)

Let $U(s)$ be Z number for which $s \sim^* U(s)R +_g (1-U(s))r$

$U(s)$ is uniquely defined eg if $U > V$ $\therefore$ $UR +_g (1-U)r \,{}^*\!\!>$ ~~$VR +_g (1-V)$~~ $VR +_g (1-V)r$. (by M)

$\therefore$ if $s \sim^* UR +_g (1-U)r$ then $s \,{}^*\!\!> VR +_g (1-V)r$ (by S)

Claim $U(\cdot)$, so defined, is a Utility function over Z rewards

\proof/ (i) Show if ~~~~ $s_1 \,{}^*\!\!> s_2$ $U(s_1) > U(s_2)$

Suppose $s_1 \,{}^*\!\!> s_2$ $\therefore$ $s_1 \sim^* U(s_1)R +_g (1-U(s_1))r$

$s_2 \sim^* U(s_2)R$ ~~$+_g$~~ $+_g (1-U(s_2))r$ $\therefore$

$U(s_1)R +_g (1-U(s_1))r \,{}^*\!\!> U(s_2)R +_g (1-U(s_2))r$ $\therefore$ by (M) $\Rightarrow$

$U(s_1) > U(s_2)$

(ii) Let $s_3 = ps_1 +_g (1-p)s_2$ $U(s_3) = pU(s_1) + (1-p)U(s_2)$

s_1, s_2 as above $\Rightarrow$ (2) $s_3 \sim^* U(s_3)R +_g (1-U(s_3))r$

~~$s_3 \sim^* ps_1 +_g (1-p)s_2$~~ $s_3 \sim^* ps_1 +_g (1-p)s_2 \sim^*$

$\sim P[U(S_1)R +_g (1-U(S_1))r] +_g (1-P)[U(S_2)R +_g (1-U(S_2))r]$ $\{by\ S\}$

$\sim (PU(S_1)+(1-P)U(S_2))R +_g (P(1-U(S_1))+(1-P)(1-U(S_2)))r$

by comparing with ⊛ : $U(S_3) = PU(S_1) + (1-P)U(S_2)$ $\square$

\thm 5.2:/ your utility is unique upto an arbit pos linear transformation

~~lemma~~ (i) suppose U is your utility funct on rewards & $a, b \in \mathbb{R}$ $a>0$ then $V(r) = aU(r)+b$ is also a utility func

(ii) Suppose U, V are both utilities on a reward set then $\exists a>0, b \in \mathbb{R}$ s.t. $V(r) = aU(r)+b$ $\forall r$

\proof/ (i) we check $v(r)$ is also a utility

U is a utility $S_1 \overset{*}{>} S_2 \Rightarrow U(S_1) > U(S_2) \Longrightarrow$

$aU(S_1)+b > aU(S_2)+b \Longrightarrow V(S_1) > V(S_2)$ let $S_3 = PS_1 +_g (1-P)S_2$

show $V(S_3) = PV(S_1)+(1-P)V(S_2)$

U is utility so $U(S_3) = PU(S_1) + (1-P)U(S_2)$.

$aU(S_3) = aPU(S_1) + a(1-P)U(S_2)$

$aU(S_3)+b = aPU(S_1) + a(1-P)U(S_2)+b$

$V(S_3) = aPU(S_1) + a(1-P)U(S_2) + (1+P-P)b = P(aU(S_1)+b)+(1-P)(aU(S_2)+b) =$
$PV(S_1) + (1-P)V(S_2)$ $\square$

Choosing a decision $\equiv$ choosing a gamble $\Longrightarrow$
Z optimal decision is Z decision with highest expected utility

if U is a utility $a>0$, $b \in \mathbb{R}$, $V = aU+b$ & our decision choices are $d_1, \dots, d_P$ then if $U(d_i) = \max_j \{U(d_j)\}$ then
$V(d_i) = E[V(d_i)] = aU(d_i)+b = \max_j \{V(d_j)\}$ $\therefore$ maximising expected utility for U or V gives Z same decision

(def 5.5)(ii) rewards r, s $U(pr +_g (1-p)s) = PU(r) + (1-P)U(S)$

\lemma 5.1/ $U\left(\sum_{i=1}^{J}{}_g\, P_i r_i\right) = U(P_1 r_1 +_g P_2 r_2 +_g \cdots +_g P_J r_J) = \sum_{i=1}^{J} P_i U(r_i)$

\Defining Expectation:/ $E[X]$: ① $\bar{x}$ s.t. $c(X-\bar{x})$ c chosen opponent

Coherence: Dont want to buy bets that certainly lose

② $\bar{x}$: you suffer loss $L=\left(\frac{X-\bar{x}}{k}\right)^2$ coherence: Dont prefer to lose an amount when could lose less for sure

you specify $P(A)=P_A$, $P(A\wedge B)=P_{AB}$, $P(A\wedge \tilde{B})=P_{A\tilde{B}}$ show $P_A > P_{AB}+P_{A\tilde{B}}$ is incoherent : set all 3 ∴ accepting bet : $G_T=C_1(A-P_A)+C_2(AB-P_{AB})+C_3(A\tilde{B}-P_{A\tilde{B}})$

suppose $B=1$ ∴ $G_T=C_1A-P_AC_1+C_2A-P_{AB}C_2+C_3P_{A\tilde{B}}$ ∴ set $C_1=-C_2$ ∴

$G_T=C_2(P_A-P_{AB})+C_3P_{A\tilde{B}}$ ∴ if $B=0$: $G_T=C_1A-P_AC_1-C_2P_{AB}+C_3A-C_3P_{A\tilde{B}}$

$=-C_2A+C_2(P_A-P_{AB})+C_3A-C_3P_{A\tilde{B}}$ ∴ if $C_3=C_2$ ∴ $=C_3\underbrace{(P_A-P_{AB}-P_{A\tilde{B}})}_{+ve}$ is not random at all ∴ if $P_A > P_{AB}+P_{A\tilde{B}}$ ∴ $P_A-P_{AB}-P_{A\tilde{B}}>0$ ∴ could keep playing without loss ∴ opponent C_3 -ve

conditional $E[\ ]/P(\ )$ event H $P(A|H)=\bar{a}$ s.t. you recieve penalty $L=H\left(\frac{A-\bar{a}}{k}\right)^2$ called off bet

i.e.e. , boundlessness

Exchangeability $X_1,\ldots,X_n$ if your joint distri for X_i doesnt depend on lables

if Exchangeability: act as though $P(X|\theta)$ exists moreover $\exists\ \pi(\theta)$ s.t. $P(X)=\int P(X|\theta)\pi(\theta)d\theta$ is Representation thm

\Computation:/ Monte Carlo for bayesian integrals

inverse CDF & rejection sampling to sample basic distri's

MCMC to convert above into correlated samples from from complex $\pi(\underline{\theta}|\underline{y})$ because of MonteCarlo

\preliminary course/ $E_1, E_2, \ldots$ mutually exclusive $E_i \cap E_j = \varnothing$, $i \neq j$. $P(E^c)=1-P(E)$ $P(E_1 \cup E_2)=P(E_1)+P(E_2)-P(E_1 \cap E_2)$ is inclusion exclusion

$P(E|F)=\frac{P(E\cap F)}{P(F)}$ ∴ $P(E\cap F)=P(E|F)P(F)=P(F|E)P(E)$

if E, F independent : $P(E\cap F)=P(E)P(F)$ ∴ $P(E|F)=P(E)$

if $E_1, E_2, \ldots$ partition S : exactly one E_i must happen for any event A: $P(A)=\sum_i P(A|E_i)P(E_i)$ LoTP

Bayes thm: $P(E|F)=\frac{P(F|E)P(E)}{P(F)}$ $\therefore$ $P(E|F)=\frac{P(E\cap F)}{P(F)}=\frac{P(F|E)P(E)}{P(E)}$

big X is r.v. & small x is Z value it takes

$F(-\infty)=0$ $(P(X\le-\infty))$ $\qquad F(\infty)=1$ $(P(X\le\infty))$

if $x_1\le x_2$: $F(x_1)\le F(x_2)$ $\qquad P(a\le X\le b)=P(X\le b)-P(X\le a)=F(b)-F(a)$

$\therefore$ discrete only: $P(X=x)=F(x)-F(x-1)=P(x)$ is pmf

For continuous X: pmf is always 0 : $P(X=x)=0$ $\because$ $P(X=x)=$

$\lim_{h\to0}P(x-h<X<x+h)=\lim_{h\to0}F(x+h)-F(x-h)=0$.

Z pdf $f(x)$ is Z derivative of Z cdf $f(x)=F'(x)=\frac{dF}{dx}$ $\therefore$ $F(x)=\int_{-\infty}^{x}f(t)\,dt$

$f(x)=F'(x)=\lim_{h\to0}\frac{F(x-h)-F(x)}{h}$ $\qquad$ small h: $P(x<X<x+h)\approx hf(x)$

discrete: $0\le P(x)\le1$ $\quad \sum_{x\in D_x}P(x)=1$ $\qquad$ continuous: $f(x)\ge0$ $\quad \int_{-\infty}^{\infty}f(x)\,dx=1$

$f(x)$ cont $[a,b]$ $\Delta x=\frac{b-a}{n}$ $\qquad \int_a^b f(x)\,dx=\lim_{n\to\infty}\sum_{j=1}^{n}f(c_j)\Delta x$ $\;: x_{j-1}\le c_j\le x_j$

$$\frac{d}{dx}\int_{a(x)}^{b(x)}f(t)\,dt=f(b(x))b'(x)-f(a(x))a'(x)$$

if X has pdf $f(x)\propto x$ $\quad x\in[0,2]$ $\therefore$ $\int_{-\infty}^{\infty}f(x)\,dx=1$, $f(x)=kx$ $\therefore$

$1=\int_{-\infty}^{\infty}kx\,dx=k\int_0^2 x\,dx=k\left[\frac{1}{2}x^2\right]_0^2=k\frac{1}{2}\left[2^2-0^2\right]=2k=1$ $\therefore$ $k=\frac{1}{2}$

$\int f(u(x))u'(x)\,dx=\int f(u)\,du$ $\qquad \int_a^b f(u(x))u'(x)\,dx=\int_{u(a)}^{u(b)}f(u)\,du$

eg $\int_a^b g(x)\,dx$ $\quad g(x)=f(u(x))u'(x)$ for substitution try guessing a sub for

Z hardest part eg $\int x(1+x^2)^3\,dx$ try $u=1+x^2$

$$\int_a^b f'(x)g(x)\,dx=\left[f(x)g(x)\right]_a^b-\int_a^b f(x)g'(x)\,dx$$

a function of n variabs $f(x_1,\dots,x_n)$ over domain D:

$$\int_{D_{x_1}}\int_{D_{x_2}}\cdots\int_{D_{x_n}}f(x_1,\dots,x_n)\,dx_n\,dx_{n-1}\cdots dx_1$$

\Ex/ $f(x,y)=x$ on circle $x^2+y^2\le1$ $\therefore$ change coords $\therefore$ let

$x=x(u,v)$, $y=y(u,v)$ functions of new coords u,v

$\iint_{D(x,y)}f(x,y)\,dx\,dy=\iint_{\Gamma(u,v)}f(x(u,v),y(u,v))\,|J(u,v)|\,du\,dv$ $\qquad$ J is Jacobian $J(u,v)=\begin{vmatrix}\frac{\partial x}{\partial u} & \frac{\partial y}{\partial u}\\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v}\end{vmatrix}$

$=\frac{\partial x}{\partial u}\frac{\partial y}{\partial v}-\frac{\partial y}{\partial u}\frac{\partial x}{\partial v}$ $\therefore$

polar coord transformation $x=r\cos\theta$, $y=r\sin\theta$ $\quad (x,y)\to(r,\theta)$ $\therefore$

$$\iint_{D:\,x^2+y^2\le 1} x\,dx\,dy = \iint_{T(D)} r\cos\theta\,|J(r,\theta)|\,dr\,d\theta = \int_0^{2\pi}\int_0^1 r\cos\theta\,|J(r,\theta)|\,dr\,d\theta$$

$\frac{\partial x}{\partial r} = \cos\theta$, $\frac{\partial y}{\partial r} = \sin\theta$, $\frac{\partial x}{\partial \theta} = -r\sin\theta$, $\frac{\partial y}{\partial \theta} = r\cos\theta$ $\therefore$

$$J(r,\theta) = \begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & +r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r(\cos^2\theta + \sin^2\theta) = r \quad \therefore$$

$$\int_0^{2\pi}\int_0^1 r^2\cos\theta\,dr\,d\theta = \int_0^{2\pi}\left[\tfrac{1}{3}r^3\cos\theta\right]_0^1 d\theta = \int_0^{2\pi}\tfrac{1}{3}\cos\theta\,d\theta = \left[\tfrac{1}{3}\sin\theta\right]_0^{2\pi} = 0$$

Binomial distri: $X \sim Bin(n,\theta)$ $\quad \theta \in [0,1]$

$P(X=k) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$ $\quad k = 0,1,2,\dots$ $\quad$ if $n=1$ $\quad X \sim Ber(\theta)$

$X_1,\dots,X_n \sim Ber(\theta)$ $\quad X = \sum X_i \sim Bin(n,\theta)$

Uniform X is discrete: "discrete unif distri": "all outcomes equally likely" $P(X=k) = \frac{|k|}{|S|}$

X cont on $[a,b]$ $\quad$ pdf $f(x) = \begin{cases} \frac{1}{b-a} & x\in[a,b] \\ 0 & \text{otherwise} \end{cases}$

Normal: $X|\mu,\sigma^2 \sim N(\mu,\sigma^2)$ $\quad f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{\frac{-1}{2\sigma^2}(x-\mu)^2\right\}$ $\quad x\in(-\infty,\infty)$

Gamma: $X|\alpha,\beta$ $\quad (\alpha \ge 1)$ $\quad X \sim Gamma(\alpha,\beta)$

pdf $f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$ $\quad \Gamma(\alpha) = \int_0^\infty u^{\alpha-1}e^{-u}\,du$ $\quad \alpha \ge 0$

$\Gamma(1) = 1$ $\quad \Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ $\forall \alpha$ $\therefore$

for integer α: $\Gamma(\alpha) = (\alpha-1)!$

Beta $X|a,b$ $\quad X \sim Beta(a,b)$ $\quad f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, & x\in[0,1] \\ 0, & \text{otherwise} \end{cases}$

Discrete X: $E(X) = \sum_{x\in D} xP(x)$ $\quad$ cont X: $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$

expectation applies to funcs: $\quad$ discrete: $E(g(x)) = \sum_{x\in D} g(x)P(x)$

cont: $E(g(x)) = \int_{-\infty}^{\infty} g(x)f(x)\,dx$

$var(x) = E([x-E(x)]^2) = E(x^2) - E(x)^2 = E(x^2 - 2xE(x) + E(x)^2) =$

$E(x^2) - 2E(x)E(x) + E(x)^2 = E(x^2) - E(x)^2$

Ex/ $x \sim unif(0,2)$ $\therefore f(x) = \frac{1}{2}$ $\therefore E(x) = \int_0^2 x f(x)\,dx = \int_0^2 \frac{x}{2}\,dx = 1$

$E(x^2) = \int_0^2 x^2 f(x)\,dx = \int_0^2 \frac{1}{2}x^2\,dx = \left[\frac{x^3}{6}\right]_0^2 = \frac{4}{3}$ $\therefore$

$var(x) = E(x^2) - E(x)^2 = \frac{4}{3} - 1^2 = \frac{1}{3}$

\Ex/ $X \sim Ber(\theta)$ $\therefore E(X) = 0\,P(X=0) + 1\,P(X=1) = 0(1-\theta) + 1(\theta) = \theta$

$E(X^2) = 0^2\,P(X=0) + 1^2\,P(X=1) = 1^2\theta = \theta$, $\therefore var(X) = \theta - \theta^2 = \theta(1-\theta)$

\Ex/ $X \sim Exp(\theta) = Gamma(1, \lambda)$ $\therefore E(X) = \int_0^\infty x\lambda e^{-\lambda x}\,dx$

$f'(x) = \lambda e^{-\lambda x}$ $\quad g(x) = x \quad f = -e^{-\lambda x}$, $g'(x) = 1$ $\therefore$

$E(X) = [f(x)g(x)]_0^\infty - \int_0^\infty f(x)g'(x)\,dx = [-xe^{-\lambda x}]_0^\infty + \int_0^\infty e^{-\lambda x}\,dx = 0 + [-\frac{1}{\lambda}e^{-\lambda x}]_0^\infty$

$= +\frac{1}{\lambda} = \frac{1}{\lambda}$

$E(X^2) = \int_0^\infty x^2\lambda e^{-\lambda x}\,dx = [-x^2 e^{-\lambda x}]_0^\infty + \int_0^\infty 2x e^{-\lambda x}\,dx =$

$0 + 2\frac{1}{\lambda}\int_0^\infty x\lambda e^{-\lambda x}\,dx = 2\frac{1}{\lambda}\frac{1}{\lambda} = 2\frac{1}{\lambda^2} = [-\frac{2x}{\lambda}e^{-\lambda x}]_0^\infty + 2\int_0^\infty \frac{1}{\lambda}e^{-\lambda x}\,dx$

$= 2[-\frac{1}{\lambda^2}e^{-\lambda x}]_0^\infty = \frac{2}{\lambda^2}$ $\therefore var(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}$

Standard deviation of X, s.d.(X) is $\sqrt{var(X)}$

\Quantiles/ Median ~~is~~ 2-quantile

quartiles = 4-quantiles

Suppose r.v. X has cdf $F(x)$ q-quantiles are Z set

$\{F^{-1}(\frac{1}{q}), F^{-1}(\frac{2}{q}), \ldots, F^{-1}(\frac{q-1}{q})\}$

$X \sim unif(0,3)$ median X? $\quad \int_0^M f(x)\,dx = \frac{1}{2} = [\frac{x}{3}]_0^M = \frac{M}{3} = \frac{1}{2}$ $\therefore M = \frac{3}{2}$

\Ex/ $X \sim Exp(\lambda) \equiv Gamma(1, \lambda)$ $\therefore$ M : $\int_0^M \lambda e^{-\lambda x}\,dx = \frac{1}{2}$ $\therefore$

$[-e^{-\lambda x}]_0^M = 1 - e^{-\lambda M} = \frac{1}{2}$ $\therefore e^{-\lambda M} = \frac{1}{2}$ $\therefore M = \frac{1}{\lambda}\ln(2)$

$X_1, \ldots, X_N$ r.v.'s joint cdf : $F_{X_1,\ldots,X_N}(x_1, \ldots, x_N) =$

$P(X_1 \le x_1 \cap X_2 \le x_2 \cap \cdots \cap X_N \le x_N)$

$\therefore$ if $\underline{X} = (X_1, \ldots, X_N)$: $F_{\underline{X}}(x)$

joint pmf when r.v.'s are discrete

$p_{\underline{x}}(x_1, \ldots, x_N) = P(X_1 = x_1 \cap X_2 = x_2 \cap \cdots \cap X_n = x_n)$

joint pdf when r.v.'s cont :

$f_{X_1,\ldots,X_N}(x_1, \ldots, x_N) = \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2}\cdots\frac{\partial}{\partial x_N}F_{X_1,\ldots,X_N}(x_1, \ldots, x_N)$

for scalars: $P(\underline{X} \in D)$ is (for cont $\underline{X}$)

$P(\underline{X} \in D) = \int\cdots\int_D f_{X_1,\ldots,X_N}(x_1, \ldots, x_N)\,dx_1\cdots dx_N$

for pmf : $0 \le P(x_1, \ldots, x_N) \le 1$, $\sum_{x_1 \in D_1}\cdots\sum_{x_N \in D_N} P(x_1, \ldots, x_N) = 1$

pdf : $f(x_1, \ldots, x_N) \ge 0$, $\int_{D_1}\cdots\int_{D_N} f(x_1, \ldots, x_N)\,dx_1\cdots dx_N = 1$

Ex/ bivariate r.v. (X,Y) joint pdf $f(x,y) \propto x^2+xy$ $0 \le x \le 1$, $0 \le y \le 1$ $\therefore$ pdf: $\therefore f(x,y) = k(x^2+xy)$ $\therefore$

$$1 = \int_0^1 \int_0^1 k(x^2+xy)\,dx\,dy = k\int_0^1 \left[\frac{x^3}{3} + \frac{x^2}{2}y\right]_0^1 dy = k\int_0^1 \frac{1}{3} + \frac{y}{2}\,dy =$$

$$k\left[\frac{y}{3} + \frac{y^2}{4}\right]_0^1 = k\left(\frac{1}{3}+\frac{1}{4}\right) = \frac{7k}{12} \quad \therefore k = \frac{12}{7} \quad \therefore f(x,y) = \frac{12}{7}(x^2+xy) \text{ on } [0,1]^2$$

$f(x_1, \dots, x_N)$ $\quad P(X_1 > 3)$? $\quad f_{X_1}(x_1)$

marginal CDF of X_i $\quad (X_1, \dots, X_N)$

$$F_{X_i}(x_i) = \lim_{x_j \to \infty} F_{X_1, \dots, X_N}(x_1, \dots, x_N) \;\therefore$$

marginal pdf of X_i : $f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \dots, X_N}(x_1, \dots, x_N)\,dx_1, \dots, dx_{i-1}, dx_{i+1}, \dots dx_N$

$$p_{X_i}(x_i) = \sum_{x_1 \in D_1} \cdots \sum_{x_{i-1} \in D_{i-1}} \sum_{x_{i+1} \in D_{i+1}} \cdots \sum_{x_N \in D_N} p_{X_1, \dots, X_N}(x_1, \dots, x_N)$$

$f(x,y) = \frac{12}{7}(x^2+xy)$ $\quad f(x)$? $\;f(y)$? $\therefore$

$$f_X(x) = \int_0^1 \frac{12}{7}(x^2+xy)\,dy = \frac{12}{7}\left[x^2y + \frac{xy^2}{2}\right]_0^1 = \frac{12}{7}\left(x^2+\frac{x}{2}\right) = \frac{12}{7}x\left(x+\frac{1}{2}\right)$$

$$f_Y(y) = \int_0^1 \frac{12}{7}(x^2+xy)\,dx = \frac{12}{7}\left[\frac{x^3}{3} + \frac{x^2y}{2}\right]_0^1 = \frac{12}{7}\left(\frac{1}{3}+\frac{y}{2}\right) = \frac{12}{21} + \frac{6y}{7}$$

T, R $\quad f_{T,R}(t,r)$ $\therefore$ $P(T>30, r<5_{mm}) = \int_{30}^{\infty}\int_0^5 f_{T,R}(t,r)\,dr\,dt$

$$P(T>30) = \int_{30}^{\infty}\int_0^{\infty} f_{T,R}(t,r)\,dr\,dt = \int_{30}^{\infty} f_T(t)\,dt$$

what is $r = 30$? : $P(T>30)$? $\therefore P(T>30 \mid r=30)$

$\therefore$ conditional pmf of $X_1, \dots, X_n$ given $X_{n+1} = x_{n+1}, \dots, X_N = x_N$ is

$$p_{X_1, \dots, X_n \mid X_{n+1}=x_{n+1}, \dots, X_N = x_N}(x_1, \dots, x_n \mid x_{n+1}, \dots, x_N) = \frac{p(x_1, \dots, x_N)}{p(x_{n+1}, \dots, x_N)}$$

conditional pdf : $f(x_1, \dots, x_n \mid x_{n+1}, \dots, x_N) = \dfrac{f(x_1, \dots, x_N)}{f(x_{n+1}, \dots, x_N)}$

$f(x,y) = \frac{12}{7}(x^2+xy)$ $\quad \therefore f(X \mid Y=0)$:

$$f(x|y) = \frac{f(x,y)}{f(y)} = \frac{\frac{12}{7}(x^2+xy)|_{y=0}}{\frac{12}{7}\left(\frac{1}{3}+\frac{y}{2}\right)|_{y=0}} = \frac{\frac{12}{7}(x^2)}{\frac{12}{7}\left(\frac{1}{3}\right)} = \frac{x^2}{\frac{1}{3}} = 3x^2$$

$$f(x|y) = \frac{\frac{12}{7}(x^2+xy)}{\frac{12}{7}\left(\frac{1}{3}+\frac{y}{2}\right)}$$

$E(g(X_1,\dots,X_N)) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\dots,x_N) f(x_1,\dots,x_N)\,dx_1\cdots dx_N$ if $\underline{X}$ is cont

if Discrete:

$E(g(x_1,\dots,x_N)) = \sum_{x_1\in D_1}\cdots\sum_{x_N\in D_N} g(x_1,\dots,x_N)\,p(x_1,\dots,x_N)$

marginal expectation of X_i is $E(X_i)$ $\therefore$ $g(\underline{X}) = X_i$

$\therefore E(X_i) = \int_{-\infty}^{\infty} x_i f_{X_i}(x_i)\,dx_i$

Ex/ $f(x,y) = \frac{12}{7}(x^2+xy)$ on $[0,1]^2$ $\cancel{f_X(x) = \frac{12}{7}(x^2+2y)}$ $f_X(x) = \frac{12}{7}x(x+\frac{1}{2})$

$E(X) = \int_0^1 x f_X(x)\,dx = \int_0^1 \frac{12}{7}x^2(x+\frac{1}{2})\,dx = \frac{12}{7}\left[\frac{x^4}{4}+\frac{x^3}{6}\right]_0^1 = \frac{12}{7}\left[\frac{5}{12}\right] = \frac{5}{7}$

$E(X) = \int_0^1\int_0^1 x f_{X,Y}(x,y)\,dx\,dy = \int_0^1\int_0^1 \frac{12}{7}(x^3+x^2y)\,dx\,dy = \int_0^1 \frac{12}{7}\left[\frac{x^4}{4}+\frac{x^3y}{3}\right]_0^1 dy =$

$\int_0^1 \frac{12}{7}\left(\frac{1}{4}+\frac{y}{3}\right)dy = \frac{12}{7}\left[\frac{y}{4}+\frac{y^2}{6}\right]_0^1 = \frac{12}{7}\left(\frac{1}{4}+\frac{1}{6}\right) = \frac{12}{7}\cdot\frac{5}{12} = \frac{5}{7}$

$E(g(X_1,\dots,X_N) \mid x_{n+1},\dots,x_N) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\dots,x_n) f_{X_{1:n}|X_{(n+1):N}}(x_1,\dots,x_n \mid x_{n+1},\dots,x_N)\,dx_1\cdots dx_n$

$E(X) = E(E(X|Y)) \longrightarrow$ wrt Y : $\int E(X|Y) f_Y(y)\,dy$

$\int x f_X(x)\,dx$ $\qquad$ $\int x f_{X|Y}(x|y)\,dx$

$E(E(X|Y)) = \int E(X|Y) f_Y(y)\,dy = \int\left[\int x f_{X|Y}(x|y)\,dx\right] f_Y(y)\,dy = \int\int x f_{X|Y}(x|y) f_Y(y)\,dx\,dy$

$= \int\int x f_{X,Y}(x,y)\,dx\,dy = E(X) = \int (x f_X(x))\,dx = E(X)$

X, Y vec valued r.v.'s covariance between X, Y is

$Cov(X,Y) = E((X-E(X))(Y-E(Y))^T) = E(XY^T) - E(X)E(Y)^T$ $\therefore$

$var(X) = Cov(X,X)$,

$var(aX+bY+c) = a\,var(X)a^T + b\,var(Y)b^T + 2a\,Cov(X,Y)b^T$

$Cov(X_i, Y_j)$ $\therefore$ let $\Sigma_{XY} = Cov(X,Y)$ $\therefore$

$\underset{\in[-1,1]}{corr}(X,Y) = diag(\Sigma_{XX})^{-1/2}\,\Sigma_{XY}\,diag(\Sigma_{YY})^{-1/2}$

$y = \{y_1,\dots,y_n\}$ $\therefore$ $Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n$ $\therefore$ $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})$ $\qquad$ $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$

Dependence between different Y's $\quad Y$ is heights $\therefore$

$n_f \therefore Y_{f,1},\dots,Y_{f,n_f} \qquad n_m \therefore Y_{m,1},\dots,Y_{m,n_m}$

$Y \sim N(\mu,\sigma^2) \quad \therefore Y_f \sim N(\mu_f,\sigma^2), \; Y_m \sim N(\mu_m,\sigma^2)$

$M_i = \mathbb{1}$ (male) $\therefore Y_i \mid M_i = m_i \sim N(\mu+\beta m_i, \sigma^2) \quad , \mu_f = \mu \; , \mu_m = \mu+\beta$

$(Y_i, X_i) \therefore Y_i \mid X_i = x_i \sim N(\alpha+\beta x_i, \sigma^2)$ linear regression

$Y_1,\dots,Y_n$ iid $N(\mu,\sigma^2)$ infer $\mu,\sigma^2 \quad (Y_1=y_1,\dots,Y_n=y_n)$

$E(Y_i)=\mu \; , var(Y_i)=\sigma^2 \quad$ set μ esti: $\hat{\mu} \; , \sigma^2 : \hat{\sigma}^2 \therefore$

$\hat{\mu}=\bar{y} \; , \hat{\sigma}^2 = s^2$

$Y_i \sim f_{Y_i}(y_i;\theta) \qquad \underline{Y}=Y_1,\dots,Y_n \therefore f_{\underline{Y}}(\underline{y};\theta)=\prod_{i=1}^{n} f_{Y_i}(y_i;\theta) \therefore$

$f_Y(\underline{y};\theta)$ is Z likelihood $\therefore$ set $\hat{\theta} = \arg\max_\theta f_Y(y;\theta) = \text{mle}$

Ex/ $Y_i \sim Ber(\theta) \quad i=1,\dots,10 \quad$ let $\sum_{i=1}^{10} Y_i = 7 \therefore$

$f_{Y_i}(y_i) = \begin{cases} \theta & \text{if } y_i=1 \\ 1-\theta & \text{if } y_i=0 \end{cases} \therefore$ likelihood : $L(\theta)=\prod_{i=1}^{n} f_{Y_i}(y_i)=\theta^7(1-\theta)^3 \therefore$

Loglikelihood : $l(\theta)=7\ln\theta+3\ln(1-\theta) \therefore$ ~~to~~ $l'(\theta)=\frac{7}{\theta}-\frac{3}{1-\theta}=0 \therefore$

$\frac{7}{\hat{\theta}}=\frac{3}{1-\hat{\theta}} \therefore 3\hat{\theta}=7(1-\hat{\theta}) \therefore 10\hat{\theta}=7 \therefore \hat{\theta}=\frac{7}{10}$

if seen $y \therefore \hat{\theta}(y)=\frac{y}{10}$

$Y \sim f(\theta) \quad \hat{\theta}(y) \quad \therefore \hat{\theta}(Y)$ is r.v. is Z estimator

$\hat{\theta}(Y)$ is Z sampling distri of an estimator

$sd(\hat{\theta}(Y))$ is called Z standard error

$\therefore y \sim Bin(10,\theta) \quad \therefore E(y)=10\theta \quad , var(y)=10\theta(1-\theta)$

$\therefore E(y^2)=var(y)+E(y)^2=10\theta(1-\theta)+100\theta^2 = 90\theta^2+10\theta$

$E(\hat{\theta})=E\left(\frac{y}{10}\right)=\frac{E(y)}{10}=\theta$

$E(\hat{\theta}^2)=\frac{E(y^2)}{100}=\frac{9}{10}\theta^2+\frac{1}{10}\theta \therefore$

$sd(\hat{\theta})=\sqrt{var(\hat{\theta})}=\sqrt{\frac{9}{10}\theta^2+\frac{1}{10}\theta-\theta^2}=\sqrt{\frac{1}{10}(\theta(1-\theta))}$

interval estimator $\theta \in [\hat{\theta}-c, \hat{\theta}+c]$

Let $\alpha = P(\theta \in [\hat{\theta}(Y)-c, \hat{\theta}(Y)+c])$ is α-level confidence interval

$\therefore 100(\alpha)\%$ CI

\Ex/ $Y_1, \dots, Y_n$ iid $N(\mu, \sigma^2)$ σ^2 known $\therefore$ mle, s.e; S_e, $\hat{\mu} \pm S_e$ CI

$L(\theta) = f_Y(y; \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right)$ $\therefore$

$l(\theta) = n \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2$

$l'(\mu) = \frac{1}{2\sigma^2} \sum_{i=1}^{n} 2(y_i - \mu) = 0$ $\therefore$ $2\sum_{i=1}^{n} Y_i = 2n\hat{\mu}$ $\therefore$ $\hat{\mu} = \bar{y}$

$\text{var}(\bar{Y}) \underset{=}{\overset{Y_i \text{ indep}}{}} \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n} \text{var}(Y_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$

$\therefore$ s.e.$(\hat{\theta}) = \sigma/\sqrt{n}$ $\therefore$ interval estimator is $\hat{\theta} \pm \sigma/\sqrt{n}$

CI is $P(\mu \in (\hat{\theta} - \sigma/\sqrt{n}, \hat{\theta} + \sigma/\sqrt{n}))$ is const $\therefore$

$P\left(\bar{Y} - \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{Y} + \frac{\sigma}{\sqrt{n}}\right) = P\left(-\frac{\sigma}{\sqrt{n}} \le \bar{Y} - \mu \le \frac{\sigma}{\sqrt{n}}\right) = P\left(-1 \le \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \le 1\right)$ $\therefore$

$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$ from CLT $\bar{Y} \sim N(\mu, \sigma^2/n)$ $\therefore Z \sim N(0,1)$

$\therefore P(-1 \le Z \le 1)$ indep of μ $\therefore$ have CI

$P(Z \le 1) - P(Z \le -1) = 0.841 - 0.159 = 0.683$ $\therefore$ 68% CI

$\bar{Y} \pm Z_\alpha \frac{\sigma}{\sqrt{n}}$ $\therefore P(-Z_\alpha \le Z \le Z_\alpha)$ $\therefore P\left(\mu \in \bar{Y} \pm Z_\alpha \frac{\sigma}{\sqrt{n}}\right)$

$\therefore P(-Z_\alpha \le Z \le Z_\alpha) = 0.95$

Z_α st $P(Z \le -Z_\alpha) = 0.025$ $\therefore Z_\alpha = 1.96$

$[\bar{Y} - Z_\alpha \sigma/\sqrt{n}, \bar{Y} + Z_\alpha \sigma/\sqrt{n}]$ $\therefore P(\mu \in [\bar{Y} - Z_\alpha\sigma/\sqrt{n}, \bar{Y} + Z_\alpha \sigma/\sqrt{n}]) = \alpha$

$Y_i \sim N(\mu, \sigma^2)$ σ^2 known

if $y_1, \dots, y_{100}$, $\bar{y} = 1$, $\sigma = 10$ $\therefore$ 95% IC: $\alpha = 1.96$ $\therefore \left[1 - \frac{1.96 \times 100}{10}, 1 + \frac{1.96 \times 100}{10}\right]$

$= [-0.96, 2.96]$ $\therefore$ either $\mu \in [-0.96, 2.96]$ or $\mu \notin [-0.96, 2.96]$

$\therefore$ coverage of 0.95

if $Y_1, \dots, Y_n$ indep; $f_{Y_i}(\theta)$ $\therefore$ for large n: Z sampling distri of

Z mle is $\theta(Y) \sim N(\theta, I(\theta)^{-1})$ $I(\theta)$ is "expected information"

$I(\theta) = E(J(\theta)) = E\left(-\frac{\partial^2 l(\theta)}{\partial\theta \partial\theta^T}\right)$ $l(\theta)$ is log likelihood

if θ is scalar: $I(\theta) = E\left(-\frac{d^2 l(\theta)}{d\theta^2}\right)$

if $I(\theta)$ is a mat $I(\theta)_{ij} = E\left(\frac{\partial^2 l(\theta)}{\partial\theta_i \partial\theta_j}\right)$

\Ex/ $Y_1, \dots, Y_n$ covariates $x_1, \dots, x_n$ $Y_i | x_i \sim N(\alpha + \beta x_i, \sigma^2)$

$\underline{\theta} = (\alpha, \beta, \sigma)$ $\hat{\theta}(Y) \sim N(\theta, I(\theta)^{-1})$

$L(\underline{\theta}) = \prod_{i=1}^{n} f_{Y_i}(y_i; \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2\right) =$

$\frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2\right)$ $\therefore$

Loglikelihood : $l(\theta) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2$

$\frac{\partial l}{\partial \alpha} = \frac{\partial}{\partial \alpha}\left[-\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i^2 + \alpha^2 + \beta^2 x_i^2 - 2\alpha y_i - 2\beta x_i y_i + 2\alpha\beta x_i)\right]$

$= \frac{\partial}{\partial \alpha}\left(\frac{1}{2\sigma^2}(2\alpha n\bar{y} - n\alpha^2 - 2\alpha\beta\bar{x}n)\right) = \frac{n\bar{y}}{\sigma^2} - \frac{n\alpha}{\sigma^2} - \frac{\beta n\bar{x}}{\sigma^2}$

$\frac{\partial l}{\partial \alpha} = \frac{n}{\sigma^2}(\bar{y} - \alpha - \beta\bar{x})$

$\frac{\partial l}{\partial \beta} = \frac{1}{2\sigma^2}\left(2\sum_{i=1}^{n}x_i y_i - 2\alpha\bar{x}n - 2\beta\sum_{i=1}^{n}x_i^2\right) = \frac{n}{\sigma^2}(\overline{xy} - \alpha\bar{x} - \beta\overline{x^2})$

$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2$

$\frac{\partial l}{\partial \alpha} = 0$ iff : $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

$\frac{\partial l}{\partial \beta} = 0$ iff : $\overline{xy} - \hat{\alpha}\bar{x} = \hat{\beta}\overline{x^2}$ $\therefore$ $\overline{xy} - (\bar{y} - \hat{\beta}\bar{x})\bar{x} = \hat{\beta}\overline{x^2}$

$\therefore$ $\overline{xy} - \bar{y}\bar{x} = \hat{\beta}(\overline{x^2} - \bar{x}^2)$ $\therefore$ $\hat{\beta} = \frac{\overline{xy} - \bar{y}\bar{x}}{\overline{x^2} - \bar{x}^2}$

$\frac{\partial l}{\partial \sigma} = 0$ iff : $\frac{n}{\hat{\sigma}} = \frac{1}{\hat{\sigma}^3}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2$ $\therefore$

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2$

$\therefore$ $\frac{\partial^2 l}{\partial \alpha^2} = -\frac{n}{\sigma^2}$, $\frac{\partial^2 l}{\partial \beta^2} = -\frac{n\overline{x^2}}{\sigma^2}$, $\frac{\partial^2 l}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2$

$\frac{\partial^2 l}{\partial \alpha \partial \beta} = -\frac{n\bar{x}}{\sigma^2}$, $\frac{\partial^2 l}{\partial \alpha \partial \sigma} = -\frac{2n}{\sigma^3}(\bar{y} - \alpha - \beta\bar{x})$, $\frac{\partial^2 l}{\partial \beta \partial \sigma} = -\frac{2n}{\sigma^3}(\overline{xy} - \alpha\bar{x} - \beta\overline{x^2})$ $\therefore$

each entry of $I(\theta)$: $E\left(-\frac{\partial^2 l}{\partial \alpha^2}\right)$ $\therefore$

$E\left(-\frac{\partial^2 l}{\partial \sigma^2}\right) = E\left(-\frac{n}{\sigma^2} + \frac{3}{\sigma^4}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2\right) =$

$-\frac{n}{\sigma^2} + \frac{3}{\sigma^4}\sum_{i=1}^{n}E((y_i - \alpha - \beta x_i)^2) = -\frac{n}{\sigma^2} + \frac{3}{\sigma^4}n\sigma^2 = -\frac{n}{\sigma^2} + \frac{3n}{\sigma^2} = \frac{2n}{\sigma^2}$ $\therefore$

$I(\alpha, \beta, \sigma) = \frac{n}{\sigma^2}\begin{pmatrix} 1 & \bar{x} & 0 \\ \bar{x} & \overline{x^2} & 0 \\ 0 & 0 & 2 \end{pmatrix}$ $\therefore$ $\begin{pmatrix} \hat{\alpha}(Y) \\ \hat{\beta}(Y) \\ \hat{\sigma}(Y) \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha \\ \beta \\ \sigma \end{pmatrix}, \frac{\sigma^2}{n}\begin{pmatrix} 1 & \bar{x} & 0 \\ \bar{x} & \overline{x^2} & 0 \\ 0 & 0 & 2 \end{pmatrix}^{-1}\right)$

$Y \sim S(\theta)$ $Y_1, \dots, Y_n$ $\therefore$ $Y \sim N(\beta_0 + \beta_1 x, \sigma^2)$ $\quad \Big\{ h(x|y) = \frac{f(x,y)}{g(y)}$

MTH3041 Bayesian CW1

1) 1/ first: propose a model, justify it & write down priors
then: fit the model & check MCMC converged, if it has then we're looking at the posteriors so we can work out whether its good or not
then: do some model checking & if it doesn't work, we need to half go back to step one & propose changes & justify the changes that we're proposing
finally then look at inference, which is what have we learnt from our model? part of that is some pictures, some is you look at the posterior distri & those come out of BRM along side the trace plots. might need to take our posterior samples & actually make probabilistic statements via Monte Carlo

Modelling step / variants-flargest data set has: country, the date for that country, the variant for that country, the cases for that country, the number of biweeks since first started measuring

highest cases is 77000 cases in UK on 2021-09-06, Delta on 19
strong signal on ~~date~~ biweek that as ~~biweek~~ increases, number of cases increases
modelling counts so use negbinomial
i.e. y = cases ∴ $y_i = \text{cases}_i$ $i = 1, \ldots, 600$ ∴
$y_i \mid \underline{\theta} \sim \text{NBinom}$ ∴ $\beta_0 + \beta_1 x_1 + \ldots \beta_p x_p$ ∴ $y_i \mid \underline{\beta}$
I think using date as one of my x's is too risky with how glitchy it is especially with biweek supposed to match up with it completely ∴
x_1 = biweek, x_2 = country, x_3 = variant (biweek grouping?)
(country grouping?) (variant grouping?)

Could do Country being a grouping for UK & the rest so seperating UK & other EU countries (since UK is land locked like the rest but is an island

I think we should group on variant since theyre different viruses with different transmission, contageours rates, and different inhabitation (asymptomatic) averages lengths

First lets group on variants ∴ :

$j = \{1, 2, 3, 4, 5, 6\}$

1 = Alpha , 2 = Beta , 3 = ~~Delta~~ Gamma , 4 = ~~Gamma~~ Delta , 5 = non-who, 6 = other

100 instances of each variant ∴

$n_1 = 100$, $n_2 = 100$, $n_3 = 100$, $n_4 = 100$, $n_5 = 100$, $n_6 = 100$

$y_{ij} \mid \beta_j, x_{ij} \sim NBin(\theta(\beta, x))$

$\log \theta(\beta, x) = \beta_{0j} + \beta_{1j} x_{country,j} + \beta_{2j} x_{biweek,j}$

$\beta \sim N(B, \Sigma)$

$\pi(B, \Sigma)$

$\cancel{\log(\theta(\beta, x)) = b_0 + b_1 x_{country} + b_2 x_{biweek}}$

$\log(\theta(\beta, x_{ij})) = b_0 + b_1 x_{country\,ij} + b_2 x_{biweek\,ij} +$

$\beta_{0j} + \beta_{1j} x_{country,j} + \beta_{2j} x_{biweek,j}$

$\beta_j \sim N(0, \Sigma)$, $\cancel{\pi(b, \Sigma)}$ $\pi(b, \Sigma)$

weakly informative priors on everything

cases is count data so use NBin to allow mean ≠ variance

$E[y] = \theta = e^{\beta_0 + b_1 x_1 + \dots + \beta_0 + \dots} = e^{b_0} e^{b_1 x_1} e^{b_2 x_2} \dots$

∴ know $E[y]$ is somewhere between 80000 & 0 ∴

Suppose need $e^{b_0} \in [0, 80000]$ ∴ $b_0 \in (\log 1, \log 80000)$ ∴

$b_0 \in (0, 11.29)$ ∴ $N(0, 4)$ dont use but use $N(3, 3)$

distri for now SA

$N(0, 4)$ because most of data is 0 ~~so use~~ or low so use 4 instead of 6 biweek ranges from 0-20

$b_1 \sim N(0, 0.05)$

$$\Sigma = \text{vec}(\underline{\sigma}) \times \Omega_k \times \text{vec}(\underline{\sigma})$$

$\underline{\sigma} = \underset{\text{intercept}}{\sigma_1}, \sigma_2, \ldots, \underset{\text{biweek}}{\sigma_?}$ set_prior(class="sd")

{note: Snowday is true or false graphed against cases ∴
Can also set priors Country?}

how much can differ from mean by group? ∴ I think variant big effect
& country big effect

∴ $N(0, 0.05)$ non intercept sds

Assessing Convergence

Make island variable to seperate UK from EU countries

the biweek will effect cases but with the variant it will
effect the cases even more so group biweek on variant
same with country so group country on variant

~~really~~ first model has really big ESS in tail with smallest being
1438 & in Bulk with smallest being 930

RHats all 1 is very good.

trace plots all good with nothing getting stuck and all the
plots converging nicely

more chance of being good is brm(iter = 10000, warmup = 8000)

Second model all RHats 1 smallest Bulk ESS is 1188
smallest Tail ESS is 1740 very good trace plots with nothing
getting stuck & them all converging

error bars on preds are way too big so need to watch
how Danny attemps to make them smaller

third Model trace plots all good

halving sd from 0.05 to 0.025 decreased error bars from e+08 to e+06 ∴ 80000 is only e+04 ∴

0.025/2 = 0.0125 ∴ do sd of 0.01

doing 0.01 still had e+06 so ill try sd of 0.0001

Sixth model smallest ESS is BulkESS 935 & tailESS 1357

None of the trace plots get stuck & they all converge

the error bars go below 200000 but above 150000 so can see a bit of the data & all data is contained in error bars

$y \mid \beta, x \sim D(\underbrace{g^{-1}(\eta)}_{\mu \text{ mean}}, \underbrace{\phi}_{\text{params}})$

$y \mid \beta, x \sim \text{NegBin}(\lambda(\beta, x), \phi)$

$p(y=k \mid \lambda, \phi) \propto \left(\frac{\mu}{\phi+\mu}\right)^k \left(\frac{\phi}{\phi+\mu}\right)^\phi$ → Poisson as $\phi \to \infty$

$E[y] = \lambda \geq 0$ ∴

$\log \lambda(x, \beta) = \sum_{i=0}^{p} \beta_i x_i$

the thing that changes is that $\text{var}[y] \neq \lambda$ but $\text{var}[y] = \lambda(1 + \frac{\lambda}{\phi}) > \text{Poisson}(\lambda)$ for $\phi \geq 0$

$\pi(\beta, \Sigma) = \pi(\beta, \sigma)$

$\pi(\beta, \sigma, \phi)$

is this model skillful at all so for lower values its fairly skillful with everything hitting the error bars is make shape param so small that allows variance to go so large

6th model very good so giving more iterations & more warmup does not effect error bars but lowest BULK ESS is 2122 & tail is 3365 so it increases the ESS's

y: his function is thinking about 2sd internals so 500 with 400 is
500±800 or 5000 with 500 is 5000±1000 is 20% error, not too
bad

max value of cases is 77000 so do N(0,5)

bweak ranges from 1 to
0.2 0.35 0.4 of 0.2

of 100 people 40 have a degree & of those 20% have quali
∴ is 8 ∴ of 100 ⇒ 8 have quali

Let B be quali ∴ A is uni degree ∴

$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$ $P(A) = 0.4$

$P(B) = 0.08 + 0.35 = 0.43$ $P(B|A) = 0.2$

~~lowest model~~ new-third model has lowest BulkESS as 785
& tail ESS as 711 all Rhats are 1 non of the traceplots
get stuck and they all converge
shap plot has mean about 0.57

fourth model has lowest BulkESS as 730 & tailESS as 666
all Rhats = 1 none of the traceplots get stuck & they all converge
still has e+06

still mean of shape distri ∴ sqrt(varys( )) of it explains
why the error bars are so large
∴ for smaller error bars ∴ lower shape priors sd

new sixth: Bulk 1020 Tail 1006 only one Rhat = 1.01
trace plots all good shap prior mean is 0.35
the model seems to be fairly skillful apart from for bweek

new sixth all good one 1.01 B1694 T1383 still not very skillful
shape mean is 0.35 300+200(2) = 700

new eight all Rhat 1 B939 T1537 still not very skillful
shape mean is still 0.34

N(100, 50)
~~a~~ sd of 0.00001 gives Rhat 1.03 B ~~188~~ T 204 so is very bad
and it didn't even effect the bars with them being
18000 still Shape Mean is still 0.34

Draft 6 Model 4 Rhat all 1 B 949 T 316 Shape Mean 0.34
traceplots all converge nothing getting stuck
Neffs still at 180000 but fairly skillful
this is my final model: Draft 6 model 4th
intercept near 6ish its Mean is 5.61 with Mode 6ish
sd trace Means have a Move from the prior as an increase
that is fairly believable
italy & Spain have a negative effect, that much is clear
so countries compared ~~to~~ France is doing alot for the Model
very unclear if biweek has a positive or negative effect
biweek has Mean Zero & Mode Zero
biweek hasnt Moved from the prior so it isnt doing anything
if α more cases if β less cases (negative correlation)
if δ More cases if γ less cases (negative correlation)
if non-who ~~mores~~ cases if Other less cases (negative correlation)

effect of biweek is highest for δ
effect of biweek is negative for γ
effect of biweek is negative for Other

effect of UK highest for δ
effect of UK negative for γ

14/1/22: Draft 1 model 2nd ~~has~~ group Country is the Same as group variant but only needs sd 0.01 for 180000 but
Draft 1 Model 3rd is 0.1 and does 1e+06 so is bad
Draft 1 Model 5th only put a sd 0.01 prior on group Country and it does 1e+06

14/1/22 Draft 1 Model 6th gives 1+e06 because ~~of~~ prior for group variant for coeffs. for bi week I put sd 0.01

Draft 1 Model 7th gives 1+e06 but with : colons in group its Rhat = 1 ∴ in prior sd = 0.01

Draft 1 Model 7th gives 1+e06 because for with : colons in group sd = 0.001 isn't enough when 0.0001 gives error bars too small but 0.001 would normally be enough but isn't because of the colons : s1 = 0.001 isn't enough

Draft 1 Model 9th is 180000 with good ESS but B1125 & T677. Rhat one is 1.~~00~~ trace plots all converging 0.638% chance biweek < 0

UK has largest increase on cases
Italy has largest decrease on cases

effect of variant all near the same apart from sign
effect of biweek all always very small
effect of biweek highest for Spain
effect of biweek is most negative for UK

Draft 1 model 10th has error bars smaller then data points but it only missing out a few (8 points at most out of 5100 apart from variant other like 50 points missed out
all Rhat = 1 ESS B561 T217 traceplots all converge
Shape mean is 0.71 & Mode is 0.71 So final model is resistant to change ∴ make Model 10th sensitivity model
in sensitivity model $p(\text{biweek} > 0) = 0.954$
still same results with biggest increase in cases being UK
effect of biweek always positive & higest for UK

14/1/22/ Draft 1 Model 11th
the form of the model is negative binomial
Rhat = 1 really good ESS
need to plot the random effects
look at summary of model with ranef( )

Draft 1 model sensi ESSB 888 T731 Rhat = 1 all traces converge

Draft 1 Model 12th Rhat = 1 B1111 T799 error bar ~~range~~ range at 150000
because this model doesn't include biweek at all variant+(1|variant:country)

Draft 1 Model 13th error bar range at 60000 (1|country)

Draft 1 Model 14th error bar range at 180000 (1|country:variant)
Draft 1 Model 15 range (1e+05) is 180000 variant+(1|country:variant)
Draft 1 Model 16 range (1e+05) is 180000 variant+(1|country:variant)
Draft 1 Model 17 range 150000 variant+(1|country:variant)
Draft 1 Model 18 range 150000 country+(1|country:variant)
Draft 1 Model 19 range 60000 country+(1|country)
Draft 1 model 20 range 150000 ~~country+(1|coun~~
country+variant+(country:variant)
Draft 1 Model 21 range 180000
biweek+variant+(biweek+variant|country)
Draft 1 Model 22 range (1e+06) 2000000
biweek+variant+(biweek|country:variant)

15/1/22/ Draft 1 Model 5 (1e+06) 2,000,000
biweek+country+variant+(biweek|country:variant)
Draft 1 Model 6 (1e+06) 2000,000
country+variant+(biweek|country:variant)
Draft 1 Model 7 150000 variant+(variant|country)
does 1e+06

15/1/22/ ~~Drag+2:~~ ~~Model1~~

~~biweek+variant+Country+(biweek|Country:variant)~~

Drag+1 Model11 variant + (1|country) 60000

~~Drag+2: Model1~~

note: needs to always be ≥180000 ~~∴ Could do:~~

Drag+1 Model8 150000

(1/country:variant)

Drag+1 Model9 60000 (1/Country)

~~(1/Country)~~ Drag+1 Model10 (1e+05) 150000 (variant|Country) 150000

danny only had one variable as the grouping so i will aswell

So in looking to do (...|Country) only ∴

biweek+variant+(biweek+variant|Country) 180000

variant+(variant|country) 150000

(variant|country) (1e+05) 150000

15/1/22/ Drag+2 Model1 180000

biweek+variant+(biweek+variant|Country)

Rhat=1 B824 T583 all trace plots converge Shape=0.35

MCMC plots all put biweek at zero P(biweek<0)=0.641

ranges of biweek always very small

Drag+2 Model2 150000 variant+(variant|country)

Rhat=1 B817 T~~8~~76 all trace plots converge Shape=0.37

MCMC not that interesting P(Delta<0)=0.473

Drag+3 Model1 150000 (variant|Country)

Rhat=1 B=468 T=368 all trace plots converge Shape=0.35

Code Q1 Markdown Drag+1 on 15/1/22/ note: doesn't seem to be any errors. I did plot(N=7) and it only showed the final plot & made the rest flat & it took 6 pages (N=7).

So maybe do (N=12) because each page has 6 lines or dont do N at all (but trace plots may too big) Currently 15 pages total

Summary() & Rane$) all take up too much space (2pages) & (6pages) so maye make it a comment or dont evaluate it

as a last resort I could just not use all those beginning plots and only use my plot with date (facet(variant) looks ugly) comments should only be for text, I should just allow all code in but do eval=FASLE or Echo=FALSE

first fix trace plots & if nothing else just dont use them the next few times maybe, like theyre not needed for sensitivity model

Rane$ & Summary are kinda need but would get rid of rane$ first because showing you can look at Rhat=1 and all ESS & say all ESS are high including the lowest Bulk & tail

first few plots take up alot of space so look into making them smaller

for trace plots theyre given in 5 in normal R have to hit enter 7 times so $7 \times 5 + 1 = 36$

if I do N=4 it will only allow 4 on first page then would do 2 big ones on the next page

if I dont do N=36 it will almost definetly show all of them so if I do N=35 atleast it'll show 2 shape plot

its 15P currently brm starts on page 5 so first model took 10pages so start takes 5pages out of 20 page limit then currently model takes 10pages so if Q1 is only 16 pages then

echo=FALSE wont show code on LATEX

eval=FALSE wont show a graph for graph code on LATEX

\15/1/22/ when setting $N=35$ and hiding the first 2 plots so the bmts only starts on page 4 the whole pdf is only 10 pages which traces only takeing up 1.5 pages because the first 30 are hidden and the next 5 are flat so the orth takes 6 pages
Can look into how to hide some useless columns of summary & how to hide some useless columns of ranef even though it might not make a difference

\16/1/22/ Count distribution $y|\beta, x \sim D(\overbrace{g^{-1}(\eta)}^{\text{mean } \mu}, \overbrace{\phi}^{\text{params}})$
So use poisson or negative binomial but if poisson then $\text{Mean}(y) = \text{var}(y)$ which makes the likelihoods flat everywhere so theres no poisson where the mean and variance are equal that actually explains the data so a poisson regression wont work
So instead use negative binomial since it allows
$\text{Mean}(y) \neq \text{var}(y)$ $\therefore$ $y|\beta, x \sim \text{NegBin}(\lambda(\beta, x), \phi)$ $\therefore$

$$P(y=k|\lambda, \phi) \propto \left(\frac{\mu}{\phi+\mu}\right)^k \left(\frac{\phi}{\phi+\mu}\right)^{\phi} \longrightarrow \text{poisson as } \phi \to \infty \quad \therefore$$

$E[y] = \lambda \geq 0$ $\therefore$ can still use $\log\lambda(x, \beta) = \sum_{i=0}^{p} \beta_i x_i$ ~~but~~ and
$\text{var}[y] \neq \lambda$ but $\text{var}[y] = \lambda(1 + \frac{\lambda}{\phi}) \geq \lambda$ for $\phi \geq 0$
$\pi(\beta, \Sigma) = \pi(B, \sigma)$
$\therefore$ distribution for $\pi(B, \sigma, \phi)$
were modelling counts $\therefore$ is y is cases $\therefore$ y_i - Cases$_i$
$i = 1, \ldots, 600$ then $y_i | \beta \sim \text{NegBin}(\lambda(x), \phi)$ (λ is inverse link) $\therefore$
$\log\lambda(x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$ believe cases will change drastically based on country because of different population sizes and goverments so wish to group by country $\therefore$
x_1 is variant x_2 is biweek $\because$ date is replicate of biweek

So for grouping by country: $j=\{1,2,3,4,5\}$
1 is France 2 is Germany 3 is Italy 4 is Spain 5 is United Kingdom ∴

$\therefore n_1=120$ $n_2=120$ $n_3=120$ $n_4=120$ $n_5=120$

$y_{ij} \mid \beta_j, x_{ij} \sim \text{NegBin}(\lambda(\beta, x), \phi)$ ∴

$\log \lambda(\beta, x) = \beta_{0j} + \beta_{1j} x_{variant,j} + \beta_{2j} x_{biweek,j}$ ∴ Model for β's:

$\beta \sim N(B, \Sigma)$ ∴ prior $\pi(B, \Sigma)$

out.width = "90%"

$\log(\lambda(\beta, x_{ij})) = b_0 + b_1 x_{variant\,ij} + b_2 x_{biweek\,ij} + \beta_{0j} + \beta_{1j} x_{variant,j} + \beta_{2j} x_{bi,j}$

So $\beta_j \sim N(0, \Sigma)$ and need prior for $\pi(b, \Sigma)$

$E[y] = \lambda = e^{b_0+b_1x_1+\ldots+\beta_0+\ldots} = e^{b_0} e^{b_1 x_1} e^{b_2 x_2} \ldots$

$e^{b_0} \in [1, 77000]$ count is a lower bound

I want to only use 16 pages for Question 1 to leave 4 pages for questions 2 and 3

$b_i \sim N(0\ \ )$ b_i priors will all have mean zero

So $b_i \sim N(0, 0.01)$ is b prior

$\Sigma = \text{vec}(\sigma) \times \Omega_k \times \text{vec}(\sigma)$

$\sigma = \sigma_0, \sigma_{variant}, \sigma_{biweek}$

first traces are 36 plots running a Markdown with broken latex runs all the brm for a really long time then stops to say broken, so write all latex first then integrate code later

still full error bars so do brm then summary then trace then errorbars

ranefs[c(4,56), 'Estimate']

Based on the variance formula, to create good prediction error bars for a 95% error bar use shape prior N(400, 200)

My new tex file has priors end at the beginning of page 4

need to make all plots smaller to 90%.

Currently full 9 pages so Model 1 took 6 pages so 3 pages for remaining two models

\ 16/1/22 / Model2 plots 5 a page hit enter 5 times remaining 3 so 5x5+3=28

note: the size of the plot seems to be the same no matter what

36-3=33=N N=28-3=25

no points missing in the error bars

\ R hats are 1 were converged, have a much smaller model. intercept & ~~categor~~ variant variables have not changed in this sensitivity step

the predictions have not changed

So my conclusions seem pretty insensitive to ~~~~ the main choices I have made

So sensitivity model is third model ∴

final model is 2nd model

need to change 90% back to 100% for trace

but 90% does help since its 2 pages now

MCMC can go to 85

pushing the limit on how small preds can be

half of page 9, no reasonable page space is wasted

So Model 1 took 5 pages and 7 remaining for Model 2 & 3 ∴ dont need ranef for 3,

use ~~MCMC~~ dont need MCMC ~~to~~ for 3

Model 3 has 5 each plot, hit enter 4 times with 3 remaining

so 4x5+3=23 so N=23-3=20

Q1 half of page 15 so 5.5 pages for Q2 & 3

\ 2 / \ 2a /

>>predict() % the model & 2 new data you want

>>case-preds gives an Estimate is a expectation of the probability

Est.~~Error~~ is sd of the samples so would have to divide

this by the ESS to get the MonteCarlo-error
So Estimate is the Expectation, so is the Monte-Carlo approx to whatever you asked
Est.Error is sd but not the standard error, so have to divide by the right ESS
also gives 95% bounds on a particular prediction
the question of whats the probability that the number is greater than some other number ~~all~~ these questions are MonteCarlo questions. So right down the integral hes asking for. See the distribution is the posterior predictive, & you get samples from that with this predict function

biweek+variant+(variant|country) 230000

>>predict(summary=F)
>>case_preds_2 is 4000 rows and 600 columns
so 600 points its predicted so each column is giving us all 4000 samples so for each column getting all the samples for the particular prediction, so can average these samples in the right way to do any MonteCarlo that you need to do

biweek+variant+(biweek|country) 80000

get similar with >>posterior_samples() is giving you 2 params if you want to do ~~it by hand~~ the Monte Carlo by hand
My prediction interval for peak Delta in UK from Model_2 so need to filter to only leave ~~at~~ Q2.5 & Q97.5?
In the data, Delta seems to be highest at biweek 19 or 20 so may need to infer when its highest, in the future outside the data

biweek+(biweek+variant|country) 1,000,000

\Similar example sleepstudy/ data contains reaction times for ~~a number~~ of subjects ~~for~~ multiple days
So they did it for subject 330 for multiple days

Notes

$e.g\ 1.04^n = 5$

$n\log(1.04) = \log(5)$

* when finding t using logs → have to log both sides

* finding the ~~CAPM~~ CML line only look at market values to calc μ_m & σ_m

* market portfolio = 1 (Beta) rf asset = 0

2a/ I wish to plot intervals for multiple UK Delta for multiple days

theyre data is sleep study with Days + (Days|Subject)
im doing variant + (variant|Country)
they have sleep_fit3, I have Model_2
sleepstudy is variants_5largest
preds_UK_Delta is 20 rows and 4 colums

Model_2
elaborate_cake3 ← lmer(angle ~ recipe + temp + (1|recipe:replicate)
Cake ⟺ variants_5largest
replicate's ⟺ variant = Delta country = Uk

I currently am able to graph from biweek 1 to 20, the predicted cases

3/ 3a/ Variant i ∴ 1 = Alpha 2 = Beta 3 = Delta
4 = Gamma 5 = non-who 6 = Other
Country C : C_1 1 = France C_2 2 = Germany C_3 3 = Italy
C_4 4 = Spain C_5 5 = Uk
fortnight up to t ∴ 14 days upto t ∴ biweek = t, ∴ t = 1, ..., 20

Q2 currently takes up 1.5 pages leaving enough for Q3

1/ ~~b2~~ Model 1 (180000) biweek + variant + (biweek + variant | country)
Model 2 (1M ~~1e+06~~) ~~biweek + (biweek|Cou~~ biweek + (biweek + variant | Country) 1000000
biweek + variant + (biweek + variant | Country) 180000
biweek + variant + (variant | Country) 230000

need to edit just after page 7 to say theres no correlation so keep all the variables then say model 1 then do sensitivity analysis for biweek + variant + (variant | country) 230 000 has traceplots of has 5 pages of 5 and one page of 4 eg 4 + ~~IIII~~ ~~5~~ vs 29 IIII + 4 ∴ 5×5+4=29 ∴ N=29-3=26

Model 1 MCMC plots show no correlation so keep all the variables so Model1 will be my final model

no need to make images smaller now to have so many pages ∴ ⟹ on this first model will do a sensitivity analysis the random effects also

~~For~~ For France Delta needs to be compared to Alpha, Beta, Gamma, non-who, Other ∴ 5 ∴ is only comparing it to Alpha, Beta ∴ so will need to store the final number ~~as~~ a vector

I need to create a mark down of question 3 a, b, c the do a bit of typing for it then ~~do~~ the montecarlo part of a, b then finally put them all together

try to look into a integral for 2a

3a is Beta > nonwho ∴

3b is Delta > Gamma

Estimate probab ∴ use Bulk ∴ use smallest ESS ∴ this is the ~~smallest~~ ESS MC error for each probab 0.0004 < 0.01 are all MC error smaller than 0.01 ⟹ TRUE, TRUE, ...

\week 3/ \2a/ unknown random quantity assigning 1 to true and 0 to false

nums whose true val is unknown with only two possible vals true & false, assigning 1 to true and 0 to false

\2b/ $P(A \vee (B \wedge C))$ $\therefore A \vee (B \wedge C) = \sim(\tilde{A} \wedge \sim(B \wedge C)) =$

$1 - ((1-A)(1-BC)) = 1 - (1 - A - BC + ABC) = A + BC - ABC$

$(A \vee B) \wedge (A \vee C) = (\sim(\tilde{A} \wedge \tilde{B}) \wedge (\sim(\tilde{A} \wedge \tilde{C})) =$ ← superfluous brackets needed

$(1 - (1-A)(1-B))(1 - (1-A)(1-C)) = (1 - 1 + A + B - AB)(1 - 1 + A + C - AC) =$

$(A^2 + AC - A^2C + BA + BC - ABC - A^2B - ABC + A^2BC)$ $\{A \text{ is } 1 \text{ or } 0 \therefore A^2 \text{ is just } A\}$

$\Rightarrow = A + (AC - AC) + BA + BC - ABC - AB + (-ABC + ABC) =$

$A + BC - ABC$ so

$P(A \vee (B \wedge C)) = P((A \vee B) \wedge (A \vee C))$ otherwise we have violated coherence

$P(A \vee (B \wedge C)) = P(A) + P(BC) - P(ABC)$ by linearity of expectation

\3/ $(A \wedge B) \vee (C \wedge D) = \sim((\widetilde{A \wedge B}) \wedge (\widetilde{C \wedge D})) =$

$1 - ((1-AB)(1-CD)) = 1 - (1 - AB - CD + ABCD) = AB + CD - ABCD$

$P((A \wedge B) \vee (C \wedge D)) = E[AB + CD - ABCD]$ by l.o.e.

$= E[AB] + E[CD] - E[ABCD] = 0.5 + 0.4 - 0.2 = 0.7$

\4/ W.L.O.G. (without loss of generality) assume X takes n possible vals $x_1, \ldots, x_n$

define events $X = x_i$ $i = 1, \ldots, n$ ~~form a~~

$X = x_i$ $i = 1, \ldots, n$ form a partition so $X = x_1(X = x_1) \vee \ldots \vee x_n(X = x_n) =$

$\sum_{i=1}^{n} x_i(X = x_i)$ (partition)

$$E[X] \underset{\substack{\uparrow \\ \text{coherence}}}{=} E\left[\sum_{i=1}^{n} x_i(X = x_i)\right] \overset{\text{l.o.e.}}{=} \sum_{i=1}^{n} x_i E[X = x_i] = \sum_{i=1}^{n} x_i P(X = x_i)$$

\5/ $E_1, \ldots, E_n$ a partition, A an r.q. $E[A] = \sum_{i=1}^{n} E[A|E_i]P(E_i) = E[E[A|E]]$

$A \underset{\text{partition}}{=} A(E_1 \vee E_2 \vee \ldots \vee E_n)$

$= A(E_1 + E_2 + \ldots + E_n)$

l.o.e. & coherence $\therefore E[A] = E[AE_1] + E[AE_2] + \ldots + E[AE_n]$

thm of compound probabs: for E event, A r.q. $E[AE] = E[A|E]P(E)$ (A event)

$P(AE) = P(A|E)P(E)$

$E[A] = E[A|E_1]P(E_1) + \dots$

\week 4/ $K = \int_0^{0.5} x^a(1-x)^b e^{-cx}\,dx$ Monte Carlo ~~Beta(a+1,...)~~ Beta(a+1, b+1)

Beta $f(x) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$

$K = \int_{-\infty}^{\infty} g(x) f(x)\,dx$ ∴ Beta$(a+1, b+1)$ $x \in [0,1]$

$f(x) = \dfrac{\Gamma(a+b+2)}{\Gamma(a+1)\Gamma(b-1)} x^a (1-x)^b$ ∴

$K = \int_{-\infty}^{\infty} g(x) f(x)\,dx = \left(\int_0^1 \mathbb{1}(x \le 0.5) e^{-cx} f(x)\,dx\right) \dfrac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+2)}$

Must have coursework in R markdown format on Ele

```
Eg >> BetaSamples <- rbeta(N, a+1, b+1)
>> tg <- (BetaSamples<0.5)*(gamma(a+1)*gamma(b+1) gamma(a+b+2))*
exp(-c*BetaSamples)
```

For Gamma:

$f(z) \sim \text{Gamma}(\underset{\alpha}{a-1}, \underset{\beta}{c})$ $f(x) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$

$K = \int_0^{\infty} g(x) f(x)\,dx = \int_0^{\infty} (1-x)^b \dfrac{\Gamma(a+1)}{c^{a+1}} \mathbb{1}(x < 0.5) f(x)\,dx$ $\mathbb{1}$ is indicator func

\Week 1/

\1a/ By Bayes thm: Z posterior: $\pi(\theta|y) \propto \pi(\theta) P(y|\theta) \propto$
$\theta^{a-1}(1-\theta)^{b-1} P(y|\theta) \propto \theta^{a-1}(1-\theta)^{b-1} \prod_{i=1}^{n} \theta(1-\theta)^{y_i - 1} \propto \theta^{a-1}(1-\theta)^{b-1} \theta^n (1-\theta)^{\sum_{i=1}^n (y_i - 1)} \propto$
$\theta^{a+n-1}(1-\theta)^{b-1}(1-\theta)^{-n+n\bar{y}} \propto \theta^{a+n-1}(1-\theta)^{b+n\bar{y}-n-1}$ ∴

$\pi(\theta|y) \propto \theta^{a+n-1}(1-\theta)^{b+n\bar{y}-n-1}$ ∴ $\theta | y \sim \text{Beta}(a+n, b+n\bar{y}-n)$

∵ $\pi(\theta|y)$ is proportional to a Beta density ∴

$\pi(\theta|y) = \dfrac{\Gamma(a+b+n\bar{y})}{\Gamma(a+n)\Gamma(b+n\bar{y}-n)} \theta^{a+n-1}(1-\theta)^{b+n\bar{y}-n-1}$ $\theta \in (0,1)$

\1b/ $E[\theta|y] = \int_0^1 \theta\, \pi(\theta|y)\,d\theta = \int_0^1 \theta \dfrac{\Gamma(a+b+n\bar{y})}{\Gamma(a+n)\Gamma(b+n\bar{y}-n)} \theta^{a+n-1}(1-\theta)^{b+n\bar{y}-n-1}\,d\theta =$

$= \dfrac{\Gamma(a+b+n\bar{y})}{\Gamma(a+n)\Gamma(b+n\bar{y}-n)} \int_0^1 \theta^{a+n}(1-\theta)^{b+n\bar{y}-n-1}\,d\theta = \dfrac{\alpha}{\alpha+\beta} = \dfrac{a+n}{a+b+n\bar{y}}$

\4a/ \three prisoners/ A, B, C Alan thinks all equally likely to go free

∴ $P(A) = P(B) = P(C) = \frac{1}{3}$ no info about A if you tell me Z name of someone to be executed Jailer says truth - B will die

Now Alan thinks its me or C ∴ $P(A) = \frac{1}{2}$ b be Event jailer says B will die:

$P(b|A) = p$ $x = P(A|b)$ Bayes: $P(A|b) = \dfrac{P(b|A) \cdot P(A)}{P(b)} = \dfrac{P(b|A)P(A)}{?}$

\week 1/ $p(A|b) = \frac{p(b|A) \cdot p(A)}{p(b)}$ one of A, B, C is true ∴

~~$p(b) = p(b|A) + p(b|B) + p(b|C) = \sum_{i=A,B,C} p($~~

$p(b) = p(b|A)p(A) + p(b|B)p(B) + p(b|C)p(C) = \sum_{i=A,B,C} p(b|i)p(i)$ ∴

$p(A|b) = \frac{p(b|A) \cdot p(A)}{\sum_{i=A,B,C} p(b|i)p(i)}$ ∴ ~~$p(b|A)p($~~ ∵ $p(A) = p(B) = p(C) = \frac{1}{3}$ ∴

$p(b|A)p(A) = 1 \cdot \frac{1}{3} = \frac{1}{3}$, $p(b|B)p(B) = 0 \cdot \frac{1}{3} = 0$. $p(b|C)p(C) = 1 \cdot \frac{1}{3} = \frac{1}{3}$ ∴

$p(b) = \frac{1}{3} + 0 + \frac{1}{3} = \frac{2}{3}$ ∴ $p(b|A) = p$ ∴

$x = p(A|b) = x = \frac{p \cdot p(A)}{p(b)} = \frac{p \cdot \frac{1}{3}}{p(b|A)p(A) + p(b|B)p(B) + p(b|C)p(C)} =$

$\frac{p \cdot \frac{1}{3}}{p \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{p \cdot \frac{1}{3}}{p \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{p}{p+1} = x$

\4b/ ~~$p($~~ ∵ $p(A) = \frac{1}{3}$ & that being told b gives no else ∴

$p(A|b) = p(A)$ ∴ $p(A|b) = x$ ∴ $x = \frac{1}{3}$ ∴ $\frac{1}{3} = \frac{p}{p+1}$ ∴ $p = \frac{1}{2} = p(b|A)$

∵ Alan thinks $p(A|b) \neq p(A)$ ∴ $p(A|b) = \frac{1}{2}$ ∴ $x = \frac{1}{2} = \frac{p}{p+1}$ ∴ $p = 1 = p(b|A)$

but this goes against Cromwell's law

\4d/ by Bayes ∵ $p(b|A) \neq 1$ ∴ $p(b|A) = \frac{1}{2}$ ∴ ∵ $p(A|b) = p(A)$ ∴

$p(A|b) = \frac{1}{3} = \frac{p}{p+1}$ ∴ $p = \frac{1}{2} = p(b|A)$ seems reasonable ∵ what Alan says

is actually what he thinks

\Week 3/

\~~2a~~/ unknown random quantities assigning 1 to true & 0 to false

numbers whose true vals is unknown with only two possible vals true

& false, assigning 1 to true & 0 to false

\2b/ $p(A \vee (B \wedge C))$ ∴ $A \vee (B \wedge C) = \sim(\tilde{A} \wedge \sim(B \wedge C)) =$

$1 - ((1-A)(1-BC)) = 1 - (1 - A - BC + ABC) = A + BC - ABC$

$p((A \vee B) \wedge (A \vee C))$ ∴ $(A \vee B) \wedge (A \vee C) = (\sim(\tilde{A} \wedge \tilde{B})) \wedge (\sim(\tilde{A} \wedge \tilde{C})) =$

$(1 - (1-A)(1-B))(1 - (1-A)(1-C)) = A^2 + AC - A^2C + BA + BC - ABC - A^2B - ABC + A^2BC$

$\{A$ is 1 or 0 ∴ A^2 is just $A\}$ ∴ $= A + (AC - AC) + BA + BC - ABC - AB + (-ABC + ABC) =$

$A + BC - ABC$ ∴ $A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C)$ ∴

$p(A \vee (B \wedge C)) = p((A \vee B) \wedge (A \vee C))$

1/ $P(A \vee B \vee C)$ $\therefore A \vee B \vee C = (A \vee B) \vee C = (\sim(\tilde{A} \wedge \tilde{B})) \vee C =$

$\sim(\sim(\sim(\tilde{A} \wedge \tilde{B})) \wedge \tilde{C}) = \sim((\tilde{A} \wedge \tilde{B}) \wedge \tilde{C}) = \sim(\tilde{A} \wedge \tilde{B} \wedge \tilde{C})$

$1-(1-A)(1-B)(1-C) = 1-(1-A-B+AB)(1-C) =$

$1-(1-A-B+AB-C+AC+BC-ABC) =$

$A+B+C-AB-AC-BC+ABC \quad \therefore$

$P(A \vee B \vee C) = P(A+B+C-AB-AC-BC+ABC) =$

$P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)$

3/ $P((A \wedge B) \vee (C \wedge D))$ $\therefore (A \wedge B) \vee (C \wedge D) = \sim((\widetilde{A \wedge B}) \wedge (\widetilde{CD})) =$

$1-((1-AB)(1-CD)) = 1-(1-AB-CD+ABCD) = AB+CD-ABCD \quad \therefore$

$P((A \wedge B) \vee (C \wedge D)) = P(AB+CD-ABCD) = P(AB)+P(CD)-P(ABCD) =$

$E(AB)+E(CD)-E(ABCD) = 0.5+0.4-0.2 = 0.7$ QED

4/ w.l.o.g. assume X takes n possible vals $x_1, \dots, x_n$ define events $X=x_i$ $i=1,\dots,n$ for a partition $\therefore X = x_1(X=x_1) \vee \dots \vee x_n(X=x_n) =$

$\sum_{i=1}^{n} x_i(X=x_i)$ $\therefore$ $E[X] =$ (by coherence) $E\left[\sum_{i=1}^{n} x_i(X=x_i)\right] \overset{\text{l.o.e}}{=}$

$\sum_{i=1}^{n} x_i E[X=x_i] = \sum_{i=1}^{n} x_i P(X=x_i)$

5/ $E_1, \dots, E_n$ a partition, A an r.q. $E[A] = \sum_{i=1}^{n} E[A|E_i]P(E_i) = E\left[E[A|E]\right]$

$A = \{\text{partition}\}\ A(E_1 \vee E_2 \vee \dots \vee E_n) = \{\text{partition}\}\ A(E_1+E_2+\dots+E_n)$

$\therefore A = A(E_1+E_2+\dots+E_n) \quad \therefore$

$A = (A|E_1)+(A|E_2)+\dots+(A|E_n) \quad \therefore$

$P(A) = P(A(E_1+E_2+\dots+E_n)) =$

$P(A|E_1)P(E_1)+P(A|E_2)P(E_2)+\dots+P(A|E_n)P(E_n) = \sum_{i=1}^{n} P(A|E_i)P(E_i) = P(A)$

week 1/

2a/ Bayes thm, the posterior density for θ given $\underline{y}$ is

$\pi(\theta|\underline{y}) \propto \pi(\theta)p(\underline{y}|\theta) \propto \pi(\theta)\prod_{i=1}^{n} f(y_i|\theta) = \pi(\theta)\prod_{i=1}^{n}\theta e^{-\theta y_i} \propto$

$\theta^{a-1}e^{-b\theta}\theta^{n}e^{-\theta n\bar{y}} \propto \theta^{a+n-1}e^{-(b+n\bar{y})\theta}$ comparing this to the density of the Gamma(a,b) prior shows that it is proportional to the density of a Gamma(a_1,b_1) distri, $a_1 = a+n$, $b_1 = b+n\bar{y}$ & $\therefore$ the posterior is a Gamma distri with $\pi(\theta|\underline{y}) = \frac{b_1^{a_1}}{\Gamma(a_1)}\theta^{a_1-1}e^{-b_1\theta}$ for $\theta > 0$ eg $\pi(\theta|\underline{y}) = \frac{(b+n\bar{y})^{a+n}}{\Gamma(a+n)}\theta^{a+n-1}e^{-(b+n\bar{y})\theta}$

$P(b|A) = p$ $\quad x = P(A|b)$ $\quad$ Bayes: $\pi(b) =$ $\quad \overline{P(b)}$

\Sheet Week 1/ \2b/ taking logs, differentiating & setting to zero, Z MAP esti, $\hat{\theta}$, satisfies $\frac{a_1-1}{\hat{\theta}} - b_1 = 0 \iff \hat{\theta} = \frac{a_1-1}{b_1}$

{Gamma density $\text{Gamma}(a,b) = \pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}$ $\theta > 0$ ∴

$\text{Gamma}(a_1, b_1) = \pi(\theta | y) = \frac{b_1^{a_1}}{\Gamma(a_1)} \theta^{a_1-1} e^{-b_1\theta}$ for $\theta > 0$ ∴

$\log \pi(\theta|y) = \log\left[\frac{b_1^{a_1}}{\Gamma(a_1)} \theta^{a_1-1} e^{-b_1\theta}\right] = \ln(b_1^{a_1}) - \ln(\Gamma(a_1)) + \ln(\theta^{a_1-1}) + \ln(e^{-b_1\theta}) =$

$a_1 \ln b_1 - \ln\Gamma(a_1) + (a_1-1)\ln\theta - b_1\theta$ ∴

$\frac{\partial[\log \pi(\theta|y)]}{\partial\theta} = 0 - 0 + \frac{(a_1-1)}{\theta} - b_1 = \frac{(a_1-1)}{\theta} - b_1$ ∴ $\frac{(a_1-1)}{\hat{\theta}} - b_1 = 0$ ∴

~~for~~ $\frac{a_1-1}{\hat{\theta}} = b_1$ ∴ $\frac{a_1-1}{b_1} = \hat{\theta}$ is Z MAP esti of $\hat{\theta}$

\3 partial sol/ $\theta | y \sim \text{Gamma}(a_1, b_1)$ with $a_1 = a+n$,

$b_1 = b + \sum_{i=1}^{n} \log y_i$ & $\pi(\theta|y) = \frac{b_1^{a_1}}{\Gamma(a_1)} \theta^{a_1-1} e^{-b_1\theta}$ for $\theta > 0$

\3/ prior distri Gamma ~~π~~ $\text{Gamma}(a,b)$: $\pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}$ for $\theta > 0$

∴ by Bayes theorem ~~Posterior~~ $\pi(\theta|y) \propto \pi(\theta) p(y|\theta) \propto \pi(\theta) \prod_{i=1}^{n} f(y_i|\theta) =$

$\pi(\theta) \prod_{i=1}^{n} \theta y_i^{-\theta-1} \propto \theta^{a-1} e^{-b\theta} \prod_{i=1}^{n} \theta y_i^{-\theta-1} \propto \theta^{a-1} e^{-b\theta} \theta^n \prod_{i=1}^{n} y_i^{-\theta-1} \propto$

$\theta^{a+n-1} e^{-b\theta} \prod_{i=1}^{n} e^{\ln(y_i^{-\theta-1})} \propto \theta^{a+n-1} e^{-b\theta} \prod_{i=1}^{n} e^{(-\theta-1)\ln y_i} \propto$

$\theta^{a+n-1} e^{-b\theta} e^{\sum_{i=1}^{n}(-\theta-1)\ln y_i} \propto \theta^{a+n-1} e^{-b\theta} e^{(-\theta-1)\sum_{i=1}^{n}\log y_i} \propto \theta^{a+n-1} e^{-b\theta + (-\theta-1)\sum_{i=1}^{n}\log y_i}$ ∴

$\pi(\theta|y) = \frac{b_1^{a_1}}{\Gamma(a_1)} \theta^{a_1-1} e^{-b_1\theta} = \frac{(b + \sum_{i=1}^{n}\log y_i)^{a+n}}{\Gamma(a+n)} \theta^{a+n-1} e^{-(b + \sum_{i=1}^{n}\log y_i)\theta}$

{$\theta^{a-1} e^{-b\theta} \prod_{i=1}^{n} \theta y_i^{-\theta-1} \propto \theta^{a+n-1} e^{-b_1\theta} e^{-(\theta-1)\ln y_i} \propto$

$\theta^{a+n-1} e^{-b_1\theta} e^{-\theta\ln y_i - \ln y_i} \propto \theta^{a+n-1} e^{-b_1\theta} e^{-\theta\ln y_i} e^{-\ln y_i} \propto \theta^{a+n-1} e^{b_1\theta} e^{-\theta\ln y_i}$ }

\4c/ Alan actually thinks $P(A|b) = x = \frac{1}{2}$ which implies $p = 1$ ie even if Alan is to go free, Z jailer would always say Bernard is to be executed rather than Charles. This seems unreasonable. How could Alan be certain that Z jailer would behave like this?

\4d/ as long as coherence isn't violated $P(b|A) = P(c|A) = \frac{1}{2}$ seems reasonable but could be $p > \frac{1}{2}$ as Alan feels it is more likely Z jailer gives Z first name he can (alphabetically)

\5a/ $1 - \frac{3}{4} = \frac{1}{4}$

\5b/ with events J, C, L, R defined naturally (eg Z phone is in Z left hand pocket), Z joint distri can be written in table form as:

	L	R
J	1/8	1/8
C	1/2	1/4

\5c/ Z answer is no but why? ⊛

\8a/ let θ denote even he dies & y event Soothsayer says false

$\therefore \pi(\theta|y) = 0.92$ $\therefore$ he should be concerned!

{ $\pi(\theta|y) \overset{\text{bayes}}{=} \frac{\pi(y|\theta)\,\pi(\theta)}{\pi(y)} = 0.92$ $\cancel{\pi(y) = \pi(y|\text{die}) + \pi(y|\text{live})} =$

π & $\pi(\theta) = 0.99$ $\pi(y|\theta)$ is says false given die ie

$\cancel{(200\ 192)}$ $\cancel{\pi(y|\theta) = \pi(\theta|y)}$ she correctly predicts false $\frac{192}{200}$

correctly predicts die $\frac{90}{100}$ $\therefore$

incorrectly predicts false $\frac{8}{200}$ &

incorrectly predicts die $\frac{10}{100}$ $\therefore$

$\pi(y|\theta)$ is incorrectly says false ~~given die~~ $\therefore \frac{8}{200} = \pi(y|\theta)$

$\therefore \frac{\pi(y|\theta)}{\pi(y)} = \frac{0.92}{0.99}$ & $\therefore$ $\pi(y) = \frac{0.99}{0.92} \cdot \frac{8}{200} = \frac{99}{2300}$

{ $\pi(y) = \frac{192}{200}$ ✗ } from 200 false said 192 false ~~from~~

from 100 end said 90 end $\therefore$ from 100 end said 10 false $\therefore$

from 300 said 202 false $\therefore \pi(y) = \frac{202}{300}$ ✗ }

\6b/ believing Soothsayer not a good idea from subjective view would only use evidence to update a prior on Z ability of Soothsayer is evidence she's guessing $\therefore p(y|\theta=1) = p(y|\theta=0) = 0.5$

$\therefore \pi(\theta|y) = 0.99 = \pi(\theta)$ makes sense, as under Z view she's guessing her data shouldn't change Z prior

salaries data:

Rank, years 0-25 $\rightarrow N(1100, 500)$ Degree, years 1-35 $\rightarrow N(750, 400)$ Salary 15000-38000 $N(6250, 3000)$

$\therefore \ (Gamma(a+n, a+b+n\bar{y}-n+1)$

\\ Week 1 / 1a / By Bayes theorem: Z posterior: $\pi(\theta|\underline{y}) \propto \pi(\theta) f(\underline{y}|\theta)$

$\left\{ \propto \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}\theta(1-\theta)^{y-1} \propto \theta^{a-1}(1-\theta)^{b-1}\theta(1-\theta)^{y-1} \propto \theta^{a}(1-\theta)^{y-b-2} \right\}$

$\propto \pi(\theta) f(\underline{y}|\theta) \propto \theta^{a-1}(1-\theta)^{b-1} p(\underline{y}|\theta) \propto \theta^{a-1}(1-\theta)^{b-1} \prod_{i=1}^{n} \theta(1-\theta)^{y_i-1} \propto$

$\theta^{a-1}(1-\theta)^{b-1}\theta^{n}(1-\theta)^{\sum_{i=1}^{n}(y_i-1)} \propto \theta^{a+n-1}(1-\theta)^{b-1}(1-\theta)^{-n+n\bar{y}} \propto \theta^{a+n-1}(1-\theta)^{b+n\bar{y}-n-1} \therefore$

$\pi(\theta|\underline{y}) \propto \theta^{a+n-1}(1-\theta)^{b+n\bar{y}-n-1} \therefore \theta|\underline{y} \sim Beta(a+n, b+n\bar{y}-n) \therefore \pi(\theta|\underline{y})$ is proportional to a Beta density $\therefore$

$\pi(\theta|\underline{y}) = \frac{\Gamma(a+b+n\bar{y})}{\Gamma(a+n)\Gamma(b+n\bar{y}-n)} \theta^{a+n-1}(1-\theta)^{b+n\bar{y}-n-1} \quad \theta \in (0,1)$

1c / $E[\theta|\underline{y}] = \int_0^1 \theta\pi(\theta|\underline{y})\,d\theta = \int_0^1 \theta \frac{\Gamma(a+b+n\bar{y})}{\Gamma(a+n)\Gamma(b+n\bar{y}-n)} \theta^{a+n-1}(1-\theta)^{b+n\bar{y}-n-1}\,d\theta =$

$\frac{\Gamma(a+b+n\bar{y})}{\Gamma(a+n)\Gamma(b+n\bar{y}-n)} \int_0^1 \theta^{a+n}(1-\theta)^{b+n\bar{y}-n-1}\,d\theta = \frac{\Gamma(a_1+b_1)}{\Gamma(a_1)\Gamma(b_1)} \int_0^1 \theta^{a_1}(1-\theta)^{b_1-1}\,d\theta =$

$\frac{\Gamma(a_1+b_1)}{\Gamma(a_1)\Gamma(b_1)} \frac{\Gamma(a_1+1)\Gamma(b_1)}{\Gamma(a_1+b_1+1)} \left(\int_0^1 \frac{\Gamma(a_1+b_1+1)}{\Gamma(a_1+1)\Gamma(b_1)} \theta^{(a_1+1)-1}(1-\theta)^{b_1-1}\,d\theta \right) =$

$\frac{\Gamma(a_1+b_1)\Gamma(a_1+1)\Gamma(b_1)}{\Gamma(a_1)\Gamma(b_1)\Gamma(a_1+b_1+1)} (1)$ {because pdf integrates to 1}

$= \frac{\Gamma(a_1+b_1)\Gamma(a_1+1)}{\Gamma(a_1)\Gamma(a_1+b_1+1)} = \left\{ \frac{\Gamma(a_1+b_1)(a_1+1)\Gamma(a_1)}{\Gamma(a_1)(a_1+b_1+1)\Gamma(a_1+b_1)} = \frac{a_1+1}{a_1+b_1+1} \right\}$

$= \frac{a_1}{a_1+b_1}$ is Z posterior expectation $E(\theta|\underline{y}) = \frac{a_1}{a_1+b_1}$

1b / Z posterior density $\pi(\theta|\underline{y})$ is proportional to Z Beta density:

$\pi(\theta|\underline{y}) = \frac{\Gamma(a+b+n\bar{y})}{\Gamma(a+n)\Gamma(b+n\bar{y}-n)} \theta^{a+n-1}(1-\theta)^{b+n\bar{y}-n-1} \therefore$

$\theta|\underline{y} \sim Beta(a+n, b+n\bar{y}-n) \therefore$ Z posterior distribution was also a beta distribution $Beta(a_1, b_1)$

1d / $a=1, b=9 \quad n=30 \quad 95\% \therefore 1-0.95 = 0.05 \therefore \frac{0.05}{2} = 0.025 \therefore$

$1-0.025 = 0.975 \quad 1-\frac{1-\alpha}{2} = \frac{2}{2}-\frac{1-\alpha}{2} = \frac{1+\alpha}{2} \quad 1-\frac{1+\alpha}{2} = 1-0.975 = 0.025$

$q(0.025, 0.975) = (q_{0.025}, q_{0.975}) = (0.00281, 0.336)$

2a / {By Bayes theorem: $\pi(\theta|\underline{y}) \propto \pi(\theta) f(\underline{y}|\theta) \propto \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta} f(\underline{y}|\theta) \propto$

$\theta^{a-1}e^{-b\theta} f(\underline{y}|\theta) \propto \theta^{a-1}e^{-b\theta} \prod_{i=1}^{n} \theta e^{-\theta y_i} \propto \theta^{a-1}e^{-b\theta}\theta^{n} e^{\sum_{i=1}^{n}(-\theta y_i)} \propto$

$\theta^{a+n-1}e^{-b\theta}e^{-\theta n\bar{y}} = \theta^{a+n-1}e^{-b\theta-\theta n\bar{y}} \propto \theta^{a+n-1}e^{-\theta(b+n\bar{y})}$ } $\propto \pi(\theta|\underline{y}) \propto$

$\pi(\theta) p(\underline{y}|\theta) \propto \pi(\theta) \prod_{i=1}^{n} \theta e^{-\theta y_i} \propto \theta^{a-1}e^{-b\theta}\theta^{n}e^{-\theta n\bar{y}} \propto \theta^{a+n-1}e^{-(b+n\bar{y})\theta} \propto \{\theta^{a+n-1} e^{\ldots}\}$

$\{\therefore$ ~~$\theta|\underline{y} \sim$~~ $\theta|\underline{y} \sim Gamma(a+n, b+n\bar{y}) \therefore \pi(\theta|\underline{y})$ is proportional to a Gamma density $\therefore \pi(\theta|\underline{y}) = \frac{(b+n\bar{y})^{(a+n)}}{\Gamma(a+n)} \theta^{a+n-1} e^{-(b+n\bar{y})\theta}$ for $\theta > 0 \therefore$

~~2a~~/ posterior distribution is: Gamma distribution $\text{Gamma}(a_1, b_1)$

& posterior density. $\pi(\theta|y) = \frac{(b+n\bar{y})^{(a+n)}}{\Gamma(a+b)} \theta^{a+n-1} e^{-(b+n\bar{y})\theta}$ for $\theta > 0$

\2b/ taking logs, differentiating & setting equal to zero:

Z MAP esti, $\hat{\theta}$, satisfies $\frac{a_1 - 1}{\theta} - b_1 = 0$

{ $\text{Gamma}(a,b) = \pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}$, $\theta > 0$ ∴ $\text{Gamma}(a_1, b_1) = \pi(\theta|y) =$
$\frac{b_1^{a_1}}{\Gamma(a_1)} \theta^{a_1 - 1} e^{-b_1\theta}$ for $\theta > 0$ ∴ $\log \pi(\theta|y) = \log\left[\frac{b_1^{a_1}}{\Gamma(a_1)} \theta^{a_1-1} e^{-b_1\theta}\right] =$
$\ln(b_1^{a_1}) - \ln(\Gamma(a_1)) + \ln(\theta^{a_1-1}) + \ln(e^{-b_1\theta}) = a_1 \ln b_1 - \ln\Gamma(a_1) + (a_1 - 1)\ln\theta - b_1\theta$ ∴
$\frac{\partial[\log \pi(\theta|y)]}{\partial\theta} = 0 - 0 + \frac{(a_1-1)}{\theta} - b_1 = \frac{(a_1-1)}{\theta} - b_1$ ∴ $\frac{(a_1-1)}{\hat{\theta}} - b_1 = 0$ ∴ $\frac{(a_1-1)}{\hat{\theta}} = b_1$ ∴
$\frac{(a_1-1)}{b_1} = \hat{\theta}$ is Z MAP esti of $\hat{\theta}$ } $\Longleftrightarrow \hat{\theta} = \frac{a_1 - 1}{b_1}$

\2e/ $a=5, b=2, m=20, \alpha=0.95$ ∴ $(0.025, 0.975)$

$(q_{0.025}, q_{0.975}) = (0.359, 0.957)$ for $\text{Gamma}(5,2)$

Monte Carlo: $(1.02, 4.6)$ for param $\text{Gamm}(5,2)$

\3/ { Gamma prior distribution ~~$\pi(\theta)$~~ for θ with density $\pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{b\theta}$
for $\theta > 0$ ∴ By Bayes thm $\pi(\theta|y) \propto \pi(\theta) f(y|\theta) \propto \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} f(y|\theta) \propto$
$\theta^{a-1} e^{-b\theta} f(y|\theta) \propto$ ~~$\theta^{a-1} e^{-b\theta}$~~ $\theta^{a-1} e^{-b\theta} \prod_{i=1}^{n} \theta y_i^{-\theta-1} \propto \theta^{a-1} e^{-b\theta} \theta^n \prod_{i=1}^{n} y_i^{-\theta-1}$
$\propto \theta^{a+n-1} e^{-b\theta} \left(\prod_{i=1}^{n} y_i\right)^{-\theta-1} \propto \theta^{a+n-1} e^{-b\theta} \left(\prod_{i=1}^{n} y_i\right)^{-\theta} \left(\prod_{i=1}^{n} y_i\right)^{-1} \propto \theta^{a+n-1} e^{-b\theta} \left(\prod_{i=1}^{n} y_i\right)^{-\theta} \propto$
$\theta^{a+n-1} e^{-b\theta} \exp\left[\ln\left[\left(\prod_{i=1}^{n} y_i\right)^{-\theta}\right]\right] \propto \theta^{a+n-1} e^{-b\theta} \exp\left[-\theta \ln\left(\prod_{i=1}^{n} y_i\right)\right] \propto$
$\theta^{a+n-1} e^{-b\theta} \exp\left[-\theta \sum_{i=1}^{n} \ln y_i\right] \propto \theta^{a+n-1} e^{-b\theta - \theta \sum_{i=1}^{n} \ln y_i} \propto \theta^{(a+n)-1} e^{-(b + \sum_{i=1}^{n} \ln y_i)\theta}$ ∴

Z posterior distribution is: Gamma distribution $\text{Gamma}(a_1, b_1) =$
$\text{Gamma}(a+n, b + \sum_{i=1}^{n} \ln y_i)$ is Z prior distribution, & Z prior density
is: $\pi(\theta|y) = \frac{(b + \sum_{i=1}^{n} \ln y_i)^{(a+n)}}{\Gamma(a+n)} \theta^{a+n-1} e^{-(b + \sum_{i=1}^{n} \ln y_i)\theta}$ ∴ $\theta|y \sim \text{Gamma}(a+n, b + \sum_{i=1}^{n} \ln y_i)$

$\theta|y \sim \text{Gamma}(a_1, b_1)$ with $a_1 = a+n$, $b_1 = b + \sum_{i=1}^{n} \ln y_i$, &
$\pi(\theta|y) = \frac{b_1^{a_1}}{\Gamma(a_1)} \theta^{a_1-1} e^{-b_1\theta}$ for $\theta > 0$ { ∴ posterior distri is $\theta|y$ &
posterior density $\pi(\theta|y) = \frac{b_1^{a_1}}{\Gamma(a_1)} \theta^{a_1-1} e^{-b_1\theta}$ for $\theta > 0$ }

\4a/ By Bayes theorem $P(A|b) = \frac{P(b|A)P(A)}{P(b)}$, Z law of total probab
gives $P(b) = P(b|A)P(A) + P(b|B)P(B) + P(b|C)P(C)$ as A, B, C form a partition.
$P(b|B) = 0$ as Z jailer is truthful, & $P(b|C) = 1$ ∵ Z jailer is truthful &

$P(b|A) = p$ $x = \pi(A|b)$ Bayes ... $P(b)$

distri: $G(a+n, \; a-b+ny-n+1)$

Week 1/ cannot inform Alan he is to be executed. So

$P(b) = \left\{ p \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = p \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = P(b|A)P(A) + P(b|C)P(C) = \right.$
$\left. P(b|A)\frac{1}{3} + P(b|C)\frac{1}{3} \right\} = \frac{1}{3}(p+1)$ & $\{P(A|b)\} = x = \left\{ \frac{P(b|A)P(A)}{P(b)} = \frac{p \cdot \frac{1}{3}}{\frac{1}{3}(p+1)} \right\} = \frac{p/3}{\frac{1}{3}(p+1)} = \frac{p}{p+1}$

\4b/ Alan argues that telling him Z identity of another that is to be executed gives no info $\therefore P(A) = P(A|b) = x = \frac{1}{3}$ & $\left\{ x = \frac{p}{p+1} \therefore p = \frac{x}{1-x} \right\} \therefore p = \frac{x}{1-x} \therefore x = \frac{1}{3} \Rightarrow p = \frac{1}{2}$. So, if Alan is to go free, Z jailer is equally likely to nominate Bernard or Charles in his conversation with Alan. This is perhaps not unreasonable for Alan to think. ie Alan doesn't believe Z jailer is more/less likely to nominate Bernard over Charles.

\4c/ Alan actually thinks $P(A|b) = x = \frac{1}{2}$ which implies $p=1$ ie even if Alan is to go free, Z jailer would always say Bernard is to be executed rather than Charles. This seems unreasonable. How could Alan be certain that Z jailer would behave like this?

\4d/ any reasoned assessment accepted as long as coherence isn't violated. $P(b|A) = P(c|A) = \frac{1}{2}$ seems reasonable but another might be $p > \frac{1}{2}$ as Alan feels $p > \frac{1}{2}$ as Alan feels it is more likely that Z jailer gives Z first name he can (alphabetically).

\5a/

	Jacket	Coat
right hand	0.125	0.25
left hand	0.125	0.5

$\therefore$ {probab of phone in Jacket is $\frac{1}{4}$} $\frac{1}{4}$ $\left\{1 - \frac{3}{4} = \frac{1}{4}\right\}$

\5b/ with events J, C, L, R defined naturally (eg phone is in Z left hand pocket) Z joint distri can be written in table form as:

	L	R
J	1/8	1/8
C	1/2	1/4

\5c/ {no could it could be elsewhere if eg stolen}
no $\therefore$ reason

\6a/ let θ denote event he dies & y event Soothsayer says safe $\therefore \pi(\theta|y) = 0.92 \therefore$ he should be concerned!

$\{\pi(\theta)=0.99$ $\quad$ θ, θ' form a partition $\therefore \pi(y) = \pi(y|\theta)\pi(\theta) + \pi(y|\theta')\pi(\theta')$

$\therefore y, y'$ form a partition $\therefore \pi(\theta) = \pi(\theta|y)\pi(y) + \pi(\theta|y')\pi(y')$ $\therefore$

$\pi(\theta|y) = 0.9$ $\quad$ $\pi(\theta'|y') = \frac{192}{200} = \frac{24}{25} = 0.96$ ✗ $\}$ $\{\pi(\theta|y)=0.92\}$)

\6b/ Shouldn't believe Sooth Sayer unless evidence they were accurate that updates a prior on Soothsayer ability beyound random chance. ∴ abstance of evidence view soothsayer as guessing ∴ $P(y|\theta=1) = P(y|\theta=0) = 0.5$ ∴ implies $\pi(\theta|y)=0.99=\pi(\theta)$ makes sense as under Z view that Soothsaying is guessing, her "data" shouldn't change your prior. ∴ have a prior on Soothsayers skill & then first view Z brochure numbers as data before hearing Z prediction will require hierarchical modelling

\Week2/ \1a/ $\{$for normal prior distri for $\mu=m, \sigma^2=v^2$:

$$\pi(\mu=m, \sigma^2=v^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} = \frac{1}{v\sqrt{2\pi}} e^{-\frac{1}{2v^2}(x-m)^2}$$

By Bayes theorem: $\pi(\theta|\underline{y}) = \pi(\mu=m, \sigma^2=v^2|\underline{y}) \propto \pi(\mu=m, \sigma^2=v^2) f(\underline{y}|\theta)$

$\propto \pi(\mu=m, \sigma^2=v^2) f(\underline{y}|\mu, \sigma^2)\}$

$\{$Normal prior distri for μ, σ^2 with density $\pi(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\}$

$\{$need posterior distri $\mu, \sigma^2|\underline{y}$ ∴ By Bayes thm $\pi(\theta|\underline{y}) \propto \pi(\theta) f(\underline{y}|\theta)$

$\therefore \pi(\mu, \sigma^2|\underline{y}) \propto \pi(\mu, \sigma^2) f(\underline{y}|\mu, \sigma^2)$ $\quad$ & $\mu = m$, $\sigma^2 = v^2$ ∴ for $\theta = (\mu, \sigma^2)$

$\pi(\theta) \propto \pi(\mu, \sigma^2) \propto \frac{1}{v\sqrt{2\pi}} e^{-\frac{1}{2v^2}(\theta-m)^2}$, $f(\underline{y}|\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2}$

$= \{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\sum_{i=1}^{n}\left[-\frac{1}{2\sigma^2}(y_i-\mu)^2\right]} = ((\sigma^2 2\pi)^{-1/2})^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2} = \cancel{(2\pi\sigma^2)^{-n/2} e}$

$(2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}\} = \prod_{i=1}^{n} f(y_i|\theta) = \{\prod_{i=1}^{n} f(y_i|\mu, \sigma^2)\}$ ∴

$\{\pi(\theta|\underline{y}) \propto \pi(\theta) f(\underline{y}|\theta) \propto \frac{1}{v\sqrt{2\pi}} e^{-\frac{1}{2v^2}(\theta-m)^2} f(\underline{y}|\theta) \propto e^{-\frac{1}{2v^2}(\theta-m)^2} f(\underline{y}|\theta) \propto$

$e^{-\frac{1}{2v^2}(\theta-m)^2} f(\underline{y}|\theta) \propto e^{-\frac{1}{2v^2}(\theta-m)^2} f(\underline{y}|\mu, \sigma^2) \propto e^{-\frac{1}{2v^2}(\theta-m)^2} \prod_{i=1}^{n} f(y_i|\mu, \sigma^2) \propto$

$e^{-\frac{1}{2v^2}(\theta-m)^2} \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \propto e^{-\frac{1}{2v^2}(\theta-m)^2} \prod_{i=1}^{n} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \propto$

$e^{-\frac{1}{2v^2}(\theta-m)^2} e^{\sum_{i=1}^{n}\left[-\frac{1}{2\sigma^2}(y_i-\mu)^2\right]} \propto e^{-\frac{1}{2v^2}(\theta-m)^2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2} \propto$

$\exp\left[-\frac{1}{2}\left(\frac{1}{v^2}(\theta-m)^2 + \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i^2 - 2\mu y_i + \mu^2)\right)\right] \propto \cancel{\exp\left[-\frac{1}{2}\left(\frac{1}{v^2}(\theta^2-2m\theta+m^2\right.\right.}$ $\quad$))

$\exp\left[-\frac{1}{2}\left(\frac{1}{v^2}(\theta^2-2m\theta+m^2) + \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} y_i^2 - 2\mu\sum_{i=1}^{n} y_i + \sum_{i=1}^{n}\mu^2\right)\right)\right] \propto$

$\exp\left[-\frac{1}{2}\left(\frac{1}{v^2}(\theta^2-2m\theta) + \frac{1}{\sigma^2}(-2n\bar{y}\mu + n\mu^2)\right]\right\}$ $\quad$ $\pi(\mu|\underline{y}) \sim N(\mu_n, \sigma_n^2)$ with

\Week 2/ $\sigma_n^2 = \frac{1}{\frac{1}{v^2}+\frac{n}{\sigma^2}}$, $\mu_n = \frac{\frac{m}{v^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{v^2}+\frac{n}{\sigma^2}}$; $\}$

$\{\pi(\mu|y) \sim N(\mu_n, \sigma_n^2) = N\left(\frac{\frac{m}{v^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{v^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{v^2}+\frac{n}{\sigma^2}}\right)\}$

$\{$posterior distri for μ : $\pi(\mu|y)$ ∴ By Bayes's thm $\pi(\mu|y) \propto \pi(\mu) f(y|\mu)$

$\propto \frac{1}{v\sqrt{2\pi}} e^{-\frac{1}{2v^2}(\mu-m)^2} f(y|\mu) \propto e^{-\frac{1}{2v^2}(\mu-m)^2} f(y|\mu) \propto e^{-\frac{1}{2v^2}(\mu-m)^2} \prod_{i=1}^{n} f(y_i|\mu)$

$\propto e^{-\frac{1}{2v^2}(\mu-m)^2} \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \propto e^{-\frac{1}{2v^2}(\mu-m)^2} e^{\sum_{i=1}^{n}\left[-\frac{1}{2\sigma^2}(y_i-\mu)^2\right]} \propto$

$\exp\left[-\frac{1}{2v^2}(\mu^2-2m\mu+m^2) + \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(y_i^2-2y_i\mu+\mu^2)\right)\right] \propto$

$\exp\left[-\frac{1}{2}\left(\frac{1}{v^2}(\mu^2-2m\mu) + \frac{1}{\sigma^2}(-2n\bar{y}\mu+n\mu^2)\right)\right] \propto$

$\exp\left[-\frac{1}{2}\left(\frac{1}{v^2}\mu^2 - \frac{2m}{v^2}\mu - \frac{2n\bar{y}}{\sigma^2}\mu + \frac{n}{\sigma^2}\mu^2\right)\right] \propto$

$\exp\left[-\frac{1}{2}\left(\left[\frac{1}{v^2}+\frac{n}{\sigma^2}\right]\mu^2 - 2\left[\frac{m}{v^2}+\frac{n\bar{y}}{\sigma^2}\right]\mu\right)\right] = \exp\left[-\frac{1}{2}(a\mu^2-2b\mu)\right] \propto$

$\exp\left[-\frac{a}{2}(\mu^2-2\frac{b}{a}\mu)\right] \propto \exp\left[-\frac{a}{2}\left(\left(\mu-\frac{b}{a}\right)^2-\frac{b^2}{a^2}\right)\right] \propto \exp\left[-\frac{a}{2}\left(\mu-\frac{b}{a}\right)^2\right] \propto$

~~exp~~ $\exp\left[-\frac{1}{2a^{-1}}\left(\mu-\frac{b}{a}\right)^2\right]$ $\}$ ∴ ~~ot~~ $\mu|y \sim N(\frac{b}{a}, a^{-1})$ ∴

$\pi(\mu|y)$ is proportional to a Normal density ∴

$\pi(\mu|y) = \frac{1}{a^{-1/2}\sqrt{2\pi}} e^{-\frac{1}{2a^{-1}}(\mu-\frac{b}{a})^2} = \frac{1}{\sigma_n\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_n^2}(\mu-\mu_n)^2\right)$

∴ $\pi(\mu|y) \sim N(\mu_n, \sigma_n^2) = N\left(\frac{\frac{m}{v^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{v^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{v^2}+\frac{n}{\sigma^2}}\right)$;

$\mu_n = \frac{\frac{m}{v^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{v^2}+\frac{n}{\sigma^2}}$, $\sigma_n^2 = \frac{1}{\frac{1}{v^2}+\frac{n}{\sigma^2}}$ ∴ $a = \frac{1}{v^2}+\frac{n}{\sigma^2}$, $b = \frac{m}{v^2}+\frac{n\bar{y}}{\sigma^2}$

\16/ ~~M1.75~~ $\{N(\mu_n, \sigma_n^2)$ $\sigma^2=2$, $m=1.75$, $v^2=0.5$

∴ $N(\mu_n, \sigma_n^2)$ has 95% credible interval $(1.59277, 3.071368)\}$

As $\mu|y \sim N(\mu_n, \sigma_n^2)$, a 95% credible interval is $\mu_n \pm \sigma_n Z_{0.975}$,

where $Z_{0.975}=1.96$ is Z 97.5th quantile of Z standard normal distri. for Z numbers in Z table, have $n=10$, $\bar{y}=2.867$ ∴

$\sigma_n^2 = 1/(2+5) = 1/7$, Z $\mu_n = (1.75/0.5+28.67/2)/7 = 2.33$ ∴

Z 95% credible interval is $2.33 \pm \frac{1.96}{\sqrt{7}} = [1.59, 3.07]$

\1c/ {>> pnorm $N(\mu_n, \sigma_n^2)$ $\therefore P(\mu > 2 \mid y) = 0.811$ }
normal tables standard way or >>(1-pnorm(2,2.33,1/sqrt(7)) $\therefore$
$1 - P(\mu < 2 \mid y)$ $\therefore$ $P(\mu > 2 \mid y) = 0.809$

\1d/ { for $M = 2$ $\therefore$ $m_2 = 2$ $\therefore$ Cred inter ~~the~~ $(1.66, 3.15)$
& $P(\mu > 2 \mid y) = 0.858$ }
under Z alternative prior for μ, σ_n is unchanged & $\mu_n = 2.41$
$\therefore$ Z 95% Confidence inter is $[1.67, 3.15]$ & $P(\mu > 2 \mid y) = 0.861$

\2a/ { $\theta \sim \text{Beta}(a,b)$ $\therefore$ Beta prior distri for θ with density
$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}$ }, $y_i \mid \theta \sim \text{Geo}(\theta)$ $\therefore$ $P(y_i \mid \theta) = \theta(1-\theta)^{y_i - 1}$
By Bayes theorem: $\pi(\theta \mid y) \propto \pi(\theta) P(y \mid \theta) \propto \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1} P(y \mid \theta) \propto$
$\theta^{a-1}(1-\theta)^{b-1} P(y \mid \theta) \propto \theta^{a-1}(1-\theta)^{b-1} \prod_{i=1}^{n} P(y_i \mid \theta) \propto \theta^{a-1}(1-\theta)^{b-1} \prod_{i=1}^{n} \theta(1-\theta)^{y_i - 1} \propto$
$\theta^{a-1}(1-\theta)^{b-1} \theta^n (1-\theta)^{\sum_{i=1}^{n}(y_i - 1)} \propto \theta^{a+n-1}(1-\theta)^{b-1}(1-\theta)^{n\bar{y}-n} \propto \theta^{a+n-1}(1-\theta)^{b+n\bar{y}-n-1}$ }
$\phi = \theta/(1-\theta)$ implies that $\theta = \phi/(1+\phi) = g^{-1}(\phi) = h(\phi)$ { $\phi(1-\theta) = \theta = \phi - \theta\phi$
$\therefore \theta + \theta\phi = \phi = \theta(1+\phi)$ $\therefore$ $\phi/(1+\phi) = \theta$ } & $\therefore$ $\frac{\partial\theta}{\partial\phi} = \left\{ \frac{(1+\phi)1 - \phi(1)}{(1+\phi)^2} = \frac{1+\phi-\phi}{(1+\phi)^2} = \frac{1}{(1+\phi)^2} \right]$
$= \frac{1}{1+\phi} - \frac{\phi}{(1+\phi)^2} = \frac{1}{(1+\phi)^2}$ { change of variables formula for $\phi = g(\theta)$:
$\pi_\phi(\phi) = \left| \det\left(\frac{\partial g^{-1}(\phi)}{\partial\phi} \right) \right| \pi_\theta(g^{-1}(\phi))$ } using Z change of variable formula
have prior: { $\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}$ $\Rightarrow$ $\pi_\theta(g^{-1}(\phi)) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} (g^{-1}(\phi))^{a-1}(1-g^{-1}(\phi))^{b-1}$
$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{\phi}{1+\phi}\right)^{a-1} \left(1 - \frac{\phi}{1+\phi}\right)^{b-1}$ $\therefore$ $\pi_\phi(\phi) = \frac{1}{(1+\phi)^2} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{\phi}{1+\phi}\right)^{a-1} \left(1 - \frac{\phi}{1+\phi}\right)^{b-1}$ }
$\pi_\phi(\phi) = \pi(h(\phi)) \left| \frac{\partial\theta}{\partial\phi} \right| = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{\phi}{1+\phi}\right)^{a-1} \left(1 - \frac{\phi}{1+\phi}\right)^{b-1} \frac{1}{(1+\phi)^2} = \cancel{\Gamma(a+b)}$
$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \phi^{a-1}(1+\phi)^{-a-b} = \left\{ \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{\phi}{1+\phi}\right)^{a-1} \left(\frac{1}{1+\phi}\right)^{b-1} \frac{1}{(1+\phi)^2} = \right.$
$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \phi^{a-1} \left(\frac{1}{1+\phi}\right)^{a-1} \left(\frac{1}{1+\phi}\right)^{b-1} \left(\frac{1}{1+\phi}\right)^2 = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \phi^{a-1} \left(\frac{1}{1+\phi}\right)^{a-1+b-1+2} =$
$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \phi^{a-1} \left((1+\phi)^{-1}\right)^{a+b} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \phi^{a-1}(1+\phi)^{-a-b}$ } for $\phi > 0$ $\therefore$ { By Bayes
theorem $\pi(\phi \mid y) \propto \pi(\phi) P(y \mid \phi) \propto \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \phi^{a-1}(1+\phi)^{-a-b} P(y \mid \phi) \propto$
$\phi^{a-1}(1+\phi)^{-a-b} P(y \mid \phi) \propto \phi^{a-1}(1+\phi)^{-a-b} \cancel{\phi(1-\phi)} \prod_{i=1}^{n} P(y_i \mid \phi) \propto$
$\phi^{a-1}(1+\phi)^{-a-b} \prod_{i=1}^{n} \phi(1-\phi)^{y_i - 1} \propto \phi^{a-1}(1+\phi)^{-a-b} \phi^n (1-\phi)^{\sum_{i=1}^{n}(y_i - 1)} \propto$
$\phi^{a-1}(1+\phi)^{-a-b} \phi^n (1-\phi)^{n\bar{y}-n} \propto \phi^{(a+n)-1}(1+\phi)^{(-a-b+n\bar{y}-n+1)-1}$ $\therefore$

\week 2/ posterior distri is Gamma distri: ~~Gamma~~$(a+n, -a-b+n\bar{y}-n+1)$

& prior density $\pi(\phi|y) = \phi^{(a+n)-1}(1+\phi)^{(-a-b+n\bar{y}-n+1)-1}$ X }

A similar argument yields the posterior $\pi_\phi(\phi|x) =$

$\frac{\Gamma(a_1+b_1)}{\Gamma(a_1)\Gamma(b_1)}\phi^{a_1-1}(1+\phi)^{-a_1-b_1}$ for $\phi>0$ $\{\pi(\phi|y) \propto \phi^{(a+n)-1}(1+\phi)^{(-a-b+n\bar{y}-n+1)-1}$

$\therefore$ posterior distri is Beta distri: ~~Beta~~ $\phi|y \sim \text{Beta}(a+n, -a-b+n\bar{y}-n+1)$

& posterior density $\pi(\phi|y) = \frac{\Gamma(a+n-a-b+n\bar{y}-n+1)}{\Gamma(a+n)\Gamma(-a-b+n\bar{y}-n+1)}\phi^{a+n-1}(1+\phi)^{-a-b+n\bar{y}-n}$ X }

$\pi(\phi|y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\phi^{(a+n)-1}(1+\phi)^{-(a+n)-(b-n\bar{y})} = \frac{\Gamma(a_1+b_1)}{\Gamma(a_1)\Gamma(b_1)}\phi^{a_1-1}(1+\phi)^{-a_1-b_1}$

for $\phi>0$ }

\3a/ { $\phi = 1/\theta$ $\therefore$ $\theta = 1/\phi = g^{-1}(\phi) = h(\phi) = \phi^{-1}$ $\therefore \frac{\partial\theta}{\partial\phi} = -\phi^{-2} = -\frac{1}{\phi^2}$

$\therefore \pi_\phi(\phi) = \left|\det\left(\frac{\partial g^{-1}(\phi)}{\partial\phi}\right)\right|\pi_\theta(g^{-1}(\phi))$ $\therefore \pi_\theta(\theta) = \pi_\theta(g^{-1}(\phi)) =$

$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(g^{-1}(\phi))^{a-1}(1-g^{-1}(\phi))^{b-1} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(\phi^{-1})^{a-1}\left(1-\frac{1}{\phi}\right)^{b-1}$ $\therefore$

$\pi_\phi(\phi) = \left|\frac{\partial\theta}{\partial\phi}\right|\pi_\theta(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}(\phi^{-1})^{a-1}\left(1-\frac{1}{\phi}\right)^{b-1}\frac{1}{\phi} =$

$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\phi^{-a+1}\left(\frac{\phi}{\phi}-\frac{1}{\phi}\right)^{b-1}\phi^{-1} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\phi^{-a}\left(\frac{\phi-1}{\phi}\right)^{b-1}$ $\therefore$

By Bayes theorem $\pi(\phi|y) \propto \pi(\phi)f(y|\phi)$

$f(y_i|\phi) = \phi e^{-\phi y_i}$; $y \geq 0$ $\therefore \pi(\phi|y) \propto \pi_\phi(\phi)f(y|\phi) \propto$

$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\phi^{-a}\left(\frac{\phi-1}{\phi}\right)^{b-1}f(y|\phi) \propto \phi^{-a}\left(\frac{\phi-1}{\phi}\right)^{b-1}f(y|\phi) \propto$

$\phi^{-a}\left(\frac{\phi-1}{\phi}\right)^{b-1}\prod_{i=1}^{n}f(y_i|\phi) \propto \phi^{-a}\left(\frac{\phi-1}{\phi}\right)^{b-1}\prod_{i=1}^{n}\phi e^{-\phi y_i} \propto$

$\phi^{-a}\left(\frac{\phi-1}{\phi}\right)^{b-1}\phi^n e^{\sum_{i=1}^{n}(-\phi y_i)} \propto \phi^{-a+n}\left(\frac{\phi-1}{\phi}\right)^{b-1}e^{-n\phi\bar{y}} \propto$

$\phi^{-a+n}e^{\ln\left[\left(\frac{\phi-1}{\phi}\right)^{b-1}\right]}e^{-n\phi\bar{y}} \propto \phi^{-a+n}e^{(b-1)(\ln[\phi-1]-\ln\phi)-n\bar{y}\phi} \propto$

$\phi^{-a+n}e^{(b-1)\ln[\phi-1]-(b-1)\ln\phi-n\bar{y}\phi}$

~~posterior~~ ~~$\pi_\phi(\phi|y)$~~ = posterior $\pi_\phi(\phi|y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\phi^{-a+n}e^{(b-1)\ln(\phi-1)-(b-1)\ln\phi-n\bar{y}\phi}$ X }

{ $\pi_\phi(\phi) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\phi^{-a+1}\left(\frac{\phi-1}{\phi}\right)^{b-1}(-1)(\phi)^{-2} = -\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\phi^{-a-1}\left(\frac{\phi-1}{\phi}\right)^{b-1}$ $\therefore$

$\pi(\phi|y) \propto \pi_\phi(\phi)\phi^n e^{-n\bar{y}\phi} \propto -\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\phi^{-a-1}\left(\frac{\phi-1}{\phi}\right)^{b-1}\phi^n e^{-n\bar{y}\phi} \propto$

$\phi^{-a+n-1}e^{(b-1)\ln(\phi-1)-(b-1)\ln\phi-n\bar{y}\phi}$ X }

{ $\pi_\theta(\theta) = \pi_\theta(g^{-1}(\phi)) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta} = \frac{b^a}{\Gamma(a)}(g^{-1}(\phi))^{a-1}e^{-b(g^{-1}(\phi))} =$

$\frac{b^a}{\Gamma(a)}(\phi^{-1})^{a-1} e^{-b(\phi^{-1})}$ ∴ $\pi_\phi(\phi) = \frac{b^a}{\Gamma(a)} \phi^{-a-1} e^{-b\phi}$ $|(-)\phi^{-2}| =$

$\frac{b^a}{\Gamma(a)} \phi^{-a-1} e^{-b\phi^{-1}}$ ∴ $f(y_i|\phi) = \phi e^{-\phi y}$ $y>0$ ∴ By Bayes theorem

$\pi(\phi|\underline{y}) \propto \pi(\phi) f(\underline{y}|\phi) \propto \frac{b^a}{\Gamma(a)} \phi^{-a-1} e^{-b\phi} \; f(\underline{y}|\phi) \propto$

$\cancel{\frac{b^a}{\Gamma(a)}} \phi^{-a-1} e^{-b\phi^{-1}} \phi^n e^{-n\bar{y}\phi} \propto \phi^{-a-n-1} e^{-b\frac{1}{\phi} - n\bar{y}\phi}$ }

$\phi = 1/\theta$ implies that $\theta = 1/\phi = g^{-1}(\phi) = h(\phi)$ & ∴ $\partial\theta/\partial\phi = -1/\phi^2$

Z prior $\pi_\phi(\phi) = \pi(h(\phi)) \left|\frac{\partial\theta}{\partial\phi}\right| = \frac{b^a}{\Gamma(a)} \phi^{-a} e^{-b/\phi} \phi^{-2} = \frac{b^a}{\Gamma(a)} \phi^{-a-1} e^{-b/\phi}$

A similar argument yields Z posterior $\pi_\phi(\phi) = \frac{b^a}{\Gamma(a)} \phi^{-a-1} e^{-b/\phi}$

{ $f(y_i|\phi) = \phi e^{-\phi y_i}$ ∴ By Bayes theorem: $\pi(\phi|\underline{y}) \propto \pi(\phi) f(\underline{y}|\phi) \propto$

$\pi_\phi(\phi) f(\underline{y}|\phi) \propto \frac{b^a}{\Gamma(a)} \phi^{-a-1} e^{-b/\phi} f(\underline{y}|\phi) \propto \phi^{-a-1} e^{-b/\phi} f(\underline{y}|\phi) \propto$

$\phi^{-a-1} e^{-b/\phi} \prod_{i=1}^{n} f(y_i|\phi) \propto \phi^{-a-1} e^{-b/\phi} \prod_{i=1}^{n} \phi e^{-\phi y_i} \propto \phi^{-a-1} e^{-b/\phi} \phi^n \prod_{i=1}^{n} e^{-\phi y_i} \propto$

$\phi^{-a+n-1} e^{-b/\phi} e^{\sum_{i=1}^{n}(-\phi y_i)} \propto \phi^{-a+n-1} e^{-b/\phi} e^{-\phi \frac{n}{n} \sum_{i=1}^{n} y_i} \propto \phi^{-(a-n)-1} e^{-b/\phi} e^{-n\bar{y}\phi} \propto$

$\phi^{-(a-n)-1} e^{-b/\phi - n\bar{y}\phi} \propto \phi^{-(a-n)-1} e^{-(b/\phi + n\bar{y}\phi^2/\phi)} \propto \phi^{-(a-n)-1} e^{-(b+n\bar{y}\phi^2)/\phi}$

$\cancel{\phi^{-a_1-1} e^{-b_1/\phi}}$ for $\phi > 0$, ~~with $a_1 = a-n$~~, $b_1 = b + n\bar{y}\phi^2$ ∴

~~posterior~~ ∴ $\pi(\phi|\underline{y}) \propto \phi^{-(a-n)-1} e^{-b_1/\phi} \propto \phi^{-a_1-1} e^{-b_1/\phi}$ for $\phi > 0$

with $a_1 = a-n$ ∴ posterior $\pi_\phi(\phi) = \frac{b_1^{a_1}}{\Gamma(a_1)} \phi^{a_1-1} e^{b_1/\phi}$ for $\phi > 0$ }

3b/ { Expected information $J(\theta) = E\left[-\frac{\partial^2 l}{\partial\theta^2}\right]$ ∴ Jeffreys prior ∴

$J(\theta)^{1/2}$ ∴ $y_i|\theta \sim Exp(\theta) = \theta e^{-\theta y_i}$ ∴

$L(\theta) = \prod_{i=1}^{n}(y_i|\theta) = \prod_{i=1}^{n} \theta e^{-\theta y_i} = \theta^n e^{-n\bar{y}\theta}$ ∴ $l(\theta) = \ln(L(\theta)) = \ln(\theta^n e^{-n\bar{y}\theta}) =$

$\ln(\theta^n) + \ln(e^{-n\bar{y}\theta}) = n\ln\theta - n\bar{y}\theta$ ∴ $\cancel{J(\theta)} = \frac{\partial l}{\partial\theta} = \frac{\partial}{\partial\theta}(n\ln\theta - n\bar{y}\theta) = n\frac{1}{\theta} - n\bar{y}$ ∴

$\frac{\partial^2 l}{\partial\theta^2} = \frac{\partial}{\partial\theta}\left(n\frac{1}{\theta} - n\bar{y}\right) = -n\frac{1}{\theta^2}$ ∴ $J(\theta) = E\left[--n\frac{1}{\theta^2}\right] = E\left[n\frac{1}{\theta^2}\right] = nE\left[\frac{1}{\theta^2}\right] = \frac{n}{\theta^2}$ ∴

$J(\theta)^{1/2} = \sqrt{\frac{n}{\theta^2}} = \frac{\sqrt{n}}{\theta}$ is Jeffrey's prior for θ }

Let $l(\theta)$ denote log likelihood then $l(\theta) = C + n\log(\theta) - \theta n\bar{y} \Longrightarrow$

$l'(\theta) = \frac{n}{\theta} - n\bar{y} \Longrightarrow l''(\theta) = -\frac{n}{\theta^2}$. As y is not involved, $J(\theta) = \frac{n}{\theta^2}$ &

Z Jeffrey's prior is $\pi_J(\theta) \propto \frac{1}{\theta}$ { Jeffrey's prior is Z distri with density proportional to $J(\theta)^{1/2}$ & $J(\theta)^{1/2} = \frac{\sqrt{n}}{\theta}$ ∴

$\pi_J(\theta) \propto J(\theta)^{1/2} = \frac{\sqrt{n}}{\theta} \propto \frac{1}{\theta}$ ∴ $\pi_J(\theta) \propto \frac{1}{\theta}$ }

Week 2 $y_i \mid \phi \sim Exp(\phi) = \phi e^{-\phi y_i}$ $\therefore L(\phi) = \phi^n e^{-n\bar{y}\phi}$ $\therefore$

$l(\phi) = n\ln\phi - n\bar{y}\phi$ $\therefore \frac{\partial l}{\partial\phi} = \frac{n}{\phi} - n\bar{y}$ $\therefore \frac{\partial l^2}{\partial\phi^2} = -n/\phi^2$ $\therefore$

$J(\phi) = E\left[n/\phi^2\right] = \frac{n}{\phi^2}$ $\therefore J(\phi)^{1/2} = \frac{\sqrt{n}}{\phi}$ $\therefore \pi_J(\phi) \propto \frac{1}{\phi}$ ✓ $\}$

Similar to 2 above argument $\pi_J(\phi) \propto \frac{1}{\phi}$

~~4a~~ $\{$ By Bayes theorem: $\pi(\theta|y) \propto \pi(\theta)\, p(y|\theta)$

θ has a Beta prior distribution $Beta(\alpha,\beta)$ $\therefore \pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$ for $0<\theta<1$

y is binomially distributed $\therefore p(y_i|\theta) = \binom{n}{y_i}\theta^{y_i}(1-\theta)^{n-y_i} = \frac{n!}{(n-y_i)!\,y_i!}\theta^{y_i}(1-\theta)^{n-y_i}$

$\therefore \pi(\theta|y) \propto \pi(\theta)\,p(y|\theta) \propto \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\cancel{p(y|\theta)} \propto \theta^{a-1}(1-\theta)^{b-1}p(y|\theta) \propto$

$\theta^{a-1}(1-\theta)^{b-1}\prod_{i=1}^{n}p(y_i|\theta) \propto \theta^{a-1}(1-\theta)^{b-1}\prod_{i=1}^{n}\frac{n!}{(n-y_i)!\,y_i!}\theta^{y_i}(1-\theta)^{n-y_i} \propto$

$\theta^{a-1}(1-\theta)^{b-1}\prod_{i=1}^{n}\theta^{y_i}(1-\theta)^{n-y_i} \propto \theta^{a-1}(1-\theta)^{b-1}\theta^{\sum_{i=1}^{n}y_i}(1-\theta)^{\sum_{i=1}^{n}(n-y_i)} \propto$

$\theta^{a-1}(1-\theta)^{b-1}\theta^{n\bar{y}}(1-\theta)^{\sum_{i=1}^{n}n+\sum_{i=1}^{n}(-y_i)} \propto \theta^{a-1}(1-\theta)^{b-1}\theta^{n\bar{y}}(1-\theta)^{n^2-n\bar{y}} \propto$

$\theta^{(a+n\bar{y})-1}(1-\theta)^{(b+n^2-n\bar{y})-1} \propto \theta^{a_1-1}(1-\theta)^{b_1-1}$ with $a_1 = a+n\bar{y}$, $b_1 = b+n^2-n\bar{y}$

$\therefore$ posterior distribution is Beta distribution $Beta(a_1,b_1)$

& posterior density $\pi(\theta|y) = \frac{\Gamma(a_1+b_1)}{\Gamma(a_1)\Gamma(b_1)}\theta^{a_1-1}(1-\theta)^{b_1-1}$ $\}$

$p(y) = \int_0^1 \binom{n}{y}\theta^y(1-\theta)^{n-y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\,d\theta =$

$\left\{ \int_0^1 \frac{n!}{(n-y)!\,y!}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}\,d\theta = \right.$

$\left. \int_0^1 \frac{\Gamma(n+1)}{\Gamma(n-y+1)\Gamma(y+1)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}\,d\theta \right\} =$

$\frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}\,d\theta$ $\therefore$

$p(y) = \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+y)\Gamma(n+\beta-y)}{\Gamma(\alpha+\beta+n)}$

$\{$ $\theta \sim Beta(\alpha,\beta)$ $\therefore p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$ for $0\le\theta\le1$

y is binomially distributed $\therefore$ $y|\theta \sim Bin(n,\theta)$ $\therefore$

$p(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$ for $y = 0,1,\dots,n$ $\therefore$

prior predictive distri; $p(y) = \int_0^1 p(y|\theta)\,p(\theta)\,d\theta$

$\because$ By Bayes theorem $\pi(\theta|y) \propto p(y|\theta)\pi(\theta)$ $\therefore p(y) = \int_{-\infty}^{\infty}p(y|\theta)\,\pi(\theta)\,d\theta$ $\therefore$

$$p(y)=\int_0^1 p(y|\theta)\pi(\theta)d\theta=\int_0^1 p(y|\theta)p(\theta)d\theta=\int_0^1\binom{n}{y}\theta^y(1-\theta)^{n-y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta$$

$$=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\binom{n}{y}\int_0^1\theta^y(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta=$$

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{n!}{(n-y)!\,y!}\int_0^1\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}d\theta=$$

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(n+1)}{\Gamma(n-y+1)\Gamma(y+1)}\frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(y+\alpha+n-y+\beta)}\int_0^1\frac{\Gamma(y+\alpha+n-y+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}d\theta$$

$$=\frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)}\frac{\Gamma(\alpha+y)\Gamma(n+\beta-y)}{\Gamma(\alpha+\beta+n)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\times 1\quad\Big\{\because\int_0^1\frac{\Gamma(y+\alpha+n-y+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}d\theta$$

$=1$ ∵ integrating Beta$(y+\alpha, n-y+\beta)$ pdf over all its domain ∴

$F(y=\infty)=F(\infty)=F(1)=P(y\le\infty)=P(y\le 1)=1\Big\}$

~~4b~~ $y_i|\theta\sim Bin(n,\theta)$ ∴ $p(y_i|\theta)=\binom{n}{y_i}\theta^{y_i}(1-\theta)^{n-y_i}$

$$p(\underline{y}|\theta)=\prod_{i=1}^n p(y_i|\theta)=\prod_{i=1}^n\binom{n}{y_i}\theta^{y_i}(1-\theta)^{n-y_i}=\prod_{i=1}^n\binom{n}{y_i}\theta^{\sum_{i=1}^n y_i}(1-\theta)^{\sum_{i=1}^n(n-y_i)}=$$

$$\prod_{i=1}^n\binom{n}{y_i}\theta^{n\bar{y}}(1-\theta)^{n^2-n\bar{y}}$$

By Bayes theorem: $\pi(\theta|\underline{y})\propto\pi(\theta)p(\underline{y}|\theta)$

prior predictive distri $P(y)=\int_{-\infty}^{\infty}p(y|\theta)\pi(\theta)d\theta$ ∴

$p(\tilde{y})=\int_0^1 p(\underline{y}|\theta)p(\theta)d\theta$

∴ $p(\underline{y}|\theta)=\prod_{i=1}^n\left(\frac{n!}{(n-y_i)!\,y_i!}\right)\theta^{n\bar{y}}(1-\theta)^{n^2-n\bar{y}}$

$\Big\{\prod_{i=1}^3(i)=1\cdot2\cdot3=\cancel{4}\cancel{2}=2\cdot3=6$ vs $\prod_{i=1}^3(i+x)=(1+x)(2+x)(3+x)$

$\prod_{i=1}^2\prod_{j=1}^3(3ij)=\prod_{i=1}^2\left[(3i1)(3i2)(3i3)\right]=(3\cdot1\cdot1)(3\cdot1\cdot2)(3\cdot1\cdot3)(3\cdot2\cdot1)(3\cdot2\cdot2)(3\cdot2\cdot3)\Big\}$

$$p(\underline{y}|\theta)=\frac{(n!)^n}{((n-y_i)!)^n(y_i!)^n}\left(\theta^{(n\bar{y})}\right)^n\left((1-\theta)^{n^2-n\bar{y}}\right)^n\Big\{$$

$\theta|\underline{y}\sim Beta(\alpha+N\bar{y},\beta+N(n-\bar{y}))\quad\Big\{y_i\sim Bin(n,\theta)$ ∴ $y_i|\theta\sim Bin(n,\theta)$ ∴

$p(y_i|\theta)=\binom{N}{y_i}\theta^{y_i}(1-\theta)^{N-y_i}$ for $y_i=0,1,\dots,N$ ∴

$$p(\underline{y}|\theta)=\prod_{i=1}^N p(y_i|\theta)=\prod_{i=1}^N\binom{n}{y_i}\theta^{y_i}(1-\theta)^{n-y_i}\propto\cancel{\theta^{N\bar{y}}(1-\theta)^{N(n-\bar{y})}}\theta^{N\bar{y}}(1-\theta)^{N(n-\bar{y})}$$

∴ $\cancel{\pi(\theta|\underline{y})=}$ $\pi(\theta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\propto\theta^{\alpha-1}(1-\theta)^{\beta-1}$ ∴

$$\pi(\theta|\underline{y})\propto\pi(\theta)p(\underline{y}|\theta)\propto\pi(\theta)\theta^{N\bar{y}}(1-\theta)^{N(n-\bar{y})}\propto\theta^{\alpha-1}(1-\theta)^{\beta-1}\theta^{N\bar{y}}(1-\theta)^{N(n-\bar{y})}$$

$$\propto\theta^{(\alpha+N\bar{y})-1}(1-\theta)^{\{\beta+N(n-\bar{y})\}-1}\ \therefore\ \pi(\theta|\underline{y})=\frac{\Gamma(\alpha+N\bar{y}+\beta+N(n-\bar{y}))}{\Gamma(\alpha+N\bar{y})\Gamma(\beta+N(n-\bar{y}))}\theta^{\alpha+N\bar{y}-1}(1-\theta)^{\beta+N(n-\bar{y})-1}$$

∴ $\theta|\underline{y}\sim Beta(\alpha+N\bar{y},\beta+N(n-\bar{y}))$ is the posterior distri ∴

$p(\tilde{y})=\int_0^1 p(\underline{y}|\theta)\pi(\theta)d\theta\Big\}$ ∴ $\tilde{y}|\underline{y}$ is Beta-Binomial & $p(\tilde{y}|\underline{y})=$

Week 2

$$\frac{\Gamma(n+1)\Gamma(\alpha+\cancel{\beta}+Nn)\Gamma(\alpha+N\bar{y}+\tilde{y})\Gamma(n+\beta+N(n-\bar{y})-\tilde{y})}{\Gamma(\tilde{y}+1)\Gamma(n-\tilde{y}+1)\Gamma(\alpha+N\bar{y})\Gamma(\beta+N(n-\bar{y}))\Gamma(\alpha+\beta+n+Nn)}$$

for $\tilde{y}=0,1,\ldots,n$

$$\left\{ \pi(\theta\mid y)=\frac{\Gamma(\alpha+\beta+Nn)}{\Gamma(\alpha+N\bar{y})\Gamma(\beta+N(n-\bar{y}))}\theta^{\alpha+N\bar{y}-1}(1-\theta)^{\beta+N(n-\bar{y})-1}\right.$$

$$p(y)=\int_0^1 p(y\mid\theta)\pi(\theta)\,d\theta$$

$$\left(\frac{\Gamma(n+1)\Gamma(\alpha+N\bar{y}+\tilde{y})\Gamma(n+\beta+N(n-\bar{y})-\tilde{y})}{\Gamma(\tilde{y}+1)\Gamma(n-\tilde{y}+1)\Gamma(\alpha+\beta+n+Nn)}\right\}$$

$\left\{\right.$ posterior predictive $p(\tilde{y}\mid y)=\int_0^1 p((\tilde{y}\mid y)\mid\theta)\,p(\theta)$

$\therefore f_{\theta}^{y}$ $\pi(\theta\mid y)\propto\pi(\theta)p(y\mid\theta)$

$\theta\mid y\sim Beta(\alpha+N\bar{y},\beta+N(n-\bar{y}))\ \therefore$

$\tilde{y}\mid y\sim Beta(\alpha+N\bar{y},\beta+N(n-\bar{y}))\ \therefore\ \cancel{\pi(\tilde{y}\mid y)=}$

$$\pi(\tilde{y}\mid y)=\frac{\Gamma(\alpha+\beta+Nn)}{\Gamma(\alpha+N\bar{y})\Gamma(\beta+N(n-\bar{y}))}\tilde{y}^{\alpha+N\bar{y}-1}(1-\tilde{y})^{\beta+N(n-\bar{y})-1}$$

$\theta\mid y\sim Beta(\alpha+N\bar{y},\beta+N(n-\bar{y}))$ $\quad$ $y_i\sim Bin(n,\theta)\ \therefore\ y_i\mid\theta\sim Bin(n,\theta)$

$\therefore p(y_i\mid\theta)=\binom{n}{y_i}\theta^{y_i}(1-\theta)^{n-y_i}$ $\quad\therefore\ \tilde{y}\mid y$ is Beta-Binomial

f_{θ}^{y} $p(y)=\int_{-\infty}^{\infty}p(y\mid\theta)\pi(\theta)\,d\theta\ \therefore$

f_{θ}^{y} $\tilde{y}\mid\theta\sim Bin(n,\theta)\ \therefore\ p(\tilde{y}\mid\theta)=\binom{n}{\tilde{y}}\theta^{\tilde{y}}(1-\theta)^{n-\tilde{y}}\ \therefore$

$$\int_0^1 p(\tilde{y}\mid\theta)\pi(\theta\mid y)\,d\theta=\int_0^1\left(\binom{n}{\tilde{y}}\theta^{\tilde{y}}(1-\theta)^{n-\tilde{y}}\right)\pi(\theta\mid y)\,d\theta=$$

$$\binom{n}{\tilde{y}}\int_0^1\theta^{\tilde{y}}(1-\theta)^{n-\tilde{y}}\pi(\theta\mid y)\,d\theta=\frac{n!}{(n-\tilde{y})!\,\tilde{y}!}\int_0^1\theta^{\tilde{y}}(1-\theta)^{n-\tilde{y}}\pi(\theta\mid y)\,d\theta=$$

$$\frac{\Gamma(n+1)}{\Gamma(n-\tilde{y}+1)\Gamma(\tilde{y}+1)}\int_0^1\theta^{\tilde{y}}(1-\theta)^{n-\tilde{y}}\frac{\Gamma(\alpha+\beta+Nn)}{\Gamma(\alpha+N\bar{y})\Gamma(\beta+N(n-\bar{y}))}\theta^{\alpha+N\bar{y}-1}(1-\theta)^{\beta+N(n-\bar{y})-1}\,d\theta=$$

$$\frac{\Gamma(n+1)\Gamma(\alpha+\beta+Nn)}{\Gamma(\tilde{y}+1)\Gamma(n-\tilde{y}+1)\Gamma(\alpha+N\bar{y})\Gamma(\beta+N(n-\bar{y}))}\int_0^1\theta^{\alpha+N\bar{y}+\tilde{y}-1}(1-\theta)^{\beta+N(n-\bar{y})-\tilde{y}-1}\,d\theta=$$

$$\frac{\Gamma(n+1)\Gamma(\alpha+\beta+Nn)}{\Gamma(\tilde{y}+1)\Gamma(n-\tilde{y}+1)\Gamma(\alpha+N\bar{y})\Gamma(\beta+N(n-\bar{y}))}\frac{\Gamma(\alpha+N\bar{y}+\tilde{y})\Gamma(n+\beta+N(n-\bar{y})-\tilde{y})}{\Gamma(\alpha+N\bar{y}+\tilde{y}+n+\beta+Nn-N\bar{y}-\tilde{y})}\int_0^1 Beta(\alpha+N\bar{y}+\tilde{y},\ n+\beta+N(n-\bar{y})-\tilde{y})\,d\theta$$

$$=\cancel{\frac{\Gamma(n+1)\Gamma(\alpha+\beta+Nn)}{\Gamma(\alpha+N\bar{y}}}\quad\frac{\Gamma(n+1)\Gamma(\alpha+\beta+Nn)\Gamma(\alpha+N\bar{y}+\tilde{y})\Gamma(n+\beta+N(n-\bar{y})-\tilde{y})}{\Gamma(\tilde{y}+1)\Gamma(n-\tilde{y}+1)\Gamma(\alpha+N\bar{y})\Gamma(\beta+N(n-\bar{y}))\Gamma(\alpha+\beta+n+Nn)}=p(\tilde{y}\mid y)$$

$\theta\mid y\sim Beta(\alpha+N\bar{y},\beta+N(n-\bar{y}))\ \therefore\ \pi(\theta\mid y)\sim Beta$, $\tilde{y}\mid\theta\sim Bin(n,\theta)\ \therefore\ p(\tilde{y}\mid\theta)\sim Bin$

$$\therefore\ p(\tilde{y}\mid y)=\int_0^1 p(\tilde{y}\mid\theta)\,\pi(\theta\mid y)\,d\theta$$

4c $\theta \sim \text{Beta}(\alpha, \beta)$ $\therefore$ prior of θ: $\pi(\theta) = \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)} \theta^{1-1}(1-\theta)^{1-1} =$

$\frac{1!}{0!\,0!} \theta^0 (1-\theta)^0 = \frac{1}{1 \cdot 1} 1 \cdot (1) = 1$ $\therefore$ $\theta \sim \text{Unif}[0,1]$ $\therefore$

Model the number of patients responding, y, within my group of 5 tested via $y \sim \text{Bin}(5, \theta)$

$\theta \sim \text{Unif}[0,1]$ $\quad p(y) = \int_0^1 p(y|\theta)\pi(\theta)\, d\theta$ $\quad \therefore \pi(\theta) = \frac{1}{\theta}$

$p(y|\theta) = \binom{5}{y}\theta^y(1-\theta)^{n-y} = \frac{5!}{(5-y)!\,5!}\theta^y(1-\theta)^{n-y}$ $\therefore$

$p(y) = \int_0^1 \frac{5!}{(5-y)!\,\cancel{y}!}\theta^y(1-\theta)^{n-y}\frac{1}{\theta}\, d\theta = \frac{5!}{(5-y)!\,y!}\int_0^1 \theta^y(1-\theta)^{n-y}\theta^{-1}\, d\theta =$

$\frac{120}{\Gamma(5-y+1)\Gamma(y)}\int_0^1 \theta^{y-1}(1-\theta)^{(n-y+1)-1}\, d\theta =$

$\frac{120}{\Gamma(5-y+1)\Gamma(y)} \frac{\Gamma(y)\Gamma(n-y+1)}{\Gamma(n+1)} \int_0^1 \frac{\Gamma(y+n-y+1)}{\Gamma(y)\Gamma(n-y+1)} \theta^{y-1}(1-\theta)^{(n-y+1)-1}\, d\theta =$

$\frac{120}{\Gamma(5-y+1)\Gamma(y)} \frac{\Gamma(y)\Gamma(n-y+1)}{\Gamma(n+1)} = p(y)$

$y \sim \text{Bin}(5, \theta)$ with $\theta \sim \text{Beta}(1,1)$. The prob for the event that none of the patients in your group respond is $P(y=0)$

$\cancel{120}\ \cancel{p(y)} =$ $n = 5$ $\therefore$ $p(y) = \frac{120\,\Gamma(y)\Gamma(5-y+1)}{\Gamma(5-y+1)\Gamma(y)\,120} = \frac{\Gamma(y)\Gamma(5-y+1)}{\Gamma(5-y+1)\Gamma(y)}$ $\therefore$

for no one responding: $P(y=0) = \frac{\Gamma(0)\Gamma(5-0+1)}{\Gamma(5-0+1)\Gamma(0)} = 1$ X

from part (a): $p(y) = \frac{\Gamma(6)\,\Gamma(2)\,\Gamma(1+y)\Gamma(5+1-y)}{\Gamma(y+1)\Gamma(5-y+1)\Gamma(1)\Gamma(1)\Gamma(1+1+5)} =$

$\frac{\Gamma(6)}{\Gamma(7)} = \frac{5!}{6!} = \frac{1}{6} = P(y=0)$

is obtained from the prior predictive distri from part (a) have:

$P(y=0) = \frac{\Gamma(6)\,\Gamma(2)\,\Gamma(1)\,\Gamma(6)}{\Gamma(1)\Gamma(6)\Gamma(1)\Gamma(1)\,\Gamma(7)} = \frac{1}{6}$

4d $\bar{y} = \frac{0+0+1}{3} = \frac{1}{3}$ $\therefore$ from part (b) $\therefore$ $P(\tilde{y}|y)$ $\therefore$ $P(\tilde{y}=0|y)$ $\therefore$ $N=3$ $\therefore$

$P(\tilde{y}=0|y) = \frac{\Gamma(5+1)\Gamma(1+1+3\cdot 5)\Gamma(1+3\frac{1}{3}+0)\Gamma(5+1+3(5-\frac{1}{3})-0)}{\Gamma(0+1)\Gamma(5-0+1)\Gamma(1+3\cdot\frac{1}{3})\Gamma(1+3(5-\frac{1}{3}))\Gamma(1+1+5+3\cdot 5)}$

$= \frac{\Gamma(6)\Gamma(17)\Gamma(2)\Gamma(22)}{\Gamma(1)\Gamma(6)\Gamma(2)\Gamma(15)\Gamma(22)} = \frac{\Gamma(17)}{\Gamma(15)} = \frac{16!}{14!} = 16 \times 15 = 240$

\Week2/ our probab having seen Z results of Z other experiments has changed through learning about θ & is given by Z posterior predictive probab $P(\tilde{y}=0|\underline{y})$ derived in part (b) with $y=(0,0,1)$, $N=3$. Here $N\bar{y}=1$ & $N(n-\bar{y})=1$ ∴ Subinte posterior predictive distri derived in part (b):

$$p(\tilde{y}=0|y)=\frac{\Gamma(6)\Gamma(17)\Gamma(2)\Gamma(20)}{\Gamma(1)\Gamma(6)\Gamma(2)\Gamma(15)\Gamma(22)}=\frac{4}{7}$$

\5/ $\{ P(y|\theta)=\theta(1-\theta)^{y-1}$ $\quad \pi(\theta)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$ $\quad 0<\theta<1$

$$p(y)=\int_0^1 p(y|\theta)\pi(\theta)\,d\theta=\int_0^1\theta(1-\theta)^{y-1}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\,d\theta=$$

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1\theta^{(a+1)-1}(1-\theta)^{(b+y-1)-1}\,d\theta=$$

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+1)\Gamma(b+y-1)}{\Gamma(a+b+y)}\int_1^{\infty}\frac{\Gamma(a+1+b+y-1)}{\Gamma(a+1)\Gamma(b+y-1)}\cancel{\theta}\,\theta^{(a+1)-1}(1-\theta)^{(b+y-1)-1}\,d\theta=$$

$$\frac{\Gamma(a+b)\Gamma(a+1)\Gamma(b+y-1)}{\Gamma(a)\Gamma(b)\Gamma(a+b+y)}$$ is Z prior predictive distribution

By Bayes theorem: $\pi(\theta|y)\propto\pi(\theta)p(y|\theta)$ ∴

$\pi(\theta)\propto\theta^{a-1}(1-\theta)^{b-1}$, $p(y|\theta)\propto\prod_{i=1}^{n}p(y_i|\theta)=\prod_{i=1}^{n}\theta(1-\theta)^{y_i-1}\propto$

$\theta^n(1-\theta)^{\sum_{i=1}^{n}(y_i-1)}\propto\theta^n(1-\theta)^{-n+n\bar{y}}$ ∴

$\pi(\theta|y)\propto\pi(\theta)p(y|\theta)\propto\theta^{a-1}(1-\theta)^{b-1}p(y|\theta)\propto$

$\theta^{a-1}(1-\theta)^{b-1}\theta^n(1-\theta)^{-n+n\bar{y}}\propto\theta^{(a+n)-1}(1-\theta)^{(b+n\bar{y}-n)-1}\propto\theta^{a_1-1}(1-\theta)^{b_1-1}$,

with $a_1=a+n$, $b_1=b+n\bar{y}-n$ ∴

posterior distribution is: Beta distribution: Beta(a_1,b_1) &

posterior density $\pi(\theta|y)=\frac{\Gamma(a_1+b_1)}{\Gamma(a_1)\Gamma(b_1)}\theta^{a_1-1}(1-\theta)^{b_1-1}=$

$$\frac{\Gamma(a+n+b+n\bar{y}-n)}{\Gamma(a+n)\Gamma(b+n\bar{y}-n)}\theta^{a+n-1}(1-\theta)^{b+n\bar{y}-n-1}=\frac{\Gamma(a+b+n\bar{y})}{\Gamma(a+n)\Gamma(b+n\bar{y}-n)}\theta^{a+n-1}(1-\theta)^{b+n\bar{y}-n-1}\},$$

have $f(y|\theta)=\theta(1-\theta)^{y-1}$ for positive integers y &

$\pi(\theta)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$ for $0<\theta<1$ So, Z prior predictive distri has mass func $\pi(y)=\int_0^1 f(y|\theta)\pi(\theta)\,d\theta=$

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1\theta(1-\theta)^{y-1}\theta^{a-1}(1-\theta)^{b-1}\,d\theta=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1\theta^{(a+1)-1}(1-\theta)^{(b+y-1)-1}\,d\theta=$$

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+1)\Gamma(b+y-1)}{\Gamma(a+b+y)}\int_0^1 \frac{\Gamma(a+b+y)}{\Gamma(a+1)\Gamma(b+y-1)}\theta^{(a+1)-1}(1-\theta)^{(b+y-1)-1}\,d\theta =$$

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+1)\Gamma(b+y-1)}{\Gamma(a+b+y)}$$ $\because$ Z integrand is a density. similarly, Z

posterior predictive distri has mass func

$$\pi(y|\underline{x}) = \frac{\Gamma(a+n+1)\Gamma(b+\sum_{i=1}^{n} x_i - n + y - 1)}{\Gamma(a+b+\sum_{i=1}^{n} x_i + y)} \frac{\Gamma(a+b+\sum_{i=1}^{n} x_i)}{\Gamma(a+n)\Gamma(b+\sum_{i=1}^{n} x_i - n)}$$

$$\left\{ p(y|\underline{x}) = \int_0^1 p(y|\theta)\,\pi(\theta|\underline{x})\,d\theta \quad \therefore \right.$$

$$\pi(\theta|\underline{x}) = \frac{\Gamma(a+b+n\bar{x})}{\Gamma(a+n)\Gamma(b+n\bar{x}-n)}\theta^{a+n-1}(1-\theta)^{b+n\bar{x}-n-1} \quad \therefore$$

$$p(y|\underline{x}) = \int_0^1 \theta(1-\theta)^{y-1}\frac{\Gamma(a+b+n\bar{x})}{\Gamma(a+n)\Gamma(b+n\bar{x}-n)}\theta^{a+n-1}(1-\theta)^{b+n\bar{x}-n-1}\,d\theta =$$

$$\frac{\Gamma(a+b+n\bar{x})}{\Gamma(a+n)\Gamma(b+n\bar{x}-n)}\int_0^1 \theta^{(a+n+1)-1}(1-\theta)^{(b+y+n\bar{x}-n-1)-1}\,d\theta =$$

$$\frac{\Gamma(a+n+1)\Gamma(b+n\bar{x}-n+y-1)}{\Gamma(a+b+n\bar{x}+y)}\frac{\Gamma(a+b+n\bar{x})}{\Gamma(a+n)\Gamma(b+n\bar{x}-n)}\int_0^1 \frac{\Gamma(a+n+1+b+y+n\bar{x}-n-1)}{\Gamma(a+n+1)\Gamma(b+n\bar{x}-n+y-1)}\theta^{(a+n+1)-1}(1-\theta)^{(b+y+n\bar{x}-n-1)-1}\,d\theta$$

$$= \frac{\Gamma(a+n+1)\Gamma(b+n\bar{x}-n+y-1)\Gamma(a+b+n\bar{x})}{\Gamma(a+b+n\bar{x}+y)\Gamma(a+n)\Gamma(b+n\bar{x}-n)}$$ $\therefore p(y|\underline{x})$ is Z posterior predictive

distri $\therefore$

$$p(y=1) = \frac{\Gamma(a+b)\Gamma(a+1)\Gamma(b+1-1)}{\Gamma(a)\Gamma(b)\Gamma(a+b+1)} = \frac{\Gamma(a+b)\Gamma(a+1)\Gamma(b)}{\Gamma(a)\Gamma(b)\Gamma(a+b+1)} = \frac{\Gamma(a+b)\Gamma(a+1)}{\Gamma(a)\Gamma(a+b+1)}$$

$$p(y=1|\underline{x}) = \frac{\Gamma(a+n+1)\Gamma(b+n\bar{x}-n+1-1)\Gamma(a+b+n\bar{x})}{\Gamma(a+b+n\bar{x}+1)\Gamma(a+n)\Gamma(b+n\bar{x}-n)} =$$

$$\frac{\Gamma(a+n+1)\Gamma(b+n\bar{x}-n)\Gamma(a+b+n\bar{x})}{\Gamma(a+b+n\bar{x}+1)\Gamma(a+n)\Gamma(b+n\bar{x}-n)} \quad \therefore \quad p(y=1|\underline{x}) > p(y=1) \quad \therefore$$

$$\frac{\Gamma(a+n+1)\Gamma(b+n\bar{x}-n)\Gamma(a+b+n\bar{x})}{\Gamma(a+b+n\bar{x}+1)\Gamma(a+n)\Gamma(b+n\bar{x}-n)} > \frac{\Gamma(a+b)\Gamma(a+1)}{\Gamma(a)\Gamma(a+b+1)} \quad \therefore$$

$$\frac{(a+n)!\,(b+n\bar{x}-n-1)!\,(a+b+n\bar{x}-1)!}{(a+b+n\bar{x})!\,(a+n-1)!\,(b+n\bar{x}-n-1)!} > \frac{(a+b-1)!\,a!}{(a-1)!\,(a+b)!} \quad \therefore$$

$$\frac{(a+n)(a+n-1)!\,(b+n\bar{x}-n-1)!\,(a+b+n\bar{x}-1)!}{(a+b+n\bar{x})!\,(a+n-1)!\,(b+n\bar{x}-n-1)!} > \frac{(a+b-1)!\,a(a-1)!}{(a-1)!\,(a+b)(a+b-1)!} \quad \therefore$$

$$\frac{(a+n)(b+n\bar{x}-n-1)!\,(a+b+n\bar{x}-1)!}{(a+b+n\bar{x})(a+b+n\bar{x}-1)!\,(b+n\bar{x}-n-1)!} > \frac{a}{a+b} \quad \therefore$$

Week 2 / $\cancel{\frac{(a+n)(a+b+n\bar{x}-1)!}{(a-}}$ $\frac{(a+n)(b+n\bar{x}-n-1)!}{(a+b+n\bar{x})} > \frac{a}{a+b} \therefore$

$(a+n)(b+n\bar{x}-n-1)! > \frac{a(a+b+n\bar{x})}{a+b} \therefore$

$\left.(b+n\bar{x}-n-1)! > \frac{a(a+b+n\bar{x})}{(a+b)(a+n)}\right\}$

$\because$ $\Gamma(t+1) = t\Gamma(t)$ $\forall t \in \mathbb{R}$: subing $y=1$ into these distris yields:

$\pi(y=1) = \frac{\Gamma(a+b)\Gamma(a+1)\Gamma(b)}{\Gamma(a)\Gamma(b)\Gamma(a+b+1)} = \frac{a}{a+b}$ Σ

$\pi(y=1|\underline{y}) = \frac{\Gamma(a+n+1)\Gamma(b+\sum_{i=1}^{n} y_i - n)}{\Gamma(a+b+\sum_{i=1}^{n} y_i + 1)} \frac{\Gamma(a+b+\sum_{i=1}^{n} y_i)}{\Gamma(a+n)\Gamma(b+\sum_{i=1}^{n} y_i - n)} =$

$\left\{\frac{(a+n)\Gamma(a+n)\Gamma(b+\sum_{i=1}^{n} y_i - n)}{(a+b+\sum_{i=1}^{n} y_i)\Gamma(a+b+\sum_{i=1}^{n} y_i)}\right. \frac{\Gamma(a+b+\sum_{i=1}^{n} y_i)}{\Gamma(a+n)\Gamma(b+\sum_{i=1}^{n} y_i - n)} =$

$\frac{(a+n)}{(a+b+\sum_{i=1}^{n} y_i} = \left.\right\} \frac{a+n}{a+b+\sum_{i=1}^{n} y_i}$ $\therefore \left\{ \frac{a+n}{a+b+n\bar{y}} > \frac{a}{a+b} \therefore \right.$

$\frac{(a+n)(a+b)}{a} > a+b+n\bar{y} \therefore$

$a+b+n\bar{y} < \frac{a^2+ab+an+bn}{a} = a+b+n+\frac{bn}{a} \therefore$

$n\bar{y} < n+\frac{bn}{a} \therefore$ $\cancel{\bar{y} \leq \frac{a+b}{a}}$ $\left.\bar{y} < 1+\frac{b}{a}\right\}$ $\therefore \pi(y_1|\underline{y}) > \pi(y=1) \Longleftrightarrow$

$(a+n)(a+b) > a(a+b+n\bar{y}) \Longleftrightarrow \bar{y} < 1+b/a$

6 / {Gamma prior for λ $\therefore \pi(\lambda) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}$ for $\lambda > 0$

$f(y_i\lambda) = \frac{\lambda^y}{y!}e^{-\lambda}$ $\therefore$ By Bayes theorem: $\pi(\lambda|\underline{y}) \propto \pi(\lambda) f(\underline{y}|\lambda) \propto$

$\frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda} f(\underline{y}|\lambda) \propto \lambda^{a-1}e^{-b\lambda} f(\underline{y}|\lambda) \propto \lambda^{a-1}e^{-b\lambda}\prod_{i=1}^{n} f(y_i|\lambda) \propto$

$\lambda^{a-1}e^{-b\lambda}\prod_{i=1}^{n}\frac{\lambda^{y_i}}{y_i!}e^{-\lambda} \propto \lambda^{a-1}e^{-b\lambda} \prod_{i=1}^{n}\left(\frac{1}{(y_i)!}\right)\lambda^{\sum_{i=1}^{n} y_i}(e^{-\lambda})^n \propto$

$\lambda^{a-1}e^{-b\lambda}\lambda^{n\bar{y}}e^{-n\lambda} \propto \lambda^{(a+n\bar{y})-1}e^{-(b+n)\lambda} \propto \lambda^{a_1-1}e^{-b_1\lambda}$, with $a_1 = a+n\bar{y}$,

~~posterior~~ $b_1 = b+n$ $\therefore$ posterior density $\left.\pi(\lambda|\underline{y}) = \frac{(b+n)^{a+n\bar{y}}}{\Gamma(a+n\bar{y})}\lambda^{a+n\bar{y}-1}e^{-(b+n)\lambda}\right\}$

$\therefore$ posterior distribution is Gamma distribution Gamma(a_1, b_1)

for $\lambda > 0$}

Week 4 1 $\left\{ \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\,dx = 1 \quad \therefore \right.$

$\left. \int_0^\infty \frac{2^{4.5}}{\Gamma(4.5)} x^{4.5-1} e^{-2x}\,dx = \cancel{\int_0^\infty} 1 = \frac{2^{4.5}}{\Gamma(4.5)} \int_0^1 x^{4.5-1} e^{-2x}\,dx \right\}$

2a $\{$ a distribution from $-\infty < x < \pi$

$\therefore$ use normal distribution $\therefore f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$\therefore$ choose $\mu = 2$, $\sigma = 1$ $\therefore \}$

$\sin(x) e^{-\frac{1}{2}(x-2)^2} = g(x) f(x) \quad \therefore g(x) = \sin(x)\sqrt{2\pi}\,\sigma$

$\therefore f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-2}{1}\right)^2}$

with $-\infty <$ sample $< \pi$, samples $\sim N(2,1)$

$\therefore -\infty < f(x) < \pi$

3a $\left\{ p(Y) = \int_{-\infty}^{\infty} p(Y|\mu)\pi(\mu)\,d\mu \right.$

By Bayes theorem $\pi(\mu|y) \propto \pi(\mu) p(y|\mu)$

$\left. p(\tilde{y}|y) = \int_{-\infty}^{\infty} p(\tilde{y}|\mu)\pi(\mu|y)\,d\mu \right\}$

from last week $\mu|Y \sim N(\mu_n, \sigma_n^2)$ with $\mu_n = 2.33$ & $\sigma_n = 1/\sqrt{7}$

$\{ \mu \sim N(1.75, 0.5) \quad \therefore E(\mu) = 1.75$, $\sigma^2(\mu) = \mathrm{var}(\mu) = 0.5$

By Bayes $\pi(\mu|y) \propto \pi(\mu) f(y|\mu) \propto \frac{1}{\sigma(\mu)\sqrt{(2\pi)}} e^{-\frac{1}{2\sigma^2(\mu)}(\mu - E(\mu))^2} f(y|\theta) \propto$

$e^{-\frac{1}{2\sigma^2(\mu)}(\mu - E(\mu))^2} f(y|\theta) \propto e^{-\frac{1}{2\sigma^2(\mu)}(\mu - E(\mu))^2} f(y|\mu, \sigma^2(y|\mu)) \propto$

$e^{-\frac{1}{2\sigma^2(\mu)}(\mu - E(\mu))^2} \prod_{i=1}^{n} e^{-\frac{1}{2\sigma^2(y|\mu)}(y_i - \mu)^2} \propto$

$\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2(\mu)}(\mu^2 - 2E(\mu)\mu) + \frac{1}{\sigma^2(y|\mu)}(-2n\bar{y}\mu + n\mu^2)\right)\right] \propto$

$\exp\left[-\frac{1}{2}\left(\left[\frac{1}{\sigma^2(\mu)} + \frac{n}{\sigma^2(y|\mu)}\right]\mu^2 - 2\left[\frac{E(\mu)}{\sigma^2(\mu)} + \frac{n\bar{y}}{\sigma^2(y|\mu)}\right]\right)\right] = \exp\left[-\frac{1}{2a^{-1}}\left(\mu - \frac{b}{a}\right)^2\right] \therefore$

$\mu|y \sim N\left(\frac{b}{a}, a^{-1}\right) \quad \therefore \pi(\mu|y) \sim N(\mu_n, \sigma_n^2)$

$\cancel{\sigma_n^2 = \frac{1}{\frac{1}{\gamma^2} + \frac{1}{\sigma^2}}, \mu_n =} \quad \sigma_n^2 = \frac{1}{\frac{1}{\sigma^2(\mu)} + \frac{n}{\sigma^2(y|\mu)}}, \; \mu_n = \frac{\frac{E(\mu)}{\sigma^2(\mu)} + \frac{n\bar{y}}{\sigma^2(y|\mu)}}{\frac{1}{\sigma^2(\mu)} + \frac{n}{\sigma^2(y|\mu)}} \therefore$

$\therefore p(\tilde{y}|y) = \int_{-\infty}^{\infty} p(\tilde{y}|\mu)\pi(\mu|y)\,d\mu$

$\therefore \mu|y \sim N(\mu_n, \sigma_n^2) \quad \therefore$ >> mean(samples > 2) $\Rightarrow 0.86 \}$

to obtain Monte Carlo esti require: $P(\mu > 2|Y) = E[g(\mu)|Y] = \int_{-\infty}^{\infty} g(\mu)\pi(\mu|Y)\,d\mu$ with $g(\mu) = \mathbb{1}(\mu > 2)$ ok

\Week4/ \3b/ {100 samples gives $P(\mu>2|Y) = 0.86$ is 2 decimal places ∴ to have it to 3 decimal places to be within 0.01 may need 1000 samples ∵ increasing Z number of samples by a factor of 10 should increase the accuracy of Z probab by a factor of 10}

Z variance of g is `>> Varg <- MCarlo*(1-MCarlo)` `>> Varg` ⇛ 0.1411

& we require $\varepsilon = \sqrt{var[g(\mu)]/N} \leq 0.01$

{∵ Monte Carlo estimate for $\pi(\theta \in S | y)$ is $\hat{p} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(\theta_i \in S)$ with standard error $\sqrt{\hat{p}(1-\hat{p})/N}$}

{∴ $Var[g(\mu)]/N \leq 0.0001$ ∴ ~~pJZt~~

$\frac{var[g(\mu)]}{0.0001} \leq N$ ∵ $var[g(\mu)] = 0.86(1-0.86) = 0.1204$ ∴

$\frac{0.1204}{0.0001} = 1204 \leq N$ ∴ $N = 1204$}

∴ $N \geq \frac{var[g(\mu)]}{0.01^2}$ ∴ `Nnew <- varg/(0.01^2)` `>> Nnew` ⇛ 1411

\3c/ {is `>> Nnew` ⇛ 1824 : `>> mean(Samples>2)` ⇛ 0.803

∴ $\sqrt{\hat{p}(1-\hat{p}/N_{new}} = 0.00931$} `>> MCerror2` ⇛ 0.0104

\3d/ {$p(\tilde{y}|Y) = \int_{-\infty}^{\infty} p(\tilde{y}|\mu)\pi(\mu|Y)\,d\mu$

$p(\tilde{y}<0|Y) = E(g(\tilde{y})|Y] = \int_{-\infty}^{\infty} g(\tilde{y})\pi(\tilde{y}|Y)\,d\tilde{y}$ with $g(\tilde{y}) = \mathbb{1}(\tilde{y}<0)$}

now we need to sample from Z posterior predictive

{$p(\tilde{y}, Y) = \int_{-\infty}^{\infty} p(\tilde{y}|\mu)\,\pi(\mu|Y)\,d\mu$ ∴ ySamples ~ N(μSamples, sd=2)

∵ $y|\mu \sim N(\mu, 4)$}

then calc how many samples we need to get Z estimate to within 0.01

finally our estimate is

\3e/ {$E[\tilde{y}|Y] = \int_{-\infty}^{\infty} \tilde{y}\,\pi(\tilde{y}|Y)\,d\tilde{y}$ of $E[\tilde{y}|Y] = \int_{-\infty}^{\infty} (\tilde{y}|Y)\,dP(\tilde{y}|Y)$

$E[\tilde{y}|Y] = \frac{1}{n}\sum_{i=1}^{n} p(\tilde{y}|Y)$ `>> var(y_tilda_Samples)`

$SE = \sqrt{var[g(\tilde{y})]/N}$ } can use our existing samples to figure out how many samples we might need ∴ So Z estimate is ⇛

2.33 (as you would expect given Z mean of μ & Z model) $\therefore$ Z plot is {a histogram with a line of Z mean & two more lines of Z $\pm 1.96\sigma$ for Z $\pm$ 97.5% quantiles for Z 95% credible) interval

\4a/ $\text{Beta}(\alpha, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}$ for $0 < \theta < 1$

$\therefore$ for $0 < x < 1$ $\therefore$ $\text{Beta}(a, b) = f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$ $\therefore$

$K = \int_0^{0.5} g(x) f(x)\,dx$ $\therefore$ $x^a(1-x)^b e^{-cx} =$

$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} x(1-x) e^{-cx}$ $\therefore$

$g(x) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} x(1-x) e^{-cx}$ $\therefore$ $0 \leq \text{Sample} \leq 0.5$,

Samples $\sim$ Beta(a, b) $\therefore$ $0 \leq f(x) \leq 0.5$

$\therefore$ let $a=1, b=1, c=1$; $\therefore$ $K = \int_0^{0.5} x^a(1-x)^b e^{-cx}\,dx = \int_0^{0.5} g(x) f(x)\,dx$

with $f(x)$ ~~$\sim$ Beta(a,b) $\therefore$ take only $0 \leq f(x) \leq 0.5$~~ $\therefore$

Samples $\sim$ Beta(a, b) & integrand = g(samples) &

$f(x)$ = samples $\therefore$ $f(x) \sim$ Beta(a, b) $\therefore$ keep only $0 \leq f(x) \leq 0.5$ $\therefore$

MonteCarlo approx for K is $K \approx$ Mean(keepers * integrand)

with keepers = $\mathbb{1}(0 \leq f(x) \leq 0.5)$

\4b/ $\text{Gamma}(\alpha, c) = \frac{c^a}{\Gamma(a)} x^{a-1} e^{-cx}$ for $\theta > 0$ $\therefore$ if

Samples $\sim$ Gamma(a, c) $\therefore$ $x^a(1-x)^b e^{-cx} =$

$\frac{c^a}{\Gamma(a)} x^{a-1} e^{-cx} \frac{\Gamma(a)}{c^a} x(1-x)^b = f(x) g(x)$ $\therefore$

$f(x) \sim$ Gamma(a, c) , $g(x) = \frac{\Gamma(a)}{c^a} x(1-x)^b$ $\therefore$ Samples $\sim$ Gamma(a, c)

integrand = g(samples) & $f(x)$ = Samples $\therefore$ keep only $0 \leq f(x) \leq 0.5$ $\therefore$

keepers = $\mathbb{1}(0 \leq f(x) \leq 0.5)$ $\therefore$ MonteCarlo approx for K is

$K \approx$ Mean(keepers * integrand)

\4c/ does it suppose to work for Gamma prior? isn't then beta distri version gives Z most accurate esti

$\sim -\infty$