Derivations for the review of probability

Dose Conditional

* Let's compute $f(x|0)$. Since $Y=0$ we can restrict our attention to the first row of the joint PMF table

		X: 5	10	20	
Y	0	0.469	0.124	0.044	0.640 $f_Y(0)$
	1				

The conditional of $X|Y=0$ just rescales the rows by their sum, so

$$P(X=5|Y=0) = \frac{0.464}{0.640} = 0.73$$

$$P(X=10|Y=0) = \frac{0.124}{0.640} = 0.19$$

$$P(X=20|Y=0) = \frac{0.044}{0.640} = 0.08$$

* Let's also do

$$P(Y=1|X=5) = \frac{f(5,1)}{f_X(5)} = \frac{0.231}{0.700} = 0.33$$

Similarly, $P(Y=1|X=10) = 0.38$ & $P(Y=1|X=15) = 0.51$

Dose $X=5$ is the best

Dose Marginal

The joint distribution is

		X			
		5	10	20	
Y	0	0.469	0.124	0.049	0.642
	1	0.231	0.076	0.051	0.358
		0.700	0.200	0.100	

$\underbrace{0.700 \quad 0.200 \quad 0.100}_{f_X(x)}$ $\qquad$ $\left.\begin{matrix}0.642\\0.358\end{matrix}\right\} f_Y(y)$

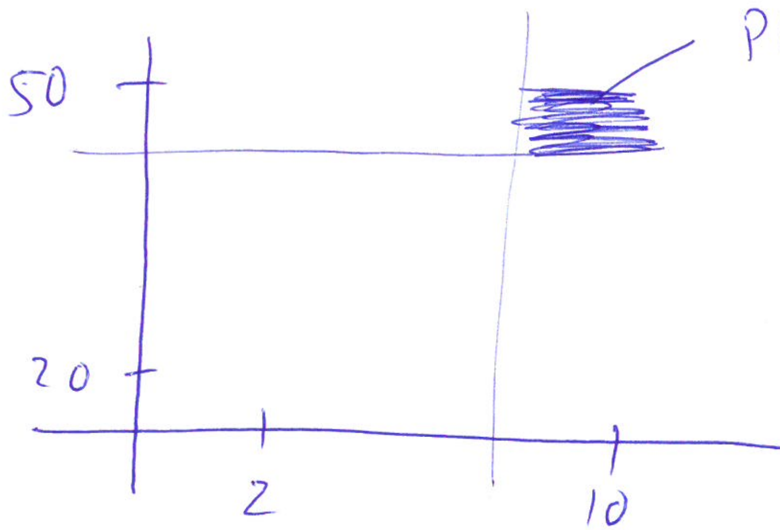
$$P(X=5) = f_X(x) = f(5,0) + f(5,1) = 0.469 + 0.231 = 0.700$$

$$P(Y=0) = \sum_{x \in \{5,10,20\}} f(x,5) = f(5,0) + f(10,0) + f(20,0)$$

$$= 0.469 + 0.124 + 0.044 = 0.642$$

BW Prob

$$f(x,y) = 0.26\, e^{-|x-7|-|y-40|} \quad \text{for} \quad \begin{array}{l} x \in (2,10) \\ y \in (20,50) \end{array}$$



P(X>7, Y>40) is the volume under this region

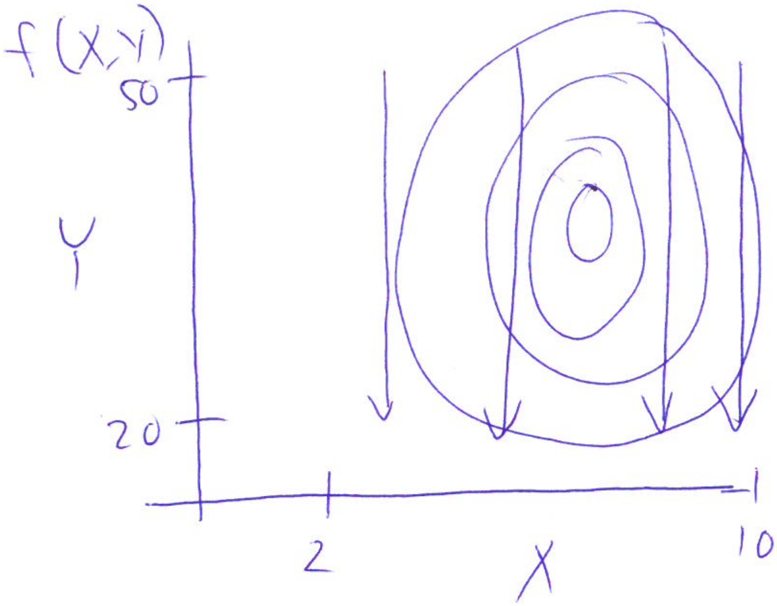
$$P(X>7, Y>40) = \int_7^{10}\int_{40}^{50} f(x,y)\,dx\,dy$$

$$= \int_7^{10}\int_{40}^{50} 0.26\, e^{-|x-7|-|y-40|}\,dx\,dy$$

$$= \int_7^{10}\int_{40}^{50} 0.26\, e^{-(x-7)-(y-40)}\,dx\,dy$$

$$= 0.25$$

BW Marginal



$f_X(x)$ is just the column sums of $f(x,y)$. The column sums are integrals

$$f_X(x) = \int f(x,y)\,dy$$

$$= \int_{20}^{50} 0.26\, e^{-|x-7| - |y-40|}\, dy$$

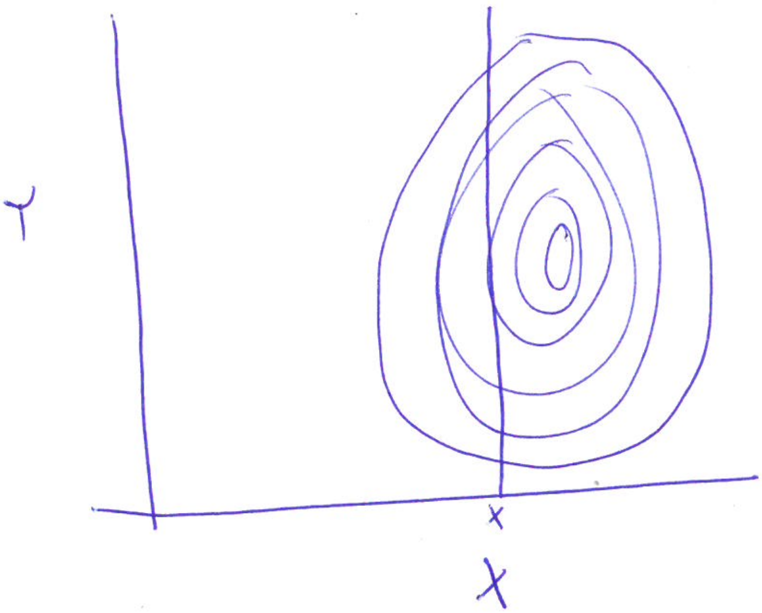
$$= 0.26\, e^{-|x-7|} \int e^{-|y-40|}\, dy$$

$$= 0.26\, e^{-|x-7|}\, \overset{\approx}{2}$$

$$= 0.52\, e^{-|x-7|}$$

$$\propto e^{-|x-7|}$$

BW Conditional



The conditional distribution of $Y|X=x$ just takes the $X=x$ column & renormalizes to integrate to one.

$$\text{So} \quad f(y|x) = \frac{f(x,y)}{f(x)} = \frac{0.26\, e^{-|x-7|-|y-40|}}{0.52\, e^{-|x-7|}}$$

$$= \frac{1}{2} e^{-|y-40|}$$

A few more lines of calculus would show that $\int_{20}^{50} f(y|x)\,dy = \int_{20}^{50} \frac{1}{2} e^{-|y-40|}\,dy = 1$.

MVN Marginal

The joint distribution is

$$\frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}\left[\frac{x^2+y^2-2\rho xy}{1-\rho^2}\right]}$$

The marginal of X is

$$f_X(x) = \int \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left[x^2+y^2-2\rho xy\right]}\,dy$$

let's rearrange this to look like a univariate Gaussian PDF for y.

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\, e^{-\frac{x^2}{2(1-\rho^2)}} \int \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2-2\rho xy}{2(1-\rho^2)}}\,dy$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\, e^{-\frac{x^2}{2(1-\rho^2)}} \int \frac{\sqrt{1-\rho^2}}{\sqrt{2\pi(1-\rho^2)}}\, e^{-\frac{y^2-2\rho xy+(\rho x)^2-(\rho x)^2}{2(1-\rho^2)}}\,dy$$

$$= \frac{1}{\sqrt{2\pi}\,\cancel{\sqrt{1-\rho^2}}}\, e^{-\frac{x^2}{2(1-\rho^2)}}\,\cancel{\sqrt{1-\rho^2}}\, e^{+\frac{\rho^2x^2}{2(1-\rho^2)}} \underbrace{\int \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\, e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}}\,dy}_{1}$$

since $Y \sim N(\rho x, 1-\rho^2)$

$$= \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{x^2-\rho^2x^2}{1-\rho^2}}$$

$$= \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \qquad \text{so } X \sim N(0,1)$$

MVN Conditional

The joint is $f(x,y) = \dfrac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{x^2+y^2-2\rho xy}{1-\rho^2}}$

X's marginal is $N(0,1)$, so $f_X(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

The conditional is the ratio

$$f(y|x) = \frac{f(x,y)}{f_X(y)} = \frac{\dfrac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{x^2+y^2-2\rho xy}{1-\rho^2}}}{\dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left[\frac{x^2+y^2-2\rho xy}{1-\rho^2} - x^2\right]}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{x^2+y^2-2\rho xy-(1-\rho^2)x^2}{1-\rho^2}}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{\rho^2x^2+y^2-2\rho xy}{1-\rho^2}}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{(y-\rho x)^2}{1-\rho^2}}$$

so $Y|X \sim N(\rho x, 1-\rho^2)$

Football

$$W = \begin{cases} 1 & \text{win} \\ 0 & \text{lose} \end{cases} \qquad H = \begin{cases} 1 & \text{home} \\ 0 & \text{away} \end{cases}$$

The problem says

$$p(H=1) = p(H=0) = \frac{1}{2} \quad (\text{"half its games at home"})$$

$$p(W=1 \mid H=1) = 0.7$$

$$p(W=1 \mid H=0) = 0.3$$

We want $p(H=1 \mid W=1)$.

$$\text{Bayes} \implies \quad p(H=1 \mid W=1) = \frac{p(W=1 \mid H=1)\, p(H=1)}{p(W=1)}$$

$$= \frac{0.7 \cdot 0.5}{p(W=1)}$$

What's the marginal prob of winning?

$$p(W=1) = \frac{1}{2} 0.7 + \frac{1}{2} 0.4 = 0.55$$

so $$p(H=1 \mid W=1) = \frac{0.7 \cdot 0.5}{0.55} = \frac{7}{11}$$

HIV

we know that

$$P(\theta=1) = p \qquad P(Y=1\,|\,\theta=0) = q_0$$
$$P(Y=1\,|\,\theta=1) = q_1$$

we are asked for $P(\theta=1\,|\,Y=1)$

we know

① $$P(\theta=1\,|\,Y=1) = \frac{P(Y=1\,|\,\theta=1)\,P(\theta=1)}{P(Y=1)}$$

$$= \frac{q_1 p}{P(Y=1)}$$

② $$P(\theta=0\,|\,Y=1) = \frac{P(Y=1\,|\,\theta=0)\,P(\theta=0)}{P(Y=1)}$$

$$= \frac{q_0(1-p)}{P(Y=1)}$$

③ $$P(\theta=1\,|\,Y=1) + P(\theta=0\,|\,Y=1) = 1$$

Combine ①–③ gives $\frac{q_1 p}{P(Y=1)} + \frac{q_0(1-p)}{P(Y=1)} = 1 \Rightarrow P(Y=1) = q_1 p + q_0(1-p)$.

So plugging this back into ① gives

$$P(\theta=1\,|\,Y=1) = \frac{q_1 p}{q_1 p + q_0(1-p)}$$

HIV (part 2)

We know

(1) $$p(\theta=1|Y=0) = \frac{p(Y=0|\theta=1)\,p(\theta=1)}{p(Y=0)}$$

$$= \frac{(1-\varepsilon_1)\,p}{p(Y=0)}$$

(2) $$p(\theta=0|Y=0) = \frac{p(Y=0|\theta=0)\,p(\theta=0)}{p(Y=0)}$$

$$= \frac{(1-\varepsilon_0)(1-p)}{p(Y=0)}$$

(3) (1) + (2) = 1. which implies

$$p(Y=0) = (1-\varepsilon_1)\,p + (1-\varepsilon_0)(1-p)$$

Therefore, plugging in (1) gives

$$p(\theta=1|Y=0) = \frac{(1-\varepsilon_1)\,p}{(1-\varepsilon_1)p + (1-\varepsilon_0)(1-p)}$$

Robins example

Y = # true birds X = # observed birds

$p(Y=y) = \frac{1}{20}$ for $y \in \{0,1,\dots,19\}$

$X|Y \sim \text{Binomial}(Y, 0.2)$

(1) We are asked to compute

$$P(Y=0|X=0) = \frac{P(X=0|Y=0)\,P(Y=0)}{P(X=0)} \quad \text{(Bayes rule)}$$

$$= \frac{1 \cdot \frac{1}{20}}{P(X=0)}$$

The marginal probability of observing no birds is

$$P(X=0) = \sum_{y=0}^{19} P(X=0, Y=y) = \sum_{y=0}^{19} P(X=0|Y=y)P(Y=y)$$

$$= \sum_{y=0}^{19} \binom{y}{0} 0.2^0 (1-0.2)^y \cdot \frac{1}{20}$$

$$= \frac{1}{20} \sum_{y=0}^{19} 0.8^y$$

So $P(Y=0|X=0) = \frac{1}{\sum_{y=0}^{19} 0.8^y} = 0.202$ (using R)

(2) ~~That would~~ If the prior for Y extened to $Y=100$ then this would decrease the prior, and thus the posterior, probability that $X=0$.

(3) If $0.2 \to 0.9$ then we are less likely to miss birds & so $P(Y=0|X=0)$ would increase.

Derivations for one-parameter models

Beta - binomial

Model: $Y|\theta \sim \text{Binomial}(n, \theta)$
$\theta \sim \text{Beta}(a, b)$

The likelihood is $f(y|\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$
The prior is $f(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$

Combining these gives the posterior

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)}$$
$$= \left[\binom{n}{y}\theta^{y}(1-\theta)^{n-y}\right]\left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\right]\frac{1}{f(y)}$$
$$= \left[\binom{n}{y}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{1}{f(y)}\right]\theta^{y+a-1}(1-\theta)^{n-y+b-1}$$

"proportional to" $\longrightarrow$

$$\propto \theta^{(y+a)-1}(1-\theta)^{(n-y+b)-1}$$

This has the form of a Beta$(y+a, n-y+b)$ PDF, therefor $\boxed{\theta|y \sim \text{Beta}(y+a, n-y+b)}$

Poisson - Gamma

First ~~o~~ consider a single observation

$Y|\lambda \sim \text{Poisson}(N\lambda)$

and prior $\lambda \sim \text{Gamma}(a,b)$

Likelihood: $f(y|\lambda) = \dfrac{e^{-N\lambda}(N\lambda)^y}{y!}$

Prior : $f(\lambda) = \dfrac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$

<u>Posterior</u> $f(\lambda|y) = \dfrac{f(y|\lambda) f(\lambda)}{f(y)}$

$$= \left\{ \frac{e^{-N\lambda}(N\lambda)^y}{y!} \right\} \left\{ \left[\frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \right] \right\} \frac{1}{f(y)}$$

$$= \left[\frac{N^y}{y!} \frac{b^a}{\Gamma(a)} \frac{1}{f(y)} \right] \lambda^{(y+a)-1} e^{-(N+b)\lambda}$$

$$= C \lambda^{(y+a)-1} e^{-(N+b)\lambda}$$

This is a Gamma$(y+a, N+b)$ PDF, therefore we must have $C = \dfrac{(N+b)^{(y+a)}}{\Gamma(y+a)}$ for it to integrate to one.

For our purposes we can forget that & just say that since $f(\lambda|y) \propto \lambda^{(y+a)-1} e^{-(N+b)\lambda}$ then $\boxed{\lambda|y \sim \text{Gamma}(y+a, N+b)}$

Poisson - Gamma (2)

Now say $Y_1, \ldots, Y_m \overset{iid}{\sim} \text{Poisson}(N\lambda)$. Then

$$f(Y_1, \ldots, Y_m \mid \lambda) = f(Y_1 \mid \lambda) \cdot \ldots \cdot f(Y_m \mid \lambda)$$

since the observations are independent. This gives

$$f(\lambda \mid Y_1, \ldots, Y_m) = \frac{f(Y_1 \mid \lambda) \cdot \ldots \cdot f(Y_m \mid \lambda) f(\lambda)}{f(Y_1, \ldots, Y_m)}$$

$$= \frac{\frac{e^{-N\lambda}(N\lambda)^{Y_1}}{Y_1!} \cdot \ldots \cdot \frac{e^{-N\lambda}(N\lambda)^{Y_m}}{Y_m!} \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}}{f(Y_1, \ldots, Y_m)}$$

$$\propto e^{-mN\lambda} \lambda^{Y_1 + \ldots + Y_m} \lambda^{a-1} e^{-b\lambda}$$

$$\propto e^{-(mN+b)\lambda} \lambda^{(Y_1 + \ldots + Y_m + a) - 1}$$

So $\lambda \mid Y_1, \ldots, Y_m \sim \text{Gamma}(Y_1 + \ldots + Y_m + a, \; mN + b)$

Normal - Normal

$Y_1, \ldots, Y_n | \mu \overset{iid}{\sim} N(\mu, \sigma^2)$ & $\mu \sim N(\theta, \frac{\sigma^2}{m})$

where σ^2 is known.

independence

$$p(\mu | Y_1, \ldots, Y_n) \propto p(Y_1|\mu) \cdot \ldots \cdot p(Y_n|\mu)\, p(\mu)$$

$$\propto e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\mu)^2}\, e^{-\frac{m}{2\sigma^2}(\mu-\theta)^2}$$

$$\propto e^{-\frac{1}{2\sigma^2}\left[\Sigma(Y_i^2 - 2Y_i\mu + \mu^2) + m(\mu^2 - 2\mu\theta + \theta^2)\right]}$$

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$

$n\bar{Y} = \sum_{i=1}^{n} Y_i$

$$\propto e^{-\frac{1}{2\sigma^2}\left[-2n\bar{Y}\mu + n\mu^2 + m\mu^2 - 2m\theta\mu\right]}$$

$$\propto e^{-\frac{1}{2\sigma^2}\left[-2(n\bar{Y} + m\theta)\mu + (n+m)\mu^2\right]}$$

$$\propto e^{-\frac{n+m}{2\sigma^2}\left[-2\frac{n\bar{Y}+m\theta}{n+m}\mu + \mu^2\right]}$$

$$\propto e^{-\frac{1}{2\left[\frac{\sigma^2}{n+m}\right]}\left(\mu - \frac{n\bar{Y}+m\theta}{n+m}\right)^2}$$

So $\mu | Y_1, \ldots, Y_n \sim N\left(\frac{n\bar{Y}+m\theta}{n+m}, \frac{\sigma^2}{n+m}\right)$

Normal - Inverse gamma

$$Y_1, \ldots, Y_n \mid \sigma^2 \sim N(\mu, \sigma^2)$$
$$\sigma^2 \sim \text{Inv Gamma}(a, b)$$

μ is known

$$p(\sigma^2 \mid Y_1, \ldots, Y_n) \propto p(Y_1 \mid \sigma^2) \cdot \ldots \cdot p(Y_n \mid \sigma^2)\, p(\sigma^2$$

$$\propto \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2} \Sigma (Y_i - \mu)^2} (\sigma^2)^{-a-1} e^{-\frac{b}{\sigma^2}}$$

$SSE = \Sigma (Y_i - \mu)^2$

$$\propto \frac{1}{(\sigma^2)^{\frac{n}{2}+a+1}} e^{-\frac{1}{\sigma^2}\left(\frac{SSE}{2} + b\right)}$$

So $\sigma^2 \mid Y_1, \ldots Y_n \sim \text{Inv Gamma}\left(\frac{n}{2} + a, \frac{SSE}{2} + b\right)$

Jeffreys - Binomial

The likelihood is $p(y|\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$

$$\log p(y|\theta) = \log\binom{n}{y} + y\log\theta + (n-y)\log(1-\theta)$$

$$\frac{\partial \log p(y|\theta)}{\partial\theta} = \frac{y}{\theta} - \frac{n-y}{1-\theta}$$

$$\frac{\partial^2 \log p(y|\theta)}{\partial\theta^2} = -\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2}$$

$$I(\theta) = -E_{y|\theta}\left(-\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2}\right) \qquad E(y) = n\theta$$

$$= \frac{E(y)}{\theta^2} + \frac{n-E(y)}{(1-\theta)^2}$$

$$= \frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2}$$

$$= \frac{n}{\theta} + \frac{n}{1-\theta}$$

$$= n\,\frac{1-\theta+\theta}{\theta(1-\theta)} = \frac{n}{\theta(1-\theta)}$$

So the Jeffreys prior is

$$p(\theta) = \sqrt{I(\theta)} = \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} = \theta^{\frac{1}{2}-1}(1-\theta)^{\frac{1}{2}-1}$$

& $\theta \sim \text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)$

Jeffreys - Normal

Say $y|\mu \sim N(\mu, \sigma^2)$ with σ^2 known.

The log likelihood is

$$\ell(\mu) = \frac{1}{2}\log(2\pi\sigma) - \frac{1}{2\sigma^2}(y-\mu)^2$$

$$\ell'(\mu) = \frac{1}{\sigma^2}(y-\mu)$$

$$\ell''(\mu) = -\frac{1}{\sigma^2}$$

so $I(\mu) = -E(\ell''(\mu)) = E\left(\frac{1}{\sigma^2}\right) = \frac{1}{\sigma^2}$

then the Jeff prior is

$$p(\mu) = \sqrt{I(\mu)} = \frac{1}{\sigma}$$

This is not a function of μ! so

the prior is $p(\mu) \propto 1$.

Derivations for (5) MCMC Sampling

$$\boxed{\text{Marginal posterior of } \mu}$$

Model $Y_1, \dots, Y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2)$ $\quad p(\mu) \propto 1 \quad \sigma^2 \sim \text{InvGamma}(a, b)$

$$\boxed{SSE = \Sigma (y_i - \mu)^2}$$

$$p(\mu \mid y) = \int p(\mu, \sigma^2 \mid y)\, d\sigma^2$$

$$\propto \int p(y \mid \mu, \sigma^2)\, p(\mu)\, p(\sigma^2)\, d\sigma^2$$

$$\propto \int \left[(\sigma^2)^{-\frac{n}{2}} e^{-\frac{SSE}{2\sigma^2}} \right] \left[(\sigma^2)^{-a-1} e^{-\frac{b}{\sigma^2}} \right] d\sigma^2$$

$$\propto \int (\sigma^2)^{-\left(\frac{n}{2}+a\right)-1} e^{-\frac{SSE/2 + b}{\cancel{2}\sigma}}\, d\sigma^2$$

$$\boxed{\begin{aligned} A &= \tfrac{n}{2} + a \\ B &= SSE/2 + b \end{aligned}}$$

$$\propto \int (\sigma^2)^{-A-1} e^{-\frac{B}{\sigma^2}}\, d\sigma^2$$

$$\propto \frac{\cancel{\Gamma(A)}}{B^A} \int \frac{B^A}{\Gamma(A)} (\sigma^2)^{A-1} e^{-\frac{B}{\sigma^2}}\, d\sigma^2$$

PDF of InvG(A, B)

$$\propto B^{-A}$$

$$\propto \left[SSE/2 + b \right]^{-A}$$

$$\boxed{\begin{aligned} \tilde{y} &= \tfrac{1}{n} \Sigma y_i^2 \\ \bar{y} &= \tfrac{1}{n} \Sigma y_i \end{aligned}}$$

$$\propto \left[\Sigma (y_i - \mu)^2 + 2b \right]^{-A}$$

$$\propto \left[n\tilde{y} - 2n\bar{y}\mu + n\mu^2 + 2b \right]^{-A}$$

$$\propto \left[\tilde{y} + 2b/n - 2\bar{y}\mu + \mu^2 \right]^{-A}$$

$$\propto \left[\tilde{y} + \frac{2b}{n} - \bar{y}^2 + \bar{y}^2 - 2\bar{y}\mu + \mu^2 \right]^{-A}$$

$$\propto \left[s^2 + \frac{2b}{n} + (\mu - \bar{y})^2 \right]^{-A}$$

$$\boxed{s^2 = \tilde{y_i} - \bar{y}^2}$$

sample variance

$$\propto \left[1 + \frac{(\mu - \bar{y})^2}{s^2 + 2b/n} \right]^{-\left(\frac{n}{2} + a\right)}$$

$$\propto \left[1 + \frac{1}{\nu} \frac{(\mu - \bar{y})^2}{\left[s^2 + 2b/n\right]/\nu} \right]^{-\left(\frac{\nu+1}{2}\right)}$$

$$\boxed{\nu = n - 1 + 2a}$$

so $\mu \mid y \sim t$ with

location : $\bar{Y}$

scale : $\dfrac{s^2 + 2b/n}{n-1+2a}$

df : $n-1+2a$

$a = b \approx 0 \Rightarrow$ scale $= \dfrac{s^2}{n-1}$ & df $= n-1$

$a, b \to \infty$ & $\frac{b}{a} \to \sigma^2$ scale $= \dfrac{\sigma^2}{n}$ df $= \infty$

Interesting special cases

Full conditional distributions for the t-test

$y_i \overset{iid}{\sim} N(\mu, \sigma^2)$

$\mu \sim N(\mu_0, \sigma_0^2)$

$\sigma^2 \sim \text{InvGamma}(a, b)$

$\boxed{\sigma^2 | \text{rest}}$

$$p(\sigma^2 | y, \mu) = \frac{p(y|\sigma^2, \mu)\, p(\sigma^2)\, \cancel{p(\mu)}}{\cancel{p(y)}}$$

$$\propto \left[\prod_{i=1}^{n} (\sigma^2)^{-\frac{1}{2}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}\right]\left[(\sigma^2)^{-a-1} e^{-\frac{b}{\sigma^2}}\right]$$

$SSE = \sum_{i=1}^{n} (y_i - \mu)^2$

$$\propto \left[(\sigma^2)^{-\frac{n}{2}} e^{-\frac{SSE}{2\sigma^2}}\right]\left[(\sigma^2)^{-a-1} e^{-\frac{b}{\sigma^2}}\right]$$

$$\propto (\sigma^2)^{-\left(\frac{n}{2}+a\right)-1} e^{-\frac{SSE/2 + b}{\sigma^2}}$$

so $\sigma^2 | \text{rest} \sim \text{InvGamma}\left(\frac{n}{2}+a, \frac{SSE}{2}+b\right)$

$$\boxed{\mu \mid \text{rest}}$$

$$f(\mu \mid Y, \sigma^2) \propto f(Y \mid \mu, \sigma^2) f(\mu)$$

$$\propto e^{-\frac{\Sigma (Y_i - \mu)^2}{2\sigma^2}} \, e^{-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}}$$

$$\propto e^{-\frac{1}{2}\left[\frac{\Sigma Y_i}{\sigma^2} - 2\left[\frac{\Sigma Y_i}{\sigma^2}\right]\mu + \frac{n}{\sigma^2}\mu^2 + \frac{\mu_0^2}{\sigma_0^2} - 2\frac{\mu_0}{\sigma_0^2}\mu + \frac{1}{\sigma_0^2}\mu^2\right]}$$

$n\bar{Y} = n\frac{1}{n}\Sigma Y_i = \Sigma Y_i$

$$\propto e^{-\frac{1}{2}\left\{-2\left[\frac{n\bar{Y}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right]\mu + \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)\mu^2\right\}}$$

$$A = \frac{n\bar{Y}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}$$

$$B = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}$$

$$\propto e^{-\frac{1}{2}\left[-2A\mu + B\mu^2\right]}$$

$$\propto e^{-\frac{B}{2}\left[-2\frac{A}{B}\mu + \mu^2\right]}$$

$$\propto e^{-\frac{B}{2}\left(\mu - \frac{A}{B}\right)^2}$$

So $\mu \mid \text{rest} \sim N\left(\frac{A}{B}, \frac{1}{B}\right)^2$

Full conditional distributions for simple linear reg

Model : $Y_i \overset{\text{indep}}{\sim} N(\alpha+\beta x_i, \sigma^2)$

Prior : $\alpha, \beta \overset{\text{iid}}{\sim} N(\mu_0, \sigma_0^2)$

$\sigma^2 \sim \text{InvGamma}(a, b)$

$\boxed{\sigma^2 \mid \text{rest}}$

$$p(\sigma^2 \mid y, \alpha, \beta) = \frac{p(y \mid \alpha, \beta, \sigma^2)\, \overbrace{\cancel{p(\alpha)\, p(\beta)}}^{\text{no } \sigma^2}\, p(\sigma^2)}{\cancel{p(y)}}$$

$$\propto p(y \mid \alpha, \beta, \sigma^2)\, p(\sigma^2)$$

$$\propto \left[\prod_{i=1}^{n} (\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{\sigma^2} \frac{(y_i-\alpha-x_i\beta)^2}{2}}\right]\left[(\sigma^2)^{-a-1} e^{-\frac{b}{\sigma^2}}\right]$$

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

$$\propto \left((\sigma^2)^{-\frac{n}{2}} e^{-\frac{SSE}{\sigma^2}}\right)\left((\sigma^2)^{-a-1} e^{-\frac{b}{\sigma^2}}\right)$$

$$\propto (\sigma^2)^{-\left(\frac{n}{2}+a\right)-1} e^{-\frac{\frac{SSE}{2}+b}{\sigma^2}}$$

So $\sigma^2 \mid \text{rest} \sim \text{InvGamma}\left(\frac{n}{2}+a, \frac{SSE}{2}+b\right)$

$\boxed{\alpha \mid \text{rest}}$

$$p(\alpha \mid Y, \beta, \sigma^2) \propto \frac{p(Y \mid \alpha, \beta, \sigma^2)\, p(\alpha)\, \cancel{p(\beta)}\, \cancel{p(\sigma^2)}}{\cancel{p(Y)}}$$

$$\propto e^{-\frac{1}{2\sigma^2}\sum (y_i - \alpha - x_i\beta)^2}\; e^{-\frac{(\alpha-\mu_0)^2}{2\sigma_0^2}}$$

$z_i = y_i - x_i\beta$

$$\propto e^{-\frac{1}{2\sigma^2}\sum (z_i - \alpha)^2}\; e^{-\frac{(\alpha-\mu_0)^2}{2\sigma_0^2}}$$

So this has all the same pieces as if we used the model

$$z_i \overset{iid}{\sim} N(\alpha, \sigma^2)$$
$$\alpha \sim N(\mu_0, \sigma_0^2).$$

So we can simple use the Normal/Normal posterior & conclude that

$$\alpha \mid \text{rest} \sim N\left(\frac{M}{V}, \frac{1}{V}\right)$$

where $M = n\bar{z}/\sigma^2 + \mu_0/\sigma_0^2$

$V = n/\sigma^2 + 1/\sigma_0^2$

$\boxed{\beta \mid \text{rest}}$

$$p(\beta \mid y) \propto p(y \mid \beta, \alpha, \sigma^2)\, p(\beta)$$

$$\propto e^{-\frac{\sum_{i=1}^{n}(y_i - \alpha - x_i\beta)^2}{2\sigma^2}} \; e^{-\frac{(\beta-\mu_0)^2}{2\sigma_0^2}} \qquad r_i = y_i - \alpha$$

$$\propto e^{-\frac{\sum r_i^2 - 2 r_i^2 x_i \beta + x_i^2\beta^2}{2\sigma^2}} \; e^{-\frac{\mu_0^2 - 2\mu_0\beta + \beta^2}{2\sigma_0^2}}$$

$$\propto e^{-\frac{1}{2}\left[-2\left[\frac{\sum r_i x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right]\beta + \left[\frac{\sum x_i}{\sigma^2} + \frac{1}{\sigma_0^2}\right]\beta^2\right]}$$

$$\propto e^{-\frac{1}{2}\left[-2M\beta + V\beta^2\right]}$$

$$M = \frac{\sum r_i x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}$$

$$V = \frac{\sum x_i}{\sigma^2} + \frac{1}{\sigma_0^2}$$

$$\propto e^{-\frac{V}{2}\left[-\frac{M}{V}\beta + \beta^2\right]}$$

$$\propto e^{-\frac{V}{2}\left[\beta - \frac{M}{V}\right]^2}$$

So $\beta \mid \text{rest} \sim N\left(\frac{M}{V}, \frac{1}{V}\right)$

Bayes linear regression with a flat prior

Likelihood: $Y \sim N(X\beta, \sigma^2 I)$ $\quad\quad$ σ^2 known

Prior : $p(\beta) = 1$

$$p(\beta|y) \propto p(y|\beta)\, p(\beta)$$

$$\propto e^{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)}$$

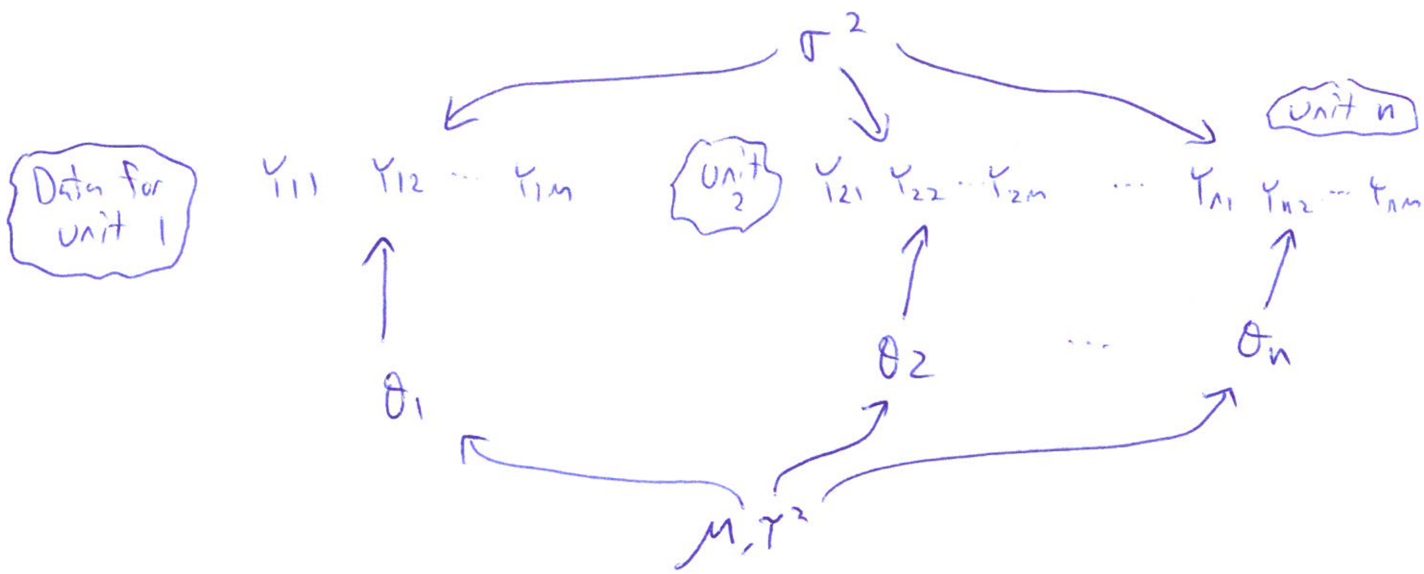
$$\propto e^{-\frac{1}{2\sigma^2}(y' - 2y'X\beta + \beta'X'X\beta)}$$

$$\propto e^{-\frac{1}{2\sigma^2}(-2\overbrace{y'X(X'X)^{-1}}^{\hat\beta}X'X\beta + \beta X'X\beta)}$$

$$\propto e^{-\frac{1}{2\sigma^2}(-2\hat\beta(X'X)\beta + \beta X'X\beta)}$$

So $\beta|Y \sim N(\hat\beta, \sigma^2(X'X)^{-1})$

DAG for a one-way random effect model



How do data from unit 1 affect θ_2?

Derivation of the Bayes factor for the Beta-binomial model

Likelihood: $Y \mid \theta \sim \text{Binomial}(n, \theta)$

Models: $M_1: \theta = \theta_0$ $\qquad$ $M_2: \theta \neq \theta_0$ & $\theta \sim \text{Beta}(a, b)$

$$BF = \frac{p(Y \mid M_2)}{p(Y \mid M_1)}$$

$$p(y \mid M_1) = p(Y \mid \theta = \theta_0) = \binom{n}{y} \theta_0^{\,y} (1-\theta_0)^{n-y}$$

$$p(y \mid M_2) = \int_0^1 p(Y, \theta)\, d\theta = \int p(Y \mid \theta)\, p(\theta)\, d\theta$$

$$= \int_0^1 \left[\binom{n}{y} \theta^{y} (1-\theta)^{n-y} \right] \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} \right] d\theta$$

$$= \binom{n}{y} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int \theta^{(y+a)-1} (1-\theta)^{(n-y+b)-1}\, d\theta$$

Looks like Beta(A, B) where $A = y + a$, $B = n - y + b$

$$= \binom{n}{y} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int \theta^{A-1} (1-\theta)^{B-1}\, d\theta$$

$$= \binom{n}{y} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(A)\Gamma(B)}{\Gamma(A+B)} \overbrace{\int_0^1 \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} \theta^{A-1} (1-\theta)^{B-1}\, d\theta}^{1}$$

$$= \binom{n}{y} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(A)\Gamma(B)}{\Gamma(A+B)}$$

$$\text{So } BF = \frac{\binom{n}{y} \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \dfrac{\Gamma(A)\Gamma(B)}{\Gamma(A+B)}}{\cancel{\binom{n}{y}} \theta_0^{\,y} (1-\theta_0)^{n-y}} = \frac{\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \dfrac{\Gamma(y+a)\Gamma(n-y+b)}{\Gamma(n+a+b)}}{\theta_0^{\,y} (1-\theta_0)^{n-y}}$$

This is messy! What if $a = b = 1$ & $\theta_0 = \frac{1}{2}$? Recall $\Gamma(1) = 1$ & $\Gamma(x) = (x-1)\Gamma(x-1)$
& $\Gamma(n) = (n-1)!$ if n is a counting number.

DIC for one-way random effects

$y_{ij} = \mu_j + \varepsilon_{ij}$ where y_{ij} = obs i for subj j

μ_j = mean for subj j

$\varepsilon_{ij} \overset{iid}{\sim} N(0, \tau_\varepsilon)$

$\mu_j \overset{iid}{\sim} N(0, \tau_\mu)$

inverse variances, assumed known

we've seen $\mu_j | y \sim N(E_j, P_j)$ where $E_j = \dfrac{n\tau_\varepsilon}{n\tau_\varepsilon + \tau_\mu} \bar{y}_j$

$P_j = n\tau_\varepsilon + \tau_\mu$

The deviance is $D(y,\mu) = \tau_\varepsilon \sum_{ij} (y_{ij} - \mu_j)^2$, so

$$\hat{D} = D(y, \mu = \hat{\mu}) = \tau_\varepsilon \sum (y_{ij} - E_j)^2$$

$$= \tau_\varepsilon \sum y_{ij} - 2\tau_\varepsilon \sum n \bar{y}_j E_j + \tau_\varepsilon n \sum E_j.$$

$$\bar{D} = E_{\mu|y}\left(D(y,\mu)\right) = \tau_\varepsilon E\left(\sum_{ij} y_{ij}^2 - 2y_{ij}\mu_j + \mu_j^2\right)$$

$$= \tau_\varepsilon\left(\sum_{ij} y_{ij}^2 - 2y_{ij}E_j + E_j^2 + 1/P_j\right)$$

so $p_D = \bar{D} - \hat{D} = \sum n\tau \dfrac{1}{P_j} = p \cdot \dfrac{n\tau_\varepsilon}{n\tau_\varepsilon + \tau_\mu}$

☆ $0 < p_D < p \quad \forall\ n, \tau_\varepsilon, \tau_\mu$

☆ $\tau_\mu = 0 \iff$ flat prior $\Rightarrow p_D = p$

☆ $\tau_\mu = \infty \iff$ tight prior $\Rightarrow p_D = 0$

DIC for linear regression

$$y \sim N(X\beta, I)$$
$$\beta \sim N(0, c(X'X)^{-1})$$

Then $\beta|y \sim N(\psi\hat{\beta}, \psi(X'X)^{-1})$ where $\psi = \frac{c}{c+1}$ & $\hat{\beta} = (X'X)^{-1}X'y$

$\boxed{\text{Computing DIC}}$

(ignoring a constant)

$$\bar{D} = E_{\beta|y}(\text{deviance}) = E(y-X\beta)'(y-X\beta)$$
$$= y'y - 2y'E(\beta|y) + E(\beta'X'X\beta)$$
$$= y'y - 2\psi X'X\hat{\beta} + E(\beta)'X'XE(\beta) + \text{trace}(X'X\,\text{cov}(\beta))$$
$$= y'y - \psi y'X\hat{\beta} + \psi^2\hat{\beta}'X'X\hat{\beta} + \text{trace}(X'X\,\psi(X'X)^{-1})$$
$$= y'y - \psi y'X\hat{\beta} + \psi^2\hat{\beta}X'X\hat{\beta} + \psi p$$

ignoring the same constant

$$\hat{D} = D(y,\bar{\beta}) = (y - XE(\beta))'(y - XE(\beta))$$
$$= y'y - 2y'X(\psi\hat{\beta}) + \psi^2\hat{\beta}'X'X\hat{\beta}$$

$\frac{c}{1+c}$ is the same Zellner's shrinkage factor as before

so $p_D = \bar{D} - \hat{D} = \psi p = \frac{c}{1+c}p$

$\text{DIC} = \bar{D} + p_D$

$\boxed{\text{Interpreting } p_D}$

$c=\infty$ means β has flat prior $\Rightarrow p_D = p$ & all parameters are "free"

$c=0$ means $\hat{\beta}$ is complete shrunk to zero $\Rightarrow p_D = 0$

Bayes rule for squared error loss

the loss is $l(\theta, \hat\theta) = (\theta - \hat\theta)^2$.

The expected loss is

$\bar\theta = E(\theta|y)$
= posterior mean

$$E_{\theta|y}\left(l(\theta,\hat\theta)\right) = E\left[(\theta-\hat\theta)^2\right]$$
$$= E\left[(\theta - \bar\theta + \bar\theta - \hat\theta)^2\right]$$
$$= E\left[(\theta-\bar\theta)^2 + 2(\theta-\bar\theta)(\bar\theta-\hat\theta) + (\bar\theta-\hat\theta)^2\right]$$
$$= E\left\{(\theta-\bar\theta)^2\right\} + 2E\left[(\theta-\bar\theta)(\bar\theta-\hat\theta)\right] + E\left[(\bar\theta-\hat\theta)^2\right]$$
$$= V(\theta|y) + 2(\bar\theta-\hat\theta)\cancelto{0}{E\left[(\theta-\bar\theta)\right]} + (\bar\theta-\hat\theta)^2$$
$$= V(\theta|y) + (\bar\theta-\hat\theta)^2$$

All expectations are with respect to θ's posterior distribution, and use $\hat\theta$ & $\bar\theta$ are treated as constants

Doesn't depend on $\hat\theta$

Minimized by setting $\hat\theta = \bar\theta$

Therefore $\hat\theta = \bar\theta$ minimizes expected loss & so the posterior mean is the Bayes rule.

Bayes rule for Hypothesis testing (+ Classification)

Say $\theta = 0$ if the null hypothesis is true
$\theta = 1$ " alternative "

$\hat{\theta} = 0$ if we do not reject the null hypothesis
$\hat{\theta} = 1$ if we reject the null

p_0 is the posterior probability of the null $= P(\theta = 0 \mid x)$
$1 - p_0$ " alternative

The loss function is

$$\ell(\theta, \hat{\theta}) = \begin{cases} \lambda_1 & \theta = 0 \;\&\; \hat{\theta} = 1 \quad \text{Type I error} \\ \lambda_2 & \theta = 1 \;\&\; \hat{\theta} = 0 \quad \text{Type II error} \end{cases}$$

If $\hat{\theta} = 1$ then the expected loss is $p_0 \lambda_1$
$\hat{\theta} = 0$ " $(1 - p_0)\lambda_2$

The Bayes rule is to reject the null hypothesis if $p_0 \lambda_1 < (1 - p_0)\lambda_2 \iff p_0 < \dfrac{\lambda_2}{\lambda_1 + \lambda_2}$.