1.1 Classical and frequentist probability

$P(X=4)$ $\quad$ $P(\text{drop packet})$

$P(\text{die} = \text{fair})$ $\quad$ $P(Y_1 > Y_2)$

$P(\text{rain})$

Classical - equally likely

$P(X=4) = 1/6$ for a die

$P(X_1+X_2=4) = \frac{3}{36} = \frac{1}{12} = \frac{\#\text{ possible}}{\#\text{ equally likely outcomes}}$

Frequentist - relative frequency in infinite sequence

$P(X=4) = 1/6$

$P(\text{drop packets}) = \frac{1}{10{,}000}$ (lose 1 in 10,000)

$P(\text{rain})$ requires an infinite sequence

$P(\text{fair}) = \{0,1\}$ (fair or not regardless of all of the rolls)

Bayesian - personal perspective

1.2 Bayesian probability and coherence

(1) If rain, win \$4 , If no rain, lose \$1 $\quad$ odds 4:1

(2) If rain lose \$4 , If no rain win \$1

If this is a fair game: $P(\text{rain}) = \frac{1}{1+4} = \frac{1}{5}$

Expected return: $4\left(\frac{1}{5}\right) - 1\left(\frac{4}{5}\right) = 0$ $\quad$ (1)

$1\left(\frac{4}{5}\right) - 4\left(\frac{1}{5}\right) = 0$ $\quad$ (2)

Coherence: Must follow all rules of probability

2.1 Conditional probability

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

30 students
9 females
12 computer science of which 4 female

$P(F) = \frac{9}{30} = \frac{3}{10}$, $P(CS) = \frac{12}{30} = \frac{2}{5}$

$P(F \cap CS) = \frac{4}{30} = \frac{2}{15}$

$$P(F|CS) = \frac{P(F \cap CS)}{P(CS)} = \frac{\frac{2}{15}}{\frac{2}{5}} = \frac{1}{3}$$

$$P(F|CS^c) = \frac{P(F \cap CS^c)}{P(CS^c)} = \frac{5/30}{18/30} = 5/18$$

Independence

$P(A|B) = P(A) \Rightarrow P(A \cap B) = P(A)P(B)$

$P(F|CS) \neq P(F) \Rightarrow$ not independent

2.2 Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} = \frac{P(A \cap B)}{P(B)}$$

$$P(CS|F) = \frac{P(F|CS)\,P(CS)}{P(F|CS)\,P(CS) + P(F|CS^c)\,P(CS^c)}$$

$$= \frac{(\frac{1}{3})(\frac{2}{5})}{(\frac{1}{3})(\frac{2}{5}) + (\frac{5}{18})(\frac{3}{5})} = \frac{4}{9}$$

$$P(CS|F) = \frac{P(CS \cap F)}{P(F)} = \frac{4/30}{9/30} = \frac{4}{9}$$

$P(+|HIV) = .977$ $\quad P(HIV) = .0026$

$P(-|\text{no } HIV) = .926$

$$P(HIV|+) = \frac{P(+|HIV)\,P(HIV)}{P(+|HIV)P(HIV) + P(+|\text{no } HIV)\,P(\text{no } HIV)}$$

$$= \frac{(.977)(.0026)}{(.977)(.0026) + (1-.926)(1-.0026)} = .033$$

3.1 Bernoulli and binomial distributions

Bernoulli

$X \sim B(p)$ $\quad P(X=1) = p$
$\quad\quad\quad\quad\;\; P(X=0) = 1-p$

$$f(X=x|p) = f(x|p)$$
$$= p^x(1-p)^{1-x} I_{\{x\in\{0,1\}\}}(x)$$

Expected value

$$E[X] = \sum_x x P(X=x) = (1)p + (0)(1-p) = p$$

$$Var(X) = p(1-p)$$

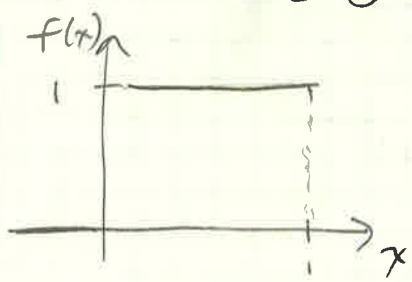
Binomial

$$X \sim Bin(n,p)$$
$$P(X=x|p) = f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \quad \text{for } x \in \{0,1,\dots,n\}$$

$$E[X] = np \quad , \quad Var(X) = np(1-p)$$

3.2 Uniform distribution

$$X \sim U[0,1]$$
$$f(x) = \begin{cases} 1 & \text{if } x\in[0,1] \\ 0 & \text{otherwise} \end{cases} = I_{\{0\le x\le 1\}}(x)$$



$$P(0 < X < \tfrac{1}{2}) = \int_0^{1/2} f(x)dx = \int_0^{1/2} dx = \frac{1}{2}$$
$$P(X=\tfrac{1}{2}) = 0$$
$$\int_{-\infty}^{\infty} f(x)dx = 1 \quad , \quad f(x) \ge 0$$

$$E[X] = \int_{-\infty}^{\infty} x f(x)\,dx$$
$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx$$
$$E[cX] = c\,E(X) \quad , \quad E[X+Y] = E[X] + E[Y]$$
$$\text{if } X \perp Y \Rightarrow E[XY] = E[X]\,E[Y]$$

3.3 Exponential and normal distributions

Exponential

$$X \sim Exp(\lambda)$$

$$f(x|\lambda) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0$$

$$E[X] = \frac{1}{\lambda}, \quad Var(X) = \frac{1}{\lambda^2}$$

Uniform $X \sim U[\theta_1, \theta_2]$

$$f(x|\theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1} I_{\{\theta_1 \leq x \leq \theta_2\}}$$

Normal

$$X \sim N(\mu, \sigma^2)$$

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(x-\mu)^2\right)$$

$$E[X] = \mu, \quad Var(X) = \sigma^2$$

4.1 Confidence Intervals

There is a fixed value of p

Flip a coin 44H, 56T. Is this a fair coin?

$X_i \sim B(p)$

By CLT $\sum_{i=1}^{100} X_i \mathrel{\dot\sim} N(100p,\ 100p(1-p))$

$100p - 1.96\sqrt{100p(1-p)}$ and $100p + 1.96\sqrt{100p(1-p)}$

Observe $\sum x_i = 44 \Rightarrow \hat{p} = \frac{44}{100} = 0.44$

CI $44 \pm 1.96\sqrt{44(.56)} = 44 \pm 9.7 = [34.3, 53.7)$

$\Leftrightarrow$ 95% confident $p \in (.343, .537)$

If we repeat this experiment an arbitrary number of times, then, on average 95% of the intervals that we make will contain the true p. Thus, the frequentist probability $P(p \in CI) = \{0, 1\}$ since p is a fixed value

4.2 Likelihood function and maximum likelihood

$Y_i \overset{iid}{\sim} B(\theta)$, $P(Y_i = 1) = \theta$

$$P(\underset{\sim}{Y} = \underset{\sim}{y} \mid \theta) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n \mid \theta)$$
$$= P(Y_1 = y_1 \mid \theta) \cdots P(Y_n = y_n \mid \theta)$$
$$= \prod_{i=1}^{n} P(Y_i = y_i \mid \theta) = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i}$$

Likelihood : $L(\theta \mid \underset{\sim}{y}) = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i}$

MLE : $\hat{\theta} = \operatorname{argmax} L(\theta \mid \underset{\sim}{y})$

$$\ell(\theta) = \log L(\theta \mid \underset{\sim}{y}) = \log\left[\prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i}\right]$$
$$= \sum_{i=1}^{n} \log\left[\theta^{y_i}(1-\theta)^{1-y_i}\right] = \sum_{i=1}^{n}\left[y_i \log\theta + (1-y_i)\log(1-\theta)\right]$$
$$= \left(\sum_{i=1}^{n} y_i\right)\log\theta + \left(\sum_{i=1}^{n}(1-y_i)\right)\log(1-\theta)$$

4.3 Computing the MLE

$$\ell'(\theta) = \frac{1}{\theta}\sum y_i - \frac{1}{1-\theta}\sum(1-y_i) \overset{set}{=} 0$$

$$\Rightarrow \frac{\sum y_i}{\hat\theta} = \frac{\sum(1-y_i)}{1-\hat\theta}$$

$$\Rightarrow \hat\theta = \frac{1}{n}\sum y_i = \hat p = \frac{72}{400} = .18$$

Approx CI : $\hat\theta \pm 1.96\sqrt{\frac{\hat\theta(1-\hat\theta)}{n}}$

$\hat\theta \mathrel{\dot\sim} N\left(\theta, \frac{1}{I(\hat\theta)}\right)$ $\quad I(\theta)$ is the Fisher Information

$I(\theta) = \frac{1}{\theta(1-\theta)}$ for a Bernoulli

4.4 Computing the MLE: Examples

$X_i \overset{iid}{\sim} Exp(\lambda)$

$$f(\underset{\sim}{x}|\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda\sum x_i}$$

$$L(\lambda|\underset{\sim}{x}) = \lambda^n e^{-\lambda\sum x_i}$$

$$\ell(\lambda) = n\log\lambda - \lambda\sum x_i$$

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum x_i \overset{set}{=} 0 \quad \Rightarrow \hat\lambda = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$$

$X_i \overset{iid}{\sim} U[0,\theta]$

$$f(\underset{\sim}{x}|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I_{\{0 \le x_i \le \theta\}}$$

$$L(\theta|\underset{\sim}{x}) = \theta^{-n} I_{\{0 \le \min x_i \le \max x_i \le \theta\}}$$

$$L'(\theta) = -n\theta^{-(n+1)} I_{\{0 \le \min x_i \le \max x_i \le \theta\}}$$

$$\hat\theta = \max x_i$$

5.1 Inference Example: Frequentist

$\theta = \{\text{fair}, \text{loaded}\}$

$X \sim Bin(5, ?)$ 5 flips, what is p?

$$f(x|\theta) = \begin{cases} \binom{5}{x}\left(\frac{1}{2}\right)^5 & \text{if } \theta = \text{fair} \\ \binom{5}{x}(.7)^x(.3)^{5-x} & \text{if } \theta = \text{loaded} \end{cases}$$

$$= \binom{5}{x}(.5)^5 I_{\{\theta=\text{fair}\}} + \binom{5}{x}(.7)^x(.3)^{5-x} I_{\{\theta=\text{loaded}\}}$$

let $X=2$

$$f(\theta|X=2) = \begin{cases} 0.3125 & \text{if } \theta = \text{fair} \\ 0.1323 & \text{if } \theta = \text{loaded} \end{cases}$$

MLE $\hat{\theta} = \text{fair}$

$P(\theta = \text{fair} | X=2) = P(\theta=\text{fair}) = \{0,1\}$ since it is a fixed coin

5.2 Inference Example: Bayesian

Prior $P(\theta = \text{loaded}) = 0.6$

$$f(\theta|x) = \frac{f(x|\theta)\,f(\theta)}{\sum_\theta f(x|\theta)\,f(\theta)}$$

$$= \frac{\binom{5}{x}\left[\left(\frac{1}{2}\right)^5(.4) I_{\{\theta=\text{fair}\}} + (.7)^x(.3)^{5-x}(.6) I_{\{\theta=\text{loaded}\}}\right]}{\sum_\theta \binom{5}{x}\left[\left(\frac{1}{2}\right)^5(.4) I_{\{\theta=\text{fair}\}} + (.7)^x(.3)^{5-x}(.6) I_{\{\theta=\text{loaded}\}}\right]}$$

$$= \frac{0.0125\, I_{\{\theta=\text{fair}\}} + 0.0079\, I_{\{\theta=\text{loaded}\}}}{0.0125 + 0.0079}$$

$$= 0.612\, I_{\{\theta=\text{fair}\}} + 0.388\, I_{\{\theta=\text{loaded}\}}$$

$P(\theta=\text{loaded}|X=2) = .388$

$P(\theta=\text{loaded}) = \frac{1}{2} \Rightarrow P(\theta=\text{loaded}|X=2) = .297$ (different priors)

$P(\theta=\text{loaded}) = 0.9 \Rightarrow P(\theta=\text{loaded}|X=2) = .792$

5.3 Continuous version of Bayes' Theorem

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)} = \frac{f(y|\theta)\,f(\theta)}{\int f(y|\theta)\,f(\theta)\,d\theta} = \frac{\text{likelihood} \times \text{prior}}{\text{normalizing constant}}$$

$$\propto \text{likelihood} \times \text{prior}$$

$$\theta \sim U[0,1] \,,\; f(\theta) = I_{\{0 \le \theta \le 1\}}$$

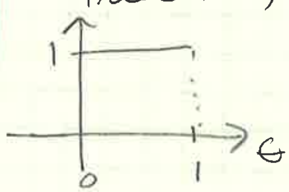
$$f(\theta|Y=1) = \frac{\theta^1(1-\theta)^0 I_{\{0\le\theta\le 1\}}}{\int_{-\infty}^{\infty} \theta^1(1-\theta)^0 I_{\{0\le\theta\le1\}}} = \frac{\theta I_{\{0\le\theta\le1\}}}{\int_0^1 \theta\, d\theta} = 2\theta I_{\{0\le\theta\le1\}}$$

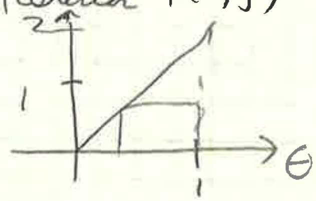
$$f(\theta|y) \propto f(y|\theta)f(\theta) \propto \theta I_{\{0\le\theta\le1\}}$$

$$\Rightarrow f(\theta|Y=1) = 2\theta I_{\{0\le\theta\le1\}}$$

5.4 Posterior Intervals

Prior $f(\theta)$ Posterior $f(\theta,y)$





Prior interval estimates

$$P(0.025 < \theta < 0.975) = .95 \quad , \quad P(\theta > 0.05) = 0.95$$

Posterior interval estimates

$$P(0.025 < \theta < 0.975) = \int_{0.025}^{0.975} 2\theta\, d\theta = .975^2 - .025^2 = .95$$

$$P(\theta > 0.05) = 1 - 0.05^2 = .9975$$

Equal tailed

$$P(\theta < q \,|\, Y=1) = \int_0^q 2\theta\, d\theta = q^2$$

$$P(\sqrt{0.025} < \theta < \sqrt{.975}) = P(0.158 < \theta < 0.987) = 0.95$$

HPD $\quad P(\theta > \sqrt{.05} \,|\, Y=1) = P(\theta > .224 \,|\, Y=1) = 0.95$

6.1 Priors and prior predictive distributions

$P(\theta \le c)$ for all $c \in \mathbb{R}$

$P(\theta = \frac{1}{2}) = 1$, think of as $\delta(\frac{1}{2})$, not a good prior

$f(\theta|y) \propto f(y|\theta) f(\theta) = f(\theta)$

$$f(y) = \int f(y|\theta) f(\theta) d\theta = \int f(y,\theta)\, d\theta$$

Prior predictive before observing any data

6.2 Prior predictive: binomial example

$$X = \sum_{i=1}^{10} Y_i$$

$$f(\theta) = I_{\{0 \le \theta \le 1\}}$$

$$f(x) = \int f(x|\theta) f(\theta) d\theta = \int_0^1 \frac{10!}{x!(10-x)!} \theta^x (1-\theta)^{10-x} (1)\, d\theta$$

$$\left[\begin{array}{l} n! = \Gamma(n+1) \\ z \sim \text{Beta}(\alpha,\beta) \Rightarrow f(z) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha-1}(1-z)^{\beta-1} \end{array} \right]$$

$$\Rightarrow f(x) = \int_0^1 \frac{\Gamma(11)}{\Gamma(x+1)\Gamma(11-x)} \theta^{(x+1)-1} (1-\theta)^{(11-x)-1} d\theta$$

$$= \frac{\Gamma(11)}{\Gamma(12)} \int_0^1 \frac{\Gamma(12)}{\Gamma(x+1)\Gamma(11-x)} \theta^{(x+1)-1} (1-\theta)^{(11-x)-1} d\theta$$

$$= \frac{\Gamma(11)}{\Gamma(12)} (1) = \frac{10!}{11!} = \frac{1}{11} \quad \text{for } x \in \{0,1,\dots,10\}$$

6.3 Posterior predictive distribution

$$f(y_2|y_1) = \int f(y_2|\theta, y_1) f(\theta|y_1) d\theta$$

$$Y_2 \perp Y_1 \Rightarrow f(y_2|y_1) = \int f(y_2|\theta) f(\theta|y_1) d\theta$$

$$f(y_2|Y_1=1) = \int_0^1 \theta^{y_2}(1-\theta)^{1-y_2}\, 2\theta\, d\theta = \int_0^1 2\theta^{y_2+1}(1-\theta)^{1-y_2} d\theta$$

$$P(Y_2=1|Y_1=1) = \int_0^1 2\theta^2 d\theta = \frac{2}{3}$$

$$P(Y_2=0|Y_1=1) = \frac{1}{3}$$

7.1 Bernoulli / binomial likelihood with uniform prior

$$f(\underset{\sim}{y}|\theta) = \theta^{\sum y_i}(1-\theta)^{n-\sum y_i} \quad , \quad f(\theta) = I_{\{0\le\theta\le 1\}}$$

$$f(\theta|\underset{\sim}{y}) = \frac{f(y|\theta)f(\theta)}{\int f(y|\theta)f(\theta)\,d\theta} = \frac{\theta^{\sum y_i}(1-\theta)^{n-\sum y_i} I_{\{0\le\theta\le1\}}}{\int_0^1 \theta^{\sum y_i}(1-\theta)^{n-\sum y_i} I_{\{0\le\theta\le1\}}\,d\theta}$$

$$= \frac{\theta^{\sum y_i}(1-\theta)^{n-\sum y_i} I_{\{0\le\theta\le1\}}}{\frac{\Gamma(\sum y_i+1)\Gamma(n-\sum y_i+1)}{\Gamma(n+2)}\int_0^1 \frac{\Gamma(n+1)}{\Gamma(\sum y_i+1)\Gamma(n-\sum y_i+1)}\theta^{\sum y_i}(1-\theta)^{n-\sum y_i}\,d\theta}$$

$$= \frac{\Gamma(n+2)}{\Gamma(\sum y_i+1)\Gamma(n-\sum y_i+1)}\theta^{\sum y_i}(1-\theta)^{n-\sum y_i} I_{\{0\le\theta\le1\}}$$

$$\Rightarrow \theta|\underset{\sim}{y} \sim \text{Beta}(\textstyle\sum y_i+1,\ n-\sum y_i+1)$$

7.2 Conjugate priors

$$f(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} I_{\{0\le\theta\le1\}}$$

$$f(\theta|\underset{\sim}{y}) \propto f(\underset{\sim}{y}|\theta)f(\theta) = \theta^{\sum y_i}(1-\theta)^{n-\sum y_i}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} I_{\{0\le\theta\le1\}}$$

$$\propto \theta^{\alpha+\sum y_i-1}(1-\theta)^{\beta+n-\sum y_i-1} I_{\{0\le\theta\le1\}}$$

$$\theta|\underset{\sim}{y} \sim \text{Beta}(\alpha+\textstyle\sum y_i,\ \beta+n-\sum y_i)$$

Conjugate family

$Y_1,\dots,Y_n \sim B(\theta)$ likelihood
$\theta|\alpha,\beta \sim \text{Beta}(\alpha,\beta)$ prior
$\alpha,\beta = \alpha_0,\beta_0$ hyperparameters

7.3 Posterior mean and effective sample size

$\text{Beta}(\alpha+\sum y_i,\ \beta+n-\sum y_i)$

Prior $\text{Beta}(\alpha,\beta)$ $\begin{cases}\text{effective sample size of prior is } \alpha+\beta \\ \text{mean of Beta is } \frac{\alpha}{\alpha+\beta}\end{cases}$

Posterior mean is
$$\frac{\alpha+\sum y_i}{\alpha+\sum y_i+\beta+n-\sum y_i} = \frac{\alpha+\sum y_i}{\alpha+\beta+n} = \frac{\alpha+\beta}{\alpha+\beta+n}\cdot\frac{\alpha}{\alpha+\beta} + \frac{n}{\alpha+\beta+n}\cdot\frac{\sum y_i}{n}$$

Posterior mean = prior weight × prior mean + data weight × data mean

95% CI for θ is $\hat{\theta} \pm 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$

$f(\theta)$ observe $y_1, \ldots, y_n$

posterior $\theta | y_1, \ldots, y_n$

more data $y_{n+1}, \ldots, y_{n+m}$

then, yesterday's posterior is today's prior

new posterior $\theta | y_1, \ldots, y_{n+m}$

can keep chaining these together : sequential updates $\Leftrightarrow$ one batch update

---

Lesson 8.1 Poisson data

8.1

$Y_i \sim \text{Pois}(\lambda)$, $f(\underset{\sim}{y}|\lambda) = \dfrac{\lambda^{\sum y_i} e^{-n\lambda}}{\prod_{i=1}^{n} y_i!}$ for $\lambda > 0$

Gamma prior

$\lambda \sim \Gamma(\alpha, \beta)$, $f(\lambda) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$

$f(\lambda|\underset{\sim}{y}) \propto f(\underset{\sim}{y}|\lambda) f(\lambda) \propto \lambda^{\sum y_i} e^{-n\lambda} \lambda^{\alpha-1} e^{-\beta\lambda}$

$\propto \lambda^{(\alpha + \sum y_i) - 1} e^{-(\beta+n)\lambda}$

posterior is $\Gamma(\alpha + \sum y_i, \beta + n)$

mean of gamma is $\frac{\alpha}{\beta}$

posterior mean is $\dfrac{\alpha + \sum y_i}{\beta + n} = \dfrac{\beta}{\beta+n} \cdot \dfrac{\alpha}{\beta} + \dfrac{n}{\beta+n} \cdot \dfrac{\sum y_i}{n}$

effective sample size of prior is β

1) prior mean $\frac{\alpha}{\beta}$

a) prior std. dev. $\frac{\sqrt{\alpha}}{\beta}$

b) effective sample size β

2) vague prior

small $\varepsilon > 0$ $\Gamma(\varepsilon, \varepsilon)$ $\Rightarrow \dfrac{\varepsilon + \sum y_i}{\varepsilon + n} \approx \dfrac{\sum y_i}{n}$

$Y \sim Exp(\lambda)$

prior mean $= \frac{1}{10}$

$\Gamma(100, 1000) \Rightarrow$ prior std. dev. $= \frac{1}{100}$

$Y = 12$

$$f(\lambda|y) \propto f(y|\lambda)f(\lambda) \propto \lambda e^{-\lambda y} \lambda^{\alpha-1} e^{-\beta\lambda}$$
$$\propto \lambda^{(\alpha+1)-1} e^{-(\beta+y)\lambda}$$

$\Rightarrow \lambda|y \sim \Gamma(\alpha+1, \beta+y)$

$\Rightarrow \lambda|y \sim \Gamma(101, 1012)$

posterior mean $= \frac{101}{1012} = 0.0998 = \frac{1}{10.02}$

10.1 Normal likelihood with variance known

$X_i \overset{iid}{\sim} N(\mu, \sigma_0^2)$

prior $\mu \sim N(m_0, s_0^2)$

$f(\mu | \underset{\sim}{x}) \propto f(\underset{\sim}{x} | \mu) f(\mu)$

$$\mu | \underset{\sim}{x} \sim N\left(\frac{\frac{n\bar{x}}{\sigma_0^2} + \frac{m_0}{s_0^2}}{\frac{n}{\sigma_0^2} + \frac{1}{s_0^2}}, \frac{1}{\frac{n}{\sigma_0^2} + \frac{1}{s_0^2}} \right)$$

$$\text{posterior mean} = \frac{\frac{n}{\sigma_0^2}}{\frac{n}{\sigma_0^2} + \frac{1}{s_0^2}} \bar{x} + \frac{\frac{1}{s_0^2}}{\frac{n}{\sigma_0^2} + \frac{1}{s_0^2}} m_0$$

$$= \frac{n}{n + \frac{\sigma_0^2}{s_0^2}} \bar{x} + \frac{\frac{\sigma_0^2}{s_0^2}}{n + \frac{\sigma_0^2}{s_0^2}} m_0$$

effective sample size of prior $= \dfrac{\sigma_0^2}{s_0^2}$

10.2 Normal likelihood with mean and variance unknown

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

$\mu | \sigma^2 \sim N\left(m, \frac{\sigma^2}{w}\right)$, $w = \dfrac{\sigma^2}{\sigma_\mu^2}$ = effective sample size of the prior

$\sigma^2 \sim \Gamma^{-1}(\alpha, \beta)$

$$\sigma^2 | \underset{\sim}{x} \sim \Gamma^{-1}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{nw}{2(n+w)}(\bar{x} - m)^2\right)$$

$$\mu | \sigma^2, \underset{\sim}{x} \sim N\left(\frac{n\bar{x} + wm}{n+w}, \frac{\sigma^2}{n+w}\right)$$

$$\frac{n\bar{x} + wm}{n+w} = \frac{w}{n+w} m + \frac{n}{n+w} \bar{x}$$

$\mu | \underset{\sim}{x} \sim t$ distribution

11.1 Noninformative priors

$Y_i \sim B(\theta)$

$\theta \sim U[0,1] = Beta(1,1)$

or $Beta(\frac{1}{2},\frac{1}{2})$ or $Beta(.001,.001)$

$\underbrace{Beta(0,0)}_{\text{improper prior}} \Rightarrow f(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ not a proper density

$f(\theta|y) \propto \theta^{y-1}(1-\theta)^{n-y-1} \sim Beta(y, n-y)$

posterior mean $\frac{y}{n} = \hat{\theta}$

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

vague prior $\mu \sim N(0, 1000000^2)$

$f(\mu) \propto 1$

$f(\mu|y) \propto f(y|\mu)f(\mu) \propto \exp\left\{\frac{-1}{2\sigma^2}\Sigma(y_i-\mu)^2\right\}\cdot(1)$

$\propto \exp\left\{\frac{-1}{2\sigma^2/n}(\mu-\bar{y})^2\right\}$

$\mu|\underset{\sim}{y} \sim N\left(\bar{y}, \frac{\sigma^2}{n}\right)$

$f(\sigma^2) \propto \frac{1}{\sigma^2} \Leftrightarrow \Gamma^{-1}(0,0)$ improper, uniform on $\log\sigma^2$

$\sigma^2|\underset{\sim}{y} \sim \Gamma^{-1}\left(\frac{n-1}{2}, \frac{1}{2}\Sigma(y_i-\bar{y})^2\right)$

11.2 Jeffreys prior

$Y_i \sim N(\mu,\sigma^2)$

$f(\sigma^2) \propto \frac{1}{\sigma^2}$

$f(\sigma^2) \propto 1$

Jeffreys prior $f(\theta) \propto \sqrt{I(\theta)} \leftarrow$ Fisher information

$Y_i \sim N(\mu,\sigma^2)$, $f(\mu) \propto 1$, $f(\sigma^2) \propto \frac{1}{\sigma^2}$

$Y_i \sim B(\theta)$, $f(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2} \sim Beta(\frac{1}{2},\frac{1}{2})$