Baysian Statistics: From Concept to Data Analysis

Wk1: Course Introduction

→ Overview

→ Frequentist vs Baysian approach

- better at handling uncertainity
  ↳ quantify & combine
- easier to describe intervals

→ Probability

- A ← event (some outcome)

  ie rolling a 4 of d6-dice

  $P(A) = 4/6$ or $P(X=4) = \boxed{4/6}$

  - All events add to 1
  - cannot be negative

- compliment, A^c vs when event doesn't happen.

  ↳ ie $P(A^c) = P(\neg A) = 1 - \frac{1}{6} = \frac{5}{6}$

· if A, B then $P(A \underset{}{\overset{\text{"or"}}{\cup}} B) = P(A) + P(B)$
$- P(A \underset{\text{"and"}}{\cap} B)$

→ **Odds**

↳ Prob. also expressed in terms of odds

$$O(A) = \frac{P(A)}{P(A^c)} = \frac{1/6}{5/6} = \frac{1}{5}$$

convert odds to prob. w/ $P(A) = \dfrac{O(A)}{O(A) + O(A^c)}$

→ **Expectation**

each outcome has a certain outcome & p of achieving outcome.

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

$$= \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix} = 3.5$$

avg of many rolls

→ Classical & Frequentist Prob.

Try to estimate uncertainty to ~~so~~ answer questions

↳ Likelihood to roll 4?

↳ is coin fair?

...

Frameworks help to formulate & answer questions

- Classical
- Frequentist
- Bayesian

→ Classical

↳ outcomes that are equally likely have equal probabilities

↳ works for well defined problems

→ Frequentist

↳ infinite sequence of events & draw prob from there

↳ possible to estimate parameters using Law of Large Numbers

↳ Difficulty: Acting objective when not really

→ Baysian Probability

↳ "Personal" perspective

P(Fair) ⟵ different people w/ different data will answer different P.

bet taking framework

	win	lose
reward	+3	−4
prob.	1/2	1/2
outcome	$\frac{3}{2}$	$\frac{-4}{2}$

← $\Sigma = \frac{-1}{2}$

if −3 instead then fair game

another example

↳ Chile has 15 regions

Size total = 756,096

How big is Atacama?

$\frac{75 \times 10^4}{5} = 50{,}000$

outcome	likelihood
$A_1 < 10000$	0.2
$10000 \le A_2 < 50000$	0.5
$50000 \le A_3 < 100000$	0.2
$100000 \le A_4$	0.1

$\underbrace{}_{\text{priors}}$

change ~~increase~~ likelihood as more information becomes clear

→ Conditional probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$ ← ~~out~~ how many ways to get A of B

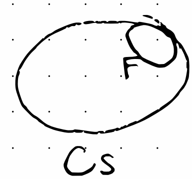
eg. Class 30 students

$9 \times F$ $\quad$ $12 \times CS$ $\quad$ $4 \times F \cap Cs$

$P(F) = \frac{9}{30}$ $\quad$ $P(Cs) = \frac{12}{30} = \frac{2}{5}$ $\quad$ $P(Cs) = \frac{2}{15}$

conditional → $P(F \mid CS) = \dfrac{P(F \cap Cs)}{P(CS)} = \dfrac{2/15}{2/5} = \dfrac{1}{3}$

- Conditionals just think of subsets of a population



- independence

if $P(A \mid B) = P(A)$ then $P(A \cap B) = P(A)P(B)$

→ Bayes Theorem

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

↳ have priors & update based on data

→ note that when multiple outcomes then we need to sum them all together

ie $$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{\sum_{i=1}^{m} P(B|A_i)P(A_i)}$$

if cont. dist. then gets replaced w/ integral.

→ Bernoulli & Binomial dist

Bernoulli → 2 outcomes
true/false

$X \sim B(p)$ $P(X=1) = p$

"is distributed as" $P(X=0) = 1-p$

$$F(X=x|p) = f(x|p) = p^{x}(1-p)^{1-x}$$

rand.   same var   given p

- Indicator function

$$I_{\{x \in \{0,1\}\}}(x) = \begin{cases} \text{true} \\ \text{false} \end{cases}$$

Always eval first.

- Expected val of Bernoulli

$$E(X) = \sum_{x} x P(X=x) = (1)p + (0)(1-p)\cancel{p} = p$$

$$var(x) = p(1-p)$$

n repeated trials creates binomial dist

$$P(X=x|p) = f(x|p) = \binom{n}{x} p^{x} (1-p)^{n-x}$$

combinatoric

$$E(x) = np$$

$$var(x) = np(1-p)$$

NB Binomial(n, x) =

$\therefore X \sim \text{Binom.}(3, 0.2)$

$$P(X=0) = \binom{3}{0} 0.2^{0} (1-0.2)^{3-0}$$

$$= (1)(1)(0.8)^3 = 0.512$$

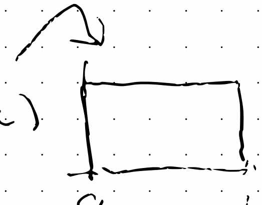
→ Uniform dist

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

↳ if between x, y

then $y = \frac{1}{y-x}$.

$= I_{\{0 \le x \le 1\}}(x)$



→ Notes on integrals & expectation vals

if $E[X] = \int_{-\infty}^{\infty} x f(x)\,dx$

then $E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

also $E[X+Y] = E[X] + E[Y]$

if $E \perp Y$ then $E[XY] = E[X]E[Y]$

→ Exponential Dist

$X \sim Exp(\lambda)$ ← waiting time between events

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

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

→ Normal Dist

$X \sim N(\mu, \sigma^2)$ $\quad f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

$Ex = \mu$ $\quad var(x) = \sigma^2$

→ W42: Statistical Inference

→ Frequentist inference

→ confidence intervals

↳ again flips of 100x coins results in 44H
56T

$x_i \sim B(\cancel{0.5}^{p})$

How do we estimate p

$\therefore \sum x_i \,\dot{\sim}\, N(100p, 100p(1-p))$

follows approx.

↳ confidence interval

$100p \pm 1.96\sqrt{100\,p(1-p)}$

$P(X<5)$ & $P(X<95)$

↳ using frequentist logic we can estimate p
w/ $\hat{p}$ & find CI ... (34.3, 53.7)

will contain
true mean 95% of the time

50 contained within interval so we have
no evidence to disprove that the coin is fair

- difficulty: coin becomes either fair or unfair "removing" "uncertainty"

→ Likelihood & Max. Likelihood

↳ 400 patients admitted over a month

↳ 72 die, 28 survive

estimate mortality rate

↳ what is our reference?

↳ population? ↳ random sample?

↳ other patients?

$Y_i \sim B(\theta) \rightarrow P(Y_i = 1) = \theta$

in vector form → $P(Y_1 = y_1, Y_2 = y_2, \dots Y_n = y_n | \theta)$

$= \prod_{i=1}^{N} \theta^{y_i}(1-\theta)^{1-y_i}$

Likelihood $= L(\theta | y) =$

↳ choose θ which gives us highest likelihood (ML)

$MLE = \hat{\theta} = \text{argmax } L(\theta | y)$

If we take log then we can convert Π to Σ

$$\ell(\theta) = \Sigma \log(\theta^{y_i}(1-\theta)^{1-y_i})$$

$$= \Sigma\ y_i(\log\theta) + (1-y_i)\log(1-\theta)$$

these are const.

$$= \log\theta\ \Sigma y_i + \log(1-\theta)\ \Sigma(1-y_i)$$

max this function by taking deriv & setting to 0

$$\ell'(\theta) = \frac{1}{\theta}\Sigma y_i - \frac{1}{1-\theta}\Sigma(1-y_i) = 0$$

$$\therefore\ \hat{\theta} = \frac{1}{n}\Sigma y_i$$

unbiased
consistent
invariante

$$CLI \Rightarrow \hat{\theta} \pm 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

0.53

→ estimating parameter for exponential

↳ $X_i \overset{iid}{\sim} Exp(\lambda)$ identical & identically distributed

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

$$\ell(\lambda | \underset{\sim}{x}) = n \log(\lambda) - \lambda \Sigma x_i$$

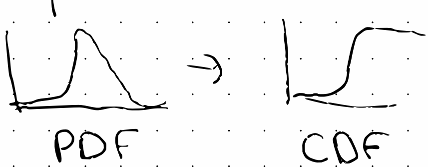
$$\ell'(\lambda) = \frac{n}{\lambda} - \Sigma x_i = 0 \quad \therefore \hat{\lambda} = \frac{1}{\bar{x}}$$! yay

→ More basic statistics

→ CDFs

↳ $F(x) = P(X \leq x)$

ie integration from 0 to x for all values of x.

↳ 

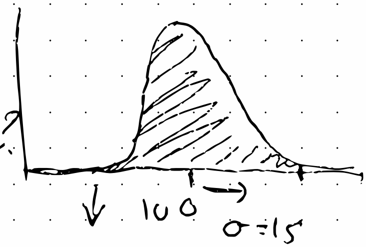
→ Quantiles

$X \to CDF(X) \to$ probability

what if we start w/ p & work back to value

$p \to CDF^{-1}(p) \to X$

ie if $IQ \sim N(100, 15)$
then what IQ has 95th percentile?



use CDF to find this value, ppf in python

↳ $100 + ppf(.95) * 15$
≈ 125

→ Bayesian Inference

↳ loaded coin $p = .7$

↳ Bet w/ unknown coin

↳ if heads is coin loaded?

$\Theta = \{f, l\} \quad X \sim Bin(5, \Theta)$

$$F(x \mid \Theta) = \begin{cases} \binom{5}{x}\left(\frac{1}{2}\right)^5 & \text{if } \Theta = \cancel{\text{fair}}\ f \\ \binom{5}{x}(.7)^x(.3)^{5-x} & \text{if } \Theta = l \end{cases}$$

If $X=2$ $\quad f(\theta \mid X=2) = \begin{cases} .3125 & \text{if } \theta = f \\ .1323 & \text{if } \theta = l \end{cases}$

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

point estimate, but how sure are we?

↳ compare this w/ Bayesian inference

- Allows for incorporation of priors

Prior $\quad P(\text{loaded}) = .6$

$\qquad\quad P(\text{fair}) = .4$

↳ update w/ data to give posterior

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

$$= \frac{\binom{5}{x}\left[\left(\frac{1}{2}\right)^5 (.4)\, I_{\{\theta=f\}} + (.7)^x (.3)^{5-x} (.6)\, I_{\{\theta=l\}}\right]}{\binom{5}{x}\left[\left(\frac{1}{2}\right)^5 (.4) + (.7)^x (.3)^{5-x} (.6)\right]}$$

$$f(\theta \mid x=2) = .612\, I_{\theta=f} + .388\, I_{\theta=l}$$

$P(\theta = l \mid x=2) = .388$ ← probability is easier to interpret.

↳ we can take different priors to get different posteriors.

$$P(\theta = 1) = 0.5 \Rightarrow P(\theta = 1 \mid x = 2) = .297$$
$$= .9 \Rightarrow \qquad = .792$$

- Baysian approach is subjective (w/ min. data)
- The values are easy to interpret

→ <u>continuous bayes</u>

$$f(\theta \mid y) = \frac{f(y \mid \theta) f(\theta)}{f(y)} = \frac{f(y \mid \theta) f(\theta)}{\int f(y \mid \theta) f(\theta)}$$

likelihood, norm. const., prior

norm. const can be v. hard to compute & so we say: $f(\theta \mid y) \propto f(y \mid \theta) f(\theta)$

ex.

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

prior ~~post~~ for com pdf

$$f(\Theta \mid y) = \frac{P(y \mid \theta)\, P(\Theta)}{\int f(y \mid \theta) f(\theta)\, d\Theta}$$

hypo data

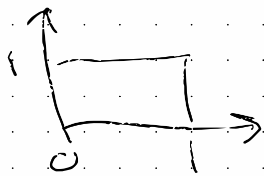
$$f(\Theta \mid Y=1) = \frac{\Theta^1 (1-\Theta^0)\, I_{0 \le \Theta \le 1}}{\int_{-\infty}^{\infty} \Theta^1 (1-\Theta^0)\, I_{0 \le \Theta \le 1}\, d\Theta}$$

$$= \frac{\Theta\, I_{0 \le \Theta \le 1}}{\int_0^1 \Theta\, d\Theta} = 2\Theta\, I_{0 \le \Theta \le 1}$$

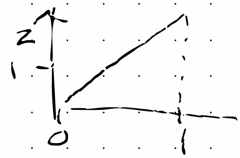
or $\propto \Theta\, I_{0 \le \Theta \le 1}$

requires normalization

Lets toy w/ plotting



Posterior



prior interval est

$P(.025 < \theta < .975)$
$= .95$

post. interv. est.

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

Credible intervals

- Equal tailed

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

$P(\sqrt{.025} < \theta \sqrt{.975}) = P(.158 < \theta < .987) = .95$

- HPD (Highest Posterior Density)
  - ↳ Shortest interval giving .95 prob

$P(\theta > \sqrt{.05} \mid Y=1) = .95$

Conclusion:

post. gives understanding of uncertainty given prior & data

↳ we can give intervals & talk about prob. being in this interval.

↳ uncertainties are represented

→ Wk 3: Priors & Models for discrete data

→ Priors

↳ How do we choose a prior

$P(\Theta \leq c)$ for all $c \in \mathbb{R}$

→ if theres enough data priors will get washed out. (As long as not 100% certain or ~~uncertain~~ 0%)

↳ thus never assign 1 or 0 prob. to priors

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

use this to create intervals w/ density in the expected regions

example

↳ flip coin 10x, what is our prior?

$$x = n. \text{ heads} = \sum_{i=1}^{10} y_i$$

$$\therefore f(\theta) = I_{(0<\theta<1)}$$ ← all probs are equally likely

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

↳ how to simplify integral

$$\binom{10}{x} = \frac{10!}{x!(10-x)!}$$

remember $n! = \Gamma(n+1)$

Gamma func.

$Z \sim \text{Beta}(\alpha, \beta)$

$$f(z) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha-1}(1-z)^{\beta-1}$$

$\boxed{\Gamma(n) = (n-1)!}$

so...

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

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

$$= \frac{\Gamma(11)}{\Gamma(12)} \underbrace{\int_0^1 \frac{\Gamma(12)}{\Gamma(x+1)\Gamma(11-x)} \theta^{(x+1)-1} (1-\theta)^{(11-x)-1} \, d\theta}_{\text{integral is 1}}$$

$$= \frac{\Gamma(11)}{\Gamma(12)} = \frac{10!}{11!} = \frac{1}{11} \quad \text{for } x \in \{0, 1, 2 \ldots 10\}$$

· posterior predictive distribution

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

ass. $y_2 \perp y_1$ $\quad = \int f(y_2 \mid \theta)\, f(\theta \mid y_1)\, d\theta$

prior = uniform $\quad y_1$ was heads

$$f(y_2 \mid y_1 = 1) = \int_0^1 \theta^{y_2} (1-\theta)^{1-y_2}\, 2\theta\, d\theta$$

calculated previously

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

$$P(y_2 = 1 \mid y_1 = 1) = \int_0^1 2\theta^2\, d\theta = \frac{2}{3}$$

$$P(y_2 = 0 \mid y_1 = 1) = \int_0^1 2\theta (1-\theta)\, d\theta = \frac{1}{3}$$

post for heads is more likely w/ 1x H

→ Bernoulli/Binomial likelihood w/ uniform prior

↳ gives beta posterior. (useful for A/B testing)

$$\underbrace{f(y|\theta) = \theta^{\Sigma y_i}(1-\theta)^{n-\Sigma y_i}} \quad \underbrace{f(\theta) = I_{0<\theta<1}}$$

$$\text{posterior} = f(\theta|y) = \frac{f(y|\theta)\, f(\theta)}{\int f(y|\theta) f(\theta)\, d\theta}$$

$$\therefore \; = \frac{\theta^{\Sigma y_i}(1-\theta)^{n-\Sigma y_i} I_{0<\theta<1}}{\int_0^1 \theta^{\Sigma y_i}(1-\theta)^{n-\Sigma y_i} I_{0<\theta<1}\, d\theta}$$

this is similar to Beta pdf
so, we add terms to make it Σ to 1

$$= \frac{\theta^{\Sigma y_i}(1-\theta)^{n-\Sigma y_i} I_{0<\theta<1}}{\frac{\Gamma(\Sigma y_i+1)\Gamma(n-\Sigma y_i+1)}{\Gamma(n+2)} \boxed{\int_0^1 \frac{\Gamma(n+2)}{\Gamma(\Sigma y_i+1)\Gamma(n-\Sigma y_i+1)} \cdots \theta^{\Sigma y_i}(1-\theta)^{n-\Sigma y_i} I_{0<\theta<1}\, d\theta}}$$

Adds to 1 →

$$= \frac{\Gamma(n+2)}{\Gamma(\Sigma y_i+1)\Gamma(n-\Sigma y_i+1)} \theta^{\Sigma y_i}(1-\theta)^{n-\Sigma y_i} I_{0<\theta<1}$$

$\therefore f(\theta|y) \sim Beta(\Sigma y_i + 1, n - \Sigma y_i + 1)$

↳ so if H, H, T then $\Sigma y_i = 2$
represented w/ Beta(3, 2)

- Conjugate priors
  - ↳ uniform ← beta(1,1)
  - ↳ Beta prior → beta posterior

- conjugate family
  - ↳ when dist in prior appears again in the posterior

Beta is conj for
- bernoulli
- binomial

→ Obviously, conj priors make things easier for us

↳ note that hyperparameters such as α or β in Beta(α, β) can also have distributions

↳ building will result in complex models
"heirarchical models"

- Posterior mean & effective sample size

Prior

$$\text{Beta}(\alpha, \beta) \xrightarrow[\text{observ.}]{y_i} B(\alpha + \Sigma y_i, \beta + n - \Sigma y_i)$$

↳ effective sample size of prior is $\alpha + \beta$ (uninformed, 1+1)

$E[\text{Beta}] = \frac{\alpha}{\alpha+\beta}$ ⇒ mean of post. is $\frac{\alpha + \Sigma y_i}{\alpha + \beta + n}$

$$= \underbrace{\frac{\alpha+\beta}{\alpha+\beta+n} \cdot \frac{\alpha}{\alpha+\beta} + \frac{n}{\alpha+\beta+n} \cdot \frac{\Sigma y_i}{n}}$$

post. mean = prior weight * prior mean + data weight * data mean

$\underbrace{\text{prior weight} \quad \text{data weight}}_{\text{must add to 1}}$

this info can tell you how much data you need to drown out priors

↳ using our posterior, we can get 95% credible interval

→ Poisson Data

↳ model n choc. chips in cookie

Y_i = n chips ~ Pois(λ)
~~in~~ per cookie

$$f(y|\lambda) = \frac{\lambda^{\sum y_i} e^{-n\lambda}}{\prod_{i=1}^{n} y_i!} = \boxed{\prod_{i=1}^{n} \frac{\lambda^{y} e^{-\lambda}}{y!}}$$

might be wrong

for $\lambda > 0$

prod of many draws of pois.

Conj. prior for Poiss?

↳ Gamma process

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

$E(\lambda) = \frac{\alpha}{\beta}$

$$f(\lambda|y) \propto f(y|\lambda) f(\lambda)$$

$$\propto \frac{\lambda^{\sum y_i} e^{-n\lambda}}{\prod_{i=1}^{n} y!} \lambda^{\alpha-1} e^{-\beta\lambda} \frac{\beta^{\alpha}}{\Gamma(\alpha)}$$

ignore due to const coeff.

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

mean(post.) = $\alpha + \sum y_i / \beta + n = \frac{\beta}{\beta+n} \cdot \frac{\alpha}{\beta} + \frac{n}{\beta+n} \cdot \frac{\sum y_i}{n}$

posterior mean is once again a weighted average of data & prior means.

How do we choose α & β for our prior?

1) prior mean $= \frac{\alpha}{\beta}$ what do we guess?

- std $= \frac{\sqrt{\alpha}}{\beta}$

- eff. sample size $= \beta$

2) uninformed prior

small $\varepsilon > 0$ $\Gamma(\varepsilon, \varepsilon)$

→ Wk 4: Models for continuous data

→ Exponential Data

↳ Wait for a bus that comes 1x every 10 min

$Y \sim Exp(\lambda)$ $\quad Exp = \frac{1}{\lambda}$ $\quad \Leftarrow$ Gama is conj for exp!

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

prior variability $= \frac{1}{100}$ $\longrightarrow$ $\boxed{\Gamma(100, 1000)} = f(\lambda)$

re $.1 \pm .02$

$y_1 = 12$, how to update post.?

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

$$\propto \lambda e^{-\lambda y} \lambda^{\alpha-1} e^{-\beta\lambda}$$

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

$$\therefore \quad f(\lambda | y) \sim \Gamma(\alpha+1, \beta+y)$$

↳ $\Gamma(101, 1012)$

↳ Barely shifts posterior

Note: α is the sample size

→ Normal Data

$X_i \overset{iid}{\sim} N(\mu, \sigma_0^2)$ ← normal is conjugate for itself

↳ prior $\mu \sim N(m_0, S_0^2)$

posterior $f(\mu|x) \propto f(x|\mu) f(\mu)$

$$\sim N\left(\frac{\frac{n\bar{x}}{\sigma_0^2} + \frac{m_0}{S_0}}{\frac{n}{\sigma_0^2} + \frac{1}{S_0^2}}, \frac{1}{\frac{n}{\sigma_0^2} + \frac{1}{S_0^2}} \right)$$

posterior mean $\dfrac{n}{n + \frac{\sigma_0^2}{S_0^2}} \bar{x} + \dfrac{\frac{\sigma_0^2}{S_0^2}}{n + \frac{\sigma_0^2}{S_0^2}} m$

once again a weighted average of two means

↳ ~~pr~~ note that prior predictive distribution is $N(m_0, S_0^2 + \sigma_0^2)$

→ Normal likelihood w/ var unknown

↳ Specify conj. prior in heirarchical fashion

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

$\mu | \sigma^2 \sim N(m, \frac{\sigma^2}{w})$ ← $\frac{\sigma^2}{\sigma_\mu^2}$

where w = effective sample size of prior

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

$$\sigma^2 | x \sim \Gamma^{-1}\left(\alpha + \frac{n}{2}, \beta + 0.5\sum(x_i - \bar{x})^2 + \frac{nw}{2(n+w)}(\bar{x} - m)^2\right)$$

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

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

... $\mu | x \sim t$

→ Alternative Priors

→ Informed & uninformed priors

↳ Alternative: explicitly uninit priors

coin flips

$y_i \sim B(\theta)$ ← Minimise information in prior ↳ uniform dist.

$\theta \sim U[0,1]$

· uniform isn't completely uninformed! → $= \text{Beta}(1,1)$

↳ sample size 2

↳ How do we get $\text{Beta}(0,0)$?

where $\boxed{f(\theta) \propto \theta^{-1}(1-\theta)^{-1}}$ ← integral is ∞

↑ improper prior

if we use then posterior gives frequentist answers.

key concepts

· improper priors are okay if post is proper

· might need restrictions on data

· returns to frequentist paradigm

another example $x_i \overset{iid}{\sim} N(\mu, \sigma^2)$

unknown

vague prior $\mu \sim N(0, 10^2)$

further still... $\mu \sim 1$

improper

but if we use $\mu \sim 1$ then

$$f(\mu|y) \propto f(y|\mu)$$

$$\propto e^{-\frac{\sum(y_i-\bar{y})^2}{2\sigma^2}}$$

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

MLE

If σ is unknown...

$$f(\sigma^2) \propto \frac{1}{\sigma^2} \quad \Gamma^{-1}(0,0)$$

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

→ Jeffrey's prior

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

what if $\sigma^2 \propto 1$

$f(\theta) \propto \sqrt{I(\theta)}$ (improper)

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

for bernoulli $y_i \sim B(\theta)$ $\quad f(\theta) \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$

$\sim Beta(\frac{1}{2}, \frac{1}{2})$

proper

→ Bayesian Linear Regression

→ 100% demonstrated

End of course!