# Bayesian data analysys

**Transformation of variable**

Suppose $u$ is a random variable.

$p_u(u)$: Distribution of random variable

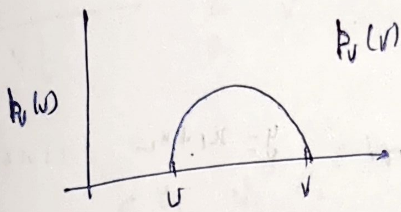
$v = f(u)$.

$p_v(v)$: Distribution of $v$ that we want to find.

Case 1: $p_u(u)$ is discrete and $v$ is one to one mapping

$v = f(u)$

$p_v(v) = p_u(f^{-1}(v))$




→ $X \sim$ geometric distribution, $p(x) = \frac{3}{4}\left(\frac{1}{4}\right)^{x-1}$, $x \in Z$

find the probability distribution of the RV $Y = x^2$

→ $Y = x^2 \Rightarrow x = \sqrt{y}$.

$$p(y) \cdot P_Y(y) = \begin{cases} \frac{3}{4}\left(\frac{1}{4}\right)^{\sqrt{y}-1}, & \text{for } y = 1, 4, 9 \\ 0, & \text{otherwise.} \end{cases}$$

Case-1 a: Suppose $X_1$ and $X_2$ are discrete RV with a joint probability distribution $p_{x_1, x_2}(x_1, x_2)$ and let $y_1$ be a function

$y_1 = u_1(x_1, x_2)$ and $y_2 = u_2(x_1, x_2)$

two one to one transformation.

so, we can solve $x_1$ and $x_2$ in terms of $y_1$ & $y_2$

$$\left.\begin{aligned} x_1 &= w_1(y_1, y_2) \\ x_2 &= w_2(y_1, y_2) \end{aligned}\right\}$$ find distribution of $p(y_1, y_2)$

$p_{y_1 y_2}(y_1, y_2)$

$$\Rightarrow \boxed{p_{y_1 y_2}(y_1, y_2) = p_{x_1 x_2}\left(w_1(y_1, y_2), w_2(y_1, y_2)\right)}$$

Q) Suppose $X_1$ & $X_2$ be two ~~rv~~ random independent random variable with poisson distribution with mean $\mu_1, \mu_2$.

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Find the distribution of rv $y = x_1 + x_2$

$y = x_1 + x_2$

$v = x_1 \qquad x_2 = y - v$

$$p_{x_1 x_2}(x_1, x_2) = \frac{\mu_1^{x_1} e^{-\mu_1}}{x_1!} \cdot \frac{\mu_2^{x_2} e^{-\mu_2}}{x_2!}$$

$$p_{y, v}(y, v) = \frac{\mu_1^{v} e^{-\mu_1}}{v!} \cdot \frac{\mu_2^{y-v} e^{-\mu_2}}{(y-v)!}$$

$$p_y(y) = \sum_v \frac{\mu_1^v}{v!} \frac{\mu_2^{y-v}}{(y-v)!} e^{-(\mu_1 + \mu_2)}$$

$$= e^{-(\mu_1+\mu_2)} \cdot \mu_2^{y} \sum_{v} \left(\frac{\mu_1}{\mu_2}\right)^{v} \cdot \frac{1}{(v!)(y-v)!}$$

$$= \frac{e^{-(\mu_1+\mu_2)} \cdot \mu_2^{y}}{y!} \sum_{v} \left(\frac{\mu_1}{\mu_2}\right)^{v}$$

$$= \frac{e^{-(\mu_1+\mu_2)} \cdot \mu_2^{y}}{y!} \sum_{v} \left(\frac{\mu_1}{\mu_2}\right)^{v} \frac{y!}{v!(y-v)!}$$

$$= \frac{e^{-(\mu_1+\mu_2)} \cdot \mu_2^{y}}{y!} \sum_{v} {}^{y}C_{v}\, \mu_1^{v}\, \mu_2^{-v}$$

$$= \frac{e^{-(\mu_1+\mu_2)}}{y!} \sum_{v} {}^{y}C_{v}\, \mu_1^{v}\, \mu_2^{y-v}$$

$$= \frac{e^{-(\mu_1+\mu_2)} (\mu_1+\mu_2)^{y}}{y!} \sim P(\mu_1+\mu_2).$$

Transformation of varibl

Case 2 Suppose $X$ is a continous RV $f_X(x)$ being its distribution and let $Y = u(X)$ define a one to one mapping from $X$ to $Y$. so $X = w(Y)$ Then the probability distribution $f_Y(y) = f_X(w(y)) \, |w'(y)|$ when $J = w'(y)$ is Jacobians an.

$$p_y(a<y<b) = \int_a^b p_y(y)\,dy = \int_{w(a)}^{w(b)} p_x(w(y))\,dx$$

$$= \int_a^b p_x(w(y))\,|w'(y)|\,dy$$

Q) Let x be a continuous RV with probability distribution

18 march (6-8)

18 march - (10-11)

The posterior is actually a compromise between the prior and data information

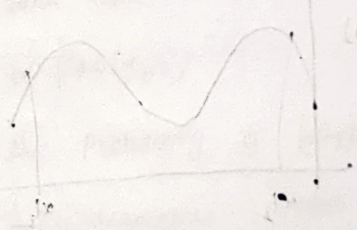
$$E(u) = \int u\, p(u)\, du$$

$$= \int_U \int_V u\, p(u,v)\, du\, dv$$

$$= \int_U \int_V p(u|v)\, p(v)\, u\, du\, dv$$

$$= \int_V dv\, p(v) \int_U u\, p(u|v)\, du$$

$$= \int_V E(u|v)\, p(v)\, dv = E\,E(u|v)$$



Simillary

$$E(\theta) = E(E(\theta|y))$$

$$var(\theta) = E(var(\theta|y)) + var(E(\theta|y))$$

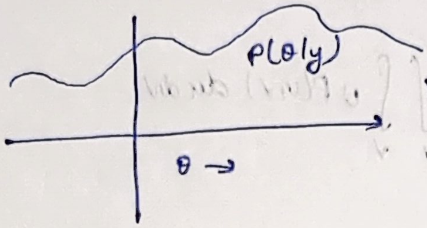
expected variance of Posterior

Variance of the expected posterior

# How to summarize the posterior inference.

- Posterior probability distribution, $p(\theta|y)$
  - contains all the information about the current parameter $\theta$

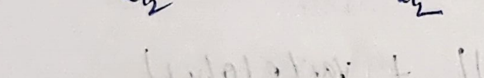
- The multiple parameters we may use a contour plot.




→ Numerical summaries like mean, median, mode.

- Variation, Interquartile plot, Range using Box plot.

- Posterior quartiles and intervals
  - Central interval of a posterior probability $100(1-\alpha)\%$ interval.




- Highest probability density
  - The part where the frequency of inner part must be larger than all the outside region.

Priors & basic interpretation of prior distribution

① Population Interpretation.

↳ Prior distribution represent a population of possible parameter values from which the correct parameter values is drawn.

stage of ② Knowledge Representation

↳

— There is an actual prior distribution and current value represents a random realization from the distribution.

Problem - the knowledge of entire population from which $\theta$ be drawn is not possible.

Ex - Industrial failure

Binomial example with different priors:

$$p(y|\theta) \propto \theta^{y}(1-\theta)^{n-y}$$

If we assume prior also to be same form

$$p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \propto \text{Beta}(\alpha,\beta)$$

This is very similar to the probability of $(\alpha-1)$ in $(\alpha+\beta-2)$ trials. $\alpha, \beta$ are hyperparameters that are assumed.

Posterior:

$$p(\theta|y) \propto p(y|\theta).p(\theta) = \theta^{y+\alpha-1}.(1-\theta)^{n-y+\beta-1}$$

$$= \text{Beta}(y+\alpha, n-y+\beta)$$

- When prior and posterior are of the same form, then they are called conjugate prior.

- If $E(\theta|y) = \frac{y+\alpha}{n+\alpha+\beta}$ $var(\theta|y) = \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n-1)}$

$1 - E(\theta|y) = \frac{\beta+n-y}{(\alpha+\beta+n)}$

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

$= \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n-1)}$

$= \frac{E(\theta|y)(1-E(\theta|y))}{(\alpha+\beta+n-1)}$

when $y, n$ tend to be Larger for fixed $\alpha, \beta$

$E(\theta|y) \to y/n$

$var(\theta|y) \to \frac{1}{n}(\frac{y}{n})(1-\frac{y}{n}) \to 0$ as $n \to \infty$

## Formal distribution of conjugate prior

If $F$ is a class of sampling distribution $p(y|\theta)$ and $P$ is a class of prior distribution $\theta$. Then $p(\theta|\theta) \in F$

$P$ is conjugate for $F$ if $p(\theta|y) \in P$ $\forall$ $p(\cdot|\theta) \in F$ and $p(\theta) \in P$.

We want $P$ to be same form as $F$, then it is called natural conjugate pair.

# Exponential family of distribution.

class $F$ is a family of exponential distribution if all its members have the form $p(y_i|\theta) = f(y_i) g(\theta) e^{\phi(\theta) u(y_i)}$.

- It's not necessarily that $\theta$ is a single parameter.
- If $\theta$ is a multiparameter then $p(y_i|\theta) = f(y_i) g(\theta) e^{\phi(\theta)^T u(y_i)}$
  - $\phi(\theta)$ & $u(y_i)$ must be of same dimension.
  - $\phi(\theta)$ is called natural parameter of $f$.

(5) Considering this principle the non uniform prior density is given as

$p(\theta) \propto |J(\theta)|^{1/2}$ where $J(\theta)$ is the fisher information for $\theta$.

$$J(\theta) = E\left[\left(\frac{d \log P(y|\theta)}{d\theta}\right)^2 \Big| \theta\right]$$

(6) Given a measure of amount of the information a parameter carries about the likelihood function.

(7) Helps quantify the sensitivity of model parameter to changes in data distribution.

$$J(\theta) = E\left(\frac{d^2 \log P(y|\theta)}{d\theta^2} \Big| \theta\right)$$

$E\left(\cancel{\frac{d\log P(y|\theta)}{}}\right.$ $$E\left(\frac{d \log P(y|\theta)}{d\theta} \Big| \theta\right) = \int_y \left[\frac{d}{d\theta} \log P(y|\theta)\right] P(y|\theta) dy$$

$$= \int_y \frac{1}{\cancel{P(y|\theta)}} \frac{\delta}{\delta\theta} P(y|\theta) \cdot \cancel{P(y|\theta)} \, dy$$

$$= \frac{d}{d\theta} \int_y P(y|\theta)\, dy = 0$$

Now

$$\frac{d^2 \log P(y|\theta)}{d\theta^2} = \frac{d}{d\theta}\left(\frac{1}{P(y|\theta)} \frac{d P(y|\theta)}{d\theta}\right) = \frac{1}{P(y|\theta)} \frac{d^2 P(y|\theta)}{d\theta^2}$$

$$= \frac{1}{P(y|\theta)} \cdot \frac{d}{d\theta^2} P(y|\theta)^2$$

$$= \frac{1}{P(y|\theta)} \frac{d^2 P(y|\theta)}{d\theta^2} = \left(\frac{d}{d\theta} \log P(y|\theta)\right)^2$$

Jeffery

$$p(\theta) = [J(\theta)]^{1/2}$$

I will come in exam for sure

Prove that Jeffery's prior is invariant to parameter

$$p(\theta) = [J(\theta)]^{1/2}$$

$J(\phi)$ at $\underline{\theta = h^{-1}(\phi)}$ show $p(\phi) = p(\theta) \left|\frac{d\theta}{d\phi}\right|$

$$J(\phi) = E\left(\frac{d^2}{d\phi^2} \log P(y|\phi) \,|\, \phi\right)$$

$$= E\left(\frac{d^2}{d\theta^2} \log P(y|\theta = h^{-1}(\phi))\,|\,\theta\right)\left(\frac{d\theta}{d\phi}\right)^2$$

$$= J(\theta) \left|\frac{d\theta}{d\phi}\right|^2$$

$$(J(\phi))^{1/2} = (J(\theta))^{1/2} \left|\frac{d\theta}{d\phi}\right|$$

$$\boxed{p(\phi) = p(\theta) \left|\frac{d\theta}{d\phi}\right|}$$ Proved

For two cases all principles (for generating non-informative prior) will give similar result.

(a) If $p(y)$ is such that $p(y-\theta|\theta)$ is free of $\theta$ & $y$, then $y-\theta$ is called a pivotal quantity, $\theta$ is called a location parameter.

$$p(y-\theta|y) \propto p(\theta)\, ~~p(y|\theta)~~\, p(y-\theta|\theta)$$

↳ constant

(b) If $p(y)$ is such that $p(y|\theta)$ is free from $\theta$ & $y$, then $\frac{y}{\theta} = u$ is called a pivotal quantity. $\theta$ in this case is a scale parameter.

Problem with Non-informative prior

It is difficult to find some prior which is flat. The CDF may not be equal to 1 in all cases.

Weakly Informative prior: Proper Prior with some general information.

Ex: Money in the Wallet: May range from 100 to 3000. So we can consider a normal distribution with mean 500.

Case study

A study was conducted to know what is the problem of the girl birth and boy birth in case the mother is suffering from placenta previa.

Proportion of ~~girl~~ girl birth in population is 0.485.

The sample of ~~pre~~ placenta previa out of 980 samples, 437 were girls.

$$P(\text{girls} \mid \text{mother has placenta previa}) = \frac{437}{980}$$

$\theta$ = Prob. of a girl birth is patient with PP.

To find. $E(\theta)$ and $Var(\theta)$

$$\overset{\text{Posterior}}{p(\theta \mid y, n)} = p(\theta) \cdot p(y \mid \theta, n)$$

↳ girl birth

$$p(y \mid \theta, n) \sim \theta^y (1-\theta)^{n-y}$$

$$p(\theta) \sim Beta(1,1)$$

$$p(\theta \mid y, n) \propto p(\theta) \cdot p(y \mid \theta, n) \propto \theta^y (1-\theta)^{n-y}$$

$$= Beta(y+1, n-y+1)$$

$$E(Beta(y+1, n-y+1)) = \frac{y+1}{n+2} = 0.446$$

$$Var(Beta(y+1, n-y+1)) = \frac{(y+1)(n-y+1)}{(n+2)^2 (n+3)} = 0.016.$$

Q) Suppose you have Beta(4,4) prior distribution on the probability $\theta$ that a coin will yield a head when spun in a specified manner. The coin is independently spun 10 times and head appear fewer than 3 times. You are not told how many heads were seen only the number is less than 3. Calculate the posterior density upto a proportionality constant.

Sol

$p(\theta) \sim Beta(4,4)$

unobserved parameter $\rightarrow \theta$

Observation $\rightarrow y<3$ (no of heed in 10 trials)

$p(\theta|y<3) \rightarrow$ To find posterior

$\propto p(y<3|\theta)\, p(\theta)$

$= p(y=0|\theta)p(\theta) + p(y=1|\theta)\, p(\theta) + p(y=2|\theta)p(\theta)$

$= \left[\binom{n}{0}\theta^0 (1-\theta)^n + \binom{n}{1}\theta^1 (1-\theta)^{n-1} + \binom{n}{2}\theta^2 (1-\theta)^{n-2}\right] p(\theta)$

$= \left[(1-\theta)^n + n(1-\theta)^{n-1}\cdot\theta + \frac{n(n-1)}{2}\theta^2 (1-\theta)^{n-2}\right] p(\theta)$

$= \left[(1-\theta)^{10} + 10\,\theta(1-\theta)^{n-1} + 45\theta^2 (1-\theta)^8\right] \theta^3 (1-\theta)^3$

$= \left[(1-\theta)^2 + 10\theta\cdot(1-\theta) + 45\theta^2\right] \theta^2 (1-\theta)^{11}$

$\Rightarrow \theta^3(1-\theta)^{11}[1+\theta^2-2\theta+10\theta-10\theta^2+45\theta^2]$

$= \theta^3(1-\theta)^{11}[36\theta^2+8\theta+1]$

(3) mind game

One person is guessing random numbers from 1 to 10 20 times and based on that the other person will guess the 10 samples whether they are greater than 5 or less than 5.

Observations $y = \{y_1, y_2, \dots y_{10}\}$

when $y_i = 0$ if count $\leq 5$

$= 1$ otherwise.

$P(\tilde{y}|y) = \int_\theta P(\tilde{y}, \theta|y) = \int_\theta P(\tilde{y}|\theta) \cdot P(\theta|y)$

$= \int_\theta \theta\, P(\theta|y) = \int_\theta \theta \frac{P(y|\theta)\, P(\theta)}{P(y)}$

$= \frac{y+1}{n+2}$

Till now

- Assume a likelihood function.
  - ↳ Binomial
  - ↳ Normal
  - ↳ Poisson
  - ↳ Exponential.

→ Priors

↳ Informative → Conjugate priors

↳ Non - Informative

↳ weakly informative

→ Posteriors

↳ Prior predictive distribution.

– posterior prediction.

Normal distribution prior

Likelihood $p(y|N(\mu, \sigma^2)) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$

Assume known variance but variance mean $\mu$.

Prior (Conjugate) → $e^{A\theta^2 + B\theta + C}$ $\sim N(\mu_0, \tau_0^2)$ ($\mu_0, \tau_0$ are hyperparameter)

Since the likelihood is normal, so the prior is also like this.

$$p(\theta) = \frac{1}{\sqrt{2\pi}\tau_0} e^{-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2} \sim N(\mu_0, \tau_0^2)$$

Posterior

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

where $\theta = \mu$

$$= \frac{1}{2\pi\sigma\tau_0}\, e^{-\frac{1}{2\sigma^2}(y-\theta)^2 - \frac{1}{2\tau_0^2}(\theta-\mu_0)^2}$$

March (10-11)

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

when $\theta = \mu$

$$= \frac{1}{2\pi \sigma \tau_0} e^{-\frac{1}{2\sigma^2}(y-\theta)^2 - \frac{1}{2\tau_0^2}(\theta-\mu_0)^2}$$

Now

$$\exp\left[-\frac{1}{2}\left(\frac{(y-\theta)^2}{\sigma^2} + \frac{(\theta-\mu_0)^2}{\tau_0^2}\right)\right] = \exp\left[\frac{-1}{2\sigma^2\tau_0^2}\left\{(\tau_0 y - \tau_0\theta)^2 + (\sigma\theta - \sigma\mu_0)^2\right\}\right]$$

$$= \exp\left[\frac{-1}{2\sigma^2\tau_0^2}\left\{\tau_0^2 y^2 + \tau_0^2\theta^2 - 2\tau_0^2 y\theta + \sigma^2\theta^2 + \sigma^2\mu_0^2 - 2\sigma^2\theta\mu_0\right\}\right]$$

$$= \exp\left[\frac{-1}{2\sigma^2\tau_0^2}\left\{(\tau_0^2+\sigma^2)\theta^2 - 2(\tau_0^2 y + \sigma^2\mu_0)\theta + (\tau_0^2 y^2 + \sigma^2\mu_0^2)\right]\right.$$

$$= \exp\left[\frac{-1}{2\frac{\sigma^2\tau_0^2}{(\tau_0^2+\sigma^2)}}\left\{\theta^2 - 2\frac{(\tau_0^2 y + \sigma^2\mu_0)}{\tau_0^2+\sigma^2}\theta + \frac{\tau_0^2 y^2 + \sigma^2\mu_0^2}{\tau_0^2+\sigma^2}\right\}\right]$$

Let's assume $\dfrac{\tau_0^2 y + \sigma^2 \mu_0}{(\tau_0^2 + \sigma^2)} \to \mu_1$, & $\dfrac{\sigma^2 \tau_0^2}{\sigma^2 + \tau_0^2} \to \tau_1^2$

$\dfrac{\tau_0^2 y^2 + \sigma^2 \mu_0^2}{\tau_0^2 + \sigma^2} \to z$

and $z \ne \mu_1^2$ #

$$= \exp\left[ \frac{-1}{\frac{2\sigma^2 \tau_0^2}{\tau_0^2 + \sigma^2}} \left\{ \theta^2 - 2\mu_1 \theta + z \right\} \right]$$

$$= \exp\left[ \frac{-1}{\frac{2\sigma^2 \tau_0^2}{\tau_0^2 + \sigma^2}} \left\{ \theta^2 - 2\mu_1 \theta + \mu_1^2 - \mu_1^2 + z \right\} \right]$$

$$= \exp\left[ \frac{-1}{2\tau_1^2} \left\{ (\theta - \mu_1)^2 - (\mu_1 - z)^2 \right\} \right]$$

$$\propto \exp\left( -\frac{1}{2} \frac{(\theta - \mu_1)^2}{\tau_1^2} \right)$$

## Multiparameter model

### Concept of "Nuisance parameter"

- Unknown and unobserved quantities are common in practical problem.
- Out of these one or few of these quantities are relevant.
- obtain a marginal probability distribution.
  - first obtain a the joint probability distribution over all unknowns.
  - Integrate the distribution over the unknown that are not in area of interest
  - These unknown are called nuisance parameters

## Averaging over nuisance parameter

$p(\theta_1, \theta_2 / y) \rightarrow$ given

→ Suppose we want to find out $p(\theta_1, y)$

$$p(\theta_1 / y) = \int_{\theta_2} p(\theta_1, \theta_2 / y)\, d\theta_2 \propto \int_{\theta_2} p(y / \theta_1, \theta_2)\, p(\theta_1, \theta_2)\, d\theta_2$$

→ Consider two coins $C_1$ & $C_2$ with the following characteristics.
$P(\text{Head} / C_1) = 0.6$ & $P(\text{Head} / C_2) = 0.4$.
Choose one of the coin at random and spin it repeatedly. Now we observe that first two spin results is "Tails". Now what is expectation of no of additional spin until Head shows up?

## Normal Data with non-Informative prior

- Consider a vector $y$ of $n$ independent observations from a universal univariate normal distribution $N(\mu, \sigma^2)$
- Consider a non-informative prior $p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$
- finding the joint posterior density,

$$p(\mu, \sigma^2 / y) \propto p(y / \mu, \sigma^2) \cdot p(\mu, \sigma^2)$$

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

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

$$\propto (\sigma)^{-n-2} \exp\left(\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \bar{y} + \bar{y} - \mu)^2\right)$$

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

$$= \propto (\sigma)^{-n-2} \exp\left(\frac{-1}{2\sigma^2}\left(\sum_{i=1}^{n}(y_i - y)^2 + \sum_{i=1}^{n}(\bar{y}-\mu)^2 + \cancel{2\sum(y_i-\bar{y})(\bar{y}-\mu)}^{\,0}\right)\right)$$

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

$$= (\sigma)^{-n-2} \exp\left(\frac{-1}{2\sigma^2}\left((n-1)s^2 + n(\bar{y}-\mu)^2\right]\right.$$

Ans

## Acceptance / Rejection Technique

- Consider an alternate distribution (A) whose inverse CDF is solvable.
- Generate sample $n$ based on this alternate distribution

$$\frac{p_0(x)}{p_A(x)}$$ , if this ratio is greater than a threshold then accept else reject.

## Normal Data with non informative prior

$$p(\mu, \sigma^2 | y) \propto p(y | \mu, \sigma^2)\, p(\mu, \sigma^2)$$

$$= (\sigma)^{n-2} \exp\left(\frac{-1}{2\sigma^2}\left((n-1)s^2 + n(\bar{y} - \hat{\mu})^2\right)\right) \quad ①$$

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

$$\mu \sim N(\bar{y}, \sigma^2/n) \quad ②$$

Now Acceptance / Rejection continued

(b) for a given $\sigma^2$ draw $\mu$ from ②

Now we have $\mu$ and $\sigma^2$ which will follow ①.

$$p(\tilde{y} | y) \propto \int_{\sigma^2}\int_{\mu} p(\tilde{y}, \mu, \sigma^2 | y)\, d(\sigma^2)\, d\mu$$

$$= \int_{\sigma^2}\int_{\mu} p(\tilde{y} | y, \sigma^2, \mu) \cdot p(\mu, \sigma^2 | y)\, d\sigma^2 = \int_{\sigma^2}\int_{\mu} N(\tilde{y}; \mu, \sigma^2)$$

The integration is difficult to calculate so we will use sampling method.

— first draw samples from joint posterior distribution of $p(\mu, \sigma^2 | y)$

— Given $\mu$ and $\sigma^2$, find $N(\tilde{y} | \mu, \sigma^2)$

ie $\tilde{y} \sim N(\mu, \sigma^2)$.

## Multinomial Models for Categorical Data

— Data ~~from~~ for which each observation has K possible outcomes.

— If $y$ is the vector of the counts of number of observations of each outcome.

$$p(y|\theta) \propto \prod_{j=1}^{K} \theta_j^{y_j} \; ; \; \underbrace{\sum_{j=1}^{K} \theta_j = 1, \quad \sum_{j=1}^{K} y_j = n}_{\text{constraints}}$$

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

$$p(\theta|y) \propto \prod_{j=1}^{K} \theta_j^{\alpha_j + y_j - 1}$$

for multinomial conjugate pair, Dirichlet distribution

$$\sim D(\alpha_1 + y_1, \alpha_2 + y_2, \ldots, \alpha_K + y_K)$$

$$D(\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_K)$$

$$\propto \prod_{j=1}^{K} \theta_j^{\alpha_j - 1}$$

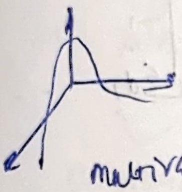
for Binomial conjugate pair

$$\text{Beta}(\alpha \mid \beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

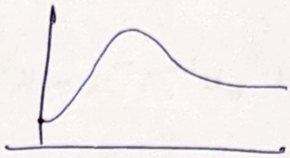
## Multivariate Normal model

### Likelihood

for univariate case $P(y : \mu, \sigma^2) \propto \frac{1}{\sigma} \exp\left(\frac{-1}{2\sigma^2}(x-\mu)^2\right)$



multivarian



univariate.

Suppose $y$ is an observable quantity with $d$ components

$y \in \mathbb{R}^d$. $P(y \mid \mu, \Sigma) \sim N(\mu, \Sigma)$

$\in \mathbb{R}^d$

$$\propto |\Sigma|^{-1/2} \exp\left(\frac{-1}{2}(y-\mu)^T \Sigma^{-1}(y-\mu)\right)$$

$$\Sigma = \begin{bmatrix} \text{Cov}(y_1, y_1) & \cdots & \text{Cov}(y_1, y_d) \\ \vdots & & \vdots \\ \text{Cov}(y_d, y_1) & \cdots & \text{Cov}(y_d, y_d) \end{bmatrix}$$

Symmetric positive definite matrix.

Suppose you are given $n$ observations.

$y_1, y_2, \ldots, y_n$

and Likelihood

$$p(y_1, y_2, \ldots, y_n \mid \mu, \Sigma) \propto |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (y_i - \mu_i)^T \Sigma^{-1} (y_i - \mu_i)\right)$$

$$\begin{bmatrix} \hat{y_1} \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1d} \\ \vdots & & & \\ \Sigma_{d1} & & \cdots & \Sigma_{dd} \end{bmatrix}^{-1} \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}$$

$= \text{trace}(\Sigma^{-1} S_0)$, where $S_0 = \sum_{i=1}^{n} (y_i - \mu)(y_i - \mu)^T$.

Conjugate Analysis

- Assume we dont know $\mu$, $\Sigma$ is known
- Prior for $\mu$ is assumed to be Normal,

$\mu \sim N(\mu_0, \Lambda_0)$

$\in \mathbb{R}^d \quad \in \mathbb{R}^{d \times d}$

posterior

$$p(\mu|y,\Sigma) \propto p(y|\mu,\Sigma)\,p(\mu|\Sigma)$$

$$\propto \mathcal{N}(y,\mu,\Sigma)\,\mathcal{N}(\mu,\mu_0,\Lambda_0).$$

$$\propto \exp\Big(-\frac{1}{2}\Big[\underbrace{(\mu-\mu_0)^T\Lambda_0^{-1}(\mu-\mu_0)}\,+ \sum_{i=1}^{n}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)\Big]\Big)$$

$$(\mu-\mu_0)^T\Lambda_0^{-1}(\mu-\mu_0)$$

$$\underbrace{\mu^T\Lambda_0^{-1}\mu}_{\text{Quadratic}} \underbrace{- \mu_0^T\Lambda_0^{-1}\mu - \mu^T\Lambda_0^{-1}\mu_0}_{\text{Scalar Quant}} + \mu_0^T\Lambda_0^{-1}\mu_0$$

$$\boxed{-\mu^T\Lambda_0^{-1}\mu_0 - \mu^T\Lambda_0^{-1}\mu_0}$$

$$= \mu_0^T\Lambda_0^{-1}\mu - 2\mu_0^T\Lambda_0^{-1}\mu_0 + \mu_0^T\Lambda_0^{-1}\mu_0\Big] \quad \text{--- (I)}$$

Quadratic in term of $\mu$ — Linear in term of $\mu$ — Constant in term of $\mu$.

2nd par

$$\sum_{i=1}^{n}(y_i-\mu)^T\Sigma^{-1}(y_i-\mu)$$

$$= \sum_{i=1}^{n}\Big[\mu^T\Sigma^{-1}\mu - 2\mu^T\Sigma^{-1}y_i + y_i^T\Sigma^{-1}y_i\Big]$$

$$= n\mu^T\Sigma^{-1}\mu - 2n\mu^T\Sigma^{-1}\bar{y} \quad + \text{constant} \quad \text{--- (II)}$$

from (i) + (ii)

$$\mu^T \Lambda_0^{-1} \mu + n\mu^T \Sigma^{-1} \mu - 2\mu^T (\Lambda_0^{-1} \mu_0 + n \Sigma^{-1} \bar{y}) + \text{constant}$$

$$= \mu^T (\Lambda_0 + n\Sigma^{-1}) \mu - 2\mu^T (\Lambda_0^{-1} \mu_0 + n \Sigma^{-1} \bar{y}) + \text{const}$$

$$\Lambda_n^{-1} = \Lambda_0^{-1} + n\Sigma^{-1}$$

$$\Lambda_n^{-1} \mu_n = (\Lambda_0^{-1} \mu_0 + n \Sigma^{-1} \bar{y})$$

$$\mu_n = \Lambda_n (\Lambda_0^{-1} \mu_0 + n \Sigma^{-1} \bar{y}) \text{ for}$$

$$= (\Lambda_0^{-1} + n\Sigma^{-1})^{-1} (\Lambda_0^{-1} \mu_0 + n\Sigma^{-1} \bar{y})$$

$$\sim N(\mu, \mu_n, \Lambda_n)$$

Posterior distribution of $\mu$ in a multivariate Normal distribution

(i) with known $\Sigma$. Assuming prior $\mu \sim N(\mu_0, \Lambda_0)$

$$p(\mu|\Sigma) \quad p(\mu|y,\Sigma) \sim N(\mu_n, \Lambda_n)$$

$$\mu_n = (\Lambda_0^{-1} + n\Sigma^{-1}) (\Lambda_0^{-1} \mu_0 + n\Sigma^{-1} \bar{y})$$

$$\Lambda_n = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1}$$

Posterior conditional and marginal distribution of subvectors of $\mu$ with known $\Sigma$:

we know $y \in \mathbb{R}^d$, $\mu \in \mathbb{R}^d$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \begin{matrix} \rightarrow \text{known} \\ \\ \rightarrow \text{unknown} \end{matrix}$$

→ we want to find their distribution.

- marginal distribution of a subset of parameters (say $\mu^{(1)}$)
- Given a known subset ($\mu^{(2)}$)
- A dataset $y$
- We want $\mu^{(1)} \mid \mu^{(2)}, y$

This will result in an multivariate normal distribution

$$\mu^{(1)}, \mu^{(2)} \mid y \sim N\left[\mu_n^{(1)} + \beta^{1/2}(\mu^2 - \mu_n^2),\ \Lambda^{1/2}\right]$$

$\mu_n \rightarrow$ Posterior mean vector

$\mu_n^{(1)} \rightarrow$ Appropriate ~~subtraction~~ subvector of $\mu_n$ with similar grouping as $\mu^{(1)}$

$\mu_n^{(2)} \rightarrow$ Appro~~ximate~~priate subvector of $\mu_n$ with similar grouping as $\mu^{(2)}$

$\Lambda_n \rightarrow$ appro~~xim~~priate submatrix of the given $n$ variance matrix

$$\Lambda^{(1|2)} = \begin{bmatrix} \Lambda_n^{11} & \Lambda_n^{12} \\ \Lambda_n^{21} & \Lambda_n^{22} \\ \Lambda_n^{31} & \Lambda_n^{32} \end{bmatrix}$$

$$\beta^{1/2} = \Lambda_n^{(12)} \left(\Lambda_n^{(22)}\right)^{-1}$$

$$\Lambda^{1/2} = \Lambda_n^{(11)} - \Lambda_n^{(12)} \left(\Lambda_n^{(22)}\right)^{-1} \Lambda_n^{(21)}$$

## Scientific problem

### Case study

- Adminstrators are testing various level of dose of a drug to some batch of test animals.

- The animal response is characterized by a dichotomary outcome (alive and dead)

- Data of the following form

$(x_i, n_i, y_i)$

$x_i$ — $i$th dose level

$n_i$ — no of animals to which $i$th dose level is given

$y_i$ — How many of the $n_i$ animal dead

sample data for test conducted on 20 animals

| Dose ($x_i$) in log (gm/ml) | No of animals | No of died |
|---|---|---|
| 0.86 | 5 | 0 |
| 0.3 | 5 | 1 |
| 0.05 | 5 | 3 |
| 0.73 | 5 | 5 |

- we want to ~~of~~ establish a dose-response relation,
- Assume that the outcomes of 5 animals within each group is exchangable.
- Reasonable to model than as "independent with equal probability.

let $\theta_i$ be the probability for group i.

So $y_i | \theta_i \sim Bin(n_i, \theta_i)$ — (1)

↳ prob. of death given dose $x_i$

- Assume that the Outcome probability of each group is independent of each other

- simple assumption is $\theta_1, \theta_2, \theta_3, \theta_4$ are exchangeable

$p(\theta_1, \theta_2, \theta_3, \theta_4) \propto 1$

Assuming $\theta_i$ in Beta posterior

Assuming $\theta_i = \alpha + \beta x_i$

$\alpha, \beta \to$ parameters of model

- Problem with $x_i \to \infty$, $\beta x_i \to \infty$ but $\theta_i \in [0, 1]$

So let $\text{logit}(\theta_i) = \alpha + \beta x_i$

$$\log\left(\frac{\theta_i}{1-\theta_i}\right) = \alpha + \beta x_i$$

$$\frac{\theta_i}{1-\theta_i} = e^{\alpha + \beta x_i}$$

$$\frac{1-\theta_i}{\theta_i} = e^{-(\alpha + \beta x_i)} \Rightarrow \theta_i = \frac{1}{1 + e^{-(\alpha + \beta x_i)}}$$

Likelihood for group $i$

$$p(y_i \mid \alpha, \beta, n_i, x_i) \propto \theta^{y_i} (1-\theta)^{n_i - y_i}$$

where $\theta_i = \dfrac{1}{1+e^{-(\alpha+\beta x_i)}} = \text{logit}^{-1}(\alpha+\beta x_i)$

$$p(y_i \mid \alpha, \beta, n_i, x_i) \propto \left[\text{logit}^{-1}(\alpha+\beta x_i)\right]^{y_i} \left[1 - \text{logit}^{-1}(\alpha+\beta x_i)\right]^{n_i - y_i}$$

Joint posterior of $\alpha, \beta$

$$p(\alpha, \beta \mid y_i, n_i, x_i) \propto p(y_i \mid \alpha, \beta, n_i, x_i)\, p(\alpha, \beta)$$

$$p(\alpha, \beta \mid y, n, x) \propto p(\alpha, \beta)\, p(y \mid \alpha, \beta, n, x)$$

$$p(\alpha, \beta \mid y, n, x) \propto p(\alpha, \beta) \prod_{i=1}^{K} p(y_i \mid \alpha, \beta, n_i, x_i)$$

As likelihood is complex, we have to rely on computation

Assumming $p(\alpha, \beta) \propto 1$ i.e. constant

for sampling

- we need to have some initial estimate of the parameter.
- Get an estimate of $\alpha$ and $\beta$ by fitting a logistic regression with MLE.

estimated $(\hat{\alpha}, \hat{\beta}) = (0.8, 7.7)$

$se(\hat{\alpha}) = 1.02$, $se(\hat{\beta}) = 4.9$

→ to find $p(\theta)$

→ start with an initial estimate of $(\hat{\alpha}, \hat{\beta})$ by fitting with a LR (Logistic Regression) using MLE.

→ Consider a range of values of $[\alpha, \beta]$

Take small values of increment and get the posterior; get the contour plot with the grid value.

$$p(\alpha, \beta \mid x_i, n_i, y) = p(\alpha, \beta)\, p(y \mid \alpha, \beta, n_i, x_i)$$

$$\propto \prod_{i=1}^{k} p(y_i \mid \alpha, \beta, x_i, n_i)$$

$$\propto \prod_{i=1}^{k} \theta_i^{y}(1-\theta_i)^{n_i - y_i}$$

$$\alpha_{new} = \alpha_{old} + \delta\alpha_{old}$$

grid size.

(3) The distribution that we get is unnormalized, so a normalization technique is required. Sampling technique can be used.

# Sampling from the joint posterior $p(\alpha, y_i, n_i, x_i)$

(1) Compute the marginal posterior distribution of $\alpha$ by numerically summing over $\beta$.

→ in the discrete distribution computed over the grid.

$$p(\alpha \mid y, n, x) \simeq \sum_{\beta=1}^{n} p(\alpha, \beta \mid y, n, x)$$

(Q)

(2) for $s = 1$ to $1000$

(a) Draw $\alpha^s$ directly from the directly computed $p(\alpha, p|y, n, n)$

(b) Draw $\beta^s$ from the discrete conditional distribution $p(\beta|\alpha, y)$

(c) for each sample $\alpha^s$ and $\beta^s$ add a random fixed $e$, $E(e) = 0$ and it is uniform.

$e = U[-\mu, +\mu]$

## Introduction to Bayesian Computation Technique

(a) Computation of the posterior $p(\theta|y)$

(b) Computation of the posterior predictive distribution $p(\tilde{y}|y)$

- standardized distribution are easier to deal with like normal, poisson, exponential, Gamma.

## Normalized & Unnormalized

The distribution (which may be multivariate) that needs to be computed (simulated) is called Target Distribution.

Let it be denoted by $p(\theta|y)$

Assume $p(\theta|y)$ to be easily computed for any value of $\theta$ upto a factor only involving $y$.

we assume that there is some easily computable function $q(\theta|y)$
which is easily computed & is unnormalized
density of $p(\theta|y)$.

$\downarrow$
Proposal distribution

$\frac{q(\theta|y)}{p(\theta|y)}$ is a constant that only depends on $y$.

and is proportional to the posterior density.