Веска Нончева

Мария Добрева

Съвременни статистически методи за вземане на решения

Това изследване е частично подкрепено от проект No: СП15-ФМИ ИТ-015, ПУ ”Паисий Хилендарски“

2016

Пловдив

Съвременни статистически методи за вземане на решения

Съдържание

Глава 1.
Еднопараметрични модели. Бейсови изводи

Томас Бейс е британски математик и презвитериански пастор, формулирал частен случай на Формулата на Бейс, която се съдържа в посмъртно издадените му трудове. Решението на Бейс на частния случай на задачата за обратната вероятност се съдържа в неговото произведение *„Съчинение за решаването на задачи, свързани с метода на случайностите“*, публикувано посмъртно от неговия приятел Ричард Прайс, в издание на Британското кралско научно дружество.

През първото десетилетие на 18 век се намира решение на следната задача: ако в кутия има определен брой бели и черни топки, каква е вероятността да бъде изтеглена бяла топка? Скоро след това, вниманието на учените е привлечено от обратната задача: ако са изтеглени определен брой бели и черни топки, каква е вероятността в кутията да е имало определен брой бели топки? Задачата е поставена от Абрам де Моавър в неговата книга *„Методът на случайностите“*, а Томас Бейс решава подобна задача, като за първи път прилага открития от него частен случай на формулата, наречена по-късно на негово име.

1.1 Биномен модел

Нека да разгледаме редица от Бернулиеви опити $y_1, y_2, \dots, y_n$; $y_i = \{0,1\}$,(Томас Бейс, 1763г.)

Данните могат да бъдат обобщени чрез случайната величина y -„броя на успехите“ в n опита, проведени по схемата на Бернули.

Биномният извадков модел е:

$$P(y \mid \theta) = Bin\,(y \mid n, \theta) = C_n^y \theta^y (1-\theta)^{n-y}, \qquad (1.1)$$

θ – вероятността за успех, постоянна във всеки опит,

n е част от планирането на експеримента и е константа.

За да направим бейсови изводи за случайната величина θ първо трябва да определим априорното вероятностно разпределение $P(\theta)$ на случайната величина θ. То може да бъде равномерно непрекъснато в [0,1].

Процесът на бейсовия анализ включва преминаване от априорно $P(\theta)$ към апостеорно $P(\theta|y)$ разпределение. Можем да очакваме, че има връзки (зависимости) между тези две разпределения.

При априорното разпределение нямаме информация от данните.

При определяне на апостеорното разпределение е отчетена информацията, която носят данните. Затова очакваме дисперсията на апостеорното разпределение да бъде по– малка от дисперсията на априорното разпределение.

$$\mathrm{E}\theta = \mathrm{E}(\mathrm{E}(\theta|y)),$$

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

Апостеорното разпределение съдържа цялата налична информация за θ.

Може да обобщим разпределенията чрез техните числови параметри:

- Средното –очакваната стойност на параметъра,
- Модата – най- вероятната стойност на параметъра.

Задача. Да се оцени вероятността да се роди момиче θ или отношението:

$$\phi = \frac{1-\theta}{\theta},$$

y – брой на момичетата в n бернулиеви опита.

Стъпка 1: Трябва да определим апостеорното вероятностно разпределение за случайната величина θ. Предполагаме, че априорното разпределение на θ е равномерно в интервала [0,1].

$$E\theta = \frac{1}{2},$$

$$D\theta = \frac{1}{12}.$$

От (1.1): $P(\theta|y) \propto \theta^y (1-\theta)^{n-y}$,

при фиксирани n и θ.

$$P(\theta|y) = \frac{P(\theta)\ .\ P(y|\theta)}{P(y)} = \frac{P(\theta)}{P(y)} {C_n}^y \theta^y (1-\theta)^{n-y}.$$

Следователно апостеорната вероятност за θ е :

$P(\theta|y) = \theta^y(1-\theta)^{n-y}$, (1.2)

- бета вероятностно разпределение $Beta(y+1, n-y+1)$,

$E(\theta|y) = \frac{y+1}{n+1}$.

Апостеорното разпределение е около точката, която представя компромиса между априорната информация и данните.

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

$\mu = \frac{\alpha}{\alpha+\beta} = \frac{y+1}{y+1+n-y+1} = \frac{y+1}{n+2}$,

$mode = \frac{\alpha-1}{\alpha+\beta-2} = \frac{y+1-1}{y+1+n-y+1-2} = \frac{y}{n}$.

За апостеорната несигурност:

- област на най-високата апостеорна плътност:
- плътността в тази област никога не е по – малка от плътността извън тази област,
- централен апостеорен интервал.

Априорното разпределение на θ може да е и $Beta(\alpha, \beta)$.

1.2 Информативни априорни разпределения

Разгледахме равномерно априорно разпределение за неизвестния параметър θ.

Априорното разпределение представя една популация от възможни стойности на θ.

Ние трябва да представим нашите знания за θ.

Функцията на правдоподобие (като функция на θ) има вида:

$P(y|\theta) \infty \theta^a(1-\theta)^b$.

Априорната и апостеорната плътност имат вида:

$P(\theta) \infty \theta^{\alpha-1}(1-\theta)^{\beta-1}$, която е $Beta(\alpha, \beta)$.

$P(y|\theta) \infty \theta^y(1-\theta)^{n-y}$, т.е априорната плътност води до α −1 успехи и β −1 неуспехи.

Хиперпараметри – параметрите на априорното разпределение.

Апостеорното разпределение за θ е:

$$P(\theta|y) \propto P(\theta) \,.\, P(y|\theta),$$

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

$$= Beta(\alpha+y, \beta+n-y).$$

Свойството, че априорното и апостеорното разпределение имат една и съща форма се нарича спрегнатост (conjugacy).

От $P(\theta|y) = Beta(\alpha+y, \beta+n-y)$ следва, че апостеорното средно е:

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

и апостеорната дисперсия е:

$$D(\theta|y) = \frac{(\alpha+y)\,(\beta+n-y)}{(\alpha+y+\beta+n-y)^2\,(\alpha+y+\beta+n-y+1)} = \frac{(\alpha+y)\,(\beta+n-y)}{(\alpha+\beta+n)^2\,(\alpha+\beta+n+1)} = \frac{E(\theta|y)\,[1-E(\theta|y)]}{\alpha+\beta+n+1},$$

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

Следователно намерихме апостеорното средно и апостеорната дисперсия.

Когато α и β са фиксирани, y и $(n-y)$ растат:

$$\mathrm{E}(\theta|y) \approx \frac{y}{n},$$

$$\mathrm{D}(\theta|y) \approx \frac{y(n-y)}{n^2(n+1)} \approx \frac{1}{n}\frac{y}{n}\frac{(n-y)}{n} \approx \frac{1}{n}\frac{y}{n}\left(1-\frac{y}{n}\right).$$

Следователно параметрите на априорното разпределение α и β нямат влияние на граничното апостеорно разпределение.

Спрегнатите априорни разпределения могат да се интерпретират като допълнителни данни (допълнителна информация) за апостеорното разпределение.

Съществуват и не спрегнати априорни разпределения.

Глава 2.
Йерархични модели

Имаме задача, която се моделира с модел с много параметри. Съвместното вероятностно разпределение на тези параметри отразява зависимостта между тези параметри. Ще използваме едно априорно разпределение, в което на параметрите се гледа като на извадка от едно популационно разпределение. Наблюдаваните данни могат да се използват, за да се оцени популационното разпределение на параметрите даже когато те са ненаблюдаеми. Наблюдаваните данни се моделират условно при дадените параметри.

В 2.1.1, ще създадем априорно разпределение като използваме йерархичните принципи. Ще анализираме един опит. Ще използваме предишните данни, за да създадем априорно разпределение.

В 2.1.2, ще създадем йерархично априорно разпределение в контекста на пълния бейсов анализ.

2.1 Създаване на априорно разпределение

2.1.1 Конструиране на априорно вероятностно разпределение, чрез анализиране на един експеримент в контекста на историческите (предишните) данни.

Ще оценим параметъра θ, като използваме данни от един опит и априорно разпределение, създадено от подобни, вече проведени опити.

Предполагаме, че данните от настоящия и предишните опити са една случайна извадка от една популация.

Пример: Ще оценим вероятността за тумори в група от плъхове.

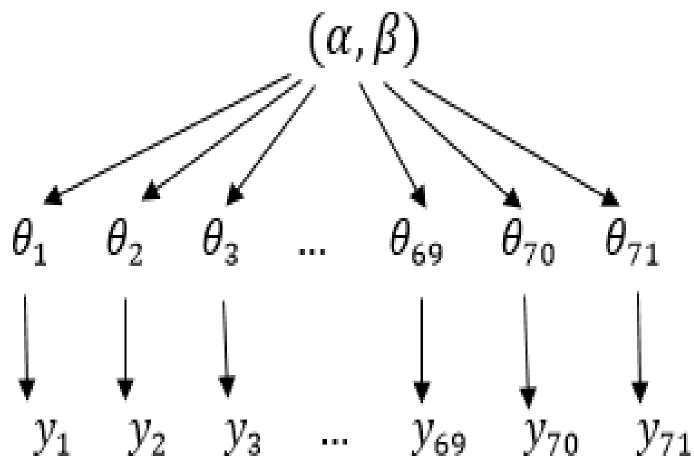
За да оценим лекарство, периодически извършваме проучвания върху гризачи.

Целта е да оценим θ – вероятността за тумори на плъхове, които не получават лекарство (контролна група). Данните показват, че 4 от 14 плъха са развили тумори. Броят на туморите е биномно разпределена случайна величина с параметър θ.

Избираме априорно разпределение на θ от спрегнатата фамилия от разпределения $Beta(\alpha,\beta)$. Т.е $\theta \sim Beta(\alpha,\beta)$.

От предишни данни предполагаме, че знаем, че вероятността за тумори θ има приблизително $Beta$ разпределение с известно средно и известно стандартно отклонение. Вероятността $Beta$ се променя и зависи от плъховете и от условията по време на опитите.

Предполагаме, че $Beta(\alpha,\beta)$ е добър модел за популационното разпределение на θ_j от историческите данни. Тогава йерархичния модел можем да представим по следния начин:



където:

$\theta_1,\dots,\theta_{70}$ и $y_1,\dots,y_{70}$ – са историческите данни,

θ_{71} и y_{71} – са данните от настоящия опит.

Като знаем средното и стандартното отклонение на $Beta$ разпределението, можем да намерим α и β.

Следователно знаем априорното разпределение на θ и то е $Beta(\alpha,\beta)$.

Тогава апостеорното разпределение на θ е:

$$Beta(\alpha+4,\ \beta+10). \qquad (2.1)$$

Но обикновено средното и стандартното отклонение на вероятността за тумор θ не са известни. Имаме данни от предишни опити – 70 групи от плъхове. Нека в j – тия опит броя на плъховете с тумори е y_j , а общия брой на плъховете е n_j.

Моделираме y_j, като независими биномно разпределени случайни величини при даден обем на извадката n_j и специфично за опита θ_j.

За извадката от 70 плъха: $\bar{X} = 0.136$ и $S = 0.103$.

Тогава за α и β получаваме: $\alpha = 1.4,\ \ \beta = 8.6$.

От (2.1) получаваме апостеорното разпределение за θ_{71} и то е $Beta(5.4, 18.6)$. Тогава апостеорното средно е 0.223 и апостеорното стандартно отклонение е 0.08.

Като използвахме априорното разпределение получихме апостеорното средно, което е значително по-малко от пропорцията: $\frac{4}{14} = 0.286$.

Използвахме 70 предишни опита, за да определим априорното разпределение на θ_{71}. *Можем ли да използваме това априорно разпределение, за да получим Бейсови изводи за $\theta_1, \theta_2, \ldots, \theta_{70}$?*

Има много по-голям смисъл да оценим популационното разпределение, като използваме всички данни $y_1, y_2, \ldots, y_{71}$ и след това да оценим всяко θ_j .

2.1.2 Взаимозаменяемост и създаване на йерархични модели

Постановка на задачата: Нека $j = 1, 2, \ldots, J$ са едно множество от опити; y_j – данни от опита j; θ_j – параметър в опита j, с функция на правдоподобие $P(y_j|\theta_j)$,

y_j и θ_j – могат да са вектори.

Ще създадем съвместен вероятностен модел за всички параметри θ. Ще използваме идеята за взаимозаменяемост.

Взаимозаменяемост: Ако нямаме друга информация освен данните y, която би ни помогнала да различим θ_j от другите параметри и не можем нито да подредим, нито да групираме параметрите, можем да предположим симетрия в параметрите в априорното им разпределение, т.е. параметрите са взаимнозаменяеми в тяхното съвместно вероятностно разпределение $P(\theta_1, \theta_2, \ldots, \theta_J)$. Или съвместното вероятностно разпределение на параметрите е инвариантно по отношение на пермутацията на индексите (1, 2, …, J).

Колкото по-малко знаем за задачата, толкова по-убедително можем да предположим взаимозаменяемост.

На параметрите $\theta_1, \theta_2, \dots, \theta_J$ гледаме като на случайна извадка от априорно разпределение с неизвестен параметър ϕ.

$$P(\theta) = \int P(\theta|\phi)\, P(\phi)\, d\phi = \int (\prod_{j=1}^{J} P(\theta_j|\phi)\, P(\phi)\, d\phi). \quad (2.2)$$

Когато $J \to \infty$ всяко взаимнозаменяемо разпределение на $(\theta_1, \theta_2, \dots, \theta_J)$ може да бъде изразено като независими и еднакви разпределения по начина (2.2).

Ние вече сме използвали взаимнозаменяемите модели за данните $y_1, y_2, \dots, y_n$ във функцията на правдоподобие, където наблюденията са независими и еднакво разпределени при даден параметър.

Ако в задачата за туморите на плъховете, част от експериментите бяха направени в друга лаборатория, ние можехме да предположим частична взаимозаменяемост и да използваме йерархичен модел.

В примера с туморите на плъховете само размерът на извадката е наличната информация, която различава опитите. Можем да построим взаимнозаменяем модел за двойките (n, y)$_j$. Една естествена първа стъпка би била да начертаем плот $\frac{y_j}{n_j}$ срещу n_j , за да видим връзката, която можем да моделираме. Но за нашия пример такъв плот не показва никаква връзка.

Необходими са модели с две нива, за да моделираме вариациите в лабораториите и вариациите между лабораториите.

Извод: Ние моделираме взаимозаменяемостта на променливите (covariates) чрез условната независимост:

$$P(\theta_1, \dots, \theta_J | x_1, \dots, x_J) = \int [\prod_{j=1}^{J} P(\theta_j | \phi, x_j)]\, P(\theta|x)\, d\phi,$$

$x = (x_1, \dots, x_J)$.

Така взаимнозаменяемите модели са универсални и навсякъде приложими.

В задачата за плъховете опитите се различават. Следователно $\boldsymbol{\theta_j}$ се различават. Но можем да предположим, че са извадка от едно и също вероятностно разпределение.

2.2 Разбиране на бейсовия йерархичен модел

Параметъра θ не е известен. $P(\phi)$ е неизвестно априорно разпределение.

Подходящото бейсово апостеорно разпределение е това на вектора (ϕ, θ).

Съвместното априорно разпределение на ϕ и θ е:

$P(\phi, \theta) = P(\phi) \,.\, P(\theta|\phi)$.

Съвместното апостеорно разпределение е:

$$P(\phi, \theta \,|y) \propto P(\phi, \theta) \,.\, P(y\,|\phi, \theta) = P(\phi, \theta) \,.\, P(y\,|\theta), \quad (2.3)$$

защото разпределението на данните е:

$$P(y\,|\phi, \theta) = P(\phi) \,.\, P(\theta\,|\phi) \,.\, P(y\,|\theta), \quad (2.4)$$

и зависи само от θ, т.е. хиперпараметрите ϕ влияят на данните y само чрез θ.

В примера за туморите на плъховете хиперпараметрите са α и β, които определят $Beta$ разпределението на θ.

2.3 Апостеорни предсказващи разпределения

Йерархичните модели се определят чрез хиперпараметрите ϕ и параметрите θ. Две са апостеорните предсказващи разпределения, от които се интересуваме:

1) Разпределението на бъдещи наблюдения $\widetilde{y}$, съответстващо на известно θ_j ($\widetilde{y}$ може да бъде допълнителен плъх в някои от проведените опити),
2) Разпределението на бъдещи наблюдения $\widetilde{y}$ съответстващо на бъдещо θ_j от същата подпопулация, т.е. разпределението на $\widetilde{y}$ при $\tilde{\theta}$ (плъхове от бъдещи опити).

И 1) и 2) могат де се използват за да оценим адекватността на модела.

В 1) апостеорната предсказваща извадка $\widetilde{y}$ се базира на апостеорната извадка на θ_j от съществуващия опит.

В 2) първо трябва да генерираме $\tilde{\theta}$ за нов опит от популационното разпределение при дадена апостеорна извадка на ϕ, и след това да генерираме извадка $\widetilde{y}$ при дадено симулирано $\tilde{\theta}$.

Ще представим подход, който обединява аналитични и числени методи за да получим симулации от съвместно апостеорно разпределение $P(\theta, \phi|y)$.

В примера за туморите на плъхове, разгледахме *Beta–Binominal* модел. В този модел популационното разпределение $P(\theta\,|\phi)$ е спрегнато на функцията на правдоподобие $P(y|\theta)$.

Подходът с използването на спрегнати разпределения е полезен за намиране на приближени оценки и начални точки.

2.3.1 Намиране на условни и маргинални разпределения по аналитичен път

Започваме с 3 аналитични стъпки:

1. Записваме съвместната апостеорна плътност $P(\theta, \phi|y)$ като произведение на хиперприорното разпределение $P(\phi)$, популационното разпределение $P(\theta\,|\phi)$ и функцията на правдоподобие $P(y\,|\theta)$ (виж (2.4)).
2. Определяме аналитично условната апостеорна плътност на θ, при даден хиперпараметър ϕ и при фиксирано y, т.е. $P(\theta\,|\phi, y)$. За спрегнатите модели $P(\theta\,|\phi, y) = \prod_{j=1}^{J} P(\theta_j\,|\phi)$.
3. Оценяваме ϕ като използваме бейсовата парадигма, т.е. получаваме маргиналното апостеорно разпределение на ϕ.
   $P(\phi\,|y) = \int P(\theta, \phi\,|y)\, d\theta$.

Получаване на симулационни извадки от съвместни апостеорни разпределения $P(\theta, \phi\,|y)$ за прости йерархични модели.

1. От маргиналното апостеорно разпределение на ϕ получаваме извадка от хиперпараметри ϕ.
2. При дадената извадка за ϕ получаваме извадка за θ от нейното условно апостеорно разпределение $P(\theta\,|\phi, y)$.
3. Правим извадка от предсказаните стойности $\widetilde{y}$ от апостеорното предсказващо разпределение при дадената извадка θ.

Тези стъпки се изпълняват L пъти за да получим множество от L извадки.

В задачата за тумори на плъхове:

$J = 71$ опита,

j = 1, 2, …, J.

Предполагаме, че $y_i \sim Bin\,(n_j, \theta_j)$, където броят на туморите n_j е известен; параметрите θ_j са независими и еднакво $Beta$ разпределени случайни величини, т.е

$\theta_j \sim Beta(\alpha, \beta)$.

Изпълняваме трите стъпки, за да определим аналитичния вид на апостеорното разпределение. Съвместното апостеорно разпределение на всички параметри е:

$$P(\theta, \alpha, \beta \mid y) \propto P(\alpha, \beta) \,.\, P(\theta \mid \alpha, \beta) \,.\, P(y \mid \theta, \alpha, \beta).$$

2.4 ABC метод

1. Calibration of model parameters to certain data sets and parameters.
2. Sensitivity of model output to changes in parameters.

Parametric space – пространство на параметрите- всички възможни комбинации от стойности на параметрите.

Методът Full Factorial Design изследва параметричното пространство (планиране на експеримента). Друг подход за изследване на параметричното пространство са sampling methods.

Целта на sampling methods е силно да редуцира броя на комбинациите от параметрите.

Последователният Monte Carlo метод (ABC–MCMC) концентрира (съсредоточава) симулациите в зони на параметричното пространство, където функцията на правдоподобие приема най – високи стойности.

Markov Chain Monte Carlo (MCMC) е sampling метод, в който изборът на следващата комбинация от параметри зависи от предишната и от близостта на симулациите и наблюденията. Само първата комбинация от параметри е от априорното разпределение. Повече априорното разпределение не се използва.

Като използваме бейсовата стратегия истинските (апостеорните) плътности на параметрите пресмятаме чрез точкова апроксимация на функцията на правдоподобие в параметричното пространство.

ABC метод:

1. Изпълняваме модела много пъти с различни стойности на параметрите получени от априорното разпределение.
2. Сравняваме моделираните данни с наблюдаваните данни, като за целта използваме обобщаващи статистики, за да намалим размерността на данните. Запазваме само тези стойности на параметрите, за които разликите между моделираните данни и реалните (наблюдаваните) данни са малки.

3. Получаваме едно приближение на апостеорното разпределение на параметрите (това са запазените стойности на параметрите).

ABC методите дават най – добри оценки на стойностите на параметрите и измерват корелацията между параметрите.

Глава 3.
Бейсов подход в статистиката

Основна идея на бейсовия подход:

3.1 Функция на правдоподобие

Разрешаваме на параметрите на функцията на правдоподобие (ФП) да бъдат случайни величини, т.е. параметрите да имат вероятностни разпределения, който наричаме априорни разпределения.

Априорните вероятностни разпределения на параметрите на ФП също имат параметри, който също могат да бъдат случайни величини, т.е. да имат свое вероятностно разпределение и т.н.

По този начин по естествен път възниква йерархия на параметрите, т.е. възникват йерархичните модели.

Като използваме моделът на максималното правдоподобие, построен за данните, и подходящо априорно вероятностно разпределение на параметрите получаваме апостеорно разпределение. Апостеорното разпределение описва поведението на параметрите, след като се видят данните (т.е. след като се направят наблюденията). Т.е апостеорното разпределение описва нашата представа за поведението на параметрите. Извършваме наблюдения, събираме данните, строим модела на максимално правдоподобие и след това обновяваме нашето разбиране за поведението на параметрите като построяваме апостеорното разпределение. Това апостеорно разпределение може да послужи за априорно разпределение на параметрите преди следващото наблюдение (преди следващия опит).

Извод: ФП изразява информацията, която се съдържа в данните. Всяко събитие, което не се е случило, няма влияние върху изводите (заключенията).

Нека y_i, $i = 1,\dots, m$ – са наблюденията (данните). Функцията на правдоподобие дефинираме като:

$$L(y|\theta) = \prod_{i=1}^{m} f(y_i|\theta), \qquad (3.1)$$

θ – е p – мерен вектор, т.е. $\theta = (\theta_1, \theta_2, \dots, \theta_p)$,

$f(y_i|\theta)$ – е вероятностна плътност.

Стойностите на y при дадени параметри са независими и затова ФП има вида (3.1), т.е предполагаме, че данните са условно независими. Забележете, че данните могат да не бъдат безусловно независими, а да бъдат корелирани. Условната независимост е едно фундаментално предположение.

$$l\ (y|\theta) = \log L(y|\theta) = \sum_{i=1}^{m} \log f(y_i|\theta).$$

Често наблюденията са корелирани в пространствените модели. Идеята е, че наблюденията от места, който са близки по местоположение, имат близки стойности, докато наблюденията от места, който са отдалечени, имат рязко различаващи се стойности. Тази пространствена корелация, известна като автокорелация, е в сила в пространствения анализ (Spatial analysis). Това ще окаже влияние върху структурата и формата на ФП.

$$L(y|\theta) = \prod_{i=1}^{m} f(y_i|y_{j\neq i}, \theta). \qquad (3.2)$$

3.2 Априорни вероятностни разпределения

Всички параметри в бейсовите модели са случайни величини и имат вероятностни разпределения. Априорните разпределения са разпределения на параметъра преди да се наблюдават данните. Те добавят информация за задачата и по този начин могат да подобрят оценките на параметрите.

Априорното разпределение на параметъра θ означаваме с $g(\theta)$. Ако θ е вектор, то с $g(\theta)$ означаваме съвместното априорно вероятностно разпределение.

- Коректност (propriety) на разпределението:

Когато $\int_{\Omega} g(\theta)\, d\theta = \infty$, априорното разпределение е некоректно (improper). Некоректността на априорното разпределението не води до некоректност на апостеорното разпределение.

- Информативни и неинформативни априорни разпределения.

Априорните разпределения, които доминират над ФП са информативни. Тяхното влияние намалява когато размерът на извадката расте.

3.3 Апостеорни разпределения

Априорните разпределения и ФП са два източника за информация за задачата (изследването). ФП носи информация за параметъра чрез данните. Априорното разпределение носи информация за нашата вяра, убеденост и предположения.

Когато обемът на извадката е много голям, тогава ФП ще допринесе повече за оценката на относителния риск (relative risk estimation). Когато обемът на извадката е малък, информацията която носи априорното разпределение ще доминира.

Апостеорното разпределение се дефинира по следния начин:

$p\ (\theta|y) = \frac{L(y|\theta)\ .\ g(\theta)}{c}$, където

$c = \int L(y|\theta) g(\theta) d\theta.$

Т.е. $p(\theta|y) \infty L(y|\theta) g(\theta).$

Спрегнатост (conjugasy):

Някои комбинации от априорни разпределения и ФП водят до същите фамилии от апостеорни разпределения.

Например:

- За поасонова ФП с параметър θ и гама априорно разпределение за θ, апостеорното разпределение на θ също е гама.
- За биномна ФП и бета априорно разпределение, апостеорното разпределение е бета.
- За нормална ФП и нормално априорно разпределение за средното, апостеорното разпределение е нормално.

Това явление наричаме спрегнатост. То може да бъде невъзможно при голяма йерархия на параметрите. Но условната спрегнатост може да бъде полезна при проверка за адекватност на модела.

Знаем, че априорните разпределения, които доминират над ФП са информативни. Тяхното влияние намалява когато размерът на извадката расте.

3.4 Предсказващи разпределения

Апостеорните разпределения обобщават нашето разбиране за параметрите при дадените данни. Апостеорните разпределения имат фундаментна роля в бейсовото моделиране.

Нека y^* е ново наблюдение. Ние можем да определим разпределението на y^* по два начина:

$$p\ (y^*|y) = \int L\ (y^*|\theta) p(\theta|y) d\theta,$$

Априорното разпределение $p(\theta|y)$ определя приноса на наблюдаваните данни за предсказването.

$p\ (y^*|y)$ се нарича апостеорно предсказващо разпределение.

Апостеорното предсказващо разпределение може да се дефинира и като маргиналното разпределение на y^*, т.е.

$$p\ (y^*|y) = \int L\ (y^*|\theta)\ p(\theta) d\theta,$$

тук не отчитаме информацията, която се носи от данните, т.е. данните нямат принос за прогнозирането.

Пример: Нека $y_i,\ i = 1,2,\dots,m$ са бройките заболели хора в относително малки области. За същите области дефинираме очакваните скорости e_i и относителния риск θ. Предполагаме, че $y_i \in\ P_0\ (e_i\theta)$ са независими при дадено θ_i, p$(\theta) \in Gamma\ (\alpha, \beta)$,

$$p\ (y^*|y, \alpha, \beta) = \int L\ (y^*|\theta)\ p(\theta|\alpha, \beta) d\theta.$$

3.5 Йерархични модели

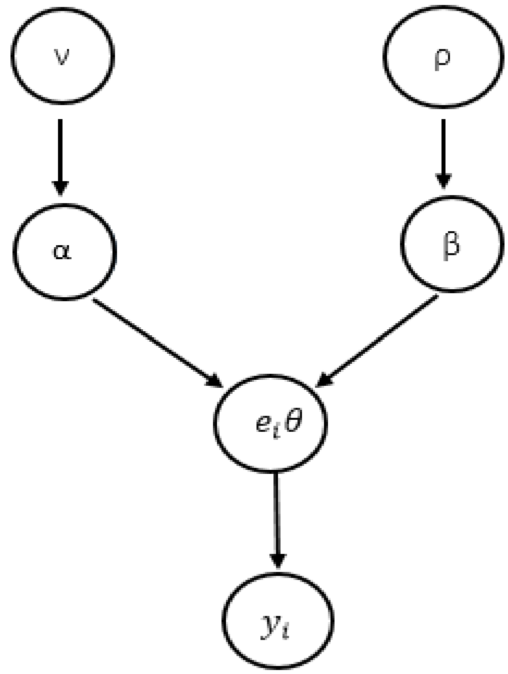
Поасон – модел на 2 нива.

$$y_i|\theta \sim P_0\ (e_i\theta),$$

$$\theta\ |\alpha\beta \sim G\ (\alpha, \beta),$$

хипер разпределения $\begin{cases} \alpha|\nu \sim h_\alpha(\nu) \\ \beta|\rho \sim h_\beta(\rho) \end{cases}.$

Йерархията на моделите може да бъде представена чрез насочени ациклични графи:



3.5.1 Апостеорни изводи

Когато използваме бейсов йерархичен модел, параметрите ще бъдат представени от апостеорните разпределения и следователно тези апостеорни разпределения трябва да бъдат намерени и оценени. При този случай ще използваме симулационни методи за да получим извадки от апостеорните разпределения. От тези извадки ще получим оценки на съответните квантили.

Пример: Имаме данни от малки географски области, т.е. популациите са крайни и данните са двоични (case/control). Нека m е броят на малките области, n_i – е обема на популацията в i-та област. Нека y_i е броят на болните и $y_i \sim B_i\,(p_i, n_i)$,

тогава ФП е : $L(y_i|p_i, n_i) = \prod_{i=1}^{m} C_{n_i}^{y_i} . p_i^{y_i}\,(1-p_i)^{n_i-y_i}$.

logit $(p_i) = \alpha_0 + \nu_i$, т.е :

$$p_i = \frac{e^{\alpha_0+\nu_i}}{1+e^{\alpha_0+\nu_i}},$$

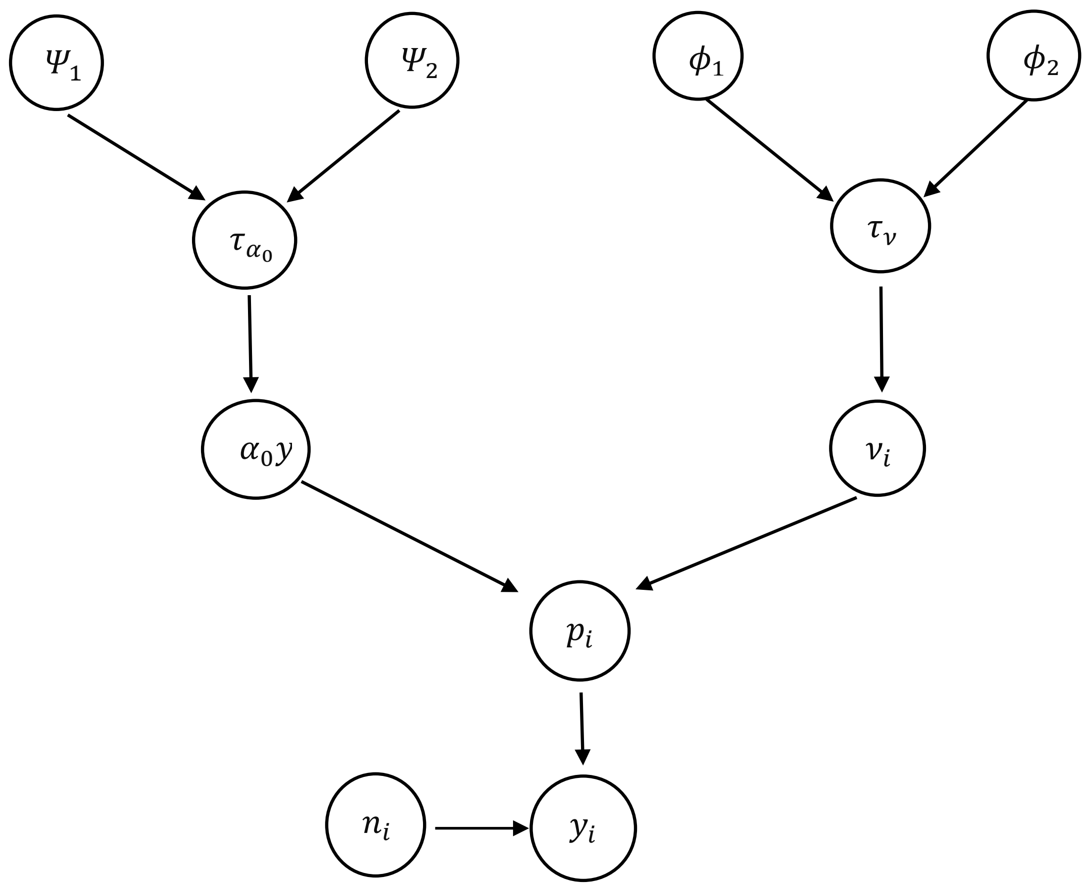
$$\alpha_0 \sim N\,(0, \tau_{\alpha_0}),$$

$$\nu_i \sim N\,(0, \tau_\nu),$$

$\tau_{\alpha_0} \sim G\ (\Psi_1, \Psi_2),$

$\tau_\nu \sim G\ (\phi_1, \phi_2).$

Йерархията на този модел може да се представи така:



3.6 Апостеорни извадки

Имаме ФП и априорни разпределения. От тях получаваме апостеорно разпределение. Ако това апостеорно разпределение трябва да бъде изследвано, то се нуждаем от данни (т.е извадка от това разпределение).

Извадките направени от апостеорното разпределение са основно средство за изучаване на това разпределение.

Нека р$(\theta|y)$ е апостеорно разпределение на данните y и параметъра –векторът θ. Правим извадка от р$(\theta|y)$ с обем m_p . Данните от тази извадка са:

$\theta_{ij}^*,\ j = 1, \dots, m_p,$

тогава можем да оценим апостеорното средно:

$$\widehat{\mathrm{E}}\theta_i = \hat{\theta}_i = \frac{\sum_{j=1}^{m_p} \theta_{ij}^*}{m_p},$$

и апостеорната дисперсия:

$$\widehat{D}\theta_i = \frac{1}{m_p - 1} \sum_{j=1}^{m_p} (\theta_{ij}^* - \hat{\theta}_i)^2.$$

Всяка функция $\gamma_j = t(\theta_j)$, може също да бъде оценена:

$$\widehat{\mathrm{E}}\gamma_j = \hat{\gamma}_j = \frac{\sum_{j=1}^{m_p} t(\theta_{ij}^*)}{m_p}.$$

Доверителните интервали също могат да бъдат построени. Маргиналната апостеорна плътност τ на θ_i също може да бъде оценена. Една Монте Карло оценка на $\pi(\theta_i)$ е:

$\hat{\pi}(\theta_i) = \frac{1}{n} \sum_{j=1}^{n} \pi(\theta_i | \theta_{j,-i})$, където

$\theta_{j,-i},\ j = 1, \dots, n$ – са извадки от маргиналното разпределение $\pi(\theta_{-i})$.

До тук извадките бяха независими.

3.7 Монте Карло методи

Често добрите бейсови модели имат две или повече нива, и апостеорните разпределения на параметрите изискват използването на извадкови алгоритми и относително сложни модели.

MCMC методите са методи, които използват симулации на стойностите на параметрите в рамките на една Марковска верига. Преобразуването на тази верига до стационарно разпределение, което приема да е апостеорно, трябва да бъде оценено.

Априорните разпределения на p компоненти на θ се дефинират като $g_i(\theta_i),\ i = 1, \dots, p$.

Апостеорното разпределение на θ и y се дефинира като:

$$\mathrm{P}(\theta|y) \propto L(y|\theta) \,.\, \prod_{i=1}^{P} g_i(\theta_i).$$

Целта е да се генерира извадка от апостеорното разпределение $\mathrm{P}(\theta|y)$.

Предполагаме, че можем да построим Марковска верига. Такива два алгоритъма са:

- Metropolis M-H метод.
- Gibbs метод.

3.8 Проверка за адекватност на модела

Ще дефинираме остатъците като стандартизирана разлика между наблюдаваните и предсказваните стойности:

$r_i = y_i - \hat{y}_i$ и $r_i{}^s = \frac{r_i}{\sqrt{\hat{D}r_i}}$.

Можем да изследваме остатъците като използваме метода Монте Карло.

Пример: Ако $y_i \sim P_0(e_i\theta_i)$, то генерираме (симулираме) $y_{ij}^*,\ j = 1, \dots, J$,

от поасоново разпределение със средно $e_i\hat{\theta}_i$. Следователно имаме множество от J+1 остатъци: $\left\{y_i - e_i\hat{\theta}_i\, ;\, \left\{y_{ij}^* - e_i\hat{\theta}_i\right\}, j = 1, \dots, J\right\}$.

3.8.1 Бейсови остатъци

Нека $\mathrm{E}(y_i|\theta_i)$ е очакваната стойност от апостеорното разпределение:

$r_i = y_i - \frac{1}{G}\sum_{g=1}^{G} E(y_i|\theta_i{}^{(g)})$, където

$\left\{\theta_i{}^{(g)}\right\}$ са стойностите на параметъра от направената извадка от апостеорното разпределение $\theta_i|y_i$. Ако $y \sim P_o(e_i\theta_i)$, то:

$r_i = y_i - \frac{1}{G}\sum_{g=1}^{G} \mathrm{e}_i\theta_i{}^{(g)}$.

3.8.2 Предсказващи остатъци

Нека $f(y^{pr}|\theta)$ е ФП на $y^{pr}|\theta$.

$f_{pr}(y^{pr}|y;\theta) = \int f(y^{pr}|\theta)\, p(\theta|y) d\theta$.

Тогава: $y_i^{pr} \sim f_{pr}(y^{pr}|y;\theta)$.

Генерираме извадка от $f_{pr}(y^{pr}|y;\theta)$.

Пример: $y_i|\theta \sim P_o(e_i\theta_i)$. На g-тата итерация (очакването ще е $e_i{\theta_i}^{(g)}$), ще бъде генерирана стойност $y_i^{pr(g)}$ от $P_o(e_i{\theta_i}^{(g)})$. Дефинираме предсказващи остатъци:

$$r_i^{pr} = y_i - y_i^{pr}.$$

Осредняваме за извадката: $r_i^{pr} = \frac{1}{G}\sum_{g=1}^{G}(y_i - y_i^{pr(g)})$.

Такива извадки могат да бъдат използвани за да генерираме:

$\{y_i^{pr}\}$, $j = 1, \dots, Y$, където $y_i^{pr} \leftarrow P_0(e_i\hat{\theta}_i)$.

Изборът на Y е между 99 и 999.

Глава 4.
Бейсов модел анализиращ ефекта от лечението на гингивална рецесия (ГР). Модел със случайни/смесени ефекти за повторени и наредени във времето измервания

4.1 Общ линеен модел (General Linear Model)

Нека y_i е случаен вектор с многомерно нормално вероятностно разпределение със средно вектора μ_i , $\mu_i(n_i \text{ x } 1)$ и ковариационна матрица Σ, $\Sigma(n_i \text{ x } n_i)$, където n_i е броят на повторените във времето наблюдения за i- тия пациент, i=1, 2, …, m.

Векторът на средните μ_i може да зависи от момента на измерване и от допълнителни променливи (covariates), които влияят и на зависимата и на независимите променливи.

В моделите на повторените във времето измервания предполагаме, че ефектът на индивида е постоянен през целия период от време, т.е няма развитие на самите индивиди. Условията на експеримента се променят.

Нека с α означим неизвестните параметри на популацията. Нека техният брой е p.

Т. е α е вектор стълб с неизвестните параметри или α(p x 1).

Нека X_i е известна матрица свързваща α с y_i. Т.е X_i $(n_i \text{ x } p)$, където n_i е броя на наблюденията за i - тия пациент, p е броя на неизвестните параметри на популацията, i =1,2,…, m.

Нека $b_i(k \text{ x } 1)$ е вектор стълб от неизвестни ефекти свързани с индивида (неизвестни ефекти на индивида), който са k на брой.

Нека Z_i е известна матрица, свързваща b_i с y_i. Т.е. $Z_i(n_i \text{ x } k)$.

За измерените данни от популация с многомерно нормално разпределение строим следния модел:

Ниво 1. За всеки индивид i:

$y_i = X_i\alpha + Z_ib_i + \varepsilon_i$, където

$\varepsilon_i -$ са независими вектори, $\varepsilon_i \in N(0, R_i)$,

R_i е ковариационната матрица, $R_i(n_i \text{ x } n_i)$,

n_i е броят на измерванията във времето за i - тия пациент.

Ковариационната матрица R_i на грешките ε_i зависи от индивида i чрез броят на направените измервания n_i на индивида i.

На ниво 1 векторите α и b_i са известни, фиксирани.

Ниво 2. Нека b_i е случаен вектор и $b_i \in N(0, D)$, D е ковариационна матрица, $D(k \text{ x } k)$, k е броя на ефектите , които се свързват с индивида, т.е. b_i са случайните ефекти. Нека

- b_i са независими по между си (независими един от друг),
- b_i не зависят от ε_i,

α са фиксираните ефекти.

Тогава векторите y_i са независими и еднакво нормално разпределени случайни вектори със средно $X_i\alpha$ и ковариационна матрица $R_i + Z_i D Z_i^T$.

Изследваме промените във времето на една характеристика на изследваното явление. Моделът, който строим, има следните характеристики:

- Вероятностните разпределения на повторените във времето измервания са от един и същи вид за всеки индивид, но параметрите на това вероятностно разпределение зависят от индивида.

Разпределението на тези параметри на модела, които се наричат случайни ефекти, е втория етап на модела (второто ниво).

Пример за такъв модел: Можем да предположим, че връзката между две променливи е линейна за всеки индивид, но регресионните параметри на линейния модел да са различни за всеки индивид.

Ако предположим, че обикновеният линеен регресионен модел може да се приложи за всеки индивид, и че регресионните параметри имат двумерно нормално разпределение, то маргиналното разпределение на серията измервания е многомерно нормално с определена структура на ковариационната матрица.

Тези двуетапни модели имат следните характерни особености:

- Данните могат да не са балансирани, т.е. експериментът може да не е балансиран (наблюденията върху индивидите могат да са различен брой).

- Моделират се различията между индивидите и различията в характеристиките на индивида във времето.
- В тези модели предполагаме, че ковариационната матрица има специална структура.
- Такива модели се наричат двустъпкови модели със случайни ефекти.
- Целта е да се изследват промените, настъпили във времето.

В двустъпковите модели със случайни ефекти имаме следните особености:

- Откликът е вектор от повторените измервания на индивидите.
- Вероятностното разпределение на отклика е многомерно нормално разпределение.
- Някой параметри на модела се променят за индивидите, т.е. зависят от индивидите.
- Вероятностните разпределения на параметрите, който зависят от индивидите, се определят на ниво 2.

Модели от този вид се наричат общи линейни модели (General Linear Models) и са въведени от Laird и Ware в (Laird, Ware, 1982).

4.1 Използване на бейсовата статистика в денталната медицина. Сравняване на два метода на лечение с бейсов подход

Гингивалната рецесия е невъзпалително състояние, при което се получава отдръпване на венеца и оголване на зъбния корен. Най- често засяга зъбите във фронталната област (резци и кучешки). Може да обхване един или няколко зъба.

Целта на настоящата задача е да се построи модел на сравнителна клинична оценка на резултатите от хирургично лечение на гингивални рецесии с помощта на мембрана от богат на тромбоцити фибрин (виж [1]). За тази цел се измерва възстановяването на гингивалните рецесии с клиничния параметър дължина на гингивалната рецесия (GRD - Gingival Recession Depth). Стойноста на дължината на гингивалната рецесия (GRD) е равна на разликата между стойностите на височината на рецесията преди операцията и тази на шестия месец след операцията. Всички измервания се извършват с електронен шублер с приспособени за целта измерващи браншове. Стойностите се измерват в милиметри. Измереният параметър се отчита предоперативно и на шестия месец след операцията.

Всеки пациент е лекуван с две различни хирургични техники в лявата и дясна половини на съответната челюст, като метода и страната са избирани произволно. Гингивалните рецесии на пациентите от едната страна бяха лекувани хирургично с метода на КПЛ + съединително тъканна присатка /СТП/(CAF+connective tissue graft /CTG/ приеман от повечето автори за "Златен стандарт" [2]. Тази група оперирани пациенти нарекохме условно контролна група (Група 1). Рецесиите от противоположната страна на същия пациент бяха лекувани с КПЛ + (PRFm), а метода нарекохме условно "Нов метод" и групата - тестова (Група 2).

Сравняват се резултатите от двата метода: Златния стандарт (Група 1) и Новия метод (Група 2).

В процеса на проучването са идентифицирани 58 сдвоени данни. Сред данните в двете групи се наблюдават рязко отклоняващи се наблюдения. Затова данните се описват с помощта на разпределения с тежки опашки, вместо с нормалното разпределение.

Таблица 1.

Данни за дължината на гингивалната рецесия от лечението по двата метода

	Дължина на гингивалната рецесия									
Златен стандарт	2.8	3.4	3.1	3.5	3.5	4.2	3.9	3.3	3.0	3.0
	3.7	3.3	4.4	3.1	3.6	3.2	3.2	3.4	3.5	2.4
	2.4	1.8	3.6	3.4	2.7	3.0	3.2	3.2	3.4	3.6
	3.8	3.3	3.8	4.4	3.4	3.4	3.2	3.3	3.9	2.5
	4.0	3.8	2.4	2.2	4.4	3.3	3.3	3.1	3.2	3.0
	3.0	2.4	2.5	2.2	2.6	2.8	2.7	3.0		
Нов метод	2.8	2.0	3.8	3.4	4.1	2.7	3.8	3.4	2.7	2.7
	3.0	3.4	3.3	2.2	2.4	2.5	2.7	3.2	2.7	2.0
	1.2	1.8	3.2	3.8	3.3	2.8	2.6	2.6	3.0	3.2
	3.5	3.7	2.9	3.2	3.0	3.7	3.4	3.2	3.5	3.8
	2.5	5.0	2.2	2.0	4.3	3.3	3.2	2.6	2.4	2.5
	2.3	2.2	1.6	2.0	1.9	2.8	2.2	2.5		

Моделът на дължината на гингивалната рецесия е вероятностно разпределение, което описва данните с пет параметъра: средна стойност (μ) и стандартно отклонение (σ) за всяка от двете групи и параметър за нормалност (ν), еднакъв за двете групи. Вероятностното разпределение на дължината на гингивалната рецесия може да се разглежда и като средство за измерване на степента на сигурност (или несигурност) на стойностите на клиничния параметър. Полученото разпределение е съвместно вероятностно разпределение на петте параметъра, като по този начин се разкриват комбинации от петте стойности на параметрите, които са достоверни, като вземем в предвид данните.

Разликата (μ_1 - μ_2) описва степента на различие на централните тенденции на групите, а разликата от стандартните отклонения (σ_1- σ_2) описва степента на разликата в разпръснатостта на данните в двете групи. Нашите основни цели са да се оценят тези величини и да се оцени нашата сигурност, убеденост в тези оценки.

Бейсовият метод дава отговори и на двете цели едновременно. Данните са измерени в метрична скала за двете групи, съдържат екстремни стойности (т.е. рязко отклоняващи се наблюдения) и са описани с помощта на разпределение, което има тежки опашки.

Относителното тегло на опашките на t разпределението се управлява от параметър обозначен с ν, който може да приема стойност от 1 до безкрайност. Чрез t разпределението могат да се опишат данни с големи различия в стойностите като се определи малка стойност на параметъра ν, но може и да се опишат данни, които са близки до нормално разпределени и без екстремни стойности с големи стойности на параметъра ν.

С представения модел на данните, описваме всяка група от данни с помощта на t разпределение, като всяка група има собствен параметър за средното и собствено стандартно отклонение. Поради факта, че екстремните стойности в двете групи са сравнително малко на брой ще се използва един и същ параметър ν и за двете групи. По този начин данните се описват с пет параметъра: средните стойности на двете групи (μ_1 и μ_2), стандартните отклонения на двете групи (σ_1 и σ_2), и нормалността на данните вътре в групите (ν). Използва се бейсовия подход за да се изчислят петте параметъра.

4.1.1 Генериране на извадки

С навлизането на достатъчно мощна изчислителна техника през последните 20 години бейсовият анализ започва да използва клас от методи за числено симулиране, известни като Монте Карло симулации с марковски вериги (MCMC – Markov Chain Monte Carlo). Те са силно интензивни от изчислителна гледна точка, но са особено ефективни за целите на бейсовия анализ. Основната идея на този метод е да се генерира достатъчно голяма извадка от апостеорната плътност като реализация от подходящо построена арковска верига. След това анализът на свойствата на апостеорното разпределение се прави на основата на така генерираната извадка.

Колкото по-голяма е генерираната извадка, толкова по-добре тя представя основното апостеорно разпределение. В бейсовия анализ генерираната по метода MCMC извадка се нарича **MCMC верига**. Размерът на генерираната извадка се нарича **дължина на веригата**. Важно е да не се бъркат двете понятия - генерираните MCMC извадки с извадките от наблюденията (емпиричните данни).

Целта на процеса MCMC е да генерира случайна извадка, която точно и надеждно представя апостеорното разпределение. Но за съжаление, MCMC алгоритмите могат да страдат от зависимост на данните (clumpiness, технически наречено, автокорелация) във веригите, които те генерират. Един от начините за разреждане на автокорелацията е чрез изтъняване на веригата, което означава използване само на данните от k-та стъпка във веригата, където k е подходящо избрано от потребителя число.

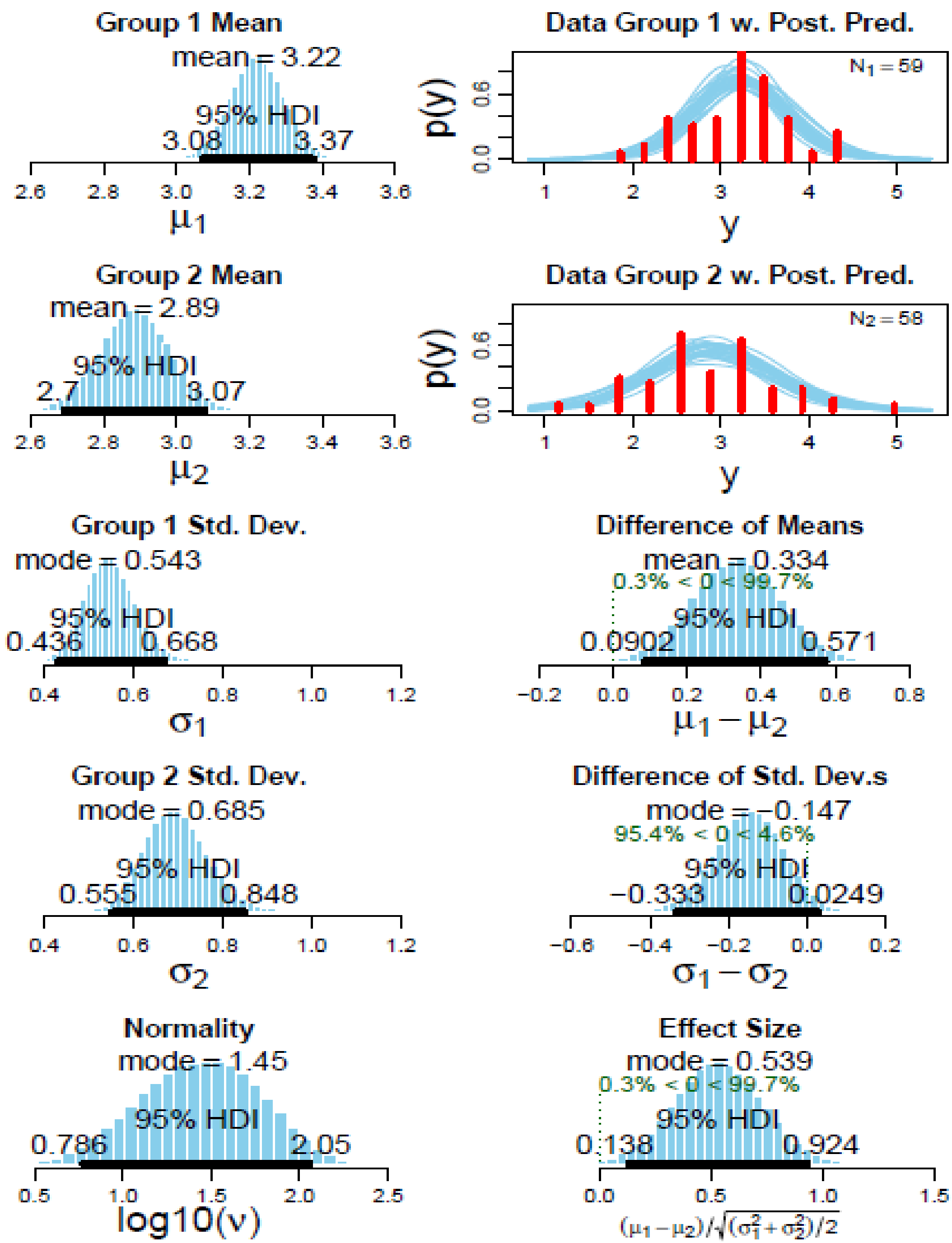
Средната стойност на генерираната извадка от Група 1 е 3.22 и средната стойност на генерираната извадка от Група 2 е 2.89, но има и много променливост в рамките на групите. Модата на стандартното отклонение на Група 1 е 0.543, а модата на стандартното отклонение на Група 2 е 0.685. Също така изглежда има няколко екстремни стойности. Можем да си зададем въпроса: *Убедителни ли са разликите в тези две групи*?

Методът наречен "Бейсова оценка", описан в [5,6], дава богата информация относно разликите между отделните групи. Методът MCMC генерира голям брой комбинации от параметри, които са надеждни. Тези комбинации от стойности на параметрите са представителни за апостеорното разпределение. На Фигура 1 са представени хистограми на 100 000 достоверни комбинации от стойностите на параметрите. Само хистограмите (горе в дясно) на Фигура 1 са хистограми на реалните данни (те не са хистограми на симулираните данни). Т.е. единствените хистограми на данните се визуализират в горния десен панел на фигурата, като техните абсцисни оси са етикетирани с y, и тези данни са фиксирани с техните емпирично наблюдавани стойности.

Всички други хистограми показват 100 000 стойности на параметрите на апостеорните разпределения. Петте хистограми в лявата колона на фигурата показват апостеорните стойности, съответстващи на следните пет априорни разпределения: Априорните разпределения на средните стойности μ_1 и μ_2, са много широки нормални разпределения, априорните разпределения на стандартните отклонения σ_1 и σ_2, са равномерни разпределения, а параметърът ν има априорно експоненциално разпределение.

Фигура 1.

Хистограми и 95% HDI интервали, получени от веригите, и хистограми на данните



В контекста на класическата статистика доверителният интервал е случаен, понеже зависи от реализациите на данните и съответно вероятностните твърдения се отнасят до вероятността, с която този интервал ще покрие неизвестната стойност на параметъра. В бейсовият анализ работим с интервали с най-висока плътност, наречени HDI интервали (highest density intervals). В HDI интервала попадат по-голямата част от най-надеждните стойности. По дефиниция, всяка стойност вътре в този интервал има по-висока плътност на вероятност, отколкото която и да е стойност извън интервала, и общата маса на точките вътре в 95% HDI интервал е 95% от разпределението.

Горният ляв панел на Фигура 1 показва, че средната стойност за μ_1 е 3.22, с 95% интервал с най-висока плътност от 3.08 до 3.37. Средната стойност на веригата MCMC за μ_2 е 2.89, с 95% HDI интервал от 2.7 до 3.07. Следователно, средното на разликата (μ_1 - μ_2) е 0,334, както е показано в средния участък от дясната колона на фигурата. Виждаме, че 95% HDI от разликата на средните е доста над нулата, и 99,7% от достоверните стойности са по-големи от нула. **Следователно можем да заключим, че средните на групите действително са различни.**

Бейсовият анализ показва едновременно надеждните стойности на стандартните отклонения за двете групи чрез хистограмите в лявата колона на Фигура 1. Разликата от стандартните отклонения е показана в дясната колона, където може да се види, че разлика от нула е сред 95-те процента най-достоверни разлики, и 95.4% от достоверните разлики са по-малки от нула.

Следователно средната стойност на първата група е убедително по-голяма от средната стойност на втората група, и стандартното отклонение на първата група е убедително по-малка от стандартното отклонение на втората група. В контекста на задачата, този резултат означава, че средната дължина на гингивалната рецесия на лекуваните по метода Златен стандарт (Група 1) пациенти на шестия месец след операцията е убедително по-голяма от средната дължина на гингивалната рецесия на лекуваните по Новия метод (Група 2). Стандартното отклонение на дължината на гингивалната рецесия на лекуваните по метода Златен стандарт пациенти на шестия месец след операцията в 95.4% от случаите е по–малко от стандартното отклонение на дължината на гингивалната рецесия на лекуваните по Новия метод, т.е. **методът известен като Златен стандарт е по-предвидим от Новия метод**.

Долният десен панел на Фигура 1 показва разпределението на размера на ефекта, като се вземат в предвид данните. Стойността на ефекта е безразмерна количествена мярка, която се изчислява като нормирана разлика на средните. По-голяма абсолютна стойност винаги показва по-силен ефект. Например, прието е стойност на ефекта 0,1 да се счита за малка стойност, практически нулевият ефект (ROPE- region of practical equivalence) може да бъде от -0,1 до 0,1.

Стойността на ефекта се изчислява по формулата:

Ефект = $(\mu_1 - \mu_2)/\sqrt{(\sigma_1^2 + \sigma_2^2)/2}$.

Хистограмата от 100 000 размери на ефекта има мода равна на 0.539, с форма, показана на Фигура 1, и 95% HDI интервал, който изключва нулата.

В долния ляв панел на Фигура 1 са показани достоверните стойности на параметъра на нормалността за t разпределението. Стойностите са зададени в логаритмичната скала, тъй като формата на t разпределението се променя забележимо за стойности на ν близо до 1, но промените са сравнително малки за ν = 30 или повече.

Горните два десни панела на Фигура 1 визуализират t разпределенията, наложени върху хистограмите на данните на двете групи. Кривите са начертани, като са избрани няколко случайни стъпки във веригите MCMC и за всяка стъпка е показано t разпределението с параметри (μ_1, σ_1, ν) на данните от Група 1, и t разпределението с параметри (μ_2, σ_2, ν) на данните от Група 2.

Чрез визуално сравняване на хистограмата на данните и типичните достоверни t разпределения, заключаваме, че представеният модел сравнително добре описва данните. Използваната идея е следната: Ако стойностите на апостериорните параметри наистина добре описват данните, тогава прогнозираните данни от модела трябва действително да "изглеждат като" реалните данни.

4.2 Общ линеен модел анализиращ ефекта от лечението

Ще построим модел изследващ (анализиращ) ефекта от лечението. Броят на измерванията във времето за всеки пациент е 4.

Нека y_i е 4x1 вектор с измерванията за i –тия пациент, $y_i = \boldsymbol{I}\alpha + \mathbf{1}b_i + \varepsilon_i$, $i = 1, 2, \dots, m$, където:

$\boldsymbol{I}$- е 4x4 единична матрица,

1 – е 4x1 вектор от единици (единичен вектор),

векторът α съдържа средните на популацията за 4 времеви точки,

b_i- са случайните отклонения от средните за i – тия пациент.

Предполагаме, че $\varepsilon_i \in N(\mathbf{0}, \sigma^2\boldsymbol{I})$ и $b_i \in N(0, \tau^2)$.

В експеримента няма липсващи данни.

Различните пациенти могат да се повлияят по различен начин от лечението. Ние можем да въведем втори случаен ефект, който да моделира/да представи (средния спад в клиничния параметър) средната реакция за всеки пациент.

$$y_i = I\alpha + Z_i b_i + \varepsilon_i,$$

където: $b_i' = (b_{1i}, b_{2i})$ и $Z_i = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$, Z_i (4x2).

Тук b_{1i} е индивидуалния ефект за i – тия пациент преди лечението, а b_{2i} измерва средното отклонение на стойностите след лечението от стойностите преди лечението. Оценките на b_{2i} могат да се използват за да се идентифицират пациентите, който не показват достатъчно подобрение и тази група пациенти подлежи на по-нататъчни изследвания. Много от тези пациенти са развили временни усложнения след лечението.

За да изучим влиянието на пола, пушенето, и други индивидуални особености на пациента върху отклика, ние строим модел, който включва тези фактори. Ако V_i е вектор от индивидуалните характеристики за i– тия пациент и b_{1i} и b_{2i} са два линейни регресионни модела, т.е:

$$b_{1i} = V_i^{\mathrm{T}}\gamma_1 + b_{1i}' \text{ и } b_{2i} = V_i^{\mathrm{T}}\gamma_2 + b_{2i}', \text{ където}$$

b_{1i}' и b_{2i}' са индивидуалните отклонения от регресионния модел за всеки индивид.

Тогава $b_i = \begin{bmatrix} V_i^{\mathrm{T}} & 0 \\ 0 & V_i^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} + b_i' = V_i\gamma + b_i'.$

Следователно моделът за y_i е:

$$y_i = I\alpha + Z_i b_i + \varepsilon_i = I\alpha + Z_i(V_i\gamma + b_i') + \varepsilon_i =$$

$$=[I,\ Z_iV_i]\begin{bmatrix} \alpha \\ \gamma \end{bmatrix} + Z_i b_i' + \varepsilon_i .$$

Следователно моделът на описаната задача има вида на общ линеен модел (ниво 1):

$$y_i = [I,\ Z_iV_i]\begin{bmatrix} \alpha \\ \gamma \end{bmatrix} + Z_i b_i' + \varepsilon_i \text{ , където } \varepsilon_i \in N(0,\ \sigma^2 I),\ b_i' = (b'_{1i}, b'_{2i})^{\mathrm{T}}.$$

4.3 Заключение

От гледна точка на бейсовата статистика няма разлика между наблюдения и параметри на статистически модел – всички те са случайни величини.

Нека с D да означим наблюденията (данните), а с θ параметрите на модела и липсващите данни. Формалните изводи изискват да се намери съвместно вероятностно разпределение $P(D, \theta)$. Това съвместно вероятностно разпределение се състои от две части: априорно разпределение $P(\theta)$ и функция на правдоподобие $P(D|\theta)$. Определянето на $P(\theta)$ и $P(D|\theta)$, ни дава пълния вероятностен модел:

$$P(D, \theta) = P(\theta).P(D, \theta).$$

Когато имаме данните теоремата на Бейс, ни дава условното разпределение на θ при условие D.

$$P(\theta|D) = \frac{P(\theta)\,.\,P(D|\theta)}{\int P(\theta)\,.\,P(D|\theta)d\theta},$$

$\mathrm{P}(\theta|D)$ се нарича апостеорно разпределение на θ.

Целта на всички бейсови изводи е намирането на апостеорното вероятностно разпределение $\mathrm{P}(\theta|D)$.

Всички квантили на апостеорното разпределение могат да се изразят в термините на условните очаквания: $\mathrm{E}\,[f(\theta)|D]$.

Много често намирането на аналитичния вид на $\mathrm{E}\,[f(\theta)|D]$ е невъзможно. Алтернативните подходи в този случай са няколко. Най- често използваният е MCMC метод.

4.4 MCMC метод

4.4.1 Monte Carlo

- Генерира извадки $\{X_t,\ t = 1, 2, \dots, n\}$ от желаното разпределение $\pi(\cdot)$.
- Апроксимира очакване $\mathrm{E}\,f(X)$ с извадковото средно, т.е. популационното средно е оценено с извадковото средно.

Понякога генерирането на независими извадки от вероятностно разпределение, което не е стандартно, не е възможно. Тогава генерираме извадки чрез марковска верига, чието стационарно разпределение е $\pi(\cdot)$. Това е основната идея на MCMC.

4.4.2 Марковски вериги

Генерираме редица от случайни величини $\{X_0, X_1, X_2, \dots\}$, така, че за всяко $t \geq 0$ следващото състояние X_{t+1} е генерирано от вероятностно разпределение $P(X_{t+1}|X_t)$, което зависи само от текущото състояние X_t.

Такава редица от случайни величини се нарича марковска верига, а $P(\cdot\,|\,\cdot)$ се наричат вероятности на преходите (transition kernel) и не зависят от t.

Как началното състояние X_0 влияе на текущото състояние X_t? Този въпрос се отнася до разпределението на X_t при дадено X_0 , т.е. $P^{(t)}(X_t|X_0)$.

Искаме $P^{(t)}(\cdot\,|X_0)$ да се доближава до единствено стационарно разпределение $\phi(\cdot)$, което не зависи от t и от X_0. Така, когато t расте точките $\{X_t\}$ ще изглеждат като зависими извадки от $\phi(\cdot)$.

Така след достатъчно много итерации точките $\{X_t, t = m+1, \dots, n\}$ ще бъдат зависими извадки от $\phi(\cdot$

Ще използваме марковската верига за да оценим очакването E $f(X)$, където X има разпределение $\phi(\cdot)$.

$$\bar{f} = \frac{1}{n-m}\sum_{i=m+1}^{n} f(X_i).$$

Оценката $\bar{f}$ се нарича ергодично средно.

Сходимостта към $E_\phi f(X)$ се осигурява от ергодичната теорема.

Как ще конструираме подходяща марковска верига? Т.е. как да конструираме марковска верига, така, че нейното стационарно разпределение $\phi(\cdot)$ да е точно разпределението $\pi(\cdot)$ от което ние се интересуваме?

За всяко t, следващото състояние X_{t+1} се избира като генерираме кандидат точка Y от предложеното разпределение $q(\cdot\,|\,X_t)$. Предложеното вероятностно разпределение може да зависи от текущото състояние X_t. Приемаме Y с вероятност $\alpha(X_t, Y)$, където

$$\alpha(X, Y) = \min\left(1,\ \frac{\pi(Y)\, q(X|Y)}{\pi(X)\, q(Y|X)}\right).$$

Ако Y е приета, то $X_{t+1} = Y$.

Ако Y е отхвърлена $X_{t+1} = X_t$, т.е. не мърдаме.

Въпрос: *Как ще изберем разпределението* $q(\cdot\,|\,\cdot)$*?*

Извод: МСМС е просто една идея с огромен потенциал. МСМС за бейсови оценки. МСМС за проверка за адекватност на бейсовия модел.

4.5 Речник на основните понятия

4.5.1 Марковски вериги

Марковската верига е дискретен вероятностен процес $\{X_0, X_1, \dots\}$ със свойството, че разпределението на X_t при дадени всички предишни реализации $X_0, X_1, \dots, X_{t-1}$ зависи само от предходното състояние X_{t-1}.

Т.е. $P(X_t \in A | X_0, X_1, \dots, X_{t-1}) = P(X_t \in A | X_{t-1})$, за $\forall A$.

Обикновено веригите на Марков приемат стойности в d −мерното Евклидово пространство $\mathbb{R}^d$. Но ние ще се ограничим до дискретно пространство на състоянията. Т.е. ще разглеждаме $P_{ij}(t) = P(X_t = j | X_0 = i)$.

Нека $f(\cdot)$ е реална функция. Ще дефинираме ергодичното средно по следния начин:

$$\overline{f_N} = \frac{\sum_{t=1}^{N} f(X_t)}{N}.$$

Асимптотичните свойства на $\overline{f_N}$ са много важни.

Нека $\pi(\cdot)$ е стационарно разпределение, т.е. ако X_0 има вероятностно разпределение $\pi(\cdot)$, то всички X_t, ще имат вероятностно разпределение $\pi(\cdot)$.

Нека τ_{ii} е времето на първото завръщане в началното състояние, т.е.

$$\tau_{ii} = \min(t > 0 :\ X_t = i | i).$$

Теорема: Нека X е повтаряща се (recurrent) и нециклична марковска верига, за която:

$E\tau_{ii} < \infty$, , за $\forall i$ (т.е. положително повтаряща).

Следователно нейното стационарно вероятностно разпределение $\pi(\cdot)$ е единственото вероятностно разпределение, за което:

$$\sum_i \pi(i) P_{ij}(t) = \pi(j), \forall j, \forall t \geq 0.$$

Тогава казваме, че марковската верига е ергодична и е в сила:

$P_{ij}(t) \to \pi(j)$, когато $t \to \infty, \ \ \forall\, i, j$.

4.5.2 Ергодична теорема

Нека $f(\cdot)$ е реална функция. Ако $E_\pi |f(X)| < \infty$, то

$P\left(\overline{f_N} \to E_\pi f(X)\right) = 1$, където

$E_\pi\, f(X) = \sum_i f(i)\, \pi(i)$ е математическото очакване на $f(X)$.

Ергодичната теорема е много важна за MCMC.

Повечето Марковски вериги генерирани от MCMC са обратими (reversible).

Казваме, че една марковска верига е обратима , ако е положително повтаряща се със стационарно разпределение $\pi(\cdot)$ и ако:

$\pi(i)\, P_{ij} = \pi(j)\, P_{ji}$.

4.6 Бейсови обобщени линейни модели със смесени ефекти (Bayesian GLMMs)

- По сложна структура на грешката

От гледна точка на бейсовата статистика не е необходимо да разбиваме вектора от предикторите (x, z) и съответно вектора на параметрите на модела (β, b) на фиксирани и случайни ефекти. На този етап всички параметри са случайни променливи, които имат съвместно нормално разпределение със средно 0.

Моделът може да се представи като прост общ линеен модел:

$\eta = X\, \beta$, където

$\beta \in N(0, \Lambda)$.

Различието в ефектите се отразява в матрицата Λ.

Разбиването на модела на фиксирани и случайни ефекти съответства на разбиването на матрицата Λ като:

$\Lambda = \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_1(\theta) \end{bmatrix}$, където θ са неизвестни параметри оценени от данните.

Взаимодействия (Смесени влияния)

Основни правила при определяне на Λ, когато имаме взаимодействащи си случайни ефекти:

- Взаимодействия между случайни ефекти и една променлива, която влияе и на зависимата и на независимата променлива (covariate).

Регресионните прави в групите имат различни свободни членове и различни коефициенти пред предиктора. Моделът е:

1 + Covariate + Group + Group*Covariate

- Взаимодействия между случайни и фиксирани ефекти.

Моделът е:

1 + Fixed + Random + Fixed*Random

- Взаимодействия между два случайни ефекта

Модели със случайни ефекти за повторени във времето измервания.

Модел 1:

1 + Covariate + Treatment + Period + Subject , където

Treatment и Period са фиксирани ефекти, а Subject е случаен ефект.

Модел 2 : Линеен ефект на времето от започване на лечението T.

1 + Covariate + Treatment + Period + Subject + T + Subject*T

Литература:

1. Dobreva M., Noncheva V., Chenchev Iv. Bayesian approach for the comparison of two methods of treatment. Proceedings of International Scientific Jubilee Conference 25 years Faculty of Mathematics and Informatics, University of Veliko Turnovo “St. Cyril and Methodius”. Bulgaria, November, 2015; pp 107-112.
2. Chambrone L, Chambrone D, Pustiglioni FE, Chambrone LA, Lima LA. Can subepithelial connective tissue grafts be considered the gold standard procedure in the treatment of Miller Class I and II recession-type defects? Journal of Dentistry, 2008; 36: 659–671.
3. Cohen Jacob, Statistical Power Analysis for Behavioral Sciences. Second Edition. Lawrence Erlbaum Associates, Publishers Hillsdale, New Jersey, 1988.

4. Gelman, A., Carlin, J.B., Stern H.S., & Rubin D.B. Bayesian data analysis. Third Edition. Boca Raton, Chapman and Hall–CRC, 2013.
5. Kruschke, J. K. Doing Bayesian data analysis: A tutorial with R and BUGS. Second Edition. Burlington, MA: Academic Press/Elsevier, 2011.
6. Kruschke, J. K. Bayesian estimation supersedes the t test. Journal of Experimental Psychology: General, 2013; 142(2), 573–603.