

Analýza data pomocí Bayesiánské statistiky
třetí vydání

Andrew Gelman, John B. Carlin, Hal S. Stern,
David B. Dunson, Aki Vehtari, and Donald B. Rubin

20. srpna 2019

Obsah

1	Pravděpodobnost a inference	7
1.1	Notace	7
1.2	Bayesiánská inference	7
1.2.1	Bayesova věta	8
1.2.2	Predikce	8
1.2.3	Věrohodnostní funkce	8
1.3	Ilustrativní příklady	9
1.3.1	Genetika	9
1.3.2	Překlepy	10
1.3.3	Americký fotbal	11
1.3.4	Párování záznamů	12
1.4	Několik užitečných poznatků z teorie pravděpodobnosti	14
1.4.1	Podmíněné pravděpodobnosti	14
1.4.2	Hypotézy	14
1.4.3	Střední hodnota a rozptyl	14
1.4.4	Modelování a podmíněná pravděpodobnost	15
1.4.5	Střední hodnota a rozptyl podmíněných distribucí	15
1.4.6	Transformace proměnných	16
2	Jednparametrové modely	17
2.1	Odhad pravděpodobnosti z binomických dat	17
2.1.1	Ilustrační příklad - odhad pravděpodobnosti odhadu narození ženy	17
2.2	Aposteriorní pravděpodobnost jako kompromis mezi daty a apriorní pravděpodobností	19
2.3	Charakteristika posteriorního rozdělení	20
2.4	Informativní apriorní distribuce	20
2.4.1	Binomický model s alternativním apriorním rozdělením	20
2.4.2	Konjugátní apriorní rozdělení	21
2.4.3	Nekonjugátní apriorní rozdělení	21
2.4.4	Konjugátní apriorní rozdělení, rodina exponenciálních pravděpodobnostních rozdělení a dostatečná statistika	22
2.4.5	Ilustrativní příklad	22
2.5	Odhad normální střední hodnoty při známém rozptylu	24
2.5.1	Věrohodnostní funkce jednoho pozorování	25
2.5.2	Normální rozdělení a vícero pozorování	26
2.6	Ostatní jednparametrové modely	27

2.6.1	Normální rozdělení se známou střední hodnotou a neznámým rozptylem	28
2.6.2	Poissonův model	28
2.6.3	Negativní binomický model	29
2.6.4	Poissonův model parametrizovaný pomocí míry a expozice	30
2.6.5	Exponenciální model	31
2.7	Ilustrativní příklad - rakovina ledvin v USA	31
2.7.1	Rozložení případů napříč USA	31
2.7.2	Bayesiánská inference	31
2.7.3	Důsledky velikosti okresu	32
2.8	Neinformativní apriorní rozdělení	33
2.8.1	Vlastní na nevlastní apriorní rozdělení	33
2.8.2	Nevlastní apriorní vs. vlastní posteriorní rozdělení	34
2.8.3	Jeffreyovo pravidlo	34
2.8.4	Neinformativní apriorní rozdělení pro binomický parametr	34
2.8.5	Klíčová hodnota, lokační a škálovací parametr	35
2.8.6	Komplikace spojené s neinformativním apriorním rozdělením	35
2.9	Slabě informativní apriorní rozdělení	36
3	Víceparametrové modely	37
3.1	Průměrování přes nadbytečné parametry	37
3.2	Normální data s neinformativním apriorním rozdělením	37
3.2.1	Neinformativní apriorní rozdělení	38
3.2.2	Sdružené posteriorní rozdělení $p(\mu, \sigma^2 y)$	38
3.2.3	Podmíněné posteriorní rozdělení $p(\mu \sigma^2, y)$	38
3.2.4	Marginální posteriorní rozdělení $p(\sigma^2 y)$	38
3.2.5	Výběr ze sdruženého posteriorního rozdělení	39
3.2.6	Analytická forma marginálního posteriorního rozdělení μ	39
3.2.7	Aposteriorní prediktivní rozdělení budoucích pozorování	40
3.3	Normální data s apriorním konjugátním rozdělením	40
3.3.1	Rodina apriorních konjugátních rozdělení	40
3.3.2	Sdružené posteriorní rozdělení $p(\mu, \sigma^2 y)$	41
3.3.3	Podmíněné posteriorní rozdělení $p(\mu \sigma^2, y)$	41
3.3.4	Marginální posteriorní rozdělení $p(\sigma^2 y)$	41
3.3.5	Výběr ze sdruženého posteriorního rozdělení	41
3.3.6	Analytická forma marginálního posteriorního rozdělení parametru μ	41
3.4	Multinomický model pro kategorická data	42
3.4.1	Ilustrativní příklad - předvolební průzkum	42
3.5	Vícerozměrný normální model se známým rozptylem	43
3.5.1	Věrohodnostní funkce	43
3.5.2	Konjugátní analýza	43
3.6	Vícerozměrný normální model s neznámou střední hodnotou a rozptylem	45
3.6.1	Konjugátní inverzní Wishartova rodina apriorních distribucí	45
3.6.2	Další neinformativní apriorní distribuce	46
3.7	Ilustrativní příklad - biotest	47
3.7.1	Popis problematiky a data	47
3.7.2	Modelování vztahu mezi dávkováním a výsledky testů	47

3.7.3	Věrohodnostní funkce	48
3.7.4	Apriorní distribuce	48
3.7.5	Hrubý odhad parametrů	48
3.7.6	Graf sdružené aposteriorní hustoty pravděpodobnosti . .	48
3.7.7	Náhodný výběr ze sdružené aposteriorní distribuce	48
3.7.8	Aposteriorní distribuce LD50	49

4	Slovník	51
---	---------	----

Kapitola 1

Pravděpodobnost a inference

1.1 Notace

Statistická inference se zabývá formulováním závěrů o nepozorovaných veličinách na základě numerických dat. Příkladem takového nepozorované veličiny může být pravděpodobnost přežití pacienta po aplikaci nového léku. Ačkoliv tato veličina popisuje pravděpodobnost přežití v rámci celé populace pacientů, je zpravidla odhadnuta na základě náhodného výběru.

V následujícím textu budeme θ používat pro označení populačních parametrů (např. výše uvedená pravděpodobnost přežití), y bude označovat pozorovaná data (např. počet přeživších v náhodném výběru dané velikosti), \tilde{y} bude označovat neznámé avšak potenciálně pozorovatelné veličiny (např. počet pacientů, kteří přežijí po té, co se podrobili léčbě).

θ může mít charakter skaláru popř. vektoru a to podle počtu parametrů, které charakterizují danou populaci. V případě vektoru tak platí $\theta = (\theta_1, \theta_2, \dots, \theta_k)$, kde k je počet parametrů.

V závislosti na počtu pozorování může mít y podobu jak skaláru tak vektoru. V případě vektoru platí $y = (y_1, y_2, \dots, y_n)$, kde n představuje počet pozorování. Pokud pro každé pozorování zaznamenáváme více než jednu charakteristiku, je y_i vektorem a y maticí. Z logiky věci vyplývá, že výše uvedené značení je možné analogicky aplikovat také na \tilde{y} .

1.2 Bayesiánská inference

V případě Bayesiánské statistiky mají závěry ohledně parametru θ popř. nepozorovaných dat \tilde{y} charakter pravděpodobnostních výroků, které jsou podmíněné pozorovanými hodnotami y , a v následujícím textu je budeme označovat jako $p(\theta|y)$ popř. jako $p(\tilde{y}|y)$. Připomeňme, že $p(\cdot|y)$ značí podmíněnou pravděpodobnost, zatímco $p(\cdot)$ značí marginální pravděpodobnost.

Abychom se vyhnuli případným nejasnostem, budeme někdy pro vyjádření pravděpodobnosti určité události používat značení $Pr(\cdot)$ ve smyslu $Pr(\theta > 2) = \int_{\theta > 2} p(\theta) d\theta$.

1.2.1 Bayesova věta

Sdružená pravděpodobnost pro θ a y je definována jako

$$p(\theta, y) = p(\theta)p(y|\theta), \quad (1.1)$$

kde $p(\theta)$ představuje tzv. apriorní rozdělení parametru θ a $p(y|\theta)$ představuje tzv. výběrové rozdělení podmíněné hodnotou parametru θ .

S využitím tzv. Bayesovy věty lze odvodit tzv. aposteriorní rozdělení jako

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)}, \quad (1.2)$$

kde $p(y) = \sum_{\theta} p(\theta)p(y|\theta)$ popř. $p(y) = \int_{\theta} p(\theta)p(y|\theta)d\theta$ v závislosti na tom, zda-li je θ diskrétní či spojitá veličina. V některých případech se aposteriorní pravděpodobnost uvádí v tzv. nenormalizované formě

$$p(\theta|y) \propto p(\theta)p(y|\theta), \quad (1.3)$$

kde jsme vynechali $p(y)$, které lze v kontextu (1.2) chápat jako normalizační konstantu.

1.2.2 Predikce

Před tím, než vezmeme v potaz pozorovaná y , je rozdělení této veličiny definováno jako

$$p(y) = \int p(y, \theta) d\theta = \int p(\theta)p(y|\theta) d\theta. \quad (1.4)$$

Toto pravděpodobnostní rozdělení často nazýváme marginálním rozdělením.

Po té, co jsme pozorovali y , je možné analogickým způsobem zkonstruovat aposteriorní prediktivní rozdělení pro \tilde{y} .

$$p(\tilde{y}|y) = \int p(\tilde{y}, \theta|y) d\theta = \int p(\tilde{y}|\theta, y)p(\theta|y) d\theta = \int p(\tilde{y}|\theta)p(\theta|y) d\theta \quad (1.5)$$

1.2.3 Věrohodnostní funkce

Aplikace Bayesovy věty spolu se zvoleným pravděpodobnostním modelem znamená, že y ovlivňuje aposteriorní inferenci (1.3) pouze skrze $p(y|\theta)$. (1.3) je věrohodnostní funkcí a lze na ni nahlížet jako na funkci θ pro fixní y .

Poměr $p(\theta|y)$ v bodech θ_1 a θ_2 pro daný model nazýváme aposteriorním pravděpodobnostním poměrem θ_1 ku θ_2

$$\frac{p(\theta_1|y)}{p(\theta_2|y)} = \frac{p(\theta_1)p(y|\theta_1)/p(y)}{p(\theta_2)p(y|\theta_2)/p(y)} = \frac{p(\theta_1)p(y|\theta_1)}{p(\theta_2)p(y|\theta_2)} \quad (1.6)$$

Výše uvedená rovnice říká, že tento aposteriorní pravděpodobnostní poměr je roven apriorní pravděpodobnosti vynásobené věrohodnostním poměrem $p(y|\theta_1)/p(y|\theta_2)$.

1.3 Ilustrativní příklady

1.3.1 Genetika

Muži mají jeden chromozom X a jeden chromozom Y. Ženy mají dva chromozomy X, přičemž každý z chromozomů zdědily po jednom z rodičů. Hemofilie je nemoc, která se dědí skrze chromozom X. Muž, který zdědí takto zasažený chromozom X, je nemocí zasažen. Aby hemofilii onemocněla žena, musely by být nemocí zasaženy oba její chromozomy X. Hemofilie je pro ženy smrtelná. Pravděpodobnost, že žena zdědí dva chromozomy X zasažené nemocí je však téměř zanedbatelná, protože je výskyt takto poškozených genů v populaci velmi malý.

Apriorní pravděpodobnost

Předpokládáme, že žena má otce, který hemofilii netrpí a bratra, který je nemocí zasažen. To znamená, že její matka je skrytá přenašečka hemofilie. Žena má tak padesátí procentní pravděpodobnost, že je také skrytou přenašečkou ($\theta = 1$), a padesátí procentní pravděpodobnost, že je zcela zdravá ($\theta = 0$). Jinými slovy platí $Pr(\theta = 1) = Pr(\theta = 0) = \frac{1}{2}$.

Model a data

Dále předpokládáme, že žena má dva syny. Nechtě $y_i = 1$ resp. $y_i = 0$ podle toho, zda-li je syn hemofilik či nikoliv. Vyšetření ukázalo, že oba synové jsou zdraví, tj. $y_1 = y_2 = 0$. Funkce věrohodnosti tak má podobu

$$Pr(y_1 = 0, y_2 = 0 | \theta = 1) = 0.50 \cdot 0.50 = 0.25$$

$$Pr(y_1 = 0, y_2 = 0 | \theta = 0) = 1.00 \cdot 1.00 = 1.00.$$

Aposteriorní pravděpodobnost

Aposteriorní pravděpodobnost, že námi uvažované žena je skrytá přenašečka, je pro $y = (y_1, y_2)$ definována jako

$$Pr(\theta = 1 | y) = \frac{p(y | \theta = 1) Pr(\theta = 1)}{p(y | \theta = 1) Pr(\theta = 1) + p(y | \theta = 0) Pr(\theta = 0)} = \frac{0.25 \cdot 0.50}{0.25 \cdot 0.50 + 1.00 \cdot 0.50} = \frac{0.125}{0.625} = 0.20.$$

Výsledek je shodný s intuicí - skutečnost, že má žena dva zdravé syny, snižuje pravděpodobnost, že je skrytá přenašečka.

Přidání vícero dat

Předpokládáme, že žena má ještě třetího syna, který je také zdravý. Celý výpočet není třeba opakovat - pouze použijeme výše uvedenou aposteriorní pravděpodobnost jako novou apriorní pravděpodobnost, čímž získáme

$$Pr(\theta = 1 | y_1, y_2, y_3) = \frac{0.50 \cdot 0.20}{0.50 \cdot 0.20 + 1.00 \cdot 0.80} = 0.111.$$

θ	$p(\theta)$
“random”	$7.60 \cdot 10^{-5}$
“radon”	$6.05 \cdot 10^{-6}$
“radom”	$3.12 \cdot 10^{-7}$

Tabulka 1.1: Apriorní pravděpodobnosti pro vybraná anglická slova

1.3.2 Překlepy

Předpokládáme, že v anglicky psaném textu narazíme na výraz “radom”. Jaká je pravděpodobnost, že autor měl na mysli “random” spíše než “radom”? Jestliže použijeme y pro označení slova, které autor skutečně napsal a θ pro označení slova, které chtěl napsat, pak je tuto pravděpodobnost možné vyjádřit v nenormalizovaném tvaru jako

$$Pr(\theta | y = \text{“radom”}) \propto p(\theta) Pr(y = \text{“radom”} | \theta). \quad (1.7)$$

Pro zjednodušení budeme uvažovat pouze tři možnosti pro zamýšlené slovo θ a to “random”, “radon” a “radom”. Normalizovanou aposteriorní pravděpodobnost je tak možné vyjádřit jako

$$p(\theta = \text{“random”} | y = \text{“radom”}) = \frac{p(\theta_1) p(y = \text{“radom”} | \theta_1)}{\sum_{j=1}^3 p(\theta_j) p(y = \text{“radom”} | \theta_j)},$$

kde $\theta_1 = \text{“random”}$, $\theta_2 = \text{“radon”}$ a $\theta_3 = \text{“radom”}$. Apriorní pravděpodobnosti $p(\theta_j)$ je možné odvodit z frekvence jejich výskytu v rozsáhlém textu. Pravděpodobnosti $p(y | \theta_j)$ je pak možné získat na základě modelu popř. empirických statistik překlepů.

Apriorní pravděpodobnost

Jak již bylo zmíněno výše, apriorní pravděpodobnost je možné odvodit z frekvencí výskytu daných slov. Pro naše účely jsme tyto pravděpodobnosti získali od společnosti Google. Protože uvažujeme pouze tři možné hodnoty parametru θ , musíme apriorní pravděpodobnosti normalizovat konstantou $p(\theta = \text{“random”}) = \frac{760}{760+60.5+3.12}$, viz. tabulka 1.1.

Ačkoliv je nejmenší ze všech tří, frekvence slova “radom” je na první pohled překvapivá. Jedná se však o název města, které hostí největší letecký den v Polsku.

Věrohodnostní pravděpodobnost

Tabulka 1.2 shrnuje podmíněné pravděpodobnosti získané na základě modelu překlepů od společnosti Google. Jedná se o podmíněné pravděpodobnosti určitého jevu (výskytu slova “radom”) ze tří rozdílných pravděpodobnostních rozdělení odpovídající třem různým hodnotám neznámého parametru θ . Konkrétně nám tato tabulka říká, že s 97.00% pravděpodobností autor skutečně zamýšlel napsat slovo “radom”, s 0.20% pravděpodobností chtěl namísto “radom” napsat “random” a s téměř zanedbatelnou pravděpodobností měl na mysli slovo “radon”.

θ	$p(y = \text{"radom"} \theta)$
"random"	0.001930
"radon"	0.000143
"radom"	0.975000

Tabulka 1.2: Podmíněné pravděpodobnosti pro vybraná anglická slova

θ	$p(\theta)p(y = \text{"radom"} \theta)$	$p(\theta y = \text{"radom"})$
"random"	$1.47 \cdot 10^{-7}$	0.325
"radon"	$8.65 \cdot 10^{-10}$	0.002
"radom"	$3.04 \cdot 10^{-7}$	0.673

Tabulka 1.3: Aposteriorní pravděpodobnosti pro vybraná anglická slova

Apsteriorní pravděpodobnost

Násobením apriorní pravděpodobnosti s věrohodností a následnou normalizací získáme aposteriorní pravděpodobnosti, které jsou zachycené v tabulce 1.3. Z této tabulky vyplývá, že slovo "radom" je s přibližně dvakrát větší pravděpodobností správné, než že se jedná o překlep ve slově "random".

1.3.3 Americký fotbal

Bodové rozpětí

Pro každý zápas amerického fotbalu stanovují sportovní experti tzv. bodové rozpětí. Jedná se o očekávaný počet bodů, o který favorizovaná strana v průměru porazí protivníka. Pokud je tedy např. tým A 3.50 bodový favorit nad týmem B, pak se sázka, že tým A porazí tým B o čtyři či více bodů, brána jako "férová".

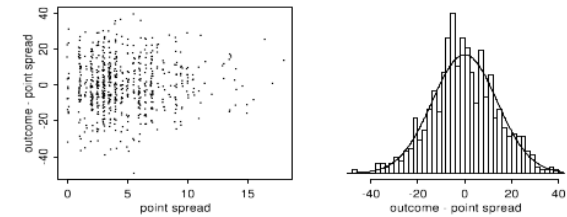
Parametrický model pro rozdíl mezi výsledkem zápasu a bodovým rozpětím

Definujeme náhodnou veličinu $d = (\text{výsledek zápasu} - \text{bodové rozpětí})$. Obrázek 1.1a ilustruje vztah mezi bodovým rozpětím (osa x) a náhodnou veličinou d (osa y) pomocí grafu pro 672 zápasů. V ideálním případě by d mělo být koncentrované okolo $y = 0$. Obrázek 1.2b pak představuje histogram náhodné veličiny d pro těchto 672 zápasů. V našem konkrétním případě lze použít aproximaci $d|x \sim N(0, 14^2)$, která je taktéž zanesena do obrázku 1.1b.

Určení pravděpodobnosti pomocí parametrického modelu

Jestliže má d normální rozdělení s nulovou střední hodnotou a je nezávislé na bodovém spreadu, pak je pravděpodobnost, že favorizovaný tým vyhraje o více než bodový spread je 50%. Pokud bodový spread označíme náhodnou veličinou x , pak lze tuto pravděpodobnost vyjádřit jako

$$Pr_{norm}(y > 0|x) = Pr_{norm}(d > -x|x) = 1 - \Phi\left(-\frac{x}{14}\right).$$



Obrázek 1.1: (a) bodové rozpětí versus rozdíl mezi skutečným výsledkem zápasu a bodovým rozpětím; (b) histogram rozdílu mezi skutečným výsledkem zápasu a bodovým rozpětím, aproximace pomocí $N(0, 14^2)$

1.3.4 Párování záznamů

Párováním záznamů rozumíme identifikaci záznamů odpovídajících stejnému objektu v různých databázích. V následujícím textu budeme tento přístup aplikovat na oficiální a testovací sčítání lidu v USA, tj. na dvou databázích.

V teorii se těší velkému zájmu párování záznamů, kdy jsou jednotlivým polím v databázi přiřazeny váhy. S pomocí těchto vah je pak zkonstruováno skóre y vyjadřující míru shody mezi dvěma záznamy. Skóre je dle zvolené metodologie vypočteno pro všechny možné páry a každý záznam je napárován se záznamem, se kterým ho poji nejvyšší skóre.

V následujícím textu použijeme Bayesianáskou pravděpodobnost, abychom získali objektivní pravděpodobnost shody pro dané rozhodovací pravidlo. Jinými slovy, pravděpodobnost shody definujeme jako funkci y .

Empirický odhad pravděpodobnosti shody

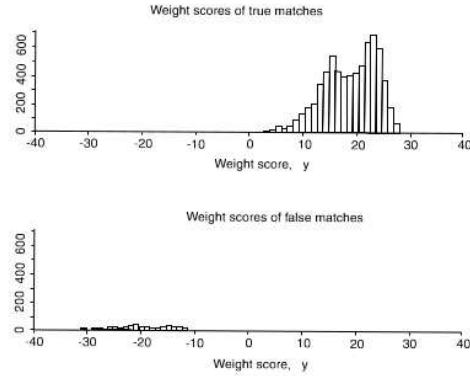
Přesné pravděpodobnosti lze získat pomocí tzv. směšného modelu (mixture model), který představíme v kapitole 22 a jehož parametry jsou odhadnuty z dat.

Tento přístup podporuje graf 1.2, který zobrazuje histogram skóre pro skutečné a falešné shody pro 2300 vybraných pozorování z testovacího a oficiálního sčítání lidu. Jak ilustruje graf 1.2, pravděpodobnost rozdělení skutečné shody $p(y|match)$ a falešné shody $p(y|non-match)$ jsou oddělené. To znamená, že ve většině případů lze identifikovat "kandidáta" pro napárování pouze na základě skóre.

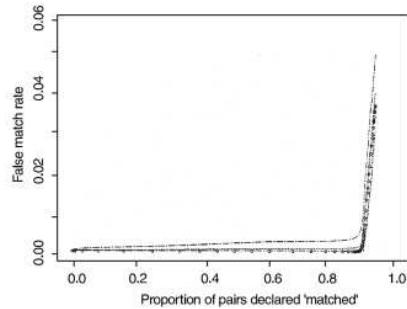
Pro směšný model lze pravděpodobnostní rozdělení skóre zapsat jako

$$p(y) = Pr(match)p(y|match) + Pr(non-match)p(y|non-match). \quad (1.8)$$

Abychom mohli použít metodu směšného modelu, zkonstruujeme křivku míry falešné shody, kdy dva záznamy jsou automaticky prohlášené za napárované, pokud jejich skóre překročí určitou hranici. Pro danou hranici je možné pomocí vztahu (1.8) odhadnout pravděpodobnost falešné shody při dosažení skóre y . Je zřejmé, že čím nižší je daná hranice, tím vyšší je počet automaticky napárovaných záznamů, což se však děje za cenu vyšší míry falešné shody.

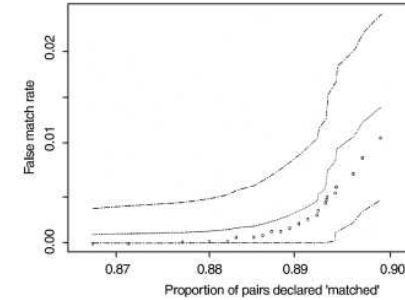


Obrázek 1.2: Histogram skóre y pro skutečné a falešné shody. Většina shod v datovém vzorku je skutečná a výše uvedená pravděpodobnostní rozdělení jsou z větší části oddělená.



Obrázek 1.3: Graf zobrazuje očekávanou falešnou míru shody (a 95% interval spolehlivosti) jako podíl na celkovém počtu napárovaných záznamů podle smíšeného modelu. Tečková čára ukazuje skutečnou míru falešné shody na pozorovaných datech.

Obrázek 1.4 představuje zvětšený výřez obrázku 1.3. Z něj je patrné, že od určité hranice automaticky napárovaných záznamů rapidně narůstá míra falešné shody.



Obrázek 1.4: Zvětšený výřez obrázku 1.4 v oblasti, pro kterou pozorujeme strmý nárůst míry falešné shody. Z obrázku je patrné, že ideálním stavem se jeví, pokud automaticky napárujeme přibližně 88% záznamů a zbylé záznamy napárujeme ručně.

1.4 Několik užitečných poznatků z teorie pravděpodobnosti

1.4.1 Podmíněné pravděpodobnosti

V Bayesiánské statistice se výpočty sdružené pravděpodobnosti $p(u, v)$ často odkazují na podmíněnou pravděpodobnost $p(u|v)$ a marginální pravděpodobnost $p(u) = \int p(u, v)dv$. Sdruženou pravděpodobnost lze totiž vyjádřit jako součin marginální a podmíněné pravděpodobnosti, tj. jako $p(u, v, w) = p(u|v, w)p(v|w)p(w)$.

1.4.2 Hypotézy

V zájmu zjednodušení zápisu však často opomíjíme podmínění pravděpodobnosti platností určité hypotézy. Koneckonců žádná pravděpodobnostní soud nemůže vzniknout ve vzduchoprázdnu. Správně bychom tak namísto $p(\theta, y) = p(\theta)p(y|\theta)$ měli psát $p(\theta, y|H) = p(\theta|H)p(y|\theta, H)$, kde H představuje soubor hypotéz a předpokladů, které definují námi uvažovaný model.

1.4.3 Střední hodnota a rozptyl

Střední hodnotu označujeme jako $E(\cdot)$ a rozptyl jako $var(\cdot)$, kde

$$E(u) = \int up(u)du, \quad var(u) = \int (u - E(u))^2 p(u)du. \quad (1.9)$$

Pokud má u podobu vektoru, je výraz pro střední hodnotu stejný a kovarianční matice je definována jako

$$\text{var}(u) = \int (u - E(u))(u - E(u))^T p(u) du. \quad (1.10)$$

1.4.4 Modelování a podmíněná pravděpodobnost

V praxi použitelné modely často vyjadřují pravděpodobnostní rozdělení pozorovaných veličin v podmíněném popř. hierarchickém tvaru spíše než v podobě složitých nepodmíněných pravděpodobnostních rozdělení. Pro ilustraci uvažujme y , které představuje výšku náhodně vybraného univerzitního studenta. Marginální pravděpodobnost $p(y)$ je výsledkem kombinace dvou normálních rozdělení se střední hodnotou 160 and 175 centimetrů. Praktičtější popis pravděpodobnosti $p(y)$ by byl však založený na sdruženém rozdělení výšky a pohlaví - $p(\text{male}) = p(\text{female}) \approx \frac{1}{2}$ spolu s informací, že $p(y|\text{male})$ a $p(y|\text{female})$ jsou přibližně normálně rozdělené se střední hodnotou 175 a 160 centimetrů. Pokud není rozptyl obou normálních rozdělení příliš vysoký, pak bude výsledné pravděpodobnostní rozdělení y bimodální.

1.4.5 Střední hodnota a rozptyl podmíněných rozdělení

Střední hodnotu u lze vypočítat jako průměrnou střední hodnotu přes marginální rozdělení v , tj. jako

$$E(u) = E(E(u|v)). \quad (1.11)$$

Výše uvedený vztah lze poměrně snadno odvodit.

$$E(u) = \int \int up(u, v) dudv = \int \int up(u|v)dup(v)dv = \int E(u|v)p(v)dv. \quad (1.12)$$

Odpovídající výsledek pro rozptyl se skládá ze dvou částí - ze střední hodnoty podmíněného rozptylu a z rozptylu podmíněné střední hodnoty.

$$\text{var}(u) = E(\text{var}(u|v)) + \text{var}(E(u|v)). \quad (1.13)$$

Tento vztah se dá dokázat, pokud pravou stranu (1.13) rozepíšeme do tvaru

$$\begin{aligned} E(\text{var}(u|v)) + \text{var}(E(u|v)) &= E\left(E(u^2|v) - (E(u|v))^2\right) + E(E(u|v))^2 - \left(E(E(u|v))\right)^2 \\ &= E(u^2) - E\left((E(u|v))^2\right) + E\left((E(u|v))^2\right) - \left(E(u)\right)^2 \\ &= E(u^2) - \left(E(u)\right)^2 \\ &= \text{var}(u) \end{aligned} \quad (1.14)$$

Výše uvedené vztahy platí také v případě, kdy u je vektor - $E(u)$ je pak vektorem a $\text{var}(u)$ maticí.

1.4.6 Transformace proměnných

Předpokládejme, že $p_u(u)$ je hustota pravděpodobnosti vektoru u a že transformace $v = f(u)$ má stejný počet členů jako u .

Jestliže je p_u diskrétní pravděpodobnostní rozdělení a f je funkce, která každému u přiřadí právě jedno v , pak je hustota pravděpodobnosti v definována jako

$$p_v(v) = p_u(f^{-1}(v)). \quad (1.15)$$

Pokud je f funkce, která několika u může přiřadit stejné v , pak

$$p_v(v) = \sum_{u \in f^{-1}(v)} p_u(f^{-1}(v)). \quad (1.16)$$

V případě, kdy je p_u spojitě a f každému u přiřadí právě jedno v , pak je sdružená hustota transformovaného vektoru u dána

$$p_v(v) = |J|p_u(f^{-1}(v)), \quad (1.17)$$

kde $|J|$ je Jacobián transformace $u = f^{-1}(v)$.¹ Pokud může f několika u přiřadit stejné v , je $p_v(v)$ reprezentováno sumou nebo integrálem.

V případě jednorozměrné veličiny často používáme přirozený logaritmus, abychom transformovali prostor z $(0, \infty)$ na $(-\infty, \infty)$. V případě veličiny definované na otevřeném jednotkovém intervalu $(0, 1)$ pak používáme tzv. logistickou transformaci

$$\text{logit}(u) = \ln\left(\frac{u}{1-u}\right), \quad (1.18)$$

jejíž inverzní transformace má podobu

$$\text{logit}^{-1}(v) = \frac{e^v}{1 + e^v}. \quad (1.19)$$

Další běžnou transformací je tzv. probit transformace $\Phi^{-1}(u)$, která se používá k přemapování prostoru $(0, 1)$ na $(-\infty, \infty)$.

¹Jacobián J je čtvercová matice parciálních derivací, jejíž (i, j) -tý člen je roven $\partial u_i / \partial v_j$.

Kapitola 2

Jednparametrové modely

2.1 Odhad pravděpodobnosti z binomických dat

V binomickém modelu odhadujeme neznámou pravděpodobnost úspěchu v populaci na základě výsledků posloupnosti Bernoulliho pokusů. Připomeňme, že výsledek pozorování v rámci Bernoulliho pokusu může nabývat pouze dvou stavů - úspěchu a neúspěchu. Ze základů statistiky víme, že binomický model je definován jako

$$p(y|\theta) = \text{Bin}(y|n, \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}, \quad (2.1)$$

kde n představuje celkový počet pokusů, y celkový počet úspěchů a θ hledanou pravděpodobnost úspěchu v populaci.

2.1.1 Ilustrační příklad - odhad pravděpodobnosti odhadu narození ženy

Aposteriorní pravděpodobnost

Předpokládejme, že počet všech narozených dětí je n , z čehož je y ženského pohlaví. Nejprve je třeba zvolit apriorní rozdělení parametru θ , které vyjadřuje podíl žen na celkovém počtu narozených dětí pro danou populaci. Jestliže má apriorní rozdělení parametru θ charakter uniformního rozdělení nad intervalem $[0, 1]$, lze aplikací Bayesova pravidla (1.3) na (2.1) odvodit aposteriorní rozdělení parametru θ jako

$$\begin{aligned} Pr(\theta \in (\theta_1, \theta_2)|y) &= \frac{Pr(\theta \in (\theta_1, \theta_2), y)}{p(y)} \\ &= \frac{\int_{\theta_1}^{\theta_2} p(y|\theta)p(\theta)d\theta}{p(y)} \\ &= \frac{\int_{\theta_1}^{\theta_2} \binom{n}{y} \theta^y (1 - \theta)^{n-y} d\theta}{p(y)}. \end{aligned} \quad (2.2)$$

Lze dokázat, že jmenovatel (2.2) je možné upravit do tvaru

$$\begin{aligned} p(y) &= \int_0^1 \binom{n}{y} \theta^y (1 - \theta)^{n-y} d\theta \\ &= \frac{1}{n+1} \quad y = 0, \dots, n. \end{aligned} \quad (2.3)$$

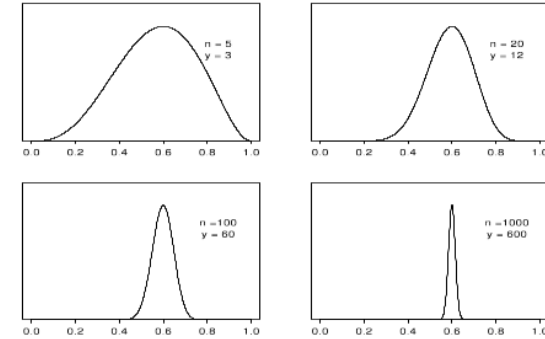
Tento výsledek nám říká, že apriori jsou všechny hodnoty y stejně pravděpodobné.¹ Aposteriorní rozdělení tak má podobu

$$p(\theta|y) = \frac{\binom{n}{y}}{n+1} \theta^y (1 - \theta)^{n-y} d\theta. \quad (2.4)$$

Pro fixní hodnoty n a y není $\frac{\binom{n}{y}}{n+1}$ funkcí neznámého parametru θ . Aposteriorní rozdělení parametru θ tak můžeme chápat jako nenormalizované beta rozdělení, tj.

$$p(\theta|y) \propto \theta^y (1 - \theta)^{n-y}. \quad (2.5)$$

Vztah (2.5) nám pro zvolené n a y umožňuje grafickou prezentaci aposteriorního rozdělení.



Obrázek 2.1: Nenormalizované aposteriorní rozdělení parametru θ pro vybrané hodnoty n a y za předpokladu apriorního uniformního rozdělení parametru θ

Historická vsuvka

Již před dvěma stoletími dokázal Laplace, že pravděpodobnost narození ženy je menší než 50.00%. Při řešení problémů aplikoval binomický model doplněný o předpoklad apriorního uniformního rozdělení parametru θ , který představoval pravděpodobnost narození ženy. V Paříži se letech 1745 až 1770 se narodilo

¹Citatel (2.2) představuje nekompletní beta integrál, který nemá analytické řešení pro velká y a $n - y$.

241 945 dětí ženského a 251 527 dětí mužského pohlaví. Na základě těchto pozorování Laplace dokázal, že

$$Pr(\theta \geq 0.50|y = 241\,945, n = 241\,945 + 251\,527) \approx 1.15 \cdot 10^{-42},$$

což mu dávalo právo tvrdit, že populační hodnota parametru θ je menší než 0.50.

Aposteriorní predikce

Apriorní prediktivní rozdělení lze vyjádřit analyticky pomocí (2.3). Tato pravděpodobnost říká, že libovolná hodnota y z intervalu nula až n má pravděpodobnost $\frac{1}{n+1}$, což je logický důsledek našeho předpokladu o uniformním rozdělení parametru θ .

Aposteriorní prediktivní rozdělení zohledňuje výsledek dodatečných Bernoulliho pokusů. Pro zjednodušení předpokládáme pouze jeden dodatečný pokus, jehož výsledek označme \tilde{y} . Platí

$$\begin{aligned} Pr(\tilde{y} = 1|y) &= \int_0^1 Pr(\tilde{y} = 1|\theta, y)p(\theta|y)d\theta \\ &= \int_0^1 \theta p(\theta|y)d\theta = E(\theta|y) = \frac{y+1}{n+2}, \end{aligned} \quad (2.6)$$

což vyplývá z vlastností beta rozdělení. Tento výsledek, založený na předpokladu apriorního uniformního rozdělení, je známý jako Laplaceův zákon posloupnosti (Laplace's law of succession).

2.2 Aposteriorní pravděpodobnost jako kompromis mezi daty a apriorní pravděpodobností

Proces Bayesovské inference zahrnuje přechod od apriorního rozdělení $p(\theta)$ k posteriornímu rozdělení $p(\theta|y)$. Je tak přirozené očekávat, že mezi těmito dvěma rozděleními mohou existovat určité vazby.

Na jednu z těchto vazeb jsme již narazili ve vztazích (1.11) a (1.13), kdy výměnou (u, v) za (θ, y) získáváme

$$E(\theta) = E(E(\theta|y)) \quad (2.7)$$

a

$$var(\theta) = E(var(\theta|y)) + var(E(\theta|y)). \quad (2.8)$$

Připomeňme, že apriorní střední hodnota parametru θ je průměrem posteriorních středních hodnot přes všechna možná data. Aposteriorní rozptyl parametru θ je pak v průměru menší než apriorní rozptyl, kdy výše rozdílu závisí na druhém členu pravé strany rovnice (2.8), tj. na rozptylu posteriorních středních hodnot přes všechna možná data. Čím vyšší je rozptyl těchto posteriorních středních hodnot, tím vyšší je potenciál redukce nejistoty ohledně parametru θ . Je nutné si však uvědomit, že vztahy (2.7) a (2.8) popisují pouze očekávané

hodnoty a v některých situacích může být např. posteriorní rozptyl i vyšší než apriorní rozptyl.

Pro výše uvedený příklad binomického modelu s apriorním uniformním rozdělením parametru θ platí, že apriorní střední hodnota $E(\theta)$ je rovna $\frac{1}{2}$. Aposteriorní střední hodnota $\frac{y+1}{n+2}$ je pak kompromisem mezi apriorní střední hodnotou a výběrovou střední hodnotou $\frac{y}{n}$. Toto je obecná vlastnost Bayesovské inference - posteriorní rozdělení je centrované na bodě, který představuje kompromis mezi apriorním rozdělením a napozorovanými daty, přičemž role dat vzrůstá s jejich objemem.

2.3 Charakteristika posteriorního rozdělení

Aposteriorní rozdělení v sobě zahrnuje všechny aktuální informace o parametru θ . Aposteriorní rozdělení $p(\theta|y)$ je vhodné graficky znázornit tak, jak jsme učinili na obrázku 2.1. V řadě praktických případů je však graf posteriorního rozdělení žádoucí doplnit o numerické statistiky jako jsou střední hodnota, medián, módus či směrodatná odchylka. Nejistotu posteriorního rozdělení lze také popsat pomocí kvantilů a rozpětím mezi těmito kvantily. Pouze v omezeném počtu případů je možné kvantily a jejich rozpětí vyjádřit analyticky. Vždy je však možné je vypočítat metodou Monte-Carlo.

2.4 Informativní apriorní rozdělení

Apriorní rozdělení lze interpretovat dvěma způsoby. Můžeme ho chápat jako rozdělení, které charakterizuje populaci možných hodnot parametru θ , z nichž je pak parametr náhodně vybrán. Nebo ho můžeme chápat jako rozdělení, které odráží všechny naše znalosti a předpoklady týkající se parametru θ , přičemž hodnotu tohoto parametru chápeme jako náhodnou realizaci z apriorního rozdělení.

Apriorní rozdělení by mělo zahrnovat všechny "uvěřitelné" hodnoty parametru θ , však nemusí být nezbytně koncentrováno kolem skutečné hodnoty. Často se totiž stává, že informace o θ obsažená v datech převáží jakoukoliv "rozumnou" specifikaci apriorního rozdělení. S tímto přístupem jsme se setkali už v předchozím textu, kdy jsme uvažovali apriorní uniformní rozdělení parametru θ na intervalu $[0, 1]$, ačkoliv bychom pravděpodobně mohli uvažovat mnohem užší interval.

2.4.1 Binomický model s alternativním apriorním rozdělením

Předpokládáme, že věrohodností funkce (2.1), kterou lze chápat jako funkci parametru θ , má tvar

$$p(y|\theta) \propto \theta^y (1-\theta)^{n-y}. \quad (2.9)$$

Jedná se tedy o beta rozdělení $y|\theta \sim \text{Beta}(y+1, n-y-1)$. Pokud má apriorní rozdělení stejnou formu, tj.

$$p(\theta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1}, \quad (2.10)$$

pak má posteriorní rozdělení taktéž shodnou formu.

Předpokládejme, že máme k dispozici “rozumné” hodnoty parametrů α a β . Aposteriorní rozdělení je pak definováno jako

$$\begin{aligned} p(\theta|y) &\propto \theta^y (1-\theta)^{n-y} \theta^{\alpha-1} (1-\theta)^{\beta-1} \\ &= \theta^{y+\alpha-1} (1-\theta)^{n+y+\beta-1} \\ &= \text{Beta}(\theta|\alpha+y, \beta+n-y) \end{aligned} \quad (2.11)$$

Jestliže má aposteriorní rozdělení stejnou parametrickou formu jako apriorní rozdělení, hovoříme o tzv. konjugaci (conjugacy).

Aposteriorní střední hodnota parametru θ má tvar

$$E(\theta|y) = \frac{\alpha+y}{\alpha+\beta+n} \quad (2.12)$$

a nachází se tak vždy mezi výběrovým podílem $\frac{y}{n}$ a apriorní střední hodnotou $\frac{\alpha}{\alpha+\beta}$. Aposteriorní rozptyl je dán

$$\text{var}(\theta|y) = \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} = \frac{E(\theta|y)[1-E(\theta|y)]}{\alpha+\beta+n+1}. \quad (2.13)$$

Pro fixní α a β a vysoké y a $n-y$ platí $E(\theta|y) \approx \frac{y}{n}$ a $\text{var}(\theta|y) \approx \frac{1}{n} \frac{y}{n} (1-\frac{y}{n})$, přičemž $\text{var}(\theta|y)$ blíží nule rychlostí $\frac{1}{n}$. Limitně tak parametry apriorního rozdělení α a β nemají vliv na podobu aposteriorního rozdělení.

Pomocí centrální limitní věty zasazené do kontextu Bayesiánské statistiky lze dokázat

$$\left(\frac{\theta - E(\theta|y)}{\sqrt{\text{var}(\theta|y)}} | y \right) \rightarrow N(0, 1), \quad (2.14)$$

což se často používá pro ospravedlnění aproximace aposteriorního rozdělení pomocí normálního rozdělení. V případě binomického parametru θ je aproximace pomocí normálního rozdělení přesnější, pokud použijeme transformaci $\ln(\theta(1-\theta))$ namísto původního θ . Touto transformací “roztáhneme” pravděpodobnostní prostor z $[0, 1]$ na $(-\infty, \infty)$, což je z pohledu normálního rozdělení vhodnější.

2.4.2 Konjugátní apriorní rozdělení

Jestliže je \mathcal{F} třída výběrového rozdělení $p(y|\theta)$ a \mathcal{P} je třída apriorního rozdělení parametru θ , pak \mathcal{P} je konjugát k \mathcal{F} , jestliže platí

$$p(\theta|y) \in \mathcal{P} \text{ pro všechna } p(\cdot|\theta) \in \mathcal{F} \text{ a } p(\cdot) \in \mathcal{P}. \quad (2.15)$$

Výše uvedená definice je formálně vágní. Lze totiž vybrat \mathcal{P} jako třídu všech rozdělení, což znamená, že \mathcal{P} je vždy konjugát bez ohledu na výběrové rozdělení. My se však především zajímáme o tzv. přirozené konjugáty. Jedná se o případ, kdy \mathcal{P} je množina všech rozdělení, které mají stejný funkcionální tvar jako věrohodnostní funkce.

2.4.3 Nekongugátní apriorní rozdělení

Zásadní výhodou konjugátního apriorního rozdělení je snadná interpretace výsledků, které je často možné vyjádřit v analytické podobě, což značně usnadňuje navazující výpočty.

V praxi je často pro složitější modely nemožné nalézt vhodné konjugátní apriorní rozdělení. Ačkoliv je interpretace jejich aposteriorních inferencí méně transparentní a jejich výpočet je složitější, nepředstavují nekongugátní apriorní rozdělení koncepční problém. V následujícím textu si představíme několik takovýchto příkladů.

2.4.4 Konjugátní apriorní rozdělení, rodina exponenciálních pravděpodobnostních rozdělení a dostatečná statistika

\mathcal{F} je rodinou exponenciálních pravděpodobnostních rozdělení, pokud všechny její členy mají podobu

$$p(y_i|\theta) = f(y_i)g(\theta)e^{\phi(\theta)^T u(y_i)}. \quad (2.16)$$

$\phi(\theta)$ a $u(y_i)$ jsou vektory se stejnou dimenzí jako θ . Vektor $\phi(\theta)$ je nazýván přirozeným parametrem rodiny \mathcal{F} . Věrohodnostní funkce odpovídající posloupnosti $y = (y_1, \dots, y_n)$ nezávislých a identicky rozdělených pozorování má tvar

$$p(y|\theta) = \left(\prod_{i=1}^n f(y_i) \right) g(\theta)^n e^{\left(\phi(\theta)^T \sum_{i=1}^n u(y_i) \right)}. \quad (2.17)$$

Pro dané n a y lze výše uvedený vztah chápat jako funkci parametru θ , tj.

$$p(y|\theta) \propto g(\theta)^n e^{\phi(\theta)^T t(y)}, \quad (2.18)$$

kde $t(y) = \sum_{i=1}^n u(y_i)$. $t(y)$ označujeme jako dostatečnou statistiku parametru θ , protože věrohodnostní funkce parametru θ závisí na pozorováních y pouze skrze $t(y)$. Pokud má apriorní rozdělení podobu

$$p(\theta) \propto g(\theta)^n e^{\phi(\theta)^T \nu}, \quad (2.19)$$

pak má aposteriorní rozdělení podobu

$$p(\theta|y) \propto g(\theta)^{\eta+n} e^{\phi(\theta)^T (\nu + t(y))}, \quad (2.20)$$

což mimo jiné dokazuje, že námi zvolené apriorní rozdělení je konjugát. Lze dokázat, že rodina exponenciálních pravděpodobnostních rozdělení je jedinou rodinou pravděpodobnostních rozdělení, které mají přirozené konjugátní apriorní rozdělení. V předchozím textu jsme diskutovali binomické rozdělení s věrohodnostní funkcí $p(y|\theta, n) = \text{Bin}(y|n, \theta)$, kde pro známé n má konjugátní apriorní rozdělení parametru θ charakter beta rozdělení. Lze dokázat, že binomické rozdělení patří do rodiny exponenciálních rozdělení s přirozeným parametrem $\logit(\theta)$.

2.4.5 Ilustrativní příklad

Jako specifický příklad faktoru, který může ovlivnit podíl narozených dětí ženského pohlaví, můžeme uvažovat tzv. placenta previa. Jedná se o situaci, kdy je během těhotenství placenta napojena v nižší části dělohy, což znemožňuje

klasický porod. V Německu proběhla studie, která se na tuto problematiku zaměřovala. V jejím průběhu bylo zkoumáno 980 porodů, z nichž připadalo 437 na novorozence ženského pohlaví. Lze na základě těchto poznatků tvrdit, že v souvislosti s placenta previa je relativní podíl novorozenců ženského pohlaví nižší než je tomu v celkové populaci, tj. než 0.485?

Apriorní uniformní rozdělení

Za předpokladu apriorního uniformního rozdělení je aposteriorní pravděpodobnost narození jedince ženského pohlaví dána beta rozdělením $Beta(438, 544)$. Aposteriorní střední hodnota a rozptyl parametru θ jsou tak 0.446 a 0.016. 95% interval spolehlivosti je pak možné získat numerickou integrací beta rozdělení nebo aproximací pomocí normálního rozdělení. S přesností na tři desetinná místa se v obou případech jedná o interval $[0.414, 0.476]$.

Jak jsme však již zmínili dříve, v případě aproximace rozdělení parametru θ pomocí normálního rozdělení je vhodnější použít logit transformaci.² Zpřesnění pomocí logit transformace je nejmarkantnější v případě výběru menšího rozsahu a v případě, kdy θ zahrnuje hodnoty 0 a 1.

Alternativní konjugátní apriorní rozdělení

Citlivost aposteriorní inference ohledně θ v závislosti na uvažovaném apriorním rozdělení je ilustrována tabulkou na obrázku (2.2)

Parameters of the prior distribution		Summaries of the posterior distribution	
$\frac{\alpha}{\alpha+\beta}$	$\alpha + \beta$	Posterior median of θ	95% posterior interval for θ
0.500	2	0.446	[0.415, 0.477]
0.485	2	0.446	[0.415, 0.477]
0.485	5	0.446	[0.415, 0.477]
0.485	10	0.446	[0.415, 0.477]
0.485	20	0.447	[0.416, 0.478]
0.485	100	0.450	[0.420, 0.479]
0.485	200	0.453	[0.424, 0.481]

Obrázek 2.2: Souhrnné statistiky aposteriorního rozdělení parametru θ ; pravděpodobnost narození jedince ženského pohlaví v případě placenta previa pro různá konjugátní apriorní rozdělení

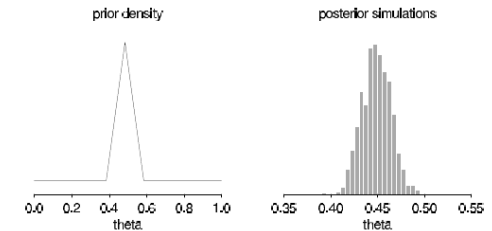
První řádek tabulky odpovídá apriornímu rozdělení s parametry $\alpha = 1$ a $\beta = 1$. Následující řádky pak reprezentují apriorní rozdělení, která jsou více a více koncentrována okolo hodnoty 0.485. První sloupec představuje apriorní hodnotu parametru θ , druhý sloupec představuje “objem” apriorní informace vyjádřené

²Nejprve vygenerujeme z $Beta(438, 544)$ rozdělení tisíc náhodných pozorování, na které aplikujeme logit transformaci. Následně z transformovaných pozorování vypočteme střední hodnotu a rozptyl. Na základě odhadnuté střední hodnoty a rozptylu pak pomocí aproximace normálním rozdělením určíme požadovaný konfidenční interval. Nakonec tento konfidenční interval “převeďme” pomocí inverze zpět do původní škály parametru θ .

jako $\alpha + \beta^3$. Tabulka také ilustruje skutečnost, že aposteriorní inference založené na velkém počtu pozorování jsou na volbu apriorního rozdělení citlivé jen v omezené míře. Pouze pro spodní řádky tabulky, kdy “objem” apriorní informace odpovídá 100 až 200 porodů, je aposteriorní rozdělení vychýlenější směrem k apriornímu rozdělení a i tomto případě stále 95% konfidenční interval nezahrnuje apriorní střední hodnotu.

Nekonjugátní apriorní rozdělení

Jako alternativu ke konjugátní rodině beta rozdělení můžeme uvažovat apriorní rozdělení, které je koncentrováno okolo hodnoty 0.485 a směrem k chvostům se zplošťuje. Jedním z příkladů takového rozdělení je rozdělení na obrázku (2.3a). Tato apriorní rozdělení má střední hodnotu 0.493 a směrodatnou odchylku 0.21, což je podobné jako v případě beta rozdělení s parametry $\alpha + \beta = 5$. Nenormalizované aposteriorní rozdělení je možné získat pomocí mřížky (0.000, 0.001, ..., 1.000) pro hodnoty parametru θ tak, že vynásobíme apriorní hustotu pravděpodobnosti s binomickou věrohodnostní funkcí pro každý bod mřížky. Aposteriorní simulace je pak možné získat normalizací rozdělení nad mřížkou uvažovaných hodnot parametru θ . Obrázek (2.3b) představuje histogram 1,000 náhodných výběrů z aposteriorní pravděpodobnostní funkce. Aposteriorní střední hodnota je 0.448 a 95% konfidenční interval je $[0.419, 0.480]$. Protože je apriorní pravděpodobnostní funkce “přebita” aposteriorními pozorováními, odpovídají tyto výsledky výsledkům z tabulky na obrázku (2.2).



Obrázek 2.3: (a) apriorní hustota pravděpodobnosti parametru θ v ilustrativním příkladě nekonjugátní analýzy; (b) histogram 1,000 náhodných výběrů z diskretní aproximace aposteriorní hustoty pravděpodobnosti

2.5 Odhad normální střední hodnoty při známém rozptylu

Pomocí centrální limitní věty lze ospravedlnit aplikaci normální věrohodnostní funkce v situacích, kdy je použití teoreticky správnější věrohodnostní funkce z analytického pohledu problematické. Nicméně v následující kapitole budeme předpokládat, že je použití normálního rozdělení nejen žádoucí ale také vhodné.

³Připomeňme si, že $\alpha + \beta - 2$ lze v jistém slova smyslu chápat jako počet apriorních pozorování.

2.5.1 Věrohodnostní funkce jednoho pozorování

Uvažujeme pozorování y náhodně vybrané z normální rozdělení se střední hodnotou θ a známým rozptylem σ^2 . Výběrové rozdělení má tvar

$$p(y|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\theta)^2}. \quad (2.21)$$

Konjugátní apriorní a aposteriorní rozdělení

Jestliže je věrohodnostní funkce exponenciála kvadratické formy v parametru θ , pak má rodina konjugátních apriorních rozdělení podobu

$$p(\theta) = e^{A\theta^2 + B\theta + C}. \quad (2.22)$$

Tuto rodinu apriorních rozdělení “parametrizujeme” jako

$$p(\theta) \propto e^{-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2}, \quad (2.23)$$

s hyperparametry μ_0 a τ_0^2 , tj. $\theta \sim N(\mu_0, \tau_0^2)$. Pro zjednodušení předpokládáme, že známe jak μ_0 tak τ_0^2 .

Zvolené konjugátní apriorní rozdělení implikuje, že aposteriorní rozdělení parametru θ je rovněž exponenciála kvadratické formy a tudíž normální.⁴ V případě aposteriorních rozdělení mají všechny parametry s výjimkou θ charakter konstanty, což vede k podmíněné hustotě pravděpodobnosti

$$p(\theta|y) \propto e^{-\frac{1}{2}\left(\frac{(y\theta)^2}{\sigma^2} + \frac{(\theta - \mu_0)^2}{\tau_0^2}\right)}, \quad (2.24)$$

kteou lze dále upravit do tvaru

$$p(\theta|y) \propto e^{-\frac{1}{2\tau_1^2}(\tau - \mu_1)^2}, \quad (2.25)$$

tj. $\theta|y \sim N(\mu_1, \tau_1^2)$, kde

$$\mu_1 = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{1}{\sigma^2}y}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}} \quad (2.26)$$

a

$$\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}. \quad (2.27)$$

Normální rozdělení a přesnost

Při práci s normálním rozdělením hraje prominentní úlohu inverze rozptylu, kterou označujeme jako přesnost (precision). Rovnice (2.27) ilustruje, že v případě pozorování, které sledují normální rozdělení, a apriorního normálního rozdělení, je aposteriorní přesnost rovna součtu apriorní přesnosti a přesnosti pozorovaných dat.

Existuje několik způsobů, jak interpretovat aposteriorní střední hodnotu μ_1 . V (2.26) je aposteriorní střední hodnota vyjádřena jako průměr apriorní střední hodnoty a pozorování y s použitím vah, které jsou proporcionální přesnostem

⁴Nicméně může být zapotřebí “trocha” algebry, abychom zjistili její konkrétní podobu.

$1/\tau_0^2$ a $1/\sigma^2$. Alternativně lze vyjádřit μ_1 jako apriorní střední hodnotu “upravenou” o pozorování y , tj. jako

$$\mu_1 = \mu_0 + (y - \mu_0) \frac{\tau_0^2}{\sigma^2 + \tau_0^2} \quad (2.28)$$

nebo také jako pozorování y “vychýlená” směrem k apriorní střední hodnotě, tj. jako

$$\mu_1 = y - (y - \mu_0) \frac{\sigma^2}{\sigma^2 + \tau_0^2}. \quad (2.29)$$

Všechny tři výše uvedené interpretace aposteriorní střední hodnoty lze chápat jako kompromis mezi apriorní střední hodnotou a pozorovanými daty.

Aposteriorní prediktivní rozdělení

Aposteriorní prediktivní rozdělení lze vyjádřit pomocí integrace (1.5) jako

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y}|\theta)p(\theta|y)d\theta \\ &\propto \int e^{-\frac{1}{2\sigma^2}(\tilde{y}-\theta)^2} e^{-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2} d\theta. \end{aligned} \quad (2.30)$$

První řádek výše uvedené rovnice je platný, protože budoucí pozorování \tilde{y} pro dané θ nezávisí na minulých pozorováních y . Pravděpodobnostní rozdělení \tilde{y} můžeme jednodušeji určit s využitím vlastností dvourozměrného normálního rozdělení. Součin v integrálu druhého řádku je exponenciála kvadratické formy (\tilde{y}, θ) . \tilde{y} a θ tak sledují aposteriorní sdružené normální rozdělení, což znamená, že marginální aposteriorní rozdělení \tilde{y} je také normální.

S využitím znalosti aposteriorního rozdělení, pro které platí $E(\tilde{y}|\theta) = \theta$ a $\text{var}(\tilde{y}) = \sigma^2$, společně s (2.7) a (2.8) lze odvodit aposteriorní střední hodnotu

$$E(\tilde{y}|y) = E(E(\tilde{y}|\theta, y)|y) = E(\theta|y) = \mu_1 \quad (2.31)$$

a aposteriorní rozptyl

$$\begin{aligned} \text{var}(\tilde{y}|y) &= E(\text{var}(y|\tilde{\theta}, y)|y) + \text{var}(E(\tilde{y}|\theta, y)|y) \\ &= E(\sigma^2|y) + \text{var}(\theta|y) \\ &= \sigma^2 + \tau_1^2. \end{aligned} \quad (2.32)$$

Jinými slovy, aposteriorní prediktivní rozdělení \tilde{y} má střední hodnotu rovnou aposteriorní střední hodnotě θ a rozptyl, který se skládá z prediktivního rozptylu σ^2 a rozptylu τ_1^2 z titulu aposteriorní nejistoty parametru θ .

2.5.2 Normální rozdělení a vícero pozorování

V předchozím textu jsme se zabývali situací, kdy jsou aposteriorní data reprezentována jedním pozorováním. Nyní uvažujeme realističtější situaci, kdy jsou

aposteriorní data reprezentována vícero pozorováními $y = (y_1, \dots, y_n)$. Aposteriorní hustota pravděpodobnosti je pak definována jako

$$\begin{aligned} p(\theta|y) &\propto p(\theta)p(y|\theta) \\ &= p(\theta) \prod_{i=1}^n p(y_i|\theta) \\ &\propto e^{-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2} \prod_{i=1}^n e^{-\frac{1}{2\sigma^2}(y_i-\theta)^2} \\ &\propto e^{-\frac{1}{2}\left(\frac{1}{\tau_0^2}(\theta-\mu_0)^2 + \frac{1}{\sigma^2} \sum_{i=1}^n \sum_{i=1}^n (y_i-\theta)^2\right)} \end{aligned} \quad (2.33)$$

Zjednodušením výše uvedeného vztahu lze dokázat, že aposteriorní rozdělení závisí na y pouze skrze výběrový průměr $\bar{y} = \frac{1}{n} \sum_i y_i$, tj. \bar{y} je dostatečná statistika. Protože $\bar{y}|\theta, \sigma^2 \sim N(\theta, \sigma^2/n)$ lze závěry pro jedno pozorování přímo aplikovat také na vícero pozorování, tj.

$$p(\theta|y_1, \dots, y_n) = p(\theta|\bar{y}) = N(\theta|\mu_n, \tau_n^2), \quad (2.34)$$

kde

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \quad (2.35)$$

a

$$\frac{1}{\tau_0^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}. \quad (2.36)$$

Stejný výsledek lze získat také tak, že budeme aposteriorní rozdělení postupně aktualizovat o pozorování y_1, \dots, y_n .

Rovnice (2.36) nám říká, že pokud $\tau_0 \rightarrow \infty$ pro fixní n nebo $n \rightarrow \infty$ pro fixní τ_0^2 , pak

$$p(\theta|y) \approx N(\theta|\bar{y}, \sigma^2/n), \quad (2.37)$$

což je praxi vhodná aproximace bez ohledu na apriorní rozdělení.

2.6 Ostatní jednoparametrové modely

V praxi obecně neplatí, že aposteriorní rozdělení $p(\theta|y)$ musí mít vždy analytickou formu. Příčinou bývá normalizační konstanta $p(y)$, která s ohledem na integrál v (1.4) často nemá analytické řešení. Nicméně mnoho formálních analýz se soustředí na situace, pro které je dipozici analytické aposteriorní rozdělení.

Standardní aposteriorní rozdělení, jako jsou binomické, normální, Poissonovo a exponenciální rozdělení, lze odvodit při řešení jednoduchých pravděpodobnostních modelů. Jak jsme již zmínili v předchozím textu, binomické rozdělení lze použít v situacích, kdy se snažíme kvantifikovat frekvenci určitého jevu v populaci. Normální rozdělení pak používáme pro náhodné veličiny, které jsou součtem mnoha nezávislých veličin. Poissonovo resp. exponenciální rozdělení se používají pro modelování počtu událostí resp. pro modelování čekací doby (waiting times).

2.6.1 Normální rozdělení se známou střední hodnotou a neznámým rozptylem

Pro $p(y|\theta, \sigma^2) = N(y|\theta, \sigma^2)$ se známou střední hodnotou θ a neznámým rozptylem σ^2 má věrohodnostní funkce vektoru $y = (y_1, \dots, y_n)$ podobu

$$\begin{aligned} p(y|\sigma^2) &\propto \sigma^{-n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta)^2} \\ &= (\sigma^2)^{-n/2} e^{-\frac{n}{2\sigma^2} v}, \end{aligned} \quad (2.38)$$

kde

$$v = \frac{1}{n} \sum_{i=1}^n (y_i - \theta)^2 \quad (2.39)$$

je dostatečná statistika.

Odpovídající konjugátní apriorní rozdělení má pak podobu inverzního gamma rozdělení

$$p(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} e^{-\beta/\sigma^2} \quad (2.40)$$

s hyperparametry (α, β) . Tento vztah lze však také vyjádřit pomocí inverzního χ^2 rozdělení s parametrem σ_0^2 a ν_0 stupni volnosti. Jinými slovy, parametr σ^2 sleduje rozdělení $\sigma_0^2 \nu_0 / X$ kde X je $\chi_{\nu_0}^2$ náhodná veličina; v následujícím textu budeme použít poněkud nestandardní označení $\sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$.

Výsledné aposteriorní rozdělení parametru σ^2 má pak tvar

$$\begin{aligned} p(\sigma^2|y) &\propto p(\sigma^2)p(y|\sigma^2) \\ &\propto \left(\frac{\sigma_0^2}{\sigma^2}\right)^{\frac{\nu_0}{2}+1} e^{-\frac{\nu_0 \sigma_0^2}{2\sigma^2}} (\sigma^2)^{-n/2} e^{-\frac{v}{2\sigma^2}} \\ &\propto (\sigma^2)^{-\frac{n+\nu_0}{2}+1} e^{-\frac{1}{2\sigma^2}(\nu_0 \sigma_0^2 + nv)} \end{aligned} \quad (2.41)$$

nebo-li

$$\sigma^2|y \sim \text{Inv-}\chi^2\left(\nu_0 + n, \frac{\nu_0 \sigma_0^2 + nv}{\nu_0 + n}\right), \quad (2.42)$$

což je inverzní χ^2 rozdělení se škálovacím parametrem $\frac{\nu_0 \sigma_0^2 + nv}{\nu_0 + n}$ a $\nu_0 + n$ stupni volnosti. Ve světle (2.42) je tak na apriorní rozdělení možné pohlížet jako na aposteriorní rozdělení založené na ν_0 pozorováních s rozptylem σ_0^2 .

2.6.2 Poissonův model

Poissonovo rozdělení se běžně používá v situacích, kdy je předmětem zájmu frekvence určité události v populaci, jako je tomu např. v epidemiologii. Jestliže jedno jednotlivé pozorování y sleduje Poissonovo rozdělení s parametrem θ , pak je jeho pravděpodobnostní rozdělení definováno jako

$$p(y|\theta) = \frac{\theta^y e^{-\theta}}{y!} \quad \text{kde } y = 0, 1, \dots \quad (2.43)$$

V případě vektoru $y = (y_1, \dots, y_n)$ má pak věrohodnostní funkce podobu

$$\begin{aligned} p(y|\theta) &= \prod_{i=1}^n \frac{1}{y_i!} \theta^{y_i} e^{-\theta} \\ &\propto \theta^{t(y)} e^{-n\theta}, \end{aligned} \quad (2.44)$$

kde $t(y) = \sum_{i=1}^n y_i$ je dostatečná statistika. Věrohodnostní funkci můžeme přepsat do tvaru rodiny exponenciálních rozdělení jako

$$p(y|\theta) \propto e^{-n\theta e^{t(y)} \ln(\theta)} \quad (2.45)$$

s přirozeným parametrem $\phi(\theta) = \ln(\theta)$.

Přirozené konjugátní apriorní rozdělení má podobu

$$p(\theta) \propto (e^{-\theta})^\eta e^{\nu \ln(\theta)} \quad (2.46)$$

s hyperparametry (η, ν) . Jinými slovy, věrohodnostní funkce má tvar $\theta^\alpha e^{-b\theta}$, a proto musí mít apriorní pravděpodobnost tvar $p(\theta) \propto \theta^A e^{-B\theta}$. Výše uvedenou formu apriorního rozdělení tak lze také vyjádřit jako

$$p(\theta) \propto e^{-\beta\theta} \theta^{\alpha-1}, \quad (2.47)$$

což je gamma rozdělení s parametry α a β . Porovnáním $p(y|\theta)$ a $p(\theta)$ zjistíme, že apriorní rozdělení lze v jistém slova smyslu chápat jako aposteriorní rozdělení založené na $\alpha - 1$ realizacích z celkového počtu β pozorování.

Pro výše uvedené apriorní rozdělení má pak aposteriorní rozdělení podobu

$$\theta|y \sim \text{Gamma}(\alpha + n\bar{y}, \beta + n). \quad (2.48)$$

2.6.3 Negativní binomický model

V případě konjugátních rodin pravděpodobnostních rozdělení lze použít znalost apriorního a aposteriorního rozdělení k odvození marginálního rozdělení $p(y)$ pomocí

$$p(y) = \frac{p(y|\theta)p(\theta)}{p(\theta|y)}. \quad (2.49)$$

Například pro Poissonův model má apriorní prediktivní rozdělení tvar

$$\begin{aligned} p(y) &= \frac{\text{Poisson}(y|\theta)\text{Gamma}(\theta|\alpha, \beta)}{\text{Gamma}(\theta|\alpha + y, 1 + \beta)} \\ &= \frac{\Gamma(\alpha + y)\beta^\alpha}{\Gamma(\alpha)y!(1 + \beta)^{\alpha+y}}, \end{aligned} \quad (2.50)$$

což lze dále zredukovat na

$$p(y) = \binom{\alpha + y + 1}{y} \left(\frac{\beta}{\beta + 1}\right)^\alpha \left(\frac{1}{\beta + 1}\right)^y. \quad (2.51)$$

Tento vztah představuje tzv. negativní binomické rozdělení. V následujícím textu pro něj budeme používat notaci $y \sim \text{Neg-bin}(\alpha, \beta)$. Výše uvedené odvození ukazuje, že negativní binomické rozdělení je kombinací Poissonových rozdělení s parametry θ , které sledují gamma rozdělení, tj.

$$\text{Neg-bin}(y|\alpha, \beta) = \int \text{Poisson}(y|\theta)\text{Gamma}(\theta|\alpha, \beta)d\theta. \quad (2.52)$$

2.6.4 Poissonův model parametrizovaný pomocí míry a expozice

V řadě praktických situací je žádoucí rozšířit Poissonův model o pozorování y_1, \dots, y_n do podoby

$$y_i \sim \text{Poisson}(x_i\theta), \quad (2.53)$$

kde x_i jsou známe kladné hodnoty vysvětlující veličiny a θ je hledaný parametr. V epidemiologii je θ často nazýván mírou a x_i expozicí. Tento model sice není zaměnitelný z pohledu y_i (tj. záleží na pořadí hodnot y_i), nicméně je zaměnitelný z pohledu $(x, y)_i$.

Věrohodnostní funkce v tomto rozšířeném Poissonově modelu má podobu

$$p(y|\theta) \propto \theta^{\sum_{i=1}^n y_i} e^{-(\sum_{i=1}^n x_i)\theta} \quad (2.54)$$

při ignorování členů, které jsou nezávislé na θ . To znamená, že gamma rozdělení parametru θ je konjugátní. Pro apriorní rozdělení

$$\theta \sim \text{Gamma}(\alpha, \beta) \quad (2.55)$$

tak získáváme aposteriorní rozdělení

$$\theta|y \sim \text{Gamma}\left(\alpha + \sum_{i=1}^n y_i, \beta + \sum_{i=1}^n x_i\right). \quad (2.56)$$

Ilustrativní příklad

Předpokládejme, že ve městě s 200,000 obyvateli zemřeli v minulém roce na astma tři lidé. Na základě této informace můžeme odhadnout míru úmrtnosti v souvislosti s astmatem na 1.50 případů na 100,000 obyvatel. Pro Poissonův model můžeme výběrovou funkci pro y definovat jako $\text{Poisson}(2.0\theta)$, kde θ představuje skutečnou dlouhodobou míru úmrtnosti na astma. V našem konkrétním případě víme, že $y = 3$ a $x = 2.0$ (protože θ je vyjádřeno na 100,000 obyvatel), a hledáme hodnotu parametru θ . Pro konstrukci apriorního rozdělení parametru θ můžeme použít celosvětové statistiky úmrtí na astma a zkombinovat je s informací $y = 3$ s cílem získat aposteriorní rozdělení.

Mortalita v zemích západní Evropy je výrazně nižší než v případě našeho města - cca 0.60 případů na 100,000 obyvatel. V podobných případech je vhodné apriorní rozdělení parametru θ aproximovat pomocí gamma rozdělení. Jako přijatelný kandidát se jeví $\text{Gamma}(3.0, 5.0)$ rozdělení, které má střední hodnotu 0.60 a pro které je 97.50% všech hodnot menších než 1.44. V praxi nejčastěji kalibrujeme hledané gamma rozdělení tak, že podíl parametrů zafixujeme v souladu s apriorní střední hodnotou a následně hledáme jejich konkrétní hodnotu tak, aby výsledné rozdělení splňovalo naše představy konkrétního kvantilu.

Vztah (2.56) nám říká, že aposteriorní rozdělení parametru θ pro apriorní rozdělení $\text{Gamma}(\alpha, \beta)$ má tvar $\text{Gamma}(\alpha + y, \beta + x)$, což v našem konkrétním případě znamená $\text{Gamma}(6.0, 7.0)$. Toto pravděpodobnostní rozdělení má střední hodnotu 0.86, což je v porovnání s napozorovanou frekvencí 1.50 případů na 100,000 obyvatel významně blíže apriorní střední hodnotě 0.60.

Předpokládejme, že namísto jednoho roku máme k dispozici pozorování za deset let, během kterých zemřelo v našem městě 30 lidí. Aposteriorní rozdělení parametru θ má pak podobu $\text{Gamma}(33.0, 25.0)$ a aposteriorní střední hodnotu 1.32, tj. mnohem blíže pozorované úmrtnosti.

2.6.5 Exponenciální model

Exponenciální rozdělení je běžně používáno pro modelování doby, která uplyne mezi dvěma událostmi, tzv. čekací doby. Výběrové rozdělení y pro daný parametr θ má tvar

$$p(y|\theta) = \theta e^{-y\theta} \text{ pro } y > 0, \quad (2.57)$$

kde $\theta = 1/E(y|\theta)$ nazýváme mírou. Matematicky je exponenciální rozdělení speciálním případem gamma rozdělení s parametry $(\alpha, \beta) = (1, \theta)$.

Exponenciální rozdělení je bez paměti, což z něj činí přirozenou volbu pro modelování přežití resp. dožití. Např. pravděpodobnost, že daný subjekt přežije dodatečný čas v délce t , je nezávislá na čase s , který již uplynul, tj. $Pr(y > t + s | y > s, \theta) = Pr(y > t | \theta)$ pro libovolné s a t . Konjugátní apriorní rozdělení pro exponenciální parametr θ je (stejně jako pro Poissonovo rozdělení) $Gamma(\theta|\alpha, \beta)$ s odpovídajícím aposteriorním rozdělením $Gamma(\theta|\alpha + 1, \beta + y)$. Výběrové rozdělení n nezávislých pozorování $y = (y_1, \dots, y_n)$ s konstantní mírou θ má tvar

$$p(y|\theta) = \theta^n e^{-n\bar{y}\theta} \text{ pro } \bar{y} \geq 0, \quad (2.58)$$

a je, pokud ho chápeme jako věrohodnostní funkci parametru θ pro fixní y , proporcionální $Gamma(n+1, n\bar{y})$ rozdělení. Proto je možné na apriorní rozdělení $Gamma(\alpha, \beta)$ parametru θ pohlížet jako na $\alpha - 1$ exponenciálních rozdělení s celkovou čekací dobou β .

2.7 Ilustrativní příklad - rakovina ledvin v USA

2.7.1 Rozložení případů napříč USA

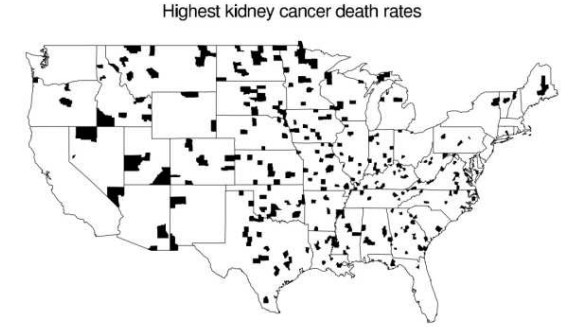
Následující mapa USA zobrazuje okresy s nejvyšší mírou úmrtí na rakovinu ledvin v průběhu let 1980 - 1989. Z mapy se zdá, že mezi nejvíce postižené okresy patří ty z oblasti Velkých plání. Pokud bychom však měli k dispozici podrobnější informace, zjistili bychom, že se na Velkých plání současně nachází také okresy s nejnižší mírou úmrtí. Důvodem tohoto zdánlivého paradoxu je velikost okresů. Uvažujme okres s 1,000 obyvatel. Rakovina ledvin je poměrně vzácné onemocnění a tak se může snadno stát, že v průběhu deseti let nikdo nezemře a okres tak bude vykazovat nulovou míru úmrtnosti. Naopak, pokud zemře jeden člověk, bude okres vykazovat míru úmrtnosti 1 na 10,000 člověkoroků, což je dostatečné na to, aby figuroval na výše zmiňované mapě. A právě Velké pláně jsou charakteristické velkým počtem okresů s nízkým počtem obyvatel.

2.7.2 Bayesiánská inference

Paradox Velkých plání indikuje, že je pro jednotlivé okresy vhodné použít model

$$y_i \sim \text{Poisson}(10 \cdot n_j \theta_j), \quad (2.59)$$

kde y_j je počet úmrtí na rakovinu ledvin v rámci okresu j , n_j je počet obyvatel okresu, a θ_j je míra úmrtnosti přepočtena na osobu a rok. Zásadní rozdíl oproti (2.53) je ten, že míra úmrtnosti je specifická pro každý okres, tj. θ_j je odhadováno separátně pro každý okres.



Obrázek 2.4: Okresy v USA s nejvyšší mírou úmrtnosti na rakovinu ledvin v průběhu let 1980 - 1989

Pro účely Bayesiánské inference potřebujeme definovat apriorní rozdělení neznámého parametru θ_j . Pro tyto účely použijeme gamma rozdělení $Gamma(20, 430,000)$, které konjuguje do Poissonova rozdělení. Apriorní rozdělení má střední hodnotu $\frac{\alpha}{\beta} = 4.65 \cdot 10^{-5}$ a směrodatnou odchylku $\frac{\sqrt{\alpha}}{\beta} = 1.04 \cdot 10^{-5}$.

Aposteriorní rozdělení má pak podobu

$$\theta_j | y_j \sim Gamma(20 + y_j, 430,000 + 10n_j) \quad (2.60)$$

se střední hodnotou

$$E(\theta_j | y_j) = \frac{20 + y_j}{430,000 + 10n_j} \quad (2.61)$$

a rozptylem

$$var(\theta_j | y_j) = \frac{20 + y_j}{(430,000 + 10n_j)^2}. \quad (2.62)$$

Aposteriorní střední hodnotu je možné chápat jako vážený průměr pozorování míry úmrtnosti $\frac{y_j}{10n_j}$ a apriorní střední hodnoty $\frac{\alpha}{\beta} = 4.65 \cdot 10^{-5}$.

2.7.3 Důsledky velikosti okresu

Relativní váha apriorní informace se odvíjí od velikosti populace n_j . Pro ilustraci uvažujme okres s $n_j = 1,000$.

- Jestliže $y_j = 0$, pak se pozorovaná míra úmrtnosti rovna nule, avšak aposteriorní střední hodnota je $\frac{20}{440,000} = 4.55 \cdot 10^{-5}$.
- Jestliže $y_j = 1$, je pozorovaná míra úmrtnosti rovna $1.00 \cdot 10^{-4}$, nicméně aposteriorní střední hodnota je pouze $\frac{21}{440,000} = 4.77 \cdot 10^{-5}$.
- Jestliže $y_j = 2$, je pozorovaná míra úmrtnosti rovna $2.00 \cdot 10^{-4}$, ale aposteriorní střední hodnota je stále $\frac{21}{440,000} = 4.77 \cdot 10^{-5}$.

Je tedy zřejmé, že pro okres s malým počtem obyvatel je aposteriorní rozdělení “definováno” apriorním rozdělením. V případě okresu, který má např. 1,000,000 obyvatel by tomu bylo opačně.

V Poissonově modelu (2.59) je rozptyl pozorované úmrtnosti $\frac{y_i}{10 \cdot n_j}$ inverzní k expozičnímu parametru n_j . Jinými slovy, okresy s malým počtem obyvatel vykazují vyšší míru fluktuace v úmrtnosti. To vysvětluje, proč okresy s malým počtem obyvatel patří současně mezi okresy s malou i velkou mírou úmrtí na rakovinu ledvin.

2.8 Neinformativní apriorní rozdělení

V případě, že je apriorní rozdělení spíše než na datech založeno na expertním odhadu, nazýváme toto apriorní rozdělení neinformativním. Vzhledem k značné nejistotě neinformativního apriorního rozdělení je přirozená snaha o minimalizaci jeho vlivu na aposteriorní rozdělení.

Kromě neinformativního apriorního rozdělení rozeznáváme také slabě informativní apriorní rozdělení zohledňujícího určitou omezenou informaci, která je např. dostatečná na to, aby vytyčila rozumné rozmezí hodnot hledaného parametru.

2.8.1 Vlastní na nevlastní apriorní rozdělení

Vraťme se k problematice odhadu střední hodnoty θ v rámci normálního modelu se známým rozptylem σ^2 a apriorním rozdělením $N(\mu_0, \tau_0^2)$ parametru θ . Jestliže je apriorní přesnost $1/\tau_0^2$ v porovnání s datovou přesností n/σ^2 malá, pak je aposteriorní rozdělení přibližně stejné jako v případě $\tau_0^2 = \infty$, tj.

$$p(\theta|y) \approx N(\theta|\bar{y}, \sigma^2/n). \quad (2.63)$$

Aposteriorní rozdělení je tak přibližně rovno rozdělení, které bychom získali za předpokladu, že $p(\theta)$ je proporcionální konstantě pro $\theta \in (-\infty, \infty)$. Takovéto rozdělení však není možné, protože integrál přes uvažované $p(\theta)$ by nebyl konečný, což je v rozporu s předpokladem, že součet pravděpodobností je roven jedné. Obecně nazýváme apriorní rozdělení $p(\theta)$ vlastním (proper), pokud je nezávislé na pozorovaných datech a jeho integrál je roven jedné.⁵ Ačkoliv je apriorní rozdělení uvedené v tomto případě nevlastní (improper), aposteriorní rozdělení je vlastní.

Jako druhý příklad neinformativního apriorního rozdělení uvažujeme normální model se známou střední hodnotou a neznámým rozptylem s konjugátním apriorním inverzním χ^2 rozdělením. Jestliže je apriorní počet stupňů volnosti ν_0 v porovnání se stupni volnosti dat n malý, pak je aposteriorní rozdělení přibližně stejné jako v případě $\nu_0 = 0$, tj.

$$p(\sigma^2|y) \approx \text{Inv-}\chi^2(\sigma^2|n, v). \quad (2.64)$$

K této limitní formě aposteriorního rozdělení lze také dojít tak, že apriorní rozdělení parametru σ^2 definujeme jako $p(\sigma^2) \propto 1/\sigma^2$. Protože je jeho integrál nad intervalem $(0, \infty)$ nekonečný, je toto rozdělení nevlastním rozdělením.

⁵Pokud je integrál $p(\theta)$ konečný a větší než jedna, nazýváme $p(\theta)$ nenormalizovaným rozdělením. Vynásobením vhodnou konstantou je možné $p(\theta)$ znormalizovat tak, aby byl integrál roven jedné.

2.8.2 Nevlastní apriorní vs. vlastní aposteriorní rozdělení

Definujeme nenormalizované aposteriorní rozdělení

$$p(\theta|y) \propto p(y|\theta)p(\theta). \quad (2.65)$$

Ve výše uvedených příkladech je aposteriorní rozdělení vlastní, tj. $\int p(\theta|y)d\theta$ je konečné pro všechna y . To však nemusí vždy platit. Při interpretaci aposteriorního rozdělení založeného na nevlastním apriorním rozdělení je třeba zvýšené opatrnosti. Je nutné zkontrolovat, že aposteriorní rozdělení má konečný integrál a smysluplnou formu. Takto získané aposteriorní rozdělení pak lze interpretovat jako aproximaci v případech, kdy kontribuce věrohodnostní funkce dominuje nad apriorním rozdělením.

2.8.3 Jeffreyovo pravidlo

Uvažujeme transformaci $\phi = h(\theta)$, která každé hodnotě θ přiřadí právě jednu hodnotu ϕ . Apriorní rozdělení $p(\theta)$ je tak informačně ekvivalentní apriornímu rozdělení $p(\phi)$, kdy

$$p(\phi) = p(\theta) \left| \frac{d\theta}{d\phi} \right| = p(\theta) |h'(\theta)|^{-1}. \quad (2.66)$$

Jakékoliv pravidlo, které určuje apriorní rozdělení $p(\theta)$, by tak mělo vést k ekvivalentnímu výsledku, pokud je aplikované na transformovaný parametr $h(\theta)$. Jinými slovy, $p(\phi)$ vypočtené na základě (2.66) by mělo odpovídat $p(\phi)$ získanému přímo pomocí transformačního modelu $p(y, \phi) = p(\phi)p(y|\phi)$.

Jeffreyovo pravidlo tak umožňuje definovat neinformativní apriorní rozdělení jako $p(\theta) \propto |J(\theta)|^{1/2}$, kde $J(\theta)$ je tzv. Fisherova informace pro parametr θ , tj.

$$J(\theta) = E \left(\left(\frac{d \ln(p(y|\theta))}{d\theta} \right)^2 \middle| \theta \right) = -E \left(\frac{d^2 \ln(p(y|\theta))}{d\theta^2} \middle| \theta \right). \quad (2.67)$$

Abychom se přesvědčili, že je Jeffreyův apriorní model invariantní, stačí $J(\phi)$ vyjádřit v bodě $\theta = h^{-1}(\phi)$, tj.

$$\begin{aligned} J(\phi) &= -E \left(\frac{d^2 \ln(p(y|\phi))}{d\phi^2} \right) \\ &= -E \left(d^2 \ln(p(y|\theta = h^{-1}(\phi))) \left| \frac{d\theta}{d\phi} \right|^2 \right) \\ &= J(\theta) \left| \frac{d\theta}{d\phi} \right|^2, \end{aligned} \quad (2.68)$$

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

2.8.4 Neinformativní apriorní rozdělení pro binomický parametr

Uvažujeme binomické rozdělení $y \sim \text{Bin}(n, \theta)$, která má logaritmickou věrohodnostní funkci

$$\ln p(y|\theta) = \text{konstanta} + y \ln(\theta) + (n - y) \ln(1 - \theta). \quad (2.69)$$

Aplikací druhé derivace a substitucí $E(y|\theta) = n\theta$ získáváme Fisherovu informaci

$$J(\theta) = -E\left(\frac{d^2 \ln(p(y|\theta))}{d\theta^2} \middle| \theta\right) = \frac{n}{\theta(1-\theta)}. \quad (2.70)$$

Nenormalizované apriorní rozdělení je pak $p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$, což je $Beta(\frac{1}{2}, \frac{1}{2})$ rozdělení.

2.8.5 Klíčová hodnota, lokační a škálovací parametr

Pro binomický model a jiné jednoparametrové modely vede aplikace rozdílných principů k rozdílným neinformativním apriorním rozdělením. Nicméně ve dvou případech - lokačních a škálovacích parametrech - se všechny principy shodují.

1. Jestliže je pravděpodobnostní funkce veličiny y taková, že $p(y - \theta|\theta)$ je funkce, která neobsahuje θ a y jako separátní proměnné - řekněme $f(u)$, kde $u = y - \theta$ - pak je $(y - \theta)$ tzv. klíčová hodnota (pivotal quantity) a θ lokační parametr (location parameter). V takovém případě je žádoucí, aby neinformativní apriorní rozdělení parametru θ implikovalo $f(y - \theta)$ pro aposteriorní rozdělení $p(y - \theta|y)$. To znamená, že i pro aposteriorní rozdělení by $(y - \theta)$ měla být klíčovou hodnotnou, a nemělo by tak neobsahovat θ ani y jako separátní proměnné. S použitím Bayesovy věty $p(y - \theta|y) \propto p(\theta)p(y - \theta|\theta)$ tato podmínka znamená, že neinformativní apriorní rozdělení je uniformní v θ , tj. $p(\theta) \propto \text{konstanta}$ nad intervalem $(-\infty, \infty)$.
2. Jestliže je pravděpodobnostní funkce veličiny y taková, že $p(\frac{y}{\theta}|\theta)$ je funkce, která neobsahuje θ a y jako separátní proměnné - řekněme $f(u)$, kde $u = \frac{y}{\theta}$ - pak je $\frac{y}{\theta}$ klíčová hodnota a θ škálovací parametr (scale parametr). V takovém případě je žádoucí, aby neinformativní apriorní rozdělení parametru θ implikovalo $f(\frac{y}{\theta})$ pro aposteriorní rozdělení $p(\frac{y}{\theta}|y)$. Transformací proměnných lze podmíněnou pravděpodobnost y pro dané θ lze vyjádřit pomocí podmíněné pravděpodobnosti u pro dané θ jako

$$p(y|\theta) = \frac{1}{\theta} p(u|\theta) \quad (2.71)$$

a podobně

$$p(\theta|y) = \frac{y}{\theta^2} p(u|\theta). \quad (2.72)$$

Předpokladem $p(u|\theta) = p(u|y) = g(u)$ se pak dostáváme k identitě $p(\theta|y) = \frac{y}{\theta} p(y|\theta)$. To implikuje apriorní rozdělení $p(\theta) \propto \frac{1}{\theta}$ popř. ekvivalentně $p(\ln(\theta)) \propto 1$ nebo $p(\theta^2) \propto \frac{1}{\theta^2}$.

2.8.6 Komplikace spojené s neinformativním apriorním rozdělením

1. Hledání “univerzálního” apriorního rozdělení je zavadějící. Jestliže věrohodnostní funkce dominuje, není volba relativně “plochého” apriorního rozdělení na škodu. Nicméně volba určitého apriorního rozdělení jako reference vede často k jeho automatické aplikaci i v situacích, kdy to není žádoucí.

2. V řadě praktických problémů neexistuje zřejmá volba pro vágní apriorní rozdělení, protože pravděpodobnostní rozdělení, které je “ploché” nebo uniformní pro jednu parametrizaci, nemusí být “ploché” nebo uniformní pro jinou parametrizaci. Například “rozhodné” apriorní rozdělení střední hodnoty θ je uniformní, zatímco v případě σ^2 se jako rozhodné apriorní rozdělení zdá $p(\sigma^2) \propto 1/\sigma^2$. Nicméně pokud definujeme transformaci $\phi = \ln(\sigma^2)$, pak má apriorní rozdělení parametru ϕ tvar

$$p(\phi) = p(\sigma^2) \left| \frac{d\sigma^2}{d\phi} \right| \propto \frac{1}{\sigma^2} \sigma^2 = 1, \quad (2.73)$$

tj. je uniformní pro $\phi = \ln(\sigma^2)$.

3. Další komplikace se objevují při průměrování několika alternativních modelů, které zahrnují nevlastní apriorní rozdělení. Touto problematikou se budeme zabývat v kapitole 7.

Navzdory výše uvedeným problémům jsou neinformativní apriorní rozdělení užitečná pokud (a) je nalezení skutečného apriorního rozdělení komplikované nebo nemožné a (b) jsme ochotni zkontrolovat smysluplnost výsledného aposteriorního rozdělení a to, že se jedná o funkci vlastni.

2.9 Slabě informativní apriorní rozdělení

Apriorní rozdělení označujeme jako slabě informativní, pokud se jedná o vlastní rozdělení, která je zkonstruována tak, že vědomě zohledňuje slabší informaci než jaká je k dispozici. V praxi často namísto úplné “informační” rezignace zkonstruujeme apriorní rozdělení, která je při relativně omezené informační hodnotě stále smysluplné. Klasickým příkladem je apriorní rozdělení poměru novorozenců ženského pohlaví v populaci. Neinformativní apriorní rozdělení může být založené na uniformním rozdělení nad intervalem $[0.00, 1.00]$; slabě informativní apriorní rozdělení by pak tento interval omezilo na $[0.40, 0.60]$. V řadě případů ani není možné popř. praktické do apriorního rozdělení zahrnout všechnu informaci, kterou máme v době konstrukce modelu k dispozici.

Slabě informativní apriorní rozdělení můžeme zkonstruovat dvěma způsoby. Buď začneme s neinformativním apriorním rozdělením, které následně “obohatíme” o část apriorní informace nebo naopak začneme s informativním apriorním rozdělením, které následně informačně “oslabíme”.

Kapitola 3

Víceparametrové modely

3.1 Průměrování přes nadbytečné parametry

Nechť je θ vektor o dvou prvcích, tj. $\theta = (\theta_1, \theta_2)$. Předpokládejme, že se zajímáme o rozdělení θ_1 podmíněné pozorovanými daty, tj. o $p(\theta_1|y)$. Toto rozdělení lze získat ze sdruženého posteriorního rozdělení

$$p(\theta_1, \theta_2|y) \propto p(y|\theta_1, \theta_2)p(\theta_1, \theta_2) \quad (3.1)$$

a to průměrováním přes θ_2 , kde

$$p(\theta_1|y) = \int p(\theta_1, \theta_2|y)d\theta_2. \quad (3.2)$$

Alternativně lze použít vztah

$$p(\theta_1|y) = \int p(\theta_1|\theta_2, y)p(\theta_2|y)d\theta_2, \quad (3.3)$$

což ilustruje skutečnost, že posteriorní rozdělení $p(\theta_1|y)$ je kombinací podmíněného posteriorního rozdělení pro daný nadbytečný parametr (nuisance parametr) θ_2 , kde $p(\theta_2|y)$ figuruje jako váha pro možné hodnoty parametru θ_2 .

Integrál (3.3) se v praxi jen zřídka vyjadřujeme analyticky. Aposteriorní rozdělení lze obvykle jednodušeji kvantifikovat pomocí marginální nebo podmíněné simulace. Nejprve náhodně vygenerujeme hodnotu parametru θ_2 na základě posteriorního rozdělení $p(\theta_2|y)$ a následně pak θ_1 na základě posteriorního rozdělení $p(\theta_1|\theta_2, y)$. Tímto způsobem lze integrál v (3.3) vyhodnotit numericky.

3.2 Normální data s neinformativním apriorním rozdělením

Uvažujme vektor y , který se skládá z n nezávislých pozorování generovaných z jednorozměrného normálního rozdělení $N(\mu, \sigma^2)$.

3.2.1 Neinformativní apriorní rozdělení

V kapitole 2 jsme si ukázali, že rozumné vágní apriorní rozdělení parametrů μ a σ má za předpokladu nezávislosti lokačního a škálovacího parametru podobu

$$p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}. \quad (3.4)$$

3.2.2 Sdružené posteriorní rozdělení $p(\mu, \sigma^2|y)$

Sdružené posteriorní rozdělení je proporcionální věrohodnostní funkci vynásobené faktorem $1/\sigma^2$, tj.

$$\begin{aligned} p(\mu, \sigma^2|y) &\propto \sigma^{-n-2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)} \\ &= \sigma^{-n-2} e^{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2 \right]} \\ &= \sigma^{-n-2} e^{-\frac{1}{2\sigma^2} [(n-1)s^2 + n(\bar{y} - \mu)^2]}, \end{aligned} \quad (3.5)$$

kde

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2. \quad (3.6)$$

Dostatečné statistiky jsou \bar{y} a s^2 .

3.2.3 Podmíněné posteriorní rozdělení $p(\mu|\sigma^2, y)$

Abychom určili posteriorní rozdělení $p(\mu|\sigma^2, y)$ podmíněné σ^2 , musíme použít výsledků kapitoly 2.5 pro střední hodnotu normálního rozdělení pro známou hodnotu rozptylu σ^2 a apriorní uniformní rozdělení, kde

$$\mu|\sigma^2, y \sim N(\bar{y}, \sigma^2/n). \quad (3.7)$$

3.2.4 Marginální posteriorní rozdělení $p(\sigma^2|y)$

Marginální posteriorní rozdělení je definováno jako průměr sdružené pravděpodobnosti (3.5) přes μ , tj.

$$p(\sigma^2|y) \propto \int \sigma^{-n-2} e^{-\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\bar{y} - \mu)^2 \right]} d\mu. \quad (3.8)$$

Výše uvedený výraz zahrnuje integrál $-\frac{1}{2\sigma^2}n(\bar{y} - \mu)^2$, což je integrál normálního rozdělení, a proto

$$\begin{aligned} p(\sigma^2|y) &\propto \sigma^{-n-2} e^{-\frac{1}{2\sigma^2}(n-1)s^2} \sqrt{2\pi\sigma^2/n} \\ &\propto (-\sigma^2)^{-(n+1)/2} e^{-\frac{(n-1)s^2}{2\sigma^2}}, \end{aligned} \quad (3.9)$$

což je škálované inverzní χ^2 rozdělení, tj.

$$\sigma^2|y \sim \text{Inv-}\chi^2(n-1, s^2). \quad (3.10)$$

Tímto způsobem jsme sdruženou pravděpodobnost (3.5) rozložili na součin podmíněné a marginální posteriorní hustoty pravděpodobnosti, tj. $p(\mu, \sigma^2|y) = p(\mu|\sigma^2, y)p(\sigma^2|y)$.

3.2.5 Výběr ze sdruženého aposteriorního rozdělení

Nejprve náhodně vybereme σ^2 z (3.10) a následně μ z (3.7).

3.2.6 Analytická forma marginálního aposteriorního rozdělení μ

Rovnice (3.3) nám říká, že aposteriorní rozdělení parametru μ lze chápat jako mix normálních rozdělení, která jsou mixována přes škálované inverzní χ^2 rozdělení parametru σ^2 . Marginální aposteriorní hustotu pravděpodobnosti parametru μ tak můžeme získat integrací sdruženého aposteriorního rozdělení přes parametr σ^2 jako

$$p(\mu|y) = \int_0^\infty p(\mu, \sigma^2|y) d\sigma^2. \quad (3.11)$$

Integrál lze vyřešit s pomocí substituce

$$z = \frac{A}{2\sigma^2}, \quad (3.12)$$

kde

$$A = (n-1)s^2 + n(\mu - \bar{y})^2 \quad (3.13)$$

a skutečností, že výsledek je nenormalizovaný gamma integrál, tj.

$$\begin{aligned} p(\mu|y) &\propto A^{-n/2} \int_0^\infty z^{(n-2)/2} e^{-z} dz \\ &\propto \left[(n-1)s^2 + n(\mu - \bar{y}) \right]^{-n/2} \\ &\propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2} \right]^{-n/2}, \end{aligned} \quad (3.14)$$

což je $t_{n-1}(\bar{y}, s^2/n)$ rozdělení.

Jinými slovy, dokázali jsme, že pro neinformativní apriorní uniformní rozdělení na $(\mu, \ln(\sigma))$ má aposteriorní rozdělení parametru μ tvar

$$\frac{\mu - \bar{y}}{s/\sqrt{n}} | y \sim t_{n-1}, \quad (3.15)$$

, kde t_{n-1} představuje standardizované t rozdělení (s lokačním parametrem 0 a škálovacím parametrem 1) s $n-1$ stupni volnosti. Toto marginální aposteriorní rozdělení demonstruje vazbu na teorii výběru (sampling theory), kde pro výběrové rozdělení $p(y|\mu, \sigma^2)$ platí

$$\frac{\bar{y} - \mu}{s/\sqrt{n}} | \mu, \sigma^2 \sim t_{n-1}. \quad (3.16)$$

V teorii výběru nezávisí rozdělení klíčové veličiny $\frac{\bar{y} - \mu}{s/\sqrt{n}}$ na nadbytečném parametru σ^2 . Aposteriorní rozdělení této veličiny nezávisí na datech.

3.2.7 Aposteriorní prediktivní rozdělení budoucích pozorování

Aposteriorní prediktivní rozdělení budoucích pozorování \bar{y} můžeme vyjádřit jako “mix” pravděpodobnostních rozdělení ve formě

$$p(\bar{y}|y) = \int \int p(\bar{y}|\mu, \sigma^2, y) p(\mu, \sigma^2|y) d\mu d\sigma^2. \quad (3.17)$$

První ze dvou činitelů výše uvedeného integrálu je normální rozdělení budoucích pozorování pro dané hodnoty parametrů (μ, σ^2) , které nezávisí na y . Abychom provedli náhodný výběr z aposteriorního prediktivního rozdělení, vybereme nejprve μ a σ^2 z jejich sdruženého aposteriorního rozdělení a následně simulujeme $\bar{y} \sim N(\mu, \sigma^2)$.

Lze dokázat, že aposteriorní rozdělení \bar{y} má charakter t rozdělení s lokačním parametrem \bar{y} a škálovacím parametrem $(1 + \frac{1}{n})^{1/2} s$ a $n-1$ stupni volnosti. Tuto analytickou formu lze získat stejným způsobem jako aposteriorní rozdělení parametru μ . Konkrétně lze rozdělení \bar{y} získat integrováním přes sdružené rozdělení parametrů μ a σ^2 s následným využitím skutečnosti, že faktorizace $p(\bar{y}|\sigma^2, y) = \int p(\bar{y}|\mu, \sigma^2, y) p(\mu|\sigma^2, y) d\mu$ implikuje $p(\bar{y}|\sigma^2, y) = N(\bar{y}|\bar{y}, (1 + \frac{1}{n})\sigma^2)$, což je až na škálovací konstantu shodné s rozdělením $\mu|\sigma^2, y$.

3.3 Normální data s apriorním konjugátním rozdělením

3.3.1 Rodina apriorních konjugátních rozdělení

Uvažujme apriorní konjugátní rozdělení dvouparametrového jednorozměrného výběrového modelu namísto výše uvedeného neinformativního apriorního rozdělení. Věrohodnostní funkce (3.5) a předchozí diskuze implikují, že apriorní konjugátní rozdělení musí mít tvar $p(\sigma^2)p(\mu|\sigma^2)$, kde marginální rozdělení parametru σ^2 má tvar škálovaného inverzního χ^2 rozdělení a podmíněné rozdělení parametru μ pro dané σ^2 má charakter normálního rozdělení.¹ Příhodná parametrizace tohoto modelu je pak dána

$$\mu|\sigma^2 \sim N(\mu_0, \sigma^2/\kappa_0) \quad (3.18)$$

$$\sigma^2 \sim Inv\text{-}\chi^2(\nu_0, \sigma_0^2), \quad (3.19)$$

což odpovídá sdružené apriorní hustotě pravděpodobnosti

$$p(\mu, \sigma^2) \propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} e^{-\frac{1}{2\sigma^2} [\nu_0\sigma_0^2 + \kappa_0(\mu_0 - \mu)^2]}. \quad (3.20)$$

Pro výše uvedené rozdělení budeme používat označení $N\text{-}Inv\text{-}\chi^2(\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2)$. Čtyři parametry tohoto rozdělení lze interpretovat jako lokační a škálovací parametr pro μ a jako počet stupňů volnosti a škálovací parametr pro σ^2 .

To, že σ^2 figuruje v podmíněné pravděpodobnosti $\mu|\sigma^2$, znamená, že μ a σ^2 jsou v rámci jejich sdruženého apriorního konjugátního rozdělení závislé - např. velké σ^2 implikuje apriorní rozdělení parametru μ s vysokým rozptylem.

¹Marginální rozdělení parametru μ má tak podobu t rozdělení.

3.3.2 Sdružené aposteriorní rozdělení $p(\mu, \sigma^2|y)$

Vynásobením apriorní hustoty pravděpodobnosti (3.20) normální věrohodnostní funkcí získáme aposteriorní hustotu pravděpodobnosti

$$\begin{aligned} p(\mu, \sigma^2|y) &\propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} e^{-\frac{1}{2\sigma^2} [\nu_0\sigma_0^2 + \kappa_0(\mu_0 - \mu)^2]} \times \\ &\times (\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} [(n-1)s^2 n(\bar{y} - \mu)^2]} \\ &= N-Inv-\chi^2(\mu_n, \sigma_n^2/\kappa_n; \nu_n, \sigma_n^2). \end{aligned} \quad (3.21)$$

S využitím algebry lze dokázat, že

$$\mu_n = \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{y} \quad (3.22)$$

$$\kappa_n = \kappa_0 + n \quad (3.23)$$

$$\nu_n = \nu_0 + n \quad (3.24)$$

$$\nu_n \sigma_n^2 = \nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n} (\bar{y} - \mu_0)^2. \quad (3.25)$$

Parametry aposteriorního rozdělení kombinují apriorní informaci s informací obsaženou v datech.

3.3.3 Podmíněné aposteriorní rozdělení $p(\mu|\sigma^2, y)$

Podmíněné aposteriorní rozdělení parametru μ pro dané σ^2 je proporcionální sdruženému aposteriornímu rozdělení (3.21) při konstantním σ^2 , tj.

$$\begin{aligned} \mu|\sigma^2, y &\sim N(\mu_n, \sigma^2/\kappa_n) \\ &N\left(\frac{\frac{\kappa_0}{\sigma_0^2} \mu_0}{\frac{\kappa_0}{\sigma_0^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{\kappa_0}{\sigma_0^2} + \frac{n}{\sigma^2}}\right), \end{aligned} \quad (3.26)$$

což odpovídá závěrům v sekci 2.5 pro μ s konstantním σ .

3.3.4 Marginální aposteriorní rozdělení $p(\sigma^2|y)$

Marginální aposteriorní rozdělení parametru σ^2 odvozené z (3.21) má charakter škálovaného inverzní χ^2 rozdělení

$$\sigma^2|y \sim Inv-\chi^2(\nu_n, \sigma_n^2). \quad (3.27)$$

3.3.5 Výběr ze sdruženého aposteriorního rozdělení

Nejprve vybereme σ^2 z marginálního aposteriorního rozdělení (3.27) a následně μ z normálního aposteriorního rozdělení (3.26) podmíněného σ^2 z předchozího kroku.

3.3.6 Analytická forma marginálního aposteriorního rozdělení parametru μ

Integrací sdružené aposteriorní hustoty pravděpodobnosti vzhledem k σ^2 získáme

$$\begin{aligned} p(\mu|y) &\propto \left(1 + \frac{\kappa_0(\mu - \mu_n)^2}{\nu_n \sigma_n^2}\right)^{-(\nu_n+1)/2} \\ &= t_{\nu_n}(\mu|\mu_n, \sigma_n^2/\kappa_n). \end{aligned} \quad (3.28)$$

3.4 Multinomický model pro kategorická data

Binomický model, kterým jsme se zabírali v kapitole 2, lze zobecnit pro více než dva možné výsledky. Předpokládejme, že pozorování může skončit jedním z k možných výsledků. Jestliže je y vektorem počtu jednotlivých výsledků, pak

$$p(y|\theta) \propto \prod_{j=1}^k \theta_j^{y_j}, \quad (3.29)$$

kde součet pravděpodobností $\sum_{j=1}^k \theta_j$ je roven jedné. U výše uvedeného rozdělení typicky předpokládáme implicitní podmínění na celkový počet pozorování $\sum_{j=1}^k y_j = n$. Konjugátní apriorní rozdělení je vícerozměrným zobecněním beta rozdělení

$$p(\theta|\alpha) \propto \prod_{j=1}^k \theta_j^{\alpha_j-1}, \quad (3.30)$$

které je omezeno na nezáporné hodnoty parametrů θ_j , kde $\sum_{j=1}^k \theta_j = 1$. Toto aposteriorní rozdělení parametrů θ_j je známo jako Dirichletovo rozdělení s parametry $\alpha_j + y_j$.

Apriorní rozdělení je matematicky ekvivalentní věrohodnostní funkci založené na $\sum_{j=1}^k \alpha_j$ pozorování s α_j pozorováními v j -té kategorii výsledků. Stejně jako v binomickém modelu, také zde existuje řada přijatelných neinformativních Dirichletových apriorních rozdělení. Uniformní hustotu pravděpodobnosti získáme, pokud $\alpha_j = 1$ pro všechna j .² Nastavením $\alpha_j = 0$ pro všechna j získáme nevlastní apriorní rozdělení, které je uniformní pro $\ln(\theta_j)$. Výsledné aposteriorní rozdělení je vlastní, pokud každá z k kategorií možných výsledků obsahuje alespoň jedno pozorování, tj. pokud je každý člen vektoru y kladný.

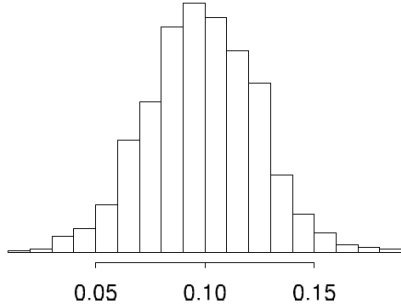
3.4.1 Ilustrativní příklad - předvolební průzkum

Uvažujme předvolební prezidentský průzkum se třemi možnými odpověďmi, který proběhl v USA v říjnu 1988. Předvolebního průzkumu se zúčastnilo 1,447 oprávněných voličů, z nichž $y_1 = 727$ vyjádřilo podporu Georgovi Bushovi, $y_2 = 583$ vyjádřilo podporu Michaelovi Dukakisovi a $y_3 = 137$ by volilo jiného kandidáta.

Je zřejmé, že data (y_1, y_2, y_3) sledují multinomické rozdělení s parametry $(\theta_1, \theta_2, \theta_3)$. Předmětem našeho zájmu je rozdíl v podpoře dvou hlavních kandidátů, tj. $\theta_1 - \theta_2$.

V případě neinformativního uniformního apriorního rozdělení pro parametr θ předpokládáme $\alpha_1 = \alpha_2 = \alpha_3 = 1$, což vede k tomu, že aposteriorní rozdělení parametru θ má charakter Dirichletova rozdělení (728, 584, 138). Aposteriorní rozdělení parametru $\theta_1 - \theta_2$ pak můžeme teoreticky určit pomocí integrace, nicméně je jednodušší náhodně vygenerovat 1,000 $(\theta_1, \theta_2, \theta_3)$ bodů z aposteriorního Dirichletova rozdělení a následně vypočítat $\theta_1 - \theta_2$ pro každý z nich. Výsledek je ilustrován obrázkem (3.1). Všechny 1,000 simulací splňují podmínku $\theta_1 > \theta_2$, a proto je odhadnutá aposteriorní pravděpodobnost, že Bush porazí Dukakisova, vyšší než 99.9%.

²Tímto způsobem přiřazujeme stejnou hustotu pravděpodobnosti libovolnému vektoru θ při splnění podmínky $\sum_{j=1}^k \theta_j = 1$



Obrázek 3.1: Histogram hodnot $(\theta_1 - \theta_2)$ pro 1,000 simulací z aposteriorního rozdělení pro ilustrativní příklad předvolebního průzkumu

3.5 Vícerozměrný normální model se známým rozptylem

3.5.1 Věrohodnostní funkce

Uvažujme vektor y , který se skládá d komponent a který lze chápat jako náhodný výběr z vícerozměrného rozdělení, tj.

$$y|\mu, \Sigma \sim N(\mu, \Sigma), \quad (3.31)$$

kde μ je sloupcový vektor délky d a Σ je $d \times d$ kovarianční matice. Věrohodnostní funkce pro jedno pozorování má podobu

$$p(y|\mu, \Sigma) \propto |\Sigma|^{-1/2} e^{-\frac{1}{2}(y-\mu)^T \Sigma^{-1}(y-\mu)} \quad (3.32)$$

a pro n nezávislých pozorování y_1, \dots, y_n sledujících identické pravděpodobnostní rozdělení pak

$$p(y_1, \dots, y_n|\mu, \Sigma) \propto |\Sigma|^{-n/2} e^{-\frac{1}{2} \sum_{i=1}^n (y_i - \mu)^T \Sigma^{-1} (y_i - \mu)} \quad (3.33)$$

$$|\Sigma|^{-n/2} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S_0)},$$

kde

$$S_0 = \sum_{i=1}^n (y_i - \mu)(y_i - \mu)^T. \quad (3.34)$$

3.5.2 Konjugátní analýza

Konjugátní apriorní rozdělení parametru μ se známou kovarianční maticí Σ

Logaritmičká věrohodnostní funkce je kvadratická v μ , a proto má konjugátní apriorní rozdělení parametru μ podobu vícerozměrného normálního rozdělení, které parametrizujeme jako $\mu \sim N(\mu_0, \Lambda_0)$.

Apsteriorní rozdělení parametru μ se známou kovarianční maticí Σ

Apsteriorní rozdělení parametru μ má tvar

$$p(\mu|y, \Sigma) \propto e^{-\frac{1}{2} \left((\mu - \mu_0)^T \Lambda_0^{-1} (\mu - \mu_0) + \sum_{i=1}^n (y_i - \mu)^T \Sigma^{-1} (y_i - \mu) \right)}, \quad (3.35)$$

což je exponenciála s kvadratickým členem v μ . Doplněním kvadratické formy a vyloučením konstantních členů získáváme

$$p(\mu|y, \Sigma) \propto e^{-\frac{1}{2} (\mu - \mu_n)^T \Lambda_n^{-1} (\mu - \mu_n)} = N(\mu|\mu_n, \Lambda_n), \quad (3.36)$$

kde

$$\mu_n = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1} (\Lambda_0^{-1} \mu_0 + n\Sigma^{-1} \bar{y}) \quad (3.37)$$

$$\Lambda_n^{-1} = \Lambda_0^{-1} + n\Sigma^{-1}. \quad (3.38)$$

Jedná se o podobné výsledky jako v případě jednorozměrného normálního modelu, který jsme představili v sekci 2.5. Apsteriorní střední hodnota je váženým průměrem apriorní střední hodnoty a dat s váhami Λ_0^{-1} a $n\Sigma^{-1}$. Apsteriorní přesnost je pak součtem apriorní a datové přesnosti.

Apsteriorní podmíněné a marginální rozdělení podmnožiny vektoru μ se známou kovarianční maticí Σ

Marginální apsteriorní rozdělení podmnožiny parametrů $\mu^{(1)}$ sleduje taktéž vícerozměrné normální rozdělení. Toto rozdělení je charakterizováno vektorem středních hodnot, který odpovídá podmnožině apsteriorních středních hodnot μ_n , a kovarianční maticí, která odpovídá podmnožině kovarianční matice Λ_n . Také apsteriorní rozdělení podmíněné zbylými parametry μ^2 sleduje vícerozměrné normální rozdělení, konkrétně

$$\mu^{(1)}|\mu^{(2)}, y \sim N(\mu_n^{(1)} + \beta^{1|2}(\mu^{(2)} - \mu_n^{(2)}), \Lambda^{1|2}), \quad (3.39)$$

kde regresní koeficient $\beta^{1|2}$ a podmíněná kovarianční matice $\Lambda^{1|2}$ jsou definovány jako

$$\beta^{1|2} = \Lambda_n^{12} \left(\Lambda_n^{22} \right)^{-1} \quad (3.40)$$

$$\Lambda^{1|2} = \Lambda_n^{(11)} - \Lambda_n^{(12)} \left(\Lambda_n^{(22)} \right)^{-1} \Lambda_n^{(21)}. \quad (3.41)$$

Apsteriorní prediktivní rozdělení pro nová data

Uvažujme nové pozorování $\tilde{y} \sim N(\mu, \Sigma)$. Sdružené rozdělení $p(\tilde{y}, \mu|y) = N(\tilde{y}|\mu_n, \Lambda_n)$ je exponenciála kvadratické formy v (\tilde{y}, μ) . Proto (\tilde{y}, μ) sleduje sdružené apsteriorní normální rozdělení, což znamená, že apsteriorní marginální rozdělení pro \tilde{y} sleduje vícerozměrné rozdělení. Připomeňme, že předpokládáme znalost kovarianční matice Σ . Stejně jako v případě jednorozměrného rozdělení lze apsteriorní střední hodnotu a rozptyl \tilde{y} určit pomocí (2.7) a (2.8), tj.

$$E(\tilde{y}|y) = E(E(\tilde{y}|\mu, y)|y) \quad (3.42)$$

a

$$\begin{aligned} \text{var}(\tilde{y}|y) &= E(\text{var}(\tilde{y}|\mu, y)|y) + \text{var}(E(\tilde{y}|\mu, y)|y) \\ &= E(\Sigma|y) + \text{var}(\mu|y) = \Sigma + \Lambda_n. \end{aligned} \quad (3.43)$$

Neinformativní apriorní hustota pravděpodobnosti parametru μ

Neinformativní uniformní apriorní hustota pravděpodobnosti parametru μ je $p(\mu) \propto \text{konstanta}$, což lze odvodit v limitní případě, kdy apriorní přesnost konverguje k nule, tj. $|\Lambda_0^{-1}| \rightarrow 0$. V případě nekonečného apriorního rozptylu se totiž stává apriorní střední hodnota irelevantní. Aposteriorní hustota pravděpodobnosti je pak proporcionální věrohodnostní funkci (3.33). Ta je vlastním aposteriorním rozdělením pouze pokud $n \geq d$, tj. pokud je velikost výběru větší nebo rovna dimenzi vícerozměrného normálního rozdělení. V opačném případě pak matice S_0 není pozitivně semidefinitní. Jestliže $n \geq d$, je aposteriorní rozdělení parametru μ , za předpokladu uniformní apriorní hustoty pravděpodobnosti, dáno $\mu|\Sigma, y \sim N(\bar{y}, \Sigma/n)$.

3.6 Vícerozměrný normální model s neznámou střední hodnotou a rozptylem

3.6.1 Konjugátní inverzní Wishartova rodina apriorních rozdělení

Připomeňme si, že konjugátní rozdělení pro jednorozměrný normální model s neznámou střední hodnotou a rozptylem má charakter $N\text{-Inv-}\chi^2$ rozdělení. Apriorní rozdělení matice Σ lze popsat pomocí inverzního Wishartova rozdělení, které je vícerozměrným zobecněním škálovaného inverzního χ^2 rozdělení. Konjugátní apriorní rozdělení pro μ, Σ je normálním inverzním Wishartovým rozdělením, které lze snadno parametrizovat skrze parametry $(\mu_0, \Lambda_0/\kappa_0; \nu_0, \Lambda_0)$

$$\Sigma \sim \text{Inv-Wishart}_{\nu_0}(\Lambda_0^{-1}) \quad (3.44)$$

$$\mu|\Sigma \sim N(\mu_0, \Sigma/\kappa), \quad (3.45)$$

což odpovídá sdružené apriorní hustotě pravděpodobnosti

$$p(\mu, \Sigma) \propto |\Sigma|^{-\left(\frac{\nu_0+d}{2}+1\right)} e^{-\frac{1}{2}\text{tr}(\Lambda_0\Sigma)^{-1} - \frac{\kappa_0}{2}(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)}. \quad (3.46)$$

Parametry ν_0 a Λ_0 představují počet stupňů volnosti a škálovací matici inverzního Wishartova rozdělení pro Σ . Zbývající parametry μ_0 a κ_0 pak představují apriorní střední hodnotu a počet apriorních pozorování na Σ škále. Vynásobením apriorní hustoty pravděpodobnosti normální věrohodnostní funkcí získáváme aposteriorní hustotu pravděpodobnosti ze stejné rodiny pravděpodobnostních rozdělení definovanou parametry

$$\mu_n = \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y} \quad (3.47)$$

$$\kappa_n = \kappa_0 + n \quad (3.48)$$

$$\nu_n = \nu_0 + n \quad (3.49)$$

$$\Lambda_n = \Lambda_0 + S + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)(\bar{y} - \mu_0)^T, \quad (3.50)$$

kde

$$S = \sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})^T. \quad (3.51)$$

Ostatní výsledky jednorozměrného normálního modelu lze snadno zobecnit pro vícerozměrný model. Aposteriorní marginální rozdělení parametru μ sleduje vícerozměrné $t_{\nu_n-d+1}\left(\nu_n, \frac{\Lambda_n}{\kappa_n(\nu_n-d+1)}\right)$ rozdělení. Aposteriorní prediktivní rozdělení pro nová pozorování \tilde{y} je taktéž vícerozměrné t rozdělení, které má navíc $\kappa_n + 1$ ve jmenovateli škálovací matice. Náhodný výběr ze sdruženého aposteriorního rozdělení pro (μ, Σ) lze snadno získat následujícím postupem. Nejprve náhodně vybereme $\Sigma|y \sim \text{Inv-Wishart}_{\nu_n}(\Lambda_n^{-1})$ a následně vybereme $\mu|\Sigma, y \sim N(\mu_n, \Sigma/\kappa_n)$. Pro náhodný výběr z aposteriorního prediktivního rozdělení pro nové pozorování použijeme $\tilde{y}|\mu, \Sigma, y \sim N(\mu, \Sigma)$ za předpokladu, že jsme před tím již náhodně vybrali μ a Σ .

3.6.2 Další neinformativní apriorní rozdělení

Inverzní Wishartovo rozdělení s $d+1$ stupni volnosti

Pokud zvolíme $\Sigma \sim \text{Inv-Wishart}_{d+1}(I)$, sleduje marginálně každá z korelací v matici Σ apriorní uniformní rozdělení.³

Inverzní Wishartovo rozdělení s $d-1$ stupni volnosti

Další možné neinformativní apriorní rozdělení je vícerozměrné Jeffreyovo rozdělení

$$p(\mu, \Sigma) \propto |\Sigma|^{-(d+1)/2}, \quad (3.52)$$

což je pro $\kappa_0 \rightarrow 0, \nu_0 \rightarrow -1$ a $|\Lambda_0| \rightarrow 0$ limita konjugátního apriorního rozdělení. Odpovídající aposteriorní rozdělení má pak podobu

$$\Sigma|y \sim \text{Inv-Wishart}_{n-1}(S^{-1}) \quad (3.53)$$

$$\mu|\Sigma, y \sim N(\bar{y}, \Sigma/n). \quad (3.54)$$

Marginální rozdělení parametru μ a aposteriorní prediktivní rozdělení \tilde{y} jsou za předpokladu vlastního aposteriorního rozdělení dány předchozím odstavcem. Např. aposteriorní marginální rozdělení parametru μ sleduje vícerozměrné $t_{n-d}\left(\frac{S}{n(n-d)}\right)$ rozdělení.

Škálovaný inverzní Wishartův model

Při modelování kovarianční matice můžeme rozšířit inverzní Wishartův model o sadu škálovacích parametrů, které mohou být modelovány odděleně. Tímto způsobem můžeme použít uniformní nebo slabé apriorní rozdělení na modelování korelací bez toho, aniž bychom definovali komplikovaná omezení pro parametry rozptylu. Tento tzv. škálovaný inverzní Wishartův model má tvar

$$\Sigma = \text{Diag}(\xi)\Sigma_\eta\text{Diag}(\xi), \quad (3.55)$$

kde matice Σ_η je dána inverzním Wishartovým apriorním rozdělením⁴ a ξ jsou škálovací parametry.

³Vzhledem k tomu, že korelační matice musí být pozitivně semidefinitní, není sdružené rozdělení těchto korelací uniformní.

⁴Jednou z možností je $\text{Inv-Wishart}_{d+1}(I)$, což implikuje marginální uniformní rozdělení korelací v matici.

3.7 Ilustrativní příklad - biotest

3.7.1 Popis problematiky

Nově vyvíjené léky se často testují na zvířatech. Reakce zvířat má pak dichotomický charakter - živé / mrtvé zvíře, přítomnost / nepřítomnost nádoru. Biotest tohoto typu tak vede k datům

$$(x_i, n_i, y_i); \quad i = 1, \dots, k, \quad (3.56)$$

kde x_i představuje i -tou z k úrovní dávek daného léku aplikovaného n_i zvířatům, z nichž y_i reagovalo pozitivně. Příklad takového experimentu představuje níže uvedená tabulka, kdy bylo testováno dvacet zvířat po pěti pro každou ze čtyř úrovní dávkování léku. Velikost dávky se exponenciálně zvyšuje, a proto je bylo zvoleno logaritmické měřítko.

Dose, x_i (log g/ml)	Number of animals, n_i	Number of deaths, y_i
-0.86	5	0
-0.30	5	1
-0.05	5	3
0.73	5	5

Obrázek 3.2: Výsledky biotestu

3.7.2 Modelování vztahu mezi dávkováním a výsledky testů

Předpokládáme, že zvířata v každé dávkovací skupině jsou vzájemně zaměnitelná. Tento předpoklad pak implikuje, že počet pozitivních reakcí y_i sleduje binomické rozdělení, tj.

$$y_i | \theta_i \sim \text{Bin}(n_i, \theta_i), \quad (3.57)$$

kde θ_i představuje pravděpodobnost úmrtí zvířete při dávce x_i . Dále je rozumné předpokládat, že výsledky čtyř testovacích skupin jsou vzájemně nezávislé, tj. že výsledky v rámci jedné skupiny jsou pro daná $\theta_1, \dots, \theta_4$ nezávislé na výsledcích jiné skupiny.

V rámci jednoduché analýzy můžeme parametry θ_i považovat za vzájemně změnitelné v rámci jejich apriorních rozdělení a to např. použitím neinformativní hustoty pravděpodobnosti $p(\theta_1, \dots, \theta_4) \propto 1$, což vede k tomu, že jednotlivé parametry θ_i sledují vzájemně nezávislé aposteriorní beta rozdělení. Nicméně předpoklad zaměnitelnosti má jeden zásadní nedostatek - pro každou skupinu známe míru dávkování x_i , a proto lze racionálně očekávat, že pravděpodobnost úmrtí / přežití se systematicky mění právě s výší dávky léku.

V nejjednodušším modelu vztahu mezi dávkou a výsledky testů můžeme předpokládat lineární vztah $\theta_i = \alpha + \beta x_i$. Tento lineární model však negarantuje, že θ_i , které lze interpretovat jako pravděpodobnost, bude pro všechny úrovně

dávků léku omezené na interval $(0, 1)$. Řešením tohoto problému je tak použít model

$$\text{logit}(\theta_i) = \alpha + \beta x_i, \quad (3.58)$$

kde $\text{logit}(\theta_i) = \ln \left(\frac{\theta_i}{1-\theta_i} \right)$. Tento model nazýváme logistickým regresním modelem.

3.7.3 Věrohodnostní funkce

Věrohodnostní funkce má pro každou skupinu i podobu

$$p(y_i | \alpha, \beta, n_i, x_i) \propto [\text{logit}^{-1}(\alpha + \beta x_i)]^{y_i} [1 - \text{logit}^{-1}(\alpha + \beta x_i)]^{n_i - y_i}. \quad (3.59)$$

Model je charakterizován parametry α a β jejichž sdružené aposteriorní rozdělení má tvar

$$\begin{aligned} p(\alpha, \beta | y, n, x) &\propto p(\alpha, \beta | n, x) p(y | \alpha, \beta, n, x) \\ &\propto p(\alpha, \beta) \prod_{i=1}^k p(y_i | \alpha, \beta, n_i, x_i). \end{aligned} \quad (3.60)$$

3.7.4 Apriorní rozdělení

Uvažujeme apriorní rozdělení pro (α, β) , která je pro tyto dva parametry nezávislá a lokálně uniformní, tj. $p(\alpha, \beta) \propto 1$. V praxi může použít uniformní apriorní rozdělení, pokud máme omezenou apriorní znalost parametrů α a β . Pokud je analýza založená na neinformativním apriorním rozdělení nedostatečně přesná, můžeme zohlednit další zdroje informací (např. výsledky podobných biotestů) a zkonstruovat informativní apriorní rozdělení.

3.7.5 Hrubý odhad parametrů

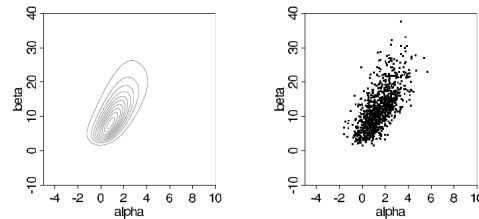
Aposterioerní rozdělení (3.60) můžeme napočítat na mřížce bodů (α, β) . Nicméně před tímto krokem je vhodné získat alespoň hrubou představu o možných hodnotách parametrů (α, β) . S pomocí logistické regrese je možné nalézt odhad (α, β) v (3.60) pro čtyři body v tabulce (3.1). Takto získaný odhad je $(\hat{\alpha}, \hat{\beta}) = (0.8, 7.7)$ s odpovídajícími směrodatnými odchylkami 1.0 a 4.9.

3.7.6 Graf sdružené aposteriorní hustoty pravděpodobnosti

V tomto kroku můžeme vypočítat aposteriorní hustotu pravděpodobnosti a mřížce bodů (α, β) . Pro tento účel použijeme rozmezí $(\alpha, \beta) \in [-5, 10] \times [-10, 40]$, které zahrnuje téměř veškerou pravděpodobnostní hmotu (probability mass) aposteriorního rozdělení.

3.7.7 Náhodný výběr ze sdruženého aposteriorního rozdělení

Po té, co jsme vypočetli nenormalizovanou aposteriorní hustotu pravděpodobnosti na mřížce hodnot, které pokrývají efektivní rozpětí parametrů (α, β) ,



Obrázek 3.3: Konturový a bodový graf posteriorní hustoty pravděpodobnosti parametrů ilustrativního biotestu

můžeme toto rozdělení normalizovat, čímž docílíme toho, že součet pravděpodobností bude roven jedné. Následně náhodně vybereme 1,000 hodnot (α^s, β^s) z posteriorního rozdělení podle následující procedury.

1. Vypočteme posteriorní marginální rozdělení parametru α součtem diskrétního rozdělení z obrázku (3.3a) přes β .
2. Pro $s = 1, \dots, 1000$
 - (a) Náhodně vybereme α^s z diskrétního rozdělení $p(\alpha|y)$ vypočtené v předchozím kroku.
 - (b) Náhodně vybereme β^s z diskrétního podmíněného rozdělení $p(\beta|\alpha, y)$ pro α náhodně vybrané v předchozím kroku.
 - (c) Pro každé náhodně vybrané α a β přidáme náhodný uniformně rozdělený člen centrovaný na bodě nula, jehož rozsah odpovídá intervallům bodů v mřížce.

Bodový graf náhodně vybraného tisíce bodů (α^s, β^s) je k dispozici na obrázku (3.3b)

3.7.8 Aposteriorní rozdělení LD50

V rámci biotestů je důležitým parametrem tzv. LD50 - výše dávky, pro kterou je pravděpodobnost úmrtí rovna 50%. V našem logistickém modelu je tato veličina definována jako

$$LD50 : E\left(\frac{y_i}{n_i}\right) = \text{logit}^{-1}(\alpha + \beta x_i) = 0.5, \quad (3.61)$$

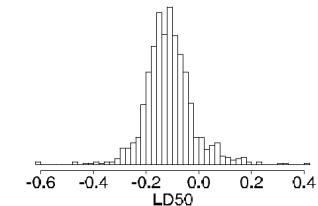
proto $\alpha + \beta x_i = \text{logit}(0.5) = 0$ a LD50 je tak rovno $x_i = -\alpha/\beta$. Výpočet posteriorního rozdělení libovolné popisné veličiny je v Bayesiánské kontextu jednoduchý - stačí pouze vypočíst $-\alpha/\beta$ pro 1,000 náhodně vybraných parametrů, které jsou znázorněny na obrázku (3.3b).

Odhad LD50 v případě, že lék má pozitivní účinek

Koncept LD50 postrádá smysl, pokud $\beta \leq 0$, protože zvyšování dávky nemá za následek nárůst pravděpodobnosti úmrtí. Pokud jsme si jisti, že určitý lék

nemůže způsobit pokles míry nádoru, je možné omezit obor parametrů tak, abychom vyloučili hodnoty β menší než nula. Nicméně je asi vhodnější povolit hodnoty $\beta \leq 0$ a konstatovat, že pro tyto případy je interpretace LD50 komplikovaná.

V případě LD50 se obvykle vykazují dva výsledky - (a) posteriorní pravděpodobnost, že $\beta > 0$, tj. že lék je škodlivý a (b) posteriorní rozdělení LD50 podmíněné předpokladem, že $\beta > 0$. V našem případě všech 1,000 náhodně vybraných (α, β) splňuje podmínku $\beta > 0$, a proto lze u posteriorní pravděpodobnosti, že $\beta > 0$, předpokládat, že přesahuje 99.90%. Histogram $LD50 = -\alpha/\beta$ pro 1,000 náhodných výběrů (α, β) je pak k dispozici na obrázku (3.4). Tento příklad ilustruje, že marginální posteriorní střední hodnota nemusí být dobrým popisem daného parametru. Obecně totiž nemáme zájem na kvantifikaci posteriorní střední hodnoty LD50, protože tato střední hodnota zahrnuje i případy, kdy je vztah mezi dávkou a výsledky testů negativní.



Obrázek 3.4: Histogram LD50 zkonstruovaný z 1,000 náhodných výběrů (α, β)

Kapitola 4

Slovník

anglicky	česky
time series	časová řada
cross-sectional data	průřezová data
pooled cross section data	sdužená průřezová data
panel data	panelová data
regression model	regresní model
explanatory variable	vysvětlující veličina
independent variable	nezávislá veličina
dependent variable	závislá veličina
error term / disturbance	chybová složka / chyba
slope parameter	parametr sklonu
intercept parameter	parametr průsečíku
population regression function (PRF)	populační regresní funkce
ordinary least square (OLS) estimates	odhad nejmenších čtverců
OLS regression line	OLS regresní přímka
sample regression function (SRF)	výběrová regresní funkce
residual	reziduum
sum of squared residuals	součet čtverců reziduí
total sum of squares (SST)	celkový součet čtverců
explained sum of squares (SSE)	vysvětlený součet čtverců
residual sum of squares (SSR)	reziduální součet čtverců
goodnes of fit	míra shody
unbiasedness of OLS	nestrannost OLS odhadů
standard error of regression (SER)	standardní směrodatná odchylka regrese
omitted variable bias	zkreslení z titulu opomenutí
	relevantní vysvětlující proměnné
variance inflation factor (VIF)	inflační faktor rozptylu
degrees of freedom	stupně volnosti
standard deviation	směrodatná odchylka
standard error	směrodatná chyba
best linear unbiased estimator (BLUE)	nejlepší nezkreslený lineární odhad
sampling distribution	výběrové rozdělení
classical liner model (CLM) assumptions	předpoklady klasického lineárního modelu
one-sided test	jednostranný test
two-sided test	dvoustranný test
significance level	hladina významnosti
p-value	p-hodnota
confidence interval	interval spolehlivosti
multiple (joint) hypothesis	vícenásobná (sdužená) hypotéza
unrestricted model	neomezený model
restricted model	omezený model
adjusted R^2	korigované R^2
resampling method	metoda opakovaného výběru
binary (dummy) variable	binární veličina
dummy variable trap	past binární veličiny
linear probability model (LPM)	lineární pravděpodobnostní model
generalized least squares (GLS) estimators	obecné odhady metodou nejmenších čtverců
weighted least squares (WLS) estimators	odhady metodou vážených nejmenších čtverců

Tabulka 4.1: Anglicko-český slovník vybraných statistických pojmů

anglicky	česky
finite distributed lag (FDL) model	model s konečným rozdělením zpoždění
impact propensity	propensitní účinek
long-run propensity	dlohodobá propensita
contemporaneous exogenous	souběžně exogenní
serial correlation aka autocorrelation	sériová korelace aka autokorelace
spurious correlation	zdánlivá korelace
stationary time series	stacionární časové řada
weakly dependent time series	slabě závislá časová řada
covariance stationary proces	kovariančně stacionární proces
independent identically distributed (iid)	nezávislé stejnoměrně rozdělené (náhodné veličiny)
moving average process of order one [MA(1)]	proces klouzavého průměru řádu jedna
autoregressive process of order one [AR(1)]	autoregresivní proces řádu jedna
trend-stationary process	trendově stacionární proces
law of iterated expectations	zákon iterované střední hodnoty
unit root process	proces s jednotkovým kořenem
integrated process of order zero	integrovaný proces řádu nula
sequentially exogenous variables	sekvenčně exogenní veličiny
strict exogeneity	striktní exogenita
contemporaneous exogeneity	souběžná exogenita
quasi-differenced data	quasi diferencovaná data
feasible GLS (FGLS) estimators	dosazitelné GLS funkce odhadu
autoregressive conditional heteroskedasticity (ARCH) model	model autoregresivní podmíněné heteroskedasticity
maximum likelihood method	metoda maximální věrohodnosti
regression specification error test (RESET)	test chybné specifikace chybového členu regresního modelu
nonested models	nevnošené modely
random coefficient / random slope model	model s náhodným sklonem
average partial effect (APE)	průměrný parciální efekt
exogenous sample selection	exogenní výběr vzorku
endogenous sample selection	endogenní výběr vzorku
stratified sampling	stratifikovaný výběr
studentized residuals	studentizovaná rezidua
least absolute deviation (LAD) estimation	odhad metodou nejmenší absolutní směrodatné odchylky
limited dependent variable (LDV)	omezená závislá veličina
corner solution response	hraniční řešení
maximum likelihood method	metoda maximální věrohodnosti
maximum likelihood estimation (MLE)	odhad na základě maximální věrohodnosti
log-likelihood function	logaritmická funkce věrohodnost
log-likelihood	logaritmická věrohodnost
likelihood ratio (LR)	věrohodnostní poměr
percent correctly predicted measure	míra správně predikovaného procenta
partial effect at the average (PEA)	parciální efekt na průměr
average partial effect (APE)	průměrný parciální efekt
quasi-maximum likelihood estimation (QMLE)	odhad metodou kvazimaximální věrohodnosti
censored regression model	cenzurovaný regresní model
truncated regression model	zkrácený regresní model
random draw	náhodné pozorování
truncated sample	omezený výběr
instrumental variable	pomocná veličina
key identification assumption	klíčový předpoklad identifikace
two stage OLS method	dvoufázová OLS metoda
two stage least squares (2SLS) estimator	odhad dvoufázovou metodou nejmenších čtverců
classical errors-in-variables (CEV) assumptions	klasické předpoklady chyby měření
sampling error	chyba výběru / výběrová chyba
overidentification restrictions	omezení nadbytečné identifikace