

Úvod do teorie statistiky, 3.vydání

Alexander M. Mood, Franklin A. Graybill, Duane C. Boes

1973

Obsah

1	Pravděpodobnost	5
1.1	Druhy pravděpodobnosti	5
1.1.1	Úvod	5
1.1.2	Klasická neboli apriorní pravděpodobnost	5
1.1.3	Posteriorní pravděpodobnost	6
1.2	Axiomy pravděpodobnosti	6
1.2.1	Pravděpodobnostní modely	6
1.2.2	Teorie množin	7
1.2.3	Definice výběrového prostoru a náhodného jevu	10
1.2.4	Definice pravděpodobnosti	11
1.2.5	Konečný výběrový prostor	14
1.2.6	Podmíněná pravděpodobnost a nezávislost	15
2	Náhodné veličiny, distribuční a momentová funkce	19
2.1	Náhodná veličina a kumulativní distribuční funkce	19
2.2	Pravděpodobnostní funkce	20
2.2.1	Nespojitá náhodná veličina	20
2.2.2	Spojité náhodné veličiny	22
2.2.3	Ostatní náhodné veličiny	23
2.3	Střední hodnoty a momenty	24
2.3.1	Střední hodnota	24
2.3.2	Rozptyl	25
2.3.3	Střední hodnota funkce náhodné veličiny	26
2.3.4	Chebysheva nerovnost	27
2.3.5	Jensenova nerovnost	28
2.3.6	Momenty a momentová funkce	28
3	Vybraná pravděpodobnostní rozdělení	33
3.1	Nespojitá pravděpodobnostní rozdělení	33
3.1.1	Nespojitá uniformní rozdělení	33
3.1.2	Bernoulliho a binomické rozdělení	34
3.1.3	Hypergeometrické rozdělení	37
3.1.4	Poissonovo rozdělení	38
3.1.5	Geometrické a negativní binomické rozdělení	42
3.1.6	Ostatní nespojitá rozdělení	44
3.2	Spojité pravděpodobnostní rozdělení	45
3.2.1	Spojité uniformní rozdělení	45
3.2.2	Normální rozdělení	46

3.2.3	Exponenciální a gamma rozdělení	49
3.2.4	Beta rozdělení	51
3.3	Ostatní spojitá rozdělení	53
3.3.1	Cauchyho rozdělení	53
3.3.2	Lognormální rozdělení	53
3.3.3	Laplaceova distribuce	53
3.3.4	Weibullovo rozdělení	54
3.3.5	Logistické rozdělení	54
3.3.6	Paretovo rozdělení	54
3.4	Poznámky	55
3.4.1	Aproximace	55
3.4.2	Vztah mezi Poissonovým a exponenciálním rozdělením . .	56
3.4.3	Smíšená pravděpodobnostní rozdělení	56
3.4.4	Zredukována pravděpodobnostní rozdělení	57
4	Vícero náhodných proměnných	59
4.1	Sdružená pravděpodobnost	59
4.1.1	Sdružená kumulativní distribuční funkce	59
4.1.2	Sdružená pravděpodobnostní funkce nespojité náhodné veličiny	60
4.1.3	Sdružená pravděpodobnostní funkce spojité náhodné veličiny	62
4.2	Podmíněné rozdělení a nezávislost	64
4.2.1	Podmíněná kumulativní distribuční funkce nespojité náhodné veličiny	64
4.2.2	Podmíněná pravděpodobnostní funkce spojitých náhodných veličin	65
4.2.3	Více o podmíněné pravděpodobnostní funkci	67
4.2.4	Nezávislost	68
4.3	Střední hodnota, kovariance a korelace	69
4.3.1	Střední hodnota	69
4.3.2	Kovariance a korelace	70
4.3.3	Podmíněná střední hodnota a rozptyl	71
4.3.4	Sdružená momentová funkce a momenty	72
4.3.5	Nezávislost, střední hodnota a rozptyl	73
4.4	Cauchy-Schwarzova nerovnost	74
4.5	Dvourozměrné normální rozdělení	74
4.5.1	Pravděpodobnostní funkce	74
4.5.2	Momentová funkce a momenty	76
4.6	Marginální a podmíněná pravděpodobnostní funkce	77
5	Distribuce funkcí náhodných veličin	79
5.1	Střední hodnota a rozptyl funkcí náhodných veličin	79
5.1.1	Dvojí pojetí	79
5.1.2	Součet náhodných veličin	80

6	Výběry a jejich rozdělení	81
6.1	Výběr a populace	81
6.1.1	Definice základních pojmů	81
6.1.2	Rozdělení výběru	82
6.1.3	Statistiky a výběrové momenty	82
6.2	Výběrová střední hodnota	85
6.2.1	Střední hodnota a rozptyl	85
6.2.2	Zákon velkých čísel	85
6.2.3	Centrální limitní věta	86
6.2.4	Bernoulliho rozdělení	88
6.2.5	Poissonovo rozdělení	88
6.2.6	Exponenciální rozdělení	88
6.2.7	Uniformní rozdělení	89
6.2.8	Cauchyho rozdělení	89
6.3	Výběr z normálního rozdělení	90
6.3.1	Střední hodnota výběru	90
6.3.2	Chi-kvadrát rozdělení	90
6.3.3	F rozdělení	93
6.3.4	Studentovo rozdělení	95
6.4	Pořadové statistiky	96
6.4.1	Rozdělení funkcí pořadových statistik	98
6.4.2	Asymptotické rozdělení	99
6.4.3	Výběrová kumulativní distribuční funkce	103
7	Bodový odhad	105
7.1	Metody bodového odhadu	105
7.1.1	Momentová metoda	105
7.1.2	Metoda maximální věrohodnosti	107
7.1.3	Ostatní metody	111
7.2	Vlastnosti funkcí odhadu	113
7.2.1	Blízkost funkce odhadu	113
7.2.2	Střední chyba čtverců	114
7.2.3	Konzistentnost a BAN	116
7.2.4	Ztrátová a riziková funkce	118
7.3	Dostatečnost	120
7.3.1	Dostatečná statistika	120
7.3.2	Kritérium rozkladu	124
7.3.3	Minimální dostatečná statistika	126
7.3.4	Rodina exponenciálních rozdělení	127
7.4	Nezkreslený odhad	128
7.4.1	Dolní mez rozptylu	128
7.4.2	Dostatečnost a kompletnost	133
7.5	Invariance	139
7.5.1	Lokační invariance	139
7.5.2	Škálovací invariance	141
7.6	Bayesovské funkce odhadu	142
7.6.1	Posteriorní rozdělení	142
7.6.2	Ztrátová funkce	145
7.6.3	Minimax funkce odhadu	149
7.7	Vektor parametrů	150

7.8	Odhad maximální věrohodnosti	155
8	Intervalový odhad	159
8.1	Interval spolehlivosti	159
8.1.1	Definice intervalu spolehlivosti	159
8.1.2	Centrální veličina	160
8.2	Náhodný výběr z normálního rozdělení	161
8.2.1	Interval spolehlivosti pro střední hodnotu	161
8.2.2	Interval spolehlivosti pro rozptyl	162
8.2.3	Region spolehlivosti pro střední hodnotu a rozptyl	163
8.2.4	Interval spolehlivosti pro rozdíl středních hodnot	164
8.3	Metody stanovení intervalu spolehlivosti	165
8.3.1	Metoda centrální veličiny	165
8.3.2	Statistická metoda	167
8.4	Interval spolehlivosti pro výběry velkého rozsahu	170
8.5	Bayesovský interval	172
9	Testování hypotéz	173
9.1	Základní pojmy	173
9.2	Testování jednoduchých hypotéz	177
9.2.1	Nejsilnější test	177
9.2.2	Ztrátová funkce	180
9.3	Testování složených hypotéz	184
9.3.1	Obecný věrohodnostní poměrový test	184
9.3.2	Uniformě nejsilnější test	185
9.3.3	Nezkreslený test	188
9.3.4	Metody pro nalézání testů	188
9.4	Testování hypotéz - výběr z normálního rozdělení	190
9.4.1	Testování středních hodnot	190
9.4.2	Testování rozptylu	193
9.4.3	Test několika středních hodnot	194
9.4.4	Testování několika rozptylů	197
9.5	Chi-kvadrát testy	199
9.5.1	Asymptotické rozdělení obecného věrohodnostního poměru	199
9.5.2	Chi-kvadrát test dobré shody	200
9.5.3	Test shodnosti dvou multinomiálních rozdělení	204
9.5.4	Test nezávislosti v kontingčních tabulkách	206
9.6	Testování hypotéz a intervaly spolehlivosti	213
9.7	Sekvenční testování hypotéz	215
9.7.1	Úvod	215
9.7.2	Definice sekvenčního poměrového testu	216
9.7.3	Přibližný sekvenční poměrový test	218
9.7.4	Přibližná střední hodnota velikosti náhodného výběru sek- venčního poměrového testu	219
10	Lineární regrese	221
10.1	Příklady lineární regrese	221
10.2	Definice regresního modelu	222
10.3	Bodový odhad - případ A	224
10.4	Intervaly spolehlivosti - případ A	227

10.4.1	Interval spolehlivosti pro σ^2	227
10.4.2	Interval spolehlivosti pro β_0	227
10.4.3	Interval spolehlivosti pro β_1	228
10.4.4	Interval spolehlivosti pro $\mu(x)$	228
10.5	Testování hypotéz - případ A	229
10.5.1	$\mathcal{H}_0 : \beta_1 = 0$ vs. $\mathcal{H}_1 : \beta_1 \neq 0$	230
10.5.2	Zobecnění pro $\mathcal{H}_0 : \beta_0 = b_0$ a $\mathcal{H}_0 : \beta_1 = b_1$	232
10.6	Bodový odhad - případ B	232
11	Neparametrické metody	237
11.1	Kumulativní distribuční funkce	237
11.1.1	Výběrová kumulativní distribuční funkce	237
11.1.2	Kolmogorov-Smirnovův test dobré shody	239
11.1.3	Hranice spolehlivosti pro kumulativní distribuční funkci	240

Kapitola 1

Pravděpodobnost

1.1 Druhy pravděpodobnosti

1.1.1 Úvod

Počátky teorie pravděpodobnosti jsou spojeny s hazardními hrami, jako jsou např. hod v kostky nebo ruleta. Ačkoliv podstatou hazardních her je náhoda, je zřejmé, že z dlouhodobého hlediska je možné výsledek odhadnout. Tato “odhadnutelnost” koneckonců umožňuje kasinu provozovat svoji činnost. S podobnou náhodností se lze setkat také v experimentální vědě, jako jsou medicína nebo fyzika.

1.1.2 Klasická neboli apriorní pravděpodobnost

Uvažujme hrací kostku. Přijetí předpokladů, že (a) při každém hození padne číslo jedna, dva, tři, čtyři, pět nebo šest a že (b) pravděpodobnost realizace jednotlivých výsledků je shodná, vytváří model, který je možné použít k analýze hry v kostky. Modely podobného typu pak tvoří základ tzv. klasické pravděpodobnosti.

Definice 1.1 (Klasická pravděpodobnost) *Uvažujme náhodný experiment, který může vyústit v n vzájemně se vylučujících výsledků se shodnou pravděpodobností realizace. Jestliže n_A z těchto výsledků má atribut A , pak je pravděpodobnost A definována jako $\frac{n_A}{n}$.*

Výše uvedená definice má několik slabých míst. Prvním z nich je situace, kdy se n popř. také n_A blíží nekonečnu¹. Další komplikací jsou případy, kdy jednotlivé výsledky nemají shodnou pravděpodobnost realizace. Vzorec $\frac{n_A}{n}$ také není vhodný, jestliže potřebujeme určit tzv. podmíněnou pravděpodobnost. Z těchto důvodů byla vyvinuta teorie posteriorní pravděpodobnosti.

¹Uvažujme pravděpodobnost, s jakou bude náhodně tažené přirozené číslo sudé. Je zřejmé, že tato pravděpodobnost je 0.5. Odpověď je však, spíše než na výše uvedeném vzorci, založena na intuici.

1.1.3 Posteriorní pravděpodobnost

V případě apriorní pravděpodobnosti jsme si vytvořili mentální model hrací kostky, které jsme přisoudili určité vlastnosti. Při každém hodu mohlo nastat šest vzájemně se vylučujících výsledků, kterým jsme přiřadili shodnou pravděpodobnost, tj. $\frac{1}{6}$.

K problému je však možné přistupovat také z druhé strany. Definujme pravděpodobnost hodu konkrétního čísla jako p_i , kde $i = 1, 2, \dots, 6$. Abychom tyto pravděpodobnosti určili pro jednu konkrétní hrací kostku, provedme velké množství hodů a jejich výsledky zaznamenejme. Frekvenci jednotlivých čísel pak použijeme jako odhad pravděpodobnosti p_i . Vzhledem k tomu, že jsme pravděpodobnost neurčili a priori (tj. pouze na základě logického zhodnocení modelu) ale ex-post (na základě zkušenosti s konkrétní hrací kostkou), hovoříme o tzv. posteriorní pravděpodobnosti. Je zřejmé, že pokud bude daná hrací kostka správně vyvážená a počet kontrolních hodů dostatečně velký, budou posteriorní pravděpodobnosti přibližně odpovídat apriorním pravděpodobnostem.

1.2 Axiomy pravděpodobnosti

1.2.1 Pravděpodobnostní modely

Apriorní a posteriorní pravděpodobnost jsou vztaženy k experimentu, který je možné opakovat v homogenním prostředí. Jako příklad jsme uváděli hod hrací kostkou. V řadě případů však pravděpodobnost není možné podrobit opakovaným experimentům. Příkladem budiž pravděpodobnost, s jakou ve vesmíru existuje inteligentní život mimo Zemi. Vzhledem k tomu, že tyto pravděpodobnosti nelze přesně kvatifikovat, hovoříme o nich jako o subjektivní pravděpodobnosti. Ačkoliv je subjektivní pravděpodobnost legitimní součástí obecné teorie pravděpodobnosti, zaměřuje se tato kniha právě na apriorní a posteriorní pravděpodobnost.

Experiment, jehož výstupem jsou náhodné výsledky, tvoří základ pro tzv. pravděpodobnostní model. V jeho rámci definujeme všechny možné výsledky experimentu, tzv. elementární jevy. Jestliže se vrátíme k našemu příkladu hrací kostky, jsou elementární jevy definované hody 1, 2, 3, 4, 5 a 6². Elementární jevy pak tvoří tzv. výběrový prostor. Množinovými operacemi nad elementárními jevy definujeme tzv. náhodné jevy. Příkladem náhodného jevu v kontextu hrací kostky je např. situace, kdy padne

- sudé číslo,
- číslo větší než dvě,
- číslo dvě³.

Všechny myslitelné náhodné jevy pak tvoří tzv. jevové pole.

²Situace, kdy se hrací kostka zastaví na hraně nebo se v průběhu hodu ztratí, ignorujeme - v našem pravděpodobnostním modelu nemohou nastat, a proto jim přiřazujeme pravděpodobnost nula.

³V tomto případě je náhodný jev tvořený jedním elementárním jevem. V přechodných příkladech byl náhodný jev tvořen množinou elementárních jevů.

1.2.2 Teorie množin

Uvažujme soubor objektů, který je dostatečně velký na to, aby obsahoval každý jednotlivý objekt relevantní pro daný problém. Každý jednotlivý objekt označme jako prvek ω a jejich úplný soubor jako pole Ω .

Skupina prvků definuje množinu. Skutečnost, že prvek ω patří do množiny A , označujeme jako $\omega \in A$.

Definice 1.2 (Podmnožina) *Uvažujme množiny A a B . Jestliže každý prvek množiny A je současně prvkem množiny B , pak o množině A hovoříme jako o podmnožině množiny B . Tuto skutečnost značíme jako $A \subset B$.*

Definice 1.3 (Ekvivalentní množiny) *Uvažujme množiny A a B . Jestliže současně platí $A \subset B$ a $B \subset A$, pak je každý prvek množiny A současně prvkem množiny B a naopak. Množiny A a B jsou tzv. ekvivalentní. Tuto skutečnost značíme jako $A = B$.*

Definice 1.4 (Prázdná množina) *Jestliže množina neobsahuje žádný prvek, nazýváme ji prázdnou množinou a označujeme jako \emptyset .*

Definice 1.5 (Doplněk množiny) *Doplněk množiny A vzhledem k poli Ω , označujeme jako A^c nebo také jako $\Omega - A$. Doplněk představuje množinu všech prvků, které jsou obsaženy v poli Ω a zároveň nejsou obsaženy v množině A .*

Definice 1.6 (Sjednocení množin) *Nechť jsou A a B libovolné podmnožiny pole Ω . Množinu tvořenou všemi prvky obsaženými v množině A nebo v množině B , nazýváme sjednocením množin A a B a značíme jako $C = A \cup B$.*

Definice 1.7 (Průnik množin) *Nechť jsou A a B libovolné podmnožiny pole Ω . Množinu tvořenou všemi prvky současně obsaženými v množině A a množině B , nazýváme průnikem množin A a B a značíme jako $C = A \cap B$.*

Definice 1.8 (Rozdíl množin) *Nechť jsou A a B libovolné podmnožiny pole Ω . Množinu tvořenou všemi prvky, které jsou obsaženy v množině A a zároveň nejsou obsaženy v množině B , nazýváme rozdílem množin A a B a značíme jako $C = A - B$.*

Věta 1.1 (Pravidlo komutace)

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

Věta 1.2 (Pravidlo asociace)

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

Věta 1.3 (Pravidlo distribuce)

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Věta 1.4 *Doplňek k doplňku množiny A je roven množině A .*

$$(A^c)^c = A$$

Věta 1.5

$$A \cap \Omega = A$$

$$A \cup \Omega = \Omega$$

$$A \cap \emptyset = \emptyset$$

$$A \cup \emptyset = A$$

Věta 1.6

$$A \cap A^c = \emptyset$$

$$A \cup A^c = \Omega$$

$$A \cap A = A$$

$$A \cup A = A$$

Věta 1.7 (De Morganovo pravidlo)

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

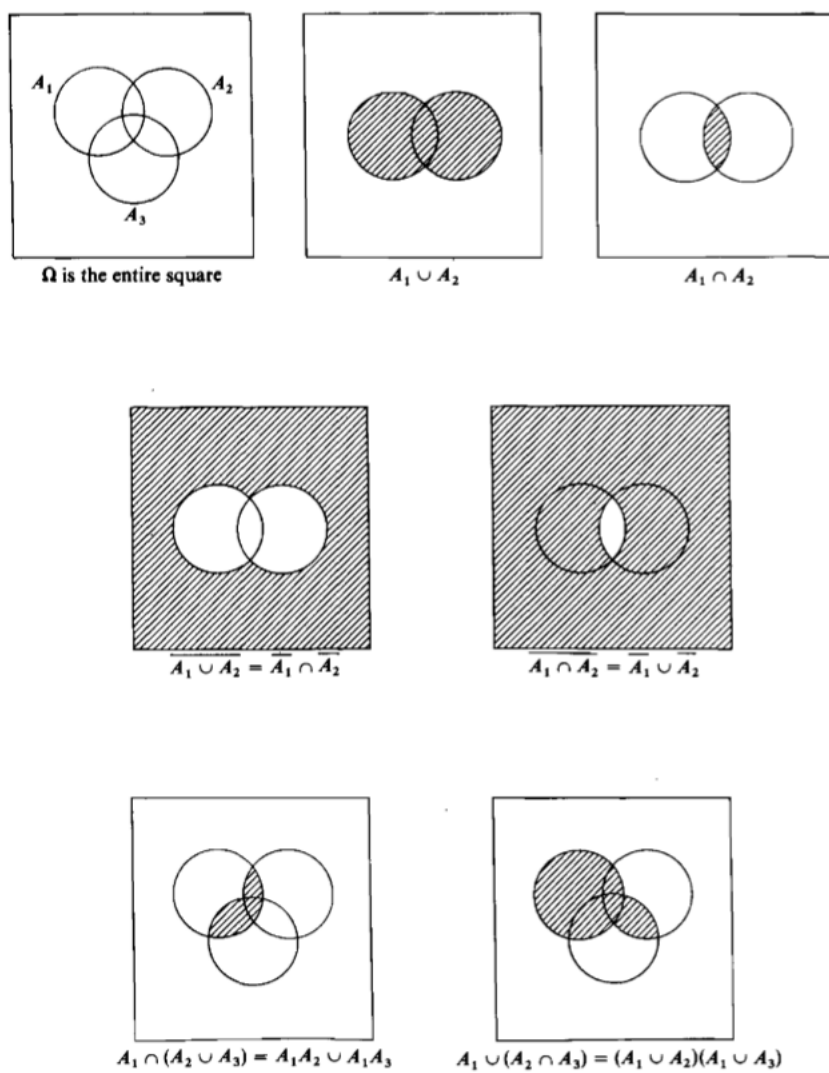
Věta 1.8

$$A - B = A \cap B^c$$

Důkaz 1.1 *Dokažme, že $(A \cup B)^c = A^c \cap B^c$. Dle definice jsou dvě množiny shodné, jestliže je každá z nich obsažena ve té druhé, tj. $A \subset B$ a zároveň $B \subset A$. Jestliže $\omega \in (A \cup B)^c$, pak $\omega \in A^c \cap B^c$. Platí tedy, že $(A \cup B)^c \subset A^c \cap B^c$. Dále $\omega \in A^c \cap B^c$ implikuje $\omega \notin A \cup B$, což implikuje $\omega \notin A$ a $\omega \notin B$, což implikuje $\omega \in A^c \cap B^c$. Dalším krokem je prokázat, že $A^c \cap B^c \subset (A \cup B)^c$. Nechť $\omega \in A^c \cap B^c$, což znamená, že ω patří do A^c i do B^c . Pak $\omega \notin A \cup B$, protože ω by muselo patřit do A nebo do B . To popírá skutečnost, že ω patří současně do A^c a B^c . Nicméně $\omega \notin A \cup B$ znamená $\omega \in (A \cup B)^c$, čímž jsme prokázali požadované.*

Je zvykem rozlišovat jednotlivé množiny v poli Ω přiřazením indexů. Označme množinu takovýchto indexů jako Λ . Jestliže bychom uvažovali pouze dvě podmnožiny pole Ω , pak bude Λ obsahovat pouze dva indexy, řekněme $\Lambda = 1, 2$.

Definice 1.9 (Sjednocení a průnik množin) *Nechť Λ je množinou indexů a $\{A_\lambda : \lambda \in \Lambda\} = \{A_\lambda\}$ je souborem podmnožin pole Ω indexovaných dle Λ . Množinu, která obsahuje všechny prvky náležící do alespoň jednoho A_λ , nazýváme sjednocením množin $\{A_\lambda\}$ a značíme jako $\bigcup_{\lambda \in \Lambda} A_\lambda$. Množinu, která obsahuje všechny prvky současně náležící do všech A_λ , nazýváme průnikem množin $\{A_\lambda\}$ a značíme jako $\bigcap_{\lambda \in \Lambda} A_\lambda$. Jestliže je Λ prázdnou množinou, pak definujeme sjednocení jako $\bigcup_{\lambda \in \Lambda} A_\lambda = \emptyset$ a průnik jako $\bigcap_{\lambda \in \Lambda} A_\lambda = \Omega$.*



Obrázek 1.1: Množinové operace nad náhodnými jevy

Věta 1.9 (De Morganův teorém) *Nechť Λ představuje množinu indexů a $\{A_\lambda\}$ je souborem všech podmnožin pole Ω indexovaných dle Λ . Pak*

$$\left(\bigcup_{\lambda \in \Lambda} A_\lambda\right)^c = \bigcap_{\lambda \in \Lambda} A_\lambda^c$$

$$\left(\bigcap_{\lambda \in \Lambda} A_\lambda\right)^c = \bigcup_{\lambda \in \Lambda} A_\lambda^c$$

Definice 1.10 (Disjunktí množiny) *Podmnožiny A a B pole Ω jsou disjunktí, jestliže $A \cap B = \emptyset$. Podmnožiny A_1, A_2, \dots jsou vzájemně disjunktí, jestliže $A_i \cap A_j = \emptyset$ pro všechna $i \neq j$.*

Věta 1.10 *Jestliže jsou A a B podmnožiny pole Ω , pak*

1. $A = (A \cap B) \cup (A \cap B^c)$
2. $(A \cap B) \cap (A \cap B^c) = \emptyset$

Důkaz 1.2

1. $A = A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c)$
2. $(A \cap B) \cap (A \cap B^c) = (A \cap A) \cap (B \cap B^c) = A \cap \emptyset = \emptyset$

Věta 1.11 *Jestliže $A \subset B$, pak $A \cap B = A$ a $A \cup B = B$.*

1.2.3 Definice výběrového prostoru a náhodného jevu

Definice 1.11 (Výběrový prostor) *Výběrový prostor, označovaný jako Ω , představuje soubor všech možných výstupů náhodného experimentu.*

Definice 1.12 (Náhodný jev a jevové pole) *Náhodný jev je podmnožinou výběrového prostoru. Množina všech náhodných jevů souvisejících s daným experimentem pak představuje jevové pole tohoto experimentu.*

Výše uvedená definice přesně nevymezuje pojem “náhodný jev”. Náhodný jev je vždy podmnožinou výběrového prostoru, avšak pro dostatečně velký výběrový prostor ne všechny jeho podmnožiny musí představovat náhodný jev. Proto soubor všech podmnožin výběrového prostoru nemusí nutně odpovídat jevovému poli. Nicméně množina všech náhodných jevů může být vždy zvolena dostatečně velká na to, aby obsahovala všechny náhodné jevy, jejichž pravděpodobností bychom se mohli chtít zabývat. Jestliže se výběrový prostor skládá z konečného počtu prvků, pak odpovídající jevové pole bude množinou všech podmnožin výběrového prostoru.

Naším primárním zájmem je pravděpodobnost, s jakou určitý náhodný jev nastane popř. nenastane. Náhodný jev A nastane, jestliže výsledek experimentu (tj. prvek odpovídajícího výběrového prostoru) náleží A . Protože prvek ω výběrového prostoru Ω je z definice jeho podmnožinou, je zároveň kandidátem pro náhodný jev. Proto můžeme na ω nahlížet jako na prvek i jako na podmnožinu Ω . Abychom tyto dva možné pohledy odlišili, používáme značení ω , jestliže mluvíme o prvku, a značení $\{\omega\}$, jestliže mluvíme o podmnožině. Také \emptyset a Ω jsou podmnožinami Ω a jsou tak náhodnými jevy. Ω někdy nazýváme jistým jevem.

Pro značení náhodných jevů se zpravidla používají velká písmena latinské abecedy (s výjimkou výše zmíněných \emptyset a Ω), tj. např. A . Jevové pole je pak nejčastěji označeno skriptovým písmenem latinské abecedy jako např. \mathcal{A} .

Jevové pole \mathcal{A} by mělo splňovat

1. $\Omega \in \mathcal{A}$,
2. je-li $A \in \mathcal{A}$, pak $A^c \in \mathcal{A}$ a
3. jestliže $A_1 \in \mathcal{A}$ a $A_2 \in \mathcal{A}$, pak $A_1 \cup A_2 \in \mathcal{A}$.

Jakýkoliv souhrn náhodných jevů, které splňují výše uvedené tři podmínky, nazýváme algebrou náhodných jevů, zkráceně též algebrou. Je zřejmé, že soubor všech podmnožin výběrového prostoru Ω splňuje tyto podmínky. Tyto požadované vlastnosti pak implikují následující vlastnosti jevového pole \mathcal{A} .

Věta 1.12 $\emptyset \in \mathcal{A}$

Důkaz 1.3 *Uvažujme vlastnost (1) $\Omega \in \mathcal{A}$ a související vlastnost (2) $\Omega^c \in \mathcal{A}$. Nicméně $\Omega^c = \emptyset$, a proto $\emptyset \in \mathcal{A}$.*

Věta 1.13 *Jestliže $A_1 \in \mathcal{A}$ a $A_2 \in \mathcal{A}$, pak $A_1 \cap A_2 \in \mathcal{A}$.*

Důkaz 1.4 *Platí $A_1^c \in \mathcal{A}$ a $A_2^c \in \mathcal{A}$, což implikuje $A_1^c \cup A_2^c \in \mathcal{A}$ a $(A_1^c \cup A_2^c)^c \in \mathcal{A}$. Zároveň, dle De Morganova pravidla, platí $(A_1^c \cup A_2^c)^c = A_1 \cap A_2 = A_1 \cap A_2$.*

Věta 1.14 *Jestliže $A_1, A_2, \dots, A_n \in \mathcal{A}$, pak $\bigcup_{i=1}^n A_i \in \mathcal{A}$ a $\bigcap_{i=1}^n A_i \in \mathcal{A}$.*

V následujícím textu budeme vždy předpokládat, že souhrn všech jevů \mathcal{A} je algebra. V praxi lze algebra \mathcal{A} vytvořit tak, že soubor jevů, o které se zajímáme, rozšíříme o

1. jistý jev,
2. všechny komplementární jevy a
3. a všechna konečná sjednocení a průniky těchto jevů.

Až dosud jsme nevysvětlili, proč \mathcal{A} není vždy souborem všech podmnožin výběrového prostoru Ω . Vysvětlení poskytneme v následující kapitole.

1.2.4 Definice pravděpodobnosti

Definice 1.13 (Funkce) *Funkce $f(\cdot)$ s definičním oborem A a oborem hodnot B je souborem uspořádaných párů (a, b) , které splňují*

1. $a \in A$ a $b \in B$,
2. pro každé $a \in A$ existuje uspořádaný pár (a, b) ⁴ a
3. žádné dva uspořádané páry nemají shodný první element⁵.

⁴Obrácené tvrzení však neplatí. Nelze tedy říci, že pro každé $b \in B$ existuje uspořádaný pár (a, b) .

⁵To znamená, že neexistují dvě uspořádané dvojice (a, b_1) a (a, b_2) , kde $b_1 \neq b_2$.

Jestliže $(a, b) \in f(\cdot)$, píšeme $b = f(a)$ a $f(a)$ nazýváme hodnotou funkce $f(\cdot)$ v bodě a . Pro libovolné $a \in A$ platí $f(a) \in B$. Označením $f(\cdot)$ rozumíme uspořádaný pár $(a, b = f(a))$. $f(a)$ někdy také nazýváme obrazem bodu a pro $f(\cdot)$.

Definice 1.14 (Indikační funkce) *Nechť Ω představuje prostor s body ω a A představuje libovolnou podmnožinu Ω . Indikační funkce A , označována jako $I_A(\cdot)$, je funkcí s definičním oborem Ω a oborem hodnot $\{0, 1\}$ a je definována jako*

$$I_A(\omega) = \begin{cases} 1 & \text{pro } \omega \in A \\ 0 & \text{pro } \omega \notin A \end{cases}$$

Nechť Ω představuje libovolný výběrový prostor a \mathcal{A} libovolný soubor podmnožin Ω . Pak

1. $I_A(\omega) = 1 - I_{A^c}(\omega)$ pro všechna $A \in \mathcal{A}$,
2. $I_{A_1 \cap A_2 \cap \dots \cap A_n}(\omega) = I_{A_1}(\omega) \cdot I_{A_2}(\omega) \cdot \dots \cdot I_{A_n}(\omega)$ pro $A_1, A_2, \dots, A_n \in \mathcal{A}$,
3. $I_{A_1 \cup A_2 \cup \dots \cup A_n}(\omega) = \max[I_{A_1}(\omega), I_{A_2}(\omega), \dots, I_{A_n}(\omega)]$ pro $A_1, A_2, \dots, A_n \in \mathcal{A}$ a
4. $I_A^2(\omega) = I_A(\omega)$ pro všechna $A \in \mathcal{A}$.

Indikační funkci lze použít také pro “indikaci” podmnožin reálné osy

$$I_{[0,1]}(x) = \begin{cases} 1 & \text{pro } 0 \leq x < 1 \\ 0 & \text{v ostatních případech} \end{cases}$$

popř. množiny přirozených čísel

$$I_{I^+}(x) = \begin{cases} 1 & \text{jestliže } x \text{ je přirozené číslo} \\ 0 & \text{v ostatních případech} \end{cases}$$

Pro následující text předpokládáme, že Ω označuje výběrový prostor a \mathcal{A} označuje souhrn jevů, které představují algebru pro určitý náhodný experiment.

Definice 1.15 (Pravděpodobnostní funkce) *Pravděpodobnostní funkce $P[\cdot]$ je funkcí s definičním oborem \mathcal{A} a oborem hodnot $[0, 1]$, která splňuje axiomy*

1. $P[A] \geq 0$ pro každé $A \in \mathcal{A}$,
2. $P[\Omega] = 1$ a
3. jsou-li A_1, A_2, \dots vzájemně neslučitelné jevy⁶ v \mathcal{A} a jestliže $A_1 \cup A_2 \cup \dots = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, pak $P[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i]$.

V naší definici náhodného jevu a jevového pole \mathcal{A} jsme zmínili, že \mathcal{A} nemůže být vždy chápáno jako souhrn všech podmnožin Ω . Důvodem je skutečnost, že pro “dostatečně velké” Ω je soubor všech podmnožin v Ω příliš velký a je tak nemožné definovat pravděpodobnostní funkci, která by byla konzistentní s výše uvedenými axiomy. Tyto axiomy implikují následující věty.

⁶Jevy A_i a A_j jsou vzájemně neslučitelné, jestliže nemohou nastat současně, tj. $A_i \cap A_j = \emptyset$ pro $i \neq j$.

Věta 1.15 $P[\emptyset] = 0$

Důkaz 1.5 Uvažujme $A_1 = \emptyset, A_2 = \emptyset, \dots$. Pak dle axiomu (3) platí

$$P[\emptyset] = P[\cup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i] = \sum_{i=1}^{\infty} P[\emptyset]$$

což platí, pouze je-li $P[\emptyset] = 0$.

Věta 1.16 Jestliže jsou A_1, \dots, A_n vzájemně neslučitelné jevy v \mathcal{A} , pak

$$P[A_1 \cup \dots \cup A_n] = \sum_{i=1}^n P[A_i]$$

Důkaz 1.6 Nechť $A_{n+1} = \emptyset, A_{n+2} = \emptyset, \dots$. Pak $\cup_{i=1}^n A_i = \cup_{i=1}^{\infty} A_i \in \mathcal{A}$ a

$$P[\cup_{i=1}^n A_i] = P[\cup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i] = \sum_{i=1}^n P[A_i]$$

Věta 1.17 Jestliže je A jevem z jevového pole \mathcal{A} , pak

$$P[A^c] = 1 - P[A]$$

Důkaz 1.7 $A \cup A^c = \Omega$ a $A \cap A^c = \emptyset$. Proto

$$P[\Omega] = P[A \cup A^c] = P[A] + P[A^c]$$

Zároveň však platí $P[\Omega] = 1$, čímž jsme dokázali požadované.

Věta 1.18 Jestliže $A \in \mathcal{A}$ a $B \in \mathcal{A}$, pak

$$P[A] = P[A \cap B] + P[A \cap B^c]$$

a

$$P[A - B] = P[A \cap B^c] = P[A] - P[A \cap B]$$

Důkaz 1.8 $A = A \cap B \cup A \cap B^c$ a $(A \cap B) \cup (A \cap B^c) = A$. Proto $P[A] = P[A \cap B] + P[A \cap B^c]$.

Věta 1.19 Pro každé dva jevy $A \in \mathcal{A}$ a $B \in \mathcal{A}$ platí $P[A \cap B] = P[A] + P[B] - P[A \cup B]$. Tuto větu lze zobecnit do tvaru

$$\begin{aligned} P[A_1 \cup A_2 \cup \dots \cup A_n] &= \sum_{j=1}^n P[A_j] - \sum_{i < j} P[A_i \cap A_j] + \\ &+ \sum_{i < j < k} P[A_i \cap A_j \cap A_k] - \dots + (-1)^{n+1} P[A_1 \cap A_2 \cap \dots \cap A_n] \end{aligned}$$

Důkaz 1.9 Platí $A \cup B = A \cup (A^c \cap B)$ a $A \cap (A^c \cap B) = \emptyset$. Proto

$$P[A \cup B] = P[A] + P[A^c \cap B] = P[A] + P[B] - P[A \cap B]$$

Zobecněnou verzi věty lze dokázat pomocí indukce.

Věta 1.20 Jestliže $A \in \mathcal{A}$ a $B \in \mathcal{A}$ a $A \subset B$, pak $P[A] \leq P[B]$.

Důkaz 1.10 Platí $B = (B \cap A) \cup (B \cap A^c)$ a $A = B \cap A$. Proto $B = A \cup (B \cap A^c)$ a $A \cap (B \cap A^c) = \emptyset$, což implikuje $P[B] = P[A] + P[B \cap A^c]$. Důkaz završíme zjištěním, že $P[B \cap A^c] \geq 0$.

Věta 1.21 (Booleova nerovnost) Jestliže $A_1, A_2, \dots, A_n \in \mathcal{A}$, pak

$$P[A_1 \cup A_2 \cup \dots \cup A_n] \leq P[A_1] + P[A_2] + \dots + P[A_n]$$

Důkaz 1.11 Platí $P[A_1 \cup A_2] = P[A_1] + P[A_2] - P[A_1 \cap A_2] \leq P[A_1] + P[A_2]$. Důkaz lze dokončit pomocí indukce.

Definice 1.16 (Pravděpodobnostní prostor) Pravděpodobnostní prostor je definován tripletem $(\Omega, \mathcal{A}, P[\cdot])$, kde Ω představuje výběrový prostor, \mathcal{A} jevoové pole a $P[\cdot]$ pravděpodobnostní funkci s definičním oborem \mathcal{A} .

1.2.5 Konečný výběrový prostor

Nechť $\omega_1, \omega_2, \dots, \omega_n$ představují elementární jevy konečného výběrového prostoru Ω . Uvažujme funkci $P[\cdot]$ s definičním oborem tvořeným souborem všech podmnožin prostoru Ω , která splňuje následující podmínky

1. $P[\{\omega_1\}] = P[\{\omega_2\}] = \dots = P[\{\omega_n\}]$
2. Je-li A podmnožinou Ω obsahující $N(A)$ elementárních jevů, pak $P[A] = N(A)/N$.

Lze relativně snadno dokázat, že funkce $P[\cdot]$ splňuje všechny tři axiomy kladené na pravděpodobnostní funkce, a proto je pravděpodobnostní funkcí.

Definice 1.17 (Pravděpodobnostní funkce se shodnou pravděpodobností)

Pravděpodobnostní funkce $P[\cdot]$, která splňuje výše uvedené dvě podmínky, je tzv. pravděpodobnostní funkcí se shodnou pravděpodobností.

Uvažujme experiment, jehož výsledkem je vzorek o velikosti n . Myšlenku lze ilustrovat na příkladu nádoby obsahující M míčků, ze které je náhodně taženo n míčků. Existují dva základní způsoby výběru vzorku a to s vracením (po té, co je míček vybrán, je vrácen zpět do nádoby) a bez vracení (po té, co je míček vybrán, je odložen stranou). Jestliže bude záležet na pořadí, v jakém jsou míčky taženy, pak v případě náhodného výběru s vracením existuje $M^n = \underbrace{M \cdots M}_n$ a

v případě náhodného výběru bez vracení $(M)_n = M(M-1) \cdots (M-n+1)$ n -tic. Jestliže na pořadí, v jakém jsou míčky taženy, nezáleží, existuje v případě výběru s vracením $\frac{M^n}{n!}$ a v případě výběru bez vracení $\binom{M}{n} = \frac{(M)_n}{n!}$ n -tic. n -tice, u kterých záleží na pořadí, nazýváme uspořádané, a n -tice, u kterých na pořadí nezáleží, nazýváme neuspořádané.

Příklad 1.1 Uvažujme množinu S o M prvcích. Celkový počet podmnožin množiny S je roven $\sum_{n=0}^M \binom{M}{n}$. Pro $a = b = 1$ je z binomické věty $(a+b)^M = \sum_{n=0}^M \binom{M}{n} a^n b^{M-n}$ zřejmé, že $2^M = \sum_{n=0}^M \binom{M}{n}$. Množina o M prvcích tak má 2^M podmnožin.

Uvažujme situaci, kdy nádoba obsahuje M míčků, z nichž právě K je bílých. Pravděpodobnost, že při výběru n míčků s vrácením jich bude právě k bílých, je rovna

$$\frac{\binom{n}{k} K^k (M-K)^{n-k}}{M^n}$$

Při výběru bez vrácení je tato pravděpodobnost rovna

$$\frac{\binom{n}{k} (K)_k (M-K)_{n-k}}{(M)_n} = \frac{\binom{K}{k} \binom{M-K}{n-k}}{\binom{M}{n}}$$

Pro úplnost dodejme, že v kontextu výše uvedených dvou rovnic je pořadí, v jakém jsou míčky taženy, irelevantní.

Až dosud jsme uvažovali, že pravděpodobnost realizace jednotlivých výsledků náhodného experimentu jsou shodné. V řadě případů však tento předpoklad nemusí být splněn. Nechť $\Omega = \{\omega_1, \dots, \omega_N\}$ a předpokládáme $p_j = P[\{\omega_j\}]$ pro $j = 1, \dots, N$. Proto

$$1 = P[\Omega] = P[\cup_{j=1}^N \{\omega_j\}] = \sum_{j=1}^N P[\{\omega_j\}] = \sum_{j=1}^N p_j$$

Pro libovolný jev A pak definujeme $P[A] = \sum p_j$, kde sumace jde přes ta ω_j , která náleží A . Lze dokázat, že takto definovaná $P[\cdot]$ splňuje všechny tři axiomy pravděpodobnostní funkce, a proto je pravděpodobnostní funkcí.

1.2.6 Podmíněná pravděpodobnost a nezávislost

Definice 1.18 (Podmíněná pravděpodobnost) *Nechť A a B jsou dva jevy z \mathcal{A} pro pravděpodobnostní prostor $(\Omega, \mathcal{A}, P[\cdot])$. Podmíněná pravděpodobnost jevu A za předpokladu realizace jevu B je definována jako*

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \quad \text{pro } P[B] > 0$$

V případě $P[B] = 0$ není tato pravděpodobnost definována.

Z výše uvedeného je zřejmé, že $P[A \cap B] = P[A|B]P[B] = P[B|A]P[A]$ pro $P[A] > 0$ a $P[B] > 0$. Tato rovnice tak definuje vztah mezi $P[A|B]$ a $P[B|A]$ na základě nepodmíněných pravděpodobností $P[A]$ a $P[B]$.

Výše uvedená definice je v konsistentní s klasickou definicí pravděpodobnosti. Jestliže máme náhodný experiment, pro který jsou definovány jevy A a B , lze $P[A|B]$ definovat jako

$$P[A|B] = \frac{N_{AB}}{N_B}$$

kde N_B představuje počet realizací jevu B a N_{AB} představuje počet realizací jevu $A \cap B$. Je-li celkový počet realizací N , platí totiž

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{N_{AB}/N}{N_B/N} = \frac{N_{AB}}{N_B}$$

Když mluvíme o podmíněné pravděpodobnosti, podmiňujeme vzhledem k určitému jevu B (tj. předpokládáme, že výstup náhodného experimentu je B).

Náhodný jev B se tak stává novým výběrovým prostorem. Vystává otázka, zda-li je $P[\cdot|B]$ pravděpodobnostní funkcí s definičním oborem \mathcal{A} , tj. zda-li splňuje tři axiomy pravděpodobnostní funkce.

1. $P[A|B] = P[A \cap B]/P[B] \geq 0$ pro všechna $A \in \mathcal{A}$
2. $P[\Omega|B] = P[\Omega \cap B]/P[B] = P[B]/P[B] = 1$
3. Je-li A_1, A_2, \dots posloupností vzájemně neslučitelných jevů v \mathcal{A} a $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, pak

$$P\left[\bigcup_{i=1}^{\infty} A_i|B\right] = \frac{P\left[\left(\bigcup_{i=1}^{\infty} A_i\right)B\right]}{P[B]} = \frac{P\left[\bigcup_{i=1}^{\infty} (A_i \cap B)\right]}{P[B]} = \frac{\sum_{i=1}^{\infty} P[A_i \cap B]}{P[B]} = \sum_{i=1}^{\infty} P[A_i|B]$$

Následující věty jsou analogií vět z kapitoly (1.2.4).

Věta 1.22 $P[\emptyset|B] = 0$

Věta 1.23 Jsou-li A_1, \dots, A_n vzájemně neslučitelné jevy v \mathcal{A} , pak $P[A_1 \cup \dots \cup A_n|B] = \sum_{i=1}^n P[A_i|B]$.

Věta 1.24 Jestliže A jevem v \mathcal{A} , pak $P[A^c|B] = 1 - P[A|B]$.

Věta 1.25 Jestliže $A_1 \in \mathcal{A}$ a $A_2 \in \mathcal{A}$, pak $P[A_1|B] = P[A_1 \cap A_2|B] + P[A_1 \cap A_2^c|B]$.

Věta 1.26 Pro každé dva jevy $A_1 \in \mathcal{A}$ a $A_2 \in \mathcal{A}$ platí $P[A_1 \cup A_2|B] = P[A_1|B] + P[A_2|B] - P[A_1 \cap A_2|B]$

Věta 1.27 Jestliže $A_1 \in \mathcal{A}$, $A_2 \in \mathcal{A}$ a $A_1 \subset A_2$, pak $P[A_1|B] \leq P[A_2|B]$.

Věta 1.28 Jestliže $A_1, A_2, \dots, A_n \in \mathcal{A}$, pak $P[A_1 \cup A_2 \cup \dots \cup A_n|B] \leq \sum_{i=1}^n P[A_i|B]$.

Věta 1.29 (Celková pravděpodobnost) Jestliže pro daný pravděpodobnostní prostor $(\Omega, \mathcal{A}, P[\cdot])$ představuje B_1, B_2, \dots, B_n soubor vzájemně neslučitelných jevů v \mathcal{A} splňujících podmínky $\Omega = \bigcup_{j=1}^n B_j$ a $P[B_j] > 0$ pro $j = 1, \dots, n$, pak pro každé $A \in \mathcal{A}$ platí $P[A] = \sum_{j=1}^n P[A|B_j]P[B_j]$.

Důkaz 1.12 Všimněme si, že $A = \bigcup_{j=1}^n (A \cap B_j)$ a že jednotlivé $A \cap B_j$ jsou vzájemně neslučitelné. Proto

$$P[A] = P\left[\bigcup_{j=1}^n (A \cap B_j)\right] = \sum_{j=1}^n P[A \cap B_j] = \sum_{j=1}^n P[A|B_j]P[B_j]$$

Tvrzení 1.1 Uvažujme pravděpodobnostní prostor $(\Omega, \mathcal{A}, P[\cdot])$. Nechť $B \in \mathcal{A}$ splňuje $0 < P[B] < 1$. Pak pro každé $A \in \mathcal{A}$ platí $P[A] = P[A|B]P[B] + P[A|B^c]P[B^c]$

Věta 1.30 (Bayesův vzorec) Uvažujme pravděpodobnostní prostor $(\Omega, \mathcal{A}, P[\cdot])$ a soubor vzájemně neslučitelných jevů B_1, B_2, \dots, B_n v \mathcal{A} . Nechť tyto jevy splňují podmínky $\Omega = \bigcup_{j=1}^n B_j$ a $P[B_j] > 0$ pro $j = 1, 2, \dots, n$. Pak pro každé $A \in \mathcal{A}$, které splňuje $P[A] > 0$, platí

$$P[B_k|A] = \frac{P[A|B_k]P[B_k]}{\sum_{j=1}^n P[A|B_j]P[B_j]}$$

Důkaz 1.13 S využitím definice podmíněné pravděpodobnosti a věty o celkové pravděpodobnosti lze dokázat, že

$$P[B_k|A] = \frac{P[B_k \cap A]}{P[A]} = \frac{P[A|B_k]P[B_k]}{\sum_{j=1}^n P[A|B_j]P[B_j]}$$

Tvrzení 1.2 Pro daný pravděpodobnostní prostor $(\Omega, \mathcal{A}, P[\cdot])$ uvažujme takové jevy $A \in \mathcal{A}$ a $B \in \mathcal{A}$, které splňují $P[A] > 0$ a $0 < P[B] < 1$. Pak platí

$$P[B|A] = \frac{P[A|B]P[B]}{P[A|B]P[B] + P[A|B^c]P[B^c]}$$

Věta 1.31 (Pravidlo multiplikace) Pro daný pravděpodobnostní prostor $(\Omega, \mathcal{A}, P[\cdot])$ uvažujme jevy A_1, A_2, \dots, A_n náležící do \mathcal{A} , pro které $P[A_1 \cap \dots \cap A_n] > 0$. Pak

$$P[A_1 \cap \dots \cap A_n] = P[A_1]P[A_2|A_1]P[A_3|A_1 \cap A_2] \cdots P[A_n|A_1 \cap \dots \cap A_{n-1}]$$

Věta 1.32 (Nezávislé jevy) Pro daný pravděpodobnostní prostor $(\Omega, \mathcal{A}, P[\cdot])$ uvažujme náhodné jevy $A \in \mathcal{A}$ a $B \in \mathcal{A}$. Jevy A a B jsou nezávislé tehdy a jen tehdy, jestliže je splněna některá z následujících ekvivalentních podmínek.

1. $P[A \cap B] = P[A]P[B]$
2. $P[A|B] = P[A]$ jestliže $P[B] > 0$
3. $P[B|A] = P[B]$ jestliže $P[A] > 0$

Abychom dokázali, že všechny tři výše uvedené podmínky jsou ekvivalentní, stačí dokázat, že (1) implikuje (2), (2) implikuje (3) a (3) implikuje (1). Jestliže $P[A \cap B] = P[A]P[B]$, pak $P[A|B] = P[A \cap B]/P[B] = P[A]P[B]/P[B] = P[A]$ pro $P[B] > 0$. Z toho plyne, že (1) implikuje (2). Jestliže $P[A|B] = P[A]$, pak $P[B|A] = P[A|B]P[B]/P[A] = P[A]P[B]/P[A] = P[B]$ pro $P[A] > 0$ a $P[B] > 0$. Z toho plyne, že (2) implikuje (3). Jestliže $P[B|A] = P[B]$, pak $P[A \cap B] = P[B|A]P[A] = P[B]P[A]$ pro $P[A] > 0$. Je tedy zřejmé, že $P[A \cap B] = P[A]P[B]$, jestliže $P[A] = 0$ nebo $P[B] = 0$.

Nezávislost a neslučitelnost jevů A a B jsou dva rozdílné pojmy. Např. dva vzájemně neslučitelné jevy A a B jsou nezávislé tehdy a jen tehdy jestliže $P[A]P[B] = 0$. To je splněno tehdy a jen tehdy, jestliže jeden z jevů má nulovou pravděpodobnost. Neboli, jestliže $P[A] \neq 0$ a $P[B] \neq 0$, nezávislost mezi jevy A a B implikuje, že tyto jevy nejsou vzájemně neslučitelné. Analogicky $P[A] \neq 0$ a $P[B] \neq 0$ a neslučitenost jevů A a B implikuje, že se nejedná o nezávislé jevy.

Příklad 1.2 Uvažujme hod hrací kostkou. Definujme jev A jako situaci, kdy padne sudé číslo a jev B jako situaci, kdy padle liché číslo. Je zřejmé, že tyto jevy jsou neslučitelné, avšak nejsou nezávislé.

Věta 1.33 Jestliže jsou A a B dva nezávislé jevy definované na daném pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P[\cdot])$, pak A a B^c jsou nezávislé, A^c a B jsou nezávislé a A^c a B^c jsou nezávislé.

Důkaz 1.14

$$P[A \cap B^c] = P[A] - P[A \cap B] = P[A] - P[A]P[B] = P[A](1 - P[B]) = P[A]P[B^c]$$

V ostatních případech je důkaz analogický.

Definice 1.19 (Nezávislost náhodných jevů) Pro daný pravděpodobnostní prostor $(\Omega, \mathcal{A}, P[\cdot])$ uvažujme náhodné jevy A_1, A_2, \dots, A_n v \mathcal{A} . Jevy A_1, A_2, \dots, A_n jsou nezávislé tehdy a jen tehdy, jestliže

$$\begin{aligned} P[A_i \cap A_j] &= P[A_i]P[A_j] \quad \text{pro } i \neq j \\ P[A_i \cap A_j \cap A_k] &= P[A_i]P[A_j]P[A_k] \quad \text{pro } i \neq j, j \neq k, i \neq k \\ &\dots \\ P\left[\bigcap_{i=1}^n A_i\right] &= \prod_{i=1}^n P[A_i] \end{aligned}$$

Pro nezávislost vícero náhodných jevů je třeba, aby byly splněny všechny podmínky. Nestačí tak např. testovat pouze párovou nezávislost, protože ta sama o sobě nezaručuje vzájemnou nezávislost všech jevů.

Kapitola 2

Náhodné veličiny, distribuční a momentová funkce

2.1 Náhodná veličina a kumulativní distribuční funkce

Definice 2.1 (Náhodná veličina) Pro daný pravděpodobnostní prostor $(\Omega, \mathcal{A}, P[\cdot])$ je náhodná veličina X , označovaná také jako $X(\cdot)$, funkcí s definičním oborem hodnot Ω a reálnou osou jako oborem hodnot. Funkce $X(\cdot)$ musí splňovat podmínku, že množina $A_r = \{\omega : X(\omega) \leq r\}$ náleží do \mathcal{A} pro každé reálné číslo r .

Jestliže tuto definici zasadíme do rámce náhodného experimentu, pak

1. Ω je úplným souborem všech možných výsledků tohoto experimentu a
2. náhodná veličina $X(\cdot)$ je funkcí s definičním oborem Ω , která přiřazuje reálné číslo každému výsledku náhodného experimentu.

Požadavek, aby ω , pro které platí $X(\omega) \leq r$, bylo náhodným jevem (tj. prvkem \mathcal{A}), není nikterak omezující, protože naším záměrem je používat pojmu náhodná veličina pouze k popisu náhodných jevů. Jen zřídka kdy se totiž budeme zajímat o náhodnou veličinu jako takovou - předmětem našeho zájmu budou spíše náhodné jevy definované skrze náhodné veličiny. V následujícím textu budeme pro označení náhodných veličin používat velkých písmen latinské abecedy (např. X) a pro označení jejich konkrétních hodnot pak malých písmen latinské abecedy (např. $X = x$).

Příklad 2.1 Uvažujme hod mincí. Nechť náhodná veličina X označuje počet hlav. Je zřejmé, že $\Omega = \{\text{orel}, \text{hlava}\}$, $X(\omega) = 1$ pro $\omega = \text{hlava}$ a $X(\omega) = 0$ jestliže $\omega = \text{orel}$. Náhodná veličina X tak asociuje reálné číslo s každým možným výsledkem náhodného experimentu. Měli bychom tedy dokázat, že $\{\omega : X(\omega) \leq r\}$ náleží do \mathcal{A} pro každé reálné číslo r . V našem případě se \mathcal{A} skládá ze čtyř podmnožin - $\emptyset, \{\text{hlava}\}, \{\text{orel}\}$ a Ω . Jestliže $r < 0$, pak $\{\omega : X(\omega) \leq r\} = \emptyset$.

Jestliže $0 \leq r < 1$, pak $\omega : X(\omega) \leq r = \{\text{orel}\}$. Jestliže $r \geq 1$, pak $\{\omega : X(\omega)\} = \Omega = \{\text{hlava}, \text{orel}\}$. Proto pro každé r náleží množina $\{\omega : X(\omega) \leq r\}$ do \mathcal{A} a $X(\cdot)$ je tak náhodnou veličinou.

Definice 2.2 (Kumulativní distribuční funkce) Kumulativní distribuční funkce $F_X(\cdot)$ náhodné veličiny X je definována jako funkce, jejímž definičním oborem je reálná osa a oborem hodnot je interval $[0, 1]$ a která splňuje podmínku $F_X(x) = P[X \leq x] = P[X \leq x] = P[\{\omega : X(\omega) \leq x\}]$ pro každé reálné číslo x .

Každá náhodná veličina je jedinečně definována skrze svou kumulativní distribuční funkci. S pomocí příslušné kumulativní distribuční funkce a s pomocí definice náhodné veličiny lze nalézt pravděpodobnost daného náhodného jevu¹.

Příklad 2.2 Uvažujme hod mincí. Nechť náhodná veličina X označuje počet hlav. Kumulativní distribuční funkce je definována jako

$$F_X(x) = \begin{cases} 0 & \text{pro } x < 0 \\ \frac{1}{2} & \text{pro } 0 \leq x < 1 \\ 1 & \text{pro } x \geq 1 \end{cases}$$

popř. s využitím idikační funkce jako

$$F_X(x) = \frac{1}{2}I_{[0,1)}(x) + I_{[1,\infty)}(x)$$

Definice 2.3 (Vlastnosti kumulativní distribuční funkce) Z výše uvedené lze vypozorovat, že kumulativní distribuční funkce má následující vlastnosti.

1. $F_X(-\infty) \equiv \lim_{x \rightarrow -\infty} F_X(x) = 0$ a $F_X(+\infty) \equiv \lim_{x \rightarrow +\infty} F_X(x) = 1$
2. $F_X(\cdot)$ je monotonní, neklesající funkcí. To znamená, že $F_X(a) \leq F_X(b)$ pro $a < b$.
3. $F_X(\cdot)$ je spojitá zprava. To znamená, že $\lim_{0 < h \rightarrow 0} F_X(x+h) = F_X(x)^2$.

Důkaz 2.1 Dokažme vlastnost (2). Platí $\{\omega : X(\omega) \leq b\} = \{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$ a $\{X \leq a\} \cap \{a < X \leq b\} = \emptyset$. Proto $F_X(b) = P[X \leq b] = P[X \leq a] + P[a < X \leq b] = F_X(a)$.

Definice 2.4 (Kumulativní distribuční funkce) Libovolná funkce $F(\cdot)$ s reálnou osou jako definičním oborem a oborem hodnot $[0, 1]$, která splňuje všechny tři výše uvedené podmínky, je kumulativní distribuční funkce.

2.2 Pravděpodobnostní funkce

2.2.1 Nespojité náhodné veličiny

Definice 2.5 (Nespojitá náhodná veličina) Náhodná veličina X je nespojitá, jestliže množina hodnot, kterých může nabývat, je spočetná. Jestliže je náhodná veličina X nespojitá, pak je nespojitá také její kumulativní distribuční funkce $F_X(\cdot)$. Vedle pojmu nespojitá náhodná veličina se někdy používá také pojem diskrétní náhodná veličina.

¹V tomto případě je využito požadavku, aby $\omega : X(\omega) \leq r$ náleželo do \mathcal{A} .

²Spojitosť zprava je dána definicí $F_X(x) = P[X \leq x]$. Pokud bychom, jako někteří jiní autoři, použili definici $F_X(x) = P[X < x]$, pak by $F_X(\cdot)$ byla spojitá zleva.

Spočetnou množinou rozumíme množinu, která má konečný popř. spočetný počet prvků x_1, x_2, \dots . Jestliže náhodná veličina X nabývá náhodných hodnot z takovéto množiny, pak $\Omega = \bigcup_n \{\omega : X(\omega) = x_n\} = \bigcup_n \{X = x_n\}$ a $\{X = x_i\} \cap \{X = x_j\} = \emptyset$ pro $i \neq j$. Proto na základě třetího axiomu pravděpodobnosti platí $1 = P[\Omega] = \sum_n P[X = x_n]$.

Definice 2.6 (Pravděpodobnostní funkce nespojité náhodné veličiny)

Uvažujme náhodnou veličinu X , která může nabývat hodnot $x_1, x_2, \dots, x_n, \dots$. Pak funkce $f_X(\cdot)$ definovaná jako

$$f_X(x) = \begin{cases} P[X = x_j] & \text{pro } x = x_j, j = 1, 2, \dots, n, \dots \\ 0 & \text{pro } x \neq x_j \end{cases}$$

je nespojitou pravděpodobnostní funkcí nespojité náhodné veličiny X .

Hodnoty x_1, x_2, \dots , kterých může nespojitá náhodná veličina X nabývat, nazýváme body koncentrace. Hodnotu pravděpodobnostní funkce $f_X(x_j)$ pak nazýváme koncentrací v bodě x_j .

Věta 2.1 *Nechť je X nespojitou náhodnou veličinou. Platí, že $F_X(\cdot)$ lze získat z $f_X(\cdot)$ a naopak.*

Důkaz 2.2 *Označme hodnoty náhodné veličiny X jako x_1, x_2, \dots*

1. *Jestliže je dáno $f_X(\cdot)$, pak $F_X(x) = \sum_{j: x_j \leq x} f_X(x_j)$.*
2. *Jestliže je dáno $F_X(x)$, pak $f_X(x_j) = F_X(x_j) - \lim_{h \rightarrow 0^+} F_X(x_j - h)$. $f_X(x_j)$ lze tímto způsobem nalézt pro všechny body koncentrace x_j a $f_X(x) = 0$ pro $x \neq x_j, j = 1, 2, \dots$, čímž jsme definovali $f_X(x)$ pro všechna reálná čísla.*

Příklad 2.3 *Uvažujme hod hrací kostkou. Náhodná veličina X , která představuje hodnotu hodu, má množinu bodů koncentrace $\{1, 2, 3, 4, 5, 6\}$. Odpovídající nespojitá pravděpodobnostní funkce a kumulativní distribuční funkce jsou pak definovány jako*

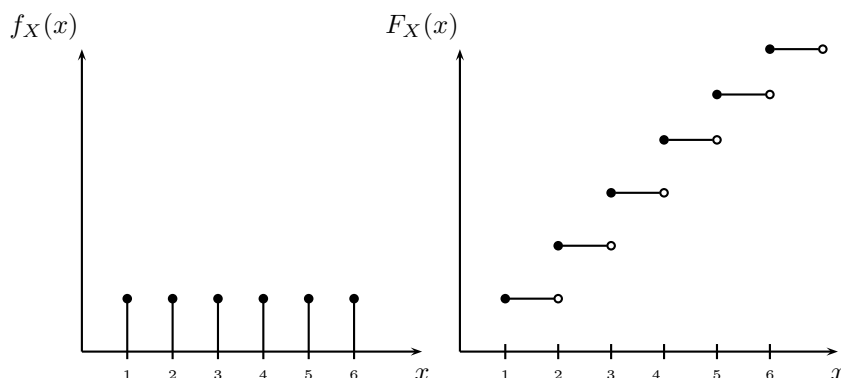
$$f_X(x_j) = \frac{1}{6} \quad \text{pro } j = 1, 2, \dots, 6$$

$$F_X(x_j) = \frac{j}{6} \quad \text{pro } j = 1, 2, \dots, 6$$

Z obrázku (2.2.1) je patrné, že kumulativní distribuční funkce nespojité náhodné veličiny X je skoková v bodech koncentrace, tj. že v bodě x_j skokově změní hodnotu o $f_X(x_j)$. Mezi jednotlivými body koncentrace je pak $F_X(\cdot)$ konstantní.

Definice 2.7 (Nespojitá pravděpodobnostní funkce) *Libovolná funkce $f(\cdot)$ s reálnou osou jako definičním oborem a oborem hodnot $[0, 1]$ je nespojitou (popř. diskrétní) pravděpodobnostní funkcí, jestliže pro určitou spočetnou množinu x_1, x_2, \dots platí*

1. $f(x_j) > 0$ pro $j = 1, 2, \dots$
2. $f(x) = 0$ pro $x \neq x_j, j = 1, 2, \dots$
3. $\sum f(x_j) = 1$ kde suma je přes body x_1, x_2, \dots



Obrázek 2.1: Hod hrací kostkou: Pravděpodobnostní funkce a kumulativní distribuční funkce

2.2.2 Spojitá náhodná veličina

Definice 2.8 (Spojitá náhodná veličina) Náhodnou veličinu X nazýváme spojitou, jestliže existuje funkce $f_X(\cdot)$ taková, že $F_X(x) = \int_{-\infty}^x f_X(u) du$ pro každé reálné číslo x .

Definice 2.9 (Pravděpodobnostní funkce spojité náhodné veličiny)

Jestliže je X spojitá náhodná veličina, pak je funkce $f_X(\cdot)$ ze vztahu $F_X(x) = \int_{-\infty}^x f_X(u) du$ nazývána pravděpodobnostní funkcí náhodné veličiny X .

Pokud bychom slepě lpěli na definici, pak pravděpodobnostní funkce $f_X(\cdot)$ spojité náhodné veličiny X není jedinečná. Např. kumulativní distribuční funkce $F_X(x) = xI_{[0,1)}(u) + I_{[1,\infty)}(x)$ je “kompatibilní” s pravděpodobnostní funkcí $f_X(u) = I_{(0,1)}(u)$ i s pravděpodobnostní funkcí $f_X(u) = I_{(0,\frac{1}{2})}(u) + 69I_{\frac{1}{2}}(u) + I_{(\frac{1}{2},1)}(u)$. V praxi je však jedinečnost pravděpodobnostní funkce daná požadavky na její spojitost.

Věta 2.2 Necht' je X spojitou náhodnou veličinou. Pak $F_X(\cdot)$ lze odvodit z $f_X(\cdot)$ a naopak.

Důkaz 2.3 Jestliže X je spojitá náhodná veličina a $f_X(\cdot)$ je dána, pak lze $F_X(x)$ odvodit integrací $f_X(\cdot)$, tj. $F_X(x) = \int_{-\infty}^x f_X(u) du$. Pokud je dáno $F_X(\cdot)$, pak lze $f_X(x)$ odvodit derivací, tj. $f_X(x) = dF_X(x)/dx$ pro ty body x , ve kterých pro $F_X(x)$ existuje derivace.

Značení pro spojitou a nespojitou pravděpodobnostní funkci jsou shodná, přesto mají odlišnou interpretaci. Pro nespojitou náhodnou veličinu je $f(x)$ definováno jako

$$f_X(x) = P[X = x]$$

Pro spojitou veličinu platí

$$f_X(x) = \frac{dF_X(x)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{F_X(x + \Delta x) - F_X(x - \Delta x)}{2\Delta x}$$

a proto $f_X(x)2\Delta x \approx F_X(x+\Delta x) - F_X(x-\Delta x) = P[x-\Delta x < X \leq x+\Delta x]$, což znamená, že pravděpodobnost, že se X nachází v malém intervalu obsahujícím hodnotu x , je rovna $f_X(x)$ krát délka tohoto intervalu. Pro nespojitou náhodnou veličinu je $f_X(\cdot)$ funkcí s reálnou osou jako definičním oborem a intervalem $[0, 1]$ jako oborem hodnot. V případě spojité náhodné veličiny je $f_X(\cdot)$ funkcí s reálnou osou jako definičním oborem a intervalem $[0, \infty)$ jako oborem hodnot.

Příklad 2.4 *Nechť náhodná veličina X představuje délku telefonního hovoru. Předpokládejme, kumulativní distribuční funkce X je $F_X(x) = (1 - e^{-\lambda x})I_{[0, \infty)}(x)$, kde $\lambda > 0$. Odpovídající pravděpodobnostní funkce je $f_X(x) = \lambda e^{-\lambda x}I_{[0, \infty)}(x)$. Jestliže je délka hovoru měřena v minutách, pak pro $\lambda = \frac{1}{5}$ je $P[5 < X \leq 10] = \int_5^{10} \lambda e^{-\lambda x} dx = e^{-5\lambda} - e^{-10\lambda} = e^{-1} - e^{-2} \approx 0.23$. Analogicky lze tuto pravděpodobnost vyjádřit jako $P[5 < X \leq 10] = P[X \leq 10] - P[X \leq 5] = (1 - e^{-10\lambda}) - (1 - e^{-5\lambda}) = e^{-1} - e^{-2}$.*

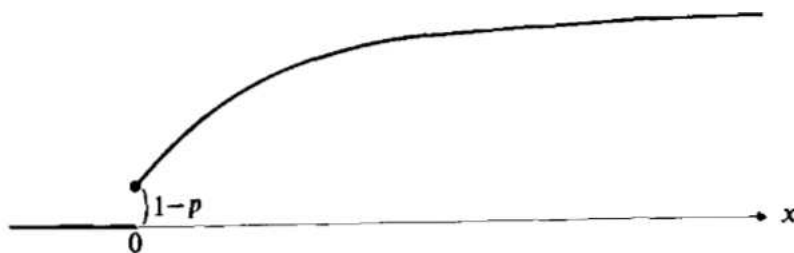
Pravděpodobnostní funkci je možné použít pro výpočet pravděpodobnosti jevů definovaných pomocí spojité náhodné veličiny X . Např. $P[a < X \leq b] = \int_a^b f_X(x) dx$ pro $a < b$.

Definice 2.10 (Spojitá pravděpodobnostní funkce) *Libovolná funkce $f(\cdot)$ s reálnou osou jako definičním oborem a $[0, \infty)$ jako oborem hodnot je spojitou pravděpodobnostní funkcí tehdy a jen tehdy, jestliže*

1. $f(x) \geq 0$ pro všechna x
2. $\int_{-\infty}^{\infty} f(x) dx = 1$

2.2.3 Ostatní náhodné veličiny

Ne všechny náhodné veličiny jsou buďto spojité nebo nespojité a ne všechny kumulativní distribuční funkce jsou buďto absolutně spojité nebo nespojité. V praxi existuje řada případů, kdy je kumulativní distribuční funkce částečně spojitá a částečně nespojitá. Příkladem může být čekání na semaforu - pokud svítí zelená, je doba čekání rovna nule; pokud svítí červená, záleží doba čekání na tom, ve které fázi jsme na křižovatku dorazili. Situaci ilustruje obrázek (2.2).



Obrázek 2.2: Kumulativní distribuční funkce se skokem v $x = 0$ a spojitá pro $x > 0$

Vedle toho existuje tzv. singularně spojitá kumulativní distribuční funkce, jejichž derivace je téměř ve všech bodech rovna nule. Tento typ kumulativní distribuční funkce je použit v následujícím tvrzení.

Tvrzení 2.1 *Jakákoliv kumulativní distribuční funkce $F(x)$ může být vyjádřena ve formě*

$$F(x) = p_1 F^d(x) + p_2 F^{ac}(x) + p_3 F^{sc}(x)$$

kde $p_i \geq 0$, $i = 1, 2, 3$, $\sum_{i=1}^3 p_i = 1$, $F^d(\cdot)$ je nespojitá, $F^{ac}(\cdot)$ je absolutně spojitá a $F^{sc}(\cdot)$ je singulárně spojitá kumulativní distribuční funkce.

Příklad 2.5 *Uvažujme $F_X(x) = (1 - pe^{-\lambda x})I_{[0, \infty)}(x)$. Tuto kumulativní distribuční funkci lze rozepsat do tvaru*

$$\begin{aligned} F_X(x) &= (1 - p)F^d(x) + pF^{ac}(x) = (1 - p)I_{[0, \infty)}(x) + p(1 - e^{-\lambda x})I_{[0, \infty)}(x) \\ &= (1 - pe^{-\lambda x})I_{[0, \infty)}(x) \end{aligned}$$

Uvažujme $F(x) = (1 - p)F^d(x) + pF^{ac}(x)$, kde $0 < p < 1$, $F^d(\cdot)$ je nespojitá a $F^{ac}(\cdot)$ je absolutně spojitá kumulativní distribuční funkce. Odpovídající pravděpodobnostní funkce je definována jako $f(x) = (1 - p)f^d(x) + pf^{ac}(x)$, kde $f^d(\cdot)$ je nespojitá pravděpodobnostní funkce odpovídající funkci $F^d(\cdot)$ a $f^{ac}(\cdot)$ je absolutně spojitá pravděpodobnostní funkce odpovídající funkci $F^{ac}(\cdot)$. Tato částečně nespojitá a částečně absolutně spojitá pravděpodobnostní funkce může být poměrně komplikovaná na interpretaci, a proto v takovýchto případech budeme spíše pracovat s kumulativní distribuční funkcí.

2.3 Střední hodnoty a momenty

2.3.1 Střední hodnota

Definice 2.11 (Střední hodnota) *Uvažujme náhodnou veličinu X . Střední hodnota X , označovaná jako $E[X]$ popř. jako μ_x , je definována následujícím způsobem.*

1. *Jestliže je X nespojité, pak*

$$E[X] = \sum x_j f_X(x_j)$$

pro body koncentrace $x_1, x_2, \dots, x_j, \dots$

2. *Jestliže je X spojité, pak*

$$E[X] = \int_0^\infty x f_X(x) dx$$

3. *Pro libovolné X pak platí*

$$E[X] = \int_0^\infty [1 - F_X(x)] dx - \int_{-\infty}^0 F_X(x) dx$$

V případě (1) $E[X]$ existuje, jestliže je daná řada absolutně konvergentní. V případě (2) $E[X]$ existuje, jestliže daný integrál existuje. Konečně, v případě (3) existuje $E[X]$, pokud jsou oba integrály konečné.

Z definicí (1) a (2) je patrné, že na $E[X]$ je možné pohlížet jako na vážený průměr jednotlivých hodnot, kterých může náhodná veličina X nabývat, přičemž vahami jsou příslušné pravděpodobnostní reprezentované pravděpodobnostní funkcí $f_X(x)$ popř. $f_X(x)dx$.

Příklad 2.6 *Uvažujme hod hrací kostkou. Střední hodnota náhodné veličiny X , která představuje hodnotu hodu, je rovna*

$$E[X] = \sum_{i=1}^6 i \frac{1}{6} = \frac{7}{2}$$

Příklad 2.7 *Nechť je X náhodná veličina s pravděpodobnostní funkcí $f_X(x) = \frac{1}{x^2} I_{[1,\infty)}(x)$. Pak*

$$E[X] = \int_1^\infty x \frac{1}{x^2} dx = \lim_{b \rightarrow \infty} \ln b = \infty$$

a proto $E[X]$ neexistuje.

Příklad 2.8 *Nechť je X náhodná veličina s kumulativní distribuční funkcí $F_X(x) = (1 - pe^{-\lambda x}) I_{[0,\infty)}(x)$. Pak*

$$E[X] = \int_0^\infty [1 - F_X(x)] dx - \int_{-\infty}^0 F_X(x) dx = \int_0^\infty pe^{-\lambda x} dx = \frac{p}{\lambda}$$

2.3.2 Rozptyl

Definice 2.12 (Rozptyl) *Nechť je X náhodná veličina se střední hodnotou $E[X]$. Rozptyl této náhodné veličiny označovaný jako $D[X]$ resp. jako σ_X^2 , je definován následujícím způsobem.*

1. *Jestliže je X nespojitý s body koncentrace $x_1, x_2, \dots, x_n, \dots$, pak*

$$D[X] = \sum_i (x_i - E[X])^2 f_X(x_i)$$

2. *Jestliže je X spojitý, pak*

$$D[X] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx$$

3. *Pro libovolnou náhodnou veličinu s kumulativní distribuční funkcí $F_X(x)$ platí*

$$D[X] = \int_0^\infty 2x[1 - F_X(x) + F_X(-x)] dx - E[X]^2$$

Rozptyl je definován pouze tehdy, jestliže je řada v (1) absolutně konvergentní nebo jestliže integrály v (2) a (3) existují. Vzhledem k tomu, že $f_X(x) > 0$ pro všechna x , je rozptyl vždy kladný.

Podobně jako střední hodnota i rozptyl je váženým průměrem. V obou případech jsou váhy shodné. V případě rozptylu se váží druhá mocnina odchylky náhodné veličiny od její střední hodnoty, kdežto v případě střední hodnoty se váží samotná hodnota náhodné veličiny.

Definice 2.13 (Směrodatná odchylka) Jestliže je X náhodná veličina, pak je její směrodatná odchylka definována jako $\sigma_X = \sqrt{D[X]}$.

Z výše uvedené definice je zřejmé, že směrodatná odchylka poskytuje stejnou informaci jako rozptyl. V řadě případů je však směrodatná odchylka preferována před rozptylem, protože je vyjádřena ve stejné měrné jednotce jako náhodná veličina.

Příklad 2.9 Uvažujme hod hrací kostkou. Rozptyl náhodné veličiny X představující hodnotu hodu, je

$$D[X] = \sum_{i=1}^6 \left(i - \frac{7}{2}\right)^2 \frac{1}{6} = \frac{35}{12}$$

Příklad 2.10 Nechť je X náhodná veličina s pravděpodobnostní funkcí $f_X(x) = \lambda e^{-\lambda x} I_{[0, \infty)}(x)$. Střední hodnota náhodné veličiny X je rovna $\frac{1}{\lambda}$ a její rozptyl je pak roven

$$D[X] = \int_{-\infty}^{\infty} \left(x - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}$$

Příklad 2.11 Nechť je X náhodná veličina s kumulativní distribuční funkcí $F_X(x) = (1 - pe^{-\lambda x}) I_{[0, \infty)}(x)$ a střední hodnotou $\frac{p}{\lambda}$. Rozptyl této náhodné veličiny je roven

$$D[X] = \int_0^{\infty} 2x[1 - F(x) + F(-x)] dx - E[X]^2 = \int_0^{\infty} 2xpe^{-\lambda x} dx - \left(\frac{p}{\lambda}\right)^2 = \frac{p(2-p)}{\lambda^2}$$

2.3.3 Střední hodnota funkce náhodné veličiny

Definice 2.14 (Střední hodnota funkce náhodné veličiny) Nechť je X náhodná veličina a $g(\cdot)$ je funkce s reálnou osou jako definičním oborem i oborem hodnot.

1. Jestliže je X nespojitý, pak

$$E[g(X)] = \sum_j g(x_j) f_x(x_j)$$

2. Jestliže je X spojitý, pak

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

Věta 2.3 Střední hodnota funkce náhodné veličiny má následující vlastnosti.

1. $E[c] = c$ pro konstantu c
2. $E[cg(X)] = cE[g(X)]$ pro konstantu c
3. $E[c_1g_1(X) + c_2g_2(X)] = c_1E[g_1(X)] + c_2E[g_2(X)]$
4. $E[g_1(X)] \leq E[g_2(X)]$ jestliže $g_1(x) \leq g_2(x)$ pro všechna x

Věta 2.4 (Výpočtový tvar rozptylu) Jestliže je X náhodná veličina, pak

$$D[X] = E[X^2] - E[X]^2$$

za předpokladu, že $E[X]$ a $E[X^2]$ existují.

Důkaz 2.4 Z definice rozptylu a střední hodnoty funkce náhodné veličiny vyplývá $D[X] = E[X - E[X]]^2$. Tento vztah lze dále upravit do podoby

$$E[X - E[X]]^2 = E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2$$

2.3.4 Chebysheva nerovnost

Věta 2.5 Nechť je X náhodnou veličinou a $g(\cdot)$ nezápornou funkcí s reálnou osou jako definičním oborem. Pak

$$P[g(X) \geq k] \leq \frac{E[g(X)]}{k} \quad \text{pro všechna } k > 0$$

Důkaz 2.5 Předpokládejme, že X je spojitá náhodná veličina s pravděpodobnostní funkcí $f_X(\cdot)$. Pak

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{\infty} g(x)f_X(x)dx = \int_{x:g(x) \geq k} g(x)f_X(x)dx + \int_{x:g(x) < k} g(x)f_X(x)dx \\ &\geq \int_{x:g(x) \geq k} g(x)f_X(x)dx \geq \int_{x:g(x) \geq k} kf_X(x)dx = kP[g(X) \geq k] \end{aligned}$$

Důkaz završíme vydělením k . Analogický důkaz lze aplikovat také na nespojitou náhodnou veličinu.

Tvrzení 2.2 (Chebysheva nerovnost) Jestliže je X náhodná veličina s konečným rozptylem, pak

$$P[|X - E[X]| > r\sigma_X] = P[(X - E[X])^2 \geq r^2\sigma_X^2] \leq \frac{1}{r^2} \quad \text{pro všechna } r > 0$$

Důkaz 2.6 Do věty (2.5) dosadíme $g(x) = (x - E[X])^2$ a $k = r^2\sigma_X^2$.

Jedna z užitečných aplikací Chebyshevovy nerovnosti je následující. Pro náhodnou veličinu s konečným rozptylem platí

$$P[|X - E[X]| < r\sigma_X] \geq 1 - \frac{1}{r^2}$$

což implikuje

$$P[E[X] - r\sigma_X < X < E[X] + r\sigma_X] \geq 1 - \frac{1}{r^2}$$

Příklad 2.12 Pro $r = 2$ např. získáme $P[E[X] - 2\sigma_X < X < E[X] + 2\sigma_X] \geq \frac{3}{4}$. Libovolná náhodná veličina X s konečným rozptylem se tak s pravděpodobností $\frac{3}{4}$ nachází do dvou směrodatých odchylek od své střední hodnoty. Toto zjištění platí pro všechny náhodné veličiny bez ohledu na jejich kumulativní distribuční funkci.

2.3.5 Jensenova nerovnost

Definice 2.15 (Konvexní funkce) Spojitá funkce $g(\cdot)$ s reálnou osou jako definičním oborem a oborem hodnot se nazývá konvexní, jestliže pro každé reálné x_0 existuje přímka, která prochází bodem $(x_0, g(x_0))$ a leží na nebo pod grafem funkce $g(\cdot)$.

Věta 2.6 (Jensenova nerovnost) Nechť je X náhodná veličina se střední hodnotou $E[X]$ a nechť je $g(\cdot)$ konvexní funkcí. Pak $E[g(X)] \geq g(E[X])$.

Důkaz 2.7 Protože je $g(x)$ spojitá a konvexní funkce, pak existuje přímka $l(x) = a + bx$, která splňuje podmínky $l(x) = a + bx \geq g(x)$ a $l(E[X]) = g(E[X])$. Uvažovaná přímka $l(x)$ tak prochází bodem $(E[X], g(E[X]))$. Platí $E[l(X)] = E[(a + bX)] = a + bE[X] = l(E[X])$. Proto platí $g(E[X]) = l(E[X]) = E[l(X)] \leq E[g(X)]$.

Jensenova nerovnost mimojiné implikuje, že obecně nemusí platit $E[g(X)] = g(E[X])$. Dalším zajímavým důsledkem je, že pro konvexní funkci $g(x) = x^2$, platí $E[X^2] \geq E[X]^2$, což implikuje nezáporný rozptyl náhodné veličiny X .

2.3.6 Momenty a momentová funkce

Definice 2.16 (Obecný moment) Jestliže je X náhodná veličina, je její r -tý obecný moment, označovaný jako μ'_r , definován jako

$$\mu'_r = E[X^r]$$

jestliže daná střední hodnota existuje.

Definice 2.17 (Centrální moment) Jestliže je X náhodná veličina, pak je r -tý centrální moment okolo a definován jako $E[(X - a)^r]$. Jestliže $a = \mu_X$, pak centrální moment náhodné veličiny X okolo μ_X , označovaný jako μ_r , je definován jako

$$\mu_r = E[(X - \mu_X)^r]$$

Z výše uvedeného je patrné, že $\mu_1 = E[(X - \mu_X)] = 0$ a $\mu_2 = E[(X - \mu_X)^2] = D[X]$. Dále platí, že liché centrální momenty náhodné veličiny X , jejíž pravděpodobnostní funkce je symetrická kolem $E[X]$, jsou rovny nule, pokud jsou tyto momenty definovány.

Definice 2.18 (Vztah mezi obecným a centrálním momentem) Platí, že z obecného momentu je možné odvodit centrální moment a naopak.

$$\mu_k = \sum_{j=0}^k \binom{k}{j} (-\mu'_1)^j \mu'_{k-j}$$

$$\mu_k = \sum_{j=0}^k \binom{k}{j} (\mu'_1)^j \mu_{k-j}$$

Definice 2.19 (Kvantil) q -tý kvantil náhodné veličiny X je označován jako ξ_q a definován jako nejmenší číslo ξ splňující podmínku $F_X(\xi) \geq q$.

Definice 2.20 (Medián) Medián náhodné veličiny X označovaný jako med_X nebo také $\xi_{0.50}$ je 0.5-tým kvantilem.

Jestliže je X spojité, pak medián splňuje podmínku

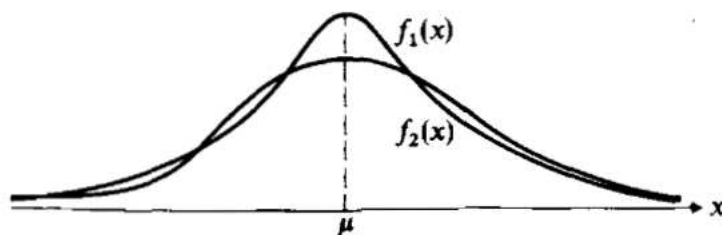
$$\int_{-\infty}^{med_x} f_X(x) dx = \frac{1}{2} = \int_{med_x}^{\infty} f_X(x) dx$$

Definice 2.21 (Modus) Další veličinou, kterou je možné použít k popisu “centra” náhodné veličiny X , je kromě střední hodnoty a mediánu tzv. modus. Modus je definován bod, pro který dosahuje $f_X(\cdot)$ maxima.

Předpokládejme, že $f_1(x)$ a $f_2(x)$ jsou dvě pravděpodobnostní funkce se stejnou střední hodnotou μ , které pro libovolné a splňují podmínku

$$\int_{\mu-a}^{\mu+a} [f_1(x) - f_2(x)] dx \geq 0$$

Takovéto pravděpodobnostní funkce jsou ilustrovány na následujícím obrázku. Lze dokázat, že rozptyl σ_1^2 pravděpodobnostní funkce $f_1(x)$ je menší než rozptyl

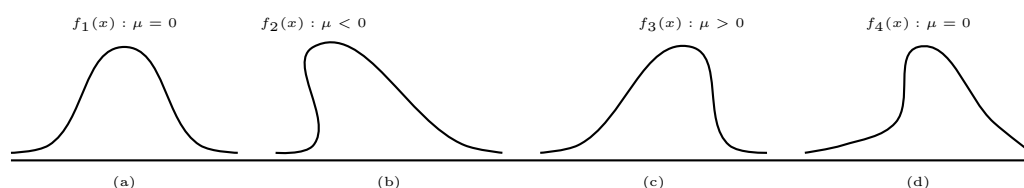


Obrázek 2.3: Pravděpodobnostní funkce $f_1(x)$ a $f_2(x)$ se shodnou střední hodnotou μ

σ_2^2 pravděpodobnostní funkce $f_2(x)$. Schéma důkazu je následující. Nechť $g(x) = f_1(x) - f_2(x)$, kde $f_1(x)$ a $f_2(x)$ splňují výše uvedenou podmínku. Protože z definice pravděpodobnostní funkce vyplývá $\int_{-\infty}^{\infty} g(x) dx = 0$, je v případě funkce $g(x)$ obsah ploch nad osou x roven obsahu ploch od osou x . Dále je z podmínky pro funkce $f_1(x)$ a $f_2(x)$ patrné, že každý kladný element $g(x') dx'$ je vyvážen záporným elementem $g(x'') dx''$, kde x'' je dále od μ než x' . Jestliže tyto elementy násobíme $(x - \mu)^2$, jsou záporné elementy násobeny větším faktorem. Proto $\int_{-\infty}^{\infty} (x - \mu)^2 g(x) dx < 0$, pokud jsou $f_1(x)$ a $f_2(x)$ rozdílné. Z toho vyplývá $\sigma_1^2 < \sigma_2^2$.

Třetí centrální moment μ_3 nazýváme mírou sešikmení. Pravděpodobnostní funkce, které jsou symetrické okolo své střední hodnoty, mají sešikmení $\mu_3 = 0$. Pravděpodobnostní funkce sešikmené do leva mají $\mu_3 < 0$. Pravděpodobnostní funkce sešikmené do prava pak mají $\mu_3 > 0$. Poněkud paradoxně nám však míra sešikmení neříká nic tvaru pravděpodobnostní funkce. Např. pravděpodobnostní funkce $f_4(x)$ na následujícím obrázku má $\mu_3 = 0$, avšak o symetrickou funkci se nejedná. Alternativním ukazatelem sešikmení je

$$s = \frac{\text{průměr} - \text{medián}}{\text{směrodatná odchylka}}$$

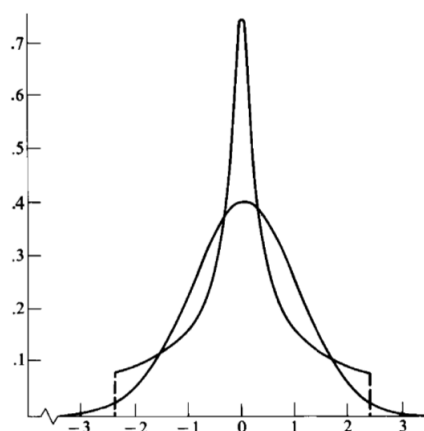


Obrázek 2.4: Pravděpodobnostní funkce s rozdílnou mírou sešikmení

Lze dokázat, že $-1 \leq s \leq 1$.

Čtvrtý centrální moment se používá jako ukazatel zakřivení, která vyjadřuje míru “plochosti” pravděpodobnostní funkce v okolí jejího středu. Pozitivní hodnota $\frac{\mu_4}{\sigma^4} - 3$ indikuje, že daná pravděpodobnostní funkce je v okolí svého středu více “špičatá” než pravděpodobnostní funkce normálního rozdělení. Naopak, negativní hodnota $\frac{\mu_4}{\sigma^4} - 3$ indikuje, že daná pravděpodobnostní funkce je v okolí svého středu více “plochá” než pravděpodobnostní funkce normálního rozdělení. Podobně jako v případě sešikmení, ani v případě zakřivení není tento ukazatel vždy směrodatný pro tvar pravděpodobnostní funkce.

Ačkoliv několik momentů může poskytnout pouze omezenou informaci o pravděpodobnostní funkci, soubor všech momentů definuje pravděpodobnostní funkci jednoznačně.



Obrázek 2.5: Dvě pravděpodobnostní funkce se shodnými prvními čtyřmi momenty

V aplikované statistice hrají klíčovou roli především první dva momenty, třetí a čtvrtý moment se používají vyjimečně a vyšší momenty pak téměř vůbec.

Definice 2.22 (Faktoriální moment) Jestliže je X náhodnou veličinou, pak r -tý faktoriální moment X je definován jako

$$E[X(X-1)\cdots(X-r+1)]$$

Pro některé, zpravidla nespojité, náhodné veličiny je snazší vypočítat faktoriální

moment než obecný popř. centrální moment. Nicméně platí, že z faktoriálního momentu lze odvodit obecný popř. centrální moment a naopak.

Definice 2.23 (Momentová funkce) *Nechť je X náhodná veličina s pravděpodobnostní funkcí $f_X(\cdot)$. Očekávaná hodnota e^{tX} je definuje tzv. momentovou funkci, jestliže tato očekávaná hodnota existuje pro každé t v intervalu $-h < t < h$ pro $h > 0$. Momentová funkce má tvar*

$$m(t) = E[e^{tx}] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

je-li X spojitě a

$$m(t) = E[e^{tx}] = \sum_x e^{tx} f_X(x)$$

je-li X nespojitě.

Existuje-li momentová funkce, pak je spojitě diferenciovatelná v okolí počátku. Jestliže momentovou funkci r krát derivujeme vzhledem k t , pak získáme

$$\frac{d^r}{dt^r} m(t) = \int_{-\infty}^{\infty} x^r e^{xt} f_X(x) dx$$

a jestliže $r \rightarrow 0$, pak

$$\frac{d^r}{dt^r} m(0) = E[X^r] = \mu'_r$$

Momenty pravděpodobnostní funkce tak lze získat z momentové funkce její derivací dle t .

Jestliže v momentové funkci nahradíme e^{tx} Taylorovým rozvojem, pak získáme

$$m(t) = E\left[1 + tX + \frac{1}{2!}(tX)^2 + \frac{1}{3!}(tX)^3 + \dots\right] = 1 + \mu'_1 t + \frac{1}{2!}\mu'_2 t^2 + \dots = \sum_{i=0}^{\infty} \frac{1}{i!} \mu'_i t^i$$

Z výše uvedených rovnic je patrné, že μ'_r lze získat z Taylorova rozvoje $m(t)$ jako koeficient prvku $\frac{t^r}{r!}$.

Příklad 2.13 *Nechť je X náhodná veličina s pravděpodobnostní funkcí $f_X(x) = \lambda e^{-\lambda x} I_{[0, \infty)}(x)$. Pak*

$$m(t) = E[e^{tX}] = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t} \quad \text{pro } t < \lambda$$

a proto

$$\begin{aligned} m'(t) &= \frac{\lambda}{(\lambda - t)^2} \quad \text{což implikuje} \quad m'(0) = E[X] = \frac{1}{\lambda} \\ m''(t) &= \frac{2\lambda}{(\lambda - t)^3} \quad \text{což implikuje} \quad m''(0) = E[X^2] = \frac{2}{\lambda^2} \end{aligned}$$

Příklad 2.14 *Uvažujme náhodnou veličinu X s pravděpodobnostní funkcí $f_X(x) = x^{-2} I_{[1, \infty)}(x)$. Kdyby momentová funkce této náhodné veličiny existovala, pak by byla dána rovnicí $\int_1^{\infty} X^{-2} e^{tx} dx$. Lze však dokázat, že takovýto integrál neexistuje pro žádné $t > 0$, a proto také neexistuje tato momentová funkce.*

Definice 2.24 (Faktoriální momentová funkce) *Nechť je X náhodná veličina. Faktoriální momentová funkce je definována jako $E[t^x]$, jestliže tato střední hodnota existuje.*

Faktoriální momentovou funkci lze použít pro generování faktoriálních momentů stejným způsobem jakým lze odvodit obecný moment z momentové funkce. Jediným rozdílem je, že $t \rightarrow 1$ namísto $t \rightarrow 0$. Tento postup někdy zjednodušuje nalezení momentu v případě nespojitě pravděpodobnostní funkce.

Příklad 2.15 *Nechť je X náhodná veličina s nespojitou pravděpodobnostní funkcí*

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

Pak

$$E[t^X] = \sum_{x=0}^{\infty} \frac{t^x e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}$$

$$\frac{d}{dt} E[t^X] = \frac{d}{dt} e^{\lambda(t-1)} = \lambda e^{\lambda(t-1)} \quad \text{a proto} \quad \left. \frac{d}{dt} E[t^X] \right|_{t=1} = \lambda$$

Momentovou funkci lze, jak její název napovídá, použít k výpočtu momentů. Vedle toho ji však lze také použít k určení pravděpodobnostní funkce.

Věta 2.7 *Nechť jsou X a Y dvě náhodné veličiny s pravděpodobnostními funkcemi $f_X(\cdot)$ a $f_Y(\cdot)$. Předpokládejme, že jejich momentové funkce $m_X(t)$ a $m_Y(t)$ existují a jsou shodné pro všechna t z intervalu $-h < t < h$ pro $h > 0$. Potom jsou shodné také pravděpodobnostní funkce $f_X(\cdot)$ a $f_Y(\cdot)$.*

Kapitola 3

Vybraná pravděpodobnostní rozdělení

Nejčastěji používané pravděpodobnostní funkce mají přiřazeno jméno. Říkáme tak např., že daná náhodná veličina sleduje normální pravděpodobnostní rozdělení. Stejně jako náhodné veličiny také pravděpodobnostní rozdělení dělíme na spojitá a nespojitá. Následující kapitola je stručným přehledem v praxi nejběžněji používaných pravděpodobnostních rozdělení.

3.1 Nespojité pravděpodobnostní rozdělení

3.1.1 Nespojité uniformní rozdělení

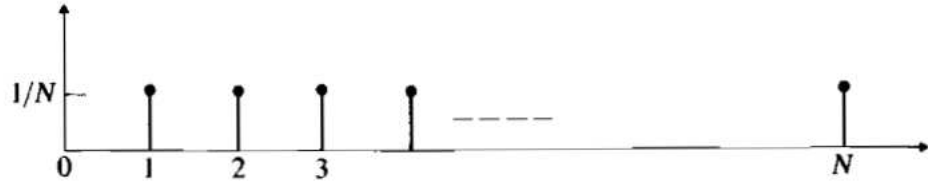
Definice 3.1 (Nespojité uniformní rozdělení) *Jestliže má náhodná veličina X pravděpodobnostní funkci*

$$f(x) = \begin{cases} \frac{1}{N} & \text{pro } x = 1, 2, \dots, N \\ 0 & \text{v ostatních případech} \end{cases}$$
$$= \frac{1}{N} I_{\{1, 2, \dots, N\}}(x)$$

kde N je přirozené číslo, říkáme, že tato náhodná veličina sleduje nespojité uniformní rozdělení.

Věta 3.1 *Jestliže náhodná veličina X sleduje nespojité uniformní rozdělení, pak*

$$E[X] = \frac{N+1}{2}$$
$$D[X] = \frac{N^2-1}{12}$$
$$m(t) = E[e^{tX}] = \sum_{j=1}^N e^{jt} \frac{1}{N}$$



Obrázek 3.1: Pravděpodobnostní funkce nespojitého uniformního rozdělení

Důkaz 3.1

$$\begin{aligned}
 E[X] &= \sum_{j=1}^N j \frac{1}{N} = \frac{N+1}{2} \\
 D[X] &= E[X^2] - E[X]^2 = \sum_{j=1}^N j^2 \frac{1}{N} - \left(\frac{N+1}{2}\right)^2 \\
 &= \frac{N(N+1)(2N+1)}{6N} - \frac{(N+1)^2}{4} = \frac{(N+1)(N-1)}{12} \\
 E[e^{tX}] &= \sum_{j=1}^N e^{jt} \frac{1}{N}
 \end{aligned}$$

Někdy je nespojité diskrétní rozdělení definována jako

$$f(x) = \frac{1}{N+1} I_{\{0,1,\dots,N\}}(x)$$

V těchto případech je třeba výše uvedené rovnice upravit.

3.1.2 Bernoulliho a binomické rozdělení**Bernoulliho rozdělení**

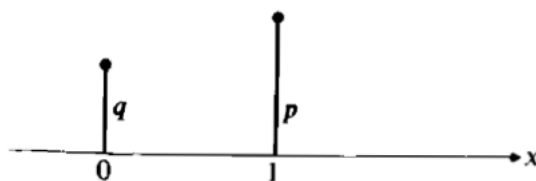
Definice 3.2 (Bernoulliho rozdělení) Říkáme, že náhodná veličina X sleduje Bernoulliho rozdělení, jestliže má její pravděpodobnostní rozdělení tvar

$$f(x) = \begin{cases} p^x(1-p)^{1-x} & \text{pro } x = 0, 1 \\ 0 & \text{v ostatních případech} \end{cases} = p^x(1-p)^{1-x} I_{\{0,1\}}(x)$$

kde $0 \leq p \leq 1$. Parameter $1-p$ se velmi často označuje jako q .

Věta 3.2 Jestliže náhodná veličina X sleduje Bernoulliho rozdělení, pak

$$\begin{aligned}
 E[X] &= p \\
 D[X] &= pq \\
 m(t) &= pe^t + q
 \end{aligned}$$



Obrázek 3.2: Pravděpodobnostní funkce Bernoulliho rozdělení

Důkaz 3.2

$$\begin{aligned}
 E[X] &= 0 \cdot q + 1 \cdot p = p \\
 D[X] &= E[X^2] - E[X]^2 = 0^2 \cdot q + 1^2 \cdot p - p^2 = pq \\
 m(t) &= E[e^{tX}] = e^{0 \cdot t} p^0 q^1 + e^{1 \cdot t} p^1 q^0 = q + e^t p
 \end{aligned}$$

Příklad 3.1 Uvažujme hod hrací kostkou. Úspěch definujme jako situaci, kdy padne číslo šest a neúspěch, jestliže padne číslo jedna až pět. Odpovídající náhodná veličina má povahu Bernoulliho rozdělení s pravděpodobností úspěchu $p = \frac{1}{6}$.

Binomické rozdělení

Definice 3.3 (Binomické rozdělení) Náhodná veličina X sleduje binomické rozdělení, jestliže její pravděpodobnostní funkce je

$$f(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & \text{pro } x = 0, 1, \dots, n \\ 0 & \text{v ostatních případech} \end{cases} = \binom{n}{x} p^x q^{n-x} I_{\{0,1,\dots,n\}}(x)$$

kde $0 \leq p \leq 1$, n je přirozené číslo a $q = 1 - p$.

Věta 3.3 Jestliže náhodná veličina X sleduje binomické rozdělení, pak

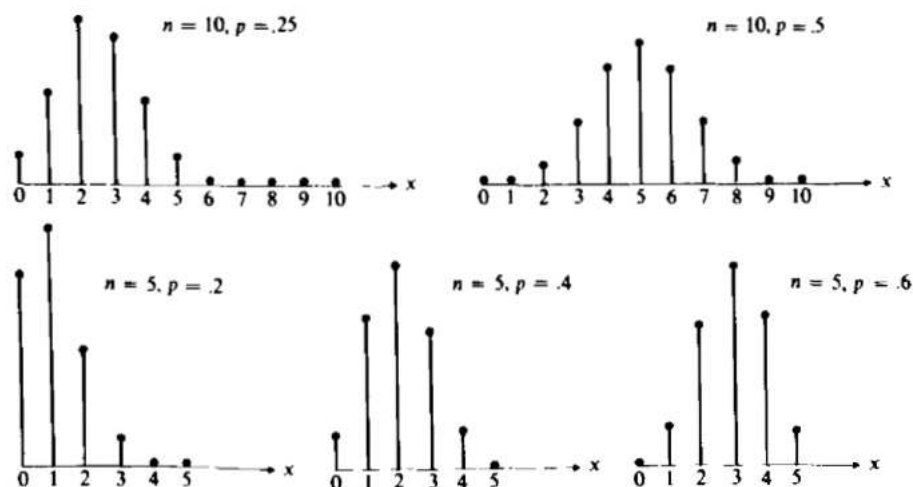
$$\begin{aligned}
 E[X] &= np \\
 D[X] &= npq \\
 m(t) &= (q + pe^t)^n
 \end{aligned}$$

Důkaz 3.3

$$\begin{aligned}
 m(t) &= E[e^{tX}] = \sum_{x=0}^n e^{tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^n \binom{n}{x} (pe^t)^x q^{n-x} = (pe^t + q)^n \\
 m'(t) &= npe^t (pe^t + q)^{n-1} \\
 m''(t) &= n(n-1)(pe^t)^2 (pe^t + q)^{n-2} + npe^t (pe^t + q)^{n-1}
 \end{aligned}$$

a proto

$$\begin{aligned}
 E[X] &= m'(0) = np \\
 D[X] &= E[X^2] - E[X]^2 = m''(0) - (np)^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)
 \end{aligned}$$



Obrázek 3.3: Pravděpodobnostní funkce binomického rozdělení

Je-li $n = 1$ zredukuje se binomické na Bernoulliho rozdělení.

Příklad 3.2 Uvažujme hod pěti hracími kostkami. Jaká je pravděpodobnost, že číslo šest padne právě dvakrát?

Jedním z možných řešení je situace, kdy v prvních třech hodech padlo číslo jedna až pět a číslo šest padlo v posledních dvou hodech. Pravděpodobnost této kombinace je $\left(\frac{5}{6}\right)^3 \left(\frac{1}{6}\right)^2$. Vzhledem k tomu, že však nezáleží na pořadí hodů, existuje $\binom{5}{2}$ takovýchto řešení. Výsledná pravděpodobnost je tak rovna

$$\binom{5}{2} \left(\frac{5}{6}\right)^3 \left(\frac{1}{6}\right)^2 = 0.1608$$

Jak vyplývá z následující věty, pravděpodobnostní funkce binomického rozdělení nejprve monotónně roste a následně monotónně klesá.

Věta 3.4 Necht' náhodná veličina X sleduje binomické rozdělení. Pak

1. $f(x-1) < f(x)$ jestliže $x < (n+1)p$,
2. $f(x-1) > f(x)$ jestliže $x > (n+1)p$ a
3. $f(x-1) = f(x)$ jestliže $x = (n+1)p$ a $(n+1)p$ je celé číslo pro $x = 1, 2, \dots, n$.

Důkaz 3.4

$$\frac{f(x)}{f(x-1)} = \frac{n-x+1}{x} \frac{p}{q} = 1 + \frac{(n+1)p-x}{xq}$$

Výše uvedené je

1. větší než 1, jestliže $x < (n+1)p$,
2. menší než 1, jestliže $x > (n+1)p$ a
3. rovno 1, jestliže $x = (n+1)p$.

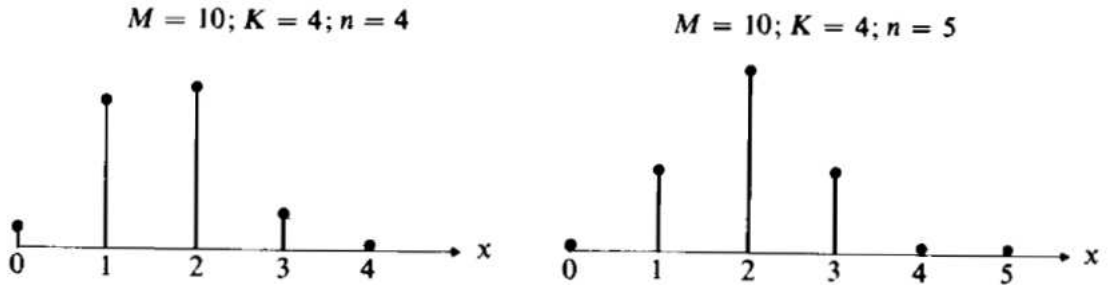
3.1.3 Hypergeometrické rozdělení

Definice 3.4 (Hypergeometrické rozdělení) Jestliže náhodná veličina X sleduje hypergeometrické rozdělení, pak je její pravděpodobnostní funkce definována jako

$$f(x) = \begin{cases} \frac{\binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}} & \text{pro } x = 0, 1, \dots, n \\ 0 & \text{v ostatních případech} \end{cases}$$

$$= \frac{\binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}} I_{\{0,1,\dots,n\}}(x)$$

kde $M > 0$, $0 \leq K \leq M$, $0 < n \leq M$.



Obrázek 3.4: Pravděpodobnostní funkce hypergeometrického rozdělení

Věta 3.5 Jestliže náhodná veličina X sleduje hypergeometrické rozdělení, pak

$$E[X] = n \frac{K}{M}$$

$$D[X] = n \frac{K}{M} \frac{M-K}{M} \frac{M-n}{M-1}$$

Důkaz 3.5

$$E[X] = \sum_{x=0}^n x \frac{\binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}} = n \frac{K}{M} \sum_{x=1}^n \frac{\binom{K-1}{x-1} \binom{M-K}{n-x}}{\binom{M-1}{n-1}}$$

$$= n \frac{K}{M} \sum_{y=0}^{n-1} \frac{\binom{K-1}{y} \binom{M-1-K+1}{n-1-y}}{\binom{M-1}{n-1}} = n \frac{K}{M}$$

s využitím vztahů $\binom{a}{b} = \frac{a}{b} \binom{a-1}{b-1}$ a $\sum_{i=0}^m \binom{a}{i} \binom{b}{m-i} = \binom{a+b}{m}$. Dále platí

$$E[X(X-1)] = \sum_{x=0}^n x(x-1) \frac{\binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}} = n(n-1) \frac{K(K-1)}{M(M-1)} \sum_{x=2}^n \frac{\binom{K-2}{x-2} \binom{M-K}{n-x}}{\frac{\binom{M-2}{n-2}}{n-2}}$$

$$= n(n-1) \frac{K(K-1)}{M(M-1)} \sum_{y=0}^{n-2} \frac{\binom{K-2}{y} \binom{M-2-K+2}{n-2-y}}{\binom{M-2}{n-2}} = n(n-1) \frac{K(K-1)}{M(M-1)}$$

A proto

$$\begin{aligned} D[X] &= E[X^2] - E[X]^2 = E[X(X-1)] + E[X] - E[X]^2 \\ &= n(n-1) \frac{K(K-1)}{M(M-1)} + n \frac{K}{M} - n^2 \frac{K^2}{M^2} = n \frac{K}{M} \left((n-1) \frac{K-1}{M-1} + 1 - \frac{nK}{M} \right) \\ &= \frac{nK}{M} \left(\frac{(M-K)(M-n)}{M(M-1)} \right) \end{aligned}$$

Jestliže dosadíme $\frac{K}{M} = p$, pak se střední hodnota hypergeometrického rozdělení shoduje se střední hodnotou binomického rozdělení a rozptyl hypergeometrického rozdělení je $\frac{M-n}{M-1}$ krát rozptyl binomického rozdělení.

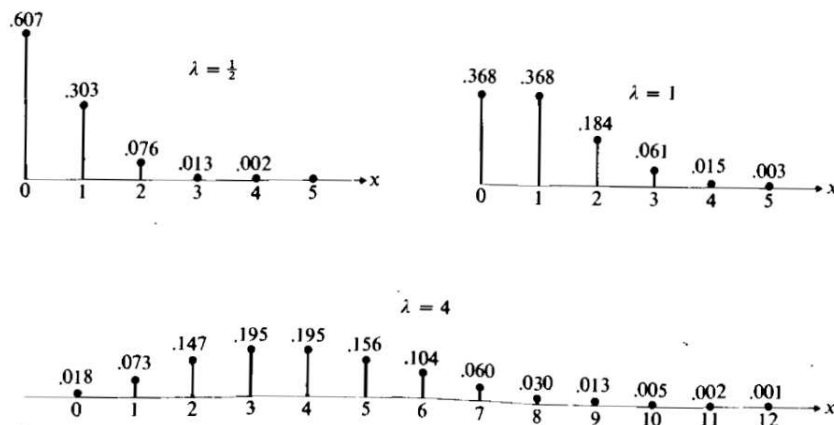
Příklad 3.3 Nechť X označuje počet vadných výrobků ve vzorku velikosti n , který byl získán výběrem bez vracení z populace M výrobků, z nichž celkem K bylo vadných. Takto definovaná náhodná veličina sleduje hypergeometrické rozdělení.

3.1.4 Poissonovo rozdělení

Definice 3.5 (Poissonovo rozdělení) Náhodná veličina X sleduje Poissonovo rozdělení, jestliže její pravděpodobnostní funkce je definována jako

$$\begin{aligned} f(x) &= \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & \text{pro } x = 0, 1, 2, \dots \\ 0 & \text{v ostatních případech} \end{cases} \\ &= \frac{e^{-\lambda} \lambda^x}{x!} I_{\{0,1,2,\dots\}}(x) \end{aligned}$$

kde $\lambda > 0$.



Obrázek 3.5: Pravděpodobnostní funkce Poissonova rozdělení

Věta 3.6 Nechť náhodná veličina X sleduje Poissonovo rozdělení. Pak

$$\begin{aligned} E[X] &= \lambda \\ D[X] &= \lambda \\ m(t) &= e^{\lambda(e^t - 1)} \end{aligned}$$

Důkaz 3.6

$$\begin{aligned}
m(t) &= E[e^{tX}] = \sum_{x=0}^{\infty} \frac{e^{tx} e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} \\
m'(t) &= \lambda e^{-\lambda} e^t e^{\lambda e^t} \\
m''(t) &= \lambda e^{-\lambda} e^t e^{-\lambda e^t} (\lambda e^t + 1)
\end{aligned}$$

Proto

$$\begin{aligned}
E[X] &= m'(0) = \lambda \\
D[X] &= E[X^2] - E[X]^2 = m''(0) - \lambda^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda
\end{aligned}$$

Poissonovo rozdělení je vhodné pro modelování počtu realizací vzájemně nezávislých náhodných událostí vztažených k času, objemu, délce nebo ploše. Příkladem může být počet radioaktivních částic emitovaných za jednotku času, počet mikroorganismů v daném objemu roztoku, množství poškození DNA na jeden nanometer nebo počet kazů na metr čtvereční vyrobené látky. V některých situacích, které jsou na první pohled analogii výše uvedených příkladů, je použití Poissonova rozdělení nevhodné. Příkladem může být počet škůdců na hektar zemědělské půdy, kde není splněn předpoklad vzájemné nezávislosti¹. Klíčové je Poissonovo rozdělení pro pojistnou matematiku neživotního pojištění, kde slouží k modelování počtu pojistných událostí.

Uvažujme $v > 0$, které vyjadřuje pravděpodobnost realizace náhodné události a které splňuje následující tři podmínky.

1. Pravděpodobnost právě jedné realizace v krátkém časovém intervalu délky h je přibližně rovno vh , nebo-li $P[\text{jedna realizace za časový interval } h] = vh + o(h)$.
2. Pravděpodobnost dvou a více realizací v časovém intervalu délky h je zanedbatelné v porovnání s pravděpodobností jedné realizace v témže časovém intervalu, nebo-li $P[\text{dvě a více realizací za časový interval } h] = o(h)$.
3. Počty realizací v nepřekrývajících se časových intervalech jsou vzájemně nezávislé.

$o(h)$ označuje funkci nižšího řádu než h , která splňuje podmínku

$$\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$$

Hodnotu v lze interpretovat jako střední míru realizace za časovou jednotku.

Věta 3.7 *Jestliže jsou výše uvedené tři podmínky splněny, pak počet realizací v čase t má Poissonovo rozdělení s parametrem $\lambda = vt$. Nebo-li jestliže náhodná veličina $Z(t)$ označuje počet realizací v časovém intervalu délky t , pak $P[Z(t) = z] = e^{-vt} \frac{(vt)^z}{z!}$ pro $z = 0, 1, 2, \dots$*

¹Pokud je škůdce nalezen na určitém místě, je pravděpodobné, že se bude nacházet také v jeho okolí. Škůdci se tak na poli nenachází náhodně, ale vytváří ohniska nákazy.

Důkaz 3.7 (První způsob) Uvažujme časový interval $(0, t]$ o délce t a časový interval $(t, t + h]$ o délce h . Definujme

$$P_n(s) = P[Z(s) = n] = P[\text{právě } n \text{ realizací v časovém intervalu délky } s]$$

Pak, na základě podmínky (3) o vzájemné nezávislosti, platí

$$\begin{aligned} P_0(t + h) &= P[0 \text{ realizací v časovém intervalu } (0, t + h)] \\ &= P[0 \text{ realizací v } (0, t] \text{ a } 0 \text{ realizací v } (t, t + h)] \\ &= P[0 \text{ realizací v } (0, t)] P[0 \text{ realizací v } (t, t + h)] \\ &= P_0(t) P_0(h) \end{aligned}$$

Dále platí

$$\begin{aligned} P[0 \text{ realizací v } (t, t + h)] &= 1 - P[\text{jedna nebo více realizací v } (t, t + h)] \\ &= 1 - P[\text{jedna realizace v } (t, t + h)] - P[\text{dvě a více realizací v } (t, t + h)] \\ &= 1 - vh - o(h) - o(h) \end{aligned}$$

Proto $P_0(t + h) = P_0(t)[1 - vh - o(h) - o(h)]$, neboli

$$\frac{P_0(t + h) - P_0(t)}{h} = -vP_0(t) - P_0(t) \frac{o(h) + o(h)}{h}$$

Pro limitu $h \rightarrow 0$ tak získáme

$$P'_0(t) = -vP_0(t)$$

Řešením této diferenciální rovnice je

$$P_0(t) = e^{-vt}$$

při počáteční podmínce $P_0(0) = 1$.

Podobně

$$P_1(t + h) = P_1(t)P_0(h) + P_0(t)P_1(h) = P_1(t)[1 - vh - o(h)] + P_0(t)[vh + o(h)]$$

což vede k diferenciální rovnici

$$P'_1(t) = -vP_1(t) + vP_0(t)$$

s řešením

$$P_1(t) = vte^{-vt}$$

pro počáteční podmínku $P_1(0) = 0$.

Jestliže bychom postupovali dále analogickým způsobem, získali bychom diferenciální rovnice

$$P'_n(t) = -vP_n(t) + vP_{n-1}(t)$$

pro $n = 2, 3, \dots$. Řešením tohoto systému diferenciálních rovnic je

$$P_n(t) = \frac{(vt)^n e^{-vt}}{n!}$$

Důkaz 3.8 (Druhý způsob) Rozdělme interval $(0, t)$ do n časových subintervalů o délce $h = t/n$. Pravděpodobnost, že dojde ke k realizacím v intervalu $(0, t)$, je přibližně rovna pravděpodobnosti právě jedné realizace v k z celkového počtu n subintervalů, přičemž pravděpodobnost takovéto realizace je vh . Na každý subinterval je možné nahlížet jako na Bernoulliho pokus - buď v něm dojde k realizaci nebo nikoliv. Z předpokladů, které jsme učinili, také vyplývá, že tyto Bernoulliho pokusy jsou vzájemně nezávislé. Pravděpodobnost k úspěchů v rámci n vzájemně nezávislých Bernoulliho pokusů je tak

$$\binom{n}{k} (vh)^k (1 - vh)^{n-k} = \binom{n}{k} \left(\frac{vt}{n}\right)^k \left(1 - \frac{vt}{n}\right)^{n-k}$$

což je aproximace požadované pravděpodobnosti k realizací v časovém intervalu $(0, t)$. Přesný vztah lze získat v limitě $n \rightarrow \infty$.

$$\binom{n}{k} \left(\frac{vt}{n}\right)^k \left(1 - \frac{vt}{n}\right)^{n-k} = \frac{1}{k!} (vt)^k \left(1 - \frac{vt}{n}\right)^{n-k} \frac{(n)_k}{n^k} \rightarrow \frac{(vt)^k e^{-vt}}{k!}$$

kde jsme využili vztahů $\lim_{n \rightarrow \infty} \left(1 - \frac{vt}{n}\right)^n = e^{-vt}$, $\lim_{n \rightarrow \infty} \left(1 - \frac{vt}{n}\right)^{-k} = 1$ a $\lim_{n \rightarrow \infty} \frac{(n)_k}{n^k} = 1$.

Příklad 3.4 Předpokládejme, že průměrný počet telefonních hovorů na ústředně menší společnosti je 30 hovorů za hodinu. Jaká je pravděpodobnost, že v průběhu tří minut neuskuteční žádný hovor? Jaká je pravděpodobnost, že v průběhu pěti minut bude skutečně více než pět hovorů?

Předpokládejme, že počet telefonních hovorů lze modelovat pomocí Poissonova rozdělení. Přepočteno na minuty je průměrný počet hovorů 0.5 hovorů za minutu. Pravděpodobnost, že průběhu tří minut nedojde k žádnému telefonnímu hovorů je tak

$$P = \frac{e^{-1.5} 1.5^0}{0!} \approx 0.223$$

Pravděpodobnost, že v průběhu pěti minut dojde k více než pěti hovorům, je pak

$$P = 1 - \sum_{x=0}^5 \frac{e^{-2.5} 2.5^x}{x!} = 0.042$$

Věta 3.8 Uvažujme pravděpodobnostní funkci Poissonova rozdělení $\frac{e^{-\lambda} \lambda^k}{k!}$ pro $k = 0, 1, 2, \dots$. Pak platí

$$\begin{aligned} \frac{e^{-\lambda} \lambda^{k-1}}{(k-1)!} &< \frac{e^{-\lambda} \lambda^k}{k!} && \text{pro } k < \lambda \\ \frac{e^{-\lambda} \lambda^{k-1}}{(k-1)!} &> \frac{e^{-\lambda} \lambda^k}{k!} && \text{pro } k > \lambda \\ \frac{e^{-\lambda} \lambda^{k-1}}{(k-1)!} &= \frac{e^{-\lambda} \lambda^k}{k!} && \text{pro } k = \lambda \end{aligned}$$

Důkaz 3.9

$$\frac{\frac{e^{-\lambda} \lambda^{k-1}}{(k-1)!}}{\frac{e^{-\lambda} \lambda^k}{k!}} = \frac{k}{\lambda}$$

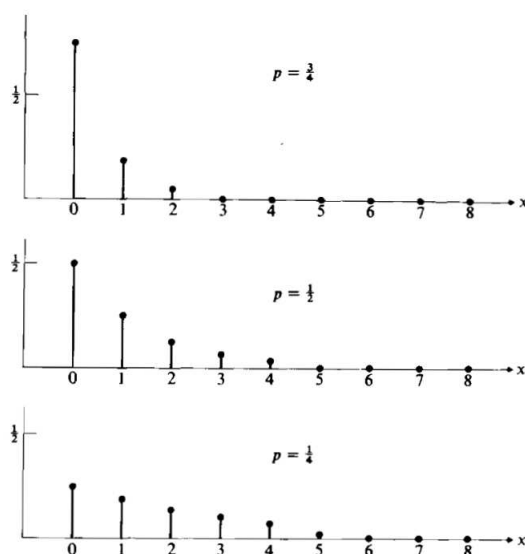
3.1.5 Geometrické a negativní binomické rozdělení

Geometrické rozdělení

Definice 3.6 (Geometrické rozdělení) Jestliže náhodná veličina X sleduje geometrické rozdělení, pak její pravděpodobnostní funkce má tvar

$$f_X(x) = \begin{cases} p(1-p)^x & \text{pro } x = 0, 1, 2, \dots \\ 0 & \text{v ostatních případech} \end{cases}$$

$$= p(1-p)^x I_{\{0,1,2,\dots\}}(x)$$



Obrázek 3.6: Pravděpodobnostní funkce geometrického rozdělení

Věta 3.9 Jestliže náhodná veličina X sleduje geometrické rozdělení, pak

$$E[X] = \frac{q}{p}$$

$$D[X] = \frac{q}{p^2}$$

$$m(t) = \frac{p}{1 - qe^t}$$

Důkaz 3.10 Vzhledem k tomu, že geometrické rozdělení se speciální případ negativního binomického rozdělení, odkazujeme se na příslušný důkaz pro negativní binomické rozdělení.

Věta 3.10 Jestliže náhodná veličina X sleduje geometrické rozdělení, pak

$$P[X \geq i + j | X \geq i] = P[X \geq j] \quad \text{pro } i, j = 0, 1, 2, \dots$$

Důkaz 3.11

$$\begin{aligned}
P[X \geq i+j | X \geq i] &= \frac{P[X \geq i+1]}{P[X \geq i]} = \frac{\sum_{x=i+j}^{\infty} p(1-p)^x}{\sum_{x=i}^{\infty} p(1-p)^x} \\
&= \frac{(1-p)^{i+j}}{(1-p)^i} = (1-p)^j = P[X \geq j]
\end{aligned}$$

Příklad 3.5 Uvažujme sérii vzájemně nezávislých Bernoulliho pokusů s pravděpodobností úspěchu p . Nechť náhodná veličina X představuje počet neúspěšných pokusů, které předcházejí prvnímu úspěšnému pokusu. Pak sleduje náhodná veličina X geometrické rozdělení.

Negativní binomické rozdělení

Definice 3.7 (Negativní binomické rozdělení) Jestliže náhodná veličina X sleduje negativní binomické rozdělení, pak její pravděpodobnostní funkce má tvar

$$\begin{aligned}
f_X(x) &= \begin{cases} \binom{r+x-1}{x} p^r q^x = \binom{-r}{x} p^r (-q)^r & \text{pro } x = 0, 1, 2, \dots \\ 0 & \text{v ostatních případech} \end{cases} \\
&= \binom{r+x-1}{x} p^r q^x I_{\{0,1,2,\dots\}}(x)
\end{aligned}$$

kde $r = 1, 2, 3, \dots$, $0 < p \leq 1$ a $q = 1 - p$.

Jestliže $r = 1$ pak se negativní binomické rozdělení zredukuje na geometrické rozdělení.

Věta 3.11 Nechť má náhodná veličina X negativní binomické rozdělení. Pak

$$\begin{aligned}
E[X] &= \frac{rq}{p} \\
D[X] &= \frac{rq}{p^2} \\
m(t) &= \left[\frac{p}{1 - qe^t} \right]^r
\end{aligned}$$

Důkaz 3.12

$$\begin{aligned}
m(t) &= E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} \binom{-r}{x} p^r (-q)^x \\
&= \sum_{x=0}^{\infty} \binom{-r}{x} p^r (-qe^t)^x = \left[\frac{p}{1 - qe^t} \right]^r
\end{aligned}$$

kde jsme využili vztah

$$(1-x)^{-n} = \sum_{j=0}^{\infty} \binom{-n}{j} (-x)^j = \sum_{j=0}^{\infty} \binom{n+j-1}{j} x^j \quad \text{pro } -1 < x < 1$$

Následnou derivaci pak získáme

$$\begin{aligned}
m'(t) &= p^r (-r)(1 - qe^t)^{-r-1} (-qe^t) \\
m''(t) &= rqp^r [q(r+1)e^{2t}(1 - qe^t)^{-r-2} + e^t(1 - qe^t)^{-r-1}]
\end{aligned}$$

a proto

$$\begin{aligned} E[X] &= m'(t) \Big|_{t=0} = \frac{rq}{p} \\ D[X] &= m''(t) \Big|_{t=0} - E[X]^2 = rqp^r [qp^{-r-2}(r+1) + p^{-r-1}] - \left(\frac{rq}{p}\right)^2 \\ &= \frac{rq^2}{p^2} + \frac{rq}{p} = \frac{rq}{p^2} \end{aligned}$$

Příklad 3.6 Uvažujme sérii vzájemně nezávislých Bernoulliho pokusů s pravděpodobností úspěchu p . Nechť náhodná veličina X představuje počet neúspěšných pokusů, které předcházely r -tému úspěšnému pokusu. V tomto případě sleduje náhodná veličina X negativní binomické rozdělení.

Odvození pravděpodobnostní funkce je následující. Poslední pokus musí skončit úspěchem. Mezi zbývajících $x + r - 1$ pokusy musí být x neúspěchů a $r - 1$ úspěchů, přičemž na jejich pořadí nezáleží. Výsledná pravděpodobnost je tak rovna

$$\binom{x+r-1}{r-1} p^{r-1} q^x = \binom{r+x-1}{x} p^{r-1} q^x$$

3.1.6 Ostatní nespojitá rozdělení

Odvození ze stávajících rozdělení

Jedním ze způsobů, jak vytvořit nová nespojitá rozdělení z výše uvedených, je pomocí “seříznutí”. Pro ilustraci uvažujme Poissonovo rozdělení “seříznuté” v bodě 0. V některých situacích nemůže náhodná veličina nabývat hodnoty nula, avšak přesto se Poissonovo rozdělení zdá být vhodným modelem. Odpovídající pravděpodobnostní funkce pak má tvar

$$f(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} (1 - e^{-\lambda}) & \text{pro } x = 1, 2, 3, \dots \\ 0 & \text{v ostatních případech} \end{cases}$$

O takto odvozených pravděpodobnostních rozděleních hovoříme jako o zredukováných pravděpodobnostních rozděleních.

Další způsob odvození nového typu nespojitého rozdělení opět ilustrujme na Poissonově rozdělení. Předpokládejme, že náhodná veličina X sleduje Poissonovo rozdělení. Dále předpokládejme, že v rámci experimentu jsme, z určitých blíže nespécifikovaných důvodů, schopni rozlišit pouze žádnou, jednu nebo více než jednu realizaci. V tomto případě má pravděpodobnostní funkce tvar

$$f(x) = \begin{cases} e^{-\lambda} & \text{pro } x = 0 \\ \lambda e^{-\lambda} & \text{pro } x = 1 \\ 1 - e^{-\lambda} - \lambda e^{-\lambda} & \end{cases}$$

O náhodné veličině X pak hovoříme jako o censorované náhodné veličině.

Beta-binomické rozdělení

Definice 3.8 (Beta-binomické rozdělení) Jestliže náhodná veličina X sleduje beta-binomické rozdělení, pak její pravděpodobnostní funkce má tvar

$$f(x) = \binom{n}{x} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(x + \alpha)\Gamma(n + \beta - x)}{\Gamma(n + \alpha + \beta)} I_{\{0,1,\dots,n\}}(x)$$

kde n je přirozené číslo větší než nula, $\alpha > 0$, $\beta > 0$ a $\Gamma(m)$ kde je tzv. gamma funkce

$$\Gamma(m) = \int_0^\infty x^{m-1} e^{-x} dx, \quad m > 0$$

Pro tuto náhodnou veličinu pak platí

$$E[X] = \frac{n\alpha}{\alpha + \beta}$$

$$D[X] = \frac{n\alpha\beta(n + \alpha + \beta)}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Jestliže $\alpha = \beta = 1$, pak se beta-binomické rozdělení zredukuje na nespojitě uniformní rozdělení.

Definice 3.9 (Logaritmické rozdělení) Jestliže náhodná veličina X sleduje logaritmické rozdělení, pak její pravděpodobnostní funkce má tvar

$$f(x) = \begin{cases} \frac{q^x}{-x \ln(p)} & \text{pro } x = 1, 2, \dots \\ 0 & \text{v ostatních případech} \end{cases}$$

$$= \frac{q^x}{-x \ln(p)} I_{\{1, 2, \dots\}}(x)$$

Pro tuto náhodnou veličinu pak platí

$$E[X] = \frac{q}{-p \ln(p)}$$

$$D[X] = \frac{q(q + \ln(p))}{-(p \ln(p))^2}$$

Označení “logaritmické rozdělení” vychází ze skutečnosti, že odpovídající pravděpodobnostní funkce je rozvojem $\ln(1 - q)$. Logaritmické rozdělení lze odvodit jako limitní rozdělení negativního binomického rozdělení zobecněného pomocí parametru r , který představuje libovolné kladné číslo (nejen celé kladné číslo), a které je seřiznuté v bodě 0.

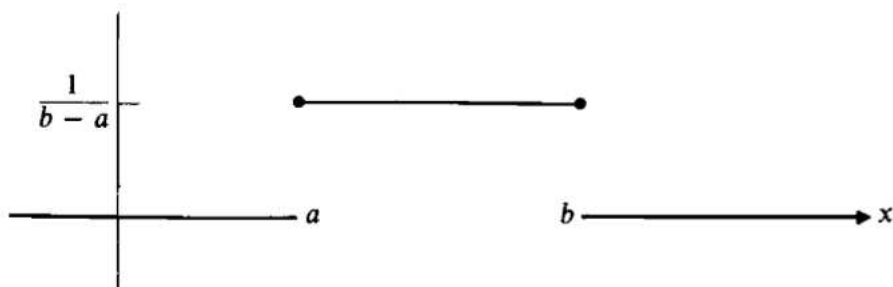
3.2 Spojité pravděpodobnostní rozdělení

3.2.1 Spojité uniformní rozdělení

Definice 3.10 (Spojitě uniformní rozdělení) Jestliže náhodná veličina X sleduje spojitě uniformní rozdělení, pak její pravděpodobnostní funkce má tvar

$$f(x) = \frac{1}{b - a} I_{[a, b]}(x)$$

kde $-\infty < a < b < \infty$.



Obrázek 3.7: Pravděpodobnostní funkce spojitého uniformního rozdělení

Věta 3.12 *Jestliže má náhodná veličina X spojité uniformní rozdělení, pak*

$$\begin{aligned} E[X] &= \frac{a+b}{2} \\ D[X] &= \frac{(b-a)^2}{12} \\ m(t) &= \frac{e^{bt} - e^{at}}{(b-a)t} \end{aligned}$$

Důkaz 3.13

$$\begin{aligned} E[X] &= \int_a^b x \frac{1}{b-a} dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} \\ D[X] &= E[X^2] - E[X]^2 = \int_a^b x^2 \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12} \\ m(t) &= E[e^{tX}] = \int_a^b e^{tx} \frac{1}{b-a} dx = \frac{e^{bt} - e^{at}}{(b-a)t} \end{aligned}$$

Věta 3.13 *Kumulativní distribuční funkce spojitého uniformního rozdělení je*

$$F(x) = \frac{x-a}{b-a} I_{[a,b]}(x) + I_{(b,\infty)}(x)$$

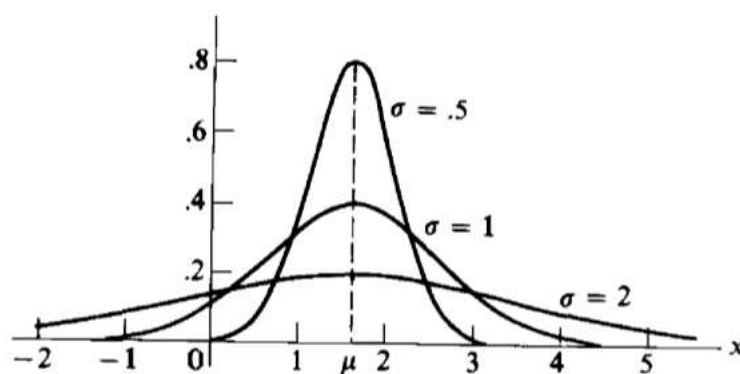
Ačkoliv jsme definovali uniformní rozdělení nad uzavřeným intervalem $[a, b]$, lze toto rozdělení definovat nad intervalem (a, b) , $[a, b)$ nebo $(a, b]$. Pravděpodobnostní funkce bychom modifikovali do podoby $f(x) = \frac{1}{b-a} I_{(a,b)}(x)$, $f(x) = \frac{1}{b-a} I_{[a,b)}(x)$ resp. $f(x) = \frac{1}{b-a} I_{(a,b]}(x)$. Kumulativní distribuční funkce je však ve všech případech shodná.

3.2.2 Normální rozdělení

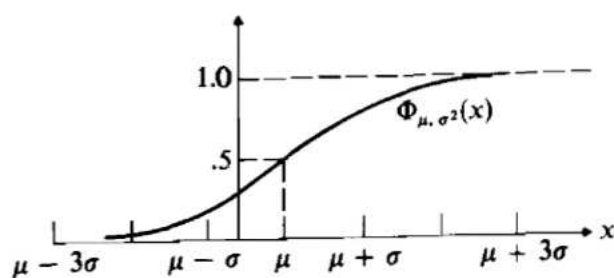
Definice 3.11 (Normální rozdělení) *Jestliže náhodná veličina X sleduje normální rozdělení, pak má její pravděpodobnostní funkce tvar*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

kde $-\infty < \mu < \infty$ a $\sigma > 0$. Jak ukážeme později, tyto parametry představuje střední hodnotu a směrodatnou odchylku tohoto pravděpodobnostního rozdělení.



Obrázek 3.8: Pravděpodobnostní funkce normálního rozdělení



Obrázek 3.9: Kumulativní distribuční funkce normálního rozdělení

Vzhledem k tomu, že normální rozdělení je velmi často používané, je pro jeho pravděpodobnostní funkci resp. kumulativní distribuční funkci používáno označení $\phi(\mu, \sigma^2)$ resp. $\Phi(\mu, \sigma^2)$. Jestliže $\mu = 0$ a $\sigma = 1$, hovoříme o tzv. standardizovaném popř. normalizovaném normálním rozdělení. V tomto případě se obvykle parametry v značení pravděpodobnostní a kumulativní distribuční funkce vypouštějí.

Protože je $\phi(\mu, \sigma^2)$ pravděpodobnostní funkcí, musí platit

$$\int_{-\infty}^{\infty} \phi(\mu, \sigma^2)(x) dx = 1$$

Důkaz 3.14 Uvažujme neurčitý integrál

$$A = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

který lze s využitím substituce $y = \frac{x-\mu}{\sigma}$ upravit do tvaru

$$A = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy$$

Potřebuje dokázat $A = 1$. Vzhledem k tomu, že e^t je kladné pro všechna t , platí $A > 0$. Pokud bychom dokázali $A^2 = 1$, pak by to implikovalo $A = 1$. A^2 lze vyjádřit jako

$$\begin{aligned} A^2 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y^2+z^2)} dy dz \end{aligned}$$

Následně tento integrál převedeme pomocí substitucí $y = r \sin(\theta)$ a $z = r \cos(\theta)$ do polárních souřadnic.

$$A^2 = \frac{1}{2\pi} \int_0^{\infty} \int_0^{2\pi} r e^{-\frac{1}{2}r^2} d\theta dr = \int_0^{\infty} r e^{-\frac{1}{2}r^2} dr = 1$$

Věta 3.14 Jestliže náhodná veličina X sleduje normální rozdělení, pak

$$\begin{aligned} E[X] &= \mu \\ D[X] &= \sigma^2 \\ m(t) &= e^{\mu + \sigma^2 t^2 / 2} \end{aligned}$$

Důkaz 3.15

$$\begin{aligned} m(t) &= E[e^{tX}] = e^{t\mu} E[e^{t(X-\mu)}] = e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{t(x-\mu)} e^{-(1/2\sigma^2)(x-\mu)^2} dx \\ &= e^{t\mu} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(1/2\sigma^2)[(x-\mu)^2 - 2\sigma^2 t(x-\mu)]} dx \end{aligned}$$

Jestliže doplníme kvadráty v závorkách

$$\begin{aligned} (x-\mu)^2 - 2\sigma^2 t(x-\mu) &= (x-\mu)^2 - 2\sigma^2 t(x-\mu) + \sigma^4 t^2 - \sigma^4 t^2 \\ &= (x-\mu - \sigma^2 t)^2 - \sigma^4 t^2 \end{aligned}$$

získáme

$$m(t) = e^{t\mu} e^{\sigma^2 t^2 / 2} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(x-\mu-\sigma^2 t)^2 / 2\sigma^2} dx$$

Integrál společně s členem $\frac{1}{\sqrt{2\pi}\sigma}$ musí být roven 1, protože se jedná o plochu pod pravděpodobnostní funkcí normálního rozdělení se střední hodnotou $\mu + \sigma^2 t$ a rozptylem σ^2 , a proto

$$m(t) = e^{\mu t + \sigma^2 t^2 / 2}$$

Derivací $m(t)$ pro $t = 0$ získáme

$$\begin{aligned} E[X] &= m'(0) = \mu \\ D[X] &= E[X^2] - E[X]^2 = m''(0) - \mu^2 = \sigma^2 \end{aligned}$$

Věta 3.15 Protože neurčitý integrál $\sigma_{\mu, \sigma^2}(x)$ nelze řešit analyticky, jsme schopni kumulativní distribuční funkci prezentovat pouze ve tvaru

$$\Phi_{\mu, \sigma^2}(x) = \int_{-\infty}^x \phi_{\mu, \sigma^2}(u) du$$

Věta 3.16 *Jestliže má náhodná veličina X normální rozdělení, pak*

$$P[a < X < b] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Důkaz 3.16

$$\begin{aligned} P[a < X < b] &= \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}[(x-\mu)/\sigma]^2} dx \\ &= \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

Věta 3.17 *Velmi užitečným vztahem je*

$$\Phi(x) = 1 - \Phi(-x)$$

což vyplývá ze symetričnosti normálního rozdělení.

Příklad 3.7 *Předpokládejme, že průměr strojově vyráběné kovové tyče má normální rozdělení se střední hodnotou 10 cm a směrodatnou odchylkou 0.1 cm. Aby mohla být tyč použita v navazující výrobě, musí být její průměr v rozmezí 9.9 až 10.2 cm. Jaké procento vyrobených tyčí splňuje tuto podmínku?*

$$\begin{aligned} P[9.9 < X < 10.2] &= \Phi\left(\frac{10.2-10}{0.1}\right) - \Phi\left(\frac{9.9-10}{0.1}\right) \\ &= \Phi(2) - \Phi(-1) \approx 0.9772 - 0.1587 = 0.8185 \end{aligned}$$

3.2.3 Exponenciální a gamma rozdělení

Exponenciální rozdělení

Definice 3.12 (Exponenciální rozdělení) *Jestliže náhodná veličina X sleduje exponenciální rozdělení, pak má její pravděpodobnostní funkce tvar*

$$f(x) = \lambda e^{-\lambda x} I_{[0,\infty)}(x)$$

kde $\lambda > 0$.

Věta 3.18 *Jestliže náhodná veličina X sleduje exponenciální rozdělení, pak*

$$\begin{aligned} E[X] &= \frac{1}{\lambda} \\ D[X] &= \frac{1}{\lambda^2} \\ m(t) &= \frac{\lambda}{\lambda - t} \end{aligned}$$

Důkaz 3.17 *Vzhledem k tomu, že exponenciální rozdělení je specifickým případem gamma rozdělení, odkazujeme se na příslušný důkaz pro toto rozdělení.*

Jestliže lze v rámci určitého experimentu modelovat počet realizací za časovou jednotku pomocí Poissonova rozdělení, pak lze délku trvání jednotlivých realizací modelovat pomocí exponenciálního rozdělení.

Věta 3.19 *Jestliže náhodná veličina X sleduje exponenciální rozdělení, pak*

$$P[x > a + b | X > a] = P[X > b] \quad \text{pro } a > 0 \text{ a } b > 0$$

Důkaz 3.18

$$P[X > a + b | X > a] = \frac{P[X > a + b]}{P[X > a]} = \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = e^{-\lambda b} = P[X > b]$$

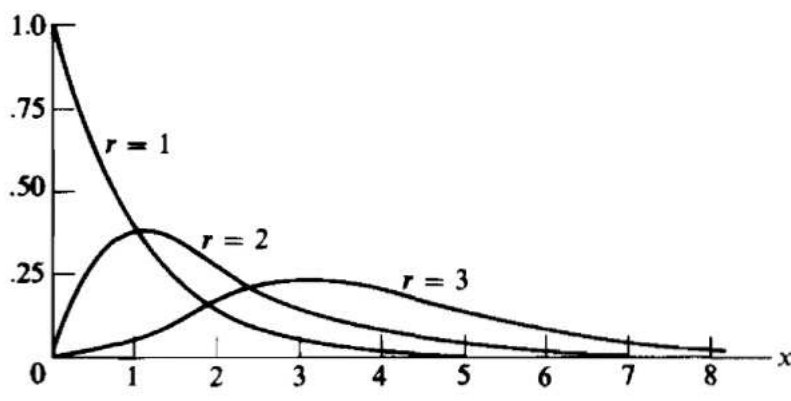
Gamma distribuce

Definice 3.13 (Gamma rozdělení) *Jestliže náhodná veličina X sleduje gamma rozdělení, pak má její pravděpodobnostní funkce tvar*

$$f(x) = \frac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} e^{-\lambda x} I_{[0, \infty)}(x)$$

kde $r > 0$, $\lambda > 0$ a $\Gamma(\cdot)$ je tzv. gamma funkce.

Jestliže je parametr r roven jedné, pak se gamma distribuce zredukuje na exponenciální rozdělení.



Obrázek 3.10: Pravděpodobnostní funkce gamma rozdělení

Věta 3.20 *Jestliže náhodná veličina X sleduje gamma rozdělení, pak*

$$\begin{aligned} E[X] &= \frac{r}{\lambda} \\ D[X] &= \frac{r}{\lambda^2} \\ m(t) &= \left(\frac{\lambda}{\lambda - t} \right)^r \quad \text{pro } t < \lambda \end{aligned}$$

Důkaz 3.19

$$\begin{aligned}
m(t) &= E[e^{tX}] = \int_0^\infty \frac{\lambda^r}{\Gamma(r)} e^{tx} x^{r-1} e^{-\lambda x} dx \\
&= \left(\frac{\lambda}{\lambda-t}\right)^r \int_0^\infty \frac{(\lambda-t)^r}{\Gamma(r)} x^{r-1} e^{-(\lambda-t)x} dx = \left(\frac{\lambda}{\lambda-t}\right)^r
\end{aligned}$$

První a druhá derivace momentové funkce jsou pak

$$\begin{aligned}
m'(t) &= r\lambda^r(\lambda-t)^{-r-1} \\
m''(t) &= r(r+1)\lambda^r(\lambda-t)^{-r-2}
\end{aligned}$$

a proto

$$\begin{aligned}
E[X] &= m'(0) = \frac{r}{\lambda} \\
D[X] &= E[X^2] - E[X]^2 = m''(0) - \left(\frac{r}{\lambda}\right)^2 = \frac{r(r+1)}{\lambda^2} - \left(\frac{r}{\lambda}\right)^2 = \frac{r}{\lambda^2}
\end{aligned}$$

Jestliže lze počet realizací v rámci určitého experimentu modelovat pomocí Poissonova rozdělení, pak lze délku časového úseku od okamžiku 0 do okamžiku r -té realizace modelovat pomocí gamma distribuce. Jedná se tedy o čas, který musíme čekat, než dojde k r -té realizaci.

Věta 3.21 *Jestliže náhodná veličina X sleduje gamma rozdělení, kde parametr r je kladné celé číslo, pak*

$$F(x) = 1 - \sum_{j=0}^{r-1} \frac{e^{-\lambda x} (\lambda x)^j}{j!}$$

Důkaz 3.20 *Výše uvedenou větu lze dokázat postupným integrováním per partes.*

3.2.4 Beta rozdělení

Definice 3.14 (Beta rozdělení) *Jestliže náhodná veličina X sleduje beta rozdělení, pak její pravděpodobnostní funkce má tvar*

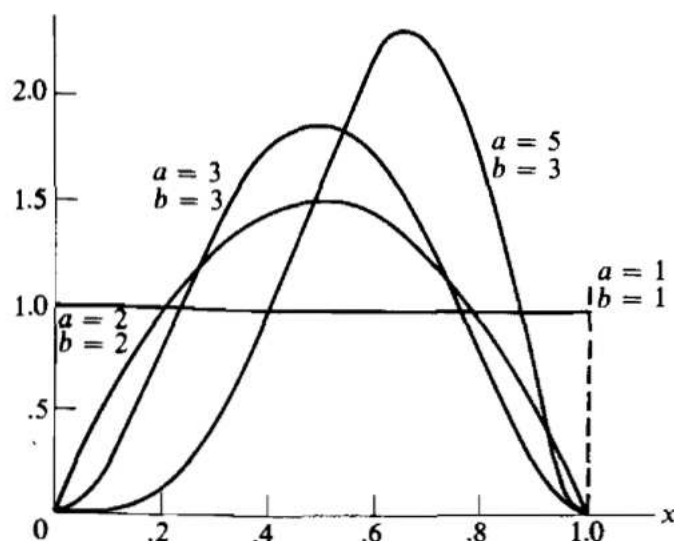
$$f(x) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} I_{(0,1)}(x)$$

kde $a > 0$, $b > 0$ a $B(a,b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$ je tzv. beta funkce.

Pro $a = b$ se beta rozdělení zredukuje na spojité uniformní rozdělení nad intervalem $(0,1)$.

Věta 3.22 *Kumulativní distribuční funkce náhodné veličiny X s beta rozdělením má tvar*

$$F(x) = I_{(0,1)}(x) \int_0^x \frac{1}{B(a,b)} u^{a-1} (1-u)^{b-1} du + I_{[1,\infty)}(x)$$



Obrázek 3.11: Pravděpodobnostní funkce beta rozdělení

Momentová funkce beta rozdělení nemá jednoduchou analytickou formu, nicméně momenty lze snadno odvodit přímo z jejich definice, tj. bez použití momentové funkce.

Věta 3.23

$$E[X] = \frac{a}{a+b}$$

$$D[X] = \frac{ab}{(a+b+1)(a+b)^2}$$

Důkaz 3.21

$$\begin{aligned} E[X^k] &= \frac{1}{B(a,b)} \int_0^1 x^{k+a-1} (1-x)^{b-1} dx \\ &= \frac{B(k+a,b)}{B(a,b)} = \frac{\Gamma(k+a)\Gamma(b)}{\Gamma(k+a+b)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\ &= \frac{\Gamma(k+a)\Gamma(a+b)}{\Gamma(a)\Gamma(k+a+b)} \end{aligned}$$

Proto

$$\begin{aligned} E[X] &= \frac{\Gamma(a+1)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+1)} = \frac{a}{a+b} \\ D[X] &= E[X^2] - E[X]^2 = \frac{\Gamma(a+2)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+2)} - \left(\frac{a}{a+b}\right)^2 \\ &= \frac{(a+1)a}{(a+b+1)(a+b)} - \left(\frac{a}{a+b}\right)^2 = \frac{ab}{(a+b+1)(a+b)^2} \end{aligned}$$

Beta rozdělení nabývá na intervalu $(0, 1)$ kladných hodnot a její pravděpodobnostní funkce je charakteristická velkým množstvím tvarů.

3.3 Ostatní spojitá rozdělení

V této kapitole budou představena některá další spojitá rozdělení. Studentovo, chi-kvadrát a F rozdělení budou představena později.

3.3.1 Cauchyho rozdělení

Definice 3.15 (Cauchyho rozdělení) *Jestliže náhodná veličina X sleduje Cauchyho rozdělení, pak její pravděpodobnostní funkce má tvar*

$$f(x) = \frac{1}{\pi\beta\left(1 + \left(\frac{x-a}{\beta}\right)^2\right)}$$

kde $-\infty < \alpha < \infty$ a $\beta > 0$.

Cauchyho rozdělení je symetrické okolo parametru α , avšak střední hodnota a vyšší momenty nejsou definovány.

Věta 3.24 *Kumulativní distribuční funkce Cauchyho rozdělení má tvar*

$$\begin{aligned} F(x) &= \frac{1}{\pi} \int_{-\infty}^x \frac{1}{\pi\beta\left(1 + \left(\frac{u-a}{\beta}\right)^2\right)} du \\ &= \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-\alpha}{\beta}\right) \end{aligned}$$

3.3.2 Lognormální rozdělení

Definice 3.16 (Lognormální rozdělení) *Nechť je X kladná náhodná veličina a definujeme novou náhodnou veličinu $Y = \ln(X)$. Sleduje-li Y normální rozdělení, pak sleduje X lognormální rozdělení. Pravděpodobnostní funkce lognormálního rozdělení má tvar*

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(\ln(x)-\mu)^2} I_{(0,\infty)}(x)$$

kde $-\infty < \mu < \infty$ a $\sigma > 0$.

Věta 3.25 *Jestliže náhodná veličina X sleduje lognormální rozdělení, pak*

$$\begin{aligned} E[X] &= e^{\mu + \frac{1}{2}\sigma^2} \\ D[X] &= e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2} \end{aligned}$$

Jestliže má X lognormální rozdělení, pak $E[\ln(X)] = \mu$ a $D[\ln(X)] = \sigma^2$.

3.3.3 Laplaceova distribuce

Definice 3.17 (Laplaceova distribuce) *Jestliže náhodná veličina X sleduje Laplaceovo rozdělení, pak má její pravděpodobnostní funkce tvar*

$$f(x) = \frac{1}{2\beta} e^{-\frac{|x-\alpha|}{\beta}}$$

kde $-\infty < \alpha < \infty$ a $\beta > 0$.

Věta 3.26 *Jestliže náhodná veličina X sleduje Laplaceovo rozdělení, pak*

$$\begin{aligned} E[X] &= \alpha \\ D[X] &= 2\beta^2 \end{aligned}$$

3.3.4 Weibullovo rozdělení

Definice 3.18 *Jestliže náhodná veličina X sleduje Weibullovo rozdělení, pak její pravděpodobnostní funkce má tvar*

$$f(x) = abx^{b-1}e^{-ax^b}I_{(0,\infty)}(x)$$

kde $a > 0$ a $b > 0$.

Věta 3.27 *Jestliže náhodná veličina X sleduje Weibullovo rozdělení, pak*

$$\begin{aligned} E[X] &= \frac{1}{a} \frac{1}{b} \Gamma(1 + b^{-1}) \\ D[X] &= \frac{1}{a} \frac{2}{b} (\Gamma(1 + 2b^{-1}) - \Gamma^2(1 + b^{-1})) \end{aligned}$$

Pro $b = 1$ se Weibullovo rozdělení zredukuje na exponenciální rozdělení.

3.3.5 Logistické rozdělení

Definice 3.19 (Logistické rozdělení) *Jestliže náhodná veličina X sleduje logistické rozdělení, pak má její kumulativní distribuční funkce tvar*

$$F(x) = \frac{1}{1 + e^{-\frac{x-\alpha}{\beta}}}$$

kde $-\infty < \alpha < \infty$ a $\beta > 0$.

Věta 3.28 *Jestliže náhodná veličina X sleduje logistické rozdělení, pak*

$$\begin{aligned} E[X] &= \alpha \\ D[X] &= \frac{\beta^2 \pi^2}{3} \end{aligned}$$

Pro kumulativní distribuční funkci logistického rozdělení platí $F(\alpha - d) = 1 - F(\alpha + d)$, a pravděpodobnostní funkce je tak symetrická kolem parametru α .

3.3.6 Paretovo rozdělení

Definice 3.20 (Pareto rozdělení) *Jestliže náhodná veličina X sleduje Paretovo rozdělení, pak má její pravděpodobnostní funkce tvar*

$$f(x) = \frac{\theta}{x_0} \left(\frac{x_0}{x}\right)^{\theta+1} I_{(x_0,\infty)}(x)$$

kde $\theta > 0$ a $x_0 > 0$.

Věta 3.29 *Jestliže náhodná veličina X sleduje Paretovo rozdělení, pak*

$$E[X] = \frac{\theta x_0}{\theta - 1} \quad \text{pro } \theta > 0$$

$$D[X] = \frac{\theta x_0^2}{\theta - 2} - \left(\frac{\theta x_0}{\theta - 1} \right)^2 \quad \text{pro } \theta > 2$$

3.4 Poznámky

3.4.1 Aproximace

Ačkoliv existuje celá řada aproximací jednoho rozdělení pomocí jiného, budeme se v této kapitole zabývat pouze třemi z nich. Některé další aproximace představíme v kapitole o Centrální limitní větě.

Aproximace binomického rozdělení Poissonovým rozdělením

Pravděpodobnostní funkce binomického rozdělení má tvar

$$\binom{n}{x} p^x (1-p)^{n-x} \quad \text{pro } x = 0, 1, \dots, n$$

Jestliže se n blíží nekonečnu a p se blíží nule, přičemž $np = \lambda$ zůstává konstantní, pak

$$\binom{n}{x} p^x (1-p)^{n-x} \rightarrow \frac{e^{-\lambda} \lambda^x}{x!}$$

pro fixní celé číslo x . Výše uvedený vztah vyplývá z

$$\begin{aligned} \binom{n}{x} p^x (1-p)^{n-x} &= \frac{(n)_x}{x!} \left(\frac{\lambda}{n} \right)^x \left(1 - \frac{\lambda}{n} \right)^{n-x} \\ &= \frac{\lambda^x}{x!} \frac{(n)_x}{n^x} \left(1 - \frac{\lambda}{n} \right)^n \left(1 - \frac{\lambda}{n} \right)^{-x} \rightarrow \frac{e^{-\lambda} \lambda^x}{x!} \end{aligned}$$

protože

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{(n)_x}{n^x} &= 1 \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n} \right)^{-x} &= 1 \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n} \right)^n &= e^{-\lambda} \end{aligned}$$

Proto pro dostatečně velká n a malá p lze tedy binomickou pravděpodobnost $\binom{n}{x} p^x (1-p)^{n-x}$ aproximovat pomocí Poissonovy pravděpodobnosti $\frac{e^{-\lambda} \lambda^x}{x!}$.

Aproximace binomického a Poissonova rozdělení normálním rozdělením

Věta 3.30 *Nechť náhodná veličina X sleduje Poissonovo rozdělení. Pak pro fixní $a < b$ platí*

$$P\left[a < \frac{X - \lambda}{\sqrt{\lambda}}\right] = P[\lambda + a\sqrt{\lambda} < X < \lambda + b\sqrt{\lambda}] \rightarrow \Phi(b) - \Phi(a) \quad \text{pro } \lambda \rightarrow \infty$$

Věta 3.31 (De Moivre-Laplaceova limitní věta) *Nechť náhodná veličina X sleduje binomické rozdělení. Pak pro fixní $a < b$ platí*

$$P\left[a \leq \frac{X - np}{\sqrt{npq}} \leq b\right] = P[np + a\sqrt{npq} \leq X \leq np + b\sqrt{npq}] \rightarrow \Phi b - \Phi(a) \quad \text{pro } n \rightarrow \infty$$

Příklad 3.8 *Uvažujme hrací kostku, která je vržena 600 krát. Nechť náhodná veličina X představuje počet hodů, při kterých padla šestka. Je zřejmé, že X sleduje binomické rozdělení s $n = 600$ a $p = \frac{1}{6}$ a že $E[X] = 100$. Nalezněte pravděpodobnost $P[90 \leq X \leq 110]$.*

Platí

$$P[90 \leq X \leq 110] = \sum_{j=90}^{110} \binom{600}{j} \left(\frac{1}{6}\right)^j \left(\frac{5}{6}\right)^{600-j}$$

S využitím aproximace se lze vyhnout rutinnímu výpočtu.

$$\begin{aligned} P[90 \leq X \leq 110] &\approx \Phi\left(\frac{110 - 100}{\sqrt{\frac{500}{6}}}\right) - \Phi\left(\frac{90 - 100}{\sqrt{\frac{500}{6}}}\right) \\ &= \Phi\left(\sqrt{\frac{6}{5}}\right) - \Phi\left(-\sqrt{\frac{6}{5}}\right) \approx \Phi(1.095) - \Phi(-1.095) \approx 0.726 \end{aligned}$$

3.4.2 Vztah mezi Poissonovým a exponenciálním rozdělením

V předchozím textu jsme ukázali, že Poissonovo rozdělení je vhodné pro modelování počtu vzájemně nezávislých realizací v čase. Předpokládejme, že k jedné takovéto realizaci došlo, a že předmětem našeho zájmu je délka, po kterou budeme muset čekat, než dojde k další realizaci. Pravděpodobnost

$$P[X > x] = P[\text{žádná realizace v časovém intervalu délky } x] = e^{-vx}$$

kde v je střední míra realizace. Proto

$$F(x) = P[X \leq x] = 1 - P[X > x] = 1 - e^{-vx} \quad \text{pro } x > 0$$

což znamená, že X má exponenciální rozdělení. Tento vztah platí také naopak. Lze dokázat, že lze-li dobu mezi vzájemně nezávislými realizacemi modelovat pomocí exponenciálního rozdělení, pak lze počet realizací modelovat pomocí Poissonova rozdělení.

3.4.3 Smíšená pravděpodobnostní rozdělení

Uvažujme pravděpodobnostní funkce $f_0(\cdot), f_1(\cdot), \dots, f_n(\cdot)$, které jsou všechny buďto spojitě nebo nespojitě. Dále uvažujme parametry p_0, p_1, \dots, p_n které splňují

$$\begin{aligned} p_i &\geq 0 \\ \sum_{i=0}^{\infty} p_i &= 1 \end{aligned}$$

Pak $\sum_{i=0}^{\infty} p_i f_i(x)$ představuje pravděpodobnostní funkci, která vznikla “smícháním” pravděpodobnostních funkcí $f_0(\cdot), f_1(\cdot), \dots, f_n(\cdot)$.

Příklad 3.9 Ilustrujme myšlenku smíšeného pravděpodobnostního rozdělení s pomocí pravděpodobnostních funkcí $f_0(x) = \phi(\mu_0, \sigma_0^2)(x)$ a $f_1(x) = \phi(\mu_1, \sigma_1^2)(x)$. Pak

$$\begin{aligned} & p_0 \phi(\mu_0, \sigma_0^2)(x) + p_1 \phi(\mu_1, \sigma_1^2)(x) \\ &= (1-p) \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2}(\frac{x-\mu_0}{\sigma_0})^2} + p \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2} \end{aligned}$$

Výše uvedené smíšené rozdělení nazýváme smíšeným normálním rozdělením. V našem konkrétním případě má toto rozdělení pět parametrů, což mu umožňuje nabývat nejrůznějších tvarů. Vhodným nastavením parametrů lze tak vytvořit např. bimodální rozdělení, tj. rozdělení se dvěma vrcholy. Díky své flexibilitě je tak “míchání” pravděpodobnostních rozdělení často využíváno při modelování některých experimentů.

Koncept smíšeného pravděpodobnostního rozdělení lze dále rozšířit. Necht $\{f(x; \theta)\}$ představuje rodinu pravděpodobnostních funkcí indexovaných parametrem θ . Množinu všech hodnot, kterých může θ nabývat, označme Θ . Je-li Θ interval a $g(\theta)$ pravděpodobnostní funkce, která je nulová pro všechny prvky mimo Θ , pak

$$\int_{\Theta} f(x; \theta) g(\theta) d\theta$$

je opět pravděpodobnostní funkce.

Příklad 3.10 Uvažujme $f(x; \theta) = \frac{e^{-\theta} \theta^x}{x!}$ pro $x = 0, 1, 2, \dots$ a $f(x; \theta) = 0$ v ostatních případech. Dále uvažujme pravděpodobnostní funkci gamma rozdělení $g(\theta) = \frac{\lambda^r}{\Gamma(r)} \theta^{r-1} e^{-\lambda \theta} I_{(0, \infty)}(\theta)$. Pak

$$\begin{aligned} \int_0^{\infty} f(x; \theta) g(\theta) d\theta &= \int_0^{\infty} \frac{e^{-\theta} \theta^x}{x!} \frac{\lambda^r}{\Gamma(r)} \theta^{r-1} e^{-\lambda \theta} d\theta = \frac{\lambda^r}{x! \Gamma(r)} \int_0^{\infty} \theta^{r+x-1} e^{-(\lambda+1)\theta} d\theta \\ &= \frac{\lambda^r}{x! \Gamma(r)} \frac{\Gamma(r+x)}{(\lambda+1)^{r+x}} \int_0^{\infty} \frac{[(\lambda+1)\theta]^{r+x-1} e^{-(\lambda+1)\theta}}{\Gamma(r+x)} d[(\lambda+1)\theta] \\ &= \left(\frac{\lambda}{\lambda+1}\right)^r \frac{\Gamma(r+x)}{x! \Gamma(r)} \frac{1}{(\lambda+1)^x} \\ &= \binom{r+x-1}{x} \left(\frac{\lambda}{\lambda+1}\right)^r \left(\frac{1}{\lambda+1}\right)^x \quad \text{pro } x = 0, 1, 2, \dots \end{aligned}$$

což je pravděpodobnostní funkce binomického rozdělení s parametry r a $p = \frac{\lambda}{\lambda+1}$. Na binomické rozdělení je tak možné nahlížet jako na gamma smíšené Poissonovo rozdělení.

3.4.4 Zredukováná pravděpodobnostní rozdělení

Pro ilustraci uvažujme normální rozdělení “seříznuté” zleva hodnotou 0 a zprava hodnotou 1. Pravděpodobnostní funkce takového zredukováného pravděpodobnostního rozdělení je

$$f(x) = \frac{\phi(\mu, \sigma^2)(x) I_{(0,1)}(x)}{\Phi(\mu, \sigma^2)(1) - \Phi(\mu, \sigma^2)(0)}$$

Obecná definice zredukováného spojitého pravděpodobnostního rozdělení je následující.

Věta 3.32 *Jestliže X je spojitá náhodná veličina s pravděpodobnostní funkcí $f(\cdot)$ a kumulativní distribuční funkcí $F(\cdot)$, pak pravděpodobnostní funkce náhodné veličiny X zredukované na interval (a, b) je*

$$\frac{f(x)I_{(a,b)}(x)}{F(b) - F(a)}$$

Kapitola 4

Vícero náhodných proměnných

4.1 Sdružená pravděpodobnost

4.1.1 Sdružená kumulativní distribuční funkce

Definice 4.1 (Sdružená kumulativní distribuční funkce) *Nechť jsou X_1, X_2, \dots, X_k náhodné veličiny definované nad pravděpodobnostním prostorem $(\Omega, \mathcal{A}, P[\cdot])$. Sdružená kumulativní distribuční funkce $F_{X_1, \dots, X_k}(\cdot, \dots, \cdot)$ je definována jako $P[X_1 \leq x_1, \dots, X_k \leq x_k]$ pro všechna x_1, x_2, \dots, x_k .*

Sdružená kumulativní distribuční funkce je tak funkcí s definičním oborem R^k a intervalem $[0, 1]$ jako oborem hodnot.

Dvourozměrná sdružená kumulativní distribuční funkce

Věta 4.1 *Dvourozměrná sdružená kumulativní distribuční funkce splňuje následující podmínky.*

1.

$$F(-\infty, y) = \lim_{x \rightarrow -\infty} F(x, y) = 0 \quad \text{pro všechna } y$$

$$F(x, -\infty) = \lim_{y \rightarrow -\infty} F(x, y) = 0 \quad \text{pro všechna } x$$

$$F(\infty, \infty) = \lim_{x \rightarrow \infty, y \rightarrow \infty} F(x, y) = 1$$

2. *Jestliže $x_1 < x_2$ a $y_1 < y_2$, pak $P[x_1 < X \leq x_2, y_1 < Y \leq y_2] = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geq 0$.*

3. *Funkce $F(x, y)$ je v obou parametrech zprava spojitá. To znamená, že $\lim_{0 < h \rightarrow 0} F(x + h, y) = \lim_{0 < h \rightarrow 0} F(x, y + h) = F(x, y)$.*

Definice 4.2 (Dvourozměrná kumulativní distribuční funkce)

Libovolná funkce, která splňuje výše uvedené tři podmínky, je dvourozměrnou kumulativní distribuční funkcí.

Definice 4.3 (Marginální kumulativní distribuční funkce) Jestliže je $F_{X,Y}(\cdot, \cdot)$ sdruženou kumulativní distribuční funkcí náhodných veličin X a Y , pak kumulativní distribuční funkce $F_X(\cdot)$ a $F_Y(\cdot)$ jsou její marginální kumulativní distribuční funkce.

Tvrzení 4.1 Platí $F_X(x) = F_{X,Y}(x, \infty)$ a $F_Y(y) = F_{X,Y}(\infty, y)$. Známe-li sdruženou kumulativní distribuční funkci, jsme schopni určit také odpovídající marginální kumulativní distribuční funkce.

Obrácené tvrzení však neplatí - znalost marginálních kumulativních distribučních funkcí neimplikuje znalost sdružené kumulativní distribuční funkce.

Věta 4.2 V případě dvourozměrné sdružené kumulativní distribuční funkce pro všechna x a y platí

$$F_X(x) + F_Y(y) - 1 \leq F_{X,Y}(x, y) \leq \sqrt{F_X(x)F_Y(y)}$$

Důkaz 4.1 Nejprve dokážeme

$$F_X(x) + F_Y(y) - 1 \leq F_{X,Y}(x, y)$$

Levou stranu této nerovnosti lze upravit do podoby

$$\begin{aligned} 1 - P[x < X < \infty, y < Y < \infty] + F_{X,Y} - 1 \\ = F_{X,Y} - P[x < X < \infty, y < Y < \infty] \end{aligned}$$

Protože $P[x < X < \infty, y < Y < \infty] \geq 0$, platí $F_X(x) + F_Y(y) - 1 \leq F_{X,Y}(x, y)$. Následně dokážeme

$$F_{X,Y}(x, y) \leq \sqrt{F_X(x)F_Y(y)}$$

Tuto nerovnost lze upravit do podoby

$$\ln(F_{X,Y}(x, y)) + \ln(F_{X,Y}(x, y)) \leq \ln(F_X(x)) + \ln(F_Y(y))$$

Protože $F_{X,Y}(x, y) \leq F_X(x)$ a $F_{X,Y}(x, y) \leq F_Y(y)$, výše uvedená nerovnost platí, čímž uzavíráme důkaz.

4.1.2 Sdružená pravděpodobnostní funkce nespojité náhodné veličiny

Jestliže jsou X_1, X_2, \dots, X_k náhodné veličiny definované nad shodným pravděpodobnostním prostorem, pak je (X_1, X_2, \dots, X_k) n -rozměrnou náhodnou veličinou.

Definice 4.4 (Sdružená nespojitá náhodná veličina) Náhodná veličina (X_1, X_2, \dots, X_k) je k -rozměrnou nespojitou náhodnou veličinou, jestliže může nabývat hodnot pouze nad spočetným množstvím bodů (x_1, x_2, \dots, x_k) z R^k .

Definice 4.5 (Sdružená nespojitá pravděpodobnostní funkce) Jestliže (X_1, X_2, \dots, X_k) je k -rozměrná nespojitá náhodná veličina, pak její sdružená nespojitá pravděpodobnostní funkce $f_{X_1, X_2, \dots, X_k}(\cdot, \dots, \cdot)$ je definována pro hodnoty (x_1, x_2, \dots, x_k) , kterých může náhodná veličina (X_1, X_2, \dots, X_k) nabývat, jako

$$f_{X_1, X_2, \dots, X_k} = P[X_1 = x_1, X_2 = x_2, \dots, X_k = x_k]$$

V ostatních případech je tato funkce rovna nule.

Stejně jako v případě jednorozměrné náhodné veličiny také v případě k -rozměrné náhodné veličiny platí $\sum f_{X_1, X_2, \dots, X_k}(x_1, x_2, \dots, x_k) = 1$, kde sčítáme přes všechny přípustné hodnoty náhodné veličiny (X_1, X_2, \dots, X_k) .

Věta 4.3 *Jestliže X a Y jsou sdruženě nespojitě náhodné veličiny, pak znalost $F_{X,Y}(\cdot, \cdot)$ je ekvivalentní znalosti $f_{X,Y}(\cdot, \cdot)$. Toto tvrzení lze zobecnit také na k -rozměrnou náhodnou veličinu.*

Důkaz 4.2 *Nechť $(x_1, y_1), (x_2, y_2), \dots$ představují všechny možné hodnoty náhodné veličiny (X, Y) . Jestliže známe $f_{X,Y}$, pak*

$$F_{X,Y}(x, y) = \sum f_{X,Y}(x_i, y_i)$$

kde sčítáme přes všechna i , pro která $x_i \leq x$ a $y_i \leq y$. Naopak je-li dáno $F_{X,Y}(\cdot, \cdot)$ pro všechna (x_i, y_i) , pak

$$f_{X,Y}(x_i, y_i) = F_{X,Y}(x_i, y_i) - \lim_{0 < h \rightarrow 0} F_{X,Y}(x_i - h, y_i) - \lim_{0 < h \rightarrow 0} F_{X,Y}(x_i, y_i - h)$$

Definice 4.6 (Marginální nespojitá pravděpodobnostní funkce)

Jestliže jsou X a Y sdružené nespojitě náhodné veličiny, pak $f_X(\cdot)$ a $f_Y(\cdot)$ představují marginální pravděpodobnostní funkce. Tuto definici lze zobecnit. Nechť X_{i_1}, \dots, X_{i_m} představují libovolnou podmnožinu sdružených nespojitých náhodných veličin X_1, \dots, X_k . Pak $f_{X_{i_1}, \dots, X_{i_m}}(x_{i_1}, \dots, x_{i_m})$ představuje marginální pravděpodobnostní funkci.

Jestliže jsou X_1, \dots, X_k sdružené nespojitě náhodné veličiny, pak lze libovolnou marginální pravděpodobnostní funkci odvodit ze sdružené pravděpodobnostní funkce. Obrácené tvrzení však neplatí.

Příklad 4.1 *Uvažujme hod dvěma čtyřstěny. Náhodná veličina X značí hodnotu hodu prvního z nich. Náhodná veličina Y pak značí největší z obou hodů. Předpokládejme, že tyto čtyřstěny nejsou dokonale vyvážené, a proto se skutečně realizované pravděpodobnosti liší od očekávaných. Jednu z možných situací zachycuje následující tabulka.*

4	$1/16 + \epsilon$	$1/16 - \epsilon$	$1/16$	$4/16$
3	$1/16 - \epsilon$	$1/16 + \epsilon$	$3/16$	0
2	$1/16$	$2/16$	0	0
1	$1/16$	0	0	0
y/x	1	2	3	4

Pro libovolné $0 \leq \epsilon \leq \frac{1}{16}$ definuje tato tabulka sdruženou pravděpodobnostní funkci. Nicméně marginální pravděpodobnosti jsou nezávislé na ϵ . Tento příklad tedy ilustruje, že znalost marginálních pravděpodobnostních funkcí neimplikuje znalost sdružené pravděpodobnostní funkce.

Uvažujme experiment, který má $k + 1$ možných výsledků. Označme tyto výsledky jako s_1, s_2, \dots, s_{k+1} . Definujme $p_i = P[s_i]$ pro $i = 1, \dots, k + 1$. Je zřejmé, že musí platit $\sum_{i=1}^{k+1} p_i = 1$. Předpokládejme, že tento pokus opakujeme n -krát a že jednotlivé pokusy jsou vzájemně nezávislé. Nechť náhodná veličina X_i

představuje počet realizací s_i . Nespojité pravděpodobnostní funkce náhodných veličin X_1, \dots, X_k je definována jako

$$f_{X_1, \dots, X_k}(x_1, \dots, x_k) = \frac{n!}{\prod_{i=1}^{k+1} x_i!} \prod_{i=1}^{k+1} p_i^{x_i}$$

kde $x_i = 0, \dots, n$ a $\sum_{i=1}^{k+1} x_i = n$. Všimněme si, že platí $X_{k+1} = n - \sum_{i=1}^k X_i$.

Výše uvedenou rovnici lze vysvětlit následovně. Levou stranu této rovnice je možné vyjádřit ve tvaru $P[X_1 = x_1, \dots, X_{k+1} = x_{k+1}]$. Zajímá nás tak pravděpodobnost, že mezi n vzájemně nezávislými pokusy bude právě x_1 realizací s_1 , x_2 realizací s_2 až x_{k+1} realizací s_{k+1} , kde $\sum_{i=1}^{k+1} x_i = n$. Za předpokladu nezávislosti jednotlivých pokusů lze pravděpodobnost jednoho takového uspořádání vyjádřit jako $p_1^{x_1} p_2^{x_2} \dots p_{k+1}^{x_{k+1}}$. Takovýchto uspořádání pak existuje $\frac{n!}{x_1! x_2! \dots x_{k+1}!}$.

Definice 4.7 (Multinomiální rozdělení) *Výše uvedená sdružená nespojité pravděpodobnostní funkce se nazývá multinomiální distribucí.*

Multinomiální rozdělení má $(k+1)$ parametrů a to n, p_1, p_2, \dots, p_k . Parametr p_{k+1} je definován jako $1 - p_1 - p_2 - \dots - p_k$. Marginální rozdělení náhodné veličiny X_i má charakter binomického rozdělení s parametry n a p_i , protože na výsledek každého experimentu je možné pohlížet jako realizaci popř. nerealizaci výsledku s_i . Na každý z n pokusů je tak možné pohlížet jako na Bernoulliho pokus a na sérii n pokusů je pak možné modelovat pomocí binomického rozdělení.

4.1.3 Sdružená pravděpodobnostní funkce spojité náhodné veličiny

Definice 4.8 (Sdružená spojitá náhodná veličina a její pravděpodobnostní funkce) *k -rozměrná náhodná veličina (X_1, X_2, \dots, X_k) je k -rozměrnou náhodnou veličinou právě tehdy a jen tehdy, jestliže existuje funkce $f_{X_1, \dots, X_k}(\cdot, \dots, \cdot) \geq 0$ taková, že*

$$F_{X_1, \dots, X_k}(x_1, \dots, x_k) = \int_{-\infty}^{x_k} \dots \int_{-\infty}^{x_1} f_{X_1, \dots, X_k}(u_1, \dots, u_k) du_1 \dots du_k$$

pro všechna (x_1, \dots, x_k) . Funkce $f_{X_1, \dots, X_k}(\cdot, \dots, \cdot)$ je tzv. sdruženou pravděpodobnostní funkcí. Tato funkce splňuje následující dvě podmínky.

1. $f_{X_1, \dots, X_k}(x_1, \dots, x_k) \geq 0$
2. $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{X_1, \dots, X_k}(x_1, \dots, x_k) dx_1 \dots dx_k = 1$

Příklad 4.2 *Uvažujme funkci*

$$f(x, y) = K(x+y)I_{(0,12)}(x)I_{(0,1)}(y)$$

Lze vhodně zvolit konstantu K tak, aby tato funkce byla dvourozměrnou pravděpodobnostní funkcí?

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(x+y) dx dy &= K \int_0^1 \int_0^1 (x+y) dx dy \\ &= K \int_0^1 \left(\frac{1}{2} + y\right) dy = K\left(\frac{1}{2} + \frac{1}{2}\right) = 1 \end{aligned}$$

pro $K = 1$. Proto je funkce $f(x, y) = (x + y)I_{(0,1)}(x)I_{(0,1)}(y)$ sdruženou pravděpodobnostní funkcí.

Věta 4.4 Jestliže jsou X a Y sdružené spojité náhodné veličiny, pak znalost sdružené kumulativní distribuční funkce $F_{X,Y}(\cdot, \cdot)$ je ekvivalentní znalosti sdružené pravděpodobnostní funkce $f_{X,Y}(\cdot, \cdot)$. Tuto větu lze rozšířit také na k -rozměrné spojité náhodné veličiny.

Důkaz 4.3 Pro danou sdruženou pravděpodobnostní funkci $f_{X,Y}(\cdot, \cdot)$ lze pro libovolné (x, y) odvodit sdruženou kumulativní distribuční funkci $F_{X,Y}(x, y)$ jako

$$F_{X,Y}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{X,Y}(u, v) du dv$$

Naopak, pro danou sdruženou kumulativní distribuční funkci $F_{X,Y}(x, y)$ lze odpovídající sdruženou pravděpodobnostní funkci $f_{X,Y}(\cdot, \cdot)$ odvodit jako

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}$$

kde x a y představují body, pro které je $F_{X,Y}(x, y)$ derivovatelná.

Definice 4.9 (Marginální pravděpodobnostní funkce) Jestliže jsou X a Y sdružené spojité náhodné veličiny, pak $f_X(\cdot)$ a $f_Y(\cdot)$ jsou nazývány marginálními pravděpodobnostními funkcemi.

Tuto definici lze zobecnit. Nechť X_{i_1}, \dots, X_{i_m} představuje libovolnou podmnožinu sdružených náhodných veličin X_1, \dots, X_k . Pak je funkce $f_{X_{i_1}, \dots, X_{i_m}}(x_{i_1}, \dots, x_{i_m})$ je nazývána marginální pravděpodobnostní funkcí náhodné veličiny $(X_{i_1}, \dots, X_{i_m})$.

Jestliže jsou X a Y sdružené spojité náhodné veličiny, pak platí

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

a

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

protože

$$f_X(x) = \frac{dF_X(x)}{dx} = \frac{d}{dx} \left[\int_{-\infty}^x \left(\int_{-\infty}^{\infty} f_{X,Y}(u, y) dy \right) du \right] = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

a

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d}{dy} \left[\int_{-\infty}^{\infty} \left(\int_{-\infty}^x f_{X,Y}(x, u) dx \right) du \right] = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

Příklad 4.3 Uvažujme sdruženou pravděpodobnostní funkci

$$f_{X,Y}(x, y) = (x + y)I_{(0,1)}(x)I_{(0,1)}(y)$$

Sdruženou kumulativní distribuční funkci lze vypočítat následujícím způsobem.

$$\begin{aligned} F_{X,Y}(x,y) &= I_{(0,1)}(x)I_{(0,1)}(y) \int_0^y \int_0^x (u+v)dudv + I_{(0,1)}(x)I_{[1,\infty)}(y) \int_0^1 \int_0^x (u+v)dudv \\ &\quad + I_{[1,\infty)}(x)I_{(0,1)}(y) \int_0^y \int_0^1 (u+v)dudv + I_{[1,\infty)}(x)I_{[1,\infty)}(y) \\ &= \frac{1}{2}[(x^2y + xy^2)I_{(0,1)}(x)I_{(0,1)}(y) + (x^2 + x)I_{(0,1)}(x)I_{[1,\infty)}(y) \\ &\quad + (y + y^2)I_{[1,\infty)}(x)I_{(0,1)}(y)] + I_{[1,\infty)}(x)I_{[1,\infty)}(y) \end{aligned}$$

Marginální pravděpodobnostní funkci pak lze vyjádřit jako

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = I_{(0,1)}(x) \int_0^1 (x+y)dy = (x + \frac{1}{2})I_{(0,1)}(x)$$

popř. jako

$$f_X(x) = \frac{\partial F_{X,Y}(x, \infty)}{\partial x} = \frac{\partial F_X(x)}{\partial x} = I_{(0,1)}(x) \frac{\partial}{\partial x} \left(\frac{x^2 + x}{2} \right) = (x + \frac{1}{2})I_{(0,1)}(x)$$

4.2 Podmíněné rozdělení a nezávislost

4.2.1 Podmíněná kumulativní distribuční funkce nespojitě náhodné veličiny

Definice 4.10 (Podmíněná nespojitá pravděpodobnostní funkce)

Nechť jsou X a Y sdružené nespojitě náhodné veličiny se sdruženou nespojitou pravděpodobnostní funkcí $f_{X,Y}(\cdot, \cdot)$. Podmíněná nespojitá pravděpodobnostní funkce $f_{Y|X}(\cdot|x)$ náhodné veličiny Y pro $X = x$ je definována jako

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

je-li $f_X(x) > 0$, kde $f_X(x)$ je marginální pravděpodobnostní funkce náhodné veličiny X v bodě x . Pro $f_X(x) = 0$ není funkce $f_{Y|X}(\cdot|x)$ definována. Podobně

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

za podmínky $f_Y(y) > 0$.

Protože náhodné veličiny X a Y jsou nespojitě, mohou nabývat hodnot některého z bodů koncentrace x_1, x_2, \dots v případě X a y_1, y_2, \dots v případě Y . Jestliže $f_X(x) > 0$, pak $x = x_i$ pro některé i a $f_X(x_i) = P[X = x_i]$. Analogicky $f_{X,Y}(x_i, y_i) = P[X = x_i, Y = y_i]$. Proto

$$f_{Y|X}(y_j|x_i) = \frac{f_{X,Y}(x_i, y_j)}{f_X(x_i)} = \frac{P[X = x_i, Y = y_j]}{P[X = x_i]} = P[Y = y_j|X = x_i]$$

Jedná se tedy o podmíněnou pravděpodobnost, tak jak jsme ji definovali v první kapitole.

Vzhledem k tomu, že $f_{Y|X}(y_j|x_i)$ je nespojitá distribuční funkce, musí být kladná a její součet přes všechny přípustné hodnoty Y musí být roven jedné. První podmínka je splněna, protože $f_{X,Y}(x, y)$ je nezáporné a $f_X(x)$ je kladné, a proto je jejich podíl kladný. Druhá podmínka je taktéž splněna, protože

$$\sum_j f_{Y|X}(y_j|x) = \sum_j \frac{f_{X,Y}(x, y_j)}{f_X(x)} = \frac{1}{f_X(x)} \sum_j f_{X,Y}(x, y_j) = \frac{f_X(x)}{f_X(x)} = 1$$

Definice 4.11 (Podmíněná nespojitá kumulativní distribuční funkce)
Jestliže jsou X a Y sdružené nespojitě náhodné veličiny, pak podmíněná nespojitá kumulativní distribuční funkce $F_{Y|X}(\cdot|x)$ je definována jako $F_{Y|X}(y|x) = P[Y \leq y|X = x]$ pro $f_X(x) > 0$. Vazba mezi podmíněnou nespojitou kumulativní distribuční funkcí a pravděpodobnostní funkcí je dána vztahem

$$F_{Y|X}(y|x) = \sum_{j: y_j \leq y} f_{Y|X}(y_j|x)$$

Definice 4.12 (Podmíněná nespojitá pravděpodobnostní funkce)

Nechť (X_1, \dots, X_k) je k -rozměrná nespojitá náhodná veličina a necht' X_{i_1}, \dots, X_{i_r} a X_{j_1}, \dots, X_{j_s} jsou dvě neslučitelné podmnožiny náhodných veličin X_1, \dots, X_k . Podmíněná pravděpodobnostní funkce r -rozměrné náhodné veličiny X_{i_1}, \dots, X_{i_r} pro dané hodnoty $(x_{j_1}, \dots, x_{j_s})$ náhodné veličiny X_{j_1}, \dots, X_{j_s} je definována jako

$$\begin{aligned} & f_{X_{i_1}, \dots, X_{i_r}|X_{j_1}, \dots, X_{j_s}}(x_{i_1}, \dots, x_{i_r}|x_{j_1}, \dots, x_{j_s}) \\ &= \frac{f_{X_{i_1}, \dots, X_{i_r}, X_{j_1}, \dots, X_{j_s}}(x_{i_1}, \dots, x_{i_r}, x_{j_1}, \dots, x_{j_s})}{f_{X_{j_1}, \dots, X_{j_s}}(x_{j_1}, \dots, x_{j_s})} \end{aligned}$$

Příklad 4.4 Vyberme 12 karet bez vracení ze standardního hracího balíčku. Necht' X_1 představuje počet tažených es, X_2 představuje počet tažených dvojek, X_3 představuje počet tažených trojek a X_4 počet tažených čtyřek. Sdružená pravděpodobnostní funkce těchto čtyř náhodných veličin je

$$f_{X_1, X_2, X_3, X_4}(x_1, x_2, x_3, x_4) = \frac{\binom{4}{x_1} \binom{4}{x_2} \binom{4}{x_3} \binom{4}{x_4} \binom{36}{12-x_1-x_2-x_3-x_4}}{\binom{52}{12}}$$

kde $x_i = 0, 1, 2, 3, 4$ při omezení $\sum x_i \leq 12$. Existuje velké množství souvisejících podmíněných pravděpodobnostních funkcí. Jednou z nich je např.

$$\begin{aligned} & f_{X_2, X_4|X_1, X_3}(x_2, x_4|x_1, x_3) \\ &= \frac{\binom{4}{x_1} \binom{4}{x_2} \binom{4}{x_3} \binom{4}{x_4} \binom{36}{12-x_1-x_2-x_3-x_4}}{\binom{4}{x_1} \binom{4}{x_3} \binom{44}{12-x_1-x_3}} \bigg/ \binom{52}{12} \\ &= \frac{\binom{4}{x_2} \binom{4}{x_4} \binom{36}{12-x_1-x_2-x_3-x_4}}{\binom{44}{12-x_1-x_3}} \bigg/ \binom{52}{12} \end{aligned}$$

4.2.2 Podmíněná pravděpodobnostní funkce spojitých náhodných veličin

Definice 4.13 (Podmíněná pravděpodobnostní funkce) Necht' jsou X a Y sdružené spojité náhodné veličiny se sdruženou pravděpodobnostní funkcí

$f_{X,Y}(x,y)$. Podmíněná pravděpodobnostní funkce náhodné veličiny Y pro $X = x$ je definována jako

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

kde $f_X(x) > 0$. Pro $f_X(x) = 0$ není tato funkce definována. Podobně $f_{X|Y}(x|y)$ je definována jako

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

pro $f_Y(y) > 0$ a nedefinováno pro $f_Y(y) = 0$.

Funkce $f_{Y|X}(\cdot|x)$ se nazývá podmíněná pravděpodobnostní funkce a proto by měla být nezáporná a její integrál přes všechny přípustné hodnoty by měl být roven jedné. První podmínka je splněna, protože $f_{X,Y}(x,y)$ je nezáporné a $f_Y(y)$ je kladné. Druhá podmínka je taktéž splněna, protože

$$\begin{aligned} \int_{-\infty}^{\infty} f_{Y|X}(y|x) dy &= \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_X(x)} dy \\ &= \frac{1}{f_X(x)} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy = \frac{f_X(x)}{f_X(x)} = 1 \end{aligned}$$

Uvažujme $f_{Y|X}(\cdot|x_0)$. Funkci $f_{X,Y}(x,y)$ lze znázornit jako plochu nad rovinou xy . Rovina kolmá na rovinu xy , která ji protíná v přímce $x = x_0$, protíná plochu $f_{X,Y}(x,y)$ v křivce $f_{X,Y}(x_0,y)$. Plocha pod touto křivkou je

$$\int_{-\infty}^{\infty} f_{X,Y}(x_0,y) dy = f_X(x_0)$$

Proto, vydělíme-li $f_{X,Y}(x_0,y)$ touto plochou $f_X(x_0)$, získáme pravděpodobnostní funkce $f_{Y|X}(y|x_0)$.

Definice 4.14 (Podmíněná spojitá kumulativní distribuční funkce)

Jestliže jsou X a Y sdružené spojitě náhodné veličiny, pak je podmíněná kumulativní distribuční funkce náhodné veličiny Y pro $X = x$ definována jako

$$F_{Y|X}(y|x) = \int_{-\infty}^y f_{Y|X}(z|x) dz$$

pro všechna taková x , kde $f_X(x) > 0$.

Příklad 4.5 Uvažujme sdruženou pravděpodobnostní funkci

$$f_{X,Y}(x,y) = (x+y)I_{(0,1)}(x)I_{(0,1)}(y)$$

Podmíněnou pravděpodobnostní funkci pak lze vypočítat jako

$$f_{Y|X}(y|x) = \frac{(x+y)I_{(0,1)}(x)I_{(0,1)}(y)}{\left(x + \frac{1}{2}\right)I_{(0,1)}(x)} = \frac{x+y}{x + \frac{1}{2}}I_{(0,1)}(y)$$

pro $0 < x < 1$. Kumulativní distribuční funkci pak lze odvodit jako

$$F_{Y|X}(y|x) = \int_{-\infty}^y f_{Y|X}(z|x) dz = \int_0^y \frac{x+z}{x + \frac{1}{2}} dz = \frac{1}{x + \frac{1}{2}} \int_0^y (x+z) dz = \frac{1}{x + \frac{1}{2}} \left(xy + \frac{y^2}{2} \right)$$

pro $0 < y < 1$.

Definici podmíněné pravděpodobnostní funkce lze analogicky rozšířit také na k -rozměrnou spojitou náhodnou veličinu. Např.

$$f_{X_1, X_2, X_4 | X_3, X_5}(x_1, x_2, x_4 | x_3, x_5) = \frac{f_{X_1, X_2, X_3, X_4, X_5}(x_1, x_2, x_3, x_4, x_5)}{f_{X_3, X_5}(x_3, x_5)}$$

pro $f_{X_3, X_5}(x_3, x_5) > 0$.

4.2.3 Více o podmíněné pravděpodobnostní funkci

Uvažujme pravděpodobnost náhodného jevu A podmíněného $X = x$, kde X může být jak spojitá tak nespojitá náhodná veličina. Předpokládejme, že jak náhodný jev A tak náhodná veličina X jsou definovány nad shodným pravděpodobnostním prostorem. Pokusme se definovat $P[A|X = x]$. Jestliže je X nespojitá a x bodem koncentrace, pak

$$P[A|X = x] = \frac{P[A, X = x]}{P[X = x]}$$

V opačném případě není $P[A, X = x]$ definováno. Jestliže je X spojitá, pak $P[X = x] = 0$. Jestliže je však x takové, že jev $x - h < X < x + h$ má kladnou pravděpodobnost pro libovolné $h > 0$, pak

$$P[A|X = x] = \lim_{0 < h \rightarrow 0} P[A|x - h < X < x + h]$$

za předpokladu existence limity.

Pro podmíněnou pravděpodobnost platí následující tvrzení.

Tvrzení 4.2

1. $P[A] = \sum_{i=1}^{\infty} P[A|X = x_i] f_X(x_i)$ jestliže X je nespojitá s body koncentrace x_1, x_2, \dots
2. $P[A] = \int_{-\infty}^{\infty} P[A|X = x] f_X(x) dx$ jestliže je X spojitá
3. $P[A, X \in B] = \sum_{i: x_i \in B} P[A|X = x_i] f_X(x_i)$ jestliže X je nespojitá s body koncentrace x_1, x_2, \dots
4. $P[A, X \in B] = \int_B P[A|X = x] f_X(x) dx$ jestliže je X spojitá
5. $P[A|X = x] = P[h(X, Y) \leq z | X = x] = P[h(x, Y) \leq z | X = x]$ kde $A = h(X, Y) \leq z$ a $h(\cdot, \cdot)$ je funkcí dvou proměnných

Z druhé rovnice vyplývá

$$F_Y(y) = \int_{-\infty}^{\infty} F_{Y|X}(y|x) f_X(x) dx$$

kde $A = Y \leq y$. Ze čtvrté rovnice pak vyplývá

$$F_{X,Y}(x, y) = \int_{-\infty}^x F_{Y|X}(y|x') f_X(x') dx'$$

kde $A = Y \leq y$ a $B = (-\infty, x]$.

V některých případech je snadné nalézt $P[A|X = x]$ a složitě nalézt $P[A]$. Jestliže však známe $f_X(\cdot)$, lze $P[A]$ vypočítat na základě některé z výše uvedených rovnic.

Příklad 4.6 Uvažujme tři náhodně zvolené body na obvodu jednotkového kruhu. Jaká je pravděpodobnost, že se všechny tři body budou nacházet na téže polokruhu?

První bod zvolíme jako orientační, a proto ho umístíme do bodu nula. Nechť X představuje polohu druhého bodu a jev A představuje situaci, kdy se všechny tři body nachází na téže polokruhu. Náhodná veličina X sleduje uniformní rozdělení nad intervalem $(0, 2\pi)$. Pro náhodný jev A platí $P[A] = \int P[A|X = x]f_X(x)dx$. Je-li $0 < x < \pi$, pak $P[A|X = x] = (\pi - x + \pi)/2\pi$. Pro $X = x$ totiž A nastane tehdy a jen tehdy, jestliže třetí bod padne mezi $x - \pi$ a π . Podobně $P[A|X = x] = (x + \pi - \pi)/2\pi$ pro $\pi \leq x \leq 2\pi$. Proto

$$P[A] = \int_0^{2\pi} P[A|X = x] \frac{1}{2\pi} dx = \frac{1}{2\pi} \left[\int_0^{\pi} \frac{2\pi - x}{2\pi} dx + \int_{\pi}^{2\pi} \frac{x}{2\pi} dx \right] = \frac{3}{4}$$

4.2.4 Nezávislost

Definice 4.15 (Nezávislost) Nechť (X_1, X_2, \dots, X_k) je k -rozměrná náhodná veličina. X_1, X_2, \dots, X_k jsou nezávislé tehdy a jen tehdy, jestliže

$$F_{X_1, \dots, X_k}(x_1, \dots, x_k) = \prod_{i=1}^k F_{X_i}(x_i)$$

pro všechna x_1, x_2, \dots, x_k .

Definice 4.16 (Nezávislost) Nechť (X_1, X_2, \dots, X_k) je k -rozměrná nespojitá náhodná veličina se sdruženou pravděpodobnostní funkcí $f_{X_1, \dots, X_k}(\cdot, \dots, \cdot)$. X_1, X_2, \dots, X_k jsou nezávislé tehdy a jen tehdy, jestliže

$$f_{X_1, \dots, X_k}(x_1, \dots, x_k) = \prod_{i=1}^k f_{X_i}(x_i)$$

pro všechna (x_1, x_2, \dots, x_k) náhodné veličiny (X_1, \dots, X_k) .

Definice 4.17 (Nezávislost) Nechť (X_1, X_2, \dots, X_k) je k -rozměrná spojitá náhodná veličina se sdruženou pravděpodobnostní funkcí $f_{X_1, \dots, X_k}(\cdot, \dots, \cdot)$. X_1, X_2, \dots, X_k jsou nezávislé tehdy a jen tehdy, jestliže

$$f_{X_1, \dots, X_k}(x_1, \dots, x_k) = \prod_{i=1}^k f_{X_i}(x_i)$$

pro všechna x_1, x_2, \dots, x_k .

Lze dokázat, že jsou-li X_1, \dots, X_k sdružené nespojité náhodné veličiny, pak jsou definice (4.15) a (4.16) ekvivalentní. Analogické tvrzení platí také pro spojité náhodné veličiny a definice (4.15) a (4.17).

Podmíněná pravděpodobnost a nezávislost jsou dva vzájemně provázané pojmy. Předpokládejme, že X a Y jsou dvě nezávislé náhodné veličiny. Pak, na základě definice nezávislosti, platí

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

Nicméně z definice podmíněné pravděpodobnosti vyplývá

$$f_{X,Y}(x, y) = f_{Y|X}(y|x)f_X(x)$$

což implikuje $f_{Y|X}(y|x) = f_Y(y)$. Abychom tedy dokázali, že dvě náhodné veličiny nejsou nezávislé, stačí dokázat, že $f_{Y|X}(y|x)$ závisí na X .

Příklad 4.7 Uvažujme hod dvěma čtyřstěny. Náhodná veličina X představuje hodnotu prvního z nich a náhodná veličina Y nejvyšší hodnotu z obou hodů. Je zřejmé, že X a Y nejsou vzájemně nezávislé, protože

$$f_{X|Y}(2|3) = 0 \neq f_X(2) = \frac{3}{16}$$

Příklad 4.8 Necht' $f_{X,Y}(x, y) = (x + y)I_{(0,1)}(x)I_{(0,1)}(y)$. Je zřejmé, že X a Y nejsou vzájemně nezávislé, protože podmíněná pravděpodobnostní funkce

$$f_{Y|X}(y|x) = \frac{x+y}{x+\frac{1}{2}}I_{(0,1)}(y)$$

je závislá na X . Připomeňme, že tato funkce je definována pro $0 < x < 1$.

Věta 4.5 Jestliže jsou X_1, \dots, X_k nezávislé náhodné veličiny a $g_1(\cdot), \dots, g_k(\cdot)$ jsou takové funkce, že $Y_j = g_j(X_j)$ pro $j = 1, \dots, k$ jsou náhodné veličiny, pak jsou Y_1, \dots, Y_k nezávislé.

Důkaz 4.4 Jestliže $g_j^{-1}(B_j) = \{z : g_j(z) \in B_j\}$, pak jsou jevy $\{Y_j \in B_j\}$ a $\{X_j \in g_j^{-1}(B_j)\}$ ekvivalentní, což implikuje

$$\begin{aligned} P[Y_1 \in B_1, \dots, Y_k \in B_k] &= P[X_1 \in g_1^{-1}(B_1), \dots, X_k \in g_k^{-1}(B_k)] \\ &= \prod_{j=1}^k P[X_j \in g_j^{-1}(B_j)] = \prod_{j=1}^k P[Y_j \in B_j] \end{aligned}$$

Uvažujme dvě náhodné nezávislé veličiny X a Y a reálné funkce $f(\cdot)$ a $g(\cdot)$. Výše uvedená věta tvrdí, že $f(X)$ a $g(Y)$ jsou taktéž nezávislé.

4.3 Střední hodnota, kovariance a korelace

4.3.1 Střední hodnota

Definice 4.18 (Střední hodnota) Necht' (X_1, \dots, X_k) je k -rozměrná náhodná veličina s pravděpodobnostní funkcí $f_{X_1, \dots, X_k}(\cdot, \dots, \cdot)$. Očekávaná hodnota funkce $g(\cdot, \dots, \cdot)$ k -rozměrné náhodné veličiny je definována jako

$$E[g(X_1, \dots, X_k)] = \sum g(x_1, \dots, x_k) f_{X_1, \dots, X_k}(x_1, \dots, x_k)$$

jestliže je (X_1, \dots, X_k) nespojitě a kde sčítáme přes všechny přípustné hodnoty (X_1, \dots, X_k) .

Je-li (X_1, \dots, X_k) spojitě, pak je střední hodnota této funkce definována jako

$$E[g(X_1, \dots, X_k)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_k) dx_1 \cdots dx_k$$

Věta 4.6 Jestliže $g(x_1, \dots, x_k) = x_i$, pak

$$E[g(X_1, \dots, X_k)] = E[X_i]$$

Důkaz 4.5 Předpokládejme, že (X_1, \dots, X_k) je spojitý. Pak

$$\begin{aligned} E[g(X_1, \dots, X_k)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i f_{X_1, \dots, X_k} dx_1 \cdots dx_k \\ &= \int_{-\infty}^{\infty} x_i f_{X_i}(x_i) dx_i = E[X_i] \end{aligned}$$

Důkaz pro nespojitou náhodnou veličinu je analogický.

Věta 4.7 Jestliže $g(x_1, \dots, x_k) = (x_i - E[X_i])^2$, pak

$$E[g(X_1, \dots, X_k)] = E[(X_i - E[X_i])^2] = D[X_i]$$

Výše uvedenou větu lze dokázat podobně jako v předchozím případě.

Příklad 4.9 Uvažujme trojrozměrnou náhodnou veličinu (X_1, X_2, X_3) s pravděpodobnostní funkcí

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = 8x_1x_2x_3I_{(0,1)}(x_1)I_{(0,1)}(x_2)I_{(0,1)}(x_3)$$

pak např.

$$E[3X_1 + 2X_2 + 6X_3] = \int_0^1 \int_0^1 \int_0^1 (3x_1 + 2x_2 + 6x_3) 8x_1x_2x_3 dx_1 dx_2 dx_3 = \frac{22}{3}$$

a

$$E[X_1X_2X_3] = \int_0^1 \int_0^1 \int_0^1 8x_1^2x_2^2x_3^2 dx_1 dx_2 dx_3 = \frac{8}{27}$$

Věta 4.8 Je-li c_i konstanta, pak

$$E\left[\sum_{i=1}^m c_i g_i(X_1, \dots, X_k)\right] = \sum_{i=1}^m c_i E[g_i(X_1, \dots, X_k)]$$

4.3.2 Kovariance a korelace

Definice 4.19 (Kovarianční koeficient) Nechtě jsou X a Y libovolné náhodné veličiny definované na shodném pravděpodobnostním prostoru. Pak je kovariance mezi X a Y , označovaná $\sigma_{X,Y}$, definována jako

$$\sigma_{X,Y} = E[(X - E[X])(Y - E[Y])]$$

za předpokladu existence obou středních hodnot.

Definice 4.20 (Korelační koeficient) Korelační koeficient náhodných veličin X a Y je definován jako

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

za předpokladu existence $\sigma_{X,Y}$, σ_X a σ_Y .

Jak kovariance tak korelace měří míru lineární závislosti dvou náhodných veličin. Oba koeficienty budou kladné, jestliže $X - E[X]$ a $Y - E[Y]$ mají s vysokou pravděpodobností stejné znaménko a naopak. Korelace pak na rozdíl od kovariance měří také sílu této závislosti, protože je “normována” směrodatnými odchylkami náhodných veličin X a Y ve jmenovateli.

Věta 4.9 *Pro korelační koeficient dvou náhodných veličin X a Y platí*

$$-1 \leq \rho_{X,Y} \leq 1$$

Věta 4.10 (Výpočtový tvar kovariančního koeficientu)

$$\sigma_{X,Y} = E[XY] - E[X]E[Y]$$

Důkaz 4.6

$$\begin{aligned} E[X - E[X]]E[Y - E[Y]] &= E[XY - E[X]Y - E[Y]X + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] = E[XY] - E[X]E[Y] \end{aligned}$$

4.3.3 Podmíněná střední hodnota a rozptyl

Definice 4.21 (Podmíněná střední hodnota) *Nechť (X, Y) je dvourozměrná náhodná veličina a $g(\cdot, \cdot)$ je funkcí dvou proměnných. Podmíněná střední hodnota $g(X, Y)$ pro $X = x$ je definována jako*

$$E[g(X, Y)|X = x] = \int_{-\infty}^{\infty} g(x, y)f_{Y|X}(y|x)dy$$

je-li (X, Y) spojitě popř. jako

$$E[g(X, Y)|X = x] = \sum g(x, y_j)f_{Y|X}(y_j|x)$$

je-li (X, Y) nespojitě, kde sčítáme přes všechny přípustné hodnoty náhodné veličiny Y .

Výše uvedený příklad lze zobecnit také na vícerozměrné veličiny. Uvažujme $k + m$ -rozměrnou spojitou náhodnou veličinu $(X_1, \dots, X_k, Y_1, \dots, Y_m)$ s pravděpodobnostní funkcí $f_{X_1, \dots, X_k, Y_1, \dots, Y_m}$, pak

$$\begin{aligned} &E[g(X_1, \dots, X_k, Y_1, \dots, Y_m)|x_1, \dots, x_k] \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_k, y_1, \dots, y_m)f_{Y_1, \dots, Y_m|X_1, \dots, X_k}(y_1, \dots, y_m|x_1, \dots, x_k)dy_1 \dots dy_m \end{aligned}$$

Příklad 4.10 *Uvažujme podmíněnou pravděpodobnostní funkci*

$$f_{Y|X}(y|x) = \frac{x+y}{x+\frac{1}{2}}I_{(0,1)}(y)$$

pro $0 < x < 1$. Podmíněná střední hodnota náhodné veličiny je tak rovna

$$E[Y|X = x] = \int_0^1 y \frac{x+y}{x+\frac{1}{2}} dy = \frac{1}{x+\frac{1}{2}} \left(\frac{x}{2} + \frac{1}{3} \right)$$

Na $E[g(Y)|x]$ lze pohlížet jako na funkci proměnné x , a proto může definovat funkci $h(\cdot)$ jako $h(x) = E[g(Y)|x]$. Pak

$$\begin{aligned} E[h(x)] &= E[E[g(Y)|X]] = \int_{-\infty}^{\infty} h(x)f_X(x)dx = \int_{-\infty}^{\infty} E[g(Y)|x]f_X(x)dx \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(y)f_{Y|X}(y|x)dy \right] f_X(x)dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y)f_{Y|X}(y|x)f_X(x)dydx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y)f_{X,Y}(x,y)dydx = E[g(Y)] \end{aligned}$$

Tímto jsme dokázali platnost následující věty.

Věta 4.11 *Nechť je (X, Y) dvourozměrná náhodná veličina. Pak*

$$E[g(Y)] = E[E[g(Y)|X]]$$

a jako specifický případ uveďme

$$E[Y] = E[E[Y|X]]$$

Definice 4.22 (Regresní křivka) $E[Y|X = x]$ se nazývá regresní křivkou Y pro $X = x$.

Definice 4.23 (Podmíněný rozptyl) Rozptyl náhodné veličiny Y pro dané $X = x$ je definován jako

$$D[Y|X = x] = E[Y^2|X = x] - [E[Y|X = x]]^2$$

Věta 4.12

$$D[Y] = E[D[Y|X]] + D[E[Y|X]]$$

Důkaz 4.7

$$\begin{aligned} E[D[Y|X]] &= E[E[Y^2|X]] - E[E[Y|X]]^2 = E[Y^2] - E[Y]^2 - E[E[Y|X]^2] + E[Y]^2 \\ &= D[Y] - E[E[Y|X]^2] + E[E[Y|X]]^2 = D[Y] - D[E[Y|X]] \end{aligned}$$

Věta 4.13 *Nechť je (X, Y) dvourozměrnou náhodnou veličinou a $g_1(\cdot)$ a $g_2(\cdot)$ jsou reálné funkce jedné proměnné. Pak*

1. $E[g_1(Y) + g_2(Y)|X = x] = E[g_1(Y)|X = x] + E[g_2(Y)|X = x]$
2. $E[g_1(Y)g_2(Y)|X = x] = g_2(x)E[g_1(Y)|X = x]$

Výše uvedenou větu lze zobecnit také na vícerozměnné náhodné veličiny.

4.3.4 Sdružená momentová funkce a momenty

Definice 4.24 (Sdružené momenty) Sdružené obecné momenty náhodných veličin X_1, \dots, X_k jsou definovány jako

$$E[X_1^{r_1} X_2^{r_2} \dots X_k^{r_k}]$$

kde jednotlivá r_i mohou nabývat přirozených čísel včetně nuly. Sdružené centrální momenty jsou pak definovány jako

$$E[(X_1 - E[X_1])^{r_1} (X_2 - E[X_2])^{r_2} \dots (X_k - E[X_k])^{r_k}]$$

Pokud $r_i = r_j = 1$ a všechna ostatní r_m jsou rovna nule, pak je příslušný sdružený moment roven $E[(X_i - E[X_i])(X_j - E[X_j])]$, což odpovídá kovariančnímu koeficientu mezi náhodnými veličinami X_i a X_j .

Definice 4.25 (Sdružená momentová funkce) *Sdružená momentová funkce náhodné veličiny (X_1, \dots, X_k) je definována jako*

$$m_{X_1, \dots, X_k}(t_1, \dots, t_k) = E\left[e^{\sum_{j=1}^k t_j X_j}\right]$$

pokud střední hodnota existuje pro všechny hodnoty t_1, \dots, t_k takové, že $-h < t_j < h$ pro některé $h > 0$ a $j = 1, \dots, k$.

Z momentové funkce $m_{X_1, \dots, X_k}(t_1, \dots, t_k)$ lze vypočíst r -tý moment náhodné veličiny X_j tak, že tuto funkci derivujeme r -krát vzhledem k t_j a následně aplikujeme limitu $t_i \rightarrow 0$ pro $i = 1, \dots, k$.

Mezi marginální a sdruženou momentovou funkcí náhodných veličin X a Y platí následující vztah.

$$m_X(t_1) = m_{X,Y}(t_1, 0) = \lim_{t_2 \rightarrow 0} m_{X,Y}(t_1, t_2)$$

Analogicky lze ze sdružené momentové funkce odvodit marginální momentovou funkci $m_Y(t_2)$.

4.3.5 Nezávislost, střední hodnota a rozptyl

Věta 4.14 *Jestliže X a Y jsou dvě nezávislé náhodné veličiny a $g_1(\cdot)$ a $g_2(\cdot)$ jsou funkce jedné proměnné, pak*

$$E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$$

Tuto větu je možné zobecnit také pro případ k rozměrné náhodné veličiny.

Důkaz 4.8 *Výše uvedenou větu dokážeme pro sdružené spojité náhodné veličiny.*

$$\begin{aligned} E[g_1(X)g_2(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x)g_2(y)f_{X,Y}(x,y)dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x)g_2(y)f_X(x)f_Y(y)dx dy \\ &= \int_{-\infty}^{\infty} g_1(x)f_X(x)dx \int_{-\infty}^{\infty} g_2(y)f_Y(y)dy \\ &= E[g_1(X)]E[g_2(X)] \end{aligned}$$

Tvrzení 4.3 *Jestliže jsou X a Y nezávislé, pak $\sigma_{X,Y} = 0$.*

Důkaz 4.9

$$\begin{aligned} \sigma_{X,Y} &= E[(X - E[X])(Y - E[Y])] = E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)] \\ &= E[X - E[X]]E[Y - E[Y]] = 0 \end{aligned}$$

protože $E[X - E[X]] = 0$.

Definice 4.26 (Nekorelované náhodné veličiny) Náhodné veličiny X a Y jsou nekorelované tehdy a jen tehdy, jestliže $\sigma_{X,Y} = 0$.

Obrácené tvrzení však nemusí vždy platit. To znamená, že $\sigma_{X,Y} = 0$ neimplikuje nezávislost mezi náhodnými veličinami X a Y .

Věta 4.15 Dvě sdružené náhodné veličiny X a Y jsou nezávislé tehdy a jen tehdy, jestliže $m_{X,Y}(t_1, t_2) = m_X(t_1)m_Y(t_2)$ pro všechna t_1, t_2 , kde $-h < t_i < h$ pro $i = 1, 2$ a $h > 0$.

Tuto větu je možné zobecnit také pro případ k rozměrné náhodné veličiny.

Důkaz 4.10 Nezávislost náhodných veličin X a Y znamená, že s využitím věty (4.14) a substitucí $g_1(x) = e^{t_1 x}$ a $g_2(y) = e^{t_2 y}$, lze rozpsat sdruženou momentovou funkci jako součin marginálních momentových funkcí.

4.4 Cauchy-Schwarzova nerovnost

Věta 4.16 (Cauchy-Schwarzova nerovnost) Necht' mají náhodné veličiny X a Y konečné druhé momenty. Pak

$$E[XY]^2 = |E[XY]|^2 \leq E[X^2]E[Y^2]$$

kde rovnost platí tehdy a jen tehdy, jestliže $P[Y = cX] = 1$ pro nějakou konstantu c .

Důkaz 4.11 Existence $E[X]$, $E[Y]$ a $E[XY]$ vyplývá z existence $E[X^2]$ a $E[Y^2]$. Definujme $0 \leq h(t) = E[(tX - Y)^2] = E[X^2]t^2 - 2E[XY]t + E[Y^2]$. Funkce $h(t)$ je kvadratickou funkcí proměnné t a je větší nebo rovna nule. Jestliže $h(t) > 0$, pak kořeny kvadratické funkce $h(t)$ nejsou reálnými čísly, což implikuje $E[XY]^2 < E[X^2]E[Y^2]$. Jestliže $h(t) = 0$ pro nějaké $t = t_0$, pak $E[(t_0X - Y)^2] = 0$, což implikuje $P[t_0X = Y] = 1$.

Tvrzení 4.4 Pro korelační koeficient platí $|\rho_{X,Y}| \leq 1$ s rovností tehdy a jen tehdy, jestliže jedna náhodná veličina je lineární funkcí druhé náhodné veličiny s pravděpodobností jedna.

Důkaz 4.12 Přepišme Cauchy-Schwarzovu nerovnost do tvaru

$$|E[UV]| \leq \sqrt{E[U^2]E[V^2]}$$

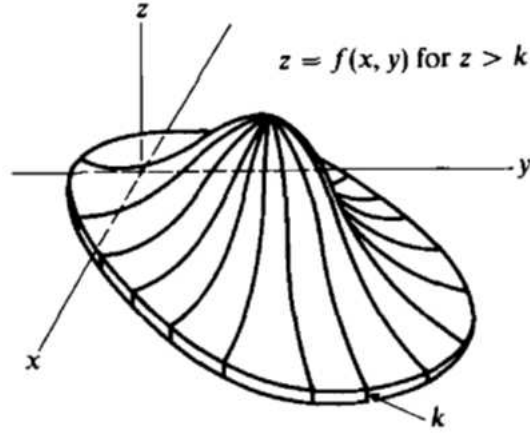
a následně použijme substituci $U = X - E[X]$ a $E[V] = Y - E[Y]$.

4.5 Dvourozměrné normální rozdělení

4.5.1 Pravděpodobnostní funkce

Definice 4.27 (Dvourozměrné normální rozdělení) Necht' má dvourozměrná náhodná veličina (X, Y) sdruženou pravděpodobnostní funkci

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-E[X]}{\sigma_X}\right)^2 - 2\rho\frac{x-E[X]}{\sigma_X}\frac{y-E[Y]}{\sigma_Y} + \left(\frac{y-E[Y]}{\sigma_Y}\right)^2\right]}$$



Obrázek 4.1: Pravděpodobnostní funkce dvourozměrného normálního rozdělení

pro $-\infty < x < \infty, -\infty < y < \infty$, kde $\sigma_Y > 0, \sigma_X > 0, -\infty < E[X] < \infty, -\infty < E[Y] < \infty$ a $-1 \leq \rho \leq 1$ jsou konstanty. Pak náhodná veličina (X, Y) sleduje dvourozměrné náhodné rozdělení.

Příslušná pravděpodobnostní funkce je znázorněna na obrázku (4.1). Libovolný řez rovnoběžný s rovinou xy představuje eliptickou křivku, zatímco libovolný řez kolmý na rovinu xy představuje pravděpodobnostní funkci jednorozměrného normálního rozdělení. Pravděpodobnost, že bod (X, Y) leží v regionu R roviny xy , je definována jako

$$P[(X, Y) \in R] = \iint_R f(x, y) dy dx$$

Aby byla pravděpodobnostní funkce dvourozměrného normálního rozdělení pravděpodobnostní funkcí, musí nabývat pouze nezáporných hodnot a musí splňovat podmínku

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx = 1$$

Splnění prvního požadavku je patrné ze zblžného pohledu na funkční předpis tohoto pravděpodobnostního rozdělení. Abychom dokázali splnění druhé podmínky, použijeme nejprve substituce

$$u = \frac{x - E[X]}{\sigma_X} \quad v = \frac{y - E[Y]}{\sigma_Y}$$

čímž pravděpodobnostní funkci zjednodušíme do podoby

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)} dv du$$

Doplněním čtverce pro u v exponentu získáváme

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}[(u-\rho v)^2 + (1-\rho^2)v^2]} dv du$$

Jestliže dosadíme

$$w = \frac{u - \rho v}{\sqrt{1 - \rho^2}} \quad dw = \frac{du}{\sqrt{1 - \rho^2}}$$

lze výše uvedený integrál rozepsat jako součin dvou jednoduchých integrálů

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv$$

které jsou oba rovny jedné.

4.5.2 Momentová funkce a momenty

Věta 4.17 *Sdružená momentová funkce dvourozměrné normální náhodné veličiny (X, Y) je rovna*

$$m(t_1, t_2) = e^{t_1 E[X] + t_2 E[Y] + \frac{1}{2}(t_1^2 \sigma_X^2 + 2\rho t_1 t_2 \sigma_X \sigma_Y + t_2^2 \sigma_Y^2)}$$

Důkaz 4.13 *Opět dosadíme za x a y veličiny u a v , což vede k rovnici*

$$m(t_1, t_2) = e^{t_1 E[X] + t_2 E[Y]} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t_1 \sigma_X u + t_2 \sigma_Y v} \frac{1}{2\pi \sqrt{1 - \rho^2}} e^{-\frac{1}{2(1 - \rho^2)}(u^2 - 2\rho uv + v^2)} dv du$$

Oba exponenty lze sloučit do tvaru

$$-\frac{1}{2(1 - \rho^2)} [u^2 - 2\rho uv + v^2 - 2(1 - \rho^2)t_1 \sigma_X u - 2(1 - \rho^2)t_2 \sigma_Y v]$$

a doplněním čtverců pro u a v získáme

$$-\frac{1}{2(1 - \rho^2)} \left[(u - \rho v - (1 - \rho^2)t_1 \sigma_X)^2 + (1 - \rho^2)(v - \rho t_1 \sigma_X - t_2 \sigma_Y)^2 - (1 - \rho^2)(t_1^2 \sigma_X^2 + 2\rho t_1 t_2 \sigma_X \sigma_Y + t_2^2 \sigma_Y^2) \right]$$

Jestliže použijeme substituce

$$w = \frac{u - \rho v + (1 - \rho^2)t_1 \sigma_X}{\sqrt{1 - \rho^2}} \quad z = v - \rho t_1 \sigma_X - t_2 \sigma_Y$$

lze tento výraz dále upravit na

$$-\frac{1}{2}w^2 - \frac{1}{2}z^2 + \frac{1}{2}(t_1^2 \sigma_X^2 + 2\rho t_1 t_2 \sigma_X \sigma_Y + t_2^2 \sigma_Y^2)$$

Momentovou funkci pak lze vyjádřit jako

$$\begin{aligned} m(t_1, t_2) &= e^{t_1 E[X] + t_2 E[Y]} e^{\frac{1}{2}(t_1^2 \sigma_X^2 + 2\rho t_1 t_2 \sigma_X \sigma_Y + t_2^2 \sigma_Y^2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{w^2}{2} - \frac{z^2}{2}} dw dz \\ &= e^{t_1 E[X] + t_2 E[Y] + \frac{1}{2}(t_1^2 \sigma_X^2 + 2\rho t_1 t_2 \sigma_X \sigma_Y + t_2^2 \sigma_Y^2)} \end{aligned}$$

Věta 4.18 *Jestliže náhodná veličina (X, Y) sleduje dvourozměrné normální rozdělení, pak*

$$\sigma_{X,Y} = \rho \sigma_X \sigma_Y$$

Důkaz 4.14 Sdružený moment $E[X^r Y^s]$ lze získat z momentové funkce $m(t_1, t_2)$ derivací r -krát dle t_1 a s -krát dle t_2 s následným dosažením $t_1 = 0$ a $t_2 = 0$. Kovarianční koeficient X a Y je pak

$$\begin{aligned} E[X - E[X]]E[Y - E[Y]] &= E[XY - XE[X] - YE[Y] - E[X]E[Y]] = E[XY] - E[X]E[Y] \\ &= \frac{\partial^2}{\partial t_1 \partial t_2} m(t_1, t_2) \Big|_{t_1=t_2=0} - E[X]E[Y] = \rho \sigma_X \sigma_Y \end{aligned}$$

Věta 4.19 Jestliže má náhodná veličina (X, Y) dvourozměrné normální rozdělení, pak jsou X a Y nezávislé tehdy a jen tehdy, jsou-li X a Y nekorelované.

Důkaz 4.15 Náhodné veličiny jsou nekorelované tehdy a jen tehdy, je-li $\sigma_{X,Y} = 0$ nebo ekvivalentně je-li $\rho_{X,Y} = 0$. Z definice dvourozměrného normálního rozdělení je zřejmé, že je-li $\rho_{X,Y} = 0$, pak se pravděpodobnostní funkce $f_{X,Y}$ stane součinem dvou jednorozměrných normálních rozdělení. Proto $\rho_{X,Y}$ implikuje, že X a Y jsou nezávislé. Víme, že nezávislost náhodných veličin X a Y implikuje jejich nekorelovanost.

4.6 Marginální a podmíněná pravděpodobnostní funkce

Věta 4.20 Jestliže (X, Y) sleduje dvourozměrné náhodné rozdělení, pak marginální rozdělení X a Y jsou jednorozměrná normální rozdělení. To znamená, že X je normálně rozdělené se střední hodnotou $E[X]$ a rozptylem σ_X^2 a Y je normálně rozdělené se střední hodnotou $E[Y]$ a rozptylem σ_Y^2 .

Důkaz 4.16 Marginální rozdělení náhodné veličiny X je z definice

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

S použitím substituce $v = \frac{y - E[Y]}{\sigma_Y}$ a doplněním čtverců pro v získáme

$$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left(\frac{x-E[X]}{\sigma_X}\right)^2 - \frac{1}{2(1-\rho)^2}\left(v - \rho\frac{x-E[X]}{\sigma_X}\right)^2} dv$$

Následnými substitucemi

$$w = \frac{v - \rho(x - E[X])/\sigma_X}{\sqrt{1-\rho^2}} \quad dw = \frac{1}{\sqrt{1-\rho^2}} dv$$

lze dokázat

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{1}{2}\left(\frac{x-E[X]}{\sigma_X}\right)^2}$$

Analogicky lze důkaz provést také pro $f_Y(y)$.

Věta 4.21 Jestliže (X, Y) sleduje dvourozměrné normální rozdělení, pak podmíněná pravděpodobnostní funkce náhodné veličiny X pro $Y = y$ sleduje normální rozdělení se střední hodnotou $E[X] + \rho\frac{\sigma_X}{\sigma_Y}(y - E[Y])$ a rozptylem $\sigma_X^2(1 - \rho^2)$. Podobně podmíněná pravděpodobnostní funkce náhodné veličiny Y pro $X = x$ sleduje normální rozdělení se střední hodnotou $E[Y] + \rho\frac{\sigma_Y}{\sigma_X}(x - E[X])$ a rozptylem $\sigma_Y^2(1 - \rho^2)$.

Důkaz 4.17 Podmíněné pravděpodobnostní funkce lze odvodit s pomocí sdružené a marginální pravděpodobnostní funkce.

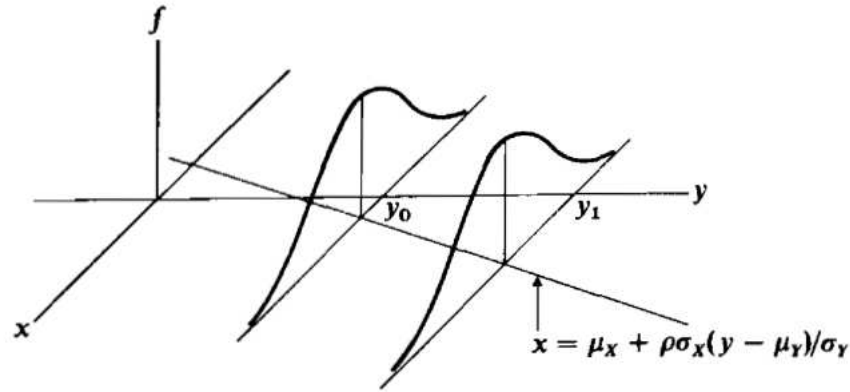
$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

Dosazením a následnými úpravami získáme

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi}\sigma_X\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_X^2(1-\rho^2)}\left[x-E[X]-\frac{\rho\sigma_X}{\sigma_Y}(y-E[Y])\right]^2}$$

což je jednorozměrné normální rozdělení se střední hodnotou $E[X] + \rho\frac{\sigma_X}{\sigma_Y}(y - E[Y])$ a rozptylem $\sigma_X^2(1 - \rho^2)$. Analogicky lze odvodit také $f_{Y|X}(y|x)$.

Již dříve jsme zmínili, že v případě podmíněného rozdělení nazýváme střední hodnotou náhodné veličiny regresní křivkou. Ve výše uvedeném případě je regrese X pro $Y = y$ rovna $E[X] + \rho\frac{\sigma_X}{\sigma_Y}(y - E[Y])$, což je lineární funkce veličiny y .



Obrázek 4.2: Regresní křivka pro podmíněné dvourozměrné normální rozdělení

Kapitola 5

Distribuce funkcí náhodných veličin

5.1 Střední hodnota a rozptyl funkcí náhodných veličin

5.1.1 Dvojití pojetí

Uvažujme funkci $g(\cdot, \dots, \cdot)$ a náhodné veličiny X_1, \dots, X_n . Definujme náhodnou veličinu $Y = g(X_1, \dots, X_n)$. Střední hodnotu náhodné veličiny Y lze vypočítat dvěma způsoby a to buď jako

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy$$

nebo jako

$$E[Y] = E[g(X_1, \dots, X_n)] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n$$

Na první pohled by se mohlo zdát, že je výhodnější řešit $E[Y]$ prvním způsobem, protože se tak vyhneme vícenásobnému intergrálu. Situaci však může značně zkomplikovat pravděpodobnostní funkce $f_Y(\cdot)$, jejíž znalost je nezbytným předpokladem pro zahájení výpočtu.

Příklad 5.1 *Nechť náhodná veličina X sleduje normované normální rozdělení a nechť $g(x) = x^2$. Střední hodnotu náhodné veličiny $Y = g(X)$ tak lze vypočítat jako*

$$E[Y] = \int_0^{\infty} y \frac{1}{\Gamma(1/2)} \left(\frac{1}{2}\right)^{\frac{1}{2}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} dy = 1$$

kde jsme využili skutečnost, že náhodná veličina Y sleduje gamma rozdělení s parametry $r = \frac{1}{2}$ a $\lambda = \frac{1}{2}$. Alternativně lze $E[Y]$ vypočítat ze vztahu

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1$$

5.1.2 Součet náhodných veličin

Věta 5.1 Pro náhodné veličiny X_1, \dots, X_n platí

$$E\left[\sum_1^n X_i\right] = \sum_1^n E[X_i]$$

a

$$D\left[\sum_1^n X_i\right] = \sum_1^n D[X_i] + 2 \sum_{i < j} \sigma_{X_i, X_j}$$

Důkaz 5.1 Dokažme tvrzení o rozptylu součtu náhodných veličin.

$$\begin{aligned} D\left[\sum_1^n X_i\right] &= E\left[\left(\sum_1^n X_i - E\left[\sum_1^n X_i\right]\right)^2\right] = E\left[\left(\sum_1^n (X_i - E[X_i])\right)^2\right] \\ &= E\left[\sum_{i=1}^n \sum_{j=1}^n (X_i - E[X_i])(X_j - E[X_j])\right] = \sum_{i=1}^n \sum_{j=1}^n E[(X_i - E[X_i])(X_j - E[X_j])] \\ &= \sum_{i=1}^n D[X_i] + 2 \sum_{i < j} \sigma_{X_i, X_j} \end{aligned}$$

Tvrzení 5.1 Jestliže jsou X_1, \dots, X_n nekorelované náhodné veličiny, pak

$$D\left[\sum_1^n X_i\right] = \sum_1^n D[X_i]$$

Kapitola 6

Výběry a jejich rozdělení

6.1 Výběr a populace

6.1.1 Definice základních pojmů

Definice 6.1 (Cílová populace) *Soubor všech prvků, které jsou předmětem našeho zájmu, nazýváme cílovou populací.*

Příklad 6.1 *V astronomii můžeme být cílovou populací soubor všech hvězd ve vesmíru a předmětem zájmu jejich hmotnost. V letectví mohou cílovou skupinu představovat všechny vyrobené letouny daného typu, kde zkoumáme životnost jejich draku.*

V praxi bývá velmi často nemožné nebo nežádoucí zjišťovat danou vlastnost na celé cílové populaci¹. Řešením je tak vybrat část populace a tu následně podrobit zkoumání - zjištění pak zobecníme na celou populaci.

Jestliže lze na každý prvek cílové populace nahlížet jako na náhodnou veličinu s danou pravděpodobnostní funkcí², pak lze pro tuto populaci definovat náhodný výběr.

Definice 6.2 (Náhodný výběr) *Nechť mají náhodné veličiny X_1, \dots, X_n sdruženou pravděpodobnostní funkci, pro kterou platí*

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f(x_1) \cdots f(x_n)$$

kde $f(\cdot)$ představuje pravděpodobnostní funkci, která je shodná pro všechna X_i . Pak definujeme X_1, \dots, X_n jako náhodný výběr velikosti n z populace s pravděpodobnostní funkcí $f(\cdot)$.

Náhodná veličina X_i reprezentuje hodnotu i -tého prvku výběru. Po té, co je výběr zanalyzován, nabývá náhodná veličina X_i konkrétní hodnoty x_i . Proto se někdy pod pojmem náhodný výběr rozumí x_1, \dots, x_n namísto X_1, \dots, X_n .

Velmi často není možné pořídit náhodný výběr z cílové populace. V těchto případech pak cílovou populaci nahrazujeme tzv. výběrovou populací.

¹Hvězd ve vesmíru je příliš mnoho na to, abychom je mohli všechny jednotlivě zkoumat. Testování odolnosti draku všech letounů by vedlo k jejich zničení, a proto postrádá smysl.

²Na hmotnost hvězdy a na životnost draku letadla lze zcela jistě pohlížet jako na náhodnou veličinu.

Definice 6.3 (Výběrová populace) *Nechť X_1, \dots, X_n je náhodný výběr z populace s pravděpodobnostní funkcí $f(\cdot)$. Tuto populaci pak nazýváme výběrovou populací.*

Závěry na základě náhodného výběru lze vztáhnout pouze na výběrovou populaci, nikoliv na cílovou populaci. Jedinou výjimkou je situace, kdy je cílová populace zároveň výběrovou populací.

Příklad 6.2 *Uvažujme marketingovou studii, jejímž cílem je zkoumat nákupní zvyky domácností dané země. Pokud budeme náhodný vzorek domácností vybírat z určitého regionu, pak tento region bude představovat naši výběrovou populaci. Naši cílovou populaci však budou domácnosti celé země. Je také zřejmé, že nákupní zvyklosti domácností se mohou mezi jednotlivými regiony lišit, a proto zobecnění z jednoho regionu na celou zemi může být problematické.*

V následujícím textu budeme někdy používat označení “populace $f(\cdot)$ ”, čímž budeme mít na mysli populaci, jejíž jednotlivé prvky mají pravděpodobnostní rozdělení $f(\cdot)$. Pokud budeme používat slovo populace bez přívlastku, budeme tím mít na mysli vždy výběrovou populaci.

6.1.2 Rozdělení výběru

Definice 6.4 (Rozdělení výběru) *Nechť X_1, \dots, X_n označují náhodný výběr velikosti n . Pak rozdělení výběru X_1, \dots, X_n je dáno sdruženým rozdělením náhodných veličin X_1, \dots, X_n .*

Definice 6.5 *Jestliže X_1, \dots, X_n představují náhodný výběr z populace $f(\cdot)$, pak pravděpodobnostní rozdělení tohoto výběru je dáno sdruženou pravděpodobnostní funkcí*

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f(x_1) \cdots f(x_n)$$

Příklad 6.3 *Uvažujme populaci, jejíž prvky sledují Bernoulliho rozdělení. Sdružená pravděpodobnostní funkce dvou prvkového náhodného výběru je tedy*

$$f_{X_1, X_2}(x_1, x_2) = f(x_1)f(x_2) = p^{x_1+x_2}q^{2-x_1-x_2}I_{(0,1)}(x_1)I_{(0,1)}(x_2)$$

Pravděpodobnostní funkce $f_{X_1, X_2}(x_1, x_2)$ z výše uvedeného příkladu definuje rozdělení uspořádaného výběru. Záleží tedy nejen na typu a počtu realizací, ale také na jejich uspořádání.

Dále je zřejmé, že naše definice náhodného výběru automaticky zahrnuje výběr bez vracení z konečné populace, protože jednotlivé “tahy” nejsou vzájemně nezávislé.

6.1.3 Statistiky a výběrové momenty

Definice 6.6 (Statistika) *Statistika je funkcí pozorovatelných náhodných veličin, a je tak sama o sobě pozorovatelnou náhodnou veličinou, která neobsahuje žádné neznámé parametry.*

Příklad 6.4 *Jestliže X_1, \dots, X_n představuje náhodný výběr z populace $f(\cdot, \theta)$, pak, za předpokladu, že X_1, \dots, X_n jsou pozorovatelné, jsou*

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

a

$$\frac{1}{2}(\min(X_1, \dots, X_n) + \max(X_1, \dots, X_n))$$

statistikami. Pokud $f(x, \theta) = \phi_{\theta,1}(x)$, kde θ je neznámý parametr, pak $\bar{X}_n - \theta$ není statistikou, právě proto, že závisí na neznámé θ .

V následující definici představíme jedny z nejdůležitějších statistik, tzv. výběrové momenty.

Definice 6.7 (Výběrové momenty) *Nechť X_1, \dots, X_n představuje náhodný výběr z populace $f(\cdot)$. Pak je r -tý obecný výběrový moment definován jako*

$$M'_r = \frac{1}{n} \sum_{i=1}^n X_i^r$$

Jestliže $r = 1$, pak získáme výběrový moment obvykle označovaný jako \bar{X} popř. jako \bar{X}_n .

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Podobně je r -tý centrální výběrový moment definován jako

$$M_r = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^r$$

Věta 6.1 *Nechť X_1, \dots, X_n představují náhodný výběr z populace $f(\cdot)$. Očekávaná hodnota r -tého výběrového momentu M'_r je rovna r -tému momentu populace μ'_r za předpokladu existence μ'_r .*

$$E[M'_r] = \mu'_r$$

Dále také platí

$$D[M'_r] = \frac{1}{n} (E[X^{2r}] - E[X^r]^2) = \frac{1}{n} [\mu'_{2r} - (\mu'_r)^2]$$

za předpokladu existence μ'_{2r} .

Důkaz 6.1

$$E[M'_r] = E\left[\frac{1}{n} \sum_{i=1}^n X_i^r\right] = \frac{1}{n} E\left[\sum_{i=1}^n X_i^r\right] = \frac{1}{n} \sum_{i=1}^n E[X_i^r] = \frac{1}{n} \sum_{i=1}^n \mu'_r = \mu'_r$$

$$\begin{aligned} D[M'_r] &= D\left[\frac{1}{n} \sum_{i=1}^n X_i^r\right] = \left(\frac{1}{n}\right)^2 D\left[\sum_{i=1}^n X_i^r\right] = \left(\frac{1}{n}\right)^2 \sum_{i=1}^n (E[X_i^{2r}] - E[X_i^r]^2) \\ &= \frac{1}{n} (E[X^{2r}] - E[X^r]^2) = \frac{1}{n} (\mu'_{2r} - (\mu'_r)^2) \end{aligned}$$

Tvrzení 6.1 *Nechť X_1, \dots, X_n představuje náhodný výběr z populace $f(\cdot)$ a $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ představuje výběrovou střední hodnotu. Pak platí*

$$\begin{aligned} E[\bar{X}_n] &= \mu \\ D[\bar{X}_n] &= \frac{1}{n} \sigma^2 \end{aligned}$$

kde μ a σ^2 jsou střední hodnota a rozptyl pravděpodobnostního rozdělení $f(\cdot)$.

Definice 6.8 (Výběrový rozptyl) *Nechť X_1, \dots, X_n představují náhodný výběr z populace $f(\cdot)$. Výběrový rozptyl je pak definován jako*

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

pro $n > 1$.

Výběrový rozptyl je definován jako $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ namísto druhého výběrového momentu $M_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, protože střední hodnota S_n^2 je rovna rozptylu populace.

Tvrzení 6.2

$$S_n^2 = \frac{1}{2n(n-1)} \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j)^2$$

Důkaz 6.2

$$\begin{aligned} \frac{1}{2n(n-1)} \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j)^2 &= \frac{1}{2n(n-1)} \sum_{i=1}^n \left(nX_i^2 - 2X_i \sum_{j=1}^n X_j + \sum_{j=1}^n X_j^2 \right) \\ &= \frac{1}{2n(n-1)} \sum_{i=1}^n \left(nX_i^2 - 2n\bar{X}X_i + n\bar{X}^2 \right) = \frac{1}{2(n-1)} \left(n\bar{X}^2 - 2\bar{X}^2 + n\bar{X}^2 \right) \\ &= \frac{1}{n-1} (n\bar{X}^2 - \bar{X}^2) = \frac{1}{n-1} nD[X] = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \end{aligned}$$

Věta 6.2 *Nechť X_1, \dots, X_n představuje náhodný výběr z populace $f(\cdot)$ a necht $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Pak platí*

$$\begin{aligned} E[S_n^2] &= \sigma^2 \\ D[S_n^2] &= \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right) \end{aligned}$$

pro $n > 1$.

Důkaz 6.3 *Nejprve dokažme tvrzení o střední hodnotě výběrového rozptylu. Připomeňme, že platí $\sigma^2 = E[(X - \mu)^2]$ a $\mu_r = E[(X - \mu)^r]$. Jádro důkazu spočívá v identitě*

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$$

Platnost této identity lze dokázat následujícím způsobem.

$$\begin{aligned} \sum (X_i - \mu)^2 &= \sum (X_i - \bar{X} + \bar{X} - \mu)^2 = \sum [(X_i - \bar{X}) + (\bar{X} - \mu)]^2 \\ &= \sum [(X_i - \bar{X})^2 + 2(X_i - \bar{X})(\bar{X} - \mu) + (\bar{X} - \mu)^2] \\ &= \sum (X_i - \bar{X})^2 + 2(\bar{X} - \mu) \sum (X_i - \bar{X}) + n(\bar{X} - \mu)^2 = \sum (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 \end{aligned}$$

S využitím této identity pak lze odvodit

$$\begin{aligned} E[S_n^2] &= E\left[\frac{1}{n-1} \sum (X_i - \bar{X})^2\right] = \frac{1}{n-1} E\left[\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2\right] \\ &= \frac{1}{n-1} \left\{ \sum_{i=1}^n E[(X_i - \mu)^2] - nE[(\bar{X} - \mu)^2] \right\} = \frac{1}{n-1} \left\{ \sum_{i=1}^n \sigma^2 - nD[\bar{X}] \right\} \\ &= \frac{1}{n-1} \left(n\sigma^2 - n\frac{\sigma^2}{n} \right) = \sigma^2 \end{aligned}$$

Důkaz rozptylu S_n^2 lze provést s využitím výše odvozené identity a vztahu

$$\bar{X} - \mu = \frac{1}{n} \sum X_i - \frac{1}{n} n\mu = \frac{1}{n} \sum X_i - \frac{1}{n} \sum \mu = \frac{1}{n} \sum (X_i - \mu)$$

Nicméně odvození je poměrně zdlouhavé, a proto jej vynecháme.

6.2 Výběrová střední hodnota

6.2.1 Střední hodnota a rozptyl

Věta 6.3 *Nechť X_1, \dots, X_n představuje náhodný výběr z populace $f(\cdot)$ se střední hodnotou μ a konečným rozptylem σ^2 . Definujme $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Pak platí*

$$E[\bar{X}] = \mu_{\bar{X}} = \mu \quad D[\bar{X}] = \sigma_{\bar{X}^2} = \frac{1}{n} \sigma^2$$

Tento teorém je v podstatě pouze reformulací tvrzení (6.1).

Výběrová střední hodnota \bar{X} má charakter náhodné veličiny. Pokud bychom ji použili jako odhad pro μ , pak platí, že její střední hodnota odpovídá hledané střední hodnotě populace μ a že její rozptyl klesá s rostoucí velikostí náhodného výběru. Tímto se dostáváme k zákonu velkých čísel.

6.2.2 Zákon velkých čísel

Označme střední hodnotu náhodné veličiny X s pravděpodobnostní funkcí $f(\cdot)$ jako μ . Předpokládejme, že předmětem našeho zájmu je odhad μ .

Na μ lze nahlížet jako na průměr nekonečného počtu hodnot náhodné veličiny X . Je zřejmé, že reálném životě nejsme schopni průměrovat nekonečný počet hodnot, a proto musíme informace o μ získat na základě náhodného výběru z populace $f(\cdot)$ velikosti n . To nám umožňuje tzv. slabý zákon velkých čísel.

Věta 6.4 (Slabý zákon velkých čísel) *Nechť $f(\cdot)$ představuje pravděpodobnostní funkci se střední hodnotou μ a rozptylem σ^2 . Dále nechť \bar{X}_n představuje střední hodnotu náhodného výběru velikosti n z populace $f(\cdot)$. Uvažujme dvě čísla $\epsilon > 0$ a $0 < \delta < 1$. Jestliže n je přirozené číslo takové, že $n > \frac{\sigma^2}{\epsilon^2} \delta$, pak*

$$P[-\epsilon < \bar{X}_n - \mu < \epsilon] \geq 1 - \delta$$

Důkaz 6.4 *Dle Chebyshevovy nerovnosti platí*

$$P[g(X) \geq k] \leq \frac{E[g(X)]}{k}$$

pro každé $k > 0$, libovolnou náhodnou veličinu X a nezápornou funkci $g(\cdot)$. Tato nerovnost je ekvivalentní s

$$P[g(X) < k] \geq 1 - \frac{E[g(X)]}{k}$$

Nechť $g(X) = (\bar{X}_n - \mu)^2$ a $k = \epsilon^2$. Pak

$$\begin{aligned} P[-\epsilon < \bar{X}_n - \mu < \epsilon] &= P[|\bar{X}_n - \mu| < \epsilon] \\ &= P[|\bar{X}_n - \mu|^2 < \epsilon^2] \geq 1 - \frac{E[(\bar{X}_n - \mu)^2]}{\epsilon^2} \\ &= 1 - \frac{\frac{1}{n}\sigma^2}{\epsilon^2} \geq 1 - \delta \end{aligned}$$

pro $\delta > \frac{\sigma^2}{n\epsilon^2}$ popř. $n > \frac{\sigma^2}{\epsilon^2}\delta$.

Příklad 6.5 Uvažujme pravděpodobnostní rozdělení s neznámou střední hodnotou a rozptylem 1. Jak velký musí být náhodný výběr, aby se jeho střední hodnota \bar{X}_n nacházela ve vzdálenosti 0.5 od střední hodnoty populace s pravděpodobností alespoň 0.95?

Dosaďme $\sigma^2 = 1, \epsilon = 0.5$ a $\delta = 0.05$. Minimální velikost n náhodného výběru je dána vztahem

$$n > \frac{\sigma^2}{\delta\epsilon} = \frac{1}{0.05 \cdot (0.5)^2} = 80$$

6.2.3 Centrální limitní věta

S centrální limitní větou jsme se již seznámili v předchozí kapitole. Nyní se k ní opět vrátíme a doplníme ji o důkaz.

Věta 6.5 (Centrální limitní věta) Nechť $f(\cdot)$ představuje pravděpodobnostní funkci se střední hodnotou μ a konečným rozptylem σ^2 . Dále nechť \bar{X}_n představuje střední hodnotu náhodného výběru velikosti n z populace $f(\cdot)$. Definujme náhodnou veličinu

$$Z_n = \frac{\bar{X}_n - E[\bar{X}_n]}{\sqrt{D[\bar{X}_n]}} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Pravděpodobnostní rozdělení náhodné veličiny Z_n se limitně blíží normovanému normálnímu rozdělení s tím, jak se n blíží nekonečnu.

Nejpozoruhodnější skutečností na centrální limitní větě je, že platí pro všechny pravděpodobnostní funkce $f(\cdot)$. Centrální limitní věta implikuje, že pravděpodobnostní funkce výběrové střední hodnoty \bar{X}_n se asymptoticky blíží normálnímu rozdělení se střední hodnotou μ a rozptylem $\frac{\sigma^2}{n}$ bez ohledu na výchozí pravděpodobnostní funkci $f(\cdot)$.

Důkaz 6.5 Centrální limitní větu nebudeme dokazovat v její obecné formě. Zaměříme se na situaci, kdy $f(\cdot)$ má momentovou funkci. Připomeňme, že momentová funkce normovaného normálního rozdělení je $m(t) = e^{\frac{1}{2}t^2}$. Pro momentovou funkci $m_{Z_n}(t)$ náhodné veličiny Z_n se pak pokusíme dokázat, že se asymptoticky blíží k $m(t)$.

Momentová funkce $m_{Z_n}(t)$ je definována jako

$$\begin{aligned} m_{Z_n}(t) &= E[e^{tZ_n}] = E\left[e^{t\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}\right] = E\left[e^{\frac{t}{n}\sum \frac{X_i-\mu}{\sigma/\sqrt{n}}}\right] \\ &= E\left[\prod_{i=1}^n e^{\frac{t}{n}\frac{X_i-\mu}{\sigma/\sqrt{n}}}\right] = \prod_{i=1}^n E\left[e^{\frac{t}{\sqrt{n}}\frac{X_i-\mu}{\sigma}}\right] \end{aligned}$$

kde jsme v posledním kroku využili nezávislosti náhodných veličin X_1, \dots, X_n . Definujme $Y_i = \frac{X_i-\mu}{\sigma}$. Momentová funkce náhodné veličiny Y_i je nezávislá na i , protože všechna Y_i mají shodné pravděpodobnostní rozdělení. Proto namísto $m_{Y_i}(t)$ budeme používat značení $m_Y(t)$. Platí

$$\begin{aligned} \prod_{i=1}^n E\left[e^{\frac{t}{\sqrt{n}}\frac{X_i-\mu}{\sigma}}\right] &= \prod_{i=1}^n E\left[e^{\frac{t}{\sqrt{n}}Y_i}\right] \\ &= \prod_{i=1}^n m_Y\left(\frac{t}{\sqrt{n}}\right) = \left[m_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n \end{aligned}$$

a proto

$$m_{Z_n}(t) = \left[m_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

r -tou derivací $m_Y(t/\sqrt{n})$ v bodě $t = 0$ získáme r -tý moment okolo střední hodnoty pro pravděpodobnostní funkci $f(\cdot)$ vydělený členem $(\sigma\sqrt{n})^r$. S pomocí Taylorova rozvoje tak lze $m_Y(t/\sqrt{n})$ vyjádřit jako

$$m_Y\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{\mu_1}{\sigma} \frac{t}{\sqrt{n}} + \frac{1}{2!} \frac{\mu_2}{\sigma^2} \left(\frac{t}{\sqrt{n}}\right)^2 + \frac{1}{3!} \frac{\mu_3}{\sigma^3} \left(\frac{t}{\sqrt{n}}\right)^3 + \dots$$

a protože $\mu_1 = 0$ a $\mu_2 = \sigma^2$, lze tento výraz dále upravit na

$$m_Y\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{1}{n} \left(\frac{1}{2}t^2 + \frac{1}{3!\sqrt{n}} \frac{\mu_3}{\sigma^3} t^3 + \frac{1}{4!n} \frac{\mu_4}{\sigma^4} t^4 + \dots\right)$$

Protože $\lim_{n \rightarrow \infty} \left(1 + \frac{u}{n}\right)^n = e^{\frac{1}{2}t^2}$, kde

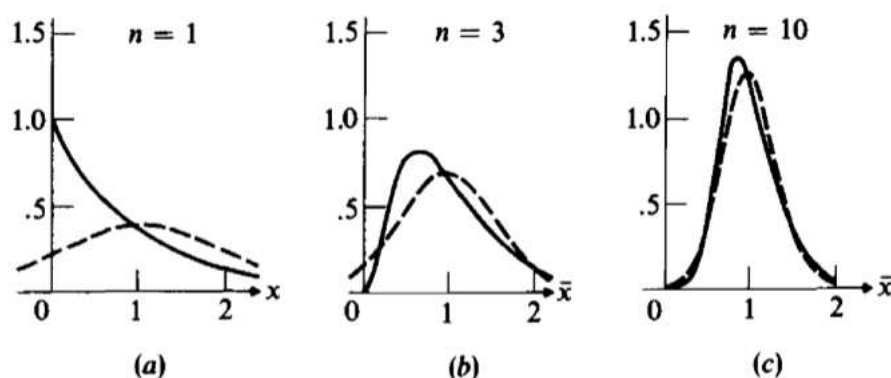
$$u = \frac{1}{2}t^2 + \frac{1}{3!\sqrt{n}} \frac{\mu_3}{\sigma^3} t^3 + \frac{1}{4!n} \frac{\mu_4}{\sigma^4} t^4 + \dots$$

představuje výraz v závorkách ve výše uvedené rovnici. Tím se dostáváme k $\lim_{n \rightarrow \infty} m_{Z_n}(t) = e^{\frac{1}{2}t^2}$. Proto má Z_n v limitě stejnou momentovou funkci jako normalizované normální rozdělení. Připomeňme, že pokud mají dvě náhodné veličiny shodnou momentovou funkci, pak sledují shodné pravděpodobnostní rozdělení.

Je zřejmé, že míra aproximace závisí na velikosti náhodného výběru a konkrétní pravděpodobnostní funkci $f(\cdot)$. Situace pro $f(x) = e^{-x}I_{(0,\infty)}(x)$ pro $n = 1, 3, 10$ je ilustrována obrázkem (6.1), kde souvislá čára představuje skutečné rozdělení a přerušovaná čára pak aproximaci normálním rozdělením.

Jak je z obrázku (6.1) patrné, aproximace se poměrně rychle zlepšuje s rostoucí velikostí náhodného výběru. Obrázek se zaměřuje na konvergenci v okolí střední hodnoty μ . Co z něj však patrné není, je konvergence v okolí chvostů, kde je míra konvergence výrazně pomalejší.

Následující kapitoly se zaměřují na pravděpodobnostní rozdělení střední hodnoty výběru pro konkrétní pravděpodobnostní funkci $f(\cdot)$.



Obrázek 6.1: Konvergence k normálnímu rozdělení

6.2.4 Bernoulliho rozdělení

Jestliže X_1, \dots, X_n představují náhodný výběr z Bernoulliho rozdělení, lze nalézt přesnou formu pravděpodobnostního rozdělení náhodné veličiny \bar{X}_n . Pravděpodobnostní funkce, ze které provádíme náhodný výběr, má formu

$$f(x) = p^x(1-p)^{1-x}I_{(0,1)}(x)$$

Z příkladu (5.7) víme, že $\sum_{i=1}^n X_i$ sleduje binomické rozdělení.

$$P\left[\sum_{i=1}^n X_i = k\right] = \binom{n}{k} p^k q^{n-k} I_{\{0,1,\dots,n\}}(k)$$

Proto je pravděpodobnostní rozdělení \bar{X}_n dáno vztahem

$$P\left[\bar{X}_n = \frac{k}{n}\right] = \binom{n}{k} p^k q^{n-k} \quad \text{pro } k = 0, 1, \dots, n$$

Proto náhodná veličina \bar{X}_n nabývá v případě Bernoulliho rozdělení hodnot $0, 1/n, \dots, 1$ s pravděpodobnostmi $\binom{n}{0}p^0q^n, \binom{n}{1}p^1q^{n-1}, \binom{n}{2}p^2q^{n-2}, \dots, \binom{n}{n}p^nq^0$.

6.2.5 Poissonovo rozdělení

Jestliže X_1, \dots, X_n představují náhodný výběr z Poissonova rozdělení se střední hodnotou λ , pak z příkladu (5.8) vyplývá, že $\sum_{i=1}^n X_i$ sleduje Poissonovo rozdělení s parametrem $n\lambda$. Proto platí

$$P\left[\bar{X}_n = \frac{k}{n}\right] = P\left[\sum_{i=1}^n X_i = k\right] = \frac{e^{-n\lambda}(n\lambda)^k}{k!} \quad \text{pro } k = 0, 1, 2, \dots$$

6.2.6 Exponenciální rozdělení

Nechť X_1, \dots, X_n představují náhodný výběr z exponenciálního rozdělení

$$f(x) = \theta e^{-\theta x} I_{(0,\infty)}(x)$$

Z příkladu (5.9) víme, že $\sum_{i=1}^n$ má gamma rozdělení s parametry n a θ . Proto platí

$$f_{\sum X_i}(z) = \frac{1}{\Gamma(n)} z^{n-1} \theta^n e^{-\theta z} I_{(0,\infty)}(z)$$

neboli

$$P[\sum X_i \leq y] = \int_0^y \frac{1}{\Gamma(n)} z^{n-1} \theta^n e^{-\theta z} dz \quad \text{pro } y > 0$$

což lze upravit do tvaru

$$\begin{aligned} P\left[\bar{X}_n \leq \frac{y}{n}\right] &= \int_0^y \frac{1}{\Gamma(n)} z^{n-1} \theta^n e^{-\theta z} dz \\ P[\bar{X}_n \leq x] &= \int_0^{nx} \frac{1}{\Gamma(n)} z^{n-1} \theta^n e^{-\theta z} dz = \int_0^x \frac{1}{\Gamma(n)} (nu)^{n-1} \theta^n e^{-n\theta u} n du \end{aligned}$$

To znamená, že \bar{X}_n sleduje gamma rozdělení s parametry n a $n\theta$.

6.2.7 Uniformní rozdělení

Nechť X_1, \dots, X_n představuje náhodný výběr z uniformního rozdělení nad intervalem $(0, 1]$. Pravděpodobnostní funkce náhodné veličiny \bar{X}_n je dána vztahem

$$\begin{aligned} f_{\bar{X}_n}(x) &= \sum_{k=0}^{n-1} \frac{n}{(n-1)!} \left[(nx)^{n-1} - \binom{n}{1} (nx-1)^{n-1} + \binom{n}{2} (nx-2)^{n-1} - \dots \right. \\ &\quad \left. + (-1)^k \binom{n}{k} (nx-k)^{n-1} \right] I_{(k/n, (k+1)/n]}(x) \end{aligned}$$

Obecné odvození výše uvedeného vztahu je založené na indukci a konvoluci a je poněkud zdlouhavé, a proto jej vynecháme.

6.2.8 Cauchyho rozdělení

Nechť X_1, \dots, X_n představuje náhodný výběr z Cauchyho rozdělení a

$$f(x) = \frac{1}{\pi\beta\{1 + [(x-\alpha)/\beta]^2\}}$$

Potom sleduje \bar{X}_n stejné Cauchyho rozdělení pro libovolné n . To znamená, že střední hodnota výběru má stejné pravděpodobnostní rozdělení jako libovolný prvek tohoto výběru.

Momentová funkce Cauchyho rozdělení neexistuje a důkaz pomocí indukce a konjunkce vede ke složitým integrálům. Nejjednodušším způsobem je tak důkaz pomocí tzv. charakteristické funkce, která je zobecněním momentové funkce. Charakteristická funkce má oproti momentové funkci tu výhodu, že vždy existuje. Samotný důkaz se pak opírá o skutečnost, že součin charakteristických funkcí nezávislých náhodných veličin se shodným pravděpodobnostním rozdělením je charakteristickou funkcí součtu těchto náhodných veličin.

6.3 Výběr z normálního rozdělení

6.3.1 Střední hodnota výběru

Věta 6.6 *Nechť \bar{X}_n označuje střední hodnotu náhodného výběru velikosti n z normálního rozdělení se střední hodnotou μ a rozptylem σ^2 . Pak má \bar{X}_n normální rozdělení se střední hodnotou μ a rozptylem σ^2/n .*

Důkaz 6.6

$$\begin{aligned} m_{\bar{X}_n}(t) &= E[e^{t\bar{X}_n}] = E\left[e^{\frac{t\sum X_i}{n}}\right] = E\left[\prod_{i=1}^n e^{\frac{tX_i}{n}}\right] = \prod_{i=1}^n E\left[e^{\frac{tX_i}{n}}\right] \\ &= \prod_{i=1}^n m_i\left(\frac{t}{n}\right) = \prod_{i=1}^n e^{\frac{\mu t}{n} + \frac{1}{2}\left(\frac{\sigma t}{n}\right)^2} = e^{\mu t + \frac{1}{2}\frac{(\sigma t)^2}{n}} \end{aligned}$$

$e^{\mu t + \frac{1}{2}\frac{(\sigma t)^2}{n}}$ je momentová funkce normálního rozdělení se střední hodnotou μ a rozptylem σ^2/n , čímž uzavíráme důkaz.

6.3.2 Chi-kvadrát rozdělení

Definice 6.9 (Chi-kvadrát rozdělení) *Jestliže je X náhodná veličina s pravděpodobnostní funkcí*

$$f_X(x) = \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{\frac{k}{2}} x^{\frac{k}{2}-1} e^{-\frac{1}{2}x} I_{(0,\infty)}(x)$$

pak tato náhodná veličina sleduje chi-kvadrát rozdělení s k stupni volnosti, kde k je přirozené číslo.

Tvrzení 6.3 *Chi-kvadrát rozdělení je specifickým případem gamma rozdělení s parametry $r = \frac{k}{2}$ a $\lambda = \frac{1}{2}$. Proto pro náhodnou veličinu X , která sleduje chi-kvadrát rozdělení, platí*

$$\begin{aligned} E[X] &= k \\ D[X] &= 2k \\ m_X(t) &= \left(\frac{1}{1-2t}\right)^{\frac{k}{2}}, \quad \text{pro } t < \frac{1}{2} \end{aligned}$$

Věta 6.7 *Jestliže náhodné veličiny X_i pro $i = 1, 2, \dots, k$ jsou nezávislé a sledují normální rozdělení se střední hodnotou μ_i a rozptylem σ_i^2 , pak náhodná veličina*

$$U = \sum_{i=1}^k \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2$$

sleduje chi-kvadrát rozdělení s k stupni volnosti.

Důkaz 6.7 *Definujme $Z_i = \frac{X_i - \mu_i}{\sigma_i}$. Pak Z_i sleduje normalizované normální rozdělení. Její momentová funkce je definována jako*

$$m_U(t) = E[e^{tU}] = E[e^{t\sum Z_i^2}] = E\left[\prod_{i=1}^n e^{tZ_i^2}\right] = \prod_{i=1}^k E\left[e^{tZ_i^2}\right]$$

Zároveň však platí

$$\begin{aligned} E \left[e^{tZ_i^2} \right] &= \int_{-\infty}^{\infty} e^{tz^2} \left(\frac{1}{\sqrt{2\pi}} \right) e^{-\frac{1}{2}z^2} dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-2t)z^2} dz \\ &= \frac{1}{\sqrt{1-2t}} \int_{-\infty}^{\infty} \frac{\sqrt{1-2t}}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-2t)z^2} dz = \frac{1}{\sqrt{1-2t}} \end{aligned}$$

pro $t < \frac{1}{2}$, kde je poslední integrál roven jedné, protože představuje plochu pod křivkou normálního rozdělení s rozptylem $\frac{1}{\sqrt{1-2t}}$.

$$\prod_{i=1}^k E \left[e^{tZ_i^2} \right] = \prod_{i=1}^k \frac{1}{\sqrt{1-2t}} = \left(\frac{1}{1-2t} \right)^{\frac{k}{2}}$$

tak představuje momentovou funkci chi-kvadrát rozdělení s k stupni volnosti.

Tvrzení 6.4 Jestliže X_1, \dots, X_n představuje náhodný výběr z normálního rozdělení se střední hodnotou μ a rozptylem σ^2 , pak $U = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2}$ sleduje chi-kvadrát rozdělení s n stupni volnosti.

Jestliže je μ nebo σ^2 neznámé, pak není náhodná veličina U statická. Jestliže je však μ známé a σ^2 neznámé, lze odhadnout σ^2 pomocí $\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ a rozdělení $\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ odvodit na základě výše uvedeného tvrzení. To je možné, protože platí

$$E \left[\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \right] = \frac{1}{n} \sum_{i=1}^n E[(X_i - \mu)^2] = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \sigma^2$$

Věta 6.8 Jestliže Z_1, \dots, Z_n představují náhodný výběr z normovaného normálního rozdělení, pak

1. \bar{Z} sleduje normální rozdělení s nulovou střední hodnotou a rozptylem $\frac{1}{n}$.
2. \bar{Z} a $\sum_{i=1}^n (Z_i - \bar{Z})^2$ jsou nezávislé.
3. $\sum_{i=1}^n (Z_i - \bar{Z})^2$ sleduje chi-kvadrát rozdělení s $n - 1$ stupni volnosti.

Důkaz 6.8 Bod (1) vychází z věty (6.6).

Bod (2) dokážeme pro $n = 2$. Jestliže $n = 2$, pak $\bar{Z} = \frac{Z_1 + Z_2}{2}$ a

$$\begin{aligned} \sum (Z_i - \bar{Z})^2 &= \left(Z_1 - \frac{Z_1 + Z_2}{2} \right)^2 + \left(Z_2 - \frac{Z_1 + Z_2}{2} \right)^2 \\ &= \frac{(Z_1 - Z_2)^2}{4} + \frac{(Z_2 - Z_1)^2}{4} = \frac{(Z_2 - Z_1)^2}{2} \end{aligned}$$

takže \bar{Z} je funkcí $Z_1 + Z_2$ a $\sum (Z_i - \bar{Z})^2$ je funkcí $Z_2 - Z_1$. Abychom prokázali, že \bar{Z} a $\sum (Z_i - \bar{Z})^2$ jsou nezávislé, stačí prokázat nezávislost $Z_1 + Z_2$ a $Z_2 - Z_1$. Pro jejich momentové funkce platí

$$m_{Z_1+Z_2}(t_1) = E \left[e^{t_1(Z_1+Z_2)} \right] = E[e^{t_1 Z_1} e^{t_1 Z_2}] = E[e^{t_1 Z_1}] E[e^{t_1 Z_2}] = e^{\frac{1}{2}t_1^2} e^{\frac{1}{2}t_1^2} = e^{t_1^2}$$

a podobně

$$m_{Z_2-Z_1}(t_2) = e^{t_2^2}$$

Dále také platí

$$\begin{aligned} m_{Z_1+Z_2, Z_2-Z_1}(t_1, t_2) &= E \left[e^{t_1(Z_1+Z_2)+t_2(Z_2-Z_1)} \right] = E \left[e^{(t_1-t_2)Z_1} e^{(t_1+t_2)Z_2} \right] \\ &= E \left[e^{(t_1-t_2)Z_1} \right] E \left[e^{(t_1+t_2)Z_2} \right] = e^{\frac{1}{2}(t_1-t_2)^2} e^{\frac{1}{2}(t_1+t_2)^2} = e^{t_1^2} e^{t_2^2} = m_{Z_1+Z_2}(t_1) m_{Z_2-Z_1}(t_2) \end{aligned}$$

což znamená, že $Z_1 + Z_2$ a $Z_2 - Z_1$ jsou nezávislé.

Abychom dokázali bod (3) využijeme nezávislosti \bar{Z} a $\sum_{i=1}^n (Z_i - \bar{Z})^2$, což jsme prokázali výše. Všimněme si, že

$$\begin{aligned} \sum Z_i^2 &= \sum (Z_i - \bar{Z} + \bar{Z})^2 = \sum (Z_i - \bar{Z})^2 + 2\bar{Z} \sum (Z_i - \bar{Z}) + \sum \bar{Z}^2 \\ &= \sum (Z_i - \bar{Z})^2 + n\bar{Z}^2 \end{aligned}$$

Proto platí

$$m_{\sum Z_i^2}(t) = m_{\sum (Z_i - \bar{Z})^2}(t) m_{n\bar{Z}^2}(t)$$

což implikuje

$$m_{\sum (Z_i - \bar{Z})^2}(t) = \frac{m_{\sum Z_i^2}(t)}{m_{n\bar{Z}^2}(t)} = \frac{\left(\frac{1}{1-2t}\right)^{\frac{n}{2}}}{\left(\frac{1}{1-2t}\right)^{\frac{1}{2}}} = \left(\frac{1}{1-2t}\right)^{\frac{n-1}{2}}$$

pro $t < \frac{1}{2}$. Protože náhodná veličina $\sqrt{n}\bar{Z}$ sleduje normované normální rozdělení, sleduje $n\bar{Z}^2$ chi-kvadrát rozdělení s jedním stupněm volnosti, čímž jsme dokázali, že momentová funkce náhodné veličiny $\sum (Z_i - \bar{Z})^2$ je momentovou funkcí chi-kvadrát rozdělení s $n - 1$ stupni volnosti.

Věta 6.9 Předchozí věta předpokládala, že náhodný výběr pochází z normovaného normálního rozdělení. Pokud bychom namísto toho předpokládali náhodný výběr z normálního rozdělení se střední hodnotou μ a rozptylem σ^2 , pak $Z_i = \frac{X_i - \mu}{\sigma}$. Výše uvedená věta se tak modifikuje do tvaru

1. $\bar{Z} = \frac{1}{n} \sum \frac{X_i - \mu}{\sigma} = \frac{\bar{X} - \mu}{\sigma}$ sleduje normální rozdělení nulovou střední hodnotou a rozptylem $\frac{1}{n}$.
2. $\bar{Z} = \frac{\bar{X} - \mu}{\sigma}$ a $\sum (Z_i - \bar{Z})^2 = \sum \left(\frac{X_i - \mu}{\sigma} - \frac{\bar{X} - \mu}{\sigma} \right)^2$ jsou nezávislé, což implikuje nezávislost mezi \bar{X} a $\sum (X_i - \bar{X})^2$.
3. $\sum (Z_i - \bar{Z})^2 = \sum \frac{(X_i - \bar{X})^2}{\sigma^2}$ sleduje chi-kvadrát rozdělení s $n - 1$ stupni volnosti.

Tvrzení 6.5 Jestliže $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ představuje rozptyl náhodného výběru z normálního rozdělení se střední hodnotou μ a rozptylem σ^2 , pak

$$U = \frac{(n-1)S_n^2}{\sigma^2}$$

sleduje chi-kvadrát rozdělení s $n - 1$ stupni volnosti.

Protože S_n^2 je lineární funkcí náhodné veličiny U , lze pravděpodobnostní rozdělení S_n^2 získat z pravděpodobnostního rozdělení U .

$$f_{S_n^2}(y) = \left(\frac{n-1}{2\sigma^2}\right)^{\frac{n-1}{2}} \frac{1}{\Gamma\left(\frac{n-1}{2}\right)} y^{\frac{n-3}{2}} e^{-\frac{(n-1)y}{2\sigma^2}} I_{(0,\infty)}(y)$$

Pojem “stupně volnosti” lze chápat jako počet nezávislých členů v součtu. Např. součet ve větě (6.7) má k nezávislých členů, zatímco součet v bodě (3) věty (6.8) má pouze $n-1$ nezávislých členů. To proto, že vztah $\sum (Z_i - \bar{Z}) = 0$ umožňuje výpočet $Z_i - \bar{Z}$ pro libovolné i za předpokladu znalosti zbývajících $n-1$ členů.

6.3.3 F rozdělení

Uvažujme dvě nezávislé náhodné veličiny U a V , které sledují chi-kvadrát rozdělení s m resp. n stupni volnosti. Jejich sdružená pravděpodobnostní funkce je pak dána rovnicí

$$f_{U,V}(u, v) = \frac{1}{\Gamma(m/2)\Gamma(n/2)2^{(m+n)/2}} u^{\frac{m-2}{2}} v^{\frac{n-2}{2}} e^{-\frac{1}{2}(u+v)} I_{(0,\infty)}(u) I_{(0,\infty)}(v)$$

Předmětem našeho zájmu je pravděpodobnostní rozdělení podílu

$$X = \frac{U/m}{V/n}$$

V prvním kroku provedeme transformaci $X = \frac{U/m}{V/n}$ a $Y = V$, nalezneme sdruženou pravděpodobnostní funkci X a Y a následně integrací získáme marginální pravděpodobnostní funkci náhodné veličiny X .

Jakobián výše uvedené transformace je $\frac{m}{n}y$, proto

$$f_{X,Y}(x, y) = \frac{m}{n} y \frac{1}{\Gamma(m/2)\Gamma(n/2)2^{(m+n)/2}} \left(\frac{m}{n}xy\right)^{\frac{m-2}{2}} y^{\frac{n-2}{2}} e^{-\frac{1}{2}\left(\frac{m}{n}xy+y\right)}$$

Následnou integrací pak získáme

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{X,Y}(x, y) dy \\ &= \frac{1}{\Gamma(m/2)\Gamma(n/2)2^{(m+n)/2}} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m-2}{2}} \int_0^\infty y^{\frac{m+n-2}{2}} e^{-\frac{1}{2}\left(\frac{m}{n}x+1\right)y} dy \\ &= \frac{\Gamma((m+2)/2)}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{x^{\frac{m-2}{2}}}{\left(1 + \frac{m}{n}x\right)^{(m+n)/2}} I_{(0,\infty)}(x) \end{aligned}$$

Definice 6.10 (F rozdělení) Jestliže náhodná veličina X má pravděpodobnostní funkci

$$f(x) = \frac{\Gamma((m+2)/2)}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{x^{\frac{m-2}{2}}}{\left(1 + \frac{m}{n}x\right)^{(m+n)/2}} I_{(0,\infty)}(x)$$

pak tato náhodná veličina sleduje F rozdělení s m a n stupni volnosti.

Pořadí, ve které jsou stupně volnosti uváděny, je důležité, protože F rozdělení není symetrické v těchto parametrech. Počet stupňů volnosti m v podílu $\frac{m}{n}$ výše uvedené pravděpodobnostní funkce je vždy uváděn jako první.

Věta 6.10 *Nechť U a V jsou nezávislé náhodné veličiny, které sledují chi-kvadrát rozdělení s m resp. n stupni volnosti. Pak náhodná veličina*

$$X = \frac{U/m}{V/n}$$

sleduje F rozdělení s m a n stupni volnosti.

Tvrzení 6.6 *Nechť X_1, \dots, X_{m+1} představuje náhodný výběr velikosti $m+1$ z normálního rozdělení se střední hodnotou μ_X a rozptylem σ^2 . Nechť Y_1, \dots, Y_{n+1} představuje náhodný výběr velikosti $n+1$ z normálního rozdělení se střední hodnotou μ_Y a rozptylem σ^2 . Nechť jsou oba náhodné výběry nezávislé. Pak*

$$\frac{1}{\sigma^2} \sum_{i=1}^{m+1} (X_i - \bar{X})^2$$

sleduje chi-kvadrát rozdělení s m stupni volnosti a

$$\frac{1}{\sigma^2} \sum_{i=1}^{n+1} (Y_i - \bar{Y})^2$$

sleduje chi-kvadrát rozdělení s n stupni volnosti. Proto

$$\frac{\sum (X_i - \bar{X})^2 / m}{\sum (Y_j - \bar{Y}) / n}$$

sleduje F rozdělení s m a n stupni volnosti.

Věta 6.11 *Jestliže náhodná veličina X sleduje F distribuci s m a n stupni volnosti, pak*

$$E[X] = \frac{n}{n-2} \quad \text{pro } n > 2$$

$$D[X] = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad \text{pro } n > 4$$

Důkaz 6.9 *Dokažme tvrzení o střední hodnotě.*

$$E[X] = E\left[\frac{U/m}{V/n}\right] = \frac{n}{m} E[U] E\left[\frac{1}{V}\right]$$

Protože U sleduje chi-kvadrát rozdělení s m stupni volnosti, platí $E[U] = m$. Pro $E\left[\frac{1}{V}\right]$ je situace o něco složitější.

$$\begin{aligned} E\left[\frac{1}{V}\right] &= \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{\frac{n}{2}} \int_0^\infty \frac{1}{v} v^{\frac{n-2}{2}} e^{-\frac{1}{2}v} dv \\ &= \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{\frac{n}{2}} \int_0^\infty v^{\frac{n-4}{2}} e^{-\frac{1}{2}v} dv \\ &= \frac{\Gamma((n-2)/2)}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{\frac{n}{2}} \left(\frac{1}{2}\right)^{-\frac{n-2}{2}} = \frac{1}{n-2} \end{aligned}$$

Proto platí

$$E[X] = \left(\frac{n}{m}\right) E[U] E\left[\frac{1}{V}\right] = \frac{n}{m} \frac{m}{n-2} = \frac{n}{n-2}$$

Tvrzení o střední hodnotě lze dokázat analogicky.

Jestliže X sleduje F rozdělení s m a n stupni volnosti, pak $1/X$ sleduje F rozdělení s n a m stupni volnosti. Tento vztah umožňuje definovat tabulky tohoto pravděpodobnostního rozdělení pouze pro horní polovinu kvantilů. Platí totiž

$$p = P[X \leq \xi_p] = P\left[\frac{1}{X} \geq \frac{1}{\xi}\right] = P\left[Y \geq \frac{1}{\xi_p}\right] = 1 - P\left[Y \leq \frac{1}{\xi_p}\right]$$

což znamená $1 - p = P[Y \leq \xi'_{1-p}]$, a proto $\xi'_{1-p} = \frac{1}{\xi_p}$.

Jestliže X sleduje F rozdělení s m a n stupni volnosti, pak

$$W = \frac{m \frac{X}{n}}{1 + m \frac{X}{n}}$$

sleduje beta rozdělení s parametry $a = m/2$ a $b = n/2$.

6.3.4 Studentovo rozdělení

Věta 6.12 (Studentovo rozdělení) *Uvažujme náhodnou veličinu Z , která sleduje normované normální rozdělení, a náhodnou veličinu U , která sleduje chi-kvadrát rozdělení s k stupni volnosti. Jestliže jsou Z a U nezávislé, pak náhodná veličina $\frac{Z}{\sqrt{U/k}}$ sleduje studentovo rozdělení s k stupni volnosti.*

Sdružená pravděpodobnostní funkce náhodných veličin Z a U je

$$f_{Z,U}(z, u) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{\frac{k}{2}} u^{\frac{k}{2}-1} e^{-\frac{1}{2}u} e^{-\frac{1}{2}z^2} I_{(0,\infty)}(u)$$

Jestliže použijeme transformaci $X = \frac{Z}{\sqrt{U/k}}$ a $Y = U$, pak je Jakobián roven $\sqrt{\frac{y}{k}}$, a proto

$$f_{X,Y}(x, y) = \sqrt{\frac{y}{k}} \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{\frac{k}{2}} y^{\frac{k}{2}-1} e^{-\frac{1}{2}y} e^{-\frac{1}{2}\frac{x^2 y}{k}} I_{(0,\infty)}(y)$$

Následnou integrací pak získáme marginální pravděpodobnostní funkci náhodné veličiny X .

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \frac{1}{\sqrt{2k\pi}} \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{\frac{k}{2}} \int_0^{\infty} y^{\frac{k}{2}-1+\frac{1}{2}} e^{-\frac{1}{2}\left(1+\frac{x^2}{k}\right)y} dy \\ &= \frac{\Gamma((k+1)/2)}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} \frac{1}{\left(1+\frac{x^2}{k}\right)^{\frac{k+1}{2}}} \end{aligned}$$

Definice 6.11 (Studentovo rozdělení) Jestliže pravděpodobnostní rozdělení náhodné veličiny X má tvar

$$f_X(x) = \frac{\Gamma((k+1)/2)}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} \frac{1}{\left(1 + \frac{x^2}{k}\right)^{\frac{k+1}{2}}}$$

pak tato náhodná veličina sleduje studentovo rozdělení s k stupni volnosti.

Tvrzení 6.7 Jestliže X_1, \dots, X_n představuje náhodný výběr z normálního rozdělení se střední hodnotou μ a rozptylem σ^2 , pak $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ sleduje normalizované normální rozdělení a $U = \frac{1}{\sigma^2} \sum (X_i - \bar{X})^2$ sleduje chi-kvadrát rozdělení s $n - 1$ stupni volnosti. Nezávislost Z a U vyplývá z věty (6.8), a proto

$$\frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{1}{\sigma^2} \sum \frac{(X_i - \bar{X})^2}{n-1}}} = \frac{\sqrt{n(n-1)}(\bar{X} - \mu)}{\sqrt{\sum (X_i - \bar{X})^2}}$$

sleduje studentovo rozdělení s $n - 1$ stupni volnosti.

V případě jednoho stupně volnosti se studentovo rozdělení zredukuje na Cauchyho rozdělení. Naopak s rostoucím počtem stupňů volnosti se studentovo rozdělení blíží normalizovanému normálnímu rozdělení. Dále platí, že druhá mocnina studentova rozdělení s k stupni volnosti odpovídá F rozdělení s jedním a k stupni volnosti.

Definice 6.12 Jestliže náhodná veličina X sleduje studentovo rozdělení s k stupni volnosti, pak

$$E[X] = 0 \quad \text{pro } k > 1$$

$$D[X] = \frac{k}{k-2} \quad \text{pro } k > 2$$

Důkaz 6.10 Postup tohoto důkazu je podobný důkazu (6.8). Připomeňme, že $X = \frac{Z}{\sqrt{U/k}}$, kde Z a U jsou nezávislé. To implikuje

$$E[X] = E\left[\frac{Z}{\sqrt{U/k}}\right] = E[Z]\sqrt{k}E\left[\frac{1}{\sqrt{U}}\right] = 0 \cdot \sqrt{k}E\left[\frac{1}{\sqrt{U}}\right] = 0$$

a

$$E[X^2] = E\left[\frac{Z^2}{U/k}\right] = E[Z^2] \cdot k \cdot E\left[\frac{1}{U}\right] = 1 \cdot k \cdot \frac{1}{k-2} = \frac{k}{k-2}$$

Rozptyl náhodné veličiny X je tedy

$$D[X] = E[X^2] - E[X]^2 = \frac{k}{k-2}$$

6.4 Pořadové statistiky

Definice 6.13 (Pořadové statistiky) Necht' X_1, \dots, X_n představuje náhodný výběr velikosti n z kumulativní distribuční funkce $F(\cdot)$. Pak $Y_1 \leq Y_2 \leq \dots \leq Y_n$, kde Y_i jsou vzestupně seřazená X_i , nazýváme pořadovou statistikou náhodného výběru X_1, \dots, X_n .

Z výše uvedené definice je zřejmé, že jednotlivá Y_i nejsou nezávislá, protože je-li $Y_j \geq y$, pak také $Y_{j+1} \geq y$.

Věta 6.13 *Nechť $Y_1 \leq Y_2 \leq \dots \leq Y_n$ představuje pořadovou statistiku z kumulativní distribuční funkce $F(\cdot)$. Marginální kumulativní distribuční funkce náhodné veličiny Y_α , kde $\alpha = 1, 2, \dots, n$, je dána*

$$F_{Y_\alpha}(y) = \sum_{j=\alpha}^n \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}$$

Důkaz 6.11 *Pro fixní y definujeme $Z_i = I_{(-\infty, y]}(X_i)$. Je zřejmé, že $\sum_{i=1}^n Z_i$ je rovno počtu náhodných veličin X_i , pro která platí $X_i \leq y$. Platí*

$$F_{Y_\alpha} = P[Y_\alpha \leq y] = P\left[\sum_i Z_i \geq \alpha\right] = \sum_{j=\alpha}^n \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}$$

Klíčovým krokem důkazu je ekvivalence jevů $\{Y_\alpha \leq y\}$ a $\{\sum Z_i \geq \alpha\}$. Jestliže je totiž α -tá pořadová statistika menší nebo rovna y , pak je počet X_i , která jsou menší nebo rovna y , větší nebo roven α a naopak.

Tvrzení 6.8

$$F_{Y_n}(y) = \sum_{j=n}^n \binom{n}{j} F(y)^j [1 - F(y)]^{n-j} = [F(y)]^n$$

$$F_{Y_1}(y) = \binom{n}{j} F(y)^j [1 - F(y)]^{n-j} = 1 - [1 - F(y)]^n$$

Po zbytek této kapitoly budeme předpokládat, že náhodný vzorek X_1, \dots, X_n pochází ze spojitého rozdělení.

Pravděpodobnostní funkci $f_{Y_\alpha}(y)$ lze získat derivací výše odvozené kumulativní distribuční funkce.

$$\begin{aligned} f_{Y_\alpha}(y) &= \lim_{\Delta y \rightarrow 0} \frac{F_{Y_\alpha}(y + \Delta y) - F_{Y_\alpha}(y)}{\Delta y} = \lim_{\Delta y \rightarrow 0} \frac{P[y < Y_\alpha \leq y + \Delta y]}{\Delta y} \\ &= \lim_{\Delta y \rightarrow 0} \frac{P[(\alpha - 1) X_i \leq y, \text{ jedno } X_i \text{ v } (y, y + \Delta y], (n - \alpha) X_i > y + \Delta y]}{\Delta y} \\ &= \lim_{\Delta y \rightarrow 0} \frac{n!}{(\alpha - 1)!1!(n - \alpha)!} \frac{[F(y)]^{\alpha-1} [F(y + \Delta y) - F(y)] [1 - F(y + \Delta y)]^{n-\alpha}}{\Delta y} \\ &= \frac{n!}{(\alpha - 1)!1!(n - \alpha)!} [F(y)]^{\alpha-1} [1 - F(y)]^{n-\alpha} f(y) \end{aligned}$$

Podobně lze odvodit sdruženou pravděpodobnostní funkci náhodných veličin Y_α a Y_β , kde $1 \leq \alpha < \beta \leq n$.

$$\begin{aligned} f_{Y_\alpha, Y_\beta}(x, y) \Delta x \Delta y &\approx P[x < Y_\alpha \leq x + \Delta x, y < Y_\beta \leq y + \Delta y] \\ &\approx P[(\alpha - 1) X_i \leq x, \text{ jedno } X_i \text{ v } (x, x + \Delta x], \\ &\quad (\beta - \alpha - 1) X_i \text{ v } (x + \Delta x, y], \text{ jedno } X_i \text{ v } (y, y + \Delta y], (n - \beta) X_i > y + \Delta y] \\ &\approx \frac{n!}{(\alpha - 1)!1!(\beta - \alpha - 1)!1!(n - \beta)!} \\ &\quad \times [F(x)]^{\alpha-1} [F(y) - F(x + \Delta x)]^{\beta-\alpha-1} [1 - F(y + \Delta y)]^{n-\beta} f(x) \Delta x f(y) \Delta y \end{aligned}$$

Proto pro $x < y$ platí

$$f_{Y_\alpha, Y_\beta}(x, y) = \frac{n!}{(\alpha-1)!(\beta-\alpha-1)!(n-\beta)!} [F(x)]^{\alpha-1} [F(y)-F(x)]^{\beta-\alpha-1} [1-F(y)]^{n-\beta} f(x)f(y)$$

Sdruženou pravděpodobnostní funkci $f_{Y_1, \dots, Y_n}(y_1, \dots, y_n)$ pak lze odvodit jako

$$\begin{aligned} f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) &= \lim_{\Delta y_i \rightarrow 0} \frac{1}{\prod_{i=1}^n \Delta y_i} P[y_1 < Y_1 \leq y_1 + \Delta y_1, \dots, y_n < Y_n \leq y_n + \Delta y_n] \\ &= \lim_{\Delta y_i \rightarrow 0} \frac{1}{\prod_{i=1}^n \Delta y_i} P[\text{jedno } X_i \text{ v } (y_1, y_1 + \Delta y_1), \dots, \text{jedno } X_i \text{ v } (y_n, y_n + \Delta y_n)] \\ &= \lim_{\Delta y_i \rightarrow 0} \frac{n!}{\prod_{i=1}^n \Delta y_i} [F(y_1 + \Delta y_1) - F(y_1)], \dots, [F(y_n + \Delta y_n) - F(y_n)] \\ &= n! f(y_1) \cdots f(y_n) \end{aligned}$$

pro $y_1 < y_2 < \dots < y_n$.

Věta 6.14 *Nechť X_1, \dots, X_n představuje náhodný výběr z populace $f(\cdot)$ s kumulativní distribuční funkcí $F(\cdot)$. Dále nechť $Y_1 \leq Y_2 \leq \dots \leq Y_n$ představuje odpovídající pořadovou statistiku. Pak platí*

$$\begin{aligned} f_{Y_\alpha}(y) &= \frac{n!}{(\alpha-1)!(n-\alpha)!} [F(y)]^{\alpha-1} [1-F(y)]^{n-\alpha} f(y) \\ f_{Y_\alpha, Y_\beta}(x, y) &= \frac{n!}{(\alpha-1)!(n-\alpha)!} [F(y)]^{\alpha-1} [F(y)-F(x)]^{\beta-\alpha-1} [1-F(y)]^{n-\beta} f(x)f(y) \\ f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) &= n! f(y_1) \cdots f(y_n) \end{aligned}$$

6.4.1 Rozdělení funkcí pořadových statistik

Jednou z možných funkcí pořadových statistik je jejich aritmetický průměr $\frac{1}{n} \sum_{j=1}^n Y_j$. Nicméně $\frac{1}{n} \sum_{j=1}^n Y_j = \frac{1}{n} \sum_{i=1}^n X_i$, což je střední hodnota náhodného výběru. Proto se zaměříme na jiné funkce pořadových statistik.

Definice 6.14 (Výběrový medián, rozpětí výběru a střed výběru)

Nechť $Y_1 \leq \dots \leq Y_n$ označuje pořadové statistiky náhodného výběru X_1, \dots, X_n z populace $f(\cdot)$. Výběrový medián je definován jako prostřední statistika, je-li n liché popř. jako průměr dvou prostředních statistik, je-li n sudé. Rozpětí výběru je definováno jako $Y_n - Y_1$ a střed výběru jako $\frac{Y_1 + Y_n}{2}$.

Pokusme se odvodit sdruženou pravděpodobnostní funkci výběrového mediánu a rozpětí výběru. Jestliže je n liché, pak je pravděpodobnostní funkce výběrového mediánu dána první rovnicí věty (6.14). Jestliže je n sudé a výběrový medián je tak definován jako $\frac{Y_{n/2} + Y_{n/2+1}}{2}$ pak je jeho pravděpodobnostní funkce dána druhou rovnicí věty (6.14). Dle třetí rovnice věty (6.14) platí

$$f_{Y_1, Y_n}(x, y) = n(n-1)[F(y) - F(x)]^n - 2f(x)f(y)$$

Jestliže použijeme transformaci $R = Y_n - Y_1$ a $\frac{Y_1 + Y_n}{2}$ popř. $r = y - x$ a $t = \frac{x+y}{2}$, získáme $x = t - \frac{r}{2}$ a $y = t + \frac{r}{2}$ a Jakobián je tak roven

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial t} \end{vmatrix} = \begin{vmatrix} -\frac{1}{2} & 1 \\ \frac{1}{2} & 1 \end{vmatrix} = -1$$

čímž se dostáváme následující větě.

Věta 6.15 Jestliže R je rozpětí výběru a T je střed výběru, pak jejich sdružená pravděpodobnostní funkce je

$$f_{R,T}(r, t) = n(n-1)[F(t+r/2) - F(t-r/2)]^{n-2} f(t-r/2) f(t+r/2)$$

a odpovídající marginální pravděpodobnostní funkce jsou

$$f_R(r) = \int_{-\infty}^{\infty} f_{R,T}(r, t) dt$$

a

$$f_T(t) = \int_0^{\infty} f_{R,T}(r, t) dr$$

Příklad 6.6 Nechť X_1, \dots, X_n představuje náhodný výběr z populace uniformního rozdělení nad intervalem $(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)$, kde μ je střední hodnota a σ^2 rozptyl populace. Z definice uniformního rozdělení vyplývá

$$f(x) = \frac{1}{2\sqrt{3}\sigma} I_{(\mu-\sqrt{3}\sigma, \mu+\sqrt{3}\sigma)}(x)$$

a

$$F(x) = \left(\frac{1}{2\sqrt{3}\sigma} x - \frac{\mu - \sqrt{3}\sigma}{2\sqrt{3}\sigma} \right) I_{(\mu-\sqrt{3}\sigma, \mu+\sqrt{3}\sigma)}(x) + I_{[\mu+\sqrt{3}\sigma, \infty)}(x)$$

Sdružená pravděpodobnostní funkce náhodných veličin R a T je tak

$$f_{R,T}(r, t) = \frac{n(n-1)}{(2\sqrt{3}\sigma)^n} r^{n-2} I_{(\mu-\sqrt{3}\sigma+r/2, \mu+\sqrt{3}\sigma-r/2)}(t) I_{(0, 2\sqrt{3}\sigma)}(r)$$

Integrací přes t získáme

$$f_R(r) = \int_{\mu-\sqrt{3}\sigma-r/2}^{\mu-\sqrt{3}\sigma+r/2} f_{R,T}(r, t) dt = \frac{n(n-1)}{(2\sqrt{3}\sigma)^n} r^{n-2} (2\sqrt{3}\sigma - r) I_{(0, 2\sqrt{3}\sigma)}(r)$$

a integrací přes r pak

$$\begin{aligned} f_T(t) &= \int_0^{\min(2t-2(\mu-\sqrt{3}\sigma), 2(\mu+\sqrt{3}\sigma)-2t)} f_{R,T}(r, t) dr \\ &= \frac{n(n-1)}{(2\sqrt{3}\sigma)^n} \int_0^{\min(2t-2(\mu-\sqrt{3}\sigma), 2(\mu+\sqrt{3}\sigma)-2t)} r^{n-2} dr \cdot I_{(\mu-\sqrt{3}\sigma, \mu+\sqrt{3}\sigma)}(t) \\ &= \frac{n}{2\sqrt{3}\sigma} \left(\frac{t-\mu}{\sqrt{3}\sigma} + 1 \right)^{n-1} I_{(-1, 0)} \left(\frac{t-\mu}{\sqrt{3}\sigma} \right) + \frac{n}{2\sqrt{3}\sigma} \left(1 - \frac{t-\mu}{\sqrt{3}\sigma} \right)^{n-1} I_{[0, 1)} \left(\frac{t-\mu}{\sqrt{3}\sigma} \right) \end{aligned}$$

6.4.2 Asymptotické rozdělení

V předchozím textu jsme se zabývali asymptotickým rozdělením náhodné veličiny \bar{X} , tj. střední hodnotou výběru. Nyní se zaměříme na asymptotické rozdělení výběrových statistik.

Věta 6.16 Nechť X_1, \dots, X_n jsou nezávislé náhodné veličiny, které sledují shodné pravděpodobnostní rozdělení s pravděpodobnostní funkcí $f(\cdot)$ a kumulativní distribuční funkcí $F(\cdot)$. Předpokládejme, že $F(\cdot)$ je pro $0 < F(x) < 1$

monotónní funkcí. Definujme ξ_p jako řešení rovnice $F(\xi_p) = p$ pro $0 < p < 1$. Vzhledem k monotónnosti $F(\cdot)$ je toto řešení jedinečné a představuje p -tý kvantil. Definujme p_n takové, že np_n je celé číslo a $n|p_n - p|$ je ohraničené. Dále definujme $Y_{np_n}^{(n)}$ jako (np_n) -tou pořadovou statistiku náhodného výběru velikosti n . Pak $Y_{np_n}^{(n)}$ asymptoticky sleduje normální rozdělení se střední hodnotou ξ_p a rozptylem $\frac{p(1-p)}{n[f(\xi_p)]^2}$.

Příklad 6.7 Necht' $p = \frac{1}{2}$. Pak ξ_p představuje medián náhodného výběru a dle výše uvedené věty asymptoticky sleduje normální rozdělení se střední hodnotou odpovídající mediánu populace a rozptylem $\frac{1}{4n}[f(\xi_{0.5})]^2$. Pokud by se populace řídila normálním rozdělením se střední hodnotou μ a rozptylem σ^2 , pak by medián náhodného výběru asymptoticky sledoval normální rozdělení se střední hodnotou μ a rozptylem $\frac{\pi\sigma^2}{2n}$. Pouze připomeňme, že střední hodnota náhodného výběru asymptoticky sleduje normální rozdělení se střední hodnotou μ a rozptylem $\frac{\sigma^2}{n}$.

Nyní se zaměříme na asymptotické rozdělení pořadové statistiky $Y_k^{(n)}$, jejíž absolutní pořadí se nemění bez ohledu na velikost náhodného výběru³. Pro zjednodušení následujících úvah se zaměříme na $Y_n^{(n)}$, tj. na největší mezi n pozorováními náhodného výběru. Podobně jako v případě střední hodnoty náhodného výběru se pokusíme nalézt asymptotickou distribuci, která by byla nezávislá na pravděpodobnostním rozdělení populace. V případě náhodné veličiny \bar{X}_n jsme prokázali, že $Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$, kde μ resp. σ představuje střední hodnotu resp. rozptyl populace, asymptoticky sleduje normované normální rozdělení. Podobně, dle výše uvedené věty, náhodná veličina $Z_n = \frac{Y_{np_n}^{(n)} - \xi_p}{\sqrt{p(1-p)/n[f(\xi_p)]^2}}$ opět sleduje normované normální rozdělení. Proto budeme podobný postup aplikovat také v případě náhodné veličiny $Y_n^{(n)}$. Postup ilustrujeme na následujících příkladech.

Příklad 6.8 Uvažujme náhodný výběr z logistického rozdělení, kde $F(x) = (1 + e^{-x})^{-1}$. Pokusme se nalézt limitní rozdělení pro $\frac{(Y_n^{(n)} - a_n)}{b_n}$. Problém rozdělíme na dva kroky. V prvním se pokusíme nalézt konstanty a_n a b_n . Ve druhém se pak pokusíme nalézt limitní rozdělení pro $\frac{Y_n^{(n)} - a_n}{b_n}$.

Vzhledem k limitním rozdělením pro \bar{X}_n a $X_{np_n}^{(n)}$ se zdá, že a_n by mělo být blízké $E[Y_n^{(n)}]$. Víme, že $F(X_1), \dots, F(X_n)$ představuje náhodný výběr z uniformního rozdělení nad intervalem $(0, 1)$. Proto $F(Y_n^{(n)})$ odpovídá největší hodnotě z náhodného výběru velikosti n z uniformního rozdělení nad intervalem $(0, 1)$. V tomto případě platí

$$E[F(Y_n^{(n)})] = \int_0^1 nx^n \cdot 1 \, dx = \frac{n}{n+1}$$

³Parameter k statistiky této statistiky nezávisí na velikosti náhodného výběru n . Naproti tomu absolutní pořadí statistiky $Y_{np_n}^{(n)}$ roste s rostoucím n , i když její relativní pořadí zůstává přibližně neměnné.

Dále platí $F(E[Y_n^{(n)}]) \approx E[F(Y_n^{(n)})]$, neboli

$$\begin{aligned} F(E[Y_n^{(n)}]) &= \left(1 + e^{-E[Y_n^{(n)}]}\right)^{-1} = 1 - \left(1 + e^{E[Y_n^{(n)}]}\right)^{-1} \\ &\approx 1 - (n+1)^{-1} = \frac{n}{n+1} = E[F(Y_n^{(n)})] \end{aligned}$$

což implikuje $n \approx e^{E[Y_n^{(n)}]}$ neboli $E[Y_n^{(n)}] \approx \ln(n)$. Proto

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left[\frac{Y_n^{(n)} - a_n}{b_n} \leq y \right] &= \lim_{n \rightarrow \infty} P \left[\frac{Y_n^{(n)} - \ln(n)}{b_n} \leq y_n \right] \\ \lim_{n \rightarrow \infty} P[Y_n^{(n)} \leq b_n y + \ln(n)] &= \lim_{n \rightarrow \infty} [F(b_n y + \ln(n))]^n \\ \lim_{n \rightarrow \infty} \left(1 + e^{-b_n y - \ln(n)}\right)^{-n} &= \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n} e^{-b_n y}\right)^{-n} \\ &= e^{-e^{-y}} \end{aligned}$$

pro $b_n \equiv 1$. Limitující distribuci pro $Y_n^{(n)} - \ln(n)$ je tedy $e^{-e^{-y}}$.

Příklad 6.9 Uvažujme náhodný výběr z exponenciálního rozdělení, kde $F(x) = (1 - e^{-\lambda x})I_{(0, \infty)}(x)$. Opět se pokusíme nalézt limitní rozdělení pro $\frac{Y_n^{(n)} - a_n}{b_n}$.

Z předchozího příkladu víme, že $E[F(Y_n^{(n)})] = n/(n+1) = 1 - 1/(n+1)$ a $F(E[Y_n^{(n)}]) \approx E[F(Y_n^{(n)})]$. Platí

$$F(E[Y_n^{(n)}]) = 1 - e^{-\lambda E[Y_n^{(n)}]}$$

a proto

$$\frac{1}{n+1} \approx e^{-\lambda E[Y_n^{(n)}]}$$

neboli

$$E[Y_n^{(n)}] \approx \frac{1}{\lambda} \ln(n+1) \approx \frac{1}{\lambda} \ln(n)$$

Tímto se dostáváme k

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left[\frac{Y_n - a_n}{b_n} \leq y \right] &= \lim_{n \rightarrow \infty} P \left[Y_n - \frac{1}{\lambda} \ln(n) \leq b_n y \right] \\ &= \lim_{n \rightarrow \infty} \left[F \left(b_n y + \frac{1}{\lambda} \ln(n) \right) \right]^n = \lim_{n \rightarrow \infty} \left(1 - e^{-\lambda b_n y - \ln(n)} \right)^n \\ &= \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n} e^{-\lambda b_n y} \right)^n = e^{-e^{-y}} \end{aligned}$$

pro $b_n \equiv \frac{1}{\lambda}$. Limitující rozdělení pro $\frac{Y_n^{(n)} - (1/\lambda) \ln(n)}{1/\lambda}$ je opět $e^{-e^{-y}}$.

Tímto se dostáváme k následující větě.

Věta 6.17 Necht jsou náhodné veličiny X_1, \dots, X_n vzájemně nezávislé a mají shodnou kumulativní distribuční funkci $F(\cdot)$. Jestliže $\frac{Y_n^{(n)} - a_n}{b_n}$ má limitní rozdělení, pak se musí jednat o jeden z následujících tří typů

$$\begin{aligned} G_1(y, \gamma) &= e^{-y^{-\gamma}} I_{(0, \infty)}(y), \quad \text{pro } \gamma > 0 \\ G_2(y, \gamma) &= e^{-|y|^\gamma} I_{(-\infty, 0)}(y) + I_{[0, \infty)}(y) \quad \text{pro } \gamma > 0 \\ G_3(y) &= e^{-e^{-y}} \end{aligned}$$

Věta 6.18 *Nechť jsou náhodné veličiny X_1, \dots, X_n vzájemně nezávislé a mají shodnou kumulativní distribuční funkci $F(\cdot)$. Předpokládejme, že $\frac{Y_n^{(n)} - a_n}{b_n}$ má limitní rozdělení. Toto limitní rozdělení má tvar*

1. $G_1(\cdot, \gamma)$ tehdy a jen tehdy, jestliže

$$\lim_{x \rightarrow \infty} \frac{1 - F(x)}{1 - F(\tau x)} = \tau^\gamma \quad \text{pro každé } \tau > 0$$

2. $G_2(\cdot, \gamma)$ tehdy a jen tehdy, jestliže existuje x_0 takové, že

$$F(x_0) = 1$$

a

$$F(X_0 - \epsilon) < 1 \quad \text{pro každé } \epsilon > 0$$

a

$$\lim_{0 < x \rightarrow 0} \frac{1 - F(x_0 - \tau x)}{1 - F(x_0 - x)} = \tau^\gamma \quad \text{pro každé } \tau > 0$$

3. $G_3(\cdot, \gamma)$ tehdy a jen tehdy, jestliže

$$\lim_{n \rightarrow \infty} n[1 - F(\beta_n x + \alpha_n)] = e^{-x} \quad \text{pro každé } x$$

kde

$$\alpha_n = \inf \left(z : \frac{n-1}{n} \leq F(z) \right)$$

a

$$\beta_n = \inf (z : 1 - (ne)^{-1} \leq F(\alpha_n + z))$$

Všimněme si, že pokud je $F(\cdot)$ striktně monotonní a spojitý, pak α_n je dáno $F(\alpha_n) = \frac{n-1}{n}$ neboli $\alpha_n = F^{-1}((n-1)/n)$ a β_n je dáno $F(\alpha_n + \beta_n) = 1 - (ne)^{-1}$ neboli $\beta_n = F^{-1}(1 - (ne)^{-1}) - \alpha_n = F^{-1}(1 - (ne)^{-1}) - F^{-1}((n-1)/n)$.

Příklad 6.10 *Uvažujme kumulativní distribuční funkci $F(x, \gamma) = [1 - (1 - x)^\gamma]I_{(0,1)}(x) + I_{[1,\infty)}(x)$. Pro $x_0 = 1$ platí $F(x_0) = 1$ a $F(x_0 - \epsilon) < 1$ pro libovolné $\epsilon > 0$. Dále také platí*

$$\lim_{0 < x \rightarrow 0} \frac{1 - F(x_0 - \tau x)}{1 - F(x_0 - x)} = \lim_{0 < x \rightarrow 0} \frac{(1 - x_0 + \tau x)^\gamma}{(1 - x_0 + x)^\gamma} = \tau^\gamma$$

Proto má $F(\cdot, \gamma)$ jako limitní funkci $G_2(\cdot, \gamma)$.

Výše uvedená věta nám s výjimkou $G_3(\cdot)$ bohužel nic neříká o parametrech a_n a b_n . Naštěstí $G(\cdot)$ představuje limitní formu pro velké množství rozdělení, jako jsou např. logistické, exponenciální, gamma a normální rozdělení.

6.4.3 Výběrová kumulativní distribuční funkce

Definice 6.15 (Výběrová kumulativní distribuční funkce) *Nechť X_1, X_2, \dots, X_n představuje náhodný výběr z kumulativní distribuční funkce $F(\cdot)$ a nechť $Y_1 \leq Y_2 \leq \dots \leq Y_n$ představuje odpovídající pořadovou statistiku. Výběrová kumulativní funkce je definována jako*

$$F_n(x) = \frac{1}{n} \times (\text{počet } X_i \text{ menších nebo rovných } x) = \frac{1}{n} \times (\text{počet } Y_i \text{ menších nebo rovných } x)$$

Následující věta ukazuje, že $F_n(\cdot)$ sleduje stejné rozdělení jako střední hodnota výběru z Bernoulliho rozdělení.

Věta 6.19 *Nechť $F_n(x)$ představuje kumulativní distribuční funkci náhodného výběru o velikosti n z populace $F(\cdot)$. Pak platí*

$$P \left[F_n(x) = \frac{k}{n} \right] = \binom{n}{k} [F(x)]^k [1 - F(x)]^{n-k}, \quad k = 0, 1, \dots, n$$

Důkaz 6.12 *Nechť $Z_i = I_{(-\infty, x]}(X_i)$. Pak náhodná veličina Z_i sleduje Bernoulliho rozdělení s parametrem $F(x)$. Proto $\sum Z_i$, které představuje počet X_i menších nebo rovných x , sleduje binomické rozdělení s parametry n a $F(x)$. Zároveň však platí $F_n(x) = \frac{1}{n} \sum Z_i$.*

Kapitola 7

Bodový odhad

7.1 Metody bodového odhadu

Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$, pro kterou známe typ pravděpodobnostního rozdělení, avšak neznáme vektor parametrů $\theta = (\theta_1, \dots, \theta_k)$. Nechť Θ představuje tzv. prostor parametrů, tj. množinu všech hodnot, kterých může θ nabývat. Cílem je nalézt statistiku, která je funkcí X_1, \dots, X_n a kterou lze použít k odhadu θ_j , kde $j = 1, \dots, k$.

Definice 7.1 (Odhad) *Statistiku¹, jejíž hodnota je použita k odhadu $\tau(\theta)$, kde $\tau(\cdot)$ je funkcí θ , nazýváme funkcí odhadu.*

Funkce odhadu je tak statistikou, která je zároveň funkcí a náhodnou veličinou. Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$. Předpokládejme, že chceme odhadnout $\tau(\theta)$. Nechť $t(X_1, \dots, X_n)$ představuje funkci odhadu pro $\tau\theta$. Na $t(X_1, \dots, X_n)$ je tak možné pohlížet jako na náhodnou veličinu $T = t(X_1, \dots, X_n)$ a zároveň jako na funkci $t(\cdot, \dots, \cdot)$. V případě konkrétní hodnoty náhodné veličiny $T = t$ pak hovoříme o odhadu $\tau(\theta)$, kdežto v případě $t(\cdot, \dots, \cdot)$ pak o funkci odhadu. Toto názvosloví ilustrujeme na příkladu střední hodnoty náhodného výběru $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Hodnota $\bar{X}_n = \bar{x}_n$ získaná na základě konkrétního náhodného výběru je odhadem střední hodnoty populace μ . V tomto případě představuje \bar{X}_n náhodnou veličinu T a \bar{x}_n její konkrétní hodnotu t . Funkce odhadu t, \dots, \cdot je pak $\frac{1}{n} \sum(\cdot)$.

V případě odhadu se pak také velmi často používá označení $\hat{\theta}$ pro odhad θ a $\hat{\tau}$ pro funkci odhadu. Právě nalezení konkrétní hodnoty popř. hodnot θ bude cílem následujícího textu.

7.1.1 Momentová metoda

Nechť $f(\cdot, \theta)$, kde $\theta = (\theta_1, \dots, \theta_k)$, představuje pravděpodobnostní funkci náhodné veličiny X . Uvažujme r -tý obecný moment

$$\mu'_r = E[X^r]$$

¹Připomeňme, že statistikou rozumíme funkci pozorovatelné náhodné veličiny a že tato statistika má sama o sobě povahu náhodné veličiny.

Obecný moment tak představuje známou funkci k parametrů $\theta_1, \dots, \theta_k$, a proto můžeme použít zápis

$$\mu'_r = \mu'_r(\theta_1, \dots, \theta_k)$$

Pro náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$ je r -tý obecný výběrový moment definován jako

$$M'_r = \frac{1}{n} \sum_{i=1}^n X_i^r$$

Definujeme k rovnic

$$M'_r = \mu'_r(\theta_1, \dots, \theta_k), \quad j = 1, \dots, k$$

a předpokládejme, že jejichž jedinečná řešení jsou $\hat{\theta}_1, \dots, \hat{\theta}_k$. $(\hat{\theta}_1, \dots, \hat{\theta}_k)$ tak představuje funkci odhadu pro $(\theta_1, \dots, \theta_k)$. Momentová metoda je tak založena na nahrazení populačních momentů výběrovými momenty.

Příklad 7.1 *Nechť X_1, \dots, X_n představuje náhodný výběr z Poissonova rozdělení s parametrem λ . Odhadněme tento parametr.*

Protože máme pouze jeden parametr, vystačíme s jednou rovnicí a to

$$M'_1 = \mu'_1 = \mu'_1(\theta) = \theta$$

Funkcí odhadu parametru λ je tedy $M'_1 = \bar{X}$, na základě které je možné odhadnout střední hodnotu populace λ pomocí střední hodnoty náhodného výběru \bar{x} .

Příklad 7.2 *Nechť X_1, \dots, X_n představuje náhodný výběr z uniformního rozdělení nad intervalem $(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)$. V tomto případě máme dva neznámé parametry a to střední hodnotu populace μ a její směrodatnou odchylku σ . Odhadněme tyto parametry.*

Pro odhad parametrů budeme potřebovat dvě rovnice a to

$$\begin{aligned} M'_1 &= \mu'_1 = \mu'_1(\mu, \sigma) = \mu \\ M'_2 &= \mu'_2 = \mu'_2(\mu, \sigma) = \sigma^2 + \mu^2 \end{aligned}$$

Proto jsou funkce odhadu \bar{X} pro μ a $\sqrt{\frac{1}{n} \sum X_i^2 - \bar{X}^2} = \sqrt{\frac{1}{n} \sum (X_i - \bar{X})^2}$ pro σ .

Momentová metoda není jednoznačná. V předchozím textu jsme vždy používali prvních k obecných momentů. V zásadě je možné použít libovolných k momentů. Navíc lze na místo obecných použít centrální momenty.

Pokud bychom namísto $(\theta_1, \dots, \theta_k)$ potřebovali odhadnout $\tau_1(\theta_1, \dots, \theta_k), \dots, \tau_r(\theta_1, \dots, \theta_k)$, můžeme postupovat několika způsoby. Prvním je získání odhadu $(\hat{\theta}_1, \dots, \hat{\theta}_k)$ pro $(\theta_1, \dots, \theta_k)$ a následné použití $\tau_j(\hat{\theta}_1, \dots, \hat{\theta}_k)$ jako odhadu pro $\tau_j(\theta_1, \dots, \theta_k)$ pro $j = 1, \dots, r$. Dalším možným způsobem je definování rovnic

$$M'_j = \mu'_j(\tau_1, \dots, \tau_r) \quad j = 1, \dots, r$$

a jejich následné řešení pro τ_1, \dots, τ_r . V obou případech hovoříme o momentové metodě odhadu, avšak výsledky obou přístupů se nemusí shodovat.

x	0	1	2	3
$f(x, \frac{3}{4})$	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$
$f(x, \frac{1}{4})$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{3}{64}$

Tabulka 7.1: Nádobu s černými a bílými míčky

7.1.2 Metoda maximální věrohodnosti

Uvažujme nádobu, která obsahuje černé a bílé míčky. Předpokládejme, že poměr míčků je $1/3$, avšak nevíme, zda-li je tento poměr ve prospěch bílých nebo černých míčků. To znamená, že pravděpodobnost tažení černého míčku je buďto $\frac{1}{4}$ nebo $\frac{3}{4}$. Jestliže bude taženo n míčků s vrácením, pak náhodná veličina X představující počet tažených černých míčků sleduje binomické rozdělení

$$f(x, p) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, \dots, n$$

kde $q = 1 - p$ a $p = \frac{1}{4}$ popř. $p = \frac{3}{4}$. Předpokládejme, že budou taženy tři míčky. Možné výsledky shrnuje tabulka (7.1).

Je zřejmé, že budou-li v náhodném výběru převládat černé kuličky, budeme se spíše klonit k $p = \frac{3}{4}$ a k $p = \frac{1}{4}$ v opačném případě. Proto bude funkce odhadu parametru p definována jako

$$\hat{p} = \hat{p}(x) = \begin{cases} 0.25 & \text{pro } x = 0, 1 \\ 0.75 & \text{pro } x = 2, 3 \end{cases}$$

Funkce odhadu tak každému x přiřadí takový odhad \hat{p} parametru p , že pro všechna p' platí

$$f(x, \hat{p}) > f(x, p')$$

kde p' představuje alternativní hodnotu pro p .

Pokud bychom vybrali 25 kuliček, z nich by šest bylo černých, pak

$$f(6, p) = \binom{25}{6} p^6 (1-p)^{19} \quad 0 \leq p \leq 1$$

Hodnotu parametru p , která maximalizuje hodnotu $f(6, p)$ lze snadno nalézt pomocí první derivace dle p . Řešením

$$\begin{aligned} \frac{df(6, p)}{dp} &= 0 \\ \binom{25}{6} p^5 (1-p)^{18} [6(1-p) - 19p] &= 0 \end{aligned}$$

jsou 0, 1 a $\frac{6}{25}$, přičemž první dva kořeny představují minimum funkce. Náš odhad je tak roven $\frac{6}{25}$ a splňuje podmínku

$$f(6, \hat{p}) > f(6, p')$$

pro všechna p' z intervalu 0 až 1.

Definice 7.2 (Funkce věrohodnosti) Funkce věrohodnosti náhodných veličin X_1, \dots, X_n je sdružená pravděpodobnostní funkce $f_{X_1, \dots, X_n}(x_1, \dots, x_n, \theta)$, která je funkcí θ . Jestliže X_1, \dots, X_n představuje náhodný výběr z populace $f(x, \theta)$, pak má funkce věrohodnosti tvar $f(x_1, \theta)f(x_2, \theta) \cdots f(x_n, \theta)$.

Abychom zdůraznili, že pravděpodobnostní funkce je funkcí θ , budeme ji v následujícím textu značit jako $L(\theta, x_1, \dots, x_n)$ popř. jako $L(\cdot, x_1, \dots, x_n)$.

Hodnota pravděpodobnostní funkce $L(\theta, x_1, \dots, x_n)$ tedy udává “pravděpodobnost”² realizace konkrétních hodnot x_1, \dots, x_n . Tímto se dostáváme k následující definici.

Definice 7.3 (Funkce maximální věrohodnosti) Nechť $L(\theta) = L(\theta, x_1, \dots, x_n)$ představuje funkci věrohodnosti pro náhodné veličiny X_1, \dots, X_n a $\hat{\theta} = \zeta(x_1, \dots, x_n)$ je hodnota θ z Θ , která maximalizuje $L(\theta)$. Pak $\hat{\theta} = \zeta(X_1, \dots, X_n)$ je funkce maximální věrohodnosti pro θ a $\hat{\theta} = \zeta(x_1, \dots, x_n)$ je odhadem dle metody maximální věrohodnosti pro θ na základě pozorovaných hodnot x_1, \dots, x_n .

Nejčastějším případem je situace, kdy X_1, \dots, X_n představuje náhodný výběr z populace $f(x, \theta)$, což implikuje funkci maximální věrohodnosti ve tvaru

$$L(\theta) = f(x_1, \theta) \cdots f(x_n, \theta)$$

Funkce $L(\theta)$ a $\ln(L(\theta))$ mají maximum v téže hodnotě θ . Jestliže tedy

$$L(\theta) = \prod_{i=1}^n f(x_i, \theta)$$

kde $\theta = (\theta_1, \dots, \theta_k)$, pak lze odhad $\hat{\theta}$ parametru θ získat řešením soustavy k rovnic

$$\begin{aligned} \frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_1} &= 0 \\ \dots \frac{\partial L(\theta_1, \dots, \theta_k)}{\partial \theta_k} &= 0 \end{aligned}$$

resp.

$$\begin{aligned} \sum_{i=1}^n \frac{\partial \ln(f(x_i, \theta_1, \dots, \theta_k))}{\partial \theta_1} &= 0 \\ \dots \sum_{i=1}^n \frac{\partial \ln(f(x_i, \theta_1, \dots, \theta_k))}{\partial \theta_k} &= 0 \end{aligned}$$

Příklad 7.3 Uvažujme náhodný výběr velikosti n z Bernoulliho rozdělení $f(x, p) = p^x q^{1-x} I_{\{0,1\}}(x)$ pro $0 \leq p \leq 1$ a $q = 1 - p$. Pozorované hodnoty x_1, \dots, x_n tak budou sekvencí nul a jedniček. Odpovídající funkce věrohodnosti je tedy rovna

$$L(p) = \prod_{i=1}^n p^{x_i} q^{1-x_i} = p^{\sum x_i} q^{n - \sum x_i}$$

²Pouze pro úplnost dodejme, že o pravděpodobnost v pravém slova smyslu se jedná pouze v případě nespojitě sdružené náhodné veličiny.

S využitím substituce $y = \sum x_i$ a následnou logaritmizací získáme

$$\ln(L(p)) = y \ln(p) + (n - y) \ln(q)$$

Derivací dle p pak tento vztah přejde do tvaru

$$\frac{d \ln(p)}{dp} = \frac{y}{p} - \frac{n - y}{q}$$

Jestliže výše uvedený výraz položíme roven nule a řešíme pro p , získáme

$$\hat{p} = \frac{y}{p} = \frac{1}{n} \sum x_i = \bar{x}$$

Příklad 7.4 Uvažujme náhodný výběr velikosti n z normálního rozdělení se sdruženou pravděpodobnostní funkcí

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x_i - \mu)^2} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

Logaritmus funkce věrohodnosti je pak

$$\ln(L) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2}(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

kde $\sigma > 0$ a $-\infty < \mu < \infty$. Maximum $\ln(L)$ nalezneme pomocí derivací

$$\begin{aligned} \frac{\partial \ln(L)}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \\ \frac{\partial \ln(L)}{\partial \sigma^2} &= -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned}$$

položením těchto derivací do rovnosti nule a řešením takto získaných rovnic pro μ a σ^2 , čímž získáme

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

Příklad 7.5 Nechť náhodná veličina X sleduje uniformní rozdělení s pravděpodobnostní funkcí $f(x, \theta) = I_{[\theta - \frac{1}{2}, \theta + \frac{1}{2}]}(x)$, kde $-\infty < \theta < \infty$. To znamená, že $\bar{\Theta}$ představuje reálnou osu. Funkce věrohodnosti pro náhodný výběr velikosti n má pak tvar

$$L(\theta, x_1, \dots, x_n) = \prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n I_{[\theta - \frac{1}{2}, \theta + \frac{1}{2}]}(x_i) = I_{[y_n - \frac{1}{2}, y_1 + \frac{1}{2}]}$$

kde y_1 představuje nejmenší a y_n největší pozorování. Poslední úprava výše uvedeného vztahu vychází ze skutečnosti, že $\prod_{i=1}^n I_{[\theta - \frac{1}{2}, \theta + \frac{1}{2}]}(x_i)$ je rovno jedné,

jsou-li všechna x_1, \dots, x_n v intervalu $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$. To je pravda, jestliže $\theta - \frac{1}{2} \leq y_1$ a $y_n \leq \theta + \frac{1}{2}$ neboli $y_n - \frac{1}{2} \leq \theta \leq y_1 + \frac{1}{2}$. Funkce věrohodnosti je tak rovna jedné pro $y_n - \frac{1}{2} \leq \theta < y_1 + \frac{1}{2}$ nebo nule v ostatních případech. Proto libovolná statistika s hodnotou $\hat{\theta}$, která splňuje podmínku $y_n - \frac{1}{2} \leq \hat{\theta} \leq y_1 + \frac{1}{2}$, je odhadem dle metody maximální věrohodnosti. Příklady takového odhadu jsou $y_n - \frac{1}{2}$, $y_1 + \frac{1}{2}$ nebo $\frac{1}{2}(y_1 + y_n)$.

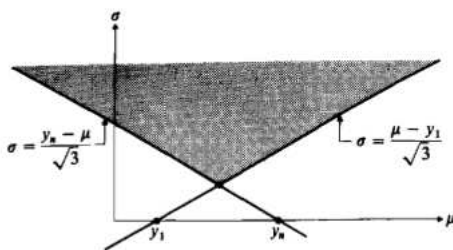
Příklad 7.6 Uvažujme náhodnou veličinu X s uniformním rozdělení a pravděpodobnostní funkcí

$$f(x, \theta) = f(x; \mu, \sigma) = \frac{1}{2\sqrt{3}\sigma} I_{[\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma]}(x)$$

kde $-\infty < \mu < \infty$ a $\sigma > 0$. Věrohodnostní funkce náhodného výběru velikosti n je

$$\begin{aligned} L(\mu, \sigma, x_1, \dots, x_n) &= \left(\frac{1}{2\sqrt{3}\sigma} \right)^n \prod_{i=1}^n I_{[\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma]}(x_i) \\ &= \left(\frac{1}{2\sqrt{3}\sigma} \right)^n I_{[\mu - \sqrt{3}\sigma, y_n]}(y_1) I_{[y_1, \mu + \sqrt{3}\sigma]}(y_n) \\ &= \left(\frac{1}{2\sqrt{3}\sigma} \right)^n I_{[(\mu - y_1)/\sqrt{3}, \infty)}(\sigma) I_{[(y_n - \mu)/\sqrt{3}, \infty)}(\sigma) I_{[y_1, \infty)}(y_n) \end{aligned}$$

kde y_1 představuje nejmenší a y_n největší pozorování. Pravděpodobnostní funkce nabývá hodnoty $(2\sqrt{3}\sigma)^{-n}$ v šedivé části grafu (7.1) a nulové hodnoty v jeho ostatních částech.



Obrázek 7.1: Konvergence k normálnímu rozdělení

Hodnota $(2\sqrt{3}\sigma)^{-n}$ nabývá maxima pro minimální σ , které leží v průsečíku přímek $\mu - \sqrt{3}\sigma = y_1$ a $\mu + \sqrt{3}\sigma = y_n$. Proto odhady pro μ a σ založené na funkci maximální věrohodnosti jsou

$$\hat{\mu} = \frac{1}{2}(y_1 + y_n)$$

a

$$\hat{\sigma} = \frac{1}{2\sqrt{3}}(y_n - y_1)$$

Věta 7.1 (Invariance funkce odhadu metody maximální věrohodnosti) Necht' $\hat{\theta} = \zeta(X_1, \dots, X_n)$ představuje funkci odhadu metody maximální věrohodnosti pro parametr θ pravděpodobnostní funkce $f(x, \theta)$. Předpokládejme, že

θ je jednorozměrné. Dále uvažujme funkci $\tau(\cdot)$, k níž existuje jednorozměrná inverzní funkce. Pak platí, že funkce odhadu metody maximální věrohodnosti $\tau(\theta)$ je $\tau(\hat{\theta})$.

Příklad 7.7 Uvažujme normální rozdělení s danou střední hodnotou μ_0 . Funkce odhadu parametru σ^2 je

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2$$

Funkce odhadu parametru $\ln(\sigma^2)$ je pak

$$\ln(\hat{\theta}) = \ln \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2 \right)$$

Výše uvedenou větu lze dále zobecnit. Nechť $\theta = (\theta_1, \dots, \theta_k)$ je k -rozměrný parametr a $\bar{\Theta}$ představuje množinu všech možných θ . Předmětem našeho zájmu je odhad $\tau(\theta) = (\tau_1(\theta), \dots, \tau_r(\theta))$, kde $1 \leq r \leq k$. Nechť T představuje obor hodnot transformace $\tau(\cdot) = (\tau_1(\cdot), \dots, \tau_r(\cdot))$. To znamená, že T je r -rozměrným prostorem. Definujme $M(\tau, x_1, \dots, x_n) = \sup_{\{\theta: \tau(\theta) = \tau\}} L(\theta, x_1, \dots, x_n)$ ³. Tímto se dostáváme k následující větě.

Věta 7.2 Uvažujme pravděpodobnostní funkci $f(\cdot, \theta_1, \dots, \theta_k)$. Nechť $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$, kde $\hat{\theta}_j = \zeta_j(X_1, \dots, X_n)$, je funkcí odhadu pro $\theta = (\theta_1, \dots, \theta_k)$ dle metody maximální věrohodnosti. Nechť $\tau(\theta) = (\tau_1(\theta), \dots, \tau_r(\theta))$ pro $1 \leq r \leq k$ představuje transformaci prostoru parametrů $\bar{\Theta}$. Pak funkce odhadu pro $\tau(\theta) = (\tau_1(\theta), \dots, \tau_r(\theta))$ dle metody maximální věrohodnosti má tvar $\tau(\hat{\theta}) = (\tau_1(\hat{\theta}), \dots, \tau_r(\hat{\theta}))$.

Důkaz 7.1 Nechť $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$ je odhadem pro $\theta = (\theta_1, \dots, \theta_k)$. Stačí dokázat, že $M(\tau(\hat{\theta}), x_1, \dots, x_n) \geq M(\tau, x_1, \dots, x_n)$ pro libovolné $\tau \in T$. To vyplývá z nerovnosti $M(\tau, x_1, \dots, x_n) = \sup_{\{\theta: \tau(\theta) = \tau\}} L(\theta, x_1, \dots, x_n) \leq \sup_{\theta \in \bar{\Theta}} L(\theta, x_1, \dots, x_n) = L(\hat{\theta}, x_1, \dots, x_n) = \sup_{\{\theta: \tau(\theta) = \tau(\hat{\theta})\}} L(\theta, x_1, \dots, x_n) = M(\tau(\hat{\theta}), x_1, \dots, x_n)$.

Příklad 7.8 Pro normální rozdělení platí $\theta = (\theta_1, \theta_2) = (\mu, \sigma^2)$. Uvažujme $\tau(\theta) = \mu + z_q \sigma$ kde $\phi(z_q) = q$ a $\tau(\theta)$ tak představuje q -tý kvantil. Dle výše uvedené věty je funkce odhadu pro $\tau(\theta)$ založená na metodě maximální věrohodnosti definována jako

$$\bar{X} + z_q \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

7.1.3 Ostatní metody

Existuje řada dalších metod bodového odhadu. Mezi tyto metody patří (a) metoda nejmenších čtverců, (b) Bayesova metoda, (c) chi-kvadrát metoda a (d) metoda nejmenší vzdálenosti. První dvě metody probereme později. V této kapitole se zaměříme na poslední dvě.

³Označení \sup v této knize má shodný význam jako v matematice. Čtenáři, kteří nejsou s tímto označením seznámeni, si namísto něj mohou dosadit pojmem \max .

Chi-kvadrát metoda

Uvažujme náhodný výběr X_1, \dots, X_n z populace $f_X(x, \theta)$. Nechť $\mathcal{S}_1, \dots, \mathcal{S}_k$ představuje rozpětí náhodné veličiny X rozdělené na k disjunktních intervalů. Pokusme se stanovit pravděpodobnost $p_j(\theta)$, s jakou pozorování připadne intervalu \mathcal{S}_j , kde $j = 1, \dots, k$.

Jestliže je $f_X(x, \theta)$ pravděpodobnostní funkcí spojitě náhodné veličiny, pak $p_j(\theta) = P[X \text{ připadne intervalu } \mathcal{S}_j] = \int_{\mathcal{S}_j} f_X(x, \theta) dx$. Z definice je zřejmé, že $\sum_{j=1}^k p_j(\theta) = 1$. Nechť náhodná veličina N_j představuje počet X_i v náhodném výběru, která připadnou intervalu \mathcal{S}_j . Pak $\sum_{j=1}^k N_j = n$, kde n představuje velikost náhodného výběru. Definujme následující sumu

$$\chi^2 = \sum_{j=1}^k \frac{[n_j - np_j(\theta)]^2}{np_j(\theta)}$$

kde n_j je hodnota náhodné veličiny N_j . Čitatel výše uvedené sumy je součtem kvadrátů rozdílů mezi skutečným a očekávaným počtem pozorování, která připadla intervalu \mathcal{S}_j . Odhad θ založený na chi-kvadrat metodě je takové $\hat{\theta}$, které minimalizuje χ^2 . Jedná se tedy o takové θ mezi všemi možnými θ , pro které počet očekávaných pozorování “nejbližší” skutečnému počtu pozorování. Je zřejmé, že odhad $\hat{\theta}$ záleží na konkrétním výběru $\mathcal{S}_1, \dots, \mathcal{S}_k$.

Příklad 7.9 Uvažujme náhodný výběr X_1, \dots, X_n z Bernoulliho rozdělení. Pravděpodobnostní funkce má tedy tvar $f_X(x, \theta) = \theta^x(1 - \theta)^{1-x}$ pro $x = 0, 1$. Náhodná veličina N_j představuje počet pozorování rovných $j = 0, 1$. Proto

$$\begin{aligned} \chi^2 &= \sum_{j=0}^1 \frac{[n_j - np_j(\theta)]^2}{np_j(\theta)} = \frac{[n_0 - n(1 - \theta)]^2}{n(1 - \theta)} + \frac{(n_1 - n\theta)^2}{n\theta} \\ &= \frac{[n - n_1 - n(1 - \theta)]^2}{n(1 - \theta)} + \frac{(n_1 - n\theta)^2}{n\theta} = \frac{(n_1 - n\theta)^2}{n} \frac{1}{\theta(1 - \theta)} \end{aligned}$$

Minimum χ^2 jako funkce θ nastává v $\chi^2 = 0$ pro $\theta = \frac{n_1}{n}$. Proto $\hat{\theta} = \frac{n_1}{n}$.

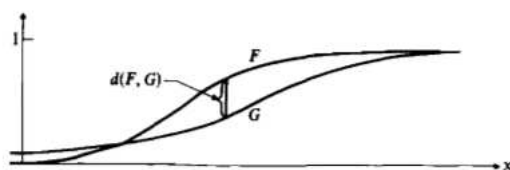
Někdy je složité nalézt $\hat{\theta}$, které minimalizuje χ^2 . Proto bývá často $np_j(\theta)$ ve jmenovateli nahrazeno n_j^4 . Pak hovoříme o tzv. modifikovaném χ^2 , které je definováno jako

$$\chi_{mod}^2 = \sum_{j=1}^k \frac{[n_j - np_j(\theta)]^2}{n_j}$$

Tento modifikovaný odhad je opět založen na hodnotě $\hat{\theta}$, která minimalizuje χ_{mod}^2 .

Metoda minimální vzdálenosti

Nechť X_1, \dots, X_n představuje náhodný výběr z populace s kumulativní distribuční funkcí $F(X, \theta)$. Definujme funkci $d(F, G)$, která představuje vzdálenost mezi dvěma kumulativními distribučními funkcemi F a G . Příkladem takovéto funkce



Obrázek 7.2: Vzdálenost mezi dvěma kumulativními pravděpodobnostními funkcemi

může být $\sup_x |F(x) - G(x)|$, která představuje největší vertikální vzdálenost mezi funkcemi F a G a je ilustrována obrázkem (7.2).

Odhad $\hat{\theta}$ parametru θ získaný pomocí metody minimální vzdálenosti je tedy představován takovou hodnotou, která minimalizuje $d(F(x, \theta), F_n(x))$, kde $F_n(x)$ je výběrovou kumulativní distribuční funkcí⁵. Bohužel ne vždy je snadné určit hodnotu $\hat{\theta}$, která minimalizuje funkci vzdálenosti $d(F(x, \theta), F_n(x))$. Následující příklad je v tomto směru výjimkou.

Důkaz 7.2 *Nechť X_1, \dots, X_n představuje náhodný výběr z Bernoulliho rozdělení. Pak platí*

$$F(x, \theta) = (1 - \theta)I_{[0,1)}(x) + I_{[1,\infty)}(x)$$

kde $0 \leq \theta \leq 1$. Dále nechť n_j představuje počet pozorování rovných j , kde $j = 0, 1$. Pak zjevně platí

$$F_n(x) = \frac{n_0}{n}I_{[0,1)}(x) + I_{[1,\infty)}(x)$$

Jestliže použijeme funkci vzdálenosti $d(F, G) = \sup_x |F(x) - G(x)|$, pak tato funkce nabývá minima v bodě $1 - \theta = \frac{n_0}{n}$ neboli $\theta = \frac{n_1}{n} = \sum \frac{x_i}{n}$. Proto $\hat{\theta} = \bar{x}$.

.

7.2 Vlastnosti funkcí odhadu

V předchozím textu jsme ukázali několik metod bodového odhadu. Tímto vystává přirozená otázka, která z těchto metod je v dané situaci nejvhodnější.

7.2.1 Blízkost funkce odhadu

Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta)$. Funkce odhadu $\tau(\theta)$ je statistika, řekněme $t(X_1, \dots, X_n)$, jejíž hodnota je použita jako odhad $\tau(\theta)$. V ideálním případě bychom chtěli, aby hodnota $t(X_1, \dots, X_n)$ byla rovna hledané hodnotě $\tau(\theta)$. To však, s výjimkou několika triviálních případů, není možné.

Příklad 7.10 *Uvažujme náhodný výběr z populace $f(x, \theta) = I_{(\theta - \frac{1}{2}, \theta + \frac{1}{2})}(x)$, kde θ je přirozené číslo. Je zřejmé, že θ je možné přesně odhadnout na základě jednoho pozorování x_1 a to tak, že $t(x_1)$ bude představovat hodnotu přirozeného čísla nejbližšího x_1 .*

⁴Pokud $n_j = 0$, pak namísto n_j použijeme 1.

⁵Pouze připomeňme, že na základě teorie velkých čísel konverguje výběrová kumulativní distribuční funkce $F_n(x)$ ke kumulativní distribuční funkci $F(x)$ dané populace.

Vzhledem k tomu, že $t(X_1, \dots, X_n)$ není ve většině případů přesným odhadem $\tau(\theta)$, požadujeme, aby bylo alespoň “blízkým” odhadem. $T = t(X_1, \dots, X_n)$ je statistika, která sleduje určité pravděpodobnostní rozdělení. Proto je logické požadovat, aby se hodnoty t náhodné veličiny T “koncentrovaly” poblíž $\tau(\theta)$.

Definice 7.4 (Koncentrovaná a nejkoncentrovanější funkce odhadu)

Uvažujme dvě funkce odhadu $T = t(X_1, \dots, X_n)$ a $T' = t'(X_1, \dots, X_n)$ pro $\tau(\theta)$. Říkáme, že T' je více koncentrované než T , jestliže

$$P_\theta[\tau(\theta) - \lambda < T' \leq \tau(\theta) + \lambda] \geq P_\theta[\tau(\theta) - \lambda < T \leq \tau(\theta) + \lambda]$$

pro všechna $\lambda > 0$ a každé $\theta \in \underline{\Theta}$. Funkci odhadu $T^* = t^*(X_1, \dots, X_n)$ nazýváme nejkoncentrovanější funkcí odhadu, jestliže je více koncentrovaná než libovolná jiná funkce odhadu.

Ve výše uvedené definici jsme použili značení $P_\theta[\cdot]$, abychom zdůraznili, že pravděpodobnost $P_\theta[\tau(\theta) - \lambda < T \leq \tau(\theta) + \lambda]$ závisí na θ skrze pravděpodobnostní rozdělení T .

Vedle koncentrované a nejkoncentrovanější funkce odhadu existuje také obdobná Pitmanova definice blízké a nejbližší funkce odhadu.

Definice 7.5 (Blízká a nejbližší funkce odhadu)

Uvažujme dvě funkce odhadu $T = t(X_1, \dots, X_n)$ a $T' = t'(X_1, \dots, X_n)$ pro $\tau(\theta)$. Říkáme, že T' je více koncentrované než T , jestliže

$$P_\theta[|T' - \tau(\theta)| < |T - \tau(\theta)|] \geq \frac{1}{2}$$

pro každé $\theta \in \underline{\Theta}$. Funkci odhadu $T^* = t^*(X_1, \dots, X_n)$ nazýváme nejbližší funkcí odhadu, jestliže je bližší než libovolná jiná funkce odhadu.

Základní slabinou výše uvedených definic je skutečnost, že nejkoncentrovanější resp. nejbližší funkce odhadu existuje pouze ve výjimečných případech. Důvodem je existence přílišného množství možných funkcí odhadu, než aby bylo možné některou z nich prohlásit za nejvhodnější. Navíc velmi často nelze ani tvrdit, že jedna funkce odhadu je koncentrovanější resp. bližší než jiná, protože to se obvykle odvíjí od konkrétní hodnoty parametru θ .

V předchozím textu jsme uvažovali fixní velikost náhodného výběru. Z pohledu “blízkosti” funkce odhadu je však také důležité, aby se přesnost odhadu zlepšovala s rostoucí velikostí náhodného výběru. V souvislosti s náhodným vzorkem dané velikosti hovoříme o vlastnostech funkce odhadu jako o vlastnostech malého výběru. O vlastnostech funkce odhadu, které jsou definované pro rostoucí velikost náhodného výběru, pak hovoříme jako o vlastnostech velkého výběru.

7.2.2 Střední chyba čtverců

Definice 7.6 (MSE (mean-squared error) - střední chyba čtverců)

Nechť $T = t(X_1, \dots, X_n)$ je funkcí odhadu pro $\tau(\theta)$. Pak $E_\theta[(T - \tau(\theta))^2]$ nazýváme MSE pro funkci odhadu $T = t(X_1, \dots, X_n)$ a značíme ji jako $MSE_t(\theta)$.

θ v případě $E_\theta[\cdot]$ indikuje pravděpodobnostní funkci, ze které pochází náhodný výběr. To znamená, že

$$\begin{aligned} E_\theta[(T - \tau(\theta))^2] &= E_\theta[(t(X_1, \dots, X_n) - \tau(\theta))^2] \\ &= \int \cdots \int (t(x_1, \dots, x_n) - \tau(\theta))^2 f(x_1, \theta) \cdots f(x_n, \theta) dx_1 \cdots dx_n \end{aligned}$$

kde $f(x, \theta)$ představuje pravděpodobnostní funkci populace.

Pokud porovnáváme funkce odhadu pomocí MSE, vybereme funkci s nejmenší MSE. Teoreticky je tak možné definovat funkci odhadu s nejmenší MSE, avšak stejně jako v předchozím případě, i zde je spíše výjimkou, pokud takováto funkce existuje.

Příklad 7.11 *Nechť X_1, \dots, X_n představuje náhodný výběr z $f(x, \theta)$, kde θ je reálné číslo. Předpokládejme, že předmětem zájmu je odhad parametru θ , tj. $\tau(\theta) = \theta$.*

Hledáme tedy funkci odhadu $T^ = t^*(X_1, \dots, X_n)$ takovou, že $MSE_{t^*}(\theta) \leq MSE_t(\theta)$ pro každé θ a libovolné $T = t(X_1, \dots, X_n)$. Uvažujme rodinu funkcí odhadu $T_{\theta_0} = t_{\theta_0}(X_1, \dots, X_n) \equiv \theta_0$ pro $\theta_0 \in \underline{\Theta}$. Pro libovolné $\theta_0 \in \underline{\Theta}$ tedy funkce odhadu T_{θ_0} “ignoruje” pozorování a parametru θ přímo přiřadí hodnotu θ_0 . Všimněme si, že*

$$MSE_{t_{\theta_0}}(\theta) = E_\theta[(t(X_1, \dots, X_n) - \theta)^2] = E_\theta[(\theta_0 - \theta)^2] = (\theta_0 - \theta)^2$$

Proto $MSE_{t_{\theta_0}}(\theta_0) = 0$. Má-li tedy existovat $T^ = t^*(X_1, \dots, X_n)$, které splňuje $MSE_{t^*}(\theta) \leq MSE_t(\theta)$ pro každé θ a funkci t , pak $MSE_{t^*} \equiv 0$. Aby funkce bodového odhadu t^* měla MSE rovnu nule, musí parametr θ odhadnout vždy správně. To znamená, že musí být z náhodného výběru schopna vždy určit správnou hodnotu parametru θ .*

Jedním z důvodů, proč nejsme schopni nalézt funkci odhadu s nejmenší chybou čtverců je skutečnost, že počet potenciálních funkcí je příliš veliký. Tato množina potenciálních funkcí také zahrnuje funkce, které jsou extrémně “předpojaté” ve prospěch určité konkrétní hodnoty parametru θ^6 . Proto je vhodné množinu všech potencionálních funkcí odhadu omezit další podmínkou. Touto další podmínkou je podmínka nezkreslení.

Definice 7.7 (Nezkreslená funkce odhadu) *O funkci odhadu $T = t(X_1, \dots, X_n)$ říkáme, že je nezkreslená, tehdy a jen tehdy, jestliže*

$$E_\theta[T] = E_\theta[t(X_1, \dots, X_n)] = \tau(\theta)$$

pro všechna $\theta \in \underline{\Theta}$.

Příklad 7.12 *Je tedy zřejmé, že funkce odhadu $t_{\theta_0}(X_1, \dots, X_n)$ z předchozího příkladu není nezkreslenou funkcí odhadu parametru θ , protože*

$$E_\theta[t_{\theta_0}(X_1, \dots, X_n)] = E_\theta[\theta_0] = \theta_0 \neq \theta$$

⁶Jako ilustrace může posloužit funkce $t_{\theta_0}(X_1, \dots, X_n)$ z předchozího příkladu, která byla “předpojatá” ve prospěch parametru θ_0 . Tato funkce totiž bez ohledu na reálný výsledek pozorování stanovila hodnotu parametru θ vždy rovnu θ_0 .

Tvrzení 7.1 *Jestliže je T nekreslenou funkcí odhadu pro $\tau(\theta)$, pak platí*

$$MSE_t(\theta) = D[T]$$

Toto tvrzení vyplývá ze vztahu

$$MSE_t(\theta) = D[T] + (\tau(\theta) - E_\theta[T])^2$$

Důkaz 7.3

$$\begin{aligned} MSE_t(\theta) &= E_\theta[(T - \tau(\theta))^2] = E_\theta \left[((T - E_\theta[T]) - (\tau(\theta) - E_\theta[T]))^2 \right] \\ &= E_\theta[(T - E_\theta[T])^2] - 2(\tau(\theta) - E_\theta[T])E_\theta[T - E_\theta[T]] + E_\theta[(\tau(\theta) - E_\theta[T])^2] \\ &= D[T] + (\tau(\theta) - E_\theta[T])^2 \end{aligned}$$

Člen $\tau(\theta) - E_\theta[T]$ nazýváme zkreslením funkce odhadu T . Tento člen může být kladný, záporný ale i nulový.

Příklad 7.13 *Uvažujme náhodný výběr X_1, \dots, X_n z $f(x, \theta) = \phi_{\mu, \sigma^2}(x)$. Z příkladu (7.5) víme, že funkce odhadu dle metody maximální věrohodnosti pro parametr μ resp. σ je \bar{X} resp. $\frac{1}{n} \sum (X_i - \bar{X})^2$. Platí $E_\theta[\bar{X}] = \mu$, a proto je \bar{X} nekreslenou funkcí odhadu pro μ . MSE funkce \bar{X} je*

$$E_\theta[(\bar{X} - \mu)^2] = D[\bar{X}] = \frac{\sigma^2}{n}$$

Z předchozího textu také víme, že $E_\theta[S_n^2] = \sigma^2$, a proto

$$\begin{aligned} E_\theta \left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right] &= \frac{n-1}{n} E_\theta \left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \right] \\ \frac{n-1}{n} E[S_n^2] &= \frac{n-1}{n} \sigma^2 \end{aligned}$$

Funkce odhadu pro σ tak je zkreslená. MSE pro $\frac{1}{n} \sum (X_i - \bar{X})^2$ lze s pomocí věty (6.2) vyjádřit jako

$$\begin{aligned} E_\theta \left[\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 - \sigma^2 \right)^2 \right] &= D \left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right] + \left(\sigma^2 - E_\theta \left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right] \right)^2 \\ &= \frac{(n-1)^2}{n^2} D[S_n^2] + \left(\sigma^2 - \frac{n-1}{n} \sigma^2 \right)^2 = \frac{(n-1)^2}{n^2} \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right) + \frac{\sigma^4}{n^2} \end{aligned}$$

V následujícím textu bude MSE hlavním nástrojem pro posouzení funkce odhadu.

7.2.3 Konzistentnost a BAN

Definice 7.8 (Konzistentnost ve smyslu MSE) *Nechť T_1, \dots, T_n představuje posloupnost funkcí odhadu pro $\tau(\theta)$, kde $T_n = t_n(X_1, \dots, X_n)$ je definováno pro náhodný výběr velikosti n . Tuto posloupnost funkcí odhadu označujeme jako konzistentní ve smyslu MSE pouze tehdy a jen tehdy, jestliže*

$$\lim_{n \rightarrow \infty} E_\theta[(T_n - \tau(\theta))^2] = 0$$

pro všechna θ z $\underline{\Theta}$.

Konzistentnost ve smyslu MSE mimojiné znamená, že zkreslení a rozptyl T_n se blíží nule s tím, jak se n blíží nekonečnu, protože $E_\theta[(T_n - \tau(\theta))^2] = D[T_n] + (\tau(\theta) - E_\theta[T_n])^2$.

Příklad 7.14 Uvažujme náhodný výběr z pravděpodobnostního rozdělení se střední hodnotou μ a rozptylem σ^2 . Necht $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ představuje posloupnost funkcí odhadu pro μ a $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ pak posloupnost funkcí odhadu pro σ^2 . Platí

$$\lim_{n \rightarrow \infty} E[(\bar{X}_n - \mu)^2] = \lim_{n \rightarrow \infty} D[\bar{X}_n] = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0$$

a proto je posloupnost $\{\bar{X}_n\}$ konzistentní funkcí odhadu pro μ . Dále platí

$$\lim_{n \rightarrow \infty} E[(S_n^2 - \sigma^2)^2] = \lim_{n \rightarrow \infty} D[S_n^2] = \lim_{n \rightarrow \infty} \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \right) = 0$$

a proto také $\{S_n^2\}$ představuje konzistentní posloupnost funkcí odhadu pro σ^2 . Všimněme si, že je-li $T_n = \frac{1}{n} \sum (X_i - \bar{X})^2$, pak je posloupnost $\{T_n\}$ konzistentní funkcí odhadu ve smyslu MSE nejen pro μ ale také pro σ^2 .

Vedle výše uvedené definice existuje také “slabší” forma konzistence.

Definice 7.9 (Slabší forma konzistence ve smyslu MSE) Uvažujme posloupnost T_1, \dots, T_n funkcí odhadu pro $\tau(\theta)$, kde $T_n = t_n(X_1, \dots, X_n)$. Posloupnost $\{T_n\}$ nazýváme “slabě” konzistentní ve smyslu MSE, jestliže je pro každé $\epsilon > 0$ a $\theta \in \underline{\Theta}$ platí

$$\lim_{n \rightarrow \infty} P_\theta[\tau(\theta) - \epsilon < T_n < \tau(\theta) + \epsilon] = 1$$

Důkaz 7.4 S použitím Chebyshevovy nerovnosti lze dokázat

$$\begin{aligned} P_\theta[\tau(\theta) - \epsilon < T_n < \tau(\theta) + \epsilon] &= P[|T_n - \tau(\theta)| < \epsilon] \\ &= P_\theta[(T_n - \tau(\theta))^2 < \epsilon^2] \geq 1 - \frac{E_\theta[(T_n - \tau(\theta))^2]}{\epsilon^2} \end{aligned}$$

S tím, jak se n blíží nekonečnu, se $E_\theta[(T_n - \tau(\theta))^2]$ blíží nule. Proto

$$\lim_{n \rightarrow \infty} P_\theta[\tau(\theta) - \epsilon < T_n < \tau(\theta) + \epsilon] = 1$$

Jestliže je funkce odhadu konzistentní, pak je také slabě konzistentní. Opačné tvrzení však neplatí.

Definice 7.10 (BAN (best asymptotically normal) funkce odhadu)

Posloupnost funkcí bodového odhadu T_1^*, \dots, T_n^* pro $\tau(\theta)$ nazýváme BAN (best asymptotically normal), jestliže splňuje následující čtyři podmínky

1. Pravděpodobnostní rozdělení $\sqrt{n}[T_n^* - \tau(\theta)]$ se blíží normálnímu rozdělení s nulovou střední hodnotou a rozptylem $\sigma^{*2}(\theta)$ s tím, jak se n blíží nekonečnu.
2. Pro každé $\epsilon > 0$ a každé $\theta \in \bar{\Theta}$ platí

$$\lim_{n \rightarrow \infty} P_\theta[|T_n^* - \tau(\theta)| > \epsilon] = 0$$

3. Nechť $\{T_n\}$ představuje libovolnou alternativní posloupnost slabě konzistentních funkcí odhadu s pravděpodobnostním rozdělením $\sqrt{n}[T_n - \tau(\theta)]$, které se blíží normálnímu rozdělení s nulovou střední hodnotou a rozptylem $\sigma^2(\theta)$.
4. $\sigma^2(\theta)$ není menší než $\sigma^{*2}(\theta)$ pro libovolný otevřený interval hodnot θ .

Namísto označení BAN se někdy používá CANE (consistent asymptotically normal efficient).

Příklad 7.15 Pro náhodné výběry z normálního rozdělení se střední hodnotou μ a rozptylem σ^2 , je posloupnost $T_n^* = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$ je posloupností BAN pro μ . Limitní pravděpodobnostní funkci $\sqrt{n}(\bar{X}_n - \mu)$ je normální rozdělení s nulovou střední hodnotou a rozptylem σ^2 . Žádná jiná funkce bodového odhadu nemůže mít menší rozptyl pro libovolný interval hodnot μ . Nicméně existují i jiné BAN funkce bodového odhadu pro μ , které mají shodné limitní normální rozdělení. Jako příklad uveďme

$$T'_n = \frac{1}{n+1} \sum_{i=1}^n X_i$$

7.2.4 Ztrátová a riziková funkce

Definice 7.11 (Ztrátová funkce) Uvažujme odhad parametru $\tau(\theta)$. Nechť t označuje hodnotu odhadu pro $\tau(\theta)$. Ztrátová funkce $l(t, \theta)$ je reálná funkce, která splňuje

1. $l(t, \theta) \geq 0$ pro všechny možné odhady t a všechna $\theta \in \bar{\Theta}$,
2. $l(t, \theta) = 0$ pro $t = \tau(\theta)$

Z logiky věci vyplývá, že ztrátová funkce by měla vracet větší hodnoty pro větší odchylky t od skutečné hodnoty $\tau(\theta)$ a naopak. Několik možných ztrátových funkcí je uvedeno v následujícím příkladu.

Příklad 7.16

1. $l_1(t, \theta) = (t - \tau(\theta))^2$
2. $l_2(t, \theta) = |t - \tau(\theta)|^2$
3. $l_3(t, \theta) = \begin{cases} A & \text{pro } |t - \tau(\theta)| > \epsilon \\ 0 & \text{pro } |t - \tau(\theta)| \leq \epsilon \end{cases}, \text{ kde } A > 0$
4. $l_4(t, \theta) = \rho(\theta)|t - \tau(\theta)|^r$ pro $\rho(\theta)$ a $r > 0$

Funkce l_1 se nazývá ztrátová funkce chyby čtverců a funkce l_2 se nazývá ztrátová funkce absolutní chyby. Jak l_1 tak l_2 je rostoucí funkce chyby $t - \tau(\theta)$. Naproti tomu funkce l_3 nabývá pouze hodnot A a 0 v závislosti na tom, zda-li se chyba $t - \tau(\theta)$ nachází v intervalu $\tau(\theta) \pm \epsilon$ či nikoliv. Konečně funkce l_4 je obecnou formou ztrátové funkce, která zahrnuje l_1 a l_2 jako specifické případy.

Při odhadu parametrů je naším cílem zvolit takovou funkci odhadu $T = t(X_1, \dots, X_n)$, která bude minimalizovat ztrátovou funkci⁷. Ztrátová funkce závisí na t , což je hodnota funkce bodového odhadu T . To znamená, že se hodnota ztrátové funkce odvíjí od konkrétního náhodného výběru $X_1 = x_1, \dots, X_n = x_n$. Je zřejmé, že ztrátovou funkci nemůžeme minimalizovat pro každý možný náhodný výběr, avšak můžeme se ji pokusit minimalizovat v “průměru”. Tímto se dostáváme k tzv. rizikové funkci.

Definice 7.12 (Riziková funkce) Pro danou ztrátovou funkci $l(t, \theta(t))$ a funkci bodového odhadu $T = t(X_1, \dots, X_n)$ je riziková funkce definována jako

$$r_t(\theta) = E_\theta[l(T, \theta)]$$

Riziková funkce tak představuje průměrnou hodnotu ztrátové funkce.

Rizikovou funkci lze v zásadě vypočítat dvěma způsoby. Prvním z nich je pomocí sdružené pravděpodobnostní funkce náhodného výběru

$$\begin{aligned} E_\theta[l(T, \theta)] &= E_\theta[l(t(X_1, \dots, X_n), \theta)] \\ &= \int \cdots \int l(t(x_1, \dots, x_n), \theta) \prod_{i=1}^n f(x_i, \theta) dx_i \end{aligned}$$

Druhý způsob se pak opírá o T a její pravděpodobnostní funkci

$$E_\theta[l(T, \theta)] = \int l(\cdot, \theta) f_T(t) dt$$

Příklad 7.17 Uvažujme ztrátové funkce z předchozího příkladu. Odpovídající rizikové funkce jsou

1. $E_\theta[(T - \tau(\theta))^2]$, což je již dříve uvedená MSE
2. $E_\theta[|T - \tau(\theta)|]$
3. $A \cdot P_\theta[|T - \tau(\theta)| > \epsilon]$
4. $\rho(\theta) E_\theta[|T - \tau(\theta)|^r]$

Definice 7.13 (Přijatelná funkce odhadu) Uvažujme dvě funkce odhadu $T_1 = t_1(X_1, \dots, X_n)$ a $T_2 = t_2(X_1, \dots, X_n)$. Funkce odhadu t_1 je považována za nadřazenou funkci odhadu t_2 tehdy a jen tehdy, jestliže

$$R_{t_1}(\theta) \geq r_{t_2}(\theta)$$

pro všechna $\theta \in \bar{\Theta}$ a

$$R_{t_1}(\theta) < r_{t_2}(\theta)$$

pro alespoň jedno $\theta \in \bar{\Theta}$.

Stejně jako v předchozích případech je také ztrátová funkce funkcí parametru θ . Určení nejlepší funkce odhadu tak nemusí být jednoznačné. Částečným řešením tohoto problému je koncept tzv. minimaxu.

⁷Předpokládáme, že ztrátovou funkci známe. Volba vhodné ztrátové funkce však nemusí být tak jednoduchá, jak by se mohlo na první pohled zdát.

Definice 7.14 (Minimax) Funkci bodového odhadu t^* nazýváme *minimaxem* tehdy a jen tehdy, jestliže

$$\sup_{\theta} \mathfrak{r}(t^*)(\theta) \leq \sup_{\theta} \mathfrak{r}_t(\theta)$$

pro každou alternativní funkci odhadu t .

7.3 Dostatečnost

7.3.1 Dostatečná statistika

Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$. Statistiku jsme definovali jako funkci náhodného výběru. Definičním oborem této funkce je množina všech hodnot, jichž může náhodný výběr X_1, \dots, X_n nabývat, a oborem hodnot je pak reálná osa. Na statistiku $T = t(X_1, \dots, X_n)$ je také možno pohlížet jako na náhodnou veličinu. Funkce $t(\cdot, \dots, \cdot)$ tedy transformuje n náhodných veličin X_1, \dots, X_n do jedné náhodné veličiny. Předmětem našeho zájmu bude, zda-li při této transformaci dochází ke ztrátě informace či nikoliv.

Nechť \mathfrak{X} označuje rozpětí hodnot, jichž může nabývat (X_1, \dots, X_n) . Pomocí statistiky lze definovat segment prostoru \mathfrak{X} . Uvažujme statistiku $T = t(X_1, \dots, X_n)$. Libovolná pozorování (x_1, \dots, x_n) , pro která platí $t(x_1, \dots, x_n) = t_0$, totiž představují segment prostoru \mathfrak{X} definovaný hodnotou $T = t_0$.

Příklad 7.18 Uvažujme náhodný výběr (X_1, X_2, X_3) z Bernoulliho rozdělení. Prostor \mathfrak{X} se skládá z osmi bodů, a to $(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (0, 1, 1), (1, 1, 0), (1, 0, 1)$ a $(1, 1, 1)$. Definujme statistiku $T = X_1 + X_2 + X_3$. Tato statistika může nabývat čtyř hodnot, a proto rozděluje prostor \mathfrak{X} na čtyři segmenty, a to $\{(0, 0, 0)\}, \{(0, 0, 1), (0, 1, 0), (1, 0, 0)\}, \{(0, 1, 1), (1, 1, 0), (1, 0, 1)\}$ a $\{(1, 1, 1)\}$. Prostor \mathfrak{X} by na shodné čtyři segmenty rozdělily také statistiky $T = 6(x_1 + x_2 + x_3)^2$ a $T = x_1^2 + x_2^2 + x_3^2$. Jestliže by tedy naším cílem bylo “zredukovat” prostor \mathfrak{X} , poskytly by nám všechny tyto tři statistiky stejnou službu.

Definice 7.15 (Dostatečná statistika) Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$, kde θ může představovat jak skalár tak vektor. Předpokládejme, že funkční předpis $f(\cdot, \theta)$ je nám, s výjimkou parametru θ , znám. Statistika $S = \mathfrak{s}(X_1, \dots, X_n)$ je dostatečná tehdy a jen tehdy, jestliže podmníněné pravděpodobnostní rozdělení X_1, \dots, X_n pro dané $S = s$ není funkcí θ .

Dostatečná statistika tedy zredukuje prostor \mathfrak{X} takovým způsobem, že nedojde ke ztrátě informace o θ . Ta je obsažena v náhodném výběru X_1, \dots, X_n , což je důsledkem požadavku, aby podmníněné pravděpodobnostní rozdělení X_1, \dots, X_n nezáviselo na θ . Nelze tedy očekávat, že získáme informaci o θ tím, že budeme provádět náhodný výběr z pravděpodobnostního rozdělení, které není funkcí θ .

Příklad 7.19 Nechť X_1, X_2, X_3 představuje náhodný výběr z Bernoulliho rozdělení. Uvažujme dvě statistiky $S = \mathfrak{s}(X_1, X_2, X_3) = X_1 + X_2 + X_3$ a $T = t(X_1, X_2, X_3) = X_1 X_2 + X_3$. Ukažme, že S je na rozdíl od T dostatečná statistika.

	hodnoty S	hodnoty T	$f_{X_1, X_2, X_3 S}$	$f_{X_1, X_2, X_3 T}$
$(0, 0, 0)$	0	0	1	$\frac{1-p}{1+p}$
$(0, 0, 1)$	1	1	$\frac{1}{3}$	$\frac{1-p}{1+2p}$
$(0, 1, 0)$	1	0	$\frac{1}{3}$	$\frac{p}{1+p}$
$(1, 0, 0)$	1	0	$\frac{1}{3}$	$\frac{p}{1+p}$
$(0, 1, 1)$	2	1	$\frac{1}{3}$	$\frac{p}{1+2p}$
$(1, 0, 1)$	2	1	$\frac{1}{3}$	$\frac{p}{1+2p}$
$(1, 1, 0)$	2	1	$\frac{1}{3}$	$\frac{p}{1+2p}$
$(1, 1, 1)$	3	2	1	1

Tabulka 7.2: Dostatečná statistika

První sloupec tabulky (7.2) představuje prostor \mathfrak{X} . Podmíněné pravděpodobnosti lze vypočítat obvyklým způsobem. Jako příklad uveďme

$$\begin{aligned}
 f_{X_1, X_2, X_3|S=1}(0, 1, 0|1) &= P[X_1 = 0, X_2 = 1, X_3 = 0|S = 1] \\
 &= \frac{P[X_1 = 0, X_2 = 1, X_3 = 0, S = 1]}{P[S = 1]} \\
 &= \frac{(1-p)p(1-p)}{\binom{3}{1}p(1-p)^2} = \frac{1}{3}
 \end{aligned}$$

a

$$\begin{aligned}
 f_{X_1, X_2, X_3|T=0}(0, 1, 0|0) &= \frac{P[X_1 = 0, X_2 = 1, X_3 = 0, T = 0]}{P[T = 0]} \\
 &= \frac{(1-p)^2p}{(1-p)^3 + 2(1-p)^2p} = \frac{p}{1-p+2p} = \frac{p}{1+p}
 \end{aligned}$$

Podmíněná pravděpodobnostní funkce náhodného výběru pro danou hodnotu $S = s$ je nezávislá na parametru p , a proto je S dostatečnou statistikou. Naproti tomu podmíněná pravděpodobnostní funkce náhodného výběru pro danou hodnotu $T = t$ je funkcí parametru p , a proto se o dostatečnou statistiku nejedná.

Všimněme si, že redukce prostoru \mathfrak{X} pomocí statistiky T je výraznější než pomocí statistiky S . Přírozenou otázkou je, zda-li existuje dostatečná statistika, která představuje větší redukci prostoru \mathfrak{X} než S . Odpověď zní “ne” a lze ji dokázat zkoumáním množin všech segmentů prostoru \mathfrak{X} , které rozdělí tento prostor na tři a méně částí.

V případě náhodného výběru z pravděpodobnostní funkce spojitě náhodné veličiny nemusí být význam pojmu “podmíněné rozdělení X_1, \dots, X_n pro $S = s$ ”, který se objevuje v předchozí definici, na první pohled patrný, protože $P[S = s] = 0$. Existují dvě možné interpretace. První je založena na skutečnosti, že dostatečnost statistiky $S = s(X_1, \dots, X_n)$ lze dokázat tím, že prokážeme nezávislost $P[X_1 \leq x_1, \dots, X_n \leq x_n|S = s]$ na θ , kde $P[X_1 \leq x_1, \dots, X_n \leq x_n|S = s]$ je definováno ve smyslu $P[A|X = x] = \lim_{0 < h \rightarrow 0} P[A|x-h < X < x+h]$. Druhá interpretace je založena na transformaci jedna ku jedné z X_1, \dots, X_n na S, Y_2, \dots, Y_n . Jestliže je pravděpodobnostní funkce Y_2, \dots, Y_n pro dané $S = s$ nezávislé na θ , pak je X_1, \dots, X_n pro $S = s$ nezávislé na θ , a S tak představuje dostatečnou statistiku.

Příklad 7.20 Uvažujme náhodný výběr X_1, X_2 z populace $f(\cdot, \theta) = \phi_{\theta,1}(\cdot)$ a statistiku $S = \mathfrak{s}(X_1, X_2) = X_1 + X_2$. Pokusme se dokázat, že T je dostatečnou statistikou.

Začněme druhou z interpretací. Transformace $z(X_1, X_2)$ na S, Y_2 , kde $S = X_1 + X_2$ a $Y_2 = X_2 - X_1$ je transformací jedna ku jedné, takže stačí dokázat, že $f_{Y_2|S}(y_2|s)$ je nezávislé na θ . S využitím nezávislosti $X_1 + X_2$ a $X_2 - X_1$ lze dokázat, že

$$f_{Y_2|S}(y_2|s) = \frac{f_{Y_2,S}(y_2, s)}{f_S(s)} = \frac{f_{Y_2}(y_2)f_S(s)}{f_S(s)} = f_{Y_2}(y_2)$$

Zároveň však platí

$$f_{Y_2}(y_2) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2}\frac{y_2^2}{2}}$$

protože $Y_2 \sim N(0, 2)$. $f_{Y_2|S}(y_2|s)$ je tedy nezávislé na θ a S je tak dostatečnou statistikou.

Důkaz pomocí druhé interpretace je poněkud složitější. V tomto případě musíme dokázat, že $P[X_1 \geq x_1, X_2 \geq x_2 | S = s]$ je nezávislé na θ . Bez ztráty obecnosti můžeme předpokládat $x_1 \leq x_2$ a rozlišovat tak tři základní případy

$$1. s \leq x_1$$

$$2. x_1 < s < x_2$$

$$3. x_2 < s$$

$P[X_1 \leq x_1, X_2 \leq x_2 | S = s]$ je pro třetí případ rovno nule a tudíž nezávislé na θ . Důkaz dále ilustrujeme pro první případ⁸.

$$\begin{aligned} P[X_1 \leq x_1, X_2 \leq x_2 | S = s] &= \lim_{h \rightarrow 0} P[X_1 \leq x_1, X_2 \leq x_2 | s - h < S < s + h] \\ &= \frac{\lim_{h \rightarrow 0} \frac{1}{2h} P[X_1 \leq x_1, X_2 \leq x_2, s - h < S < s + h]}{\lim_{h \rightarrow 0} \frac{1}{2h} P[s - h < S < s + h]} \\ &= \frac{\lim_{h \rightarrow 0} \frac{1}{2h} P[X_1 \leq x_1, X_2 \leq x_2, s - h < S < s + h]}{f_S(s)} \end{aligned}$$

Všimněme si, že

$$\begin{aligned} &\lim_{h \rightarrow 0} \frac{1}{2h} \int_{s+h-x_2}^{x_1} \int_{s-h-u}^{s+h-u} f_{X_1}(u) f_{X_2}(v) dv du \\ &\leq \lim_{h \rightarrow 0} \frac{1}{2h} P[X_1 \leq x_1, X_2 \leq x_2, s - h < X_1 + X_2 < s + h] \\ &\leq \lim_{h \rightarrow 0} \frac{1}{2h} \int_{s-h-x_2}^{x_1} \int_{s-h-u}^{s+h-u} f_{X_1}(u) f_{X_2}(v) dv du \end{aligned}$$

a proto

$$\begin{aligned} &\lim_{h \rightarrow 0} \frac{1}{2h} P[X_1 \leq x_1, X_2 \leq x_2, s - h < X_1 + X_2 < s + h] \\ &= \int_{s-x_2}^{x_1} f_{X_1}(u) f_{X_2}(s-u) du \end{aligned}$$

⁸Postup v druhém případě je analogický.

Platí tedy

$$\begin{aligned}
 P[X_1 \leq x_1, X_2 \leq x_2 | S = s] &= \frac{\int_{s-x_2}^{x_1} f_{X_1}(u) f_{X_2}(s-u) du}{f_S(s)} \\
 &= \frac{\int_{s-x_2}^{x_1} \frac{1}{2\pi} e^{-\frac{1}{2}[u^2 - 2u\theta + \theta^2 + (s-u)^2 - 2(s-u)\theta + \theta^2]} du}{\frac{1}{\sqrt{4\pi}} e^{-\frac{1}{2}\left(\frac{s-2\theta}{\sqrt{2}}\right)^2}} \\
 &= \frac{\frac{1}{\sqrt{2\pi}} \int_{s-x_2}^{x_1} e^{-\frac{1}{2}[u^2 + (s-u)^2]} du}{\frac{1}{\sqrt{2}} e^{-\frac{1}{2}\frac{s^2}{2}}}
 \end{aligned}$$

a podmíněná pravděpodobnost tak není funkcí θ .

Výše uvedená definice dostatečné statistiky není v praxi příliš použitelná, protože vyžaduje odvození podmíněné pravděpodobnostní funkce, což nemusí být vždy snadný úkol.

Následující definice je ekvivalentní s předcházející. Toto tvrzení však dokazovat nebudeme.

Definice 7.16 (Dostatečná statistika) Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$. Statistika $S = s(X_1, \dots, X_n)$ je dostatečnou statistikou tehdy a jen tehdy, jestliže podmíněná pravděpodobnost T pro dané S nezávisí na θ pro žádnou statistiku T .

Tato definice je vhodná v situacích, kdy potřebujeme dokázat, že konkrétní statistika není dostatečná. Např. k dokázání toho, že $T' = t'(X_1, \dots, X_n)$ není dostatečná, stačí nalézt jinou statistiku $T = t(X_1, \dots, X_n)$, jejíž podmíněná pravděpodobnost pro dané T' bude záviset na θ .

V některých případech neexistuje dostatečná statistika. Vždy však existují sdružené dostatečné statistiky.

Definice 7.17 (Sdružené dostatečné statistiky) Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$. Statistiky S_1, \dots, S_r jsou sdruženě dostatečné tehdy a jen tehdy, jestliže podmíněná pravděpodobnost X_1, \dots, X_n pro $S_1 = s_1, \dots, S_r = s_r$ nezávisí na θ .

Náhodný výběr X_1, \dots, X_n je sám o sobě vždy sdruženě dostatečný, protože jeho podmíněná pravděpodobnost nezávisí na θ ⁹. Také pořadové statistiky Y_1, \dots, Y_n jsou sdruženě dostatečné. Pro dané $Y_1 = y_1, \dots, Y_n = y_n$ totiž X_1, \dots, X_n musí být některou z permutací y_1, \dots, y_n . Vzhledem k tomu, že těchto permutací je $n!$ a vzhledem k tomu, že se jedná o náhodný výběr, je pravděpodobnost realizace jednotlivých permutací shodná, a proto $P[X_1 = x_1, \dots, X_n = x_n | Y_1 = y_1, \dots, Y_n = y_n] = \frac{1}{n!}$, což je nezávislé na θ .

Připomeňme, že nejdůležitější úlohou dostatečné statistiky je rozdělit prostor \mathcal{X} a množinu segmentů. Hodnoty, kterých přitom nabývá, pravidla nejsou předmětem zájmu. Platnost následující věty je pak patrná.

Definice 7.18 Jestliže statistiky $S_1 = s_1(X_1, \dots, X_n), \dots, S_r = s_r(X_1, \dots, X_n)$ jsou sdruženě dostatečné, pak je libovolný soubor jejich funkcí jedna ku jedné popř. transformací jedna ku jedné je taktéž sdruženě dostatečný.

⁹Podmíněná pravděpodobnost tohoto náhodného výběru je totiž podmíněna sama sebou. Je totiž zřejmé, že $P[X_1 = x_1, \dots, X_n = x_n | X_1 = x_1, \dots, X_n = x_n] = 1$.

Příklad 7.21 Jsou-li $\sum X_i$ a $\sum X_i^2$ sdruženě dostatečné, pak \bar{X} a $\sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}$ jsou také sdruženě dostatečné. Naproti tomu \bar{X}^2 a $\sum (X_i - \bar{X})^2$ nemusí být sdruženě dostatečné, protože se nejedná o funkce jedna ku jedné pro $\sum X_i$ a $\sum X_i^2$.

7.3.2 Kritérium rozkladu

Věta 7.3 (Kritérium rozkladu pro dostatečnou statistiku) Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$. Statistika $S = \mathfrak{s}(X_1, \dots, X_n)$ je dostatečná právě tehdy a jen tehdy, jestliže sdruženou pravděpodobnost X_1, \dots, X_n , která má tvar $\prod_{i=1}^n f(x_i, \theta)$, lze vyjádřit pomocí rozkladu

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n, \theta) &= g(\mathfrak{s}(x_1, \dots, x_n), \theta) h(x_1, \dots, x_n) \\ &= g(s, \theta) h(x_1, \dots, x_n) \end{aligned}$$

kde $h(x_1, \dots, x_n)$ je nezáporná funkce, která neobsahuje θ a $g(\mathfrak{s}(x_1, \dots, x_n), \theta)$ je nezáporná funkce, která závisí na x_1, \dots, x_n pouze skrze funkci $\mathfrak{s}(\cdot, \dots, \cdot)$.

Věta 7.4 (Kritérium rozkladu pro sdruženě dostatečné statistiky)

Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$. Statistiky $S_1 = s_1(X_1, \dots, X_n), \dots, S_r = s_r(X_1, \dots, X_n)$ jsou sdruženě dostatečné právě tehdy a jen tehdy, jestliže sdruženou pravděpodobnost X_1, \dots, X_n lze vyjádřit pomocí rozkladu

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n, \theta) &= g(\mathfrak{s}_1(x_1, \dots, x_n), \dots, \mathfrak{s}_r(x_1, \dots, x_n), \theta) h(x_1, \dots, x_n) \\ &= g(s_1, \dots, s_r, \theta) h(x_1, \dots, x_n) \end{aligned}$$

kde $h(x_1, \dots, x_n)$ je nezáporná funkce, která neobsahuje θ a $g(\mathfrak{s}_1, \dots, \mathfrak{s}_r, \theta)$ je nezáporná funkce, která závisí na x_1, \dots, x_n pouze skrze funkce $\mathfrak{s}_1, \dots, \mathfrak{s}_r$.

Výše uvedené dvě věty získají intuitivní základ, pokud si uvědomíme následující. Jestliže lze provést rozklad sdružené pravděpodobnostní funkce, pak je funkce maximální věrohodnosti proporcionální funkci $g(s, \theta)$, která závisí na pozorováních x_1, \dots, x_n pouze skrze funkci \mathfrak{s} . Funkce $h(x_1, \dots, x_n)$ vystupuje v rámci funkce maximální věrohodnosti jako konstanta. Funkce maximální věrohodnosti je tak pouze funkcí θ . To znamená, že informace ohledně θ , které funkce maximální věrohodnosti “obsahuje”, je obsažena ve statistice $\mathfrak{s}(\cdot, \dots, \cdot)$.

Příklad 7.22 Uvažujme náhodný výběr X_1, \dots, X_n z Bernoulliho rozdělení s parametrem θ , tj.

$$f(x, \theta) = \theta^x (1 - \theta)^{1-x} I_{(0,1)}(x) \quad 0 \leq \theta \leq 1$$

Pak platí

$$\prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} I_{(0,1)}(x_i) = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i} \prod_{i=1}^n I_{(0,1)}(x_i)$$

Jestliže dosadíme $g(\mathfrak{s}(x_1, \dots, x_n)) = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}$, $h(x_1, \dots, x_n) = \prod_{i=1}^n I_{(0,1)}(x_i)$ a $\mathfrak{s}(x_1, \dots, x_n) = \sum x_i$, pak je sdružená pravděpodobnostní funkce X_1, \dots, X_n dána větou (7.3), což znamená, že statistika $S = \mathfrak{s}(X_1, \dots, X_n) = \sum X_i$ je dostatečná.

Příklad 7.23 Uvažujme náhodný výběr z populace $\phi_{\theta,1}(\cdot)$. Sdružená pravděpodobnostní funkce je rovna

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n, \theta) &= \prod_{i=1}^n \phi_{\theta,1}(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - \theta)^2} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} (\sum x_i^2 - 2\theta \sum x_i + n\theta^2)} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{\theta \sum x_i - \frac{n}{2} \theta^2} e^{-\frac{1}{2} \sum x_i^2} \end{aligned}$$

Jestliže dosadíme $h(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-1/2 \sum x_i^2}$, $g(\mathbf{s}(x_1, \dots, x_n), \theta) = e^{\theta \sum x_i - (n/2)\theta^2}$ a $\mathbf{s}(x_1, \dots, x_n) = \sum x_i$, pak je sdružená pravděpodobnostní funkce X_1, \dots, X_n dána větou (7.4), což znamená, že statistika $S = \mathbf{s}(X_1, \dots, X_n) = \sum X_i$ je dostatečná.

Připomeňme, že statistika \bar{X}_n je taktéž dostatečná, protože libovolná funkce jedna ku jedné dostatečné statistiky je také dostatečná statistika.

Příklad 7.24 Uvažujme náhodný výběr X_1, \dots, X_n z populace ϕ_{μ, σ^2} . V tomto případě je θ vektor dvou parametrů, tj. $\theta = (\mu, \sigma)$. Sdružená pravděpodobnost X_1, \dots, X_n je dána

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n, \theta) &= \prod_{i=1}^n \phi_{\mu, \sigma^2}(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma}\right)^2} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \sigma^{-n} e^{-\frac{1}{2} \sum \left(\frac{x_i - \mu}{\sigma}\right)^2} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \sigma^{-n} e^{-\frac{1}{2\sigma^2} (\sum x_i^2 - 2\mu \sum x_i + n\mu^2)} \end{aligned}$$

Sdružená pravděpodobnostní funkce tak závisí na pozorováních x_1, \dots, x_n pouze skrze statistiky $\mathbf{s}_1(x_1, \dots, x_n) = \sum x_i$ a $\mathbf{s}_2(x_1, \dots, x_n) = \sum x_i^2$ a je dána větou (7.5), kde $h(x_1, \dots, x_n) = 1$. Statistiky $\sum X_i$ a $\sum X_i^2$ jsou tak sdruženě dostatečné.

Vzhledem k tomu, že \bar{X}_n a $S_n^2 = \frac{1}{n-1} \sum (X_i - \bar{X}_n)^2$ jsou funkcemi jedna ku jedné sdruženě dostatečných statistik $\sum X_i$ a $\sum X_i^2$, jsou také \bar{X}_n a S_n^2 sdruženě dostatečné.

Příklad 7.25 Uvažujme náhodný výběr X_1, \dots, X_n z uniformního rozdělení nad intervalem $[\theta_1, \theta_2]$, tj. $\theta = (\theta_1, \theta_2)$. Sdružená pravděpodobnostní funkce X_1, \dots, X_n je dána

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n, \theta) &= \prod_{i=1}^n \frac{1}{\theta_2 - \theta_1} I_{[\theta_1, \theta_2]}(x_i) \\ &= \frac{1}{(\theta_2 - \theta_1)^n} \prod_{i=1}^n I_{[\theta_1, \theta_2]}(x_i) = \frac{1}{(\theta_2 - \theta_1)^n} I_{[\theta_1, y_n]}(y_1) I_{[y_1, \theta_2]}(y_n) \end{aligned}$$

kde $y_1 = \min(x_1, \dots, x_n)$ a $y_n = \max(x_1, \dots, x_n)$. Sdružená pravděpodobnost tak závisí na x_1, \dots, x_n pouze skrze y_1 a y_n , a její rozklad je dán větou (7.5), kde $h(x_1, \dots, x_n) = 1$. Statistiky Y_1 a Y_n jsou sdruženě dostatečné.

Všimněme si, že pro $\theta_1 = \theta^*$ a $\theta_2 = \theta^* + 1$, jsou Y_1 a Y_n stále sdruženě dostatečné. Jestliže však $\theta_1 = 0$ a $\theta_2 = \theta^*$, pak se náš rozklad změní na

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n, \theta^*) = \frac{1}{\theta^{*n}} \prod_{i=1}^n I_{[0, \theta^*]}(x_i) = \frac{1}{\theta^{*n}} I_{[0, \theta]}(y_n) I_{[0, y_n]}(y_1)$$

Jestliže dosadíme $g(\mathfrak{s}(x_1, \dots, x_n), \theta) = \frac{1}{\theta^{*n}} I_{[0, \theta]}(y_n)$ a $h(x_1, \dots, x_n) = I_{[0, y_n]}(y_1)$ a $h(x_1, \dots, x_n) = I_{[0, y_n]}(y_1)$, je zřejmé, že statistika Y_n je dostatečná sama o sobě.

Věta 7.5 Funkce (popř. soubor funkcí) maximální věrohodnosti závisí na náhodném výběru skrze libovolný soubor sdruženě dostatečných statistik.

Důkaz 7.5 Jestliže $S_1 = \mathfrak{s}_1(X_1, \dots, X_n), \dots, S_k = \mathfrak{s}_k(X_1, \dots, X_n)$ jsou sdruženě dostatečné statistiky, pak lze funkci maximální věrohodnosti zapsat ve tvaru

$$\begin{aligned} L(\theta, x_1, \dots, x_n) &= \prod_{i=1}^n f(x_i, \theta) \\ &= g(\mathfrak{s}_1(x_1, \dots, x_n), \dots, \mathfrak{s}_k(x_1, \dots, x_n), \theta) h(x_1, \dots, x_n) \end{aligned}$$

Funkce maximální věrohodnosti tedy má maximum ve stejném bodě jako funkce $g(s_1, \dots, s_k)$. Maximum funkce $g(\cdot, \dots, \cdot)$ může být ovlivněno hodnotami s_1, \dots, s_k , které jsou však dány pozorováními x_1, \dots, x_n .

7.3.3 Minimální dostatečná statistika

Hlavní myšlenkou dostatečné statistiky je redukce dat bez ztráty informace o parametrech výběrové populace. Z předchozího textu je zřejmé, že existuje více než jedna dostatečná statistika popř. více než jeden soubor sdruženě dostatečných statistik. V případě normálního rozdělení s neznámou střední hodnotou a rozptylem mohou roli dostatečných sdružených statistik plnit (a) samotný náhodný výběr X_1, \dots, X_n , (b) jeho pořadová statistika Y_1, \dots, Y_n a (c) statistiky \bar{X}_n a S_n^2 . Logicky preferujeme sdruženě dostatečné statistiky \bar{X}_n a S_n^2 , protože míra, s kterou redukují původní data X_1, \dots, X_n , je větší než u prvních dvou příkladů. Přirozenou otázkou je, zda-li existuje nějaký jiný soubor dostatečných statistik, které redukují data více než \bar{X}_n a S_n^2 . Odpověď zní nikoliv, avšak nemáme (a ve zbytku knihy a ani nebudeme mít) aparát, s jehož pomocí bychom toto tvrzení dokázali. Nicméně tyto úvahy nás přivedly k pojmu minimální dostatečná statistika.

Jak jsme již dříve zmínili, každá statistika odpovídá určitému regionu v prostoru \mathfrak{X} . Zjednodušeně řečeno platí, že míra redukce dat dané statistiky je dána počtem podmnožin v tomto regionu. Minimální dostatečná statistika je pak taková statistika, které má v daném regionu menší počet podmnožin než jakákoliv jiná alternativní statistika. Analogické tvrzení platí pro soubor statistik.

Věta 7.6 (Minimální dostatečná statistika) Soubor sdruženě dostatečných statistik je minimálně dostatečný tehdy a jen tehdy, jestliže se jedná o funkci všech ostatních souborů dostatečných statistik.

Výše uvedená definice nám bohužel není příliš nápomocná při hledání sdruženě minimálních dostatečných statistik. Ty je možné nalézt vhodným rozkladem jejich sdružené pravděpodobnostní funkce. Konkrétní postup však přesahuje rámec této knihy.

7.3.4 Rodina exponenciálních rozdělení

Definice 7.19 (Rodina jednoparametrických exponenciálních rozdělení)

Je-li θ skalárem, pak pravděpodobnostní funkce $f(\cdot, \theta)$ spojité náhodné veličiny, kterou lze pro $-\infty < x < \infty$, všechna $\theta \in \underline{\Theta}$ a vhodně zvolené funkce $a(\cdot), b(\cdot), c(\cdot), d(\cdot)$ vyjádřit ve tvaru

$$f(x, \theta) = a(\theta)b(x)e^{c(\theta)d(x)}$$

patří do rodiny exponenciálních rozdělení.

Příklad 7.26 Je-li $f(x, \theta) = \theta e^{-\theta x} I_{(0, \theta)}(x)$, pak $f(x, \theta)$ náleží do rodiny exponenciálních funkcí, protože $a(\theta) = \theta, b(x) = I_{(0, \infty)}(x), c(\theta) = -\theta$ a $d(x) = x$.

Příklad 7.27 Je-li $f(x, \theta) = f(x, \lambda)$ Poissonovo rozdělení, pak

$$f(x, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} I_{\{0, 1, \dots\}}(x) = e^{-\lambda} \left(\frac{1}{x!} I_{\{0, 1, \dots\}}(x) \right) e^{x \ln(\lambda)}$$

Poissonovo rozdělení tak náleží do rodiny exponenciálních rozdělení, protože $a(\lambda) = e^{-\lambda}, b(x) = \frac{1}{x!} I_{\{0, 1, \dots\}}(x), c(\lambda) = \ln(\lambda)$ a $d(x) = x$.

Jestliže $f(x, \theta) = a(\theta)b(x)e^{c(\theta)d(x)}$, pak

$$\prod_{i=1}^n f(x_i, \theta) = a^n(\theta) \left[\prod_{i=1}^n b(x_i) \right] e^{c(\theta) \sum_{i=1}^n d(x_i)}$$

a proto je $\sum d(X_i)$ dostatečnou statistikou. To znamená, že náleží-li pravděpodobnostní rozdělení do rodiny jednoparametrických exponenciálních rozdělení, pak vždy existuje dostatečná statistika. Lze dokázat, že takto získaná dostatečná statistika je zároveň minimální.

Definice 7.20 (Rodina víceparametrických exponenciálních rozdělení)

Pravděpodobnostní rozdělení $f(\cdot, \theta_1, \dots, \theta_k)$ spojitých náhodných veličin, které lze pro vhodně zvolené funkce $a(\cdot, \dots, \cdot), b(\cdot), c_j(\cdot, \dots, \cdot)$ a $d_j(\cdot)$ vyjádřit ve tvaru

$$f(x, \theta_1, \dots, \theta_k) = a(\theta_1, \dots, \theta_k)b(x)e^{\sum_{j=1}^k c_j(\theta_1, \dots, \theta_k)d_j(x)}$$

náleží do rodiny exponenciálních rozdělení.

Příklad 7.28 Je-li $f(x, \theta_1, \theta_2) = \phi_{\mu, \sigma^2}(x)$, kde $(\theta_1, \theta_2) = (\mu, \sigma)$, pak $f(x; \theta_1, \theta_2)$ náleží do rodiny exponenciálních rozdělení.

$$\begin{aligned} f(x; \theta_1, \theta_2) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{\mu^2}{\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x} \end{aligned}$$

protože $a(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-1/2\mu^2/\sigma^2}, b(x) = 1, c_1(\mu, \sigma) = -1/2\sigma^2, c_2(\mu, \sigma) = \mu^2/\sigma^2, d_1(x) = x^2$ a $d_2(x) = x$.

Příklad 7.29 *Jestliže*

$$f(x; \theta_1, \theta_2) = \frac{1}{B(\theta_1, \theta_2)} x^{\theta_1-1} (1-x)^{\theta_2-1} I_{(0,1)}(x)$$

pak

$$f(x; \theta_1, \theta_2) = \frac{1}{B(\theta_1, \theta_2)} I_{(0,1)}(x) e^{(\theta_1-1) \ln(x) + (\theta_2-1) \ln(1-x)}$$

kde $a(\theta_1, \theta_2) = \frac{1}{B(\theta_1, \theta_2)}$, $b(x) = I_{(0,1)}(x)$, $c_1(\theta) = \theta_1 - 1$, $c_2(\theta_1, \theta_2) = \theta_1 - 1$, $c_2(\theta_1, \theta_2) = \theta_2 - 1$, $d_1(x) \ln(x)$ a $d_2(x) = \ln(1-x)$.

Uvažujme náhodný výběr z populace

$$f(x; \theta_1, \dots, \theta_k) = a(\theta_1, \dots, \theta_k) b(x) e^{\sum_{j=1}^k c_j(\theta_1, \dots, \theta_k) d_j(x)}$$

pak platí

$$\prod_{i=1}^n f(x_i; \theta_1, \dots, \theta_k) = a^n(\theta_1, \dots, \theta_k) \left[\prod_{i=1}^n b(x_i) \right] e^{\sum_{j=1}^k c_j(\theta_1, \dots, \theta_k) \sum_{i=1}^n d_j(x_i)}$$

a proto $\sum_{i=1}^n d_1(X_i), \dots, \sum_{i=1}^n d_k(X_i)$ představuje soubor sdruženě dostatečných statistik. Zároveň se jedná o minimální sdruženě dostatečné statistiky.

Příklad 7.30 *Pro výše uvedený příklad náhodného výběru z beta rozdělení jsou statistiky $\sum_{i=1}^n \ln(X_i)$ a $\sum_{i=1}^n \ln(1-X_i)$ minimální sdruženě dostatečné statistiky.*

7.4 Nezkreslený odhad

Připomeňme, že dle tvrzení (7.1) platí

$$E_\theta[(T - \tau(\theta))^2] = D_\theta[T] + (\tau(\theta) - E_\theta[T])^2$$

Jestliže je tedy T nezkreslenou funkcí odhadu pro $\tau(\theta)$, pak $E_\theta[T] = \tau(\theta)$, a proto $E_\theta[(T - \tau(\theta))^2] = D_\theta[T]$.

Definice 7.21 (UMVUE (uniformly minimum variance unbiased estimator))

Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$. Funkce odhadu $T^* = t^*(X_1, \dots, X_n)$ pro $\tau(\theta)$ je UMVUE (uniformly minimum-variance unbiased estimator) pro $\tau(\theta)$ tehdy a jen tehdy, jestliže

1. $E_\theta[T^*] = \tau(\theta)$, tj. T^* je nezkreslenou funkcí odhadu
2. $D_\theta[T^*] \leq D_\theta[T]$ pro libovolnou alternativní funkci odhadu $T = t(X_1, \dots, X_n)$, která splňuje $E_\theta[T] = \tau(\theta)$.

7.4.1 Dolní mez rozptylu

Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$, kde $\theta \in \bar{\Theta}$. Předpokládejme, že $\bar{\Theta}$ je podmnožinou reálné osy a že $T = t(X_1, \dots, X_n)$ je nezkreslená funkce odhadu pro $\tau(\theta)$. V následujícím textu je $f(\cdot, \theta)$ pravděpodobnostní funkcí spojitě náhodné veličiny; v případě nespojitě náhodné veličiny je postup analogický. Předpokládejme, že jsou splněny tzv. podmínky regularity.

1. Pro všechna x a θ existuje

$$\frac{\partial}{\partial \theta} \ln(f(x, \theta))$$

2.

$$\begin{aligned} & \frac{\partial}{\partial \theta} \int \cdots \int \prod_{i=1}^n f(x_i, \theta) dx_1 \cdots dx_n \\ &= \int \cdots \int \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i, \theta) dx_1 \cdots dx_n \end{aligned}$$

3.

$$\frac{\partial}{\partial \theta} \int \cdots \int \mathbf{t}(x_1, \dots, x_n) \prod_{i=1}^n f(x_i, \theta) dx_1 \cdots dx_n = \int \cdots \int \mathbf{t}(x_1, \dots, x_n) \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i, \theta) dx_1 \cdots dx_n$$

4. Pro všechna $\theta \in \underline{\Theta}$ platí

$$0 < E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \ln(f(X, \theta)) \right)^2 \right] < \infty$$

Věta 7.7 (Cramér-Raova nerovnost) Při splnění výše uvedených čtyř podmínek regularity platí

$$D_{\theta}[T] \geq \frac{[\tau'(\theta)]^2}{n E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \ln(f(X, \theta)) \right)^2 \right]}$$

kde $T = \mathbf{t}(X_1, \dots, X_n)$ je nezkreslenou funkcí odhadu pro $\tau(\theta)$. Rovnost ve výše uvedeném vztahu nastane tehdy a jen tehdy, jestliže existuje funkce $K(\theta, n)$ taková, že

$$K(\theta, n)[\mathbf{t}(x_1, \dots, x_n) - \tau(\theta)] = \sum_{i=1}^n \frac{\partial}{\partial \theta} \ln(f(x_i, \theta))$$

Nerovnost uvedená v této větě je tzv. Cramér-Raovou nerovností a její pravá strana se nazývá Cramér-Raovou dolní hranicí rozptylu nezkreslené funkce odhadu pro $\tau(\theta)$.

Důkaz 7.6

$$\begin{aligned}
\tau'(\theta) &= \frac{\partial}{\partial \theta} \tau(\theta) = \frac{\partial}{\partial \theta} \int \cdots \int \mathfrak{t}(x_1, \dots, x_n) \prod_{i=1}^n f(x_i, \theta) dx_1 \cdots dx_n \\
&= \int \cdots \int \mathfrak{t}(x_1, \dots, x_n) \frac{\partial}{\partial \theta} \left[\prod_{i=1}^n f(x_i, \theta) \right] dx_1 \cdots dx_n - \tau(\theta) \frac{\partial}{\partial \theta} \int \cdots \int \prod_{i=1}^n [f(x_i, \theta) dx_i] \\
&= \int \cdots \int \mathfrak{t}(x_1, \dots, x_n) \frac{\partial}{\partial \theta} \left[\prod_{i=1}^n f(x_i, \theta) \right] dx_1 \cdots dx_n - \tau(\theta) \int \cdots \int \frac{\partial}{\partial \theta} \left[\prod_{i=1}^n f(x_i, \theta) \right] dx_1 \cdots dx_n \\
&= \int \cdots \int [\mathfrak{t}(x_1, \dots, x_n) - \tau(\theta)] \frac{\partial}{\partial \theta} \left[\prod_{i=1}^n f(x_i, \theta) \right] dx_1 \cdots dx_n \\
&= \int \cdots \int [\mathfrak{t}(x_1, \dots, x_n) - \tau(\theta)] \left[\frac{\partial}{\partial \theta} \ln \left(\prod_{i=1}^n f(x_i, \theta) \right) \right] \times \prod_{i=1}^n f(x_i, \theta) dx_1 \cdots dx_n \\
&= E_\theta \left[[\mathfrak{t}(X_1, \dots, X_n) - \tau(\theta)] \left[\frac{\partial}{\partial \theta} \ln \left(\prod_{i=1}^n f(x_i, \theta) \right) \right] \right]
\end{aligned}$$

S využitím Cauchy-Schwarzovy nerovnosti platí

$$[\tau'(\theta)]^2 \leq E_\theta[(\mathfrak{t}(X_1, \dots, X_n) - \tau(\theta))^2] E_\theta \left[\left(\frac{\partial}{\partial \theta} \ln \left(\prod_{i=1}^n f(X_i, \theta) \right) \right)^2 \right]$$

neboli

$$D_\theta[T] \geq \frac{[\tau'(\theta)]^2}{E_\theta \left[\left(\frac{\partial}{\partial \theta} \ln \left(\prod_{i=1}^n f(x_i, \theta) \right) \right)^2 \right]}$$

Zároveň však, s využitím nezávislosti X_i a X_j a vztahu

$$\begin{aligned}
E_\theta \left[\frac{\partial}{\partial \theta} \ln(f(X, \theta)) \right] &= \int \left(\frac{\partial}{\partial \theta} \ln(f(x, \theta)) \right) f(x, \theta) dx \\
&= \int \frac{\partial}{\partial \theta} f(x, \theta) dx = \frac{\partial}{\partial \theta} \int f(x, \theta) dx = \frac{\partial}{\partial \theta} (1) = 0
\end{aligned}$$

lze dokázat

$$\begin{aligned}
E_\theta \left[\left(\frac{\partial}{\partial \theta} \ln \left(\prod_{i=1}^n f(X_i, \theta) \right) \right)^2 \right] &= E_\theta \left[\left(\sum_{i=1}^n \frac{\partial}{\partial \theta} \ln(f(X_i, \theta)) \right)^2 \right] \\
&= \sum_i \sum_j E_\theta \left[\left(\frac{\partial}{\partial \theta} \ln(f(X_i, \theta)) \right) \left(\frac{\partial}{\partial \theta} \ln(f(X_j, \theta)) \right) \right] = n E_\theta \left[\left(\frac{\partial}{\partial \theta} \ln(f(X, \theta)) \right)^2 \right]
\end{aligned}$$

Cauchy-Schwarzova nerovnost se stane rovností tehdy a jen tehdy, jestliže je jedna funkce násobkem druhé. To v našem případě znamená, že $\frac{\partial}{\partial \theta} \ln \left(\prod_{i=1}^n f(x_i, \theta) \right)$ je násobkem $\mathfrak{t}(x_1, \dots, x_n) - \tau(\theta)$. Jinými slovy musí existovat taková funkce $K(\theta, n)$, aby platilo

$$\frac{\partial}{\partial \theta} \ln \left(\prod_{i=1}^n f(x_i, \theta) \right) = K(\theta, n) [\mathfrak{t}(x_1, \dots, x_n) - \tau(\theta)]$$

V případě nespojitých náhodných veličin, je třeba modifikovat podmínky regularity, věta o Cramér-Raově nerovnosti však zůstává beze změny.

Z Cramér-Raovy nerovnosti vyplývá, že pokud má daná funkce odhadu rozptyl, který odpovídá dolní hranici rozptylu dle této nerovnosti, pak se jedná o UMVUE. Jestliže existuje $T^* = \mathbf{t}^*(X_1, \dots, X_n)$ taková, že

$$\sum_{i=1}^n \frac{\partial}{\partial \theta} \ln(f(x_i, \theta)) = K(\theta, n)[\mathbf{t}^*(x_1, \dots, x_n) - \tau^*(\theta)]$$

pro vhodně zvolenou funkci $K(\theta, n)$ a $\tau^*(\theta)$, pak je T^* je UMVUE pro $\tau^*(\theta)$.

Příklad 7.31 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \theta e^{-\theta x} I_{(0, \infty)}(x)$. Lze dokázat, že pro uvažovanou pravděpodobnostní funkci jsou splněny podmínky regularity.

Nechť $\tau(\theta) = \theta$. Protože $\tau'(\theta) = 1$ platí

$$D_\theta[T] \geq \frac{1}{n E_\theta \left[\left(\frac{\partial}{\partial \theta} \ln(f(X, \theta)) \right)^2 \right]}$$

Všimněme si, že $\frac{\partial}{\partial \theta} \ln(f(x, \theta)) = \frac{\partial}{\partial \theta} (\ln(\theta) - \theta x) = \frac{1}{\theta} - x$, a proto

$$E_\theta \left[\left(\frac{\partial}{\partial \theta} \ln(f(X, \theta)) \right)^2 \right] = E_\theta \left[\left(\frac{1}{\theta} - X \right)^2 \right] = D[X] = \frac{1}{\theta^2}$$

Cramér-Raova dolní hranice rozptylu nezkreslené funkce odhadu pro θ je tak

$$D_\theta[T] \geq \frac{1}{n(1/\theta^2)} = \frac{\theta^2}{n}$$

Nechť $\tau(\theta) = 1/\theta$. Cramér-Raova dolní hranice rozptylu nezkreslené funkce odhadu je dána

$$D_\theta[T] \geq \frac{[\tau'(\theta)]^2}{n(1/\theta^2)} = \frac{1}{n\theta^2}$$

Levá strana druhé rovnice věty (7.7) je

$$\sum_{i=1}^n \frac{\partial}{\partial \theta} \ln(f(x_i, \theta)) = \sum_{i=1}^n \frac{\partial}{\partial \theta} (\ln(\theta) - \theta x_i) = \sum_{i=1}^n \left(\frac{1}{\theta} - x_i \right) = -n \left(\bar{x}_n - \frac{1}{\theta} \right)$$

Vzhledem k tomu, že $K(\theta, n) = -n$, je \bar{X} UMVUE pro $1/\theta$, protože se její rozptyl shoduje s Cramér-Raovou dolní hranicí rozptylu.

Příklad 7.32 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = f(x, \lambda) = e^{-\lambda} \lambda^x / x!$ pro $x = 0, 1, 2, \dots$. Platí

$$\begin{aligned} \frac{\partial}{\partial \lambda} \ln(f(x, \lambda)) &= \frac{\partial}{\partial \lambda} \ln \left(\frac{e^{-\lambda} \lambda^x}{x!} \right) \\ &= \frac{\partial}{\partial \lambda} (-\lambda + x \ln(\lambda) - \ln(x!)) = -1 + \frac{x}{\lambda} \end{aligned}$$

Proto

$$\begin{aligned} E \left[\left(\frac{\partial}{\partial \lambda} \ln(f(X, \lambda)) \right)^2 \right] &= E \left[\left(\frac{X}{\lambda} - 1 \right)^2 \right] = \frac{1}{\lambda^2} E[(X - \lambda)^2] \\ &= \frac{1}{\lambda^2} D[X] = \frac{1}{\lambda^2} \lambda = \frac{1}{\lambda} \end{aligned}$$

Jmenovatel Cramér-Raovy nerovnosti je tedy n/λ . Jestliže $\tau(\lambda) = e^{-\lambda} = P[X = 0]$, pak je Cramér-Raova dolní hranice rozptylu pro nezkreslenou funkci odhadu dána

$$D[T] \geq \frac{\lambda e^{-2\lambda}}{n}$$

Statistika $T = \frac{1}{n} \sum_{i=1}^n I_{\{0\}}(X_i)$ je nezkresleným odhadem $\tau(\lambda) = e^{-\lambda}$, protože

$$E[T] = \frac{1}{n} \sum_{i=1}^n E[I_{\{0\}}(X_i)] = \frac{1}{n} \sum_{i=1}^n e^{-\lambda} = e^{-\lambda}$$

Jestliže $X_i = 0$, pak $I_{\{0\}}(X_i) = 1$; v ostatních případech je $I_{\{0\}}(X_i)$ rovno nule. Proto $E[I_{\{0\}}(X_i)] = 1 \cdot P[X_i = 0] + 0 \cdot P[X_i \neq 0] = e^{-\lambda}$. T tak představuje relativní podíl pozorování rovných nule. Pro její rozptyl platí

$$D[T] = \frac{1}{n} e^{-\lambda} (1 - e^{-\lambda})$$

Odpovídající Cramér-Raova dolní mez pro rozptyl je $\frac{1}{n} \lambda e^{-2\lambda}$. Všimněme si, že skutečně platí

$$\frac{1}{n} e^{-\lambda} (1 - e^{-\lambda}) \geq \frac{1}{n} \lambda e^{-2\lambda}$$

Dále platí

$$\sum_{i=1}^n \frac{\partial}{\partial \lambda} \ln(f(x, \lambda)) = \sum_{i=1}^n \left(-1 + \frac{x_i}{\lambda} \right) = \frac{n}{\lambda} (\bar{x} - \lambda)$$

a proto je \bar{X} UMVUE pro λ .

Za určitých předpokladů zahrnujících existenci druhých derivací a zaměnitelnosti pořadí pro vybrané integrace a derivace platí

$$E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \ln(f(X, \theta)) \right)^2 \right] = -E_{\theta} \left[\frac{\partial^2}{\partial \theta^2} \ln(f(X, \theta)) \right]$$

Tento vztah může být užitečný v případech, kde je složitější vypočítat střední hodnotu na levé straně rovnosti než střední hodnotu na pravé straně rovnosti.

Věta 7.8 Jestliže je odhad $\hat{\theta} = \hat{\zeta}(x_1, \dots, x_n)$ parametru θ získaný metodou maximální věrohodnosti dán řešením rovnice

$$\frac{\partial}{\partial \theta} \ln(L(\theta, x_1, \dots, x_n)) \equiv \frac{\partial}{\partial \theta} \ln \left(\prod_{i=1}^n f(x_i, \theta) \right) = 0$$

a jestliže $T^* = \mathbf{t}^*(X_1, \dots, X_n)$ je nezkreslenou funkcí odhadu pro $\tau^*(\theta)$ a její rozptyl se shoduje s Cramér-Raovou dolní mezí rozptylu, pak $\mathbf{t}^*(x_1, \dots, x_n) = \tau^*(\hat{\zeta}(x_1, \dots, x_n))$.

Důkaz 7.7 *S využitím druhé rovnice věty (7.7) lze dokázat*

$$0 = \frac{\partial}{\partial \theta} \ln \left(\prod_{i=1}^n f(x_i, \theta) \right) \Big|_{\theta=\hat{\theta}} = K(\theta, n)[t^*(x_1, \dots, x_n) - \tau^*(\theta)] \Big|_{\theta=\hat{\theta}}$$

Výše uvedená věta nám tedy říká, že při splnění zde uvedených podmínek je funkce odhadu získaná metodou maximální věrohodnosti UMVUE.

Následující tvrzení ponecháme bez důkazu. Uvažujme pravděpodobnostní funkci $f(\cdot, \theta)$. Nechť je $T^* = t^*(X_1, \dots, X_n)$ nezkreslenou funkcí odhadu pro $\tau^*(\theta)$, jejíž rozptyl se shoduje s Cramér-Raovou dolní mezí rozptylu. Pak je $f(\cdot, \theta)$ členem rodiny exponenciálních rozdělení. Toto tvrzení platí také naopak. Je-li $f(\cdot, \theta)$ členem rodiny exponenciálních rozdělení, pak existuje nezkreslená funkce odhadu T^* pro $\tau^*(\theta)$, jejíž rozptyl se shoduje s Cramér-Raovou dolní mezí rozptylu. Bohužel nám toto tvrzení také nepřímou říká, že statistiku, jejíž rozptyl se shoduje s Cramér-Raovou dolní mezí rozptylu, jsme schopni nalézt pouze v případě, že pravděpodobnostní funkce populace náleží do rodiny exponenciálních rozdělení. Cramér-Raova dolní mez rozptylu je tak užitečná pro nalezení UMVUE pouze, jestliže náhodný výběr pochází z rodiny exponenciálních rozdělení s jedním parametrem. V takovém případě navíc existuje pouze jedna jediná funkce (kromě lineárních funkcí této funkce) tohoto parametru, pro kterou existuje nezkreslená funkce odhadu a jejíž rozptyl se shoduje s Cramér-Raovou dolní mezí rozptylu. Proto je třeba při identifikaci UMVUE aplikovat jiný postup.

7.4.2 Dostatečnost a kompletnost

Věta 7.9 (Rao-Blackwellova) *Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$. Nechť $S_1 = s_1(X_1, \dots, X_n), \dots, S_k = s_k(X_1, \dots, X_n)$ jsou sdruženě dostatečné statistiky a $T = t(X_1, \dots, X_n)$ představuje nezkreslenou funkci odhadu pro $\tau(\theta)$. Definujme $T' = E[T|S_1, \dots, S_k]$. Pak*

1. T' je statistikou a jedná se o funkci sdruženě dostatečných statistik S_1, \dots, S_k . Proto $T' = t'(S_1, \dots, S_k)$.
2. $E_\theta[T'] = \tau(\theta)$, což znamená, že T' je nezkreslenou funkcí odhadu pro $\tau(\theta)$.
3. $D_\theta[T'] \leq D_\theta[T]$ pro každé θ a $D_\theta[T'] < D_\theta[T]$ pro alespoň některá θ , pokud není T rovno T' s pravděpodobností 1.

Důkaz 7.8

1. S_1, \dots, S_k jsou sdruženě dostatečné statistiky. Pro daná S_1, \dots, S_k je tak statistika T nezávislá na θ . Proto je $T' = E[T|S_1, \dots, S_k]$ taktéž nezávislé na θ a T' lze tedy chápat jako statistiku, která je funkcí S_1, \dots, S_k .
2. *S pomocí věty (4.11) lze dokázat*

$$E_\theta[T'] = E_\theta[E[T|S_1, \dots, S_k]] = E[T] = \tau(\theta)$$

3. *Rozptyl statistiky T lze vyjádřit jako*

$$\begin{aligned} D_\theta[T] &= E_\theta[(T - E_\theta[T'])^2] = E_\theta[(T - T' + T' - E_\theta[T'])^2] \\ &= E_\theta[(T - T')^2] + 2E_\theta[(T - T')(T' - E_\theta[T'])] + D_\theta[T'] \end{aligned}$$

Zároveň však platí

$$E_{\theta}[(T - T')(T' - E_{\theta}[T'])] = E_{\theta}[E[(T - T')(T' - E_{\theta}[T'])|S_1, \dots, S_k]]$$

a

$$\begin{aligned} & E[(T - T')(T' - E_{\theta}[T'])|S_1 = s_1, \dots, S_k = s_k] \\ &= (t'(s_1, \dots, s_k) - E_{\theta}[T']) E[(T - T')|S_1 = s_1, \dots, S_k = s_k] \\ &= (t'(s_1, \dots, s_k) - E_{\theta}[T']) (E[T|S_1 = s_1, \dots, S_k = s_k] - E[T'|S_1 = s_1, \dots, S_k = s_k]) \\ &= (t'(s_1, \dots, s_k) - E_{\theta}[T']) (t'(s_1, \dots, s_k) - t'(s_1, \dots, s_k)) = 0 \end{aligned}$$

a proto

$$D_{\theta}[T] = E_{\theta}[(T - T')^2] + D_{\theta}[T'] \geq D_{\theta}[T']$$

Všimněme si, že $D_{\theta}[T] > D_{\theta}[T']$ pokud se T nerovná T' s pravděpodobností 1.

Příklad 7.33 Uvažujme náhodný výběr X_1, \dots, X_n z Bernoulliho rozdělení $f(x, \theta) = \theta^x(1 - \theta)^{1-\theta}$ pro $x = 0, 1$. X_1 je nezkreslená funkce odhadu pro $\tau(\theta) = \theta$. Použijme X_1 jako $T = t(X_1, \dots, X_n)$. Vzhledem k tomu, že $\sum X_i$ je dostatečná statistika, použijme $S = \sum X_i$ jako dostatečnou statistiku. Dle výše uvedené věty je $T' = E[T|S] = E[X_1|\sum X_i]$ nezkreslenou funkcí odhadu pro θ s rozptylem ne větším než má statistika $T = X_1$. Pokusme se vypočíst $E[T|S]$.

V prvním kroku je třeba vypočíst podmíněnou pravděpodobnost X_1 pro $\sum X_i = s$. Připomeňme, že X_1 může nabývat pouze hodnot 0 a 1.

$$\begin{aligned} P[X_1 = 0 | \sum_{i=1}^n X_i = s] &= \frac{P[X_1 = 0, \sum_{i=1}^n X_i = s]}{P[\sum_{i=1}^n X_i = s]} = \frac{P[X_1 = 0, \sum_{i=2}^n X_i = s]}{P[\sum_{i=1}^n X_i = s]} \\ &= \frac{P[X_1 = 0] \cdot P[\sum_{i=2}^n X_i = s]}{P[\sum_{i=1}^n X_i = s]} = \frac{(1 - \theta) \binom{n-1}{s} \theta^s (1 - \theta)^{n-1-s}}{\binom{n}{s} \theta^s (1 - \theta)^{n-s}} = \frac{n - s}{n} \end{aligned}$$

$$\begin{aligned} P\left[X_1 = 1 | \sum_{i=1}^n X_i = s\right] &= \frac{P[X_1 = 1, \sum_{i=1}^n X_i = s]}{P[\sum_{i=1}^n X_i = s]} = \frac{P[X_1 = 1, \sum_{i=2}^n X_i = s - 1]}{P[\sum_{i=1}^n X_i = s]} \\ &= \frac{P[X_1 = 1] \cdot P[\sum_{i=2}^n X_i = s - 1]}{P[\sum_{i=1}^n X_i = s]} = \frac{\theta \binom{n-1}{s-1} \theta^{s-1} (1 - \theta)^{n-1-s+1}}{\binom{n}{s} \theta^s (1 - \theta)^{n-s}} = \frac{s}{n} \end{aligned}$$

Podmíněná pravděpodobnost X_1 pro dané $\sum X_i = s$ je nezávislá na θ . Definujme

$$T' = \frac{\sum_{i=1}^n X_i}{n}$$

Rozptyl X_1 je $\theta(1 - \theta)$ a rozptyl T' je $\frac{\theta(1-\theta)}{n}$. Pro $n > 2$ je tak rozptyl T' menší než rozptyl $T = X_1$.

Jestliže je funkce odhadu T již pouze funkcí S_1, \dots, S_k , pak bude odvozená statistika T' totožná s T . Ačkoliv S_1, \dots, S_k představuje libovolný soubor sdruženě dostatečných statistik, přirozeně preferujeme minimální sdruženě dostatečné statistiky.

Definice 7.22 (Kompletní rodina rozdělení) Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$ a statistiku $T = t(X_1, \dots, X_n)$. Rodina pravděpodobnostních rozdělení statistiky T je kompletní tehdy a jen tehdy, jestliže $E_\theta[z(T)] \equiv 0$ pro všechna $\theta \in \underline{\Theta}$ implikuje $P_\theta[z(T) = 0] \equiv 1$ pro všechna $\theta \in \underline{\Theta}$, pro která je $z(T)$ statistikou. O statistice T říkáme, že je kompletní tehdy a jen tehdy, jestliže je její rodina pravděpodobnostních rozdělení kompletní.

Příklad 7.34 Uvažujme náhodný výběr X_1, \dots, X_n z Bernoulliho rozdělení.

Statistika $T = X_1 - X_2$ není kompletní, protože $E_\theta[X_1 - X_2] = 0$, ale $X_1 - X_2$ není rovno nule s pravděpodobností 1.

Uvažujme statistiku $T = \sum_{i=1}^n X_i$. Nechť $z(T)$ je libovolná statistika, která je funkcí T a pro kterou platí $E_\theta[z(T)] \equiv 0$ pro všechna $0 \leq \theta \leq 1$. Abychom dokázali, že je T kompletní, musíme dokázat, že $z(T) = 0$ pro všechna $t = 0, 1, \dots, n$. Pro střední hodnotu $z(T)$ platí

$$E_\theta[z(T)] = \sum_{t=0}^n z(t) \binom{n}{t} \theta^t (1-\theta)^{n-t} = (1-\theta)^n \sum_{t=0}^n z(t) \binom{n}{t} \left(\frac{\theta}{1-\theta}\right)^t$$

$E_\theta[z(T)] \equiv 0$ tak pro všechna $0 \leq \theta \leq 1$ implikuje

$$\sum_{t=0}^n z(t) \binom{n}{t} \left(\frac{\theta}{1-\theta}\right)^t \equiv 0$$

neboli

$$\sum_{t=0}^n z(t) \binom{n}{t} \alpha^t \equiv 0$$

pro všechna α , kde $\alpha = \frac{\theta}{1-\theta}$. Aby byl polynom $\sum_{t=0}^n z(t) \binom{n}{t} \alpha^t \equiv 0$ roven nule pro $t = 0, \dots, n$, musí $z(t) \binom{n}{t} = 0$ pro $t = 0, \dots, n$. Zároveň však platí $\binom{n}{t} \neq 0$, a proto musí platit $z(t) = 0$ pro $t = 0, \dots, n$. Statistika $T = \sum X_i$ je tedy kompletní.

Příklad 7.35 Uvažujme náhodný výběr X_1, \dots, X_n z uniformního rozdělení nad intervalem $(0, \theta)$, kde $\underline{\Theta} = \{\theta : \theta > 0\}$. Pokusme se dokázat, že statistika Y_n je kompletní.

Musíme dokázat, že $E_\theta[z(Y_n)] \equiv 0$ pro všechna $\theta > 0$ implikuje $P_\theta[z(Y_n) = 0] \equiv 1$ pro všechna $\theta > 0$. Střední hodnota $z(Y_n)$ je

$$E_\theta[z(Y_n)] = \int z(y) f_{Y_n}(y) dy = \int_0^\theta z(y) \theta^{-n} n y^{n-1} dy$$

a $E_\theta[z(Y_n)] \equiv 0$ pro všechna $\theta > 0$ tehdy a jen tehdy, jestliže

$$\frac{n}{\theta^n} \int_0^\theta z(y) y^{n-1} dy \equiv 0 \quad \text{pro všechna } \theta > 0$$

neboli

$$\int_0^\theta z(y) y^{n-1} dy \equiv 0 \quad \text{pro všechna } \theta > 0$$

Derivací obou stran poslední identity vzhledem k θ pak získáme $z(\theta) \theta^{n-1} \equiv 0$, což implikuje $z \equiv 0$ pro všechna $\theta > 0$.

Následující větu prezentujeme bez důkazu.

Věta 7.10 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$, $\theta \in \bar{\Theta}$, kde $\bar{\Theta}$ představuje interval (a to i nekonečný). Jestliže má pravděpodobnostní funkce tvar

$$f(x, \theta) = a(\theta)b(x)e^{c(\theta)d(x)}$$

tj. je členem jedno parametrické rodiny exponenciálních rozdělení, pak $\sum d(X_i)$ představuje kompletní minimální dostatečnou statistiku.

Věta 7.11 (Lehmann-Schefféova) Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$. Jestliže $S = s(X_1, \dots, X_n)$ je kompletní dostatečná statistika a je-li $T^* = t^*(S)$ nezkreslenou funkcí odhadu pro $\tau(\theta)$, pak T^* je UMVUE pro $\tau(\theta)$.

Důkaz 7.9 Uvažujme nezkreslenou funkci odhadu $T' = t'(S)$ pro $\tau(\theta)$. Pak $E_\theta[T^* - T'] \equiv 0$ pro všechna $\theta \in \bar{\Theta}$ a $T^* - T'$ je funkcí S . Z kompletnosti statistiky S tedy vyplývá $P_\theta[t^*(S) = t'(S)] \equiv 1$ pro všechna $\theta \in \bar{\Theta}$. Proto existuje pouze jedna nezkreslená funkce odhadu pro $\tau(\theta)$, která je funkcí S . Nechť T představuje libovolnou nezkreslenou funkci odhadu pro $\tau(\theta)$. T^* musí být rovno $E[T|S]$, protože $E[T|S]$ je nezkreslenou funkcí odhadu pro $\tau(\theta)$ závisící na S . Na základě věty (7.9) platí $D_\theta[T^*] \leq D_\theta[T]$ pro všechna $\theta \in \bar{\Theta}$. Proto je T^* UMVUE.

Výše uvedená věta tedy říká, že

1. jestliže existuje kompletní dostatečná statistika S a nezkreslená funkce odhadu pro $\tau(\theta)$, pak existuje také UMVUE pro $\tau(\theta)$ a
2. tato UMVUE je jedinečná nezkreslená funkce odhadu pro $\tau(\theta)$, která je funkcí S .

Nezkreslenou funkci odhadu pro $\tau(\theta)$, která je sama funkcí S , lze nalézt několika způsoby. Prvním z nich je kvalifikovaný odhad možného tvaru nezkreslené funkce odhadu. Druhý způsob spočívá v nalezení libovolné nezkreslené funkce odhadu pro $\tau(\theta)$ a následném výpočtu její podmíněné pravděpodobnosti pro dané S . Třetí způsob je založen na vyjádření t^* z rovnice $E_\theta[s^*] \equiv \tau(\theta)$.

Příklad 7.36 Uvažujme náhodný výběr X_1, \dots, X_n z Poissonova rozdělení

$$f(x, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, \dots$$

V příkladu (7.27) jsme ukázali, že $f(x, \lambda)$ patří do rodiny exponenciální pravděpodobnostních rozdělení s $d(x) = x$. S ohledem na větu (7.10) je tak statistika $\sum X_i$ kompletní a dostatečná. Abychom našli UMVUE pro λ , stačí odhadnout funkci této statistiky tak, aby střední hodnota této funkce byla λ . Vzhledem k tomu, že λ představuje střední hodnotu populace, je $\frac{1}{n} \sum X_i$ zřejmou volbou. Statistika $\frac{1}{n} \sum X_i$ je tak UMVUE pro λ .

Pokusme se nyní odhadnout $\tau(\theta) = e^{-\lambda} = P[X_i = 0]$ ¹⁰ pomocí třetího z výše jmenovaných postupů. Nejprve je třeba nalézt libovolnou nezkreslenou funkci odhadu pro $e^{-\lambda}$. Zvolme si např. $I_{\{0\}}(X_1)$. Na základě věty (7.11) je

¹⁰Viz. příklad (7.32).

$E[I_{\{0\}}(X_1)|\sum X_i]$ UMVUE pro $e^{-\lambda}$. Abychom odvodili požadovanou podmíněnou pravděpodobnost, musíme nejprve nalézt podmíněnou pravděpodobnost X_1 pro $\sum X_i$.

$$\begin{aligned} P\left[X_1 = 0 \mid \sum_{i=1}^n X_i = s\right] &= \frac{P[X_1 = 0, \sum_{i=1}^n X_i = s]}{P[\sum_{i=1}^n X_i = s]} = \frac{P[X_1 = 0, \sum_{i=2}^n X_i = s]}{P[\sum_{i=1}^n X_i = s]} \\ &= \frac{P[X_1 = 0]P[\sum_{i=2}^n X_i = s]}{P[\sum_{i=1}^n X_i = s]} = \frac{e^{-\lambda}e^{-(n-1)\lambda}[(n-1)\lambda]^s/s!}{e^{-n\lambda}(n\lambda)^s/s!} = \left(\frac{n-1}{n}\right)^s \quad \text{pro } n > 1 \end{aligned}$$

Proto platí

$$E[I_{\{0\}}(X_1) \mid \sum_{i=1}^n X_i = s] = P[X_1 = 0 \mid \sum_{i=1}^n X_i = s] = \left(\frac{n-1}{n}\right)^s$$

a

$$\left(\frac{n-1}{n}\right)^{\sum_{i=1}^n X_i}$$

tak představuje UMVUE pro $e^{-\lambda}$ pro $n > 1$. Pro $n = 1$ je $I_{\{0\}}(X_1)$ nezkreslenou funkcí odhadu založenou na dostatečné statistice X_1 , a proto je $I_{\{0\}}(X_1)$ UMVUE pro $e^{-\lambda}$.

Příklad 7.37 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \theta e^{-\theta x} I_{(0, \infty)}(x)$. Pokusme se nalézt UMVUE pro každou z následujících funkcí - $\theta, 1/\theta$ a $e^{-K\theta} = P[X > K]$ pro dané K .

Vzhledem k tomu, že $\theta e^{-\theta x} I_{(0, \infty)}(x)$ náleží do rodiny exponenciálních rozdělení¹¹, je statistika $S = \sum_{i=1}^n X_i$ kompletní a dostatečná.

Statistika $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, která je funkcí kompletní dostatečné statistiky $S = \sum_{i=1}^n X_i$, je nezkreslenou funkcí podhadu pro $1/\theta$, a proto je \bar{X}_n na základě věty (7.11) UMVUE pro $1/\theta$.

Abychom našli UMVUE pro θ , zaměřme pozornost na funkci odhadu ve tvaru $\frac{c}{\sum_{i=1}^n X_i}$, kde c představuje konstantu, která může záviset na n .

$$\begin{aligned} E_\theta \left[\frac{c}{\sum X_i} \right] &= c E_\theta \left[\frac{1}{S} \right] = c \int_0^\infty \frac{1}{s} f_S(s) ds = c \int_0^\infty \frac{1}{s} \frac{1}{\Gamma(n)} \theta^n s^{n-1} e^{-\theta s} ds \\ &= \frac{c}{\Gamma(n)} \int_0^\infty \theta^n s^{n-2} e^{-\theta s} ds = \frac{c\theta}{\Gamma(n)} \int_0^\infty u^{n-2} e^{-u} du = \frac{c\theta}{\Gamma(n)} \Gamma(n-1) = \frac{c\theta}{n-1} \end{aligned}$$

pro $n > 1$. Pro $c = n-1$ platí $E_\theta \left[\frac{c}{\sum X_i} \right] = \theta$ a $\frac{n-1}{\sum X_i}$ je tak pro $n > 1$ UMVUE pro θ .

Uvažujme $I_{(K, \infty)}(X_1)$ jako jednoduchou funkci odhadu pro $e^{-K\theta}$. Z vět (7.9) a (7.11) vyplývá, že $E_\theta[I_{(K, \infty)}(X_1)|S]$ je UMVUE pro $e^{-K\theta}$. Platí $E_\theta[I_{(K, \infty)}(X_1)|S = s] = P[I_{(K, \infty)}(X_1) = 1|S = s] = P[X_1 > K|S = s]$. Abychom získali $P[X_1 > K|S = s]$, je třeba nejprve nalézt podmíněné pravdě-

¹¹Viz. příklad (7.26).

podobnostní rozdělení X_1 pro dané $S = s$.

$$\begin{aligned}
 f_{X_1|S=s}(x_1|s)\Delta x_1 &= \frac{f_{X_1,S}(x_1,s)\Delta x_1\Delta s}{f_S(s)\Delta s} \\
 &\approx \frac{P[x_1 < X_1 < x_1 + \Delta x_1, s < \sum_{i=1}^n X_i < s + \Delta s]}{\frac{1}{\Gamma(n)}\theta^n s^{n-1}e^{-\theta s}\Delta s} \\
 &\approx \frac{P[x_1 < X_1 < x_1 + \Delta x_1, s - x_1 < \sum_{i=2}^n X_i < s - x_1 + \Delta s]}{\frac{1}{\Gamma(n)}\theta^n s^{n-1}e^{-\theta s}\Delta s} \\
 &= \frac{P[x_1 < X_1 < x_1 + \Delta x_1]P[s - x_1 < \sum_{i=2}^n X_i < s - x_1 + \Delta s]}{\frac{1}{\Gamma(n)}\theta^n s^{n-1}e^{-\theta s}\Delta s} \\
 &\approx \frac{\theta e^{-\theta x_1} \frac{1}{\Gamma(n-1)}\theta^{n-1}(s-x_1)^{n-2}e^{-\lambda(s-x_1)}\Delta x_1\Delta s}{\frac{1}{\Gamma(n)}\theta^n s^{n-1}e^{-\theta s}\Delta s} \\
 &= \frac{\Gamma(n)}{\Gamma(n-1)} \frac{(s-x_1)^{n-2}}{s^{n-1}} \Delta x_1
 \end{aligned}$$

pro $x_1 < s$ a $n > 1$.

$$\begin{aligned}
 E_\theta[I_{(K,\infty)}(X_1)|S=s] &= P[X_1 > K|S=s] = \int_K^\infty f_{X_1|S=s} = \int_K^\infty f_{X_1|S=s}(x_1|s)dx_1 \\
 &= \int_K^s \frac{\Gamma(n)}{\Gamma(n-1)} \frac{(s-x_1)^{n-2}}{s^{n-1}} dx_1 = \frac{n-1}{s^{n-1}} \int_K^s (s-x_1)^{n-2} dx_1 = \frac{n-1}{s^{n-1}} \int_{s-K}^0 y^{n-2}(-dy) \\
 &= \frac{n-1}{s^{n-1}} \frac{y^{n-1}}{n-1} \Big|_0^{s-K} = \left(\frac{s-K}{s}\right)^{n-1}
 \end{aligned}$$

pro $s > K$ a $n > 1$ a kde jsme využili substituci $y = s - x_1$. Statistika $\left(\frac{\sum X_i - K}{\sum X_i}\right)^{n-1} I_{(K,\infty)}(\sum X_i)$ je tak pro $n > 1$ UMVUE pro $e^{-K\theta}$. Pro kontrolu provedme výpočet střední hodnoty

$$\begin{aligned}
 E_\theta \left[\left(\frac{\sum X_i - K}{\sum X_i} \right)^{n-1} I_{(K,\infty)}(\sum X_i) \right] &= \int_K^\infty \left(\frac{s-K}{s} \right)^{n-1} f_S(s) ds \\
 &= \int_K^\infty \left(\frac{s-K}{s} \right)^{n-1} \frac{1}{\Gamma(n)} \theta^n s^{n-1} e^{-\theta s} ds = \int_0^\infty u^{n-1} \frac{1}{\Gamma(n)} \theta^n e^{-\theta(u+K)} du = e^{-\theta K}
 \end{aligned}$$

kde jsme využili substituce $u = s - K$.

Pro některé funkce daného parametru neexistuje nezkreslená funkce odhadu. Např. pro náhodný výběr o jednom prvku z binomického rozdělení neexistuje nezkreslená funkce odhadu pro $1/\theta$. Předpokládejme, že by taková funkce existovala a označme ji $T = t(X)$. Pak by platilo $E_\theta[T] = \sum_{x=0}^n t(x) \binom{n}{x} \theta^x (1-\theta)^{n-x} \equiv 1/\theta$, neboli polynom n -tého řádu v θ by byl identický s $1/\theta$, což není možné.

V některých případech existuje nezkreslená statistika, avšak neexistuje UMVUE. Pro ilustraci uvádíme následující příklad.

Příklad 7.38 Uvažujme náhodný výběr X_1, \dots, X_n z uniformního rozdělení nad intervalem $(\theta, \theta + 1]$. Předpokládejme, že chceme odhadnout θ .

Např. $\bar{X}_n - \frac{1}{2}$ a $\frac{Y_1+Y_n}{2} - \frac{1}{2}$ představují nezkreslené funkce odhadu θ , avšak neexistuje UMVUE pro θ . Pro fixní $0 \leq p < 1$ uvažujme funkci odhadu $g(X_1 - p) + p$, kde funkce $g(y)$ představuje nejvyšší celé číslo větší než y . Platí

$$E[g(X_1 - p) + p] = \int_{\theta}^{\theta+1} g(x_1 + p) dx_1 + p = \int_{\theta-p}^{\theta+1-p} g(y) dy + p$$

Pro fixní θ a p existuje celé číslo $N = N(\theta, p)$, které splňuje podmínku $\theta - p < N\theta + 1 - p$. Proto

$$E[g(X_1 - p) + p] = \int_{\theta-p}^{\theta+1-p} g(y) dy + p = \int_{\theta-p}^N (N-1) dy + \int_N^{\theta+1-p} N dy + p = \theta$$

Funkce $g(X_1 - p) + p$ je tak nezkreslenou funkcí odhadu pro θ . Navíc, je-li $\theta + 1 - p$ celé číslo, řekněme J , pak $g(x_1 - p) = J - 1$ pro všechna x_1 splňující podmínku $\theta - p < x_1 - p \leq \theta + 1 - p$. Proto $g(x_1 - p) + p = J - 1 + p = \theta + 1 - p - 1 + p = \theta$ pro všechna $\theta < x_1 \leq \theta + 1$. To znamená, že $g(X_1 - p) + p$ odhaduje θ bezchybně, a proto má nulový rozptyl, je-li $\theta + 1 - p$ celé číslo. Nicméně $0 \leq p < 1$ je arbitrární a tak pro libovolné fixní $\theta = \theta_0$ lze nalézt nezkreslenou funkci odhadu pro θ s nulovým rozptylem. Aby tak některá funkce odhadu byla UMVUE pro θ , musela by mít nulový rozptyl pro všechna θ ¹². Je zřejmé, že takováto funkce odhadu neexistuje.

7.5 Invariance

7.5.1 Lokační invariance

Definice 7.23 (Lokační invariance) Funkce odhadu $T = t(X_1, \dots, X_n)$ je lokačně invariantní tehdy a jen tehdy, jestliže $t(x_1 + c, \dots, x_n + c) = t(x_1, \dots, x_n) + c$ pro všechny hodnoty x_1, \dots, x_n a všechna c .

Příklad 7.39 Řada statistik, na které jsme již narazili, jsou lokačně invariantní - např. \bar{X}_n a $\frac{Y_1+Y_n}{2}$. Na druhou stranu existuje velké množství statistik, které lokačně invariantních nejsou - např. S_n^2 nebo $Y_n - Y_1$.

Naše využití lokační invariance bude podobné jako v případě nezkreslenosti statistiky, tj. při hledání vhodné funkce odhadu se omezíme pouze na ty statistiky, které jsou lokačně invariantní, a z nich s pokusíme vybrat tu s nejmenším rozptylem.

Definice 7.24 (Lokační parametr) Necht' $\{f(\cdot, \theta), \theta \in \bar{\Theta}\}$ je rodina pravděpodobnostních rozdělení indexovaných parametrem θ , kde $\bar{\Theta}$ je reprezentováno reálnou osou. Parametr θ je lokačním parametrem tehdy a jen tehdy, jestliže pravděpodobnostní rozdělení $f(x, \theta)$ lze zapsat jako funkci parametru $x - \theta$, tj. $f(x, \theta) = h(x - \theta)$. Ekvivalentně, θ je lokačním parametrem pro pravděpodobnostní rozdělení $f(x, \theta)$ náhodné veličiny X pouze tehdy a jen tehdy, jestliže pravděpodobnostní rozdělení náhodné veličiny $X - \theta$ není funkcí θ .

Všimněme si, že je-li θ lokační parametr pro rodinu pravděpodobnostních rozdělení $\{f(\cdot, \theta), \theta \in \bar{\Theta}\}$, pak je funkce $h(\cdot)$ pravděpodobnostní funkcí definovanou jako $h(\cdot) = f(\cdot, 0)$.

¹²To znamená, že by musela odhadnout všechna θ zcela bez chyby.

Příklad 7.40 Je-li $f(x, \theta) = \phi_{\theta,1}(x)$, pak je θ lokačním parametrem, protože

$$\phi_{\theta,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2} = \phi_{0,1}(x - \theta)$$

Je-li $f(x, \theta) = I_{(\theta-\frac{1}{2}, \theta+\frac{1}{2})}(x)$, pak je θ lokační parametr, protože

$$I_{(\theta-\frac{1}{2}, \theta+\frac{1}{2})}(x) = I_{(-\frac{1}{2}, \frac{1}{2})}(x - \theta)$$

je funkcí $x - \theta$.

Je-li $f(x, \alpha) = \frac{1}{\pi[1+(x-\alpha)^2]}$, pak je α lokačním parametrem, protože $f(x, \alpha)$ je funkcí $x - \alpha$.

Větu, kterou nyní uvedeme bez důkazu, lze použít k nalezení funkce odhadu s nejmenším rozptylem pro lokačně invariantní parametr. Hledaná funkce odhadu je vybrána z lokačně invariantních statistik.

Věta 7.12 (Pitmannova funkce odhadu) Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$, kde θ je lokační parametr a $\underline{\Theta}$ je reálná osa. Funkce odhadu

$$t(X_1, \dots, X_n) = \frac{\int \theta \prod_{i=1}^n f(X_i, \theta) d\theta}{\int \prod_{i=1}^n f(X_i, \theta) d\theta}$$

je funkcí odhadu pro θ s nejmenším rozptylem v rámci všech lokačně invariantních statistik.

Příklad 7.41 Uvažujme náhodný výběr X_1, \dots, X_n z normálního rozdělení se střední hodnotou θ a jednotkovým rozptylem. V předchozím textu jsme ukázali, že θ je lokační parametr. Pokusme se nalézt Pitmannovu funkci odhadu pro θ . V následujícím odvození probíhá současně krácení a ukládání členů jak v čitateli tak ve jmenovateli.

$$\begin{aligned} \frac{\int \theta \prod_{i=1}^n \phi_{\theta,1}(X_i) d\theta}{\int \prod_{i=1}^n \phi_{\theta,1}(X_i) d\theta} &= \frac{\int \theta (1/\sqrt{2\pi})^n e^{-\frac{1}{2} \sum (X_i - \theta)^2} d\theta}{\int (1/\sqrt{2\pi})^n e^{-\frac{1}{2} \sum (X_i - \theta)^2} d\theta} \\ &= \frac{\int \theta e^{-\frac{n}{2}\theta^2 + \theta \sum X_i} d\theta}{\int e^{-\frac{n}{2}\theta^2 + \theta \sum X_i} d\theta} = \frac{\int \theta e^{-\frac{n}{2}(\theta - \bar{X}_n)^2} d\theta}{\int e^{-\frac{n}{2}(\theta - \bar{X}_n)^2} d\theta} \\ &= \frac{\int \theta [1/\sqrt{2\pi}(1/\sqrt{n})] e^{-\frac{1}{2}[(\theta - \bar{X}_n)/(1/\sqrt{n})]^2} d\theta}{\int [1/\sqrt{2\pi}(1/\sqrt{n})] e^{-\frac{1}{2}[(\theta - \bar{X}_n)/(1/\sqrt{n})]^2} d\theta} = \bar{X}_n \end{aligned}$$

Při poslední úpravě jsme využili dvou skutečností. Jmenovatel představuje integrál normálního rozdělení se střední hodnotou \bar{X}_n a rozptylem $1/n$ a tudíž je roven jedné. Podobně čítec představuje střední hodnotu tohoto normálního rozdělení, a proto je roven \bar{X}_n .

Příklad 7.42 Uvažujme náhodný výběr X_1, \dots, X_n z uniformního rozdělení nad intervalem $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$. V předchozím textu jsme ukázali, že θ je lokační parametr. Pitmannova funkce odhadu pro θ je

$$\begin{aligned} \frac{\int \theta \prod_{i=1}^n I_{\theta-\frac{1}{2}, \theta+\frac{1}{2}}(X_i) d\theta}{\int \prod_{i=1}^n I_{\theta-\frac{1}{2}, \theta+\frac{1}{2}}(X_i) d\theta} &= \frac{\int \theta \prod_{i=1}^n I_{X_i-\frac{1}{2}, X_i+\frac{1}{2}}(\theta) d\theta}{\int \prod_{i=1}^n I_{X_i-\frac{1}{2}, X_i+\frac{1}{2}}(\theta) d\theta} \\ &= \frac{\int_{Y_n-\frac{1}{2}}^{Y_1+\frac{1}{2}} \theta d\theta}{\int_{Y_n-\frac{1}{2}}^{Y_1+\frac{1}{2}} d\theta} = \frac{1}{2} \frac{(Y_1 + \frac{1}{2})^2 - (Y_n - \frac{1}{2})^2}{(Y_1 + \frac{1}{2}) - (Y_n - \frac{1}{2})} = \frac{Y_1 + Y_n}{2} \end{aligned}$$

Věta 7.13 Pitmannova funkce odhadu je funkcí dostatečné statistiky.

Důkaz 7.10 Jestliže je $S_1 = s_1(X_1, \dots, X_n), \dots, S_k = s_k(X_1, \dots, X_n)$ soubor dostatečných statistik, pak dle pravidla rozkladu platí $\prod_{i=1}^n f(x_i, \theta) = g(s_1, \dots, s_k, \theta)h(x_1, \dots, x_n)$. Pitmannova funkce odhadu tak může být vyjádřena jako

$$\frac{\int \theta \prod_{i=1}^n f(X_i, \theta) d\theta}{\int \prod_{i=1}^n f(X_i, \theta) d\theta} = \frac{\int \theta g(S_1, \dots, S_k, \theta) h(X_1, \dots, X_n) d\theta}{\int g(S_1, \dots, S_k, \theta) h(X_1, \dots, X_n) d\theta} = \frac{\int \theta g(S_1, \dots, S_k, \theta) d\theta}{\int g(S_1, \dots, S_k, \theta) d\theta}$$

což je funkce dostatečných statistik S_1, \dots, S_k .

7.5.2 Škálovací invariance

Definice 7.25 (Škálovací invariance) Funkce odhadu $T = t(X_1, \dots, X_n)$ je škálově invariantní právě tehdy a jen tehdy, jestliže $t(cx_1, \dots, cx_n) = ct(x_1, \dots, x_n)$ pro všechny hodnoty x_1, \dots, x_n a všechna $c > 0$.

Příklad 7.43 Mezi škálově invariantní funkce odhadu patří např. $\bar{X}_n, \sqrt{S_n^2}, \frac{Y_1 + Y_n}{2}$ a $Y_n - Y_1$.

Definice 7.26 (Škálovací parametr) Uvažujme rodinu pravděpodobnostních funkcí $\{f(\cdot, \theta), \theta > 0\}$ indexovaných reálným parametrem θ . Parametr θ je škálovací právě tehdy a jen tehdy, jestliže je možné pravděpodobnostní funkci $f(x, \theta)$ zapsat ve tvaru $\frac{1}{\theta} h\left(\frac{x}{\theta}\right)$, kde $h(\cdot)$ představuje pravděpodobnostní funkci. Ekvivalentně, θ je škálovací parametr pravděpodobnostní funkce $f(x, \theta)$ náhodné veličiny X právě tehdy a jen tehdy, jestliže je pravděpodobnostní funkce náhodné veličiny $\frac{X}{\theta}$ nezávislá na θ .

Všimněme si, že je-li θ škálovací parametr pro rodinu pravděpodobnostních funkcí $\{f(\cdot, \theta), \theta > 0\}$, pak je pravděpodobnostní funkce $h(\cdot)$ dána vztahem $h(x) = f(x, 1)$.

Příklad 7.44 Je-li $f(x, \lambda) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} I_{(0, \infty)}(x)$, pak je λ škálovací parametr, protože $e^y I_{(0, \infty)}(y)$ je pravděpodobnostní funkcí.

Je-li $f(x, \theta) = \phi_{0, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$, pak je σ škálovací parametr, protože $\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ je pravděpodobnostní funkce.

Je-li $f(x, \theta) = \frac{1}{\theta} I_{(0, \theta)}(x) = \frac{1}{\theta} I_{(0, 1)}\left(\frac{x}{\theta}\right)$, pak je θ škálovací parametr, protože $I_{(0, 1)}(y)$ je pravděpodobnostní funkce.

Je-li $f(x, \theta) = \frac{1}{\theta} I_{(0, 2\theta)}(x) = \frac{1}{\theta} I_{(1, 2)}\left(\frac{x}{\theta}\right)$, pak je θ škálovací parametr, protože $I_{(1, 2)}(y)$ je pravděpodobnostní funkce.

Škálovací parametr vyžaduje odlišnou definici měření chybové odchylky. Namísto obvyklého čtverce rozdílu použijeme ztrátovou funkci $l(t, \theta) = \frac{(t-\theta)^2}{\theta^2} = (t/\theta - 1)^2$. Následující větu uvádíme bez důkazu.

Věta 7.14 (Pitmannova funkce odhadu) Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$, kde $\theta > 0$ je škálovací parametr. Předpokládejme, že $f(x, \theta) = 0$ pro $x \leq 0$. V rámci funkcí odhadu, které jsou škálově invariantní, vykazuje funkce odhadu

$$t(X_1, \dots, X_n) = \frac{\int_0^\infty \frac{1}{\theta^2} \prod_{i=1}^n f(X_i, \theta) d\theta}{\int_0^\infty \frac{1}{\theta^3} \prod_{i=1}^n f(X_i, \theta) d\theta} d\theta$$

nejmenší míru odchylky měřenou ztrátovou funkcí $l(t, \theta) = \frac{(t-\theta)^2}{\theta^2}$. Tato funkce je navíc funkcí dostatečných statistik.

Příklad 7.45 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \frac{1}{\theta} I_{(0, \theta)}(x)$. Pitmannova funkce odhadu škálovacího parametru θ je

$$\begin{aligned} \frac{\int_0^\infty \frac{1}{\theta^2} \prod_{i=1}^n \frac{1}{\theta} I_{(0, \theta)}(X_i) d\theta}{\int_0^\infty \frac{1}{\theta^3} \prod_{i=1}^n \frac{1}{\theta} I_{(0, \theta)}(X_i) d\theta} &= \frac{\int_{Y_n}^\infty \theta^{-n-2} d\theta}{\int_{Y_n}^\infty \theta^{-n-3} d\theta} \\ &= \frac{\frac{1}{(n+2)-1} Y_n^{-(n+2)+1}}{\frac{1}{(n+3)-1} Y_n^{-(n+3)+1}} = \frac{n+2}{n+1} Y_n \end{aligned}$$

Víme, že Y_n je kompletní statistika a že $E[Y_n] = \frac{n}{n+1} \theta$. Z Lehmann-Schefféovy věty tak vyplývá, že $\frac{n+1}{n} Y_n$ je UMVUE pro θ .

Příklad 7.46 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \lambda) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} I_{(0, \infty)}(x)$. Pitmannova funkce odhadu škálovacího parametru θ je

$$\begin{aligned} \frac{\int_0^\infty \frac{1}{\lambda^2} \prod_{i=1}^n f(X_i, \lambda) d\lambda}{\int_0^\infty \frac{1}{\lambda^3} \prod_{i=1}^n f(X_i, \lambda) d\lambda} &= \frac{\int_0^\infty \frac{1}{\lambda^{n+2} e^{-\sum X_i / \lambda}} d\lambda}{\int_0^\infty \frac{1}{\lambda^{n+3} e^{-\sum X_i / \lambda}} d\lambda} \\ &= \frac{\int_0^\infty (\alpha / \sum X_i)^{n+2} e^{-\alpha (\sum X_i / \alpha)} d\alpha}{\int_0^\infty (\alpha / \sum X_i)^{n+3} e^{-\alpha (\sum X_i / \alpha)} d\alpha} = \sum X_i \frac{\int_0^\infty \alpha^n e^{-\alpha} d\alpha}{\int_0^\infty \alpha^{n+1} e^{-\alpha} d\alpha} \\ &= \sum X_i \frac{\Gamma(n+1)}{\Gamma(n+2)} = \frac{\sum X_i}{n+1} \end{aligned}$$

7.6 Bayesovské funkce odhadu

V předchozím textu jsme předpokládali, že náhodný výběr pochází z pravděpodobnostní funkce $f(\cdot, \theta)$, jejíž předpis, s výjimkou θ , známe. O parametru θ jsme uvažovali jako o konstantě. V reálném životě však θ může mít charakter náhodné veličiny, která je charakterizována určitým pravděpodobnostním rozdělením. Tuto náhodnou veličinu označme jako Θ , hodnotu Θ jako θ , kumulativní distribuční funkci Θ jako $G(\cdot) = G_\Theta(\cdot)$ a pravděpodobnostní funkci Θ jako $g(\cdot) = g_\Theta(\cdot)$. Předpokládejme, že $G(\cdot)$ a $g(\cdot)$ neobsahují žádný neznámý parametr.

7.6.1 Posteriorní rozdělení

Až dosud jsme používali $f(x, \theta)$ k označení pravděpodobnostní funkce náhodné veličiny X pro všechna θ z $\underline{\Theta}$. Nyní však potřebuje vyjádřit skutečnost, že θ je hodnota náhodné veličiny Θ . Proto budeme pravděpodobnostní funkci náhodné veličiny X zapisovat jako $f(x|\theta)$ namísto $f(x, \theta)$. To znamená, že $f(x|\theta)$ je podmíněnou pravděpodobností - jedná se o pravděpodobnostní funkci náhodné veličiny X pro dané $\Theta = \theta$.

Uvažujme náhodný výběr X_1, \dots, X_n z pravděpodobnostního rozdělení $f(\cdot|\theta)$, kde θ představuje hodnotu náhodné veličiny Θ . Předpokládejme, že pravděpodobnostní funkce náhodné veličiny Θ je $g_\Theta(\cdot)$, která je nám známa a která neobsahuje žádné neznámé parametry. O $g_\Theta(\cdot)$ hovoříme jako o apriorním rozdělení

náhodné veličiny Θ , protože “shrnuje” všechny informace, které máme k dispozici o Θ , před realizací náhodného výběru. Dále předpokládejme, že chceme odhadnout $\tau(\theta)$. To znamená, že potřebujeme nalézt výraz, který “shrnuje” vše, co víme o θ , po realizaci náhodného výběru. Naším cílem je tedy nalézt posteriorní rozdělení náhodné veličiny Θ pro $X_1 = x_1, \dots, X_n = x_n$.

Definice 7.27 (Apriorní a posteriorní rozdělení) *Pravděpodobnostní funkci $g_\Theta(\cdot)$ nazýváme apriorním rozdělením náhodné veličiny Θ . Podmíněnou pravděpodobnostní funkci náhodné veličiny Θ pro dané $X_1 = x_1, \dots, X_n = x_n$ označovanou jako $f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n)$ nazýváme posteriorním rozdělením náhodné veličiny Θ .*

Připomeňme si, že pro podmíněnou pravděpodobnostní funkci platí

$$f_{Y|X=x}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X|Y=y}(x|y)f_Y(y)}{f_X(x)}$$

Pro náhodný výběr tak platí

$$\begin{aligned} f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n) &= \frac{f_{X_1, \dots, X_n|\Theta=\theta}(x_1, \dots, x_n|\theta)g_\Theta(\theta)}{f_{X_1, \dots, X_n}(x_1, \dots, x_n)} \\ &= \frac{[\prod_{i=1}^n f(x_i|\theta)] g_\Theta(\theta)}{\int [\prod_{i=1}^n f(x_i|\theta)] g_\Theta(\theta) d\theta} \end{aligned}$$

Posteriorní rozdělení nahrazuje funkci maximální věrohodnosti jako výraz, který “shrnuje” všechny dostupné informace. Pokud bychom “načrtli” paralelu s funkcí maximální věrohodnosti, vybrali bychom takové θ , které maximalizuje hodnotu posteriorního rozdělení. Nicméně posteriorní rozdělení je pravděpodobnostní funkcí, a proto je stejně dobře možné odhadnout θ pomocí střední hodnoty popř. mediánu. V následujícím textu tak budeme používat střední hodnotu posteriorního rozdělení jako odhad parametru θ , tj. jako odhad pro $\tau(\theta)$ použijeme $E[\tau(\Theta)|X_1 = x_1, \dots, X_n = x_n]$.

Definice 7.28 (Posteriorní Bayesovská funkce odhadu) *Nechť X_1, \dots, X_n představuje náhodný výběr z $f(x|\theta)$, kde θ je hodnota náhodné veličiny Θ , která sleduje známé pravděpodobnostní rozdělení $g_\Theta(\cdot)$. Posteriorní Bayesovská funkce odhadu pro $\tau(\theta)$ vzhledem k $g_\Theta(\cdot)$ je definována jako*

$$\begin{aligned} E[\tau(\Theta)|X_1 = x_1, \dots, X_n = x_n] &= \int \tau(\theta) f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n) d\theta \\ &= \frac{\int \tau(\theta) [\prod_{i=1}^n f(x_i|\theta)] g_\Theta(\theta) d\theta}{\int [\prod_{i=1}^n f(x_i|\theta)] g_\Theta(\theta) d\theta} \end{aligned}$$

Všimněme si podobnosti mezi posteriorní Bayesovskou funkcí odhadu pro $\tau(\theta) = \theta$ a Pitmannovou funkcí odhadu lokálního parametru.

Příklad 7.47 *Uvažujme náhodný výběr X_1, \dots, X_n z Bernoulliho rozdělení $f(x|\theta) = \theta^x(1-\theta)^{1-x}$ pro $x = 0, 1$. Předpokládejme, že apriorní rozdělení náhodné veličiny Θ je dáno $g_\Theta(\theta) = I_{(0,1)}(\theta)$, tj. Θ sleduje uniformní rozdělení nad intervalem $(0, 1)$. Pokusme se odhadnout θ a $\tau(\theta) = \theta(1-\theta)$.*

Posteriorní pravděpodobnostní funkce je

$$f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n) = \frac{\theta^{\sum x_i} (1-\theta)^{n-\sum x_i} I_{(0,1)}(\theta)}{\int_0^1 \theta^{\sum x_i} (1-\theta)^{n-\sum x_i} d\theta}$$

a posteriorní Bayesovská funkce odhadu pro θ je tak rovna

$$\begin{aligned} E[\Theta|X_1 = x_1, \dots, X_n = x_n] &= \int \theta f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n) d\theta \\ &= \frac{\int_0^1 \theta \theta^{\sum x_i} (1-\theta)^{n-\sum x_i} d\theta}{\int_0^1 \theta^{\sum x_i} (1-\theta)^{n-\sum x_i} d\theta} = \frac{B(\sum_{i=1}^n x_i + 2, n - \sum_{i=1}^n x_i + 1)}{B(\sum_{i=1}^n x_i + 1, n - \sum_{i=1}^n x_i + 1)} \\ &= \frac{\Gamma(\sum_{i=1}^n x_i + 2) \Gamma(n - \sum_{i=1}^n x_i + 1)}{\Gamma(n + 3)} \frac{\Gamma(n + 2)}{\Gamma(\sum_{i=1}^n x_i + 1) \Gamma(n - \sum_{i=1}^n x_i + 1)} \\ &= \frac{\sum_{i=1}^n x_i + 1}{n + 2} \end{aligned}$$

Posteriorní Bayesovská funkce odhadu pro θ je tedy $\frac{\sum X_i + 1}{n + 2}$. Funkce odhadu pro θ založená na metodě maximální věrohodnosti je naproti tomu rovna $\frac{\sum X_i}{n}$. Funkce odhadu $\frac{\sum X_i}{n}$ je nezkreslená a UMVUE, proto nemůže být posteriorní Bayesovská funkce odhadu taktéž nezkreslená.

Posteriorní Bayesovská funkce odhadu pro $\theta(1-\theta)$ je rovna

$$\begin{aligned} E[\tau(\Theta)|X_1 = x_1, \dots, X_n = x_n] &= \int \theta(1-\theta) f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n) d\theta \\ &= \frac{\int_0^1 \theta(1-\theta) \theta^{\sum x_i} (1-\theta)^{n-\sum x_i} d\theta}{\int_0^1 \theta^{\sum x_i} (1-\theta)^{n-\sum x_i} d\theta} \\ &= \frac{\Gamma(\sum_{i=1}^n x_i + 2) \Gamma(n - \sum_{i=1}^n x_i + 2)}{\Gamma(n + 4)} \frac{\Gamma(n + 2)}{\Gamma(\sum_{i=1}^n x_i + 1) \Gamma(n - \sum_{i=1}^n x_i + 1)} \\ &= \frac{(\sum_{i=1}^n x_i + 1)(n - \sum_{i=1}^n x_i + 1)}{(n + 3)(n + 2)} \end{aligned}$$

Posteriorní Bayesovská funkce odhadu pro $\theta(1-\theta)$ je tak vzhledem k uniformnímu apriornímu pravděpodobnostnímu rozdělení rovna $\frac{(\sum X_i + 1)(n - \sum X_i + 1)}{(n + 3)(n + 2)}$.

Věta 7.15 Nechť $T_G^* = \mathfrak{t}_G^*(X_1, \dots, X_n)$ označuje posteriorní Bayesovskou funkci odhadu pro $\tau(\theta)$ s ohledem na apriorní kumulativní distribuční funkci $G(\cdot)$. Jestliže T_G^* a $\tau(\Theta)$ mají konečný rozptyl, pak buď $D[T_G^*|\theta] = 0$ nebo T_G^* není nezkreslenou funkcí odhadu pro $\tau(\theta)$. To znamená, že buď T_G^* odhaduje $\tau(\theta)$ správně s pravděpodobností jedna nebo T_G^* není nezkreslenou funkcí odhadu pro $\tau(\theta)$.

Důkaz 7.11 Předpokládejme, že T_G^* je nezkreslenou funkcí odhadu pro $\tau(\theta)$, tj. $E[T_G^*|\theta] = \tau(\theta)$. Z definice vyplývá $T_G^* = \mathfrak{t}_G^*(X_1, \dots, X_n) = E[\tau(\Theta)|X_1, \dots, X_n]$. Pro rozptyl náhodné veličiny T_G^* platí

$$D[T_G^*] = E[D[T_G^*|\Theta]] + D[E[T_G^*|\Theta]] = E[D[T_G^*|\Theta]] + D[\tau(\Theta)]$$

Pro $\tau(\Theta)$ analogicky platí

$$\begin{aligned} D[\tau(\Theta)] &= E[D[\tau(\Theta)|X_1, \dots, X_n]] + D[E[\tau(\Theta)|X_1, \dots, X_n]] \\ &= E[D[\tau(\Theta)|X_1, \dots, X_n]] + D[T_G^*] \end{aligned}$$

Proto

$$\begin{aligned} D[T_G^*] &= E[D[T_G^*|\theta]] + D[\tau(\theta)] \\ D[T_G^*] &= E[D[T_G^*|\theta]] + E[D[\tau(\theta)|X_1, \dots, X_n]] + D[T_G^*] \\ E[D[T_G^*|\theta]] + E[D[\tau(\theta)|X_1, \dots, X_n]] &= 0 \end{aligned}$$

Vzhledem k tomu, že jak $E[D[T_G^*|\theta]]$ tak $E[D[\tau(\theta)|X_1, \dots, X_n]]$ jsou nezáporné a jejich součet je roven nule, musí být oba tyto výrazy rovny nule. Protože $E[D[T_G^*|\theta]] = 0$ a $D[T_G^*|\theta]$ je nezáporné, musí platit $D[T_G^*|\theta] = 0$.

7.6.2 Ztrátová funkce

Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x|\theta)$, kde $\theta \in \bar{\Theta}$ a kde známe funkční předpis $f(\cdot|\theta)$ s výjimkou θ . Předpokládejme, že θ je hodnota náhodné veličiny Θ , jejíž pravděpodobnostní rozdělení známe a to včetně parametrů. Naším cílem je, na základě náhodného výběru X_1, \dots, X_n , odhadnout $\tau(\theta)$. K tomuto účelu specifikujme ztrátovou funkci $l(t, \theta)$, která představuje ztrátu informace z titulu odhadu $\tau(\theta)$. Pro funkci odhadu $T = t(X_1, \dots, X_n)$ představuje $E_\theta[l(T, \theta)]$ průměrnou ztrátu informace, neboli tzv. riziko. Je zřejmé, že při porovnávání dvou funkcí odhadu budeme upřednostňovat tu s menším “rizikem”. Jistou komplikací však představuje skutečnost, že se dvě rizikové funkce, jakožto funkce parametru θ , mohou křížit - pro určitý interval hodnot parametru θ tak může být méně “riziková” jedna funkce odhadu, zatímco pro jiný interval může být méně “riziková” funkce druhá. Protože je hodnota θ neznámá, je obtížné zvolit vhodnější funkci odhadu. Tímto se dostáváme k následující definici.

Definice 7.29 (Bayesovské riziko) Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x|\theta)$, kde θ je hodnota náhodné veličiny Θ s kumulativní distribuční funkcí $G(\cdot) = G_\Theta(\cdot)$ a pravděpodobnostní funkcí $g(\cdot) = g_\Theta(\cdot)$. Při odhadu $\tau(\theta)$ použijme ztrátovou funkci $l(t, \theta)$. Riziko funkce odhadu $T = t(X_1, \dots, X_n)$ označujeme jako $r_t(\theta)$. Bayesovské riziko funkce odhadu $T = t(X_1, \dots, X_n)$ vzhledem ke ztrátové funkci $l(\cdot, \cdot)$ a apriorní kumulativní distribuční funkci $G(\cdot)$ je definován jako

$$r(t) = r_{l,G}(t) = \int_{\bar{\Theta}} r_t(\theta) g(\theta) d\theta$$

Bayesovské riziko funkce odhadu tak představuje průměrné riziko této funkce nad $\bar{\Theta}$ s ohledem na apriorní pravděpodobnostní funkci $g(\cdot)$. Dvě funkce odhadu je tak možné porovnat právě na základě Bayesovského rizika.

Definice 7.30 (Bayesovská funkce odhadu) Bayesovská funkce odhadu pro $\tau(\theta)$ označovaná jako $T_{l,G}^* = t_{l,G}^*(X_1, \dots, X_n)$ vzhledem ke ztrátové funkci $l(\cdot, \cdot)$ a apriorní kumulativní distribuci $G(\cdot)$ je definována jako funkce odhadu s nejmenším Bayesovským rizikem. To znamená, že Bayesovská funkce odhadu $T_{l,G}^* = t_{l,G}^*(X_1, \dots, X_n)$ pro $\tau(\theta)$ splňuje podmínku

$$r_{l,G}(t^*) = r_{l,G}(t_{l,G}^*) \leq r_{l,G}(t)$$

pro libovolnou alternativní funkci odhadu $T = t(X_1, \dots, X_n)$.

Posteriorní Bayesovská funkce odhadu v definici (7.28) nevyžadovala, na rozdíl od výše uvedené definice, specifikaci ztrátové funkce.

Nalezení Bayesovské funkce odhadu pro obecnou ztrátou funkci nemusí být snadný úkol. Vyjímkou je ztrátová funkce založená na čtverci rozdílů. Řekněme, že hledáme funkci odhadu $t^*(\cdot, \dots, \cdot)$, která minimalizuje výraz

$$\int_{\underline{\Theta}} \mathfrak{r}_t(\theta) g(\theta) d\theta = \int_{\underline{\Theta}} E_{\theta} [[t(X_1, \dots, X_n) - \tau(\theta)]^2] g(\theta) d\theta$$

nad možnými funkcemi odhadu $t(\cdot, \dots, \cdot)$.

$$\begin{aligned} & \int_{\underline{\Theta}} E_{\theta} [[t(X_1, \dots, X_n) - \tau(\theta)]^2] g(\theta) d\theta \\ &= \int_{\underline{\Theta}} \left\{ \int_{\mathfrak{X}} [t(x_1, \dots, x_n) - \tau(\theta)]^2 f_{X_1, \dots, X_n}(x_1, \dots, x_n | \theta) \prod_{i=1}^n dx_i \right\} g(\theta) d\theta \\ &= \int_{\mathfrak{X}} \left\{ \int_{\underline{\Theta}} [\tau(\theta) - t(x_1, \dots, x_n)]^2 \frac{f_{X_1, \dots, X_n}(x_1, \dots, x_n | \theta) g(\theta) d\theta}{f_{X_1, \dots, X_n}(x_1, \dots, x_n)} \right\} \\ & \quad \times f_{X_1, \dots, X_n}(x_1, \dots, x_n) \prod_{i=1}^n dx_i \\ &= \int_{\mathfrak{X}} \left\{ \int_{\underline{\Theta}} [\tau(\theta) - t(x_1, \dots, x_n)]^2 f_{\Theta | X_1=x_1, \dots, X_n=x_n}(\theta | x_1, \dots, x_n) d\theta \right\} \\ & \quad \times f_{X_1, \dots, X_n}(x_1, \dots, x_n) \prod_{i=1}^n dx_i \end{aligned}$$

Protože je člen, který je předmětem integrace, nezáporný, lze výše uvedený dvojitý integrál minimalizovat tím, že minimalizujeme výraz ve složených závorkách pro všechna x_1, \dots, x_n . Tento výraz však zároveň představuje podmíněnou střední hodnotu pro $[\tau(\Theta) - t(x_1, \dots, x_n)]^2$ s ohledem na posteriorní rozdělení náhodné veličiny Θ pro dané $X_1 = x_1, \dots, X_n = x_n$. Je tedy zřejmé, že řešením je funkce $t^*(x_1, \dots, x_n)$ rovná podmíněné střední hodnotě $\tau(\Theta)$ vzhledem k posteriornímu rozdělení náhodné veličiny Θ a pro dané $X_1 = x_1, \dots, X_n = x_n$ ¹³. Proto je Bayesovská funkce odhadu pro $\tau(\theta)$ a ztrátovou funkci založenou na čtverci rozdílů dána vztahem

$$E[\tau(\Theta) | X_1 = x_1, \dots, X_n = x_n] = \frac{\int \tau(\theta) [\prod_{i=1}^n f(x_i | \theta)] g(\theta) d\theta}{\int [\prod_{i=1}^n f(x_i | \theta)] g(\theta) d\theta}$$

což je identické s funkcí odhadu dle definice (7.28).

V případě obecné ztrátové funkce, hledáme funkci odhadu, která minimali-

¹³Připomeňme, že funkce $E[(Z - x)^2]$ je minimalizována pro x rovné $E[Z]$.

zuje výraz

$$\begin{aligned}
 & \int_{\underline{\Theta}} \mathfrak{r}_{\mathfrak{t}}(\theta) g(\theta) d\theta \\
 &= \int_{\underline{\Theta}} \left\{ \int_{\mathfrak{X}} \mathfrak{l}(\mathfrak{t}(x_1, \dots, x_n), \theta) f_{X_1, \dots, X_n}(x_1, \dots, x_n | \theta) \prod_{i=1}^n dx_i \right\} g(\theta) d\theta \\
 &= \int_{\mathfrak{X}} \left\{ \int_{\underline{\Theta}} \mathfrak{l}(\mathfrak{t}(x_1, \dots, x_n), \theta) f_{\Theta | X_1=x_1, \dots, X_n=x_n}(\theta | x_1, \dots, x_n) d\theta \right\} \\
 & \quad \times f_{X_1, \dots, X_n}(x_1, \dots, x_n) \prod_{i=1}^n dx_i
 \end{aligned}$$

Minimalizace výše uvedeného dvojitého integrálu je ekvivalentní minimalizaci výrazu ve dvojitých závorkách. Tento výraz někdy nazýváme posteriorním rizikem. Obecně tedy platí, že Bayesovská funkce odhadu pro $\tau(\theta)$ vzhledem ke ztrátové funkci $\mathfrak{l}(\cdot, \cdot)$ a apriorní pravděpodobnostní funkce $g(\cdot)$ je funkce odhadu, která minimalizuje posteriorní riziko, což je očekávaná ztráta s ohledem na posteriorní rozdělení náhodné veličiny Θ a pozorování x_1, \dots, x_n . Shrňme výše uvedený text do následující věty a tvrzení.

Věta 7.16 *Uvažujme náhodný výběr z populace $f(x|\theta)$. Nechť $g(\theta)$ je pravděpodobnostní funkci náhodné veličiny Θ . Dále nechť $\mathfrak{l}(\cdot, \cdot)$ představuje ztrátovou funkci pro odhad $\tau(\theta)$. Bayesovská funkce odhadu pro $\tau(\theta)$ je funkce odhadu $\mathfrak{t}^*(\cdot, \dots, \cdot)$, která minimalizuje výraz*

$$\int_{\underline{\Theta}} \mathfrak{l}(\mathfrak{t}(x_1, \dots, x_n), \theta) f_{\Theta | X_1=x_1, \dots, X_n=x_n}(\theta | x_1, \dots, x_n) d\theta$$

pro funkci $\mathfrak{t}(\cdot, \dots, \cdot)$.

Tvrzení 7.2 *Při splnění předpokladů výše uvedené věty je Bayesovská funkce odhadu pro $\tau(\theta)$ dána*

$$E[\tau(\Theta) | X_1 = x_1, \dots, X_n = x_n] = \frac{\int \tau(\theta) [\prod_{i=1}^n f(x_i|\theta)] g(\theta) d\theta}{\int [\prod_{i=1}^n f(x_i|\theta)] g(\theta) d\theta}$$

pro ztrátovou funkci ve smyslu čtverce rozdílů.

Tvrzení 7.3 *Při splnění předpokladů výše uvedené věty je Bayesovská funkce odhadu pro θ dána mediánem posteriorního rozdělení náhodné veličiny Θ pro ztrátovou funkci ve smyslu absolutní hodnoty rozdílů.*

Důkaz 7.12 *Tvrzení (7.2) vyplývá ze skutečnosti, že*

$$\int_{\underline{\Theta}} |\theta - \mathfrak{t}(x_1, \dots, x_n)| f_{\Theta | X_1=x_1, \dots, X_n=x_n}(\theta | x_1, \dots, x_n) d\theta$$

je minimalizováno pro funkci $\mathfrak{t}^*(\cdot, \dots, \cdot)$, která rovna mediánu posteriorního rozdělení náhodné veličiny Θ ¹⁴.

¹⁴Připomeňme, že funkce $E[|Z - x|]$ je minimalizována pro x rovno mediánu Z .

Příklad 7.48 Uvažujme náhodný výběr X_1, \dots, X_n z normálního rozdělení se střední hodnotou θ a jednotkovým rozptylem. Pokusme se odhadnout θ s pomocí ztrátové funkce ve smyslu čtverce rozdílů za předpokladu, že Θ má normální rozdělení se střední hodnotou μ_0 a jednotkovým rozptylem.

Z předchozího textu víme, že Bayesovská funkce odhadu je dána střední hodnotou posteriorního rozdělení náhodné veličiny Θ . V následujícím odvození použijeme $\mu_0 = x_0$ všude tam, kde je to žádoucí.

$$\begin{aligned}
 & f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n) \\
 &= \frac{f_{X_1, \dots, X_n|\theta}(x_1, \dots, x_n|\theta)g(\theta)}{f_{X_1, \dots, X_n}(x_1, \dots, x_n)} = \frac{[\prod_{i=1}^n f(x_i|\theta)]g(\theta)}{\int [\prod_{i=1}^n f(x_i|\theta)]g(\theta)d\theta} \\
 &= \frac{\frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta - \mu_0)^2}}{\int_{-\infty}^{\infty} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta - \mu_0)^2} d\theta} \\
 &= \frac{e^{-\frac{1}{2} \sum_{i=0}^n (x_i - \theta)^2}}{\int_{-\infty}^{\infty} e^{-\frac{1}{2} \sum_{i=0}^n (x_i - \theta)^2} d\theta} \\
 &= \frac{e^{-\frac{1}{2}((n+1)\theta^2 - 2\theta \sum_{i=0}^n x_i + \sum_{i=0}^n x_i^2)}}{\int_{-\infty}^{\infty} e^{-\frac{1}{2}((n+1)\theta^2 - 2\theta \sum_{i=0}^n x_i + \sum_{i=0}^n x_i^2)} d\theta} \\
 &= \frac{e^{-\frac{n+1}{2}[\theta^2 - 2\theta \sum_{i=0}^n \frac{x_i}{n+1} + (\sum_{i=0}^n \frac{x_i}{n+1})^2]}}{\int_{-\infty}^{\infty} e^{-\frac{n+1}{2}[\theta^2 - 2\theta \sum_{i=0}^n \frac{x_i}{n+1} + (\sum_{i=0}^n \frac{x_i}{n+1})^2]} d\theta} \\
 &= \frac{\frac{1}{\sqrt{2\pi/(n+1)}} e^{-\frac{n+1}{2}(\theta - \sum_{i=0}^n \frac{x_i}{n+1})^2}}{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi/(n+1)}} e^{-\frac{n+1}{2}(\theta - \sum_{i=0}^n \frac{x_i}{n+1})^2} d\theta} \\
 &= \frac{1}{\sqrt{2\pi/(n+1)}} e^{-\frac{n+1}{2}(\theta - \sum_{i=0}^n \frac{x_i}{n+1})}
 \end{aligned}$$

Poslední úprava využila skutečnosti, že integrál ve jmenovateli je integrálem pravděpodobnostní funkce, a proto je roven jedné. Podařilo se nám tedy dokázat, že posteriorní pravděpodobnostní rozdělení náhodné veličiny Θ je normální rozdělení se střední hodnotou $\sum_{i=1}^n \frac{x_i}{n+1}$ a rozptylem $\frac{1}{n+1}$. Proto je Bayesovská funkce odhadu s ohledem na ztrátovou funkci ve smyslu čtverce rozdílů rovna

$$\frac{x_0 + \sum_{i=1}^n X_i}{n+1} = \frac{\mu_0 + \sum_{i=1}^n X_i}{n+1}$$

Protože je posteriorní rozdělení náhodné veličiny Θ normální, jsou střední hodnota a medián shodné. Proto je $\frac{\mu_0 + \sum_{i=1}^n X_i}{n+1}$ také Baysovskou funkcí odhadu s ohledem na ztrátovou funkci ve smyslu absolutní hodnoty rozdílů.

Příklad 7.49 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x|\theta) = \frac{1}{\theta} I_{(0,\theta)}(x)$. Předpokládejme, že Θ sleduje pravděpodobnostní funkci $g(\theta) = I_{(0,1)}(\theta)$. Odhadněme θ pro ztrátovou funkci $l(t, \theta) = \frac{(t-\theta)^2}{\theta^2}$.

Nechť $y_n = \max(x_1, \dots, x_n)$. Posteriorní pravděpodobnostní funkce náhodné

veličiny Θ je

$$\begin{aligned} f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n) &= \frac{\frac{1}{\theta^n} \prod_{i=1}^n I_{(0,\theta)}(x_i) I_{(0,1)}(\theta)}{\int_0^1 \frac{1}{\theta^n} \prod_{i=1}^n I_{(0,\theta)}(x_i) d\theta} \\ &= \frac{\frac{1}{\theta^n} I_{(y_n,1)}(\theta)}{\int_{y_n}^1 \frac{1}{\theta^n} I_{(y_n,1)}(\theta) d\theta} = \frac{\frac{1}{\theta^n} I_{(y_n,1)}(\theta)}{\frac{1}{n-1} \left(\frac{1}{y_n^{n-1}} - 1 \right)} \end{aligned}$$

Hledáme funkci odhadu $t(\cdot)$, která minimalizuje výraz

$$\frac{\int \frac{(t(y_n) - \theta)^2}{\theta^2} \frac{1}{\theta^n} I_{(y_n,1)}(\theta) d\theta}{\frac{1}{n-1} \left(\frac{1}{y_n^{n-1}} - 1 \right)}$$

neboli funkci odhadu minimalizující

$$\begin{aligned} &\int_{y_n}^1 [t(y_n) - \theta]^2 \frac{1}{\theta^{n+2}} d\theta \\ &= [t(y_n)]^2 \int_{y_n}^1 \frac{1}{\theta^{n+2}} d\theta - 2t(y_n) \int_{y_n}^1 \frac{1}{\theta^{n+1}} d\theta + \int_{y_n}^1 \frac{1}{\theta^n} d\theta \end{aligned}$$

Výše uvedená rovnice je kvadratická v $t(\cdot)$ a nabývá minima pro

$$t^*(y_n) = \frac{\int_{y_n}^1 \frac{1}{\theta^{n+1}} d\theta}{\int_{y_n}^1 \frac{1}{\theta^{n+2}} d\theta} = \frac{-\frac{1}{n} \left(1 - \frac{1}{y_n^n} \right)}{-\frac{1}{n+1} \left(1 - \frac{1}{y_n^{n+1}} \right)} = \frac{n+1}{n} y_n \frac{1 - y_n^n}{1 - y_n^{n+1}}$$

Lze dokázat, že Bayesovská funkce odhadu je vždy funkcí minimální dostatečné statistiky. Při splnění určitých obecných podmínek lze také dokázat, že Bayesovská funkce odhadu pro libovolné apriorní rozdělení, které je kladné pro všechna $\theta \in \underline{\Theta}$, je konsistentní a BAN. Jestliže tedy známe správné apriorní pravděpodobnostní rozdělení a přijmeme kritérium, že nejlepší funkce odhadu je ta s nejmenší průměrnou ztrátou, pak je Bayesovská funkce odhadu optimální.

7.6.3 Minimax funkce odhadu

V předchozím textu jsme definovali funkci odhadu, jejíž maximální riziko je menší nebo rovno maximálnímu riziku libovolné alternativní funkce odhadu.

Věta 7.17 *Jestliže $T^* = t^*(X_1, \dots, X_n)$ je Bayesovská funkce odhadu s konstantním rizikem, tj. $r_{t^*}(\theta) \equiv \text{konstanta}$, pak je T^* minimax funkcí odhadu.*

Důkaz 7.13 *Uvažujme apriorní funkci $g^*(\cdot)$ odpovídající Bayesovské funkci odhadu $t^*(\cdot, \dots, \cdot)$.*

$$\begin{aligned} \sup_{\theta \in \underline{\Theta}} r_{t^*}(\theta) &= \text{konstanta} = r_{t^*}(\theta) \\ &= \int_{\underline{\Theta}} r_{t^*}(\theta) g^*(\theta) d\theta \leq \int_{\underline{\Theta}} d\theta \leq \sup_{\theta \in \underline{\Theta}} r_t(\theta) \end{aligned}$$

pro libovolnou alternativní funkci odhadu $t(\cdot, \dots, \cdot)$.

Příklad 7.50 Uvažujme náhodný výběr z Bernoulliho rozdělení. Pokusme se nalézt minimax funkce odhadu pro θ vzhledem k ztrátové funkci založené na čtverci rozdílu.

Potřebujeme nalézt Bayesovskou funkci odhadu s konstantním rizikem. Jako apriorní pravděpodobnostní funkci můžeme použít beta rozdělení. Odpovídající Bayesovská funkce odhadu je pak

$$\begin{aligned} & \frac{\int_0^1 \theta \theta^{\sum x_i} (1-\theta)^{n-\sum x_i} \frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1} d\theta}{\int_0^1 \theta^{\sum x_i} (1-\theta)^{n-\sum x_i} \frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1} d\theta} \\ &= \frac{\int_0^1 \theta^{\sum x_i + a} (1-\theta)^{n-\sum x_i + b-1} d\theta}{\int_0^1 \theta^{\sum x_i} (1-\theta)^{n-\sum x_i} \frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1} d\theta} \\ &= \frac{B(\sum_{i=1}^n x_i + a + 1, n - \sum_{i=1}^n x_i + b)}{B(\sum_{i=1}^n x_i + a, n - \sum_{i=1}^n x_i + b)} = \frac{\sum_{i=1}^n x_i + a}{n + a + b} \end{aligned}$$

Bayesovská funkce odhadu vzhledem k apriornímu beta rozdělení s parametry a a b je tedy

$$\frac{\sum_{i=1}^n x_i + a}{n + a + b}$$

Nyní se pokusme kvantifikovat riziko funkce $\frac{\sum_{i=1}^n x_i + a}{n + a + b}$ v naději, že se nám podaří zvolit a a b tak, že toto riziko bude konstantní. Definujme

$$\mathbf{t}_{A,B}^*(x_1, \dots, x_n) = A \sum_{i=1}^n x_i + B = \frac{\sum_{i=1}^n x_i + a}{n + a + b}$$

Pak

$$\begin{aligned} \mathbf{r}_{\mathbf{t}_{A,B}^*}(\theta) &= E[(A \sum_{i=1}^n X_i + B - \theta)^2] = E\left[\left[A \left(\sum_{i=1}^n X_i - n\theta\right) + B - \theta + nA\theta\right]^2\right] \\ &= A^2 E\left[\left(\sum_{i=1}^n X_i - n\theta\right)^2\right] + (B - \theta + nA\theta)^2 = nA^2\theta(1-\theta) + (B - \theta + nA\theta)^2 \\ &= \theta^2[(nA-1)^2 - nA^2] + \theta[nA^2 + 2(nA-1)B] + B^2 \end{aligned}$$

Tento výraz je konstantní pokud

$$\begin{aligned} (nA-1)^2 - nA^2 &= 0 \\ nA^2 + 2(nA-1)B &= 0 \end{aligned}$$

Výraz $(nA-1)^2 - nA^2$ je roven nule pro $A = \frac{1}{\sqrt{n}(\sqrt{n}+1)}$ a výraz $nA^2 + 2(nA-1)B$ je roven nule pro $B = -n\frac{A^2}{2}(nA-1)$, což odpovídá $\frac{1}{2}(\sqrt{n}+1)$ pro $A = \frac{1}{\sqrt{n}(\sqrt{n}+1)}$. Řešením pro a a b získáme $a = b = \frac{\sqrt{n}}{2}$. $\frac{\sum_{i=1}^n x_i + \sqrt{n}/2}{n + \sqrt{n}}$ je tak Bayesovskou funkcí odhadu s konstantním rizikem a tím pádem minimaxem.

7.7 Vektor parametrů

Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta)$, kde $\theta = (\theta_1, \dots, \theta_k)$ a $\bar{\Theta}$ je tak k -rozměrný prostor. Předpokládejme, že chceme současně odhadnout

$\tau_1(\theta), \dots, \tau_r(\theta)$, kde $r \leq k$ a $\tau_j(\theta)$ představuje funkci vektoru parametrů $\theta = (\theta_1, \dots, \theta_k)$. Funkce odhadu pro $(\tau_1(\theta), \dots, \tau_r(\theta))$ je vektor statistik (T_1, \dots, T_r) , kde $T_j = t_j(X_1, \dots, X_n)$ je funkcí odhadu pro $\tau_j(\theta)$.

Definice 7.31 (Nezkreslená funkce odhadu) *Funkce odhadu (T_1, \dots, T_r) , kde $T_j = t_j(X_1, \dots, X_n)$, je nazývána nezkreslenou funkcí odhadu pro $(\tau_1(\theta), \dots, \tau_r(\theta))$ právě tehdy a jen tehdy, jestliže $E_\theta[T_j] = \tau_j(\theta)$ pro všechna $j = 1, \dots, r$ a všechna $\theta \in \underline{\Theta}$.*

V předchozím textu, kde jsme taktéž uvažovali nezkreslenou funkci odhadu pro $\tau(\theta)$, jsme použili rozptyl této funkce jako měřítko její “blízkosti” k $\tau(\theta)$. Nyní potřebujeme zobecnit pojem rozptyl pro r dimenzí. K tomuto účelu lze použít (a) vektor rozptylů, (b) lineární kombinaci rozptylů, (c) elipsoid koncentrace a (d) Wilksovo zobecnění rozptylu.

Nejjednodušším vyjádřením rozptylu v r dimenzích je použití vektoru rozptylů nezkreslených funkcí odhadu T_1, \dots, T_r . To znamená, že k vyjádření “blízkosti” funkce odhadu (T_1, \dots, T_r) k $(\tau(\theta), \dots, \tau(\theta))$ slouží vektor $(D_\theta[T_1], \dots, D_\theta[T_r])$. Nevýhodou tohoto přístupu je skutečnost, že tato míra “blízkosti” má charakter vektoru, se kterým se ne vždy dobře pracuje. Jedním z možných řešení je použít lineární kombinaci rozptylů ve vektoru. Míra “blízkosti” pak má podobu $\sum_{j=1}^r a_j D_\theta[T_j]$, kde $a_j \leq 0$.

Jestliže jsou jednotlivá T_j vzájemně korelována, pak by míra “blízkosti” měla tuto korelaci zahrnovat.

Definice 7.32 *Nechť je (T_1, \dots, T_r) nezkreslenou funkcí odhadu pro $(\tau_1(\theta), \dots, \tau_r(\theta))$. Dále nechť $\sigma_{ij}(\theta) = \text{cov}_\theta[T_i, T_j]$. Čtvercová matice, jejíž ij -tý prvek je $\sigma_{ij}(\theta)$, nazýváme kovarianční maticí funkce odhadu (T_1, \dots, T_r) . ij -tý prvek inverzní matice ke kovarianční matici označme jako $\sigma^{ij}(\theta)$.*

Definice 7.33 (Elipsoid koncentrace) *Uvažujme nezkreslenou funkci odhadu (T_1, \dots, T_r) pro $(\tau_1(\theta), \dots, \tau_r(\theta))$. Elipsoid koncentrace pro (T_1, \dots, T_r) je definován jako vnitřní a obvodová část elipsoidu*

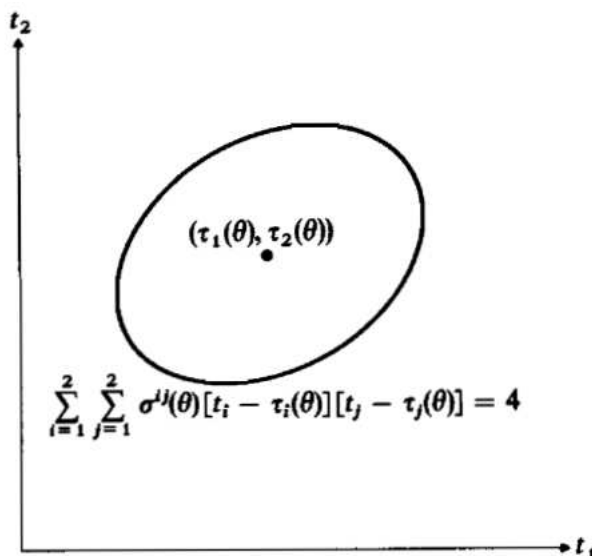
$$\sum_{i=1}^r \sum_{j=1}^r \sigma^{ij}(\theta) [t_i - \tau_i(\theta)] [t_j - \tau_j(\theta)] = r + 2$$

Situaci pro $r = 2$ ilustruje obrázek (7.3).

Elipsoid koncentrace tedy měří, jak je (T_1, \dots, T_r) “koncentrováno” okolo $(\tau_1(\theta), \dots, \tau_r(\theta))$. Funkce odhadu (T_1, \dots, T_r) , jejíž elipsoid koncentrace je obsažen v elipsoidu koncentrace jiné funkce odhadu (T'_1, \dots, T'_r) , je více “koncentrován” okolo $\tau_1(\theta), \dots, \tau_r(\theta)$ než (T'_1, \dots, T'_r) .

Determinant kovarianční matice funkce odhadu je proporcionální čtverci plochy odpovídajícímu elipsoidu koncentrace. Tímto se dostáváme k Wilksově zobecnění rozptylu.

Definice 7.34 (Wilksovo zobecnění rozptylu) *Uvažujme nezkreslenou funkci odhadu (T_1, \dots, T_r) pro $(\tau_1(\theta), \dots, \tau_r(\theta))$. Wilkosův zobecněný rozptyl funkce odhadu (T_1, \dots, T_r) je definován jako determinant její kovarianční matice.*

Obrázek 7.3: Elipsoid koncentrace pro $r = 2$

Následující věta představuje zobecnění věty (7.9) pro r dimenzí. Tuto větu uvádíme bez důkazu.

Věta 7.18 *Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x; \theta_1, \dots, \theta_k)$. Nechť $S_1 = \mathfrak{s}_1(X_1, \dots, X_n), \dots, S_m = \mathfrak{s}_m(X_1, \dots, X_n)$ jsou sdruženě dostatečné statistiky a nechť (T_1, \dots, T_r) představuje nezkreslenou funkci odhadu pro $(\tau_1(\theta), \dots, \tau_r(\theta))$. Definujme $T'_j = E[T_j | S_1, \dots, S_m]$ pro $j = 1, \dots, r$. Pak*

1. (T'_1, \dots, T'_r) je statistika a nezkreslená funkce odhadu pro $(\tau_1(\theta), \dots, \tau_r(\theta))$ a $T'_j = \mathfrak{t}'_j(S_1, \dots, S_m)$, tj. T'_j je funkcí sdruženě dostatečných statistik S_1, \dots, S_m .
2. $D_\theta[T'_j] \leq D_\theta[T_j]$ pro každé $\theta \in \bar{\Theta}$.
3. Elipsoid koncentrace pro (T'_1, \dots, T'_r) je obsažen v elipsoidu koncentrace pro (T_1, \dots, T_r) pro každé $\theta \in \bar{\Theta}$.

Bod (2) implikuje

$$\sum_{j=1}^r a_j D_\theta[T'_j] \leq \sum_{j=1}^r a_j D_\theta[T_j] \quad \text{pro } a_j \leq 0$$

a bod (3) implikuje, že Wilksův zobecněný rozptyl pro (T'_1, \dots, T'_r) je menší než Wilksův zobecněný rozptyl pro (T_1, \dots, T_r) .

Větu (7.11) lze taktéž zobecnit pro r dimenzí. Nejprve je však třeba zobecnit koncept úplnosti.

Definice 7.35 (Sdružená kompletnost) *Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x; \theta_1, \dots, \theta_k)$. Nechť (T_1, \dots, T_m) představuje soubor statistik. Statistiky T_1, \dots, T_m nazýváme sdruženě kompletní, jestliže $E_\theta[z(T_1, \dots, T_m)] \equiv 0$ pro všechna $\theta \in \bar{\Theta}$ implikuje $P_\theta[z(T_1, \dots, T_m) = 0] \equiv 1$ pro všechna $\theta \in \bar{\Theta}$, kde $z(T_1, \dots, T_m)$ je statistikou.*

Příklad 7.51 Uvažujme náhodný výběr z populace $f(x; \theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1} I_{(\theta_1, \theta_2)}(x)$, kde $\theta_1 < \theta_2$. Definujme $\theta = (\theta_1, \theta_2)$, $Y_1 = \min(X_1, \dots, X_n)$ a $Y_n = \max(X_1, \dots, X_n)$. Pokusme se dokázat, že X_1 a X_n jsou sdruženě kompletní¹⁵.

Uvažujme nezkreslenou funkci odhadu $z(Y_1, Y_n)$ pro θ , tj.

$$E_\theta[z(Y_1, \dots, Y_n)] \equiv 0 \quad \text{pro všechna } \theta \in \Theta$$

Dále

$$\begin{aligned} E_\theta[z(Y_1, Y_n)] &= \int \int z(y_1, y_n) f_{Y_1, Y_n}(y_1, y_n) dy_1 dy_n \\ &= \int_{\theta_1}^{\theta_2} \left(\int_{\theta_1}^{y_n} z(y_1, y_n) n(n-1) \left(\frac{y_n - \theta_1}{\theta_2 - \theta_1} - \frac{y_1 - \theta_1}{\theta_2 - \theta_1} \right)^{n-2} \frac{1}{\theta_2 - \theta_1} \frac{1}{\theta_2 - \theta_1} dy_1 \right) dy_n \end{aligned}$$

což je shodou okolností rovno nule jen tehdy, jestliže

$$\int_{\theta_1}^{\theta_2} \int_{\theta_1}^{y_n} z(y_1, y_n) (y_n - y_1)^{n-2} dy_1 dy_n \equiv 0 \quad \text{pro } \theta_1 < \theta_2$$

Derivací obou stran vzhledem k θ_2 získáme

$$\int_{\theta_1}^{\theta_2} z(y_1, \theta_2) (\theta_2 - y_1)^{n-2} dy_1 \equiv 0 \quad \text{pro všechna } \theta_1 < \theta_2$$

Následnou derivací dle θ_1 získáme

$$-z(\theta_1, \theta_2) (\theta_2 - \theta_1)^{n-2} \equiv 0 \quad \text{pro všechna } \theta_1 < \theta_2$$

a proto $z(\theta_1, \theta_2) = 0$ pro $\theta_1 < \theta_2$, tj. $z(\theta_1, \theta_2) = 0$ pro $y_1 < y_n$. Y_1 a Y_n jsou tak sdruženě kompletní.

Jestliže je $f(x; \theta_1, \dots, \theta_k)$ je členem k -parametrické rodiny exponenciálních rozdělení, pak lze soubor sdruženě kompletních a dostatečných statistik nalézt s pomocí následující věty. Tato věta, kterou uvádíme bez důkazu, je zobecněním věty (7.10). Navíc jsou opomenuty některé podmínky regularity, a proto není zcela přesně formulována.

Věta 7.19 Uvažujme náhodný výběr z populace $f(x; \theta_1, \dots, \theta_k)$. Jestliže je $f(x; \theta_1, \dots, \theta_k)$ členem k -parametrické rodiny exponenciálních rozdělení, tj.

$$f(x; \theta_1, \dots, \theta_k) = a(\theta_1, \dots, \theta_k) b(x) e^{\sum_{j=1}^k c_j(\theta_1, \dots, \theta_k) d_j(x)}$$

pak $(\sum d_1(X_i), \dots, \sum d_k(X_i))$ představuje soubor minimálních sdruženě kompletních a dostatečných statistik.

Příklad 7.52 Uvažujme náhodný výběr X_1, \dots, X_n z populace

$$f(x; \theta_1, \theta_2) = \phi_{\theta_1, \theta_2^2}(x) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{1}{2} \left(\frac{x - \theta_1}{\theta_2} \right)^2}$$

¹⁵Z předchozího textu již víme, že jsou sdruženě dostatečné.

Výše uvedenou pravděpodobnostní funkci lze také vyjádřit jako

$$f(x; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi}\theta_2} e^{-\frac{1}{2}\left(\frac{\theta_1}{\theta_2}\right)^2} e^{-\frac{x^2}{2\theta_2^2} + \frac{\theta_1 x}{\theta_2^2}}$$

a proto jsou, dle výše uvedené věty, $\sum X_i$ a $\sum X_i^2$ sdruženě kompletní a dostatečné statistiky.

Následující věta je vektorovou analogií věty (7.11). Jestliže definujeme UM-VUE jako optimum, pak lze s pomocí následující věty identifikovat optimální funkci odhadu pro vektor funkcí parametrů.

Věta 7.20 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x; \theta_1, \dots, \theta_k)$. Definujme $\theta = (\theta_1, \dots, \theta_k)$. Jestliže $S_1 = \mathbf{s}_1(X_1, \dots, X_n), \dots, S_m = \mathbf{s}_m(X_1, \dots, X_n)$ jsou sdruženě kompletní a dostatečné statistiky a jestliže existuje nezkreslená funkce odhadu pro $(\tau_1(\theta), \dots, \tau_r(\theta))$, pak existuje jedinečná nezkreslená funkce odhadu $T_1^* = \mathbf{t}_1^*(S_1, \dots, S_m), \dots, T_r^* = \mathbf{t}_r^*(S_1, \dots, S_m)$ pro $(\tau_1(\theta), \dots, \tau_r(\theta))$, která splňuje

- $D_\theta[T_j^*] \leq D_\theta[T_j]$ pro každé $\theta \in \overline{\Theta}$, $j = 1, \dots, r$ a libovolnou alternativní funkci odhadu (T_1, \dots, T_r) pro $(\tau_1(\theta), \dots, \tau_r(\theta))$,
- Elipsoid koncentrace pro (T_1^*, \dots, T_r^*) je obsažen v elipsoidu koncentrace pro (T_1, \dots, T_r) , kde (T_1, \dots, T_r) je libovolná alternativní nezkreslená funkce odhadu pro $(\tau_1(\theta), \dots, \tau_r(\theta))$.

Funkci odhadu (T_1^*, \dots, T_r^*) lze nalézt dvěma způsoby. Prvním je kvalifikovaný odhad $\mathbf{t}_1^*, \dots, \mathbf{t}_r^*$, které představují funkce sdruženě kompletních a dostatečných statistik S_1, \dots, S_m a které budou nezkreslenými funkcemi odhadu pro $(\tau_1(\theta), \dots, \tau_r(\theta))$. Druhým způsobem je nalezení libovolného souboru nezkreslených funkcí odhadu pro $(\tau_1(\theta), \dots, \tau_r(\theta))$ a následný výpočet jejich podmíněné střední hodnoty pro daný soubor sdruženě kompletních a dostatečných statistik S_1, \dots, S_m .

Příklad 7.53 Uvažujme náhodný výběr z populace $f(x; \theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1} I_{(\theta_1, \theta_2)}(x)$. Pokusme se “sdruženě” odhadnout $\tau_1(\theta) = \theta_2 - \theta_1$ a $\tau_2(\theta) = \frac{\theta_1 + \theta_2}{2}$.

Z předchozího textu víme, že $Y_1 = \min(X_1, \dots, X_n)$ a $Y_n = \max(X_1, \dots, X_n)$ jsou sdruženě kompletní a dostatečné statistiky. Abychom tak našli nezkreslenou funkci odhadu (T_1^*, T_2^*) s uniformě nejmenším rozptylem, stačí nalézt nezkreslenou funkci odhadu, která je funkcí těchto statistik. Protože $E[Y_1] = \theta_1 + \frac{\theta_2 - \theta_1}{n+1}$ a $E[Y_n] = \theta_2 - \frac{\theta_2 - \theta_1}{n+1}$, představuje $\left(\frac{n+1}{n-1}(Y_n - Y_1), \frac{Y_1 + Y_n}{2}\right)$ námi hledanou nezkreslenou funkci odhadu pro $(\theta_2 - \theta_1, \frac{\theta_1 + \theta_2}{2})$.

Příklad 7.54 Uvažujme náhodný výběr X_1, \dots, X_n z normálního rozdělení $f(x; \theta_1, \theta_2) = \phi_{\mu, \sigma^2}(x)$. Z předchozího textu víme, že $\sum X_i$ a $\sum X_i^2$ jsou sdruženě kompletní a dostatečné statistiky. Proto na základě věty (7.20) je $\left(\sum \frac{X_i}{n}, \sum \frac{(X_i - \bar{X})^2}{n-1}\right)$ nezkreslenou funkcí odhadu pro (μ, σ^2) jejíž elipsoid koncentrace je obsažen v elipsoidu koncentrace libovolné alternativní funkce odhadu¹⁶.

¹⁶Protože $\sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$, je funkce odhadu $\frac{\sum (X_i - \bar{X})^2}{n-1}$ funkcí sdruženě kompletních a dostatečných statistik $\sum X_i$ a $\sum X_i^2$.

Nyní předpokládejme, že chceme získat funkci odhadu pro $\theta = (\mu, \sigma^2)$ při splnění podmínky

$$\int_{\tau(\theta)}^{\infty} \phi_{\mu, \sigma^2}(x) dx = \alpha$$

pro známé a fixní α . $\tau(\theta)$ je $(1 - \alpha)$ -tý kvantil, tj. splňuje podmínku $P[X_i > \tau(\theta)] = \alpha$. To implikuje $1 - \alpha = \Phi\left(\frac{\tau(\theta) - \mu}{\sigma}\right)$ neboli $\tau(\theta) = \mu + z_{1-\alpha}\sigma$, kde $z_{1-\alpha}$ je definováno vztahem $\Phi(z_{1-\alpha}) = 1 - \alpha$. Abychom našli UMVUE pro $\tau(\theta)$, stačí nalézt nezkreslenou funkci odhadu pro $\mu + z_{1-\alpha}\sigma$, která je sama funkcí $\sum X_i$ a $\sum X_i^2$. Protože \bar{X} je UMVUE pro μ a protože lze dokázat, že

$$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{2}} \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} = T^*$$

je UMVUE pro σ , je $\bar{X} + z_{1-\alpha}T^*$ UMVUE pro $\tau(\theta)$. Při odvození jsme aplikovali větu (7.20) pro $r = 1$.

7.8 Odhad maximální věrohodnosti

Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$, kde θ je reálné číslo. Pro pozorované hodnoty x_1, \dots, x_n je odhad $\hat{\theta}$ parametru θ dán maximem funkce věrohodnosti

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i, \theta)$$

Označme funkci odhadu založenou na principu maximální věrohodnosti jako $\hat{\Theta}_n = \hat{\zeta}_n(X_1, \dots, X_n)$. Z předchozího textu víme, že je žádoucí, aby funkce odhadu splňovala některé podmínky.

Věta 7.21 *Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta)$, kde $f(x, \theta)$ splňuje určité podmínky regularity. Je-li $\hat{\Theta} = \hat{\zeta}_n(X_1, \dots, X_n)$ funkcí odhadu dle maximální věrohodnosti pro θ , pak*

1. $\hat{\Theta}_n$ je asymptoticky normálně rozdělené se střední hodnotou θ a rozptylem $\frac{1}{n} E \left[\left(\frac{\partial}{\partial \theta} \ln(f(X, \theta)) \right)^2 \right]$,
2. posloupnost funkcí odhadu $\hat{\Theta}_1, \dots, \hat{\Theta}_n$ je BAN.

Výše uvedená věta, kterou nebudeme dokazovat, říká, že pro náhodné výběry velkého rozsahu je funkce odhadu pro θ založená na metodě maximální věrohodnosti nejlepší možnou funkcí odhadu¹⁷. Dále platí, že rozptyl asymptotického normálního rozdělení v bodě (1) je Cramér-Raovou dolní mezí rozptylu.

Příklad 7.55 *Uvažujme náhodný výběr X_1, \dots, X_n z negativního exponenciálního rozdělení $f(x, \theta) = \theta e^{-\theta x} I_{[0, \infty)}(x)$. Lze odvodit, že funkce odhadu dle maximální věrohodnosti pro θ je $\frac{\sum X_i}{n} = \frac{1}{\bar{X}}$. Dle výše uvedené věty má tato funkce odhadu asymptoticky normální rozdělení se střední hodnotou θ a rozptylem¹⁸*

$$\frac{1}{n E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \ln(f(X, \theta)) \right)^2 \right]} = \frac{\theta^2}{n}$$

¹⁷Mohou existovat jiné funkce odhadu, které budou stejně dobré ale ne lepší.

¹⁸Viz. příklad (7.31)

Dle věty (7.2) je funkce odhadu pro $\tau(\theta)$ daná metodou maximální věrohodnosti $\tau(\hat{\Theta})$. Jestliže má $\tau(\cdot)$ derivaci prvního stupně, lze dokázat, že $\tau(\hat{\Theta})$ má asymptoticky normální rozdělení se střední hodnotou $\tau(\theta)$ a rozptylem

$$\frac{[\tau'(\theta)]^2}{nE_\theta \left[\left(\frac{\partial}{\partial \theta} \ln(f(X, \theta)) \right)^2 \right]}$$

který představuje Cramér-Raovu dolní mez rozptylu.

Funkce odhadu dle metody maximální věrohodnosti si zachovává výše uvedené vlastnosti také v případě, kdy je θ k -rozměrným vektorem. Lze totiž opět dokázat, že při splnění určitých podmínek regularity, lze funkce odhadu popsat pomocí asymptotického vícerozměrného normálního rozdělení. Je-li např. $k = 2$, tj. $\theta = (\theta_1, \theta_2)$, sledují $\hat{\Theta}_1$ a $\hat{\Theta}_2$ asymptoticky dvourozměrné normální rozdělení s parametry

$$\begin{aligned} \mu_1 &= \theta_1 \\ \sigma_1^2 &= -\frac{E_\theta \left[\frac{\partial^2}{\partial \theta_1^2} \ln(f(X, \theta)) \right]}{n\Delta} \\ \mu_2 &= \theta_2 \\ \sigma_2^2 &= -\frac{E_\theta \left[\frac{\partial^2}{\partial \theta_2^2} \ln(f(X, \theta)) \right]}{n\Delta} \\ \rho_{\sigma_1, \sigma_2} &= \frac{E_\theta \left[\frac{\partial^2}{\partial \theta_1 \partial \theta_2} \ln(f(X, \theta)) \right]}{n\Delta} \end{aligned}$$

kde

$$\Delta = E_\theta \left[\frac{\partial^2}{\partial \theta_1^2} \ln(f(X, \theta)) \right] E_\theta \left[\frac{\partial^2}{\partial \theta_2^2} \ln(f(X, \theta)) \right] - \left(E_\theta \left[\frac{\partial^2}{\partial \theta_1 \partial \theta_2} \ln(f(X, \theta)) \right] \right)^2$$

Příklad 7.56 Uvažujme náhodný výběr z populace

$$f(x, \theta) = f(x; \theta_1, \theta_2) = \phi_{\theta_1, \theta_2}(x) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{1}{2\theta_2}(x-\theta_1)^2}$$

V předchozích příkladech jsme odvodily funkce odhadu dle metody maximální věrohodnosti pro θ_1 a θ_2 .

$$\begin{aligned} \hat{\Theta}_1 &= \frac{1}{n} \sum_{i=1}^n X_i \\ \hat{\Theta}_2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\Theta}_1)^2 \end{aligned}$$

Asymptotické dvourozměrné normální rozdělení $\hat{\Theta}_1$ a $\hat{\Theta}_2$ má střední hodnoty θ_1

a θ_2 . Protože $\ln(f(X, \theta)) = -\frac{1}{2} \ln(\theta_2) - \frac{1}{2\theta_2}(X - \theta_1)^2$, jsou hledané derivace

$$\begin{aligned}\frac{\partial^2}{\partial \theta_1^2} \ln(f(X, \theta)) &= -\frac{1}{\theta_2} \\ \frac{\partial^2}{\partial \theta_2 \partial \theta_1} \ln(f(X, \theta)) &= -\frac{X - \theta_1}{\theta_2^2} \\ \frac{\partial^2}{\partial \theta_2^2} \ln(f(X, \theta)) &= \frac{1}{2\theta_2^2} - \frac{(X - \theta_1)^2}{\theta_2^3}\end{aligned}$$

Platí

$$\begin{aligned}E[X] &= \theta_1 \\ E[(X - \theta_1)^2] &= \theta_2 \\ -E_\theta \left[\frac{\partial^2}{\partial \theta_1^2} \ln(f(X, \theta)) \right] &= \frac{1}{\theta_2} \\ -E_\theta \left[\frac{\partial^2}{\partial \theta_2 \partial \theta_1} \ln(f(X, \theta)) \right] &= 0 \\ -E_\theta \left[\frac{\partial^2}{\partial \theta_2^2} \ln(f(X, \theta)) \right] &= \frac{1}{2\theta_2^2}\end{aligned}$$

což implikuje $\Delta = \frac{1}{2\theta_2^2}$. Zbývající parametry dvourozměrného normálního rozdělení tak jsou $\sigma_1^2 = \frac{\theta_2}{n}$, $\sigma_2^2 = \frac{2\theta_2^2}{n}$ a $\rho = 0$.

Kapitola 8

Intervalový odhad

8.1 Interval spolehlivosti

8.1.1 Definice intervalu spolehlivosti

Definice 8.1 (Interval spolehlivosti) Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$. Nechť $T_1 = t_1(X_1, \dots, X_n)$ a $T_2 = t_2(X_1, \dots, X_n)$ představují dvě statistiky, které splňují podmínku $T_1 \leq T_2$ a pro které platí $P_\theta[T_1 < \tau(\theta) < T_2] \equiv \gamma$, kde γ není funkcí θ . Náhodný interval (T_1, T_2) nazýváme 100γ procentním intervalem spolehlivosti pro $\tau(\theta)$ a T_1 resp. T_2 nazýváme jeho dolním resp. horním limitem. Hodnotu (t_1, t_2) náhodného intervalu (T_1, T_2) taktéž nazýváme 100γ procentním intervalem spolehlivosti pro $\tau(\theta)$.

Jedna ze statistik $t_1(X_1, \dots, X_n)$ a $t_2(X_1, \dots, X_n)$ může být konstantou.

Definice 8.2 (Jednostranný interval spolehlivosti) Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$. Nechť $T_1 = t_1(X_1, \dots, X_n)$ je statistika, pro kterou platí $P_\theta[T_1 < \tau(\theta)] \equiv \gamma$. Pak T_1 nazýváme levostranným limitem spolehlivosti pro $\tau(\theta)$. Podobně, je-li $T_2 = t_2(X_1, \dots, X_n)$ statistika, pro kterou platí $P_\theta[\tau(\theta) < T_2] \equiv \gamma$, pak nazýváme T_2 pravostranným intervalem spolehlivosti pro $\tau(\theta)$.

Příklad 8.1 Uvažujme náhodný výběr z populace $f(x, \theta) = \phi_{0,9}(x)$. Definujme $T_1 = t_1(X_1, \dots, X_n) = \bar{X} - \frac{6}{\sqrt{n}}$ a $T_2 = t_2(X_1, \dots, X_n) = \bar{X} + \frac{6}{\sqrt{n}}$. Pak (T_1, T_2) představuje interval spolehlivosti pro $\tau(\theta) = \theta$ s koeficientem spolehlivosti $\gamma = P_\theta[\bar{X} - \frac{6}{\sqrt{n}} < \theta < \bar{X} + \frac{6}{\sqrt{n}}] = P_\theta[-2 < \frac{\bar{X} - \theta}{3/\sqrt{n}} < 2] = \Phi(2) - \Phi(-2) = 0.9772 - 0.0228 = 0.9544$. Jestliže by střední hodnota náhodného výběru θ , řekněme, 25 prvcích byla 17.5, pak je interval $(17.5 - \frac{6}{\sqrt{25}}, 17.5 + \frac{6}{\sqrt{25}})$ nazýván 95.44 procentním intervalem spolehlivosti pro θ .

Z daného 100γ procentního intervalu spolehlivosti pro θ je možné odvodit 100γ procentní interval spolehlivosti pro $\tau(\theta)$ za předpokladu, že $\tau(\cdot)$ je striktně monotonní funkcí. Jestliže je $\tau(\cdot)$ rostoucí funkcí a (T_1, T_2) je 100γ procentním intervalem spolehlivosti pro θ , pak $(\tau(T_1), \tau(T_2))$ je 100γ procentním intervalem spolehlivosti pro $\tau(\theta)$, protože

$$P_\theta[\tau(T_1) < \tau(\theta) < \tau(T_2)] = P_\theta[T_1 < \theta < T_2]$$

Stejně jako v případě bodového odhadu čelíme dvojímu problému. Prvním problémem je definování metod pro nalezení intervalu spolehlivosti. Druhým problémem je nalezení vhodného kritéria pro porovnání všech možných intervalů spolehlivosti. Tímto se dostáváme k následující kapitole.

8.1.2 Centrální veličina

Definice 8.3 (Centrální veličina) Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta)$. Definujme $Q = q(X_1, \dots, X_n; \theta)$. Jestliže pravděpodobnostní funkce náhodné veličiny Q nezávisí na θ , pak ji nazýváme centrální veličinou.

Příklad 8.2 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \phi_{\theta,9}(x)$. $\bar{X} - \theta$ je centrální veličinou, protože $\bar{X} - \theta$ sleduje normální rozdělení se střední hodnotou nula a rozptylem $9/n$. Také $\frac{\bar{X} - \theta}{3/\sqrt{n}}$ je centrální veličinou, protože sleduje normalizované normální rozdělení. Naproti tomu \bar{X}/θ není centrální veličinou, protože sleduje normální rozdělení s jednotkovou střední hodnotou a rozptylem $9/\theta^2 n$.

Definice 8.4 (Metoda centrální veličiny) Jestliže $Q = q(X_1, \dots, X_n; \theta)$ je centrální veličinou, pak pro libovolné $0 < \gamma < 1$ existují q_1 a q_2 taková, že $P[q_1 < Q < q_2] = \gamma$. Jestliže pro libovolné realizace náhodného výběru x_1, \dots, x_n platí $q_1 < q(x_1, \dots, x_n; \theta) < q_2$ tehdy a jen tehdy, jestliže $t_1(x_1, \dots, x_n) < \tau(\theta) < t_2(x_1, \dots, x_n)$ ¹, pak (T_1, T_2) představuje 100γ procentní interval spolehlivosti pro $\tau(\theta)$, kde $T_i = t_i(X_1, \dots, X_n)$ pro $i = 1, 2$.

Je zřejmé, že q_1 a q_2 jsou nezávislé na θ , protože Q je taktéž nezávislé na θ .

Pro každou hodnotu γ existuje mnoho možných párů q_1 a q_2 , které splňují podmínku $P[q_1 < Q < q_2] = \gamma$. Různá q_1 a q_2 definují různé funkce t_1 a t_2 . Je zřejmé, že preferujeme taková t_1 a t_2 , která jsou si pokud možno co “nejblíže”. Definujme tuto “blízkost” jako $t_2(X_1, \dots, X_n) - t_1(X_1, \dots, X_n)$. Jestliže vzdálenost mezi t_1 a t_2 nemá charakter náhodné veličiny, pak vybereme taková q_1 a q_2 , která tuto vzdálenost minimalizují. Pokud má tato vzdálenost charakter náhodné veličiny, pak vybereme taková q_1 a q_2 , která minimalizují její střední hodnotu.

Základním požadavkem na centrální veličinu je, že nerovnost $\{q_1 < q(x_1, \dots, x_n; \theta) < q_2\}$ lze přepsat do tvaru $\{t_1(x_1, \dots, x_n) < \tau(\theta) < t_2(x_1, \dots, x_n)\}$ pro libovolnou realizaci náhodného výběru x_1, \dots, x_n . Tato vlastnost však není implikována výše uvedenou definicí centrální veličiny.

Příklad 8.3 Uvažujme náhodný výběr X_1, \dots, X_n z $\phi_{\theta,1}$. Nechť $\tau(\theta) = \theta$. Náhodná veličina $Q = q(X_1, \dots, X_n; \theta) = \frac{\bar{X} - \theta}{\sqrt{1/n}}$ sleduje normalizované normální rozdělení, a proto představuje centrální veličinu. Platí $f_Q(q) = \phi(q)$. Pro dané $0 < \gamma < 1$ existuje nekonečně párů q_1 a q_2 , které splňují podmínku $P[q_1 < Q < q_2] = \gamma$.

$\{q_1 < \frac{\bar{x} - \theta}{\sqrt{1/n}} < q_2\}$ implikuje $\{\bar{x} - q_2\sqrt{1/n} < \theta < \bar{x} - q_1\sqrt{1/n}\}$. Proto je interval $\{\bar{X} - q_2\sqrt{1/n} < \theta < \bar{X} - q_1\sqrt{1/n}\}$ 100γ procentním intervalem spolehlivosti pro θ . Délka tohoto intervalu je $(\bar{X} - q_1\sqrt{1/n}) - (\bar{X} - q_2\sqrt{1/n}) = (q_2 - q_1)\sqrt{1/n}$ a je proto nejkratší pro nejmenší hodnotu $q_2 - q_1$ při splnění

¹Připomeňme, že funkce $(t)_1$ a $(t)_2$ nezávisí na θ .

podmínky $\gamma = P[q_1 < Q < q_2] = \Phi(q_2) - \Phi(q_1)$. Délka tohoto intervalu tak dosahuje minima pro $q_1 = -q_2$.

8.2 Náhodný výběr z normálního rozdělení

8.2.1 Interval spolehlivosti pro střední hodnotu

Uvažujme náhodný výběr X_1, \dots, X_n z populace, která sleduje normální rozdělení, jejíž střední hodnotu ani rozptyl neznáme. To znamená, že $\theta = (\mu, \sigma)$ a $\tau(\theta) = \mu$. Náhodná veličina $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ sleduje normované normální rozdělení, a proto je centrální veličinou. Nicméně $\{q_1 < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < q_2\}$ nelze převést do tvaru $\{t_1(x_1, \dots, x_n) < \mu < t_2(x_1, \dots, x_n)\}$. Důvodem je, že $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ obsahuje neznámé σ . Připomeňme, že hledáme centrální veličinu, jejíž jedinou neznámou je μ . Nicméně víme, že

$$\frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)\sigma^2}}} = \frac{\bar{X} - \mu}{S_n/\sqrt{n}}$$

sleduje studentovo rozdělení s $n-1$ stupni volnosti². Pravděpodobnostní funkce náhodné veličiny $\frac{\bar{X} - \mu}{S_n/\sqrt{n}}$ je tak nezávislá na μ a σ^2 , a proto se jedná o centrální veličinu. Nerovnosti $\{q_1 < \frac{\bar{x} - \mu}{s/\sqrt{n}} < q_2\}$ implikují $\{\bar{x} - q_2 \frac{s}{\sqrt{n}} < \mu < \bar{x} - q_1 \frac{s}{\sqrt{n}}\}$, kde q_1 a q_2 splňují podmínku $P[q_1 < \frac{\bar{X} - \mu}{S_n/\sqrt{n}} < q_2] = \gamma$. Proto je $(\bar{X} - q_2 \frac{S_n}{\sqrt{n}}, (\bar{X} - q_1 \frac{S_n}{\sqrt{n}}))$ 100 γ procentním intervalem spolehlivosti pro μ . Délka tohoto intervalu je $\frac{q_2 - q_1}{S/\sqrt{n}}$, což je náhodná veličina. Pro libovolný náhodný výběr je jeho délka minimalizována pro nejmenší $q_2 - q_1$. To je splněno pro $q_1 = -q_2$, jak dokazuje následující text.

Naším cílem je minimalizovat

$$L = \frac{S_n}{\sqrt{n}}(q_2 - q_1)$$

při splnění podmínky

$$\int_{q_1}^{q_2} f_T(t) dt = \gamma$$

kde $f_T(t)$ je pravděpodobnostní funkce studentova rozdělení s $n-1$ stupni volnosti. Derivací výše uvedené rovnice podle q_1 získáme

$$f_T(q_2) \frac{dq_2}{dq_1} - f_T(q_1) = 0$$

Abychom minimalizovali L , musíme řešit rovnici $\frac{dL}{dq_1}$, tj.

$$\frac{dL}{dq_1} = \frac{S_n}{\sqrt{n}} \left(\frac{dq_2}{dq_1} - 1 \right) = 0$$

²Připomeňme, že $S_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$.

Zároveň však platí

$$\frac{S_n}{\sqrt{n}} \left(\frac{dq_2}{dq_1} - 1 \right) = \frac{S_n}{\sqrt{n}} \left(\frac{f_t(q_1)}{f_t(q_2)} - 1 \right) = 0$$

Tato rovnost je tedy splněna pro $f_T(q_2) = f_T(q_1)$, což v kontextu symetrického studentova rozdělení znamená $q_1 = -q_2$.

8.2.2 Interval spolehlivosti pro rozptyl

Uvažujme náhodný výběr X_1, \dots, X_n z populace, která sleduje normální rozdělení, jejíž střední hodnotu ani rozptyl neznáme. Víme, že náhodná veličina

$$Q = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)S_n^2}{\sigma^2}$$

sleduje chi-kvadrát rozdělení s $n-1$ stupni volnosti. Náhodná veličina Q je tedy centrální veličinou. Je zřejmé, že $\{q_1 < \frac{(n-1)s^2}{\sigma^2} < q_2\}$ implikuje $\{\frac{(n-1)s^2}{q_2} < \sigma^2 < \frac{(n-1)s^2}{q_1}\}$. To znamená, že $\left(\frac{(n-1)S_n^2}{q_2}, \frac{(n-1)S_n^2}{q_1}\right)$ je 100γ procentní interval spolehlivosti pro σ^2 , q_1 a q_2 splňují podmínku $P[q_1 < Q < q_2] = \gamma$.

q_1 a q_2 jsou často zvoleny tak, aby $P[Q < q_1] = P[Q > q_2] = \frac{1-\gamma}{2}$. O takovémto intervalu spolehlivosti pak hovoříme jako o intervalu spolehlivosti se shodnými chvosty.

Další možností je nalézt taková q_1 a q_2 , která minimalizují délku intervalu L , kde

$$L = (n-1)S_n^2 \left(\frac{1}{q_1} - \frac{1}{q_2} \right)$$

Nechť $f_Q(q)$ představuje pravděpodobnostní funkci chi-kvadrát rozdělení s $n-1$ stupni volnosti. Derivací podmínky

$$\int_{q_1}^{q_2} f_Q(q) dq = \gamma$$

dle q_1 získáváme

$$\frac{dq_2}{dq_1} f_Q(q_2) - f_Q(q_1) = 0$$

a proto

$$\frac{dL}{dq_1} = (n-1)S_n^2 \left(-\frac{1}{q_1^2} + \frac{1}{q_2^2} \frac{dq_2}{dq_1} \right) = (n-1)S_n^2 \left(-\frac{1}{q_1^2} + \frac{1}{q_2^2} \frac{f_Q(q_1)}{f_Q(q_2)} \right) = 0$$

což implikuje $q_1^2 f_Q(q_1) = q_2^2 f_Q(q_2)$. Délka intervalu spolehlivosti tak bude minimální pro q_1 a q_2 taková, že

$$q_1^2 f_Q(q_1) = q_2^2 f_Q(q_2)$$

při splnění podmínky

$$\int_{q_1}^{q_2} f_Q(q) dq = \gamma$$

Řešení pro q_1 a q_2 lze nalézt pomocí numerické integrace.

Závěrem uvedme, že pro libovolná q_1 a q_2 taková, že $\int_{q_1}^{q_2} f_Q(q) dq = \gamma$ představuje $\left(\sqrt{\frac{(n-1)S_n^2}{q_2}}, \sqrt{\frac{(n-1)S_n^2}{q_1}} \right)$ 100γ procentní interval spolehlivosti pro σ .

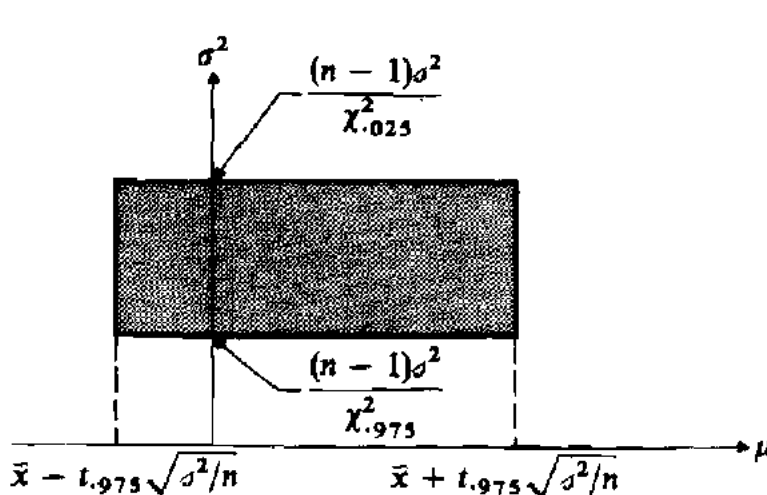
8.2.3 Region spolehlivosti pro střední hodnotu a rozptyl

Při konstrukci regionu spolehlivosti pro střední hodnotu μ a rozptyl σ^2 normálního rozdělení by se na první pohled mohlo zdát, že stačí rutinně aplikovat postup předešlých dvou kapitol. To znamená, že např. pro konstrukci 95 procentního regionu spolehlivosti ilustrovaného obrázkem (8.3) by stačilo použít vztahů

$$P \left[\bar{X} - t_{0.975} \frac{S_n^2}{n} < \mu < \bar{X} + t_{0.975} \frac{S_n^2}{n} \right] = 0.95$$

$$P \left[\frac{(n-1)S_n^2}{\chi_{0.975}^2} < \sigma^2 < \frac{(n-1)S_n^2}{\chi_{0.025}^2} \right] = 0.95$$

kde $t_{0.975}$ představuje 97.5 procentní kvantil studentova rozdělení s $n-1$ stupni volnosti a $\chi_{0.025}^2$ resp. $\chi_{0.975}^2$ představuje 2.5 resp. 97.5 procentní kvantil chi-kvadrát rozdělení s $n-1$ stupni volnosti. Region ilustrovaný obrázkem (8.3) je sice regionem spolehlivosti pro (μ, σ^2) , ale jeho koeficient spolehlivosti neznáme. Víme pouze, že tento koeficient není 0.95^2 , protože výše uvedené dva vztahy nejsou nezávislé.



Obrázek 8.1: Region spolehlivosti pro střední hodnotu a rozptyl

Při stanovení regionu spolehlivosti je třeba vzít v potaz vzájemnou nezávislost \bar{X} a S_n^2 . Protože $Q_1 = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ a $Q_2 = \frac{(n-1)S_n^2}{\sigma^2}$ představují centrální veličinu, lze nalézt q_1, q'_2 a q''_2 taková, že

$$P \left[-q_1 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < q_1 \right] = \gamma_1$$

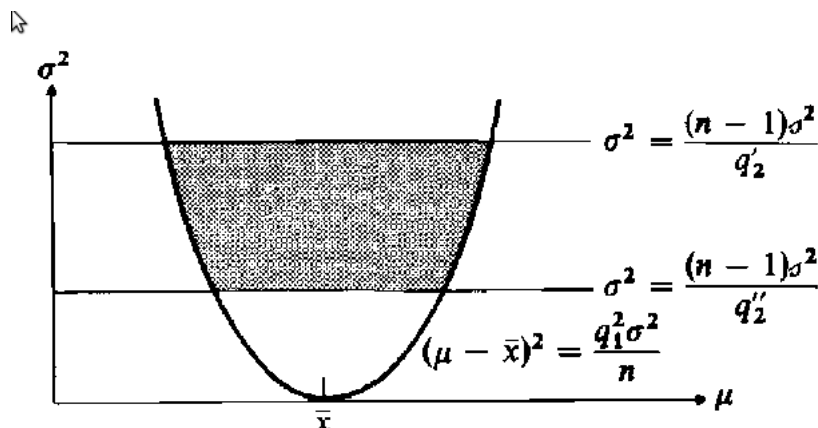
$$P \left[q'_2 < \frac{(n-1)S_n^2}{\sigma^2} < q''_2 \right] = \gamma_2$$

Protože jsou Q_1 a Q_2 vzájemně nezávislé, je jejich sdružená pravděpodobnost

rovna

$$P \left[-q_1 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < q_1, q'_2 < \frac{(n-1)S_n^2}{\sigma^2} < q'_2 \right] = \gamma_1 \gamma_2$$

Čtyři nerovnosti ve výše uvedeném výrazu určují region spolehlivosti, který ilustruje šedivá plocha obrázku (8.2).



Obrázek 8.2: Region spolehlivosti pro střední hodnotu a rozptyl

Pro (μ, σ) bychom odvodili region spolehlivosti stejně jako v případě (μ, σ^2) . Jediným rozdílem by bylo, že bychom pro osu y použili σ namísto σ^2 a parabola z obrázku (8.2) by se změnila na dvojici polopřímek daných rovnicemi $\mu = \bar{x} \pm q_1 \frac{\sigma}{\sqrt{n}}$, které se protínají v bodě \bar{x} a ose μ .

Závěrem je třeba zdůraznit, že region, který jsme zkonstruovali, nemá pro daná γ_1 a γ_2 minimální plochu. Výhoda výše popsaného postupu však spočívá v jeho jednoduchosti a ve skutečnosti, že pro přijatelně veliký náhodný výběr se takto získaný region spolehlivosti příliš neliší od minimálního regionu spolehlivosti. Minimální region spolehlivosti má přibližně eliptický tvar a jeho odvození je poměrně komplikované.

8.2.4 Interval spolehlivosti pro rozdíl středních hodnot

Dva nezávislé náhodné výběry

Uvažujme náhodný výběr X_1, \dots, X_m z normálního rozdělení se střední hodnotou μ_1 a rozptylem σ^2 a náhodný výběr Y_1, \dots, Y_n z normálního rozdělení se střední hodnotou μ_2 a rozptylem σ^2 . Předpokládejme, že tyto dva náhodné výběry jsou vzájemně nezávislé. Pokusme se nalézt interval spolehlivosti pro $\mu_2 - \mu_1$.

Náhodná veličina $\bar{Y} - \bar{X}$ je normálně rozdělená se střední hodnotou $\mu_2 - \mu_1$ a rozptylem $\frac{\sigma^2}{n} + \frac{\sigma^2}{m}$. Náhodná veličina $\frac{\sum (X_i - \bar{X})^2}{\sigma^2}$ sleduje chi-kvadrát rozdělení s $m - 1$ stupni volnosti a náhodná veličina $\frac{\sum (Y_i - \bar{Y})^2}{\sigma^2}$ sleduje chi-kvadrát rozdělení s $n - 1$ stupni volnosti. Proto náhodná veličina $\frac{\sum (X_i - \bar{X})^2}{\sigma^2} + \frac{\sum (Y_i - \bar{Y})^2}{\sigma^2}$ sleduje

chi-kvadrát rozdělení s $m + n - 2$ stupni volnosti. A konečně náhodná veličina

$$\begin{aligned} Q &= \frac{\frac{(\bar{Y} - \bar{X}) - (\mu_2 - \mu_1)}{\sqrt{\sigma^2/m + \sigma^2/n}}}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (Y_i - \bar{Y})^2}{\sigma^2(m+n-2)}}} \\ &= \frac{(\bar{Y} - \bar{X}) - (\mu_2 - \mu_1)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (Y_i - \bar{Y})^2}{m+n-2}}} \\ &= \frac{(\bar{Y} - \bar{X}) - (\mu_2 - \mu_1)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) S_n'^2}} \end{aligned}$$

sleduje studentovo rozdělení s $m + n - 2$ stupni volnosti. Z toho vyplývá $\gamma = P[-t_{(1+\gamma)/2} < Q < t_{(1+\gamma)/2}]$, kde $t_{(1+\gamma)/2}$ představuje $(1+\gamma)/2$ procentní kvantil studentova rozdělení s $m+n-2$ stupni volnosti. $S_n'^2$ je nezkreslená funkce odhadu společného rozptylu σ^2 . Dvojice nerovností

$$-t_{(1+\gamma)/2} < \frac{(\bar{y} - \bar{x}) - (\mu_2 - \mu_1)}{\sqrt{(1/m + 1/n)s'^2}} < t_{(1+\gamma)/2}$$

implikuje

$$(\bar{y} - \bar{x}) - t_{(1+\gamma)/2} \sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) s'^2} < \mu_2 - \mu_1 < (\bar{y} - \bar{x}) + t_{(1+\gamma)/2} \sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) s'^2}$$

a proto

$$\left((\bar{Y} - \bar{X}) - t_{(1+\gamma)/2} \sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) S_n'^2}, (\bar{Y} - \bar{X}) + t_{(1+\gamma)/2} \sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) S_n'^2} \right)$$

představuje 100γ procentní interval spolehlivosti pro $\mu_2 - \mu_1$.

Až dosud jsme předpokládali dva nezávislé náhodné výběry. Nyní předpokládejme, že $(X_1, Y_1), \dots, (X_n, Y_n)$ je náhodný výběr z dvourozměrného normálního rozdělení s parametry $\mu_1 = E[X], \mu_2 = E[Y], \sigma_1^2 = D[X], \sigma_2^2 = D[Y]$ a $\rho = \frac{\text{cov}[X, Y]}{\sigma_1 \sigma_2}$. Pokusme se nalézt interval spolehlivosti pro $\mu_2 - \mu_1$. Definujme $D_i = Y_i - X_i$ pro $i = 1, \dots, n$. Pak D_1, \dots, D_n jsou nezávislé náhodné veličiny, které sledují normální rozdělení se střední hodnotou $\mu_D = \mu_2 - \mu_1$ a rozptylem $\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$. Na náhodné veličiny D_1, \dots, D_n tak lze aplikovat postup z kapitoly (8.2.1). 100γ procentní interval spolehlivosti pro $\mu_2 - \mu_1$ je tedy

$$\left(\bar{D} - t_{(1+\gamma)/2} \sqrt{\frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n(n-1)}}, \bar{D} + t_{(1+\gamma)/2} \sqrt{\frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n(n-1)}} \right)$$

kde $t_{(1+\gamma)/2}$ představuje $(1+\gamma)/2$ procentní kvantil studentova rozdělení s $n-1$ stupni volnosti.

8.3 Metody stanovení intervalu spolehlivosti

8.3.1 Metoda centrální veličiny

V předchozím textu jsme popsali metodu centrální veličiny pro stanovení intervalu spolehlivosti. Otázkou však zůstává, zda-li vůbec, v kontextu daného

problému, centrální veličina existuje.

Tvrzení 8.1 *Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta)$. Předpokládejme, že kumulativní distribuční funkce $F(x, \theta)$, která odpovídá pravděpodobnostní funkci $f(x, \theta)$, je spojitá v x . Pak je zřejmé, že $F(X_i, \theta)$ sleduje uniformní rozdělení nad intervalem $(0, 1)$, a proto $-\ln(F(X_i, \theta))$ sleduje pravděpodobnostní rozdělení $e^{-u}I_{(0, \infty)}(u)$, což vyplývá ze vztahu*

$$P[-\ln(F(X_i, \theta)) \geq u] = P[\ln(F(X_i, \theta)) \leq -u] = P[F(X_i, \theta) \leq e^{-u}] = e^{-u}$$

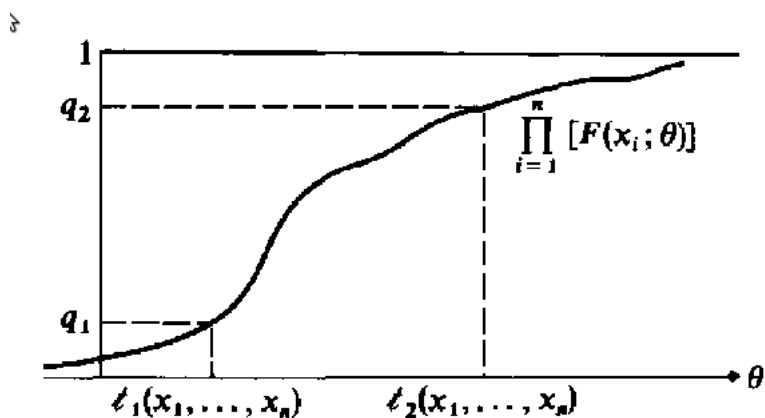
kde $u > 0$. Lze dokázat, že $-\sum \ln(F(X_i, \theta))$ sleduje gamma rozdělení s parametry n a 1 , tj.

$$\begin{aligned} & P \left[-\ln(q_2) < -\sum_{i=1}^n \ln(F(X_i, \theta)) < -\ln(q_1) \right] \\ &= \int_{-\ln(q_2)}^{-\ln(q_1)} \frac{1}{\Gamma(n)} z^{n-1} e^{-z} dz = P \left[q_1 < \prod_{i=1}^n F(X_i, \theta) < q_2 \right] \end{aligned}$$

pro $0 < q_1 < q_2 < 1$. Proto jsou $\prod_{i=1}^n F(X_i, \theta)$ a $-\sum_{i=1}^n \ln(F(X_i, \theta))$ centrální veličiny.

Z výše uvedeného tvrzení vyplývá, že pokud je kumulativní pravděpodobnostní funkce výběrové populace spojitá, centrální veličina existuje.

Tvrzení 8.2 *Jestliže je kumulativní pravděpodobnostní funkce výběrové populace monotónní v θ pro libovolné x , pak je $\prod_{i=1}^n F(x_i, \theta)$ taktéž monotónní v θ pro libovolná x_1, \dots, x_n . Tato monotónnost umožňuje nalezení intervalu spolehlivosti pro θ . Nerovnosti $q_1 < \prod_{i=1}^n F(x_i, \theta) < q_2$ totiž, jak ilustruje obrázek (??), pak implikují $t_1(x_1, \dots, x_n) < \theta < t_2(x_1, \dots, x_n)$.*



Obrázek 8.3: Inverze pravděpodobnostní funkce

Příklad 8.4 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \theta x^{\theta-1} I_{(0,1)}(x)$. Odpovídající kumulativní distribuční funkce má tedy tvar $F(x, \theta) = x^\theta I_{(0,1)}(x) + I_{[1,\infty)}(x)$. Jestliže jsou q_1 a q_2 vybrány tak, že

$$\begin{aligned} \gamma &= P \left[q_1 < \prod_{i=1}^n F(X_i, \theta) < q_2 \right] = P \left[q_1 < \prod_{i=1}^n X_i^\theta < q_2 \right] \\ &= P \left[\ln(q_1) < \theta \ln \left(\prod_{i=1}^n X_i \right) < \ln(q_2) \right] = P \left[-\ln(q_2) < -\theta \ln \left(\prod_{i=1}^n X_i \right) < -\ln(q_1) \right] \\ &= P \left[\frac{\ln(q_2)}{\ln \left(\prod_{i=1}^n X_i \right)} < \theta < \frac{\ln(q_1)}{\ln \left(\prod_{i=1}^n X_i \right)} \right] \end{aligned}$$

pak

$$\left(\frac{\ln(q_2)}{\ln \left(\prod_{i=1}^n X_i \right)}, \frac{\ln(q_1)}{\ln \left(\prod_{i=1}^n X_i \right)} \right)$$

představuje 100γ procentní interval spolehlivosti pro θ .

Závěrem uvedme poznámku ohledně existence centrální veličiny. Jestliže je θ lokální parametr, pak pravděpodobnostní funkce náhodné veličiny $X_i - \theta$ nezávisí na θ , a proto je centrální veličinou stejně jako např. $\sum X_i - n\theta, Y_j - \theta$ a $Y_1 + Y_n - 2\theta$. Jestliže je θ škálovací parametr, pak je pravděpodobnostní funkce náhodné veličiny X_i/θ nezávislá na θ , a proto se jedná centrální veličinu. Centrálními veličinami jsou v tomto případě také $\sum X_i/\theta, Y_j/\theta$ apod.

8.3.2 Statistická metoda

Opět uvažujme náhodný výběr X_1, \dots, X_n z populace $f(\cdot, \theta_0)$. Předpokládejme, že θ_0 , které v následujícím textu představuje skutečnou hodnotu parametru θ , je reálné číslo a že $\underline{\Theta}$ představuje interval na reálné ose. Pokusme se stanovit interval spolehlivosti pro θ_0 .

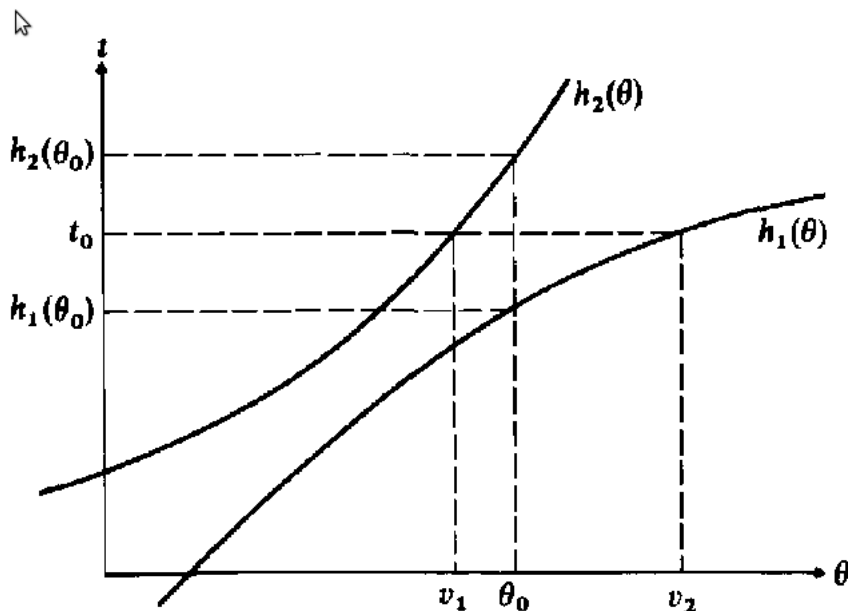
Nechť $T = t(X_1, \dots, X_n)$ představuje statistiku, kterou vybereme jedním z následujících dvou způsobů. Jestliže existuje jednorozměrná dostatečná statistika, pak tuto statistiku použijeme namísto T . Jestliže dostatečná statistika neexistuje, pak pro T použijeme bodovou funkci odhadu pro θ . Konkrétní volba statistiky T může také záviset na tom, jak složitá je její aplikace při konstrukci intervalu spolehlivosti - kritickým bodem je zejména odvození její pravděpodobnostní funkce.

Nechť $f_T(t, \theta)$ představuje pravděpodobnostní funkci statistiky T^3 . Definujme funkce $h_1(\theta)$ a $h_2(\theta)$, pro které platí

$$\begin{aligned} \int_{-\infty}^{h_1(\theta)} f_T(t, \theta) dt &= p_1 \\ \int_{h_2(\theta)}^{\infty} f_T(t, \theta) dt &= p_2 \end{aligned}$$

³V následujícím textu budeme předpokládat, že T je spojitá náhodná veličina. Analogický postup však lze aplikovat také v případě nespojité náhodné veličiny.

kde p_1 a p_2 jsou daná reálná čísla, která splňují podmínky $0 < p_1, 0 < p_2$ a $p_1 + p_2 < 1$. $h_1(\theta)$ a $h_2(\theta)$ lze znázornit jako funkce parametru θ . Předpokládejme, že $h_1(\theta)$ a $h_2(\theta)$ jsou rostoucí funkce, pro které navíc platí $h_1(\theta) < h_2(\theta)$. Příklad dvou takovýchto funkcí ilustruje obrázek (8.4). Nechť t_0 označuje hod-



Obrázek 8.4: Graf funkcí $h_1(\cdot)$ a $h_2(\cdot)$

notu statistiky T , tj. $t_0 = t(x_1, \dots, x_n)$. Z této hodnoty lze, jak ilustruje obrázek (8.4), získat hodnoty v_1 a v_2 takové, že $v_1 = v_1(t_0)$ a $v_2 = v_2(t_0)$. Z obrázku je dále zřejmé, že $h_1(\theta_0) < t_0 = t(x_1, \dots, x_n) < h_2(\theta_0)$ platí pouze tehdy a jen tehdy, jestliže $v_1 = v_1(x_1, \dots, x_n) < \theta_0 < v_2 = v_2(x_1, \dots, x_n)$ pro libovolnou realizaci náhodného výběru x_1, \dots, x_n . Z definice $h_1(\cdot)$ a $h_2(\cdot)$ však vyplývá

$$P_{\theta_0}[h_1(\theta_0) < t(X_1, \dots, X_n) < h_2(\theta_0)] = 1 - p_1 - p_2$$

a proto

$$P_{\theta_0}[v_1(X_1, \dots, X_n) < \theta_0 < v_2(X_1, \dots, X_n)] = 1 - p_1 - p_2$$

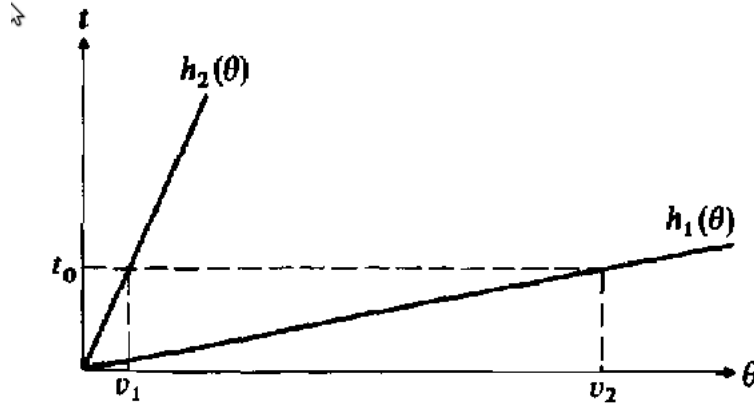
Interval (V_1, V_2) pak představuje $100(1 - p_1 - p_2)$ procentní interval spolehlivosti pro θ_0 , kde $V_i = v_i(X_1, \dots, X_n)$ pro $i = 1, 2$.

Příklad 8.5 Uvažujme náhodný výběr z populace $f(x, \theta_0) = \frac{1}{\theta_0} I_{(0, \theta_0)}(x)$. Pokusme se stanovit interval spolehlivosti pro θ_0 .

O $Y_n = \max(X_1, \dots, X_n)$ víme, že se jedná o dostatečnou statistiku a že se jedná o funkci odhadu parametru θ_0 založenou na metodě maximální věrohodnosti. Jestliže použijeme Y_n jakožto statistiku T , pak

$$f_T(t, \theta) = n \left(\frac{t}{\theta} \right)^{n-1} \frac{1}{\theta} I_{(0, \theta)}(t)$$

$\int_0^{h_1(\theta)} nt^{n-1}\theta^{-n}dt = p_1$ implikuje $\int_0^{h_1(\theta)} t^{n-1}dt = \frac{\theta^n p_1}{n}$, což implikuje $\frac{(h_1(\theta))^n}{n} = \frac{\theta^n p_1}{n}$ neboli $h_1(\theta) = \theta p_1^{1/n}$. Podobně $p_2 = \int_{h_2(\theta)}^{\theta} nt^{n-1}\theta^{-n}dt$ implikuje $\theta^n - [h_2(\theta)]^n = \theta^n p_2$ neboli $h_2(\theta) = \theta(1 - p_2)^{1/n}$. Situaci ilustruje obrázek (8.5), který je konkretizací obrázku (8.4) v kontextu našeho příkladu. Pro $t_0 =$

Obrázek 8.5: Graf funkcí $h_1(\cdot)$ a $h_2(\cdot)$

$\max(x_1, \dots, x_n)$ je v_1 definováno jako $h_2(v_1) = t_0$, tj. $h_2(v_1) = v_1(1 - p_2)^{1/n} = t_0$ neboli $v_1 = t_0(1 - p_2)^{-1/n}$. Podobně lze odvodit $v_2 = t_0 p_1^{-1/n}$. To znamená, že $100(1 - p_1 - p_2)$ procentní interval spolehlivosti pro θ_0 je dán intervalem $(Y_n(1 - p_2)^{-1/n}, Y_n p_1^{-1/n})$.

Naším dalším cílem je stanovit p_1 a p_2 tak, aby interval spolehlivosti

$$L = Y_n[p_1^{-1/n} - (1 - p_2)^{-1/n}]$$

byl pokud možno co nejkratší při splnění podmínek $1 - p_1 - p_2 = \gamma$ a $0 < p_1 + p_2 < 1$. To je splněno pro $p_2 = 0$ a $p_1 = 1 - \gamma$.

Z výše uvedeného příkladu je zřejmé, že funkce $h_1(\theta)$ a $h_2(\theta)$ nejsou vlastně potřeba. Pro danou hodnotu $t_0 = t(x_1, \dots, x_n)$ statistiky T je zapotřebí nalézt $v_1 = v_1(x_1, \dots, x_n)$ a $v_2 = v_2(x_1, \dots, x_n)$. v_2 lze získat řešením rovnice $p_1 = \int_{-\infty}^{h_1(\theta)=t_0} f_T(t, \theta)dt$ pro θ . Podobně v_2 lze získat řešením rovnice $p_2 = \int_{h_2(\theta)=t_0}^{\infty} f_T(t, \theta)dt$.

Výše uvedený postup lze aplikovat také pro nespojitě náhodné veličiny. Jediným rozdílem bude nahrazení integrálů sumami.

Příklad 8.6 Uvažujme náhodný výběr X_1, \dots, X_n z Bernoulliho rozdělení $P[X = 1] = \theta_0 = 1 - P[X = 0]$. Víme, že $T = \sum X_i$ je dostatečná statistika s binomickým rozdělením, tj.

$$P[T = t] = \binom{n}{t} \theta_0^t (1 - \theta_0)^{n-t}$$

pro $t = 0, 1, \dots, n$. Pokusme se stanovit interval spolehlivosti pro θ_0 .

Předpokládejme, že $T = t_0$. Řešme rovnice

$$p_1 = \sum_{t=0}^{t_0} \binom{n}{t} \theta^t (1-\theta)^{n-1}$$

$$p_2 = \sum_{t=0}^{t_0} \binom{n}{t} \theta^t (1-\theta)^{n-1}$$

pro θ . Jestliže např. $p_1 = 0.0509, p_2 = 0.0159, n = 20$ a $T = 4$, pak je 93.33 procentní interval spolehlivosti parametru θ_0 dán intervalem $(0.05, 0.40)$.

8.4 Interval spolehlivosti pro výběry velkého rozsahu

V případě bodového odhadu jsme ukázali, že v některých případech je možné nalézt posloupnost funkcí odhadu $T_n = t_n(X_1, \dots, X_n)$ pro parametr θ pravděpodobnostního rozdělení $f(\cdot, \theta)$, které sledují asymptoticky normální rozdělení. To znamená, že T_n asymptoticky sleduje normální rozdělení se střední hodnotou θ a rozptylem $\sigma_n^2(\theta)$. V kapitole (7.8) jsme ukázali, že v případě náhodného výběru velkého rozsahu sleduje funkce odhadu $\hat{\Theta}_n = \hat{\zeta}_n(X_1, \dots, X_n)$ založená na metodě maximální věrohodnosti přibližně normální rozdělení se střední hodnotou θ a rozptylem

$$\sigma_n^2(\theta) = \frac{1}{nE_\theta \left[\left(\frac{\partial}{\partial \theta} \ln(f(X, \theta)) \right)^2 \right]}$$

V tomto případě lze $\frac{T_n - \theta}{\sigma_n(\theta)}$ použít jako přibližnou centrální veličinu a příslušný 100γ procentní interval spolehlivosti získat z nerovností

$$-z < \frac{T_n - \theta}{\sigma_n(\theta)} < z$$

kde $z = z_{(1+\gamma)/2}$ je definováno jako $\Phi(z_{(1+\gamma)/2}) = (1+\gamma)/2$ neboli $\Phi(z) - \Phi(-z) = \gamma$. Tento postup lze aplikovat vždy, když lze z výše uvedených nerovností izolovat θ .

Příklad 8.7 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \theta e^{-\theta x} I_{(0, \infty)}(x)$. Z příkladu (7.55) víme, že funkce odhadu dle metody maximální věrohodnosti pro θ má tvar $1/\bar{X}_n$ a asymptoticky sleduje normální rozdělení se střední hodnotou θ a rozptylem

$$\sigma_n^2(\theta) = \frac{1}{nE_\theta \left[\left(\frac{\partial}{\partial \theta} \ln(f(X, \theta)) \right)^2 \right]} = \frac{\theta^2}{n}$$

Proto

$$\begin{aligned}\gamma &\approx P \left[-z < \frac{1/\bar{X}_n - \theta}{\sqrt{\theta^2/n}} < z \right] \\ &= P \left[-\frac{z\theta}{\sqrt{n}} < \frac{1}{\bar{X}_n} - \theta < \frac{z\theta}{\sqrt{n}} \right] \\ &= P \left[\frac{1/\bar{X}_n}{1 + z/\sqrt{n}} < \theta < \frac{1/\bar{X}_n}{1 - z/\sqrt{n}} \right]\end{aligned}$$

V případě náhodného výběru velkého rozsahu tak má 100γ procentní interval spolehlivosti pro parametr θ tvar

$$\left(\frac{1/\bar{X}}{1 + z/\sqrt{n}}, \frac{1/\bar{X}}{1 - z/\sqrt{n}} \right)$$

kde z je definováno jako $\Phi(z) - \Phi(-z) = \gamma$.

Příklad 8.8 Uvažujme náhodný výběr velkého rozsahu z Bernoulliho rozdělení s parametry $\theta = P[X = 1] = 1 - P[X = 0]$. Funkce odhadu pro θ dle metody maximální věrohodnosti má tvar $\hat{\Theta} = \bar{X}_n$ a její rozptyl je $\sigma_n^2(\theta) = \theta(1 - \theta)/n$. 100γ procentní interval spolehlivosti pro θ lze získat z nerovností

$$P \left[\frac{2n\hat{\Theta} + z^2 - z\sqrt{4n\hat{\Theta} + z^2 - 4n\hat{\Theta}^2}}{2(n + z^2)} < \theta < \frac{2n\hat{\Theta} + z^2 + z\sqrt{4n\hat{\Theta} + z^2 - 4n\hat{\Theta}^2}}{2(n + z^2)} \right] \approx \gamma$$

Vzhledem k tomu, že předmětem našeho zájmu je náhodný výběr velkého rozsahu, lze všechny členy výše uvedených nerovností, které obsahují $1/\sqrt{n}$ resp. $1/n$, zanedbat. Tím se dostáváme k

$$P \left[\hat{\Theta} - z\sqrt{\frac{\hat{\Theta} - (1 - \hat{\Theta})}{n}} < \theta < \hat{\Theta} + z\sqrt{\frac{\hat{\Theta}(1 - \hat{\Theta})}{n}} \right] \approx \gamma$$

Vztah odvozený ve výše uvedeném příkladě je vztahem, který bychom získali, pokud bychom v $\sigma_n^2(\theta)$ použili $\hat{\Theta}$ namísto θ . Tato substituce by implikovala, že $\frac{\hat{\Theta} - \theta}{\sqrt{\hat{\Theta}(1 - \hat{\Theta})/n}}$ je přibližně normálně rozdělené s nulovou střední hodnotou a jednotkovým rozptylem. Tento závěr, který však nebudeme dokazovat, je obecně platný. To znamená, že funkce odhadu $\hat{\Theta}$ dle metody maximální věrohodnosti asymptoticky sleduje normální rozdělení se střední hodnotou θ a rozptylem $\sigma_n^2(\theta)$, který lze v případě náhodného výběru dostatečně velkého rozsahu aproximovat pomocí $\sigma_n^2(\hat{\Theta})$. Je zřejmé, že toto zjištění velmi zjednoduší konstrukci intervalu spolehlivosti, protože

$$P \left[-z < \frac{\hat{\Theta} - \theta}{\sigma_n(\hat{\Theta})} < z \right] \approx \gamma$$

lze snadno převést do tvaru

$$P[\hat{\Theta} - z\sigma_n(\hat{\Theta}) < \theta < \hat{\Theta} + z\sigma_n(\hat{\Theta})] \approx \gamma$$

Závěrem uvedme, že interval spolehlivosti pro náhodný výběr velkého rozsahu založený na funkci odhadu dle metody maximální věrohodnosti je optimální. To znamená, že je v průměru kratší než interval spolehlivosti založený na alternativních funkcích odhadu.

8.5 Bayesovský interval

Následující text navazuje na kapitolu (7.6). Uvažujme náhodný výběr z pravděpodobnostní funkce $f(\cdot|\theta)$, jejíž funkční předpis je znám. Připomeňme, že θ je hodnota náhodné veličiny Θ , která je charakterizována apriorní pravděpodobnostní funkcí $g_\Theta(\cdot)$. Posteriorní pravděpodobnostní funkce náhodné veličiny Θ pro dané $(X_1, \dots, X_n) = (x_1, \dots, x_n)$ má tvar

$$f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n) = \frac{[\prod_{i=1}^n f(x_i|\theta)]}{\int [\prod_{i=1}^n f(x_i|\theta)] g_\Theta(\theta) d\theta}$$

Pro fixní γ je libovolný interval (t_1, t_2) splňující podmínku

$$\int_{t_1}^{t_2} f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n) d\theta = \gamma$$

100γ procentním Bayesovským intervalem odhadu pro parameter θ . V praxi bychom se snažili zvolit taková t_1 a t_2 , aby délka intervalu byla co nejmenší. Dále platí, že $t_i = t_i(x_1, \dots, x_n)$, tj. na t_i lze pohlížet jako na funkci pozorování x_1, \dots, x_n .

Příklad 8.9 Uvažujme náhodný výběr X_1, \dots, X_n z normálního rozdělení se střední hodnotou θ a jednotkovým rozptylem. Předpokládejme, že Θ má normální rozdělení se střední hodnotou x_0 a rozptylem 1. Pokusme se odhadnout θ .

Z příkladu (7.48) víme, že posteriorní rozdělení náhodné veličiny Θ sleduje normální rozdělení se střední hodnotou $\frac{\sum_{i=0}^n x_i}{n+1}$ a rozptylem $\frac{1}{n+1}$. Hledáme t_1 a t_2 , která splňují podmínku

$$\begin{aligned} \gamma &= \int_{t_1}^{t_2} f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n) d\theta \\ &= \Phi\left(\frac{t_2 - \sum_{i=0}^n \frac{x_i}{n+1}}{\sqrt{1/(n+1)}}\right) - \Phi\left(\frac{t_1 - \sum_{i=0}^n \frac{x_i}{n+1}}{\sqrt{1/(n+1)}}\right) \end{aligned}$$

Jestliže zvolíme z takové, že $\Phi(z) - \Phi(-z) = \gamma$, pak

$$\begin{aligned} t_2 &= \frac{\sum_{i=0}^n x_i}{n+1} + z\sqrt{\frac{1}{n+1}} \\ t_1 &= \frac{\sum_{i=0}^n x_i}{n+1} - z\sqrt{\frac{1}{n+1}} \end{aligned}$$

definuje nejkratší Bayesovský interval odhadu pro θ . Odpovídající interval spolehlivosti založený na metodě maximální věrohodnosti je dán intervalem $\left(\frac{\sum x_i}{n} - z\sqrt{\frac{1}{n}}, \frac{\sum x_i}{n} + z\sqrt{\frac{1}{n}}\right)$. Jediný rozdíl mezi těmito dvěma metodami spočívá v tom, že v případě Baysovského intervalu je velikost náhodného výběru zdánlivě navýšena o jedna, kde toto “dodatečné” pozorování je představováno střední hodnotou apriorního normálního rozdělení náhodné veličiny Θ .

Kapitola 9

Testování hypotéz

9.1 Základní pojmy

Definice 9.1 (Statistická hypotéza) *Statistickou větou se rozumí tvrzení a pravděpodobnostním rozdělení jedné nebo více náhodných veličin. Jestliže statistická hypotéza plně specifikuje toto pravděpodobnostní rozdělení, nazýváme ji jednoduchou statistickou hypotézou. V opačném případě mluvíme o tzv. složené statistické hypotéze.*

Příklad 9.1 *Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \phi_{\theta, 25}(x)$. Statistickou hypotézu, že střední hodnota tohoto normálního rozdělení je menší nebo rovna 17, zapisujeme jako $\mathcal{H} : \theta \leq 17$. Takto definovaná hypotéza je složená, protože plně nespecifikuje normální rozdělení. Naproti tomu hypotéza, že střední hodnota normálního rozdělení je rovna 17, tj. $\mathcal{H} : \theta = 17$, je jednoduchou statistickou hypotézou.*

Definice 9.2 (Test statistické hypotézy) *Testem statistické hypotézy \mathcal{H} s rozumí pravidlo, na základě kterého se tato hypotéza přijme popř. zamítne. Test statistické hypotézy značíme jako \mathcal{T} .*

Příklad 9.2 *Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \phi_{\theta, 25}(x)$. Definujme statistickou hypotézu $\mathcal{H} : \theta \leq 17$. Jedním z možných testů této hypotézy je zamítnutí \mathcal{H} , jestliže $\bar{X} > 17 + 5/\sqrt{n}$.*

Test hypotézy může být náhodný popř. nenáhodný. Test \mathcal{T} z výše uvedeného příkladu je testem nenáhodným. Alternativní test \mathcal{T}' hypotézy \mathcal{H} z výše uvedeného příkladu by mohl spočívat v hodu mincí a následném zamítnutí \mathcal{H} , pokud by padla panna. Takovýto test by byl testem náhodným.

Stejně jako v předchozím textu budeme \mathfrak{X} používat pro označení výběrového prostoru pozorování, tj. $\mathfrak{X} = \{(x_1, \dots, x_n) : (x_1, \dots, x_n) \text{ představují všechny možné kombinace náhodných veličin } (X_1, \dots, X_n)\}$.

Definice 9.3 (Nenáhodný test a kritický region) *Uvažujme test \mathcal{T} statistické hypotézy H , který tuto hypotézu zamítne, jestliže $(x_1, \dots, x_n) \in C_{\mathcal{T}}$, kde $C_{\mathcal{T}}$ je podmnožinou \mathfrak{X} . Pak test \mathcal{T} nazýváme nenáhodným testem a $C_{\mathcal{T}}$ jeho kritickým regionem.*

Příklad 9.3 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \phi_{\theta, 25}(x)$, tj. \mathfrak{X} je n -rozměrný euklidovský prostor. Dále uvažujme hypotézu $\mathcal{H} : \theta \leq 17$ a test \mathcal{T} , který zavrhuje tuto hypotézu pro $\bar{x} > 17 + 5/\sqrt{n}$. Pak je \mathcal{T} nenáhodným testem a $C_{\mathcal{T}} = \{(x_1, \dots, x_n) : \bar{x} > 17 + 5/\sqrt{n}\}$.

V případě nenáhodného testu \mathcal{T} hypotézy \mathcal{H} je tak \mathfrak{X} rozděleno na $C_{\mathcal{T}}$ a $\overline{C}_{\mathcal{T}}$ takové, že pro $(x_1, \dots, x_n) \in C_{\mathcal{T}}$ je \mathcal{H} zamítnuto. Nenáhodný test je tak specifikován svým kritickým regionem.

Definice 9.4 (Náhodný test) Test \mathcal{T} hypotézy \mathcal{H} je náhodný, jestliže je \mathcal{T} definován funkcí

$$\psi_{\mathcal{T}}(x_1, \dots, x_n) = P[\mathcal{H} \text{ je zamítnuta pro realizaci náhodného výběru } (x_1, \dots, x_n)]$$

Tuto funkci nazýváme kritickou funkcí testu \mathcal{T} .

Při aplikaci náhodného testu \mathcal{T} hypotézy \mathcal{H} je pro realizovaný náhodný výběr (x_1, \dots, x_n) ohodnocena funkce $\psi_{\mathcal{T}}(x_1, \dots, x_n)$. Následně je proveden pomocný výběr z Bernoulliho rozdělení s pravděpodobností úspěchu $\psi_{\mathcal{T}}(x_1, \dots, x_n)$. Jestliže tento výběr indikuje “úspěch”, je \mathcal{H} zamítnuto. Z tohoto popisu je zřejmé, že náhodné testy nejsou v praxi příliš používány. V některých případech je \mathfrak{X} rozděleno na tři části, kde pro jednu je \mathcal{H} přijato, pro druhou je \mathcal{H} zamítnuto a pro třetí, která představuje “šedou zónu”, je aplikován náhodný test. Situaci ilustruje následující příklad.

Příklad 9.4 Uvažujme náhodný výběr X_1, \dots, X_{10} z populace $f(x, \theta) = \theta^x(1 - \theta)^{1-x}$ pro $x = 0, 1$. Předpokládejme, že chceme testovat hypotézu $\mathcal{H} : \theta < \frac{1}{2}$. Jedním z možných testů je \mathcal{T} definovaný jako

1. zamítni \mathcal{H} , jestliže $\sum_{i=1}^{10} x_i > 5$,
2. přijmi \mathcal{H} , jestliže $\sum_{i=1}^{10} x_i < 5$ a
3. hod minci¹ a rozhodni, jestliže $\sum_{i=1}^{10} x_i = 5$

Test \mathcal{T} tedy rozděluje \mathfrak{X} na tři části A, B a C , kde

$$\begin{aligned} A &= \left\{ (x_1, \dots, x_{10}) : \sum_{i=1}^{10} x_i < 5 \right\} \\ B &= \left\{ (x_1, \dots, x_{10}) : \sum_{i=1}^{10} x_i = 5 \right\} \\ C &= \left\{ (x_1, \dots, x_{10}) : \sum_{i=1}^{10} x_i > 5 \right\} \end{aligned}$$

Kritická funkce testu \mathcal{T} je pak definována jako

$$\psi_{\mathcal{T}}(x_1, \dots, x_n) = \begin{cases} 1 & \text{pro } (x_1, \dots, x_{10}) \in C \\ \frac{1}{2} & \text{pro } (x_1, \dots, x_{10}) \in B \\ 0 & \text{pro } (x_1, \dots, x_{10}) \in A \end{cases}$$

¹Hod mincí zde zastupuje pomocný výběr z Bernoulliho rozdělení.

V předchozím textu jsme uvedli, že nenáhodný test je specifikován svým kritickým regionem a že stejnou úlohu v případě náhodného testu pak plní kritická funkce. Platí, že libovolná funkce $\psi(\cdot, \dots, \cdot)$ s definičním oborem \mathfrak{X} a oborem hodnot $[0, 1]$ je možnou kritickou funkcí, která definuje náhodný test.

Jestliže je test \mathcal{T} definován kritickou funkcí

$$\psi_{\mathcal{T}}(x_1, \dots, x_n) = \begin{cases} 1 & \text{pro } (x_1, \dots, x_n) \in C_{\mathcal{T}} \\ 0 & \text{v ostatních případech} \end{cases}$$

pak je \mathcal{T} nenáhodným testem s kritickým regionem $C_{\mathcal{T}}$.

Při testování hypotéz se rozhodujeme mezi tzv. nulovou hypotézou \mathcal{H}_0 a alternativní hypotézou \mathcal{H}_1 . Logika rozhodování je taková, že je-li \mathcal{H}_0 zamítnuta, pak je přijata \mathcal{H}_1 a naopak. Při tomto postupu se můžeme dopustit dvou druhů chyb.

Definice 9.5 (Typy chyb při testování hypotéz) *Chyba, které se dopustíme, jestliže zamítneme \mathcal{H}_0 , která je ve skutečnosti pravdivá, je tzv. chybou typu I. Naopak přijetím \mathcal{H}_0 , která je však ve skutečnosti nepravdivá, se dopouštíme chyby typu II. Velikosti chyb typu I a II jsou definovány pravděpodobnostmi, že se těchto chyb dopustíme.*

Jestliže je pravděpodobnostní rozdělení, ze kterého provádíme náhodný výběr, funkcí parametru θ , kde $\theta \in \overline{\Theta}$, pak lze každému testu přiřadit tzv. funkci charakteristické veličiny.

Definice 9.6 (Funkce charakteristické veličiny) *Uvažujme test \mathcal{T} nulové hypotézy \mathcal{H}_0 . Funkce charakteristické veličiny testu \mathcal{T} je označována $\pi_{\mathcal{T}}(\theta)$ a je definována jako pravděpodobnost zamítnutí \mathcal{H}_0 , jestliže pravděpodobnostní rozdělení, ze kterého byl prováděn náhodný výběr, je funkcí parametru θ .*

Funkce charakteristické veličiny plní stejnou roli jako MSE v případě bodového odhadu - v následujícím textu bude představovat standardní nástroj pro porovnání daného testu s alternativními testy. Z definice je zřejmé, že ideální funkce charakteristické veličiny nabývá nulové hodnoty pro θ , která odpovídají nulové hypotéze, a jednotkové hodnoty pro θ , která odpovídají alternativní hypotéze.

Platí, že $\pi_{\mathcal{T}}(\theta) = P_{\mathcal{T}}[\text{zamítnutí } \mathcal{H}_0]$, kde θ představuje skutečnou hodnotu parametru. Jestliže je \mathcal{T} nenáhodným testem, pak $\pi_{\mathcal{T}}(\theta) = P_{\theta}[(X_1, \dots, X_n) \in C_{\mathcal{T}}]$, kde $C_{\mathcal{T}}$ je kritickým regionem testu \mathcal{T} . Jestliže je \mathcal{T} náhodným testem s kritickou funkcí $\psi_{\mathcal{T}}(\cdot, \dots, \cdot)$, pak

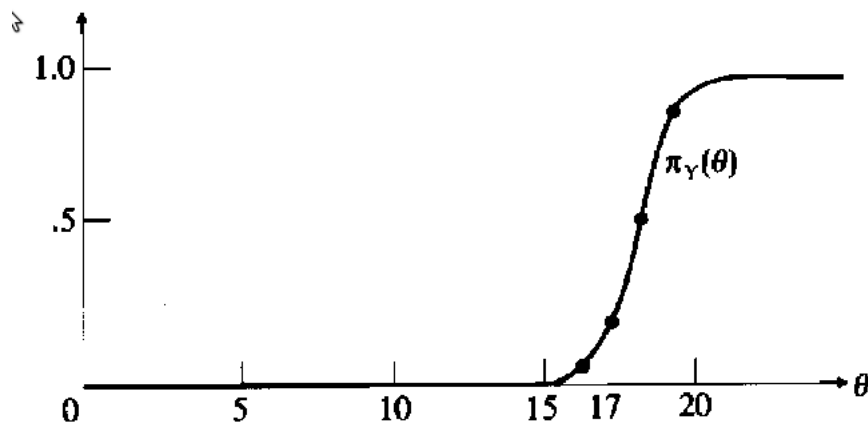
$$\begin{aligned} \pi_{\mathcal{T}} &= P_{\mathcal{T}}[\text{zamítnutí } \mathcal{H}_0] \\ &= \int \cdots \int P[\text{zamítnutí } \mathcal{H}_0 | x_1, \dots, x_n] f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) \prod_{i=1}^n dx_i \\ &= \int \cdots \int \psi_{\mathcal{T}}(x_1, \dots, x_n) f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) \prod_{i=1}^n dx_i = E_{\theta}[\psi_{\mathcal{T}}(X_1, \dots, X_n)] \end{aligned}$$

Analogické odvození lze provést také v případě nespojitě náhodné veličiny.

Příklad 9.5 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \phi_{\theta, 25}(x)$. Uvažujme nulovou hypotézu $\mathcal{H}_0 : \theta \leq 17$ a test \mathcal{T} , který zavrhuje \mathcal{H}_0 , jestliže $\bar{X} > 17 + 5/\sqrt{n}$. Funkce charakterické veličiny je dána rovnicí

$$\pi_{\mathcal{T}}(\theta) = P_{\theta} \left[\bar{X} > 17 + \frac{5}{\sqrt{n}} \right] = 1 - \Phi \left(\frac{17 + 5/\sqrt{n} - \theta}{5/\sqrt{n}} \right)$$

Pro $n = 25$ je $\pi_{\mathcal{T}}(\theta)$ ilustrována obrázkem (9.2). Funkce charakteristiké veličiny indikuje, nakolik je daný test spolehlivý. Jestliže je v našem konkrétním případě θ větší než 20, pak test \mathcal{T} správně zamítne \mathcal{H}_0 . Naopak, jestliže je θ menší než 16, pak test \mathcal{T} , opět správně, s vysokou pravděpodobností \mathcal{H}_0 nezamítne. Jestliže je však $17 < \theta < 18$, je pravděpodobnost zamítnutí \mathcal{H}_0 na základě testu \mathcal{T} méně než poloviční, ačkoliv je hypotéza \mathcal{H}_0 nepravdivá.



Obrázek 9.1: Funkce charakteristické veličiny

Definice 9.7 (Velikost testu) Uvažujme test \mathcal{T} hypotézy $\mathcal{H}_0 : \theta \in \bar{\Theta}_0$, kde $\bar{\Theta}_0 \subset \bar{\Theta}$. Velikost testu \mathcal{T} je definována jako $\sup_{\theta \in \bar{\Theta}_0} [\pi_{\mathcal{T}}(\theta)]$. Velikost nenáhodného testu je někdy také nazývána velikostí kritického regionu.

Příklad 9.6 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \phi_{\theta, 25}(x)$. Dále uvažujme hypotézu $\mathcal{H}_0 : \theta \leq 17$ a test \mathcal{T} , který zavrhuje \mathcal{H}_0 jestliže $\bar{X} > 17 + 5/\sqrt{n}$. Platí $\bar{\Theta}_0 = \{\theta : \theta \leq 17\}$ a velikost testu \mathcal{T} je

$$\sup_{\theta \in \bar{\Theta}_0} [\pi_{\mathcal{T}}(\theta)] = \sup_{\theta \leq 17} \left[1 - \Phi \left(\frac{17 + 5/\sqrt{n} - \theta}{5/\sqrt{n}} \right) \right] = 1 - \Phi(1) \approx 0.159$$

Věta 9.1 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta)$, kde $\theta \in \bar{\Theta}$ a soubor dostatečných statistik $S_1 = \mathbf{s}_1(X_1, \dots, X_n), \dots, S_r = \mathbf{s}_r(X_1, \dots, X_n)$. Pak pro libovolný test \mathcal{T} s kritickou funkcí $\psi_{\mathcal{T}}(X_1, \dots, X_n)$ existuje test \mathcal{T}' s kritickou funkcí $\psi_{\mathcal{T}'}(S_1, \dots, S_r)$, která splňuje $\pi_{\mathcal{T}}(\theta) = \pi_{\mathcal{T}'}(\theta)$ pro všechna $\theta \in \bar{\Theta}$.

Důkaz 9.1 Definujme $\psi_{\mathcal{T}'}(s_1, \dots, s_r) = E[\psi_{\mathcal{T}}(X_1, \dots, X_n) | S_1 = s_1, \dots, S_r = s_r]$, pak $\psi_{\mathcal{T}'}$ je kritickou funkcí. Dále platí $\pi_{\mathcal{T}'}(\theta) = E_{\theta}[\psi_{\mathcal{T}'}(S_1, \dots, S_r)] = E_{\theta}[E[\psi_{\mathcal{T}}(X_1, \dots, X_n) | S_1, \dots, S_r]] = E_{\theta}[\psi_{\mathcal{T}}(X_1, \dots, X_n)] = \pi_{\mathcal{T}}(\theta)$.

Z výše uvedené věty vyplývá, že pro libovolný test existuje alternativní test založený pouze na souboru dostatečných statistik a že tento test má stejnou funkci charakteristické veličiny jako test původní. Při našem pátrání po vhodném testu se tak můžeme omezit pouze na testy založených právě na souboru dostatečných statistik.

9.2 Testování jednoduchých hypotéz

Definice 9.8 (Jednoduchý věrohodnostní poměrový test) *Uvažujme náhodný výběr X_1, \dots, X_n , o kterém víme, že pochází buď z populace $f_0(\cdot)$ nebo $f_1(\cdot)$. Test \mathcal{T} , který testuje $\mathcal{H}_0 : X_i \sim f_0(\cdot)$ proti $\mathcal{H}_1 : X_i \sim f_1(\cdot)$, je nazýván jednoduchým věrohodnostním poměrovým testem, jestliže je definován jako*

1. zamítne \mathcal{H}_0 , jestliže $\lambda < k$,
2. přijme \mathcal{H}_0 , jestliže $\lambda > k$,
3. přijme / zamítne \mathcal{H}_0 popř. rozhodne na základě náhodného testu, jestliže $\lambda = k$

kde

$$\lambda = \lambda(x_1, \dots, x_n) = \frac{\prod_{i=1}^n f_0(x_i)}{\prod_{i=1}^n f_1(x_i)} = \frac{L_0(x_1, \dots, x_n)}{L_1(x_1, \dots, x_n)} = \frac{L_0}{L_1}$$

a k je nezáporná konstanta. Připomeňme, že $L_j = L_j(x_1, \dots, x_n)$ je funkce maximální věrohodnosti pro náhodný výběr z populace $f_j(\cdot)$.

Pro fixní k tedy test zavrhuje \mathcal{H}_0 pro nízké hodnoty λ a naopak. Výsledek testu tedy závisí na zvolené hodnotě k .

Jednoduchý věrohodnostní poměrový test je v souladu s intuicí. Navíc lze očekávat, že “optimální” test bude mít podobu právě jednoduchého věrohodnostního poměrového testu. K hledání “optimálního” testu pro jednoduché hypotézy lze přistoupit dvěma způsoby - prvním z nich je koncept tzv. nejsilnějšího testu a druhý spočívá ve využití ztrátové funkce.

9.2.1 Nejsilnější test

Uvažujme náhodný výběr X_1, \dots, X_n , o kterém víme, že pochází buď z populace $f_0(\cdot)$ nebo $f_1(\cdot)$, kde $f_0(x) = f(x, \theta_0)$ a $f_1(x) = f(x, \theta_1)$ a θ_0 a θ_1 jsou známé konstanty. To znamená, že $\Theta = \{\theta_0, \theta_1\}$ je prostor, který se skládá z pouhých dvou bodů. Testujme hypotézu $\mathcal{H}_0 : \theta = \theta_1$ proti hypotéze $\mathcal{H}_1 : \theta = \theta_0$. Tomuto testu lze přiřadit funkci charakteristické veličiny $\pi_{\mathcal{T}}(\theta)$. Dobrý test je takový, pro který je $\pi_{\mathcal{T}}(\theta_0) = P[\text{zamítnutí } \mathcal{H}_0 | \mathcal{H}_0 \text{ je pravdivá}]$ malé (ideálně rovno nule) a $\pi_{\mathcal{T}}(\theta_1) = P[\text{zamítnutí } \mathcal{H}_0 | \mathcal{H}_0 \text{ je nepravdivá}]$ vysoké (ideálně rovno jedné). $\pi_{\mathcal{T}}(\theta_0)$ odpovídá chybě typu I a $1 - \pi_{\mathcal{T}}(\theta_1) = P[\text{přijetí } \mathcal{H}_0 | \mathcal{H}_0 \text{ je nepravdivá}]$ odpovídá chybě typu II. Naším cílem je minimalizovat obě chyby. Možných postupů je několik.

Definice 9.9 (Nejsilnější test) *Test \mathcal{T}^* , který testuje $\mathcal{H}_0 : \theta = \theta_0$ proti $\mathcal{H}_1 : \theta = \theta_1$ je nejsilnějším testem velikosti α , kde $0 < \alpha < 1$, jestliže*

1. $\pi_{\mathcal{T}^*}(\theta_0) = \alpha$

2. $\pi_{\mathcal{T}^*}(\theta_1) \geq \pi_{\mathcal{T}}(\theta_1)$ pro libovolný alternativní test \mathcal{T} , pro který platí $\pi_{\mathcal{T}}(\theta_0) \leq \alpha$.

Důvod, proč fixujeme velikost chyby typu I na úroveň α^2 zatímco velikost chyby typu II nikoliv, je skutečnost, že nulová a alternativní hypotéza jsou zpravidla formulovány tak, že realizace chyby typu I má závažnější následky než realizace chyby typu II.

Věta 9.2 (Neyman-Pearsonova) Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta)$, kde θ nabývá hodnoty θ_0 nebo θ_1 . Předpokládejme, že α , kde $0 < \alpha < 1$, je fixní a známé. Nechť k^* je pozitivní konstanta a C^* je podmnožina \mathfrak{X} , která splňuje

1. $P_{\theta_0}[(X_1, \dots, X_n) \in C^*] = \alpha$
2. $\lambda = \frac{L(\theta_0; x_1, \dots, x_n)}{\theta_1; x_1, \dots, x_n} = \frac{L_0}{L_1} \leq k^*$ pro $(x_1, \dots, x_n) \in C^*$
3. $\lambda \geq k^*$ pro $(x_1, \dots, x_n) \in \overline{C^*}$

Pak je test \mathcal{T}^* , který je charakterizován kritickým regionem C^* , nejsilnějším testem velikosti α pro testování $\mathcal{H}_0: \theta = \theta_0$ proti $\mathcal{H}_1: \theta = \theta_1$. Připomeňme, že $L_j = L(\theta_j; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i, \theta_j)$ pro $j = 0, 1$. $\overline{C^*}$ je doplněk k C^* , tj. $\overline{C^*} = \mathfrak{X} - C^*$

Důkaz 9.2 Předpokládejme, že existují taková k^* a C^* , která splňují podmínky (1) až (3) výše uvedené věty. Uvažujme alternativní test \mathcal{T} se silou alespoň α a kritickým regionem C , tj. $P_{\theta_0}[(X_1, \dots, X_n) \in C] \leq \alpha$.

V následujícím textu budeme pro libovolné $R \subset \mathfrak{X}$ namísto $\int \dots \int_R [\prod_{i=1}^n f(x_i, \theta_j) dx_i]$ používat zkrácený zápis $\int_R L_j$, kde $j = 0, 1$. Ze zápisu vyplývá, že $f_0(\cdot)$ a $f_1(\cdot)$ jsou pravděpodobnostní funkce spojitě náhodné veličiny. Stejný důkaz nicméně platí také pro nespojitě náhodné veličiny.

Musíme dokázat $\pi_{\mathcal{T}^*}(\theta_1) \geq \pi_{\mathcal{T}}(\theta_1)$, což je ekvivalentní důkazu $\int_{C^*} L_1 \geq \int_C L_1$. Následující postup ilustruje obrázek (??). Platí

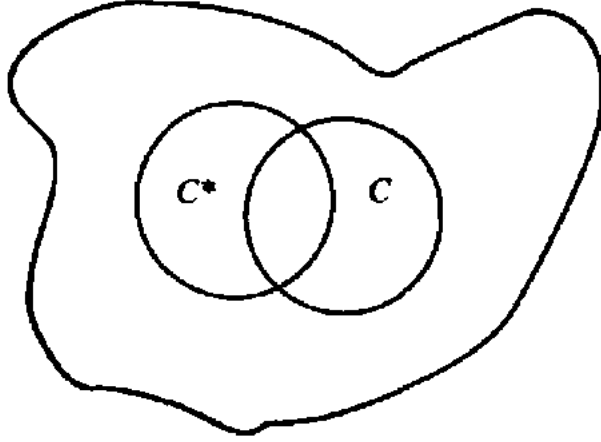
$$\int_{C^*} L_1 - \int_C L_1 = \int_{C^* \setminus C} L_1 - \int_{C \setminus C^*} L_1 \leq \frac{1}{k^*} \int_{C^* \setminus C} L_0 - \frac{1}{k^*} \int_{C \setminus C^*} L_0$$

protože $L_1 \geq \frac{L_0}{k^*}$ na množině C^* (a tím pádem také na množině $C^* \setminus C$) a $L_1 \leq \frac{L_0}{k^*}$ neboli $-L_1 \geq -\frac{L_0}{k^*}$ na množině $\overline{C^*}$ (a tím pádem také na množině $C \setminus C^*$). Nicméně

$$\begin{aligned} \frac{1}{k^*} \left(\int_{C^* \setminus C} L_0 - \int_{C \setminus C^*} L_0 \right) &= \frac{1}{k^*} \left(\int_{C^* \setminus C} L_0 - \int_{C^*} L_0 + \int_{C^*} L_0 - \int_{C \setminus C^*} L_0 \right) \\ &= \frac{1}{k^*} \left(\int_{C^*} L_0 - \int_C L_0 \right) = \frac{1}{k^*} (\alpha - \text{velikost testu } \mathcal{T}) \end{aligned}$$

což implikuje $\int_{C^*} L_1 - \int_C L_1 \geq 0$.

² α je nastaveno zpravidla na nízkou úroveň - obvyklé hodnoty jsou 0.01 a 0.05.



Obrázek 9.2: Ilustrační obrázek

Obecně platí, že ne vždy musí existovat k^* a C^* , která splňují podmínky Neyman-Pearsovy věty. V takovém případě Neyman-Pearsova věta není schopná dát nejsilnější test síly α . Nicméně jestliže jsou $f_0(\cdot)$ a $f_1(\cdot)$ pravděpodobnostními funkcemi spojitě náhodné veličiny, pak odpovídající k^* a C^* vždy existují. Ačkoliv Neyman-Pearsova věta explicitně neříká, jak nalézt k^* a C^* , jsou k^* a C^* dány implicitně její druhou podmínkou.

Příklad 9.7 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \theta e^{-\theta x} I_{0, \infty}(x)$, kde $\theta = \theta_0$ popř. $\theta = \theta_1$. θ_0 a θ_1 jsou fixní a známé. Předpokládejme $\theta_0 < \theta_1$. Definujme test \mathcal{T} , který testuje hypotézu $\mathcal{H}_0 : \theta = \theta_0$ proti hypotéze $\mathcal{H}_1 : \theta = \theta_1$.

Potřebné funkce maximální věrohodnosti mají tvar $L_0 = \theta_0^n e^{-\theta_0 \sum x_i}$ a $L_1 = \theta_1^n e^{-\theta_1 \sum x_i}$. Dle Neyman-Pearsonovy věty zamítne nejsilnější test \mathcal{H} , jestliže $\lambda \leq k^*$, což lze v kontextu našeho příkladu vyjádřit jako

$$\left(\frac{\theta_0}{\theta_1}\right)^n e^{-(\theta_0 - \theta_1) \sum_{i=1}^n x_i} \leq k^*$$

$$\sum_{i=1}^n x_i \frac{1}{\theta_1 - \theta_0} \ln \left[\left(\frac{\theta_1}{\theta_0}\right)^n k^* \right] = k'$$

kde k' je opět konstanta. Nerovnost $\lambda \leq k^*$ tak byla zjednodušena a vyjádřena ve tvaru $\sum x_i \leq k'$. První podmínka má tvar

$$\alpha = P_{\theta_0}[\text{zamítnutí } \mathcal{H}_0] = P_{\theta_0} \left[\sum_{i=1}^n X_i \leq k' \right]$$

Víme, že $\sum x_i$ sleduje gamma rozdělení s parametry n a θ , a proto lze k' vypočítat z rovnice

$$P_{\theta_0} \left[\sum_{i=1}^n X_i \leq k' \right] = \int_0^{k'} \frac{1}{\Gamma(n)} \theta_0^n x^{n-1} e^{-x\theta_0} dx = \alpha$$

Nejsilnější test \mathcal{T} velikosti α pro $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_1$, kde $\theta_0 < \theta_1$ zamítne \mathcal{H}_0 jestliže $\sum x_i \leq k'$, kde k' je α -tý kvantil gamma rozdělení s parametry n a θ_0 .

Příklad 9.8 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \theta^x(1 - \theta)^{1-x}I_{\{0,1\}}(x)$, kde $\theta = \theta_0$ popř. $\theta = \theta_1$ a $\theta_0 < \theta_1$. Potřebné funkce maximální věrohodnosti mají tvar $L_0 = \theta_0^{\sum_{i=1}^n x_i} (1 - \theta_0)^{n - \sum_{i=1}^n x_i}$ a $L_1 = \theta_1^{\sum_{i=1}^n x_i} (1 - \theta_1)^{n - \sum_{i=1}^n x_i}$. Nerovnost $\lambda < k^*$ je tedy splněna, jestliže

$$\frac{\theta_0^{\sum_{i=1}^n x_i} (1 - \theta_0)^{n - \sum_{i=1}^n x_i}}{\theta_1^{\sum_{i=1}^n x_i} (1 - \theta_1)^{n - \sum_{i=1}^n x_i}} \leq k^*$$

$$\left[\frac{\theta_0(1 - \theta_1)}{(1 - \theta_1)\theta_1} \right]^{\sum_{i=1}^n x_i} \left(\frac{1 - \theta_0}{1 - \theta_1} \right)^n \leq k^*$$

neboli, jestliže $\sum x_i \geq k'$, kde k' je konstanta³.

Předpokládejme, že $\theta_0 = 1/4$, $\theta_1 = 3/4$ a $n = 10$. Musíme nalézt k' takové, že

$$\alpha = P_{\theta_0=1/4}[\text{zamítnutí } \mathcal{H}_0] = P_{\theta_0=1/4} \left[\sum_{i=1}^n X_i \geq k' \right] = \sum_{i=k'}^{10} \binom{10}{i} \left(\frac{1}{4} \right)^i \left(\frac{3}{4} \right)^{10-i}$$

Jestliže $\alpha = 0.0197$, pak $k' = 6$. Jestliže $\alpha = 0.0781$, pak $k' = 5$. Pro $\alpha = 0.05$ však neexistuje kritický region C^* a konstanta k^* dle Neyman-Pearsonovy věty. V tomto příkladě je totiž námi uvažovaná náhodná veličina diskrétní a pro diskrétní náhodné veličiny není vždy možné nalézt C^* a k^* pro libovolné α . Tento problém je možné obejít např. změnou velikosti náhodného výběru. Nicméně nejsilnější test velikosti α existuje vždy, pouze se může jednat o náhodný test. V našem konkrétním případě lze pro $\alpha = 0.05$ definovat tento test jako

$$\psi(x_1, \dots, x_{10}) = \begin{cases} 1 & \text{pro } \sum_{i=1}^n x_i \geq 6 \\ \frac{0.05 - 0.0197}{0.0584} & \text{pro } \sum_{i=1}^n x_i = 5 \\ 0 & \text{pro } \sum_{i=1}^n x_i \leq 4 \end{cases}$$

9.2.2 Ztrátová funkce

Stejně jako v předchozím textu uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta)$, kde $\theta = \theta_0$ nebo $\theta = \theta_1$ a testujme $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_1$. Můžeme tedy učinit dvě rozhodnutí - rozhodnutí d_0 , že náhodný výběr pochází z populace $f(x, \theta_0)$ a rozhodnutí d_1 , že náhodný výběr pochází z populace $f(x, \theta_1)$. Předpokládejme, že máme k dispozici ztrátovou funkci.

Definice 9.10 (Ztrátová funkce) Pro účely testování $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_1$ definujme $l(d_i, \theta_j) =$ ztráta z rozhodnutí d_i jestliže správná hodnota parametru je θ_j . Přijmeme konvenci $l(d_i, \theta_i) = 0$ a $l(d_i, \theta_j) > 0$ pro $i \neq j$.

³Připomeňme, že $\ln \left(\frac{\theta_0(1-\theta_1)}{(1-\theta_0)\theta_1} \right) < 0$.

V kontextu nenáhodného testu \mathcal{T} s kritickým regionem $C_{\mathcal{T}}$ je učiněno rozhodnutí d_1 , jestliže (x_1, \dots, x_n) náleží do $C_{\mathcal{T}}$ a rozhodnutí d_0 , jestliže (x_1, \dots, x_n) náleží do $\bar{C}_{\mathcal{T}}$. Jestliže je přijato rozhodnutí d_1 , avšak správná hodnota parametru je θ_0 , dopouštíme se chyby typu I. Podobně, jestliže je přijato rozhodnutí d_0 , avšak správná hodnota parametru je θ_1 , dopouštíme se chyby typu II.

Při výběru testu pochopitelně preferujeme ten s nejmenší “ztrátou”. Bohužel v praxi jen zřídka existuje jen jeden test, který má nejmenší “ztrátu” pro obě možná rozhodnutí a všechna možná θ_0 a θ_1 . Tímto se dostáváme k tzv. rizikové funkci.

Definice 9.11 (Riziková funkce) Uvažujme náhodný výběr X_1, \dots, X_n , o kterém víme, že pochází s populace $f(\cdot, \theta_0)$ nebo $f(\cdot, \theta_1)$. Nechť test \mathcal{T} s kritickým regionem $C_{\mathcal{T}}$ testuje $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_1$. Uvažujme danou ztrátovou funkci $l(\cdot, \cdot)$. Riziková funkce $\mathcal{r}_{\mathcal{T}}(\theta)$ testu \mathcal{T} je definována jako očekávaná ztráta, tj.

$$\mathcal{r}_{\mathcal{T}}(\theta) = \int \cdots \int_{C_{\mathcal{T}}} l(d_1, \theta) \left[\prod_{i=1}^n f(x_i, \theta) dx_i \right] + \int \cdots \int_{\bar{C}_{\mathcal{T}}} l(d_0, \theta) \left[\prod_{i=1}^n f(x_i, \theta) dx_i \right]$$

Všimněme si, že riziková funkce je lineární funkcí funkce charakteristické veličiny, kde koeficienty v této lineární funkci jsou dány hodnotou ztrátové funkce.

$$\begin{aligned} \mathcal{r}_{\mathcal{T}}(\theta) &= l(d_1, \theta) P_{\theta}[(X_1, \dots, X_n) \in C_{\mathcal{T}}] + l(d_0, \theta) P_{\theta}[(X_1, \dots, X_n) \in \bar{C}_{\mathcal{T}}] \\ &= l(d_1, \theta) \pi_{\mathcal{T}}(\theta) + l(d_0, \theta) [1 - \pi_{\mathcal{T}}(\theta)] \end{aligned}$$

Vzhledem k tomu, že parametr θ může nabývat pouze dvou hodnot, může pouze dvou hodnot nabývat také riziková funkce. Konkrétně se jedná o hodnoty

$$\begin{aligned} \mathcal{r}_{\mathcal{T}}(\theta_0) &= l(d_1, \theta_0) \pi_{\mathcal{T}}(\theta_0) \\ \mathcal{r}_{\mathcal{T}}(\theta_1) &= l(d_0, \theta_1) [1 - \pi_{\mathcal{T}}(\theta_1)] \end{aligned}$$

Ideální test je charakterizován nejmenším “rizikem” v porovnání se všemi možnými alternativními testy. Problém však spočívá v tom, že “riziko” nabývá dvou hodnot a test, který by v porovnání se všemi možnými alternativními testy dosahoval minima pro obě tyto hodnoty, existuje pouze ve vyjimečných případech. Tímto se dostáváme ke konceptu minimax testu.

Definice 9.12 (Minimax test) Test \mathcal{T}_m pro $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_1$ je minimaxem, jestliže

$$\max[\mathcal{r}_{\mathcal{T}_m}(\theta_0), \mathcal{r}_{\mathcal{T}_m}(\theta_1)] \leq \max[\mathcal{r}_{\mathcal{T}}(\theta_0), \mathcal{r}_{\mathcal{T}}(\theta_1)]$$

pro libovolný alternativní test.

Následující větu je někdy možné použít při hledání minimax testu.

Věta 9.3 Uvažujme náhodný výběr X_1, \dots, X_n , o kterém víme, že pochází s populace $f(\cdot, \theta_0)$ nebo $f(\cdot, \theta_1)$. Dále uvažujme test \mathcal{T}_m , který testuje $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_1$. Jestliže kritický region testu \mathcal{T}_m je definován jako $C_m = \{(x_1, \dots, x_n) : \lambda \leq k_m\}$, kde k_m je kladná konstanta taková, že $\mathcal{r}_{\mathcal{T}_m}(\theta_0) = \mathcal{r}_{\mathcal{T}_m}(\theta_1)$, pak je \mathcal{T}_m minimaxem. Připomeňme, že

$$\lambda = \frac{L_0}{L_1} = \frac{\prod_{i=1}^n f(x_i, \theta_0)}{\prod_{i=1}^n f(x_i, \theta_1)}$$

Důkaz 9.3 Předpokládejme, že $f(\cdot, \theta_0)$ a $f(\cdot, \theta_1)$ jsou pravděpodobnostní funkce spojitých náhodných veličin. Důkaz pro nespojité náhodné veličiny by byl analogický.

Uvažujme alternativní test \mathcal{T} s kritickým regionem $C_{\mathcal{T}}$, který splňuje podmínku $\mathfrak{r}_{\mathcal{T}}(\theta_0) \leq \mathfrak{r}_{\mathcal{T}_m}(\theta_0)^4$, což implikuje

$$\begin{aligned} \mathfrak{l}(d_1, \theta_0) \pi_{\mathcal{T}}(\theta_0) &\leq \mathfrak{l}(d_1, \theta_0) \pi_{\mathcal{T}_m}(\theta_0) \\ \pi_{\mathcal{T}}(\theta_0) &\leq \pi_{\mathcal{T}_m}(\theta_0) \end{aligned}$$

tj. \mathcal{T} má velikost menší nebo rovnou velikosti testu \mathcal{T}_m . Nicméně na základě Neyman-Pearsonovy věty je \mathcal{T}_m nejsilnějším testem velikosti $\pi_{\mathcal{T}_m}(\theta_0)$. Proto platí

$$\begin{aligned} \pi_{\mathcal{T}}(\theta_1) &\leq \pi_{\mathcal{T}_m}(\theta_1) \\ 1 - \pi_{\mathcal{T}}(\theta_1) &\geq 1 - \pi_{\mathcal{T}_m}(\theta_1) \\ \mathfrak{l}(d_0, \theta_1)[1 - \pi_{\mathcal{T}}(\theta_1)] &\geq \mathfrak{l}(d_0, \theta_1)[1 - \pi_{\mathcal{T}_m}(\theta_1)] \\ \mathfrak{r}_{\mathcal{T}}(\theta_1) &\geq \mathfrak{r}_{\mathcal{T}_m}(\theta_1) \end{aligned}$$

a proto

$$\max[\mathfrak{r}_{\mathcal{T}_m}(\theta_0), \mathfrak{r}_{\mathcal{T}_m}(\theta_1)] = \mathfrak{r}_{\mathcal{T}_m}(\theta_1) \leq \mathfrak{r}_{\mathcal{T}}(\theta_1) \leq \max[\mathfrak{r}_{\mathcal{T}}(\theta_0), \mathfrak{r}_{\mathcal{T}}(\theta_1)]$$

tj. \mathcal{T}_m je minimaxem.

Příklad 9.9 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \theta e^{-\theta x} I_{(0, \infty)}(x)$. Pro $\theta_1 > \theta_0$ testujme $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_1$. V příkladu (9.7) jsme našli nejsilnější test velikosti α . Pokusme se nyní nalézt minimax test pro ztrátovou funkci definovanou jako $\mathfrak{d}_1, \theta_0 = a$ a $\mathfrak{d}_0, \theta_1 = b$.

Dle výše uvedené věty je minimax test \mathfrak{T}_m dán $C_m = \{(x_1, \dots, x_n) : \lambda \leq k_m\}$, kde k_m definováno rovností $\mathfrak{r}_{\mathcal{T}_m}(\theta_0) = \mathfrak{r}_{\mathcal{T}_m}(\theta_1)$. Nerovnost $\lambda \leq k_m$ lze vyjádřit jako $\sum x_i \leq k$, kde k je konstanta. Rovnost $\mathfrak{r}_{\mathcal{T}_m}(\theta_0) = \mathfrak{r}_{\mathcal{T}_m}(\theta_1)$ je splněna, jestliže $a\pi_{\mathcal{T}_m}(\theta_0) = b[1 - \pi_{\mathcal{T}_m}(\theta_1)]$, tj. hledáme takové k , které splňuje

$$\begin{aligned} aP_{\theta_0}[\sum_{i=1}^n X_i \leq k] &= bP_{\theta_1}[\sum_{i=1}^n X_i > k] \\ a \int_0^k \frac{1}{\Gamma(n)} \theta_0^n x^{n-1} e^{-\theta_0 x} dx &= b \int_k^\infty \frac{1}{\Gamma(n)} \theta_1^n x^{n-1} e^{-\theta_1 x} dx \end{aligned}$$

Jestliže jsou $f_0(\cdot)$ a $f_1(\cdot)$ pravděpodobnostní funkce diskrétní náhodné veličiny, pak se může stát, že neexistuje takové k_m , pro které $\mathfrak{r}_{\mathcal{T}_m}(\theta_0) = \mathfrak{r}_{\mathcal{T}_m}(\theta_1)$, pokud nepřistoupíme na náhodné testy. Z věty (9.3) je zřejmé, že minimax je jednoduchým věrohodnostním poměrovým testem.

V předchozím textu jsme předpokládali, že funkční předpis $f(\cdot, \theta)$ je znám a že parametr θ může nabývat pouze dvou hodnot a to θ_0 a θ_1 . Dále jsme předpokládali, že známe příslušnou ztrátovou funkci. Nyní však předpokládejme, že θ_0 a θ_1 jsou hodnoty náhodné veličiny Θ , jejíž apriorní pravděpodobnostní funkci známe. Jedná se tedy o stejný koncept jako v případě Bayesovského odhadu. Náhodná veličina Θ je tedy nespojitá a nabývá pouze dvou hodnot a to θ_0 a θ_1 . Apriorní pravděpodobnostní funkce Θ má tedy tvar $g = P[\Theta = \theta_1] = P[\Theta = \theta_0]$.

⁴Pokud by platilo $\mathfrak{r}_{\mathcal{T}}(\theta_0) > \mathfrak{r}_{\mathcal{T}_m}(\theta_0)$, pak by test \mathcal{T} nemohl být ani kandidátem na minimax.

Definice 9.13 (Bayesovský test) Test \mathcal{T}_g , který testuje $\mathcal{H} : \theta = \theta_0$ proti $\mathcal{H} : \theta = \theta_1$ je Bayesovským testem s apriorní pravěpodobnostní funkcí

$$g = P[\Theta = \theta_1]$$

právě tehdy a jen tehdy je-li splněna nerovnost

$$(1 - g)\mathfrak{r}_{\mathcal{T}_g}(\theta_0) + g\mathfrak{r}_{\mathcal{T}_g}(\theta_1) \leq (1 - g)\mathfrak{r}_{\mathcal{T}}(\theta_0) + g\mathfrak{r}_{\mathcal{T}}(\theta_1)$$

pro libovolný alternativní test \mathcal{T} . Abychom našli Bayesovský test, hledáme kritický region C_g , který minimalizuje

$$\begin{aligned} (1 - g)\mathfrak{r}_{\mathcal{T}}(\theta_0) + g\mathfrak{r}_{\mathcal{T}}(\theta_1) &= (1 - g)\mathfrak{l}(d_1, \theta_0)\pi_{\mathcal{T}}(\theta_0) + g\mathfrak{l}(d_0, \theta_1)[1 - \pi_{\mathcal{T}}(\theta_1)] \\ &= (1 - g)\mathfrak{l}(d_1, \theta_0) \int_C L_0 + g\mathfrak{l}(d_0, \theta_1) \int_{\bar{C}} L_1 \end{aligned}$$

jakožto funkci regionu C . Zároveň však platí, že

$$\begin{aligned} (1 - g)\mathfrak{l}(d_1, \theta_0) \int_C L_0 + g\mathfrak{l}(d_0, \theta_1) \int_{\bar{C}} L_1 \\ = g\mathfrak{l}(d_0, \theta_1) + \int_C [(1 - g)\mathfrak{l}(d_1, \theta_0)L_0 - g\mathfrak{l}(d_0, \theta_1)L_1] \end{aligned}$$

je minimální pro region C , který obsahuje všechna (x_1, \dots, x_n) , pro která je poslední integrant záporný, tj.

$$C_g = \{(x_1, \dots, x_n) : (1 - g)\mathfrak{l}(d_1, \theta_0)L_0 - g\mathfrak{l}(d_0, \theta_1)L_1 < 0\}$$

Tímto jsme zároveň dokázali následující větu.

Věta 9.4 Bayesovský test \mathcal{T}_g , který testuje $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_1$ s ohledem na apriorní rozdělení $g = P[\Theta = \theta_1]$ má kritický region definovaný jako

$$C_g = \left\{ (x_1, \dots, x_n) : \lambda < \frac{g\mathfrak{l}(d_0, \theta_1)}{(1 - g)\mathfrak{l}(d_1, \theta_0)} \right\}$$

To znamená, že Bayesovský test má opět charakter jednoduchého věrohodnostního poměrového testu.

Příklad 9.10 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \theta e^{-\theta x} I_{(0, \infty)}(x)$. Testujme $\mathcal{H}_0 : \theta = \theta_0$ proti $\mathcal{H}_1 : \theta = \theta_1$. Kritický region Bayesovského testu je dán

$$\begin{aligned} C_g &= \left\{ (x_1, \dots, x_n) : \frac{\theta_0^n e^{-\theta_0 \sum_{i=1}^n x_i}}{\theta_1^n e^{-\theta_1 \sum_{i=1}^n x_i}} < \frac{g\mathfrak{l}(d_0, \theta_1)}{(1 - g)\mathfrak{l}(d_1, \theta_0)} \right\} \\ &= \left\{ (x_1, \dots, x_n) : \sum_{i=1}^n x_i < \frac{1}{\theta_1 - \theta_0} \ln \left(\frac{g\theta_1^n \mathfrak{l}(d_0, \theta_1)}{(1 - g)\theta_0^n \mathfrak{l}(d_1, \theta_0)} \right) \right\} \end{aligned}$$

pro $\theta_1 > \theta_0$.

9.3 Testování složených hypotéz

Předpokládejme náhodný výběr z populace $f(x, \theta)$, kde $\theta \in \bar{\Theta}$. Dále předpokládejme, že chceme testovat $\mathcal{H}_0 : \theta \in \bar{\Theta}_0$ vs. $\mathcal{H}_1 : \theta \in \bar{\Theta}_1$, kde $\bar{\Theta}_0 \subset \bar{\Theta}$, $\bar{\Theta}_1 \subset \bar{\Theta}$ a $\bar{\Theta}_0$ a $\bar{\Theta}_1$ jsou disjunktní množiny. Obvykle platí $\bar{\Theta}_1 = \bar{\Theta} - \bar{\Theta}_0$.

9.3.1 Obecný věrohodnostní poměrový test

Pro náhodný výběr X_1, \dots, X_n z populace $f(x, \theta)$, kde $\theta \in \bar{\Theta}$ hledáme test $\mathcal{H}_0 : \theta \in \bar{\Theta}_0$ vs. $\mathcal{H}_1 : \theta \in \bar{\Theta}_1 = \bar{\Theta} - \bar{\Theta}_0$.

Definice 9.14 (Obecný věrohodnostní poměr) *Uvažujme funkci maximální věrohodnosti $L(\theta; x_1, \dots, x_n)$ náhodného výběru X_1, \dots, X_n se sdruženou pravděpodobnostní funkcí $f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta)$, kde $\theta \in \bar{\Theta}$. Obecný věrohodnostní poměr λ_n je definován jako*

$$\lambda_n = \lambda(x_1, \dots, x_n) = \frac{\sup_{\theta \in \bar{\Theta}_0} L(\theta; x_1, \dots, x_n)}{\sup_{\theta \in \bar{\Theta}_1} L(\theta; x_1, \dots, x_n)}$$

Všimněme si, že λ_n je funkcí pozorování x_1, \dots, x_n , tj. $\lambda_n = \lambda(x_1, \dots, x_n)$. Jestliže tato pozorování nahradíme náhodnými veličinami X_1, \dots, X_n , pak se dle zvyklosti změni zápis na $\Lambda = \lambda(X_1, \dots, X_n)$. Λ je tak funkcí náhodných veličin X_1, \dots, X_n , a proto se jedná o náhodnou veličinu resp. o statistiku, protože nezávisí na neznámých parametrech.

V souvislosti s obecným věrohodnostním poměrem uvedme následujících pět poznámek.

1. Pro obecný věrohodnostní poměr používáme symbol λ , který jsme v předchozím textu používali pro jednoduchý věrohodnostní poměr. Avšak pro $\bar{\Theta} = \{\theta_0, \theta_1\}$ se obecný věrohodnostní poměr nezredukuje na jednoduchý věrohodnostní poměr, jak by bylo možné se domnívat.
2. Obecný věrohodnostní poměr λ splňuje podmínku $0 \leq \lambda \leq 1$. Nerovnost $\lambda \geq 0$ vyplývá ze skutečnosti, že λ je poměrem dvou nezáporných čísel. Nerovnost $\lambda \leq 1$ pak vyplývá ze skutečnosti, že supremum v čitateli je přes $\bar{\Theta}_0$ a supremum ve jmenovateli přes $\bar{\Theta}$, přičemž $\bar{\Theta}_0 \subset \bar{\Theta}$.
3. Parametr Θ může mít podobu vektoru.
4. Jmenovatel Λ je představován hodnotou funkce maximální věrohodnosti pro funkci odhadu založené na metodě maximální věrohodnosti.
5. V našich úvahách bude mít náhodný výběr X_1, \dots, X_n často podobu náhodného výběru z populace $f(x, \theta)$, kde $\theta \in \bar{\Theta}$.

Hodnoty λ statistiky Λ slouží k formulaci testu $\mathcal{H}_0 : \theta \in \bar{\Theta}_0$ vs. $\mathcal{H}_1 : \theta \in \bar{\Theta}_1 = \bar{\Theta} - \bar{\Theta}_0$, kde \mathcal{H}_0 je zamítnuta, jestliže $\lambda \leq \lambda_0$, kde λ_0 představuje konstantu, která splňuje podmínku $0 \leq \lambda_0 \leq 1$ a která je často odvozena od síly testu.

V některých případech výše popsaná metoda naráží na skutečnost, že může být poměrně komplikované nalézt $\sup L(\theta; x_1, \dots, x_n)$ nebo pravděpodobnostní rozdělení Λ , které je nezbytné pro stanovení síly testu.

Příklad 9.11 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \theta e^{-\theta x} I_{(0, \infty)}(x)$, kde $\underline{\Theta} = \{\theta; \theta > 0\}$. Testujme $\mathcal{H}_0: \theta \leq \theta_0$ vs. $\mathcal{H}_1: \theta > \theta_0$.

$$\sup_{\theta \in \underline{\Theta}} L(\theta; x_1, \dots, x_n) = \sup_{\theta > 0} \left(\theta^n e^{-\sum_{i=1}^n x_i} \right) = \left(\frac{n}{\sum_{i=1}^n x_i} \right)^n e^{-n}$$

a

$$\begin{aligned} \sup_{\theta \in \underline{\Theta}_0} L(\theta; x_1, \dots, x_n) &= \sup_{0 < \theta \leq \theta_0} (\theta^n e^{-\sum_{i=1}^n x_i}) \\ &= \begin{cases} \left(\frac{n}{\sum_{i=1}^n x_i} \right)^n e^{-n} & \text{pro } \frac{n}{\sum_{i=1}^n x_i} \leq \theta_0 \\ \theta_0^n e^{-\theta_0 \sum_{i=1}^n x_i} & \text{pro } \frac{n}{\sum_{i=1}^n x_i} > \theta_0 \end{cases} \end{aligned}$$

Proto

$$\lambda = \begin{cases} 1 & \text{pro } \frac{n}{\sum_{i=1}^n x_i} \leq \theta_0 \\ \frac{\theta_0^n e^{-\theta_0 \sum_{i=1}^n x_i}}{\left(\frac{n}{\sum_{i=1}^n x_i} \right)^n e^{-n}} & \text{pro } \frac{n}{\sum_{i=1}^n x_i} > \theta_0 \end{cases}$$

Pro $0 < \lambda_0 < 1$ tak obecný věrohodnostní poměrový test zamítne \mathcal{H}_0 , jestliže

$$\frac{n}{\sum_{i=1}^n x_i} > \theta_0 \text{ a } \left(\frac{\theta_0 \sum_{i=1}^n x_i}{n} \right)^n e^{-\theta_0 \sum_{i=1}^n x_i + n} \leq \lambda_0$$

což je ekvivaletní podmínkám

$$\theta_0 \bar{x} < 1 \text{ a } (\theta_0 \bar{x})^n e^{-n(\theta_0 \bar{x} - 1)} \leq \lambda_0$$

Jestliže použijeme substituci $y = \theta_0 \bar{x}$, pak má $y^n e^{-n(y-1)}$ maximum v bodě $y = 1$. Proto $y < 1$ a $y^n e^{-n(y-1)}$ je menší nebo rovno λ_0 pouze je-li $y \leq k$, kde k je konstanta splňující podmínku $0 < k < 1$. Situaci ilustruje obrázek (9.3).

Obecný věrohodnostní poměrový test se tak zjednoduší do podoby, kdy \mathcal{H}_0 je zamítnuto, jestliže $\theta_0 \bar{x} < k$, kde $0 < k < 1$. Jestliže má tento test mít sílu α , pak je k řešením následující rovnice⁵

$$\alpha = P_{\theta_0}[\theta_0 \bar{X} < k] = P_{\theta_0}\left[\theta_0 \sum_{i=1}^n X_i < nk\right] = \int_0^{nk} \frac{1}{\Gamma(n)} u^{n-1} e^{-u} du$$

Na závěr uvedme, že vzhledem ke kritériu rozkladu musí obecný věrohodnostní poměrový test záviset pouze na minimální dostatečné statistice.

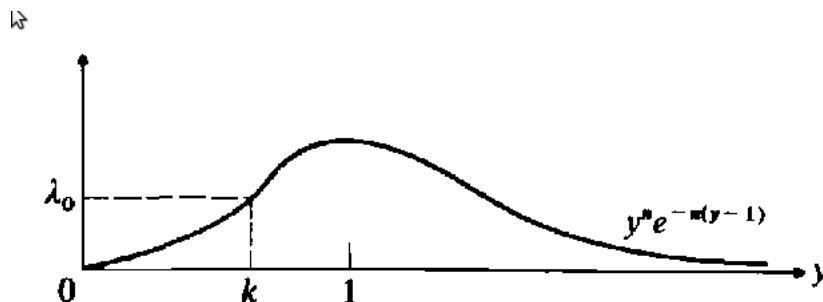
9.3.2 Uniformě nejsilnější test

Definice 9.15 (Uniformě nejsilnější test) Test \mathcal{T}^* pro $\mathcal{H}_0: \theta \in \underline{\Theta}_0$ vs. $\mathcal{H}_1: \theta \in \underline{\Theta} - \underline{\Theta}_0$ je uniformě nejsilnější test právě tehdy a jen tehdy, jestliže

$$1. \sup_{\theta \in \underline{\Theta}_0} \pi_{\mathcal{T}^*}(\theta) = \alpha$$

$$2. \pi_{\mathcal{T}^*}(\theta) \geq \pi_{\mathcal{T}}(\theta) \text{ pro všechna } \theta \in \underline{\Theta} - \underline{\Theta}_0 \text{ a libovolný alternativní test } \mathcal{T} \text{ velikosti menší nebo rovnou } \alpha.$$

⁵Připomeňme, že $P_{\theta}[\theta_0 \bar{X} < k] \leq P_{\theta_0}[\theta_0 \bar{X} < k]$ pro $\theta \leq \theta_0$.



Obrázek 9.3: Funkce charakteristické veličiny

Test \mathcal{T}^* je uniformě nejsilnějším testem s velikostí α , jestliže mezi všemi ostatními testy s velikostí rovnou nebo menší než α má největší sílu pro všechny alternativní hodnoty θ . Prívlastek “uniformní” je váže právě ke “všem” alternativním hodnotám θ . Uniformě nejsilnější test pro dané hypotézy nemusí vždy existovat.

Příklad 9.12 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \theta e^{-\theta x} I_{(0, \infty)}(x)$, kde $\bar{\Theta} = \{\theta : \theta \geq \theta_0\}$. Pokusme se nalézt uniformě nejsilnější test pro $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta > \theta_0$.

Z příkladu (9.7) víme, že pro fixní $\theta_1 > \theta_0$ a $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_0$ nejsilnější test zamítne \mathcal{H}_0 , jestliže $\sum x_i \leq k'$, kde k' je dáno řešením rovnice

$$\alpha = \int_0^{k'} \frac{1}{\Gamma(n)} \theta_0^n x^{n-1} e^{-\theta_0 x} dx$$

Tento test vyplynul z Neyman-Pearsonovy věty. Vzhledem k tomu, že s výjimkou podmínky $\theta_1 > \theta_0$ tento test nezávisí na θ_1 , zikali bychom stejný test pro libovolné $\theta_1 > \theta_0$. Proto se jedná o uniformě nejsilnější test.

Následující větu uvádíme bez důkazu.

Věta 9.5 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta)$, kde $\theta \in \bar{\Theta}$ a $\bar{\Theta}$ je interval. Předpokládejme, že

$$f(x, \theta) = a(\theta)b(x)e^{c(\theta)d(x)}$$

a definujme statistiku

$$t(x_1, \dots, x_n) = \sum_{i=1}^n d(x_i)$$

1. Jestliže je $c(\theta)$ monotónní rostoucí funkce v θ a jestliže existuje k^* takové, že $P_{\theta_0}[t(X_1, \dots, X_n) > k^*] = \alpha$, pak test \mathcal{T}^* s kritickým regionem $C^* = \{(x_1, \dots, x_n) : t(x_1, \dots, x_n) > k^*\}$ je uniformě nejsilnějším testem velikosti α pro $\mathcal{H}_0 : \theta \leq \theta_0$ vs. $\mathcal{H}_1 : \theta > \theta_0$ nebo $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta > \theta_0$.
2. Jestliže je $c(\theta)$ monotónní klesající funkce v θ a jestliže existuje k^* takové, že $P_{\theta_0}[t(X_1, \dots, X_n) < k^*] = \alpha$, pak test \mathcal{T}^* s kritickým regionem $C^* = \{(x_1, \dots, x_n) : t(x_1, \dots, x_n) < k^*\}$ je uniformě nejsilnějším testem velikosti α pro $\mathcal{H}_0 : \theta \leq \theta_0$ vs. $\mathcal{H}_1 : \theta > \theta_0$ nebo $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta > \theta_0$.

Příklad 9.13 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x, \theta) = \theta e^{-\theta x} I_{(0, \infty)}(x)$, kde $\bar{\Theta} = \{\theta > 0\}$. Testované hypotézy mají tvar $\mathcal{H}_0 : \theta \leq \theta_0$ vs. $\mathcal{H}_1 : \theta > \theta_0$. Pravděpodobnostní funkci lze vyjádřit ve tvaru

$$f(x, \theta) = \theta I_{(0, \infty)}(x) e^{-\theta x} = a(\theta) b(x) e^{c(\theta) d(x)}$$

a proto $t(x_1, \dots, x_n) = \sum x_i$, $c(\theta) = -\theta$. $c(\theta)$ je monotónní klesající funkce, a proto dle bodu (2) výše uvedené věty uniformě nejsilnější test zamítne \mathcal{H}_0 pro $\sum x_i < k^*$, kde k^* je řešením rovnice

$$\alpha = P_{\theta_0} \left[\sum_{i=1}^n X_i < k^* \right] = \int_0^{k^*} \frac{1}{\Gamma(n)} \theta_0^n u^{n-1} e^{-\theta_0 u} du$$

Definice 9.16 (Monotónní věrohodnostní poměr) O rodině pravděpodobnostních rozdělení $\{f(x, \theta) : \theta \in \bar{\Theta}, \text{ kde } \bar{\Theta} \text{ je interval}\}$ říkáme, že mají monotónní věrohodnostní poměr, jestliže existuje statistika $T = t(X_1, \dots, X_n)$ taková, že poměr

$$\frac{L(\theta'; x_1, \dots, x_n)}{L(\theta''; x_1, \dots, x_n)}$$

je buďto nerostoucí funkcí $t(x_1, \dots, x_n)$ pro každé $\theta' < \theta''$ nebo neklesající funkcí $t(x_1, \dots, x_n)$ pro každé $\theta' < \theta''$.

Příklad 9.14 Jestliže $\{f(x; \theta) : \theta \in \bar{\Theta}\} = \{\theta e^{-\theta x} I_{(0, \infty)}(x) : \theta > 0\}$, pak

$$\frac{L(\theta'; x_1, \dots, x_n)}{L(\theta''; x_1, \dots, x_n)} = \frac{(\theta')^n e^{-\theta' \sum_{i=1}^n x_i}}{(\theta'')^n e^{-\theta'' \sum_{i=1}^n x_i}} = \left(\frac{\theta'}{\theta''} \right)^n e^{-(\theta' - \theta'') \sum_{i=1}^n x_i}$$

což je monotónní rostoucí funkce v $\sum x_i$.

Příklad 9.15 Jestliže $\{f(x; \theta) : \theta \in \bar{\Theta}\} = \{(1/\theta) I_{(0, \infty)}(x) : \theta > 0\}$, pak

$$\begin{aligned} \frac{L(\theta'; x_1, \dots, x_n)}{L(\theta''; x_1, \dots, x_n)} &= \frac{\left(\frac{1}{\theta'}\right)^n \prod_{i=1}^n I_{(0, \theta')}(x_i)}{\left(\frac{1}{\theta''}\right)^n \prod_{i=1}^n I_{(0, \theta'')}(x_i)} \\ &= \begin{cases} \left(\frac{\theta''}{\theta'}\right)^n & \text{pro } 0 < y_n < \theta' \\ 0 & \text{pro } \theta' \leq y_n < \theta'' \end{cases} \end{aligned}$$

což je monotónní nerostoucí funkce v $y_n = \max(x_1, \dots, x_n)$. Připomeňme, že y_n nemůže padnout mimo interval $(0, \theta'')$, jestliže θ nabývá hodnoty θ' nebo θ'' .

Následující větu uvádíme bez důkazu.

Věta 9.6 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x; \theta)$, kde $\bar{\Theta}$ představuje interval. Předpokládejme, že rodina pravděpodobnostních rozdělení $\{f(x; \theta) : \theta \in \bar{\Theta}\}$ má monotónní věrohodnostní poměr pro statistiku $t(X_1, \dots, X_n)$.

1. Jestliže je monotónní věrohodnostní poměr neklesající v $t(x_1, \dots, x_n)$ a jestliže existuje k^* takové, že $P_{\theta_0}[t(X_1, \dots, X_n) < k^*] = \alpha$, pak je test odpovídající kritickému regionu $C^* = \{(x_1, \dots, x_n) : t(x_1, \dots, x_n) < k^*\}$ uniformě nejsilnější test síly α pro $\mathcal{H}_0 : \theta \leq \theta_0$ vs. $\mathcal{H}_1 : \theta > \theta_0$.

2. Jestliže je monotónní věrohodnostní poměr nerostoucí v $t(x_1, \dots, x_n)$ a jestliže existuje k^* takové, že $P_{\theta_0}[t(X_1, \dots, X_n) > k^*] = \alpha$, pak je test odpovídající kritickému regionu $C^* = \{(x_1, \dots, x_n) : t(x_1, \dots, x_n) > k^*\}$ uniformě nejsilnější test síly α pro $\mathcal{H}_0 : \theta \leq \theta_0$ vs. $\mathcal{H}_1 : \theta > \theta_0$.

Příklad 9.16 Uvažujme náhodný výběr X_1, \dots, X_n z populace $f(x; \theta) = \frac{1}{\theta} I_{(0, \theta)}(x)$, kde $\theta > 0$ a hypotézy $\mathcal{H}_0 : \theta \leq \theta_0$ vs. $\mathcal{H}_1 : \theta > \theta_0$. Z předchozího příkladu víme, že toto pravděpodobnostní rozdělení má monotónní nerostoucí věrohodnostní poměr pro $(x_1, \dots, x_n) = y_n = \max(x_1, \dots, x_n)$. Dle bodu (2) výše uvedené věty je uniformě nejsilnější test velikosti α definován jako zamítnutí \mathcal{H}_0 , jestliže $y_n > k^*$, kde k^* je řešením rovnice

$$\alpha = P_{\theta_0}[Y_n > k^*] = \int_{k^*}^{\theta_0} n \left(\frac{y}{\theta_0} \right)^{n-1} \frac{dy}{\theta_0} = \frac{1}{\theta_0^n} [\theta_0^n - (k^*)^n] = 1 - \left(\frac{k^*}{\theta_0} \right)^n$$

což implikuje $k^* = \theta_0 \sqrt[n]{1 - \alpha}$.

Závěrem uveďme následující tři poznámky. Za prvé, jak ve větě (9.5) tak ve větě (9.6) je nulová hypotéza formulována jako $\theta \leq \theta_0$. Jestliže by nulová hypotéza byla formulována jako $\theta \geq \theta_0$, zůstaly by obě věty v platnosti, pokud bychom obrátily nerovnosti, které definují kritický region. Za druhé, věta (9.5) je důsledkem věty (9.6). Za třetí, obě věty uvažují pouze jednostranné hypotézy.

9.3.3 Nezkreslený test

V řadě případů uniformě nejsilnější test neexistuje. V takovém případě můžeme soubor testů omezit a následně nalézt uniformě nejsilnější test na této omezené množině. Tímto se dostáváme ke konceptu nezkresleného testu.

Definice 9.17 (Nezkreslený test) Test \mathcal{T} hypotéz $\mathcal{H}_0 : \theta \in \overline{\Theta}_0$ vs. $\mathcal{H}_1 : \theta \in \overline{\Theta}_1$ je nezkreslený test, jestliže

$$\sup_{\theta \in \overline{\Theta}_0} \pi_{\mathcal{T}}(\theta) \leq \inf_{\theta \in \overline{\Theta}_1} \pi_{\mathcal{T}}(\theta)$$

V případě nezkresleného testu je tedy pravděpodobnost zamítnutí nepravdivé \mathcal{H}_0 přinejmenším stejně velká jako pravděpodobnost zamítnutí pravdivé \mathcal{H}_0 . Jestliže v rámci takto omezené množiny testů existuje test, který je uniformě nejsilnější, pak se jedná o uniformě nejsilnější nezkreslený test. Pro nalezení uniformě nejsilnějšího nezkresleného testu byla vyvinuta poměrně složitá teorie, která však přesahuje záběr této knihy.

9.3.4 Metody pro nalézání testů

V rámci testu testujeme nulovou hypotézu \mathcal{H}_0 proti alternativní hypotéze \mathcal{H}_1 . Uvažujme statistiku, která pro obě hypotézy chová "odlišně", a pokusme se využít toto odlišné chování pro definici testu. Jako ilustraci uvažujme testování $\mathcal{H}_0 : \theta \leq \theta_0$ vs. $\mathcal{H}_1 : \theta > \theta_1$, kde náhodný výběr X_1, \dots, X_n pochází z $f(x; \theta) = \phi_{\theta, 1}(x)$. Statistika \bar{X} sleduje normální rozdělení se střední hodnotou θ a rozptylem $1/n$. Lze tedy očekávat, že \bar{X} bude menší pro pravdivé \mathcal{H}_0 a větší pro pravdivé \mathcal{H}_1 - statistika \bar{X} se tak chová odlišně pro nulovou a alternativní

hypotézu. Rozumným testem se tedy zdá zamítnutí \mathcal{H}_0 , je-li $\bar{X} > k$, kde k je dáno velikostí testu⁶.

Klíčem této techniky je tedy nalezení statistiky, která vykazuje “odlišné” chování pro nulovou a alternativní hypotézu. Jestliže existuje dostatečná statistika, pak se jedná o přirozeného kandidáta. V opačné případě je možné použít vhodnou funkci odhadu⁷. Ve výše uvedeném případě tuto funkci plní \bar{X} , což zároveň dostatečná statistika a funkce odhadu parametru θ založená na metodě maximální věrohodnosti.

Příklad 9.17 Uvažujme náhodný výběr X_1, \dots, X_n z Poissonova rozdělení se střední hodnotou θ . Testujeme $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta \neq \theta_1$. Předpokládejme $n = 10$ a $\theta_1 = 1$.

Víme, že \bar{X} je funkce odhadu pro θ založená na metodě maximální věrohodnosti. To znamená, že \bar{X} bude mít tendenci se nacházet v okolí θ_0 , jestliže je \mathcal{H}_0 pravdivá. Přirozená formulace testu tedy zní: “Přijmi \mathcal{H}_0 , jestliže $c_1 < \bar{x} < c_2$ ”, kde c_1 a c_2 jsou zvoleny tak, aby měl test požadovanou velikost. Např. test “přijmi \mathcal{H}_0 , jestliže $0.4 < \bar{x} < 1.6$ ” má velikost

$$\begin{aligned} 1 - P_{\theta=1}[0.4 < \bar{X} < 1.6] &= 1 - P_{\theta=1}[4 < \sum_{i=1}^{10} X_i < 16] \\ &= 1 - \sum_{j=5}^{15} \frac{e^{-10} 10^j}{j!} \approx 0.078 \end{aligned}$$

Test, který je běžně používán v praxi, testuje $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta \neq \theta_0$. Pro ilustraci předpokládejme, že θ je rozdílem výnosů dvou odrůd pšenice. Na první pohled by se mohlo zdát, že kandidátem na vhodný test je $\mathcal{H}_0 : \theta = 0$ vs. $\mathcal{H}_1 : \theta \neq 0$, který testuje, zda-li existuje rozdíl ve výnosech obou odrůd. Nicméně θ je spojitá veličina, a proto není možné, aby nabývala právě hodnoty nula. Jako vhodnější se proto jeví test $\mathcal{H}_0 : \theta_1 \leq \theta \leq \theta_2$ vs. $\mathcal{H}_1 : \theta < \theta_1$ nebo $\theta > \theta_2$. Nalézt odpovídající uniformě nejsilnější test může být složité či dokonce nemožné. Nicméně jestliže $f(x; \theta)$ je pravděpodobnostním rozdělením s jedním parametrem, pak lze pro konstrukci testu použít funkci odhadu $\hat{\theta}$ založenou na metodě maximální věrohodnosti a síla tohoto testu je pak porovnána s ideální funkcí charakteristické veličiny testu velikosti α . Pro některá pravděpodobnostní rozdělení lze použít test, který zamítne θ_0 , jestliže se $\hat{\theta}$ nenachází v určitém intervalu (c_1, c_2) , kde c_1 a c_2 mohou být vybrána tak, aby platilo

$$\int_{c_1}^{c_2} f_{\hat{\theta}}(\hat{\theta}; \theta_1) d\hat{\theta} = \int_{c_1}^{c_2} f_{\hat{\theta}}(\hat{\theta}; \theta_2) d\hat{\theta} = 1 - \alpha$$

kde $f_{\hat{\theta}}(\hat{\theta}, \theta)$ je pravděpodobnostní funkce $\hat{\theta}$ pro parametr θ . Odpovídající funkce charakteristické veličiny je pak

$$\pi(\theta) = 1 - \int_{c_1}^{c_2} f_{\hat{\theta}}(\hat{\theta}; \theta) d\hat{\theta} \text{ pro } \theta \in \bar{\Theta}$$

Tuto funkci charakteristické veličiny je možné porovnat s ideální funkcí charakteristické veličiny a pokud její odchylka nepřesahuje přijatelnou mez, může být příslušný test použit, i když se nejedná o uniformě nejsilnější test.

⁶Z předchozí kapitoly víme, že takovýto test je uniformě nejsilnější.

⁷Např. funkci odhadu založenou na metodě maximální věrohodnosti.

Příklad 9.18 Uvažujme náhodný výběr X_1, \dots, X_n z populace $\phi_{\theta,1}(x)$. Testujme $\mathcal{H}_0 : 1 \leq \theta \leq 2$ vs. $\mathcal{H}_1 : \theta < 1$ nebo $\theta > 2$.

Víme, že \bar{X} je funkce odhadu pro θ dle metody maximální věrohodnosti. Tato statistika sleduje normální rozdělení se střední hodnotou θ a rozptylem $1/n$. Naším cílem je nalézt c_1 a c_2 takové, aby platilo

$$1 - \alpha = \int_{c_1}^{c_2} \phi_{1,1/n}(x) dx = \int_{c_1}^{c_2} \phi_{2,1/n}(x) dx$$

tj.

$$\Phi\left(\frac{c_2 - 1}{\sqrt{1/n}}\right) - \Phi\left(\frac{c_1 - 1}{\sqrt{1/n}}\right) = \Phi\left(\frac{c_2 - 2}{\sqrt{1/n}}\right) - \Phi\left(\frac{c_1 - 2}{\sqrt{1/n}}\right) = 1 - \alpha$$

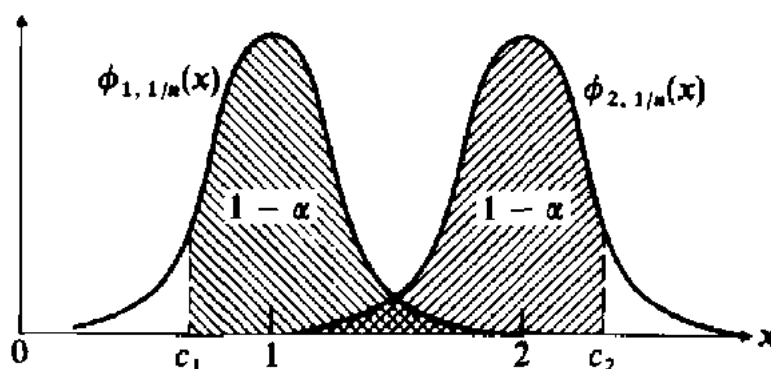
Jak je patrné z obrázku (9.4), platí $c_1 = \frac{3}{2} - d$ a $c_2 = \frac{3}{2} + d$, kde d je dáno

$$\Phi\left(\frac{d + \frac{1}{2}}{\sqrt{1/n}}\right) - \Phi\left(\frac{\frac{1}{2} - d}{\sqrt{1/n}}\right) = 1 - \alpha$$

Jestliže, např. $\alpha = 0.05$ a $n = 16$, pak $d \approx 0.911$, $c_1 \approx 0.589$ a $c_2 \approx 2.411$. Funkce charakteristické veličiny je

$$\pi(\theta) = 1 - P_\theta[c_1 < \bar{X} < c_2] = 1 - P_\theta[0.589 < \bar{X} < 2.411]$$

a je ilustrována obrázkem (9.5).



Obrázek 9.4: Ilustrační obrázek

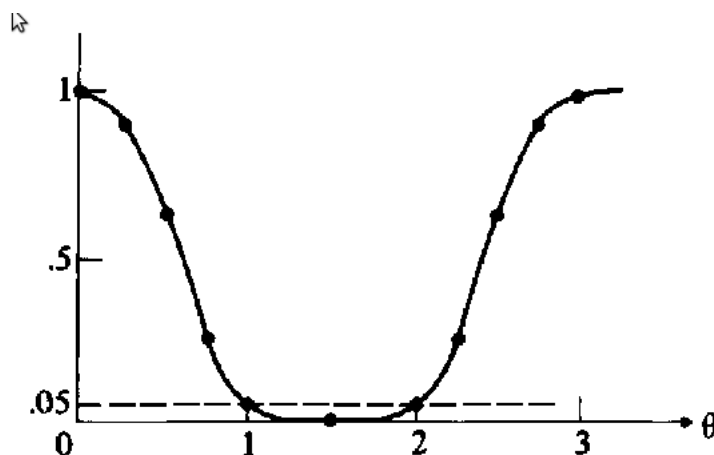
9.4 Testování hypotéz - výběr z normálního rozdělení

9.4.1 Testování středních hodnot

$\mathcal{H}_0 : \mu \leq \mu_0$ vs. $\mathcal{H}_1 : \mu > \mu_0$

Nejprve předpokládejme, že známe σ^2 . Normální pravděpodobnostní rozdělení

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



Obrázek 9.5: Ilustrační obrázek

je člen rodiny exponenciálních rozdělení s

$$a(\mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2}, \quad b(x) = e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}, \quad c(\mu) = \frac{\mu}{\sigma} \quad \text{a} \quad d(x) = x$$

Podmínky věty (9.5) jsou splněny, a proto uniformě nejsilnější test existuje a zahrnuje \mathcal{H}_0 pro $t(x_1, \dots, x_n) = \sum x_i > k^*$, kde k^* je řešením $P_{\mu_0}[\sum X_i > k^*] = \alpha$. Protože $\alpha = P_{\mu_0}[\sum X_i > k^*] = 1 - \Phi\left(\frac{k^* - n\mu_0}{\sqrt{n}\sigma}\right)$, tak $\frac{k^* - n\mu_0}{\sqrt{n}\sigma} = z_{1-\alpha}$, kde $z_{1-\alpha}$ je $(1-\alpha)$ -tý kvantil normovaného normálního rozdělení. Test tedy zamítne \mathcal{H}_0 , jestliže $\sum x_i > n\mu_0 + \sqrt{n}\sigma z_{1-\alpha}$ neboli $\bar{x} > \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$.

Jestliže je σ^2 neznáme, pak je testování $\mathcal{H}_0 : \mu \leq \mu_0$ vs. $\mathcal{H}_1 : \mu > \mu_0$ ekvivalentní testování $\mathcal{H}_0 : \theta \in \underline{\Theta}_0$ vs. $\mathcal{H}_1 : \theta \notin \underline{\Theta}_0$, kde $\theta = (\mu, \sigma^2)$, $\underline{\Theta} = \{(\mu, \sigma^2) : -\infty < \mu < \infty; \sigma^2 > 0\}$ a $\underline{\Theta}_0 = \{(\mu, \sigma^2) : \mu \leq \mu_0; \sigma^2 > 0\}$. Pro formulaci testu můžeme (a) použít obecný věrohodnostní poměr nebo (b) nalézt statistiku, která se pro obě hypotézy chová “odlišně” a založit na ní test. Příkladem takovéto statistiky je $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$, kde \bar{X} je střední hodnota výběru a S^2 je výběrový rozptyl. Protože T bude větší pro $\mu > \mu_0$ než pro $\mu \leq \mu_0$, bude test zahrnovat \mathcal{H}_0 pro vysoké hodnoty T , tj. zamítne \mathcal{H}_0 , jestliže $T > k$. Jestliže $\mu = \mu_0$, pak T sleduje studentovo rozdělení s $n-1$ stupni volnosti. k je tak možné získat z $\alpha = P_{\mu_0}[T > k]$, což implikuje $k = t_{1-\alpha}(n-1)$, tj. k je rovno kvantilu $(1-\alpha)$ studentova rozdělení s $(n-1)$ stupni volnosti. Lze dokázat, že takto odvozený test je zároveň testem velikosti α založeným na obecném věrohodnostním poměru.

$\mathcal{H}_0 : \mu = \mu_0$ vs. $\mathcal{H}_1 : \mu \neq \mu_0$

Jestliže známe σ^2 , pak je $(\bar{X} - z_{(1+\gamma)/2}(\sigma/\sqrt{n}), \bar{X} + z_{(1+\gamma)/2}(\sigma/\sqrt{n}))$ je 100γ procentní interval spolehlivosti pro μ , kde $z_{(1+\gamma)/2}$ je $(1+\gamma)/2$ procentní kvantil normovaného normálního rozdělení. Test pak zamítne \mathcal{H}_0 , jestliže tento interval nezahrnuje μ_0 . Takto definovaný test má navíc velikost $1-\gamma$, protože

$$P_{\mu=\mu_0} \left[\bar{X} - z_{(1+\gamma)/2} \frac{\sigma}{\sqrt{n}} < \mu_0 < \bar{X} + z_{(1+\gamma)/2} \frac{\sigma}{\sqrt{n}} \right] = \gamma$$

Jestliže je σ^2 neznámé, pak lze odvodit test podobný výše uvedenému, který vychází z 100γ procentního intervalu spolehlivosti

$$\left(\bar{X} - t_{(1+\gamma)/2}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{(1+\gamma)/2}(n-1) \frac{S}{\sqrt{n}} \right)$$

Pokusme se však nalézt test založený na obecném věrohodnostním poměru. Funkce maximální věrohodnosti má tvar

$$L(\mu, \sigma^2; x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2}$$

$\bar{\Theta}_0 = \{(\mu, \sigma^2) : \mu = \mu_0; \sigma^2 > 0\}$ a $\bar{\Theta} = \{(\mu, \sigma^2) : -\infty < \mu < \infty; \sigma^2 > 0\}$. Víme, že hodnoty pro μ a σ^2 , které maximalizují $L(\mu, \sigma^2; x_1, \dots, x_n)$ nad $\bar{\Theta}$, jsou $\hat{\mu} = \bar{x}$ a $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$, proto

$$\sup_{\bar{\Theta}} L(\mu, \sigma^2; x_1, \dots, x_n) = \left(\frac{n}{2\pi \sum_{i=1}^n (x_i - \bar{x})^2} \right)^{n/2} e^{-n/2}$$

Abychom maximalizovali $L(\mu, \sigma^2; x_1, \dots, x_n)$ nad $\bar{\Theta}_0$, dosadíme $\mu = \mu_0$. Jediným neznámým parametrem tak zůstává σ^2 , který maximalizuje $L(\mu, \sigma^2; x_1, \dots, x_n)$ pro $\frac{\sum_{i=1}^n (x_i - \mu_0)^2}{n}$, čímž získáváme

$$\sup_{\bar{\Theta}_0} L(\mu, \sigma^2; x_1, \dots, x_n) = \left(\frac{n}{2\pi \sum_{i=1}^n (x_i - \mu_0)^2} \right)^{n/2} e^{-n/2}$$

Obecný věrohodnostní poměr je tak

$$\begin{aligned} \lambda &= \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \mu_0)^2} \right)^{n/2} = \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu_0)^2} \right)^{n/2} \\ &= \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2} \right)^{n/2} = \left(\frac{1}{1 + n(\bar{x} - \mu_0)^2 / \sum_{i=1}^n (x_i - \bar{x})^2} \right)^{n/2} \end{aligned}$$

Všimněme si, že λ je monotónní funkcí pro statistiku

$$t^2 = t^2(x_1, \dots, x_n) = \frac{n(n-1)(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \left(\frac{\bar{x} - \mu_0}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)}} \right)^2$$

a proto je kritický region ve tvaru $\lambda \leq \lambda_0$ ekvivalentní kritickému regionu ve tvaru $t^2(x_1, \dots, x_n) \geq k^2$. Test založený obecném věrohodnostním poměru tak zamítne \mathcal{H}_0 , jestliže

$$T^2 = \left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \right)^2 \geq k^2$$

neboli přijme \mathcal{H}_0 , jestliže $-k < T < k$. Protože T sleduje studentovo rozdělení s $n-1$ stupni volnosti pro $\mu = \mu_0$, je k vybráno tak, aby splňovalo

$$\int_{-k}^k f_T(t; n-1) dt = 1 - \alpha$$

kde α je požadovaná velikost testu. To znamená, že k je rovno $t_{(1-\alpha)/2}(n-1)$, tj. kvantilu $(1-\alpha)/2$ studentova rozdělení s $n-1$ stupni volnosti. Takto odvozený test se shoduje s testem založeným na intervalu spolehlivosti, který jsme definovali na začátku jako

$$\left(\bar{X} - t_{(1+\gamma)/2}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{(1+\gamma)/2}(n-1) \frac{S}{\sqrt{n}} \right)$$

kde $\gamma = 1 - \alpha$. Bez důkazu uvedme, že tento test je také uniformě nejsilnějším testem.

Ostatní případy

Ve výše uvedeném textu jsme odvodili test střední hodnoty normálního rozdělení pro jednostranné i oboustranné hypotézy. Jednostrannou nulovou hypotézu $\mu \leq \mu_0$ lze obrátit a získat příslušný test. Je možné definovat také další typy hypotéz, jako např. $\mathcal{H}_0 : \mu_1 \leq \mu \leq \mu_2$ vs. $\mathcal{H}_1 : \mu < \mu_1$ nebo $\mu > \mu_2$.

9.4.2 Testování rozptylu

$\mathcal{H}_0 : \sigma^2 \leq \sigma_0^2$ vs. $\sigma^2 > \sigma_0^2$

Předpokládejme, že μ je známo. Pak je parametrický prostor reprezentovaný intervalem. Námi uvažovaný test je jednostranný, a proto máme naději nalézt uniformě nejsilnější test velikosti α . Pravděpodobnostní funkce normálního rozdělení

$$f(x; \theta) = f(x; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

je člen rodiny exponenciálních rozdělení s $a(\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}}$, $b(x) = 1$, $c(\sigma^2) = -\frac{1}{2}\sigma^2$ a $d(x) = (x - \mu)^2$. $c(\sigma^2)$ je monotónní rostoucí funkce v σ^2 , a proto dle věty (9.5) je test s kritickým regionem $\{(x_1, \dots, x_n) : \sum (x_i - \mu)^2 > k^*\}$ uniformě nejsilnějším testem velikosti α , přičemž k^* je dáno rovnicí $P_{\sigma^2=\sigma_0^2}[\sum (X_i - \mu)^2 > k^*] = \alpha$. To implikuje $k^* = \sigma_0^2 \chi_{1-\alpha}^2(n)$, kde $\chi_{1-\alpha}^2(n)$ je $1 - \alpha$ procentní kvantil chi-kvadrát rozdělení s n stupni volnosti.

Jestliže μ není známé, lze test odvodit s pomocí statistiky $V = \frac{\sum (X_i - \bar{X})^2}{\sigma_0^2}$. Statistika V bude nabývat vysokých hodnot pro $\sigma^2 > \sigma_0^2$, a proto test bude zamítat \mathcal{H}_0 pro vysoká V . Jestliže $\sigma^2 = \sigma_0^2$, pak bude V sledovat chi-kvadrát rozdělení s $n-1$ stupni volnosti a $P_{\sigma^2=\sigma_0^2}[V > \chi_{1-\alpha}^2(n-1)] = \alpha$, kde $\chi_{1-\alpha}^2(n-1)$ je $(1-\alpha)$ procentní kvantil chi-kvadrát rozdělení s $n-1$ stupni volnosti. Lze dokázat, že test, který zamítne \mathcal{H}_0 pro $\frac{\sum (X_i - \bar{X})^2}{\sigma_0^2} > \chi_{1-\alpha}^2(n-1)$, je také testem velikosti α založeným na obecném věrohodnostním poměru.

$\mathcal{H}_0 : \sigma^2 = \sigma_0^2$ vs. $\sigma^2 \neq \sigma_0^2$

Vynechme situaci, kdy je μ známo a zaměříme se na situaci, kdy μ neznáme. V tomto případě je $\bar{\Theta}_0 = \{(\mu, \sigma) : -\infty < \mu < \infty; \sigma^2 = \sigma_0^2\}$. Pokusme se nalézt test velikosti α s využitím metody intervalu spolehlivosti. Z kapitoly (8.2.2) víme, že 100γ procentní interval spolehlivosti pro σ^2 je definován jako

$$\left(\frac{(n-1)S^2}{q_2}, \frac{(n-1)S^2}{q_1} \right)$$

kde q_1 a q_2 jsou kvantily chi-kvadrát rozdělení $f_Q(q; n-1)$ s $n-1$ stupni volnosti, kde

$$\int_{q_1}^{q_2} f_Q(q; n-1) dq = \gamma$$

Test velikosti $\alpha = 1 - \gamma$ pak přijme \mathcal{H}_0 , jestliže σ_0^2 je obsaženo ve výše uvedeném intervalu. Lze dokázat, že tento test odpovídá testu založenému na obecném věrohodnostním poměru.

9.4.3 Test několika středních hodnot

Test dvou středních hodnot

Uvažujeme dvě vzájemně nezávislé populace, které sledují normální rozdělení - jednu se střední hodnotou μ_1 a rozptylem σ_1^2 a druhou se střední hodnotou μ_2 a rozptylem σ_2^2 . Testujeme hypotézy $\mathcal{H}_0 : \mu_1 = \mu_2, \sigma_1^2 > 0, \sigma_2^2 > 0$ vs. $\mathcal{H}_1 : \mu_1 \neq \mu_2, \sigma_1^2 > 0, \sigma_2^2 > 0$. Parametrický prostor $\bar{\Theta}$ je tedy čtyřrozměrný a příslušné sdružené pravděpodobnostní rozdělení je definováno skrze $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$. Podprostor \bar{H}_0 je pak pouze třírozměrný, protože $\mu_1 = \mu_2 = \mu$. Uvažujeme náhodný výběr X_{11}, \dots, X_{1n_1} z první a náhodný výběr X_{21}, \dots, X_{2n_2} z druhé populace. Funkce věrohodnosti je pak definována jako

$$\begin{aligned} L(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2; x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{2n_2}) &= L \\ &= \left(\frac{1}{2\pi\sigma_1^2} \right)^{n_1/2} e^{-\frac{1}{2} \sum_{i=1}^{n_1} \left(\frac{x_{1i} - \mu_1}{\sigma_1} \right)^2} \left(\frac{1}{2\pi\sigma_2^2} \right)^{n_2/2} e^{-\frac{1}{2} \sum_{j=1}^{n_2} \left(\frac{x_{2j} - \mu_2}{\sigma_2} \right)^2} \end{aligned}$$

a její maximum nad $\bar{\Theta}$ je

$$\sup_{\bar{\Theta}} L = \left(\frac{n_1}{2\pi \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2} \right)^{n_1/2} \left(\frac{n_2}{2\pi \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2} \right)^{n_2/2} e^{-n_1/2} e^{-n_2/2}$$

Jestliže použijeme $\mu_1 = \mu_2 = \mu$ a pokusíme se maximalizovat L vzhledem k μ, σ_1^2 a σ_2^2 , zjistíme, že odhad μ má charakter kořene polynomu třetího řádu a že se jedná o poměrně složitou funkci. To znamená, že také obecný věrohodnostní poměr λ je složitou funkcí, která navíc obsahuje podíl rozptylů σ_1^2 a σ_2^2 a její pravděpodobnostní funkce je velmi náročná na odvození. Je tedy nemožné analyticky odvodit kritický region $0 < \lambda < k$ pro danou pravděpodobnost chyby typu I, protože poměr rozptylů uvažovaných populací není znám. Kořen polynomu třetího řádu je možné vypočítat numericky, což nám následně umožní výpočet λ . Pro velké populace navíc, jak si ukážeme v kapitole (9.5), platí, že $-\ln(\Lambda)$ přibližně sleduje chi-kvadrát rozdělení s jedním stupněm volnosti. To nám umožní konstrukci testu, který zamítne \mathcal{H}_0 pro velká $-2\ln(\lambda)$.

Jestliže mají obě uvažované populace shodný rozptyl, problematika se značně zjednoduší. Parametrický prostor $\bar{\Theta}$ se stává třírozměrným se souřadnicemi (μ_1, μ_2, σ^2) a parametrický prostor $\bar{\Theta}_0$ dvourozměrným se souřadnicemi (μ, σ^2) . Pro $\bar{\Theta}$ jsou odhady pro μ_1, μ_2 a σ^2 dány \bar{x}_1, \bar{x}_2 a $\frac{1}{n_1+n_2} (\sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{2j} - \bar{x}_2)^2)$, a proto

$$\sup_{\bar{\Theta}} L = \left(\frac{n_1 + n_2}{2\pi (\sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 + \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2)} \right)^{(n_1+n_2)/2} e^{-(n_1+n_2)/2}$$

Pro $\bar{\Theta}_0$ jsou odhady pro μ a σ^2 založené na metodě maximální věrohodnosti dány

$$\hat{\mu} = \frac{1}{n_1 + n_2} \left(\sum_{i=1}^{n_1} x_{1i} + \sum_{j=1}^{n_2} x_{2j} \right) = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

a

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n_1 + n_2} \left(\sum_{i=1}^{n_1} (x_{1i} - \hat{\mu})^2 + \sum_{j=1}^{n_2} (x_{2j} - \hat{\mu})^2 \right) \\ &= \frac{1}{n_1 + n_2} \left(\sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 + \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2 + \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)^2 \right) \end{aligned}$$

čímž získáme

$$\sup_{\bar{\Theta}_0} L = \left(\frac{n_1 + n_2}{2\pi \left(\sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 + \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2 + \frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)^2 \right)} \right)^{\frac{n_1 + n_2}{2}} e^{-\frac{n_1 + n_2}{2}}$$

Obecný věrohodnostní poměr je tak roven

$$\lambda = \left(1 + \frac{\frac{n_1 n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)^2}{\sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 + \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2} \right)^{-\frac{n_1 + n_2}{2}}$$

Tento výraz je velmi podobný tomu, který jsme odvodili v kapitole (9.4.1). Příslušný test se pak také opírá o studentovo rozdělení. Víme, že \bar{X}_1 a \bar{X}_2 jsou nezávislé náhodné veličiny, které sledují normální rozdělení se střední hodnotou μ_1 resp. μ_2 a rozptylem σ^2/n a σ^2/n_2 . Je zřejmé, že $\bar{X}_1 - \bar{X}_2$ je normálně rozdělené se střední hodnotou $\mu_1 - \mu_2$ a rozptylem $\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$. Pro nulovou hypotézu je střední hodnota $\bar{X}_1 - \bar{X}_2$ rovna 0. $\frac{\sum (X_{1i} - \bar{X}_1)^2}{\sigma^2}$ a $\frac{\sum (X_{2j} - \bar{X}_2)^2}{\sigma^2}$ jsou nezávislé a sledují chi-kvadrát rozdělení s $n_1 - 1$ resp. $n_2 - 1$ stupni volnosti. Její součet tedy také sleduje chi-kvadrát rozdělení s $n_1 + n_2 - 2$ stupni volnosti. Protože pro nulovou hypotézu sleduje $Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma \sqrt{1/n_1 + 1/n_2}}$ normální rozdělení s nulovou střední hodnotou a jednotkovým rozptylem, sleduje

$$T = \frac{\sqrt{\frac{n_1 n_2}{n_1 + n_2}} (\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2 + \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2}{n_1 + n_2 - 2}}}$$

studentovo rozdělení s $n_1 + n_2 - 2$ stupni volnosti⁸. Obecný věrohodnostní poměr je pak

$$\lambda = \left(\frac{1}{1 + \frac{t^2}{n_1 + n_2 - 2}} \right)^{\frac{n_1 + n_2}{2}}$$

⁸Ve výše uvedené rovnici jsou číselník a jmenovatel vzájemně nezávislí.

a sleduje studentovo rozdělení. Např. 5 procentní kritický region pro T je $T^2 > [t_{0.975}(n_1 + n_2 - 2)]^2$, kde $t_{0.975}(n_1 + n_2 - 2)$ je 0.975 procentní kvantil studentova rozdělení s $n_1 + n_2 - 2$ stupni volnosti.

Pokud bychom chtěli testovat $\mathcal{H}_0 : \mu_1 = \mu_2$ vs. $\mathcal{H}_1 : \mu_1 > \mu_2$ nebo $\mathcal{H}_0 : \mu_1 \leq \mu_2$ vs. $\mathcal{H}_1 : \mu_1 > \mu_2$, pak by test s velikostí α zamítl \mathcal{H}_0 jestliže $T > t_{1-\alpha}(n_1 + n_2 - 2)$.

Test několika středních hodnot

Výše uvedený test lze rozšířit ze dvou na k normálních rozdělení. Uvažujme náhodný výběr X_{j1}, \dots, X_{jn_j} z j -té populace se střední hodnotou μ_j a rozptylem σ^2 . Předpokládejme, že všech k populací je vzájemně nezávislých. Testujme nulovou hypotézu, že střední hodnoty všech k populací jsou shodné proti alternativní hypotéze, že tomu tak není. Pro konstrukci testu použijme obecný věrohodnostní poměr.

Funkce maximální věrohodnosti má tvar

$$L(\mu_1, \dots, \mu_k, \sigma^2; x_{11}, \dots, x_{1n_1}, \dots, x_{k1}, \dots, x_{kn_k}) \\ = \prod_{j=1}^k \prod_{i=1}^{n_j} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_{ji}-\mu_j}{\sigma}\right)^2} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji}-\mu_j)^2}$$

kde $n = \sum_{j=1}^k n_j$. Parametrický prostor $\bar{\Theta}$ je $(k+1)$ -rozměrný prostor se souřadnicemi $(\mu_1, \dots, \mu_k, \sigma^2)$ a $\bar{\Theta}_0$ je dvourozměrným parametrickým prostorem se souřadnicemi μ, σ^2 , kde $\mu = \mu_1 = \dots = \mu_k$. Pro $\bar{\Theta}$ jsou odhady pro $\mu_1, \dots, \mu_k, \sigma^2$ založené na metodě maximální věrohodnosti definovány jako

$$\hat{\mu}_j = \bar{x}_{j\cdot} = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ji}, \quad j = 1, \dots, k$$

a

$$\hat{\sigma}_{\bar{\Theta}}^2 = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_{j\cdot})^2$$

a proto

$$\sup_{\bar{\Theta}} L = \left(\frac{2\pi \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_{j\cdot})^2}{n} \right)^{-\frac{n}{2}} e^{-\frac{n}{2}}$$

Pro $\bar{\Theta}_0$ jsou odhady pro μ a σ^2 založené na metodě maximální věrohodnosti definovány jako

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} x_{ji}$$

a

$$\hat{\sigma}_{\bar{\Theta}_0}^2 = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x})^2$$

a proto

$$\sup_{\bar{\Theta}_0} L = \left(\frac{2\pi \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x})^2}{n} \right)^{-\frac{n}{2}} e^{-\frac{n}{2}}$$

Obecný věrohodnostní poměr je pak definován jako

$$\begin{aligned}\lambda &= \frac{\sup_{\Theta_0} L}{\sup_{\Theta}} = \left(\frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x})^2}{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_{j\cdot})^2} \right)^{-\frac{n}{2}} \\ &= \left(\frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_{j\cdot} + \bar{x}_{j\cdot} - \bar{x})^2}{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_{j\cdot})^2} \right)^{-\frac{n}{2}} = \left(\frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_{j\cdot})^2 + \sum_{j=1}^k n_j (\bar{x}_{j\cdot} - \bar{x})^2}{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_{j\cdot})^2} \right)^{-\frac{n}{2}} \\ &= \left(1 + \frac{k-1}{n-k} \frac{\sum_{j=1}^k n_j (\bar{x}_{j\cdot} - \bar{x})^2 / (k-1)}{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_{j\cdot})^2 / (n-k)} \right)^{-\frac{n}{2}}\end{aligned}$$

Test založený na obecném věrohodnostním poměru zamítne \mathcal{H}_0 , jestliže $\lambda \leq \lambda_0$. Nicméně $\lambda \leq \lambda_0$ pouze jestliže

$$r = \frac{\sum_{j=1}^k n_j (\bar{x}_{j\cdot} - \bar{x})^2 / (k-1)}{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_{j\cdot})^2 / (n-k)} \geq c$$

kde c je konstanta. Poměr r nazýváme rozptylovým poměr nebo také F poměrem. Konstanta c je zvolena tak, aby $P[R \geq c | \mathcal{H}_0] = \alpha$. Vzhledem k tomu, že $\bar{X}_{j\cdot}$ a $\sum_{i=1}^{n_j} (X_{ji} - \bar{X}_{j\cdot})^2$ jsou vzájemně nezávislé, proto jsou také číselník a jmenovatel v r nezávislí. Pro hypotézu \mathcal{H}_0 sleduje číselník vydělený σ^2 chi-kvadrát rozdělení s $k-1$ stupni volnosti a jmenovatel vydělený σ^2 pak chi-kvadrát rozdělení s $n-k$ stupni volnosti. Je-li tedy \mathcal{H}_0 pravdivé, sleduje R rozdělení F s $k-1$ a $n-k$ stupni volnosti a konstanta c tak představuje $(1-\alpha)$ procentní kvantil F rozdělení s $k-1$ a $n-k$ stupni volnosti.

Výše popsany způsob testování se nazývá analýza rozptylu. Smysl tohoto pojmu je patrný, jestliže si uvědomíme, že jmenovatel r je odhadem rozptylu v rámci jednotlivých populací a číselník je odhadem rozptylu mezi populacemi za předpokladu shodné střední hodnoty. Při testování středních hodnot tak paradoxně tak analyzujeme rozptyl.

9.4.4 Testování několika rozptylů

Testování dvou rozptylů

Uvažujme náhodné výběry ze dvou populací, které sledují normální rozdělení s parametry (μ_1, σ_1^2) a (μ_2, σ_2^2) . V následujícím textu budeme uvažovat následující hypotézy

1. $\mathcal{H}_0 : \sigma_1^2 \leq \sigma_2^2$ vs. $\mathcal{H}_1 : \sigma_1^2 > \sigma_2^2$
2. $\mathcal{H}_0 : \sigma_1^2 \geq \sigma_2^2$ vs. $\mathcal{H}_1 : \sigma_1^2 < \sigma_2^2$
3. $\mathcal{H}_0 : \sigma_1^2 = \sigma_2^2$ vs. $\mathcal{H}_1 : \sigma_1^2 \neq \sigma_2^2$

Uvažujme náhodný výběr X_{11}, \dots, X_{1n_1} z normálního rozdělení se střední hodnotou μ_1 a rozptylem σ_1^2 a náhodný výběr X_{21}, \dots, X_{2n_2} z normálního rozdělení se střední hodnotou μ_2 a rozptylem σ_2^2 . Předpokládejme nezávislost těchto dvou náhodných výběrů. Víme, že poměr

$$\frac{\sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1) / (n_1 - 1) \sigma_1^2}{\sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2) / (n_2 - 1) \sigma_2^2}$$

sleduje F rozdělení s $n_1 - 1$ a $n_2 - 1$ stupni volnosti a konkrétně statistika

$$R = \frac{(n_2 - 1) \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2}{(n_1 - 1) \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2}$$

sleduje F rozdělení s $n_1 - 1$ a $n_2 - 1$ stupni volnosti pro $\sigma_1^2 = \sigma_2^2$. Statistika R bude nabývat větších hodnot pro $\sigma_1^2 > \sigma_2^2$ a menších hodnot pro $\sigma_1^2 < \sigma_2^2$, čehož lze využít při formulaci testu. Např. test $\mathcal{H}_0 : \sigma_1^2 \leq \sigma_2^2$ vs. $\mathcal{H}_1 : \sigma_1^2 > \sigma_2^2$ zamítne \mathcal{H}_0 , jestliže R přesáhne $F_\alpha(n_1 - 1, n_2 - 1)$, tj. α procentní kvantil F rozdělení s $n_1 - 1$ a $n_2 - 1$ stupni volnosti. Test $\mathcal{H}_0 : \sigma_1^2 = \sigma_2^2$ vs. $\mathcal{H}_1 : \sigma_1^2 \neq \sigma_2^2$ přijme \mathcal{H}_0 , jestliže $k_1 < R < k_2$, kde k_1 a k_2 jsou vybrány tak, aby test měl velikost α . Běžně se volí $k_1 = F_{\alpha/2}(n_1 - 1, n_2 - 1)$ a $k_2 = F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$, tj. oba chvosty mají plochu $\alpha/2$, ačkoliv se nejedná o nejlepší možný test.

Na závěr uvedme, že ke stejným výsledkům bychom dospěli také s pomocí obecného věrohodnostního poměru.

Testování rovnosti několika rozptylů

Uvažujme náhodný výběr X_{j1}, \dots, X_{jn_j} z populace, která sleduje normální rozdělení se střední hodnotou μ_j a rozptylem σ_j^2 , kde $j = 1, \dots, k$. Předpokládejme, že jednotlivé výběry jsou vzájemně nezávislé. Testujme nulovou hypotézu $\mathcal{H}_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ proti alternativní hypotéze \mathcal{H}_1 , že rozptyly jednotlivých náhodných výběrů nejsou shodné.

Funkce maximální věrohodnosti je

$$L(\mu_1, \dots, \mu_k, \sigma_1^2, \dots, \sigma_k^2; x_{11}, \dots, x_{1n_1}, \dots, x_{k1}, \dots, x_{kn_k}) = \prod_{j=1}^k \prod_{i=1}^{n_j} \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2}\left(\frac{x_{ji}-\mu_j}{\sigma_j}\right)^2}$$

a odhady pro μ_j , σ_j^2 založené na metodě maximální věrohodnosti jsou

$$\hat{\mu}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ji} = \bar{x}_j.$$

a

$$\hat{\sigma}_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_j)^2$$

Dle nulové hypotézy jsou všechny rozptyly shodné, tj. $\sigma^2 = \sigma_1^2, \dots, \sigma_k^2$. Parametrický prostor $\underline{\Theta}_0$ je tedy definován jako $\{(\mu_1, \dots, \mu_k, \sigma^2) : -\infty < \mu_j < \infty; \sigma^2 > 0\}$ a odhady pro $\mu_1, \dots, \mu_k, \sigma^2$ dle metody maximální věrohodnosti nad $\underline{\Theta}_0$ jsou

$$\hat{\mu}_j = \bar{x}_j$$

a

$$\hat{\sigma}^2 = \frac{1}{\sum_{j=1}^k n_j} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji} - \bar{x}_j)^2 = \frac{\sum_{j=1}^k n_j \hat{\sigma}_j^2}{\sum_{j=1}^k n_j}$$

Proto

$$\begin{aligned} \lambda &= \frac{\sup_{\bar{\Theta}_0} L}{\sup_{\bar{\Theta}} L} = \frac{\left(\frac{1}{\hat{\sigma}^2}\right)^{\sum_{j=1}^k n_j/k} e^{-\sum_{j=1}^k n_j/2}}{\prod_{j=1}^k \left(\frac{1}{\hat{\sigma}_j^2}\right)^{n_j/2} e^{-\sum_{j=1}^k n_j/2}} \\ &= \frac{\prod_{j=1}^k (\hat{\sigma}_j^2)^{n_j/2}}{\left(\sum_{j=1}^k n_j \hat{\sigma}_j^2 / \sum_{j=1}^k n_j\right)^{\sum_{j=1}^k n_j/2}} \end{aligned}$$

Test zamítne \mathcal{H}_0 , jestliže $\lambda \leq \lambda_0$. Nalezení konkrétní hodnoty λ_0 , která by odpovídala testu velikosti α , je však problematické, protože pravděpodobnostní rozdělení statistiky Λ nelze analyticky vyjádřit. Avšak, stejně jako v předchozím textu, lze tento problém obejít v případě výběrů dostatečně velkého rozsahu, proto které $-2 \ln(\Lambda)$ přibližně sleduje chi-kvadrát rozdělení s $k - 1$ stupni volnosti. Výše popsany test tak zamítne \mathcal{H}_0 , jestliže $-2 \ln(\lambda) > \chi_{1-\alpha}(k - 1)$, kde $\chi_{1-\alpha}(k - 1)$ představuje $1 - \alpha$ procentní kvantil chi-kvadrát rozdělení s $k - 1$ stupni volnosti.

9.5 Chi-kvadrát testy

V následujícím textu představíme testy, které jsou postaveny na chi-kvadrát rozdělení. Soustředíme se pouze na nalezení testu pro určitou hypotézu bez ohledu na otázku jejich optimality, a proto se nebudeme zabývat funkcí charakteristické veličiny.

9.5.1 Asymptotické rozdělení obecného věrohodnostního poměru

V předchozím textu jsme dvakrát konstatovali, že pravděpodobnostní rozdělení obecného věrohodnostního poměru je složité odvodit, nicméně že existuje asymptotické rozdělení, které lze použít pro jeho aproximaci. Následující věta, kterou však nebudeme dokazovat, nám dává toto asymptotické rozdělení.

Věta 9.7 *Uvažujme náhodný výběr X_1, \dots, X_n se sdruženou pravděpodobnostní funkcí $f_{X_1, \dots, X_n}(\cdot, \dots, \cdot; \theta)$, kde $\theta = (\theta_1, \dots, \theta_k)$ splňuje podmínky regularity. Předpokládejme, že parametrický prostor $\bar{\Theta}$ je k rozměrný. Pro hypotézu*

$$\mathcal{H}_0 : \theta_1 = \theta_1^0, \dots, \theta_r = \theta_r^0, \theta_{r+1}, \dots, \theta_k$$

kde $\theta_1^0, \dots, \theta_r^0$ jsou známé a $\theta_{r+1}, \dots, \theta_k$ nejsou specifikovány, sleduje $-2 \ln(\Lambda_n)$ pro pravdivé \mathcal{H}_0 přibližně chi-kvadrát rozdělení s r stupni volnosti za předpokladu dostatečně velkého náhodného výběru.

Ve výše uvedené větě jsme předpokládali, že $1 \leq r \leq k$. Jestliže $r = k$, jsou všechny parametry specifikovány. Parametrický prostor $\bar{\Theta}$ je k rozměrný, a protože \mathcal{H}_0 specifikuje r parametrů, je parametrický prostor $\bar{\Theta}_0$ $k - r$ rozměrný. Připomeňme, že Λ_n je náhodná veličina, která má hodnotu

$$\lambda_n = \frac{\sup_{\bar{\Theta}_0} L(\theta_1, \dots, \theta_k; x_1, \dots, x_n)}{\sup_{\bar{\Theta}} L(\theta_1, \dots, \theta_k; x_1, \dots, x_n)}$$

což je obecný věrohodnostní poměr náhodného výběru velikosti n . $\bar{\Theta}$ je podmnožinou $\bar{\Theta}$ specifikovanou skrze hypotézu \mathcal{H}_0 . V případě testu založeného na obecném věrohodnostním poměru zamítáme hypotézu \mathcal{H}_0 pro malá λ_n . Nicméně protože $-2\ln(\lambda_n)$ roste s klesajícím λ_n , zamítáme \mathcal{H}_0 pro vysoké hodnoty $-2\ln(\lambda_n)$. Odpovídající test o přibližné⁹ síle α tedy zamítá \mathcal{H}_0 , jestliže $-2\ln(\lambda_n) > \chi^2_{1-\alpha}(r)$, kde $\chi^2_{1-\alpha}(r)$ představuje $(1 - \alpha)$ procentní kvantil chi-kvadrát rozdělení s r stupni volnosti.

Vzhledem ke tvaru nulové hypotézy by se mohlo zdát, že výše uvedená věta má pouze omezené použití. Nicméně, jak ilustruje níže uvedené příklady, lze někdy upravit pravděpodobnostní rozdělení tak, aby bylo možné formulovat nulovou hypotézu v požadovaném tvaru.

Příklad 9.19 V kapitole (9.4.3) jsme testovali střední hodnoty dvou normálních rozdělení pomocí hypotéz $\mathcal{H}_0 : \mu_1 = \mu_2, \sigma_1^2 > 0, \sigma_2^2 > 0$ vs. $\mathcal{H}_1 : \mu_1 \neq \mu_2, \sigma_1^2 > 0, \sigma_2^2 > 0$. Parametrický prostor je čtyř rozměrný a ačkoliv se nulová hypotéza \mathcal{H}_0 na první pohled nezdá být v požadovaném tvaru, lze ji snadno upravit. Nechť $\theta_1 = \mu_1 - \mu_2, \theta_2 = \mu_2, \theta_3 = \sigma_1^2$ a $\theta_4 = \sigma_2^2$. Nulová hypotéza tak přejde do tvaru $\mathcal{H}_0 : \theta_1 = \theta_1^0 = 0, \theta_2, \theta_3, \theta_4$, tj. první parametr je roven nule a zbývající tři nejsou specifikovány. To znamená, že je-li \mathcal{H}_0 pravdivé, pak $-2\ln(\Theta')$ sleduje chi-kvadrát rozdělení s jedním stupněm volnosti, kde Θ' je statistika, která představuje obecný věrohodnostní poměr pro pozměněnou \mathcal{H}_0 . Nicméně vzhledem k invarianci funkcí odhadu založených na metodě maximální věrohodnosti je Θ' stejné jako Θ pro původní formulaci \mathcal{H}_0 .

Příklad 9.20 V kapitole (9.4.4) jsme testovali $\mathcal{H}_0 : \sigma_1^2 = \dots = \sigma_m^2, \mu_1, \dots, \mu_m$. Nulovou hypotézu \mathcal{H}_0 lze upravit pomocí substitucí

$$\theta_1 = \frac{\sigma_1^2}{\sigma_m^2}, \dots, \theta_{m-1} = \frac{\sigma_{m-1}^2}{\sigma_m^2}, \theta_m = \sigma_m^2, \theta_{m+1} = \mu_1, \dots, \theta_{m+m} = \mu_m$$

do tvaru

$$\mathcal{H}_0 : \theta_1 = 1, \dots, \theta_{m-1} = 1, \theta_m, \theta_{m+1}, \dots, \theta_{2m}$$

tj. prvních $m - 1$ parametrů je rovno jedné a zbylé parametry nejsou specifikovány. Stejně jako v předchozím případě i zde jsou obecné věrohodnostní poměry před a po transformaci shodné a statistika $-2\ln(\Theta)$, je-li \mathcal{H}_0 pravdivé, sleduje chi-kvadrát rozdělení s $m - 1$ stupni volnosti.

9.5.2 Chi-kvadrat test dobré shody

Uvažujme populaci, která sleduje multinomiální pravděpodobnostní rozdělení

$$f(x_1, \dots, x_k; p_1, \dots, p_k) = \prod_{j=1}^{k+1} p_j^{x_j}$$

kde $j = 1, \dots, k + 1$, x_j nabývá hodnot 0 nebo 1, $0 \leq p_j \leq 1, \sum x_j = 1$ a $\sum p_j = 1$. Jedná se tedy o stejnou situaci, jakou je výběr s vrácením z populace, jejíž prvky mohou být rozděleny do $k+1$ skupin. Klasickým předmětem testování je, zda-li pravděpodobnost p_j odpovídá konkrétní hodnotě¹⁰. Výsledek jednoho

⁹Připomeňme, že se jedná o aproximaci.

¹⁰Pokud bychom házeli hrací kostkou, mohli bychom se ptát, zda-li je tato kostka správně vyvážená, tj. zda-li $p_j = \frac{1}{6}$ pro $j = 1, \dots, 6$.

experimentu lze popsat pomocí náhodných veličin (X_1, \dots, X_k) , kde $X_j = 1$ jestliže výsledek experimentu odpovídá skupině j , jinak $X_j = 0$. Výsledek n pokusů tak lze popsat jako

$$(X_{11}, \dots, X_{1k}), (X_{21}, \dots, X_{2k}), \dots, (X_{n1}, \dots, X_{nk})$$

Definujeme $N_j = \sum_{i=1}^n X_{ij}$, tj. N_j představuje počet experimentů, jejichž výsledky odpovídaly kategorii j . Z kapitoly (4.1.2) víme, že náhodná veličina (N_1, \dots, N_k) sleduje multinomiální rozdělení.

Pro testování nulové hypotézy $\mathcal{H}_0 : p_j = p_j^0, j = 1, \dots, k+1$, kde p_j^0 představují známé pravděpodobnosti a jejichž součet je roven jedné, použijeme obecný věrohodnostní moment. Funkce maximální věrohodnosti je rovna

$$L = L(p_1, \dots, p_k; x_{11}, \dots, x_{1k}, \dots, x_{n1}, \dots, x_{nk}) = \prod_{i=1}^n \prod_{j=1}^{k+1} p_j^{x_{ij}}$$

Parametrický prostor $\bar{\Theta}$ je k rozměrný¹¹, zatímco parametrický prostor $\underline{\Theta}$ je jednorozměrný. Lze snadno dokázat, že L je nad $\underline{\Theta}$ maximalizováno pro

$$p_j = \sum_{i=1}^n \frac{x_{ij}}{n} = \frac{n_j}{n}$$

kde n_j je hodnota náhodné veličiny N_j . Proto

$$\sup_{\underline{\Theta}} L = \frac{1}{n^n} \prod_{j=1}^{k+1} n_j^{n_j}$$

Maximum L nad $\bar{\Theta}_0$ je rovno $\prod_{j=1}^{k+1} (p_j^0)^{n_j}$ a obecný věrohodnostní poměr je tak roven

$$\lambda = n^n \prod_{j=1}^{k+1} \prod_{j=1}^{k+1} \left(\frac{p_j^0}{n_j} \right)^{n_j}$$

Nulová hypotéza \mathcal{H}_0 je tak zamítnuta, jestliže $\lambda < \lambda_0$, kde λ je konstanta zvolená s ohledem na požadovanou velikost chyby typu I. Pro dostatečně velká n lze použít aproximaci pomocí $-2 \ln(\lambda)$, které sleduje chi-kvadrát rozdělení s k stupni volnosti. Tato aproximace je překvapivě dobrá také pro malá n za předpokladu, že $k > 0$.

Další test, který se běžně používá pro testování \mathcal{H}_0 , byl navržen Karlem Pearsonem ještě před obecnou teorií testování hypotéz. Tento test se opírá o statistiku

$$Q_k^0 = \sum_{j=1}^{k+1} \frac{(N_j - np_j^0)^2}{np_j^0}$$

která je malá pro pravdivou \mathcal{H}_0 malá a velká pro nepravdivou \mathcal{H}_0 . N_j představuje pozorovaný počet experimentů, jejichž realizace odpovídala skupině j a np_j^0 pak představuje očekávaný počet realizací za předpokladu pravdivosti \mathcal{H}_0 . Lze snadno dokázat, že

$$E[Q_k^0] = \sum_{j=1}^{k+1} \frac{1}{np_j^0} [np_j(1 - p_j) + n^2(p_j - p_j^0)^2]$$

¹¹Parametrů p_j je sice $k+1$, avšak jeden rozměr "ubírá" podmínka $\sum p_j = 1$.

kde p_j představuje skutečnou hodnotu parameteru. Jestliže je \mathcal{H}_0 pravdivé, pak $E[Q_k^0] = \sum(1 - p_j^0) = k + 1 - 1 = k$. Následující věta uvádí limitující pravděpodobnostní rozdělení pro Q_k^0 za předpokladu pravdivosti \mathcal{H}_0 .

Věta 9.8 Předpokládejme, že možné výsledky určitého experimentu je možné rozdělit do $k + 1$ vzájemně neslučitelných skupin A_1, \dots, A_{k+1} . Definujme $p_j = P[A_j]$, kde $j = 1, \dots, k + 1$. Necht v rámci n náhodných experimentů představuje N_j počet realizací, které odpovídají skupině A_j . Platí tedy $\sum_{j=1}^{k+1} N_j = n$. Pak náhodná veličina $\frac{(N_j - np_j)^2}{np_j}$ sleduje pravděpodobnostní rozdělení, které se s rostoucím n limitně blíží chi-kvadrát rozdělení s k stupni volnosti.

Důkaz 9.4 Výše uvedenou větu nedokážeme v její obecné podobě, ale pouze pro $k = 1$. Potřebujeme dokázat, že pro každé x konverguje $F_{Q_k(x)}$ k $F_{\chi^2(k)}(x)$ s tím, jak se n blíží k nekonečnu. $F_{Q_k}(\cdot)$ představuje kumulativní distribuční funkci náhodné veličiny Q_k a $F_{\chi^2(k)}(\cdot)$ představuje kumulativní distribuční funkci chi-kvadrát rozdělení s k stupni volnosti. Pro $k = 1$ platí

$$\begin{aligned} Q_k = Q_1 &= \frac{(N_1 - np_1)^2}{np_1} + \frac{(N_2 - np_2)^2}{np_2} \\ &= \frac{(N_1 - np_1)^2}{np_1} + \frac{(n - N_1 - n + np_1)^2}{n(1 - p_1)} \\ &= \frac{(N_1 - np_1)^2}{np_1(1 - p_1)} \end{aligned}$$

Víme, že N_1 sleduje binomické rozdělení s parametry n a p_1 a že $Y_n = \frac{N_1 - np_1}{\sqrt{np_1(1 - p_1)}}$ tudíž limitně sleduje normované normální rozdělení. Protože druhá mocnina normované normální veličiny sleduje chi-kvadrát rozdělení, a protože $Y_n^2 = Q_1$, sleduje také Q_1 chi-kvadrát rozdělení.

Výše uvedená věta nám poskytuje limitní rozdělení pro statistiku

$$Q_k^0 = \sum_{j=1}^{k+1} \frac{(N_j - np_j^0)^2}{np_j^0}$$

za předpokladu pravdivosti nulové hypotézy $\mathcal{H}_0 : p_j = p_j^0, j = 1, \dots, k + 1$. Test s přibližnou velikostí α tak zamítne \mathcal{H}_0 , jestliže $Q_k^0 > \chi_{1-\alpha}^2(k)$.

Ve výše uvedeném textu jsme uvedli věty (9.7) a (9.8) s jejichž pomocí lze testovat shodnou hypotézu \mathcal{H}_0 . Ačkoliv jsou odpovídající testy na první pohled rozdílné, lze dokázat, že jsou, v případě náhodného výběru velkého rozsahu, ekvivalentní.

Příklad 9.21 Při svých pokusech přírodopisec Gregor Mendel zjistil, že kuličky hrášku lze rozdělit do čtyř skupin podle barvy a tvaru a to konkrétně na (a) žlutý a kulatý, (b) zelený a kulatý, (c) žlutý a hranatý a (d) zelený a hranatý a to přibližně v poměru 9/3/3/1. Pro $n = 556$ kuliček hrášku jsme zjistily

Typ	pozorovaný počet	očekávaný počet
žlutý a kulatý	315	312.75
zelený a kulatý	108	104.25
žlutý a hranatý	101	104.25
zelený a hranatý	32	34.75

Test velikosti α zamítne nulovou hypotézu $\mathcal{H}_0 : p_1 = \frac{9}{16}, p_2 = \frac{3}{16}, p_3 = \frac{3}{16}, p_4 = \frac{1}{16}$, jestliže

$$Q_3^0 = \sum_{j=1}^4 \frac{(N_j - np_j^0)^2}{np_j^0} > \chi_{1-\alpha}^2(k) = \chi_{0.95}^2(3) = 7.81$$

Pozorované Q_3^0 má hodnotu

$$\frac{(315 - 312.75)^2}{312.75} + \frac{(108 - 104.25)^2}{104.25} + \frac{(101 - 104.25)^2}{104.25} + \frac{(32 - 34.75)^2}{34.75} \approx 0.470$$

a proto můžeme konstatovat dobrou shodu mezi modelem a skutečnými pozorováními.

Následující věta, jejíž důkaz však přesahuje záběr knihy, je zobecněním věty (9.8) pro případ, kdy pravděpodobnosti p_1 závisí na neznámém parametru.

Věta 9.9 *Uvažujme náhodný experiment, jehož realizace lze rozdělit do $k+1$ vzájemně neslučitelných skupin A_1, \dots, A_{k+1} . Definujme $p_j = P[A_j]$, $j = 1, \dots, k+1$ a předpokládejme, že p_j závisí na r neznámých parametrech $\theta_1, \dots, \theta_r$, tj. $p_j = \mathbf{p}_j(\theta_1, \dots, \theta_r)$, $j = 1, \dots, k+1$. Nechť pro n vzájemně nezávislých pokusů představuje N_j počet realizací, které náleží do skupiny A_j , $j = 1, \dots, k+1$, přičemž $\sum N_j = n$. Dále nechť $\hat{\Theta}_1, \dots, \hat{\Theta}_r$ jsou BAN funkce odhadu¹² pro $\theta_1, \dots, \theta_r$, které jsou založené na N_1, \dots, N_k . Pak, jestliže \mathbf{p}_j splňují podmínky regularity, limitně sleduje*

$$Q'_k = \sum_{j=1}^{k+1} \frac{(N_j - n\hat{P}_j)^2}{n\hat{P}_j}$$

chi-kvadrát rozdělení s $k-r$ stupni volnosti, kde $\hat{P}_j = \mathbf{p}_j(\hat{\Theta}_1, \dots, \hat{\Theta}_r)$, $j = 1, \dots, k+1$.

Výše uvedená věta neobsahuje žádnou explicitní zmínku o testování hypotéz. V následujícím textu si ukážeme, jak lze závěry této věty použít pro formulaci testu dobré shody. Předpokládejme, že potřebujeme otestovat, zda-li náhodný výběr X_1, \dots, X_n pochází z populace $f(x; \theta_1, \dots, \theta_r)$, kde parametry $\theta_1, \dots, \theta_r$ jsou neznámé, avšak funkční předpis f známe. Nulová hypotéza má charakter složené hypotézy $\mathcal{H}_0 : X_i$ má pravděpodobnostní funkci $f(x; \theta_1, \dots, \theta_r)$ pro daná $\theta_1, \dots, \theta_r$. Nulová hypotéza tedy tvrdí, že náhodný výběr pochází z populace $f(\cdot; \theta_1, \dots, \theta_r)$. Dále předpokládejme, že hodnoty, kterých může náhodná veličina X_i nabývat, jsou rozděleny na $k+1$ vzájemně neslučitelných skupin A_1, \dots, A_{k+1} , že $p_j = P[X_i \in A_j]$ a že N_j je počet X_i náležících do A_j . Za předpokladu pravdivosti \mathcal{H}_0 a pro dostatečně velké n pak z věty (9.9) vyplývá, že náhodná veličina

$$Q'_k = \sum_{j=1}^{k+1} \frac{(N_j - n\hat{P}_j)^2}{n\hat{P}_j}$$

¹²Např. funkce odhadu založené na metodě maximální věrohodnosti.

přibližně sleduje chi-kvadrát rozdělení s $k - r$ stupni volnosti, kde $\hat{P}_j = \mathbf{p}_j(\hat{\Theta}_1, \dots, \hat{\Theta}_r)$ a $\hat{\Theta}_i$ je funkce odhadu dle metody maximální věrohodnosti pro $\theta_i, i = 1, \dots, r$, která je založená na statistikách N_1, \dots, N_k ¹³. Test tak zamítne \mathcal{H}_0 pro $Q'_k > \chi^2_{1-\alpha}(k-r)$, kde $\chi^2_{1-\alpha}(k-r)$ je $1-\alpha$ procentní kvantil chi-kvadrát rozdělení s $k - r$ stupni volnosti. Tento test nazýváme testem dobré shody, protože testuje, zda-li jsou pozorování konzistentní s předpokladem, že pocházejí z pravděpodobnostního rozdělení $f(x; \theta_1, \dots, \theta_r)$, tj. že se s tímto pravděpodobnostním rozdělením shodují.

Parametry $\theta_1, \dots, \theta_r$ jsme namísto pozorování X_1, \dots, X_n odhadli pomocí statistik N_1, \dots, N_k . Tyto parametry je však možné odhadnout také pomocí funkce odhadu dle metody maximální věrohodnosti s využitím X_1, \dots, X_n . Limitní pravděpodobnostní rozdělení náhodné veličiny Q'_k tak již nesleduje chi-kvadrát rozdělení s $k - r$ stupni volnosti, ale je ohraničena chi-kvadrát rozdělením s $k - r$ stupni volnosti a chi-kvadrát rozdělením s k stupni volnosti.

Příklad 9.22 Uvažujme náhodný výběr X_1, \dots, X_n . Testujme, zda-li tento náhodný výběr pochází z normálního rozdělení.

Rozdělme pozorování x_1, \dots, x_n do $k+1$ skupin, kde j -tá skupina představuje všech pozorování, která spadají do intervalu $(z_{j-1}, z_j]$, $j = 1, \dots, k+1$ pro $z_0 < z_1 < \dots < z_k < z_{k+1}$, kde $z_0 = -\infty$ a $z_{k+1} = \infty$. Pak platí

$$p_j = \mathbf{p}_j(\mu, \sigma^2) = \int_{z_{j-1}}^{z_j} \phi_{\mu, \sigma^2}(x) dx = \Phi\left(\frac{z_j - \mu}{\sigma}\right) - \Phi\left(\frac{z_{j-1} - \mu}{\sigma}\right)$$

Nechť jsou $\hat{\mu}$ a $\hat{\sigma}$ odhady dle metody maximální věrohodnosti založené na n_1, \dots, n_k , kde n_j představuje počet pozorování v j -tém intervalu. Pak lze

$$\hat{p}_j = \Phi\left(\frac{z_j - \hat{\mu}}{\hat{\sigma}}\right) - \Phi\left(\frac{z_{j-1} - \hat{\mu}}{\hat{\sigma}}\right)$$

vypočíst na základě pozorování stejně jako hodnota

$$q'_k = \sum_{j=1}^{k+1} \frac{(n_j - n\hat{p}_j)^2}{n\hat{p}_j}$$

náhodné veličiny Q'_k . Nulová hypotéza, že vzorek pochází z normálního rozdělení bychom na hladině α zamítly, jestliže $q'_k > \chi^2_{1-\alpha}(k-2)$.

Jestliže bychom μ a σ odhadli dle metody maximální věrohodnosti na základě x_1, \dots, x_n , pak by asymptotické rozdělení náhodné veličiny Q'_k bylo ohraničeno chi-kvadrát rozdělením s $k-2$ stupni volnosti a chi-kvadrát rozdělením s k stupni volnosti. Nulovou hypotézu bychom tedy zamítli, jestliže $q'_k > C$, kde C se nachází mezi $\chi^2_{1-\alpha}(k-2)$ a $\chi^2_{1-\alpha}(k)$. Na závěr dodejme, že pro velké k je rozdíl mezi $\chi^2_{1-\alpha}(k-2)$ a $\chi^2_{1-\alpha}(k)$ zanedbatelný.

9.5.3 Test shodnosti dvou multinomiálních rozdělení

Uvažujme dvě multinomiální rozdělení, jejichž realizace možné rozdělit do $k+1$ skupin. Nechť jsou pravděpodobnosti odpovídající prvnímu z rozdělení $p_{11}, \dots, p_{1k}, p_{1,k+1}$ a pravděpodobnosti odpovídající druhému rozdělení

¹³Připomeňme, že $\mathbf{p}_j(\theta_1, \dots, \theta_r) = P[X_i \in A_j]$, což je v případě spojité náhodné veličiny X_i rovno $\int_{A_j} f(x; \theta_1, \dots, \theta_r) dx$.

pak $p_{21}, \dots, p_{2k}, p_{2,k+1}$. Nulová hypotéza má tvar $\mathcal{H}_0 : p_{1j} = p_{2j} = p_j$ pro $j = 1, \dots, k+1$. Předpokládejme, že první náhodný výběr má velikost n_1 a druhý náhodný výběr pak velikost n_2 . Definujme N_{1j} jako počet realizací v j -té skupině prvního výběru a N_{2j} jako počet realizací v j -té skupině druhého výběru, kde $j = 1, \dots, k+1$. Víme, že statistika

$$\sum_{j=1}^{k+1} \frac{(N_{ij} - n_i p_{ij})^2}{n_i p_{ij}}$$

limitně sleduje chi-kvadrát rozdělení s k stupni volnosti pro $i = 1, 2$. Proto, jsou-li náhodné výběry vzájemně nezávislé a \mathcal{H}_0 pravdivá, sleduje

$$Q_{2k} \sum_{i=1}^2 \sum_{j=1}^{k+1} \frac{(N_{ij} - n_i p_{ij})^2}{n_i p_{ij}}$$

limitně chi-kvadrát rozdělení $2k$ stupni volnosti. Jestliže \mathcal{H}_0 specifikuje hodnotu p_j , pak je Q_{2k} statistikou a její hodnotu lze použít přímo při konstrukci testu. Jestliže však \mathcal{H}_0 hodnotu p_j nespecifikuje, pak musí být p_j odhadnuto. Jestliže je \mathcal{H}_0 pravdivá, pak lze oba náhodné výběry považovat za jeden náhodný výběr z multinomiálního rozdělení s pravděpodobnostmi p_1, \dots, p_{k+1} . Funkce odhadu dle metody maximální věrohodnosti je $p_j = \frac{N_{1j} + N_{2j}}{n_1 + n_2}$, $j = 1, \dots, k+1$, čímž získáme

$$Q'_{2k} = \sum_{i=1}^2 \sum_{j=1}^{k+1} \frac{\left(\frac{N_{ij} - n_i(N_{1j} + N_{2j})}{n_1 + n_2} \right)^2}{\frac{n_i(N_{1j} + N_{2j})}{n_1 + n_2}}$$

Lze dokázat, že Q'_{2k} limitně sleduje chi-kvadrát rozdělení s $2k - k = k$ stupni volnosti¹⁴.

Alternativní test shody dvou multinomiálních rozdělení lze odvodit pomocí obecného věrohodnostního poměru Λ s využitím věty (9.7) a limitního rozdělení $-2 \ln(\Lambda)$.

Příklad 9.23 Předpokládejme, že provádíme výzkum veřejného mínění a zajímá nás, zda-li voliči pod 25 let zastávají odlišné názory než voliči nad 25 let. Dále předpokládejme, že bylo osloveno 1000 voličů do 25 let a 1500 voličů nad 25 let a že jejich odpovědi na jednu konkrétní otázku shrnuje následující tabulka.

skupina	proti	neví	pro	celkem
pod 25 let	400	100	500	1000
nad 25 let	600	400	500	1500
celkem	1000	500	1000	2500

Nulová hypotéza zní, že neexistuje rozdíl mezi voliči do a nad 25 let, tj. $\mathcal{H}_0 : p_{1j} = p_{2j} = p_j$, $j = 1, 2, 3$. Parametry p_1 a p_2 je třeba odhadnout¹⁵. Hodnota

¹⁴Počet stupňů volnosti limitního pravděpodobnostního rozdělení pro Q'_{2k} byl snížen o jedna pro každý odhadovaný parametr.

¹⁵Hodnota parametru p_3 je pak dána vztahem $p_1 + p_2 + p_3 = 1$.

testovací statistiky je rovna

$$\begin{aligned} & \frac{(400 - 1000 \cdot 1000/2500)^2}{1000 \cdot 1000/2500} + \frac{(100 - 1000 \cdot 500/2500)^2}{1000 \cdot 500/2500} \\ & + \frac{(500 - 1000 \cdot 1000/2500)^2}{1000 \cdot 1000/2500} + \frac{(600 - 1500 \cdot 1000/2500)^2}{1500 \cdot 1000/2500} \\ & + \frac{(400 - 1500 \cdot 500/2500)^2}{1500 \cdot 500/2500} + \frac{(500 - 1500 \cdot 1000/2500)^2}{1500 \cdot 1000/2500} = 125 \end{aligned}$$

99% kvantil chi-kvadrát rozdělení se dvěma stupni volnosti je pouze 9.21, a proto lze konstatovat, že se volební preference voličů do 25 let výrazně liší od volebních preferencí nad 25 let.

Výše uvedený postup lze zobecnit ve dvou směrech. V první rovině lze postup zobecnit ze dvou na vícero multinomiálních rozdělení. Ve druhé rovině, jak ilustruje následující příklad, pak lze postup zobecnit z multinomiálního na pravděpodobnostní rozdělení jiného typu.

Příklad 9.24 Uvažujme dva náhodné výběry z Poissonova rozdělení, které shrnuje následující tabulka.

skupina	0	1	2	3	4	5	6	7	8	9 a více	celkem
výběr 1	11	25	28	20	9	3	3	0	1	0	100
výběr 2	13	27	28	17	11	1	2	1	0	0	100
celkem	24	52	56	37	20	11					200

Defnujme nulovou hypotézu, že oba náhodné výběry pochází z téhož Poissonova rozdělení. Rozdělme pozorování do šesti skupin, kdy poslední sdružuje všechny realizace s hodnotou větší než 4. Za předpokladu pravdivosti nulové hypotézy stačí odhadnout jeden parametr a to střední hodnotu společného Poissonova rozdělení. Odhad dle metody maximální věrohodnosti je

$$\frac{0 \cdot 24 + 1 \cdot 52 + 2 \cdot 56 + 3 \cdot 37 + 4 \cdot 20 + 5 \cdot 4 + 6 \cdot 5 + 7 \cdot 1 + 8 \cdot 1}{200} = \frac{420}{200} = 2.1$$

Očekávaný počet pozorování pro každou skupinu v rámci jedné populace je tak

0	1	2	3	4	5 a více
12.25	25.72	27.00	18.90	9.92	6.21

S pomocí těchto údajů lze spočítat hodnotu statistiky Q_{2k} , kde $n_i p_j$ nahradíme hodnotami z výše uvedené tabulky. Hodnota této statistiky je přibližně 1.68. Protože jsme odhadovali jeden parametr, je počet stupňů volnosti roven $2k - 1$, tj. 9. Není tedy důvod se domnívat, že oba náhodné výběry pochází z rozdílných Poissonových rozdělení.

9.5.4 Test nezávislosti v kontingčních tabulkách

Kontingenční tabulky lze chápat jako vícenásobnou klasifikaci, jako je např. rozdělení dotazovaných podle jejich věku, pohlaví a výše příjmu. Pokud bychom

definovali čtyři a tři příjmové kategorie, získali bychom kontingenční tabulku typu $4 \times 2 \times 3$, a každého z respondentů by bylo možné zařadit do jedné z výsledných 24 kategorií. Kontingenční tabulky jsou pak vhodné pro následnou analýzu případných souvislostí, jako je např. závislost příjmu na věku a pohlaví.

Dvourozměrné kontingenční tabulky

Předpokládejme, že n osob lze roztrždit na základě kritérií A a B , kde A se skládá z kategorií A_1, \dots, A_r a B se skládá z kategorií B_1, \dots, B_s . Počet jedinců, kteří současně náležejí do A_i a B_j označme N_{ij} . Je zřejmé, že $\sum N_{ij} = n$.

	B_1	B_2	B_3	\dots	B_s
A_1	N_{11}	N_{12}	N_{13}	\dots	N_{1s}
A_2	N_{21}	N_{22}	N_{23}	\dots	N_{2s}
A_3	N_{31}	N_{32}	N_{33}	\dots	N_{3s}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
A_r	N_{r1}	N_{r2}	N_{r3}	\dots	N_{rs}

V následujícím textu budeme řádkové součty značit jako $N_{i\cdot}$ a sloupcové jako $N_{\cdot j}$, tj.

$$N_{i\cdot} = \sum_j N_{ij} \quad N_{\cdot j} = \sum_i N_{ij} \quad \sum_i N_{i\cdot} = \sum_j N_{\cdot j} = n$$

Námi uvažovaných n jedinců lze chápat jako náhodný výběr z multinomiálního rozdělení s pravděpodobnostmi p_{ij} , kde $i = 1, \dots, r$ a $j = 1, \dots, s$. Pravděpodobnostní funkce jednoho pozorování je tak

$$f(x_{11}, x_{12}, \dots, x_{rs}; p_{11}, \dots, p_{rs}) = \prod_{i,j} p_{ij}^{x_{ij}}$$

kde $x_{ij} = 0$ popř. $x_{ij} = 1$ a $\sum_{i,j} x_{ij} = 1$.

Definujme nulovou hypotézu jako vzájemnou nezávislost kritéria A a B , tj.

$$P[B_j|A_i] = P[B_j] \text{ a } P[A_i|B_j] = P[A_i]$$

nebo také

$$P[A_i \cap B_j] = P[A_i]P[B_j]$$

Jestliže marginální pravděpodobnosti $P[A_i]$ označíme jako $p_{i\cdot}$, kde $i = 1, \dots, r$, a marginální pravděpodobnosti $P[B_j]$ jako $p_{\cdot j}$, kde $j = 1, \dots, s$, lze nulovou hypotézu formulovat jako

$$\mathcal{H}_0 : p_{ij} = p_{i\cdot}p_{\cdot j} \quad \sum_{i=1}^r p_{i\cdot} = 1, \sum_{j=1}^s p_{\cdot j} = 1$$

Parametrický prostor $\overline{\Theta}$ distribuce N_{11}, \dots, N_{rs} má $rs - 1$ rozměrů¹⁶, zatímco pro \mathcal{H}_0 má parametrický prostor $\overline{\Theta}_0$ $r - 1 + s - 1$ rozměrů¹⁷. Funkce maximální věrohodnosti pro náhodný výběr velikosti n je

$$L = \prod_{i,j} p_{ij}^{n_{ij}}$$

¹⁶Jedna z marginálních pravděpodobností p_{ij} je dána rovností $\sum_{i,j} p_{ij} = 1$ a zbývajících $rs - 1$ marginálních pravděpodobnostmi.

¹⁷Nulová hypotéza je sice specifikována pravděpodobnostmi $p_{i\cdot}, i = 1, \dots, r$ a $p_{\cdot j}, j = 1, \dots, s$, ale počet dimenzí je $r - 1 + s - 1$, protože $\sum p_{i\cdot} = 1$ a $\sum p_{\cdot j} = 1$.

a její maximum nad $\bar{\Theta}$ nastává pro $\hat{p}_{ij} = \frac{n_{ij}}{n}$. Pro $\bar{\Theta}_0$ je funkce maximální věrohodnosti definována jako

$$L = \prod_{i,j} (p_{i\cdot}, p_{\cdot j})^{n_{ij}} = \left(\prod_i p_{i\cdot}^{n_{i\cdot}} \right) \left(\prod_j p_{\cdot j}^{n_{\cdot j}} \right)$$

s maximem v $\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n}$ a $\hat{p}_{\cdot j} = \frac{n_{\cdot j}}{n}$. Obecný věrohodnostní poměr je tak

$$\lambda = \frac{(\prod_i n_{i\cdot}^{n_{i\cdot}}) (\prod_j n_{\cdot j}^{n_{\cdot j}})}{n^n \prod_{i,j} n_{ij}^{n_{ij}}}$$

Pravděpodobnostní rozdělení Λ pro nulovou hypotézu není jedinečné, protože se jedná o složenou hypotézu a pravděpodobnostní rozdělení Λ je funkcí neznámých parametrů $p_{i\cdot}$ a $p_{\cdot j}$. Proto je velmi složité nalézt λ_0 , které je řešením $\sup_{\bar{\Theta}_0} P_\theta[\Lambda \leq \lambda_0] = \alpha$. Pro dostatečně velké náhodné výběry však lze opět použít statistiku $-2 \ln(\Theta)$, která přibližně sleduje chi-kvadrát rozdělení s $rs - 1 - (r + s - 2) = (r - 1)(s - 1)$ stupni volnosti. S její pomocí lze určit jedinečný kritický region pro λ .

Formulace nulové hypotézy $\mathcal{H}_0 : p_{ij} = p_{i\cdot} p_{\cdot j}$ není v souladu s požadavky věty (9.7), nicméně ji lze přeformulovat. Pro zjednodušení předpokládejme $r = s = 2$. V tomto případě $\bar{\Theta} = \{\theta_1, \theta_2, \theta_3\} = (p_{11}, p_{12}, p_{21}) : p_{11} \geq 0, p_{12} \geq 0, p_{21} \geq 0, p_{11} + p_{12} + p_{21} \leq 1$. Necht' $\bar{\Theta}' = \{\theta'_1, \theta'_2, \theta'_3\}$ označuje nový parametrický prostor, kde $\theta'_1 = p_{11} - p_{1\cdot} p_{\cdot 1}, \theta'_2 = p_{12} - p_{1\cdot} p_{\cdot 2}$ a $\theta'_3 = p_{21} - p_{2\cdot} p_{\cdot 1}$. Lze dokázat, že $\bar{\Theta}'$ je transformací jedna ku jedné parametrického prostoru $\bar{\Theta}$. Původní nulová hypotéza $\mathcal{H}_0 : p_{11} = p_{1\cdot} p_{\cdot 1}, p_{12} = p_{1\cdot} (1 - p_{\cdot 1})$ pro parametrický prostor $\bar{\Theta}$ se tak změni na $\mathcal{H}'_0 : \theta'_1 = 0$ a nespecifikované θ'_2 a θ'_3 pro parametrický prostor $\bar{\Theta}'$ ¹⁸.

Obecně lze bod $rs - 1$ rozměrného parametrického prostoru $\bar{\Theta}$ vyjádřit jako

$$\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1,s-1} & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2,s-1} & p_{2s} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{r-1,1} & p_{r-1,2} & \cdots & p_{r-1,s-1} & p_{r-1,s} \\ p_{r1} & p_{r2} & \cdots & p_{r,s-1} & p_{rs} \end{pmatrix}$$

a odpovídající bod v parametrickém prostoru $\bar{\Theta}'$ pak jako

$$\begin{pmatrix} p_{11} - p_{1\cdot} p_{\cdot 1} & p_{12} - p_{1\cdot} p_{\cdot 2} & \cdots & p_{1,s-1} - p_{1\cdot} p_{\cdot,s-1} & p_{1s} - p_{1\cdot} p_{\cdot s} \\ p_{21} - p_{2\cdot} p_{\cdot 1} & p_{22} - p_{2\cdot} p_{\cdot 2} & \cdots & p_{2,s-1} - p_{2\cdot} p_{\cdot,s-1} & p_{2s} - p_{2\cdot} p_{\cdot s} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{r-1,1} - p_{r-1\cdot} p_{\cdot 1} & p_{r-1,2} - p_{r-1\cdot} p_{\cdot 2} & \cdots & p_{r-1,s-1} - p_{r-1\cdot} p_{\cdot,s-1} & p_{r-1,s} - p_{r-1\cdot} p_{\cdot s} \\ p_{r1} - p_{r\cdot} p_{\cdot 1} & p_{r2} - p_{r\cdot} p_{\cdot 2} & \cdots & p_{r,s-1} - p_{r\cdot} p_{\cdot,s-1} & p_{rs} - p_{r\cdot} p_{\cdot s} \end{pmatrix}$$

Přirozená otázka při formulaci testu zní, jak je možné, že kritický region pro dostatečně velký náhodný výběr je jednoznačný a pro náhodný výběr malého rozsahu nikoliv, ačkoliv oba tyto regiony závisí na stejné sadě neznámých parametrů. Ve skutečnosti tyto parametry však neznámé nejsou - lze je odhadnout z pozorování a tyto odhady se blíží jejich skutečným hodnotám s tím, jak

¹⁸Výraz $p_{12} = p_{1\cdot} (1 - p_{\cdot 1})$ je ekvivalentní s $p_{11} - p_{1\cdot} p_{\cdot 1}$, což je ekvivalentní s $p_{11} - p_{1\cdot} p_{\cdot 1} = 0$. Podobně platí $p_{21} = (1 - p_{1\cdot}) p_{\cdot 1}$.

roste velikost náhodného výběru. V limitě, kdy se velikost výběru blíží nekonečnu, lze parametry odhadnout přesně a pravděpodobnostní funkce Λ je dána jednoznačně.

Mohlo by zdát racionální použít podobný přístup také v případě náhodných výběru malého rozsahu, tj. definovat pravděpodobnostní rozdělení Λ s pomocí odhadů neznámých parametrů. Odhady p_i resp. p_j jsou $\hat{p}_i = \frac{n_{i\cdot}}{n}$ resp. $\hat{p}_j = \frac{n_{\cdot j}}{n}$ a lze je přímo dosadit do distribuční funkce N_{ij} a tu využít pro získání pravděpodobnostního rozdělení Λ . Nicméně kritický region závisí na marginálních součtech $N_{i\cdot}$ a $N_{\cdot j}$ a pravděpodobnost chyby typu I se tak liší od náhodného výběru k náhodnému výběru pro libovolný fixní kritický region $0 < \lambda < \lambda_0$.

Existuje způsob, jak tento problém vyřešit. Označme sdruženou pravděpodobnostní funkci pro všechna N_{ij} jako $f(n_{ij})$, sdruženou marginální pravděpodobnostní funkci pro všechna $N_{i\cdot}$ a $N_{\cdot j}$ jako $g(n_{i\cdot}, n_{\cdot j})$ a podmíněnou pravděpodobnostní funkci N_{ij} pro dané $N_{i\cdot}$ a $N_{\cdot j}$ jako $f(n_{ij}|n_{i\cdot}, n_{\cdot j}) = \frac{f(n_{ij})}{g(n_{i\cdot}, n_{\cdot j})}$. Pro uvažovanou nulovou hypotézu je tato podmíněná pravděpodobnost nezávislá na neznámých parametrech a funkce odhadu $N_{i\cdot}/n$ a $N_{\cdot j}/n$ představují soubor dostatečných statistik pro p_i a p_j . Tato skutečnost, kterou dokážeme níže, nám umožní zkonstruovat test.

Sdružená pravděpodobnostní funkce pro N_{ij} má tvar multinomiálního rozdělení

$$f(n_{ij}) = f(n_{11}, n_{12}, \dots, n_{rs}) = \frac{n!}{\prod_{i,j} n_{ij}!} \prod_{i,j} p_{ij}^{n_{ij}}$$

pro $\bar{\Theta}$ a pro $\bar{\Theta}_0$ pak tvar¹⁹

$$f(n_{11}, n_{12}, \dots, n_{rs}) = \frac{n!}{\prod_{i,j} n_{ij}!} \left(\prod_i p_{i\cdot}^{n_{i\cdot}} \right) \left(\prod_j p_{\cdot j}^{n_{\cdot j}} \right) \quad (9.1)$$

Abychom získali požadovanou podmíněnou pravděpodobnost, musíme nejprve nalézt pravděpodobnostní rozdělení pro $N_{i\cdot}$ a $N_{\cdot j}$, čehož lze dosáhnout sumou (9.1) přes všechny soubory n_{ij} takové, že

$$\sum_i n_{ij} = n_{\cdot j} \text{ a } \sum_j n_{ij} = n_{i\cdot}. \quad (9.2)$$

Pro fixní $N_{i\cdot}$ a $N_{\cdot j}$ je předmětem tohoto součtu pouze člen $1/\prod n_{ij}!$ v (9.1). Jedná se tedy vlastně o součet tohoto členu pro všechna n_{ij} , která splňují podmínky (9.2). Požadovaný součet lze získat porovnáním koeficientů členu $\prod_i x_i^{n_{i\cdot}}$ ve výrazu

$$(x_1 + \dots + x_r)^{n_{\cdot 1}} (x_1 + \dots + x_r)^{n_{\cdot 2}} \dots (x_1 + \dots + x_r)^{n_{\cdot s}} = (x_1 + \dots + x_r)^n \quad (9.3)$$

Na pravé straně tohoto výrazu je koeficient $\prod_i x_i^{n_{i\cdot}}$ roven

$$\frac{n!}{\prod_i n_{i\cdot}!} \quad (9.4)$$

a na jeho levé straně pak

$$\frac{n_{\cdot 1}!}{\prod_i n_{i1}!} \frac{n_{\cdot 2}!}{\prod_i n_{i2}!} \dots \frac{n_{\cdot s}!}{\prod_i n_{is}!} = \frac{\prod_j n_{\cdot j}!}{\prod_{i,j} n_{ij}!} \quad (9.5)$$

¹⁹Připomeňme, že se zajímáme o pravděpodobnostní rozdělení Λ za předpokladu pravdivosti \mathcal{H}_0 .

kde n_{ij} představuje exponent x_i v j -tém činiteli. n_{ij} splňuje obě podmínky (9.2) - splnění první podmínky vyplývá z věty multinominálního rozdělení a splnění druhé vyplývá ze skutečnosti, že exponent x_i je $n_{i\cdot}$. Je zřejmé, že součet dle (9.5) musí být roven (9.4), a proto platí

$$\sum \frac{1}{\prod n_{ij}!} = \frac{n!}{\left(\prod_i n_{i\cdot}! \prod_j n_{\cdot j}!\right)}$$

To je přesně součet, který hledáme, protože pro každý prvek kontingenční tabulky s danými marginálními součty existuje právě jeden koeficient ve tvaru (9.5), který se nachází na levé straně (9.3). Pravděpodobnostní rozdělení $N_{i\cdot}$ a $N_{\cdot j}$ je tedy

$$g(n_{i\cdot}, n_{\cdot j}) = \frac{(n!)^2}{(\prod_i n_{i\cdot}!) (\prod_j n_{\cdot j}!)} \left(\prod_i p_{i\cdot}^{n_{i\cdot}}\right) \left(\prod_j p_{\cdot j}^{n_{\cdot j}}\right) \quad (9.6)$$

z čehož shodou okolností vyplývá, že $N_{i\cdot}$ a $N_{\cdot j}$ jsou, za předpokladu pravdivosti \mathcal{H}_0 , vzájemně nezávislé. To je poněkud překvapivé, jestliže si uvědomíme, že např. $N_{1\cdot}$ a $N_{\cdot 1}$ sdílí náhodnou veličinu N_{11} . Podmíněná pravděpodobnost N_{ij} pro dané $N_{i\cdot}$ a $N_{\cdot j}$ je pak dána podílem (9.1) a (9.6), tj.

$$f(n_{11}, n_{12}, \dots, n_{rs} | n_{1\cdot}, n_{2\cdot}, \dots, n_{\cdot s}) = \frac{(\prod_i n_{i\cdot}!) (\prod_j n_{\cdot j}!)}{n! \prod n_{ij}!} \quad (9.7)$$

a neobsahuje žádné neznámé parametry, a proto jsou funkce odhadu N_{\cdot}/n a $N_{\cdot j}/n$ dostatečné.

Abychom ukázali, jak lze zkonstruovat test, uvažujme obecnou situaci, kdy má statistika Λ pravděpodobnostní rozdělení $f_{\Lambda}(\lambda; \theta)$, které zahrnuje neznámý parametr θ . Jestliže má θ dostatečnou statistiku T , pak lze sdruženou pravděpodobnost Λ a T vyjádřit jako

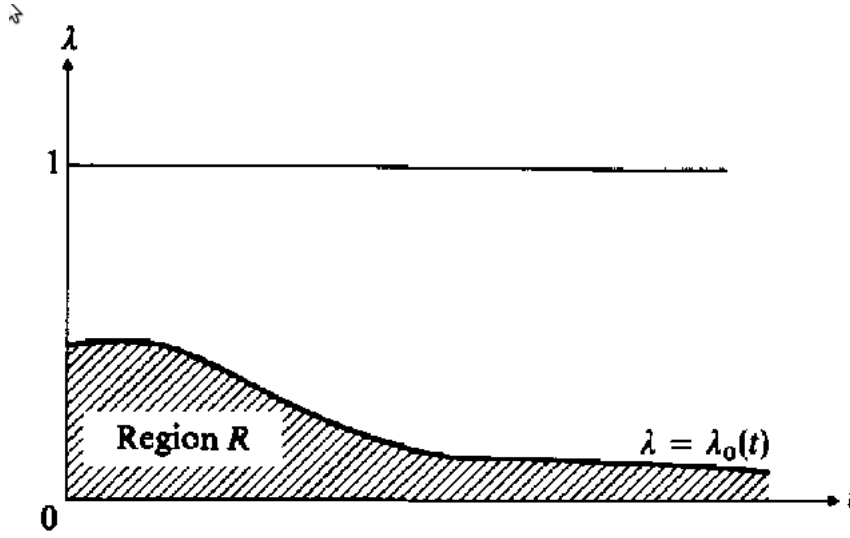
$$f_{\Lambda, T}(\lambda, t; \theta) = f_{\Lambda|T}(\lambda|t) f_T(t; \theta)$$

kde podmíněná pravděpodobnost náhodné veličiny Λ pro dané T nezávisí na θ . S využitím podmíněné pravděpodobnosti lze pro každé t nalézt $\lambda_0(t)$ takové, že

$$\int_0^{\lambda_0(t)} f_{\Lambda|T}(\lambda|t) d\lambda = \alpha$$

V rovině $\lambda \times t$ křivka $\lambda = \lambda_0(t)$ společně s přímkou $\lambda = 0$ vymezují kritický region R , jak ilustruje obrázek (9.6). Pravděpodobnost, že prvek náhodného výběru odpovídá páru hodnot (λ, t) , který spadá pod region R je právě α , protože

$$\begin{aligned} P[(\Lambda, T) \in R] &= \int_{-\infty}^{\infty} \int_0^{\lambda_0(t)} f_{\Lambda, T}(\lambda, t; \theta) d\lambda dt \\ &= \int_{-\infty}^{\infty} \left(\int_0^{\lambda_0(t)} f_{\Lambda|T}(\lambda|t) d\lambda \right) f_T(t; \theta) dt \\ &= \int_{-\infty}^{\infty} \alpha f_T(t; \theta) dt \\ &= \alpha \end{aligned}$$

Obrázek 9.6: Kritický region R vymezený $\lambda = \lambda_0(t)$ a $\lambda = 0$

Proto můžeme testovat hypotézu na základě Λ podmíněného T - nejprve určíme hodnotu T a následně provede test s pomocí intervalu $0 < \lambda < \lambda_0(t)$ s využitím pravděpodobnosti Λ podmíněné T . Kritický region tak bude mít podobu plochy namísto intervalu $0 < \lambda < \lambda_0$. Tento region je navíc zkonstruován tak, že bez ohledu na skutečnou hodnotu θ , má chyba typu I stále pravděpodobnost realizace α . Připomeňme, že abychom byli schopni aplikovat tento postup, musí existovat dostatečná statistika pro θ .

S využitím výše uvedeného obecného věrohodnostního poměru

$$\lambda = \frac{(\prod_i n_{i\cdot}^{n_{i\cdot}}) (\prod_j n_{\cdot j}^{n_{\cdot j}})}{n^n \prod_{i,j} n_{ij}^{n_{ij}}}$$

a podmíněných pravděpodobností (9.7) určit interval $0 < \lambda < \lambda_0(n_{i\cdot}; n_{\cdot j})$, který má požadovanou pravděpodobnost chyby typu I pro dané marginální součty. Při aplikaci tohoto testu však narazíme na velmi pracné odvození pravděpodobnostního rozdělení Λ pokud r, s a marginální součty nejsou relativně malé. Pro dostatečně velké náhodné výběry však lze použít aproximaci s výjimkou případů, kdy r a s jsou rovny 2, kdy je možné použít specifickou alternativní aproximaci. Tato problematika však přesahuje záběr knihy.

Další test nulové hypotézy $\mathcal{H}_0 : p_{ij} = p_{i\cdot}p_{\cdot j}; \sum p_{i\cdot} = 1, \sum p_{\cdot j} = 1$ lze odvodit nahrazením pravděpodobnostního rozdělení v (9.7) vícerozměrnou normální aproximací, kdy lze dokázat, že statistika

$$Q = \sum_{i,j} \frac{[N_{ij} - n(N_{i\cdot}/n)(N_{\cdot j}/n)]^2}{n(N_{i\cdot}/n)(N_{\cdot j}/n)}$$

přibližně sleduje chi-kvadrát rozdělení s $rs - 1 - (r - 1 + s - 1) = (r - 1)(s - 1)$ stupni volnosti. Test zamítne \mathcal{H}_0 pro velká Q . Toto kritérium bylo poprvé navrženo Karlem Pearsonem a od $-2 \ln(\lambda)$ se liší členem řádu $1/\sqrt{n}$. Oba testy

jsou tak pro dostatečně velké náhodné výběry v podstatě totožné. Statistika Q má intuitivní základ. N_{ij} představuje počet pozorování v ij -té buňce dvourozměrné kontingenční tabulky a $n(N_{i\cdot}/n)(N_{\cdot j}/n)$ představuje očekávaný počet pozorování v ij -té buňce za předpokladu pravdivosti \mathcal{H}_0 . Proto dosahuje statistika Q nízkých hodnot pro pravdivé \mathcal{H}_0 a vysokých hodnot pro nepravdivé \mathcal{H}_0 .

Třírozměrné kontingenční tabulky

Jestliže lze každý prvek náhodného výběru klasifikovat podle tří kritérií, řekněme A , B a C , jsme schopni tyto prvky uspořádat do tzv. třírozměrné kontingenční tabulky. Jestliže kritérium A nabývá hodnot $A_i, i = 1, \dots, s_1$, kritérium B hodnot $B_j, j = 1, \dots, s_2$ a kritérium C hodnot $C_k, k = 1, \dots, s_3$ pak lze tuto kontingenční tabulku chápat jako kvádr, který se skládá z $s_1 \times s_2 \times s_3$ buněk. Pravděpodobnost, že prvek náhodného výběru připadne ijk -té buňce označme jako p_{ijk} a počet prvků v této buňce pak jako N_{ijk} . Marginální součty budeme, stejně jako v případě dvourozměrné kontingenční tabulky, značit tak, že příslušný index nahradíme znakem (\cdot) , tj.

$$N_{i\cdot k} = \sum_{j=1}^{s_2} N_{ijk} \text{ a } N_{\cdot\cdot k} = \sum_{i=1}^{s_1} \sum_{j=1}^{s_2} N_{ijk}$$

V souvislosti s třírozměrnou kontingenční tabulkou lze testovat čtyři hypotézy. První hypotézou je, že jsou všechna tři kritéria vzájemně nezávislá, tj.

$$\mathcal{H}_0 : p_{ijk} = p_{i\cdot} p_{\cdot j} p_{\cdot k} \quad (9.8)$$

nebo lze testovat zda-li jedno ze tří kritérií je nezávislé na zbývajících dvou kritériích. Např. hypotéza, která testuje nezávislost B na A a C , má tvar

$$\mathcal{H}_0 : p_{ijk} = p_{i\cdot k} p_{\cdot j} \quad (9.9)$$

Postup testování těchto hypotéz je analogický postupům v případě dvourozměrných kontingenčních tabulek. Věrohodnostní funkce náhodného výběru je

$$L = \prod_{i,j,k} p_{ijk}^{n_{ijk}} \quad (9.10)$$

kde $\sum_{i,j,k} p_{ijk} = 1$ a $\sum_{i,j,k} n_{ijk} = n$. Pro $\bar{\Theta}$ nastává maximum L pro $\hat{p}_{ijk} = \frac{n_{ijk}}{n}$, a proto

$$\sup_{\bar{\Theta}} L = \frac{1}{n^n} \prod_{i,j,k} n_{ijk}^{n_{ijk}} \quad (9.11)$$

Např. pro testování hypotézy v (9.9) použijeme (9.9) jako substituci do (9.10) a maximalizujeme L s ohledem na $p_{i\cdot k}$ a $p_{\cdot j}$, čímž získáme

$$\hat{p}_{i\cdot k} = \frac{n_{i\cdot k}}{n} \text{ a } \hat{p}_{\cdot j} = \frac{n_{\cdot j}}{n}$$

a

$$\sup_{\bar{\Theta}_0} L = \frac{1}{n^{2n}} \left(\prod_{i,k} n_{i\cdot k}^{n_{i\cdot k}} \right) \left(\prod_j n_{\cdot j}^{n_{\cdot j}} \right) \quad (9.12)$$

Příslušný obecný věrohodnostní poměr λ je pak dán podílem (9.11) a (9.12) a v případě náhodných výběrů velkého rozsahu je možné jej aproximovat pomocí $-2 \ln(\Theta)$, který sleduje chi-kvadrát rozdělení s $s_1 s_2 s_3 - 1 - [(s_1 s_3 - 1) + s_2 - 1] = (s_1 s_3 - 1)(s_2 - 1)$ stupni volnosti. $(s_1 s_3 - 1) + (s_2 - 1)$ je dimenze parametrického prostoru $\underline{\Theta}_0$ a $s_1 s_2 s_3 - 1$ je dimenze prostoru $\underline{\Theta}$.

Pro hypotézu nezávislosti B na A a C , tj.

$$\mathcal{H}_0 : p_{ijk} = p_{i \cdot k} p_{\cdot j}.$$

má odpovídající statistika Q tvar

$$Q = \sum_i \sum_j \sum_k \frac{[N_{ijk} - n(N_{i \cdot k}/n)(N_{\cdot j}/n)]^2}{n(N_{i \cdot k}/n)(N_{\cdot j}/n)}$$

Pro platné \mathcal{H}_0 sleduje Q asymptoticky chi-kvadrát rozdělení s $s_1 s_2 s_3 - 1 - (s_1 s_3 - 1) - (s_2 - 1) = (s_1 s_3 - 1)(s_2 - 1)$ stupni volnosti. Statistika Q má opět intuitivní základ, protože N_{ijk} představuje počet prvků v buňce ijk a $n(N_{i \cdot k}/n)(N_{\cdot j}/n)$ představuje očekávaný počet prvků v téže buňce za předpokladu pravdivosti \mathcal{H}_0 .

9.6 Testování hypotéz a intervaly spolehlivosti

V předchozím textu jsme interval spolehlivosti jednorozměrného parametru θ použili pro konstrukci testu $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta \neq \theta_0$. Tento postup však lze také obrátit a z rodiny testů $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta \neq \theta_0$ ²⁰ získat interval spolehlivosti pro θ . Následující text je možné chápat jako úvod do problematiky.

Pojem interval spolehlivosti nahradíme obecnějším pojmem množina spolehlivosti. Stejně jako v předchozím textu označuje \mathfrak{X} výběrový prostor, $\underline{\Theta}$ parametrický prostor a (x_1, \dots, x_n) konkrétní realizaci náhodného výběru.

Definice 9.18 (Množina spolehlivosti) Rodina podmnožin parametrického prostoru $\underline{\Theta}$ indexovaná skrze $(x_1, \dots, x_n) \in \mathfrak{X}$, kterou označujeme jako $\vartheta = \{\underline{\Theta}(x_1, \dots, x_n) : \underline{\Theta}(x_1, \dots, x_n) \subseteq \underline{\Theta}; (x_1, \dots, x_n) \in \mathfrak{X}\}$, je rodinou množin spolehlivosti s koeficientem spolehlivosti γ , jestliže pro všechna $\theta \in \underline{\Theta}$ platí

$$P_\theta[\underline{\Theta}(X_1, \dots, X_n) \text{ obsahuje } \theta] = \gamma$$

Libovolný člen $\underline{\Theta}(x_1, \dots, x_n)$ z rodiny množin spolehlivosti je podmnožinou parametrického prostoru $\underline{\Theta}$. $\underline{\Theta}(X_1, \dots, X_n)$ je náhodnou podmnožinou - pro konkrétní realizaci (x_1, \dots, x_n) náhodných veličin (X_1, \dots, X_n) nabývá $\underline{\Theta}(X_1, \dots, X_n)$ hodnoty $\underline{\Theta}(x_1, \dots, x_n)$, což je člen rodiny ϑ . Pro dané θ je $\{\underline{\Theta}(X_1, \dots, X_n) \text{ obsahuje } \theta\}$ náhodným jevem. Toto θ , které taktéž figuruje v P_θ výše uvedené věty, indexuje pravděpodobnostní rozdělení náhodných veličin (X_1, \dots, X_n) vystupujících v $\underline{\Theta}(X_1, \dots, X_n)$.

Příklad 9.25 Uvažujme náhodný výběr X_1, \dots, X_n z populace $N(\theta, 1)$, kde $\underline{\Theta} = \{\theta : -\infty < \theta < \infty\}$. Nechť má podmnožina $\underline{\Theta}(x_1, \dots, x_n)$ podobu intervalu $(\bar{x} - z/\sqrt{n}, \bar{x} + z/\sqrt{n})$, kde z je dáno vztahem $\Phi(z) - \Phi(-z) = \gamma$. Pak je rodina podmnožin $\vartheta = \{\underline{\Theta}(x_1, \dots, x_n) : \underline{\Theta}(x_1, \dots, x_n) = (\bar{x} - z/\sqrt{n}, \bar{x} + z/\sqrt{n})\}$ rodinou

²⁰Rodinu testů vytvoříme jednoduše tak, že postupně měníme θ_0 .

množin spolehlivosti s koeficientem spolehlivosti γ , protože pro všechna $\theta \in \bar{\Theta}$ platí

$$\begin{aligned} P_\theta[\bar{\Theta}(X_1, \dots, X_n) \text{ obsahuje } \theta] &= P_\theta\left[\bar{X} - \frac{z}{\sqrt{n}} < \theta < \bar{X} + \frac{z}{\sqrt{n}}\right] \\ &= P_\theta\left[-z < \frac{\bar{X} - \theta}{1/\sqrt{n}} < z\right] = \gamma \end{aligned}$$

Rodina ϑ je rodinou intervalů spolehlivosti pro θ s koeficientem spolehlivosti γ .

Jak nyní ukážeme, lze množiny spolehlivosti zkonstruovat na základě testů hypotéz. Uvažujme nenáhodný test \mathcal{T}_{θ_0} velikosti α pro nulovou hypotézu $\mathcal{H}_0: \theta = \theta_0$. Nechť $\mathfrak{X}(\theta_0)$ představuje akceptační region testu \mathcal{T}_{θ_0} ²¹. Protože má test \mathcal{T}_{θ_0} má velikost α platí

$$P_{\theta_0}[(X_1, \dots, X_n) \in \mathfrak{X}(\theta_0)] = 1 - \alpha$$

S tím, jak měníme θ_0 nad množinou $\bar{\Theta}$, získáváme sérii testů \mathcal{T}_{θ_0} a jim odpovídající akceptačních regiony $\mathfrak{X}((\theta_0) : \theta_0 \in \bar{\Theta})$. Lze definovat

$$\bar{\Theta}(x_1, \dots, x_n) = \{\theta_0 : (x_1, \dots, x_n) \in \mathfrak{X}(\theta_0)\} \quad (9.13)$$

Je zřejmé, že $\bar{\Theta}(x_1, \dots, x_n)$ je podmnožinou $\bar{\Theta}$. Dále platí, že $\{\bar{\Theta}(x_1, \dots, x_n)\}$ je rodinou množin spolehlivosti s koeficientem spolehlivosti $\gamma = 1 - \alpha$, protože $\{\bar{\Theta}(X_1, \dots, X_n) \text{ obsahuje } \theta_0\}$ pouze, jestliže $\{(X_1, \dots, X_n) \in \mathfrak{X}(\theta_0)\}$, a proto $P_{\theta_0}[\bar{\Theta}(X_1, \dots, X_n) \text{ obsahuje } \theta_0] = P_{\theta_0}[(X_1, \dots, X_n) \in \mathfrak{X}(\theta_0)] = 1 - \alpha$.

Příklad 9.26 Uvažujme náhodný výběr X_1, \dots, X_n z populace $N(\theta, 1)$ a testujme $\mathcal{H}: \theta = \theta_0$. Test velikosti α zamítne \mathcal{H}_0 , jestliže $|\bar{x} - \theta_0| \geq z/\sqrt{n}$, kde z je dáno vztahem $\Phi(z) - \Phi(-z) = 1 - \alpha$. Akceptační region je dán

$$\mathfrak{X}(\theta_0) = \left((x_1, \dots, x_n) : \theta_0 - \frac{z}{\sqrt{n}} < \bar{x} < \theta_0 + \frac{z}{\sqrt{n}} \right)$$

Dále, stejně jako v případě (9.13) definujeme

$$\begin{aligned} \bar{\Theta}(x_1, \dots, x_n) &= \{\theta_0 : (x_1, \dots, x_n) \in \mathfrak{X}(\theta_0)\} \\ &= \left(\theta_0 : \theta_0 - \frac{z}{\sqrt{n}} < \bar{x} < \theta_0 + \frac{z}{\sqrt{n}} \right) \\ &= \left(\theta_0 : \bar{x} - \frac{z}{\sqrt{n}} < \theta_0 < \bar{x} + \frac{z}{\sqrt{n}} \right) \end{aligned}$$

$\bar{\Theta}(x_1, \dots, x_n)$ je interval spolehlivosti s koeficientem spolehlivosti $\gamma = 1 - \alpha$.

Výše uvedený obecný postup ukazuje, jak lze testování hypotéz použít při konstrukci množin spolehlivosti. Tento postup však lze také obrátit a z rodiny množin spolehlivosti získat test hypotézy. Pro rodinu množin spolehlivosti $\{\bar{\Theta}(x_1, \dots, x_n)\}$ s koeficientem spolehlivosti γ jsme definovali

$$\mathfrak{X}(\theta_0) = \{(x_1, \dots, x_n) : \theta_0 \in \bar{\Theta}(x_1, \dots, x_n)\}$$

²¹Připomeňme, že akceptační region je doplňkem kritického regionu. Pro kritický region $C(\theta_0)$ tak platí $\mathfrak{X}(\theta_0) = \mathfrak{X} - C(\theta_0)$.

Nenáhodný test s výše uvedeným akceptačním regionem $\mathfrak{X}(\theta_0)$ je testem hypotézy $\mathcal{H}_0 : \theta = \theta_0$ s velikostí $\alpha = 1 - \gamma$.

Vztah mezi testováním hypotéz a množinami spolehlivosti není užitečný pouze k tomu, že z prvního lze odvodit druhé a naopak. Další důležitou oblastí je optimalita testu resp. množiny spolehlivosti. Jestliže je daný test optimální, pak je odpovídající množina spolehlivosti taktéž optimální a naopak. Test toho tvrzení překračuje záběr této knihy, a proto se pouze omezíme na následující definici.

Definice 9.19 (Uniformně nejvíce přesná rodina množin spolehlivosti)

Rodina množin spolehlivosti $\{\bar{\Theta}^(x_1, \dots, x_n)\}$ s koeficientem spolehlivosti γ je uniformně nejpřesnější rodinou množin spolehlivosti pro koeficient γ , jestliže pro libovolnou alternativní rodinu množin spolehlivosti $\{\bar{\Theta}(x_1, \dots, x_n)\}$ s koeficientem spolehlivosti γ a všechna θ a θ' platí*

$$P_\theta[\bar{\Theta}^*(X_1, \dots, X_n) \text{ obsahuje } \theta'] \leq P_\theta[\bar{\Theta}(X_1, \dots, X_n) \text{ obsahuje } \theta']$$

Výše uvedená definice říká, že $\bar{\Theta}^*(X_1, \dots, X_n)$ obsahuje nesprávné θ' s menší pravděpodobností než $\bar{\Theta}(X_1, \dots, X_n)$, přičemž $\bar{\Theta}^*(X_1, \dots, X_n)$ a $\bar{\Theta}(X_1, \dots, X_n)$ obsahují správné θ se stejnou pravděpodobností. Problémem je, že obecné uniformně nejpřesnější množina spolehlivosti zpravidla neexistuje. Nicméně ji lze definovat nad omezeným souborem množin spolehlivosti, kde jsou šance na její existenci větší. K tomuto účelu lze opět využít vztahu mezi testem hypotéz a množinami spolehlivosti. Jestliže je \mathcal{T}^* uniformně nejsilnějším testem velikosti α nad omezeným souborem testů pro hypotézu $\mathcal{H} : \theta = \theta_0$, pak je pro tento omezený soubor odpovídající množina spolehlivosti uniformně nejpřesnější s koeficientem $\gamma = 1 - \alpha$. Tímto způsobem lze přetransformovat uniformně nejsilnější test v uniformně nejpřesnější množinu spolehlivosti.

9.7 Sekvenční testování hypotéz

9.7.1 Úvod

Sekvenční testování hypotéz je charakteristické tím, že se velikost náhodného výběru odvíjí od průběhu testu a namísto toho, aby byla dána před samotným testováním. V následujícím textu se zaměříme na jednu z forem sekvenčního testování hypotéz a to na sekvenční pravděpodobnostní poměr.

Uvažujme test ve formě $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_1$. Dle Neyman-Pearsonovy věty platí, že v případě náhodného výběru velikosti n a chyby typu II velikosti α lze velikost β chyby typu I minimalizovat pomocí testu založeného na věrohodnostním poměru. To znamená, že pro dané α a n minimalizujeme β . Nyní předpokládejme, že pro dané α i β chceme stanovit velikost náhodného výběru n , pro kterou má odpovídající věrohodnostní poměr chybu I resp. II typu o velikosti α resp. β . Řešení takového problému ilustruje následující příklad

Příklad 9.27 *Výrobce ví, že doba životnosti jeho výrobku je dána $N(100, 100)$. Aby přijal nový výrobní proces, chce si být jistý, že tento proces povede k navýšení střední doby životnosti alespoň o 5%. K tomuto účelu podrobí testu výrobky z zkušebního provozu, které jsou vyrobeny novým postupem.*

Předpokládejme, že doba životnosti výrobku ze zkušebního provozu je $N(\theta, 100)$ a že testujeme $\mathcal{H}_0 : \theta \leq 100$ vs. $\mathcal{H}_1 : \theta > 100$. Zafixujme velikosti chyby typu I a II na 1% a 5% a pokusme se nalézt velikost zkušební vzorku. Platí

$$0.01 = P_{\theta=100}[\bar{X}_n > k] \text{ a } 0.05 = P_{\theta=105}[\bar{X} \leq k]$$

neboli

$$0.01 = 1 - \Phi\left(\frac{k-100}{10/\sqrt{n}}\right) \text{ a } 0.05 = \Phi\left(\frac{k-105}{10/\sqrt{n}}\right)$$

První z rovnic lze převést do tvaru $\Phi\left(\frac{k-100}{10/\sqrt{n}}\right) = 0.99$, což implikuje

$$\frac{k-100}{10/\sqrt{n}} \approx 2.326$$

Druhá rovnice $\Phi\left(\frac{k-105}{10/\sqrt{n}}\right) = 0.05$ pak implikuje

$$\frac{k-105}{10/\sqrt{n}} \approx -1.645$$

Jestliže z obou vztahů vyjádříme k a dáme je do rovnosti, získáme

$$100 + 10 \cdot 2.326/\sqrt{n} \approx 105 - 10 \cdot 1.645/\sqrt{n}$$

Řešením je pak $n \approx 63.08$, tj. potřebujeme vzorek o velikosti 64 výrobků.

Ve výše uvedeném případě nemusí být vždy nutné otestovat všech 64 výrobků, abychom byli schopni přijmout či zamítnout nulovou hypotézu. Může se totiž stát, že již během prvních 20, 30 či 40 pozorování získáme dostatečné důkazy k tomu, abychom pro dané α a β nulovou hypotézu přijali popř. zamítl. To znamená, že je zde možnost zkonstruovat takový test, který by v průměru umožnil ukončit testování před dosažením počtu 64 pozorování. Tímto se dostáváme k následující kapitole.

9.7.2 Definice sekvenčního poměrového testu

Uvažujme dvě populace a náhodný výběr, o kterém nevíme, ze které populace pochází. Předpokládejme, že naším cílem je tento náhodný výběr přiřadit jedné z populací. Testujeme tedy $\mathcal{H}_0 : X_i \sim f_0(\cdot)$ vs. $\mathcal{H}_1 : X_i \sim f_1(\cdot)$. Připomeňme, že věrohodnostní poměrový test zamítl \mathcal{H}_0 , jestliže

$$\lambda = \frac{L_0}{L_1} \leq k$$

pro určitou konstantu $k > 0$. Sekvenční poměrový test je založen na sérii takovýchto testů. Definujme

$$\lambda_m = \lambda_m(x_1, \dots, x_m) = \frac{L_0(x_1, \dots, x_m)}{L_1(x_1, \dots, x_m)} = \frac{L_0(m)}{L_1(m)} = \frac{\prod_{i=1}^m f_0(x_i)}{\prod_{i=1}^m f_1(x_i)}$$

pro $m = 1, 2, \dots$ a postupně vypočteme $\lambda_1, \lambda_2, \dots$. Pro fixní k_0 a k_1 , které splňují podmínku $0 < k_0 < k_1$, aplikujme následující postup.

1. Realizujeme pozorování x_1 a vypočteme pro něj poměr λ_1 .
2. Jestliže $\lambda_1 \leq k_0$, zamítneme \mathcal{H}_0 a ukončíme testování.
3. Jestliže $\lambda_1 \geq k_1$, přijmeme \mathcal{H}_0 a ukončíme testování.
4. Jestliže $k_0 < \lambda_1 < k_1$, realizujeme pozorování x_2 a vypočteme λ_2 .
5. Jestliže $\lambda_2 \leq k_0$, zamítneme \mathcal{H}_0 a ukončíme testování.
6. Jestliže $\lambda_2 \geq k_1$, přijmeme \mathcal{H}_0 a ukončíme testování.
7. Jestliže $k_0 < \lambda_2 < k_1$, realizujeme pozorování x_3 a vypočteme λ_3 .
8. Postupně navyšujeme počet pozorování a aplikujeme výše popsany postup, dokud
 - (a) $\lambda_i \leq k_0$, kdy zamítneme \mathcal{H}_0 a ukončíme testování
 - (b) $\lambda_i \geq k_1$, kdy přijmeme \mathcal{H}_0 a ukončíme testování

Kritický region výše popsaného sekvenčního testu je $C = \cup_{n=1}^{\infty} C_n$, kde

$$C_n = \{(x_1, \dots, x_n) : k_0 < \lambda_j(x_1, \dots, x_j) < k_1, j = 1, \dots, n-1, \lambda_n(x_1, \dots, x_n) \leq k_0\}$$

Podobně akceptační region je $A = \cup_{n=1}^{\infty} A_n$, kde

$$A_n = \{(x_1, \dots, x_n) : k_0 < \lambda_j(x_1, \dots, x_j) < k_1, j = 1, \dots, n-1, \lambda_n(x_1, \dots, x_n) \geq k_1\}$$

Definice 9.20 (Sekvenční poměrový test) Pro dané $0 < k_0 < k_1$ je výše popsaný test označován jako *sekvenční poměrový test*.

V předchozím textu jsme se pro dané n snažili stanovit k tak, aby výsledný test měl velikost α . Nyní se pokusíme stanovit k_0 a k_1 takové, aby výsledný sekvenční poměrový test měl chyby typu I resp. II předem dané velikosti α resp. β . Všimněme si, že

$$\alpha = P[\text{zamítnutí } \mathcal{H}_0 | \mathcal{H}_0 \text{ je pravdiv}] = \sum_{n=1}^{\infty} \int_{C_n} L_0(n)$$

a

$$\beta = P[\text{přijetí } \mathcal{H}_0 | \mathcal{H}_0 \text{ je nepravdiv}] = \sum_{n=1}^{\infty} \int_{A_n} L_1(n)$$

kde, stejně jako dříve, představuje $\int_{C_n} L_0(n)$ zkrácený zápis pro $\int \dots \int_{C_n} (\prod_{i=1}^n f_0(x_i) dx_i)$. Protože A_n a C_n jsou vyjádřeny ve formě k_0 a k_1 , lze pro fixní α a β výše uvedené dvě rovnice chápat jako rovnice o dvou neznámých k_0 a k_1 . Jejich řešením pak získáme sekvenční poměrový test s chybami o velikosti α a β . Jak je však při pohledu na tyto dvě rovnice patrné, je výpočet k_0 a k_1 poměrně komplikovanou záležitostí. Naštěstí však existuje poměrně jednoduchá aproximace, kterou si představíme následující kapitole.

V rámci sekvenčního poměrového testu navyšujeme počet pozorování, dokud se $\lambda_n(x_1, \dots, x_n)$ nenachází mimo interval (k_0, k_1) . Velikost náhodného vzorku v rámci testu tak závisí na realizacích x_1, x_2, \dots náhodných veličin X_1, X_2, \dots , a proto se jedná taktéž o náhodnou veličinu. Označme ji písmenem N . Jedním ze způsobů hodnocení sekvenčních poměrových testů je porovnání střední hodnoty N , kde preferujeme test s nižší střední hodnotou. Tímto se dostáváme k následující větě, kterou uvádíme bez důkazu.

Věta 9.10 *Sekvenční poměrový test, který splňuje*

$$P[\mathcal{H}_0 \text{ je zamítnuto} | \mathcal{H} \text{ je pravdivé}] \leq \alpha$$

a

$$P[\mathcal{H}_0 \text{ je přijmuto} | \mathcal{H} \text{ je nepravdivé}] \leq \beta$$

minimalizuje v porovnání se všemi alternativními testy s chybami velikosti α a β jak

$$E[N | \mathcal{H}_0 \text{ je pravdivé}]$$

tak

$$E[N | \mathcal{H}_0 \text{ je nepravdivé}]$$

V tomto směru je tedy tento sekvenční poměrový test optimální.

9.7.3 Přibližný sekvenční poměrový test

V předchozím textu jsme konstatovali, že výpočet parametrů k_0 a k_1 , které jsou nezbytné pro definování sekvenčního poměrového testu, je poměrně komplikovaný. Naštěstí existuje relativně jednoduchá aproximace těchto parametrů.

Tvrzení 9.1 *Nechť jsou k_0 a k_1 definovány tak, že jim odpovídající sekvenční poměrový test má chyby velikosti α a β . Parametry k_0 a k_1 lze aproximovat pomocí*

$$k'_0 = \frac{\alpha}{1-\beta} \quad a \quad k'_1 = \frac{1-\alpha}{\beta}$$

Důkaz 9.5 *Předpokládejme, že $\sum_{n=1}^{\infty} P[N = n | \mathcal{H}_i] = 1$ pro $i = 0, 1$. Z definice chyby typu I a poměrového testu vyplývá*

$$\begin{aligned} \alpha &= P[\text{zamítnutí } \mathcal{H}_0 | \mathcal{H}_0 \text{ je pravdivé}] = \sum_{n=1}^{\infty} \int_{C_n} L_0(n) \leq \sum_{n=1}^{\infty} \int_{C_n} k_0 L_1(n) \\ &= k_0 \sum_{n=1}^{\infty} \int_{C_n} L_1(n) = k_0 P[\text{zamítnutí } \mathcal{H}_0 | \mathcal{H}_1 \text{ je pravdivé}] = k_0(1-\beta) \end{aligned}$$

a proto $k_0 \geq \frac{\alpha}{1-\beta}$. Zároveň však také platí

$$\begin{aligned} 1-\alpha &= P[\text{přijetí } \mathcal{H}_0 | \mathcal{H}_0 \text{ je pravdivé}] = \sum_{n=1}^{\infty} \int_{A_n} L_0(n) \geq \sum_{n=1}^{\infty} \int_{A_n} k_1 L_1(n) \\ &= k_1 P[\text{přijetí } \mathcal{H}_0 | \mathcal{H}_1 \text{ je pravdivé}] = k_1 \beta \end{aligned}$$

a proto $k_1 \leq \frac{1-\alpha}{\beta}$. Všimněme si, že aproximace $k'_0 = \frac{\alpha}{1-\beta}$ a $k'_1 = \frac{1-\alpha}{\beta}$ splňují

$$k'_0 = \frac{\alpha}{1-\beta} \leq k_0 < k_1 \leq \frac{1-\alpha}{\beta} = k'_1$$

Tvrzení 9.2 *Jestliže jsou α' a β' velikosti chyb sekvenčního poměrového testu definovaného pomocí k'_0 a k'_1 , pak platí $\alpha' + \beta' \leq \alpha + \beta$.*

Důkaz 9.6 *Nechť A' resp. C' označuje akceptační resp. kritický region sekvenčního poměrového testu definovaného pomocí k'_0 a k'_1 . Pak platí*

$$\alpha' = \sum_{n=1}^{\infty} \int_{C'_n} L_0(n) \leq \frac{\alpha}{1-\beta} \sum_{n=1}^{\infty} \int_{C'_n} L_1(n) = \frac{\alpha}{1-\beta} (1-\beta')$$

a

$$1-\alpha' = \sum_{n=1}^{\infty} \int_{A'_n} L_0(n) \geq \frac{1-\alpha}{\beta} \sum_{n=1}^{\infty} \int_{A'_n} L_1(n) = \frac{1-\alpha}{\beta} \beta'$$

a proto $\alpha'(1-\beta) \leq \alpha(1-\beta')$ a $(1-\alpha)\beta' \leq (1-\alpha')\beta$, což v souhrnu implikuje $\alpha'(1-\beta) + (1-\alpha)\beta' \leq \alpha(1-\beta') + (1-\alpha')\beta$ neboli $\alpha' + \beta' \leq \alpha + \beta$.

Ideálně bychom preferovali sekvenční poměrový test, který má chyby o předem daných velikostech α a β . Protože je však složité nalézt odpovídající k_0 a k_1 , použijeme testt definovaný pomocí jejich aproximací k'_0 a k'_1 s vědomím, že součet chyb α' a β' je menší nebo roven součtu požadovaných chyb α a β .

9.7.4 Přibližná střední hodnota velikosti náhodného výběru sekvenčního poměrového testu

Sekvenční poměrový test probíhá tak dlouho, dokud $k_0 < \lambda_m < k_1$ a skončí jakmile $\lambda_m \leq k_0$ nebo $\lambda_m \geq k_1$. Definujme $z_i = \ln \left(\frac{f_0(x_i)}{f_1(x_i)} \right)$. V rámci ekvivalentního testu pak navyšujeme velikost vzorku dokud $\ln(k_0) < \sum_{i=1}^m z_i < \ln(k_1)$ a test ukončíme, jestliže $\sum_{i=1}^m z_i \leq \ln(k_0)$, kdy \mathcal{H}_0 zamítneme, nebo $\sum_{i=1}^m z_i \geq \ln(k_1)$, kdy \mathcal{H}_0 přijmeme. Stejně jako v předchozím textu i zde představuje náhodná veličina N velikost vzorku sekvenčního poměrového testování. Dále definujme $Z_i = \ln \left(\frac{f_0(X_i)}{f_1(X_i)} \right)$. Následující věta slouží k určení přibližné střední hodnoty velikosti výběru pro sekvenční poměrový test.

Věta 9.11 (Waldova rovnice) *Uvažujme nezávislé náhodné veličiny Z_1, \dots, Z_n, \dots s identickým pravděpodobnostním rozdělením. Nechť tyto náhodné veličiny splňují podmínku $E[|Z_i|] < \infty$ a nechť N je náhodná veličina, která nabývá hodnot z oboru přirozených čísel. Pak platí*

$$E[Z_1 + \dots + Z_n] = E[N] \cdot E[Z_i]$$

Důkaz 9.7

$$\begin{aligned} E[Z_1 + \dots + Z_N] &= E[E[Z_1 + \dots + Z_N] | N] \\ &= \sum_{n=1}^{\infty} E[Z_1 + \dots + Z_n | N = n] P[N = n] = \sum_{n=1}^{\infty} \sum_{i=1}^n E[Z_i | N = n] P[N = n] \\ &= \sum_{i=1}^{\infty} \sum_{n=1}^{\infty} E[Z_i | N = n] P[N = n] = \sum_{i=1}^{\infty} E[Z_i | N \geq i] P[N \geq i] \\ &= \sum_{i=1}^{\infty} E[Z_i] P[N \geq i] = E[Z_i] \sum_{i=1}^{\infty} P[N \geq i] = E[Z_i] E[N] \end{aligned}$$

Při úpravách jsme použili vztah $E[Z_i] = E[Z_i | N \geq i]$. Tento vztah platí, protože $\{N \geq i\}$ závisí pouze na Z_1, \dots, Z_{i-1} , a proto je nezávislý na Z_i . Vztah $E[N] = \sum_{i=1}^{\infty} P[N \geq i]$ pak vychází z bodu (3) definice střední hodnoty (2.11).

Jestliže sekvenční pravděpodobnostní test vede k zamítnutí \mathcal{H}_0 , pak $Z_1 + \dots + Z_N \leq \ln(k_0)$, nicméně $Z_1 + \dots + Z_N$ je relativně blízké $\ln(k_0)$, protože $Z_1 + \dots + Z_N$ se stane poprvé menší nebo rovno $\ln(k_0)$ pro N -té pozorování. Proto platí $E[Z_1 + \dots + Z_N] \approx \ln(k_0)$. Podobně, jestliže test vede k přijetí \mathcal{H}_0 , platí $E[Z_1 + \dots + Z_N] \approx \ln(k_1)$. Proto $E[Z_1 + \dots + Z_N] \approx \rho \ln(k_0) + (1 - \rho) \ln(k_1)$, kde $\rho = P[\mathcal{H}_0 \text{ je zamítnuto}]$. S využitím

$$E[N] = \frac{E[Z_1 + \dots + Z_N]}{E[Z_1]} \approx \frac{\rho \ln(k_0) + (1 - \rho) \ln(k_1)}{E[Z_i]}$$

získáme

$$\begin{aligned} E[N | \mathcal{H}_0 \text{ je pravdivé}] &\approx \frac{\alpha \ln(k_0) + (1 - \alpha) \ln(k_1)}{E[Z_i | \mathcal{H}_0 \text{ je pravdivé}]} \\ &\approx \frac{\alpha \ln[\alpha/(1 - \beta)] + (1 - \alpha) \ln[(1 - \alpha)/\beta]}{E[Z_i | \mathcal{H}_0 \text{ je pravdivé}]} \end{aligned} \quad (9.14)$$

a

$$\begin{aligned} E[N | \mathcal{H}_0 \text{ je nepravdivé}] &\approx \frac{(1 - \beta) \ln(k_0) + \beta \ln(k_1)}{E[Z_i | \mathcal{H}_0 \text{ je nepravdivé}]} \\ &\approx \frac{(1 - \beta) \ln[\alpha/(1 - \beta)] + \beta \ln[(1 - \alpha)/\beta]}{E[Z_i | \mathcal{H}_0 \text{ je nepravdivé}]} \end{aligned} \quad (9.15)$$

Příklad 9.28 Uvažujme náhodný výběr z $N(\theta, \sigma^2)$, kde σ^2 známe. Testujme $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_1$. Platí

$$\begin{aligned} z_i &= \ln \left(\frac{f_0(x_i)}{f_1(x_i)} \right) = \ln \left(\frac{(1/\sqrt{2\pi}\sigma) e^{-\frac{1}{2}[(x_i - \theta_0)/\sigma]^2}}{(1/\sqrt{2\pi}\sigma) e^{-\frac{1}{2}[(x_i - \theta_1)/\sigma]^2}} \right) \\ &= -\frac{1}{2\sigma^2} [(x_i - \theta_0)^2 - (x_i - \theta_1)^2] = -\frac{1}{2\sigma^2} [(\theta_0^2 - \theta_1^2) - 2x_i(\theta_0 - \theta_1)] \end{aligned}$$

a proto

$$E[Z_i | \mathcal{H}_0 \text{ je pravdivé}] = -\frac{1}{2\sigma^2} [(\theta_0^2 - \theta_1^2) - 2\theta_0(\theta_0 - \theta_1)] = \frac{1}{2\sigma^2} (\sigma_1 - \sigma_0)^2$$

a

$$E[Z_i | \mathcal{H}_0 \text{ je nepravdivé}] = -\frac{1}{2\sigma^2} (\theta_1 - \theta_0)^2$$

Vraťme se k příkladu (9.27). Dle zadání tohoto příkladu je $\alpha = 0.01, \beta = 0.05, \sigma^2 = 100, \theta_0 = 100$ a $\theta_1 = 105$. Dosazením do (9.14) a (9.15) získáváme

$$E[N | \mathcal{H}_0 \text{ je pravdivé}] \approx \frac{0.01 \ln(0.01/0.95) + 0.99 \ln(0.99/0.05)}{25/200} \approx 24$$

a

$$E[N | \mathcal{H}_0 \text{ je nepravdivé}] \approx 34$$

Pro sekvenční poměrový test jsou tedy průměrné velikosti vzorku 24 resp. 34 výrobků. Pro test s pevně danou velikostí vzorku jsem v příkladě (9.27) dospěli k počtu 64 produktů.

Kapitola 10

Lineární regrese

10.1 Příklady lineární regrese

Příklad 10.1 Vzdálenost s , kterou částice urazí v čase t je dána rovnicí $s = \beta_0 + \beta_1 t$, kde β_1 je průměrnou rychlostí a β_0 je pozice v čase $t = 0$. Jestliže β_0 a β_1 neznáme, pak lze pozorovat s pro dvě rozdílná t a z odpovídajících dvou rovnic vyřešit β_0 a β_1 . Pokud např. $s = 2$ pro $t = 1$ a $s = 11$ pro $t = 4$, získáme rovnice $2 = \beta_0 + \beta_1$ a $11 = \beta_0 + 4\beta_1$, jejichž řešením je $\beta_0 = -1$ a $\beta_1 = 3$, tj. $s = -1 + 3t$.

Předpokládejme, že měření vzdálenosti je zatíženou náhodnou chybou. Proto namísto s pozorujeme $Y = s + E$, kde E představuje chybu s nulovou střední hodnotou. Substitucí za s tak získáme

$$Y = \beta_0 + \beta_1 t + E$$

kde Y je náhodná veličina, jejíž realizaci jsme schopni pozorovat, a E je náhodná veličina, jejíž realizaci schopni pozorovat nejsme. t je nenáhodná veličina a β_0 a β_1 představují neznámé parametry. Hodnotu β_0 a β_1 nejsme schopni vypočítat ze dvou pozorování (t, Y) , jako tomu bylo v případě (t, s) , protože neexistuje žádná funkční závislost mezi Y a t . Cílem našeho snažení je nalézt β_0 a β_1 a odvodit tak vztah $s = \beta_0 + \beta_1 t$ pro různé hodnoty t . Protože s je zatíženo chybami a nelze ho přímo pozorovat, nemůžeme znát ani β_0 a β_1 . Nicméně na základě pozorování různých (t, Y) lze s využitím statistických metod odhadnout β_0 , β_1 a s .

Příklad 10.2 Uvažujme vztah mezi výškou jedince h a jeho hmotností w . Je zřejmé, že neexistuje žádný funkční vztah mezi h a w , nicméně se dá očekávat, že mezi nimi existuje určitá závislost.

Předpokládejme, že výška a váha jsou náhodné veličiny (W, H) , které sledují dvourozměrné normální rozdělení. Očekávaná hodnota H pro danou hodnotu w náhodné veličiny W , je dána vztahem

$$E[H|W = w] = \beta_0 + \beta_1 w$$

kde β_0 a β_1 jsou parametry dvourozměrného normálního rozdělení. Ačkoliv mezi H a W neexistuje funkční vztah, existuje mezi nimi, pokud předpokládáme sdruženě normální rozdělení, lineární závislost. Proto platí

$$E[H|W = w] = \beta_0 + \beta_1 w$$

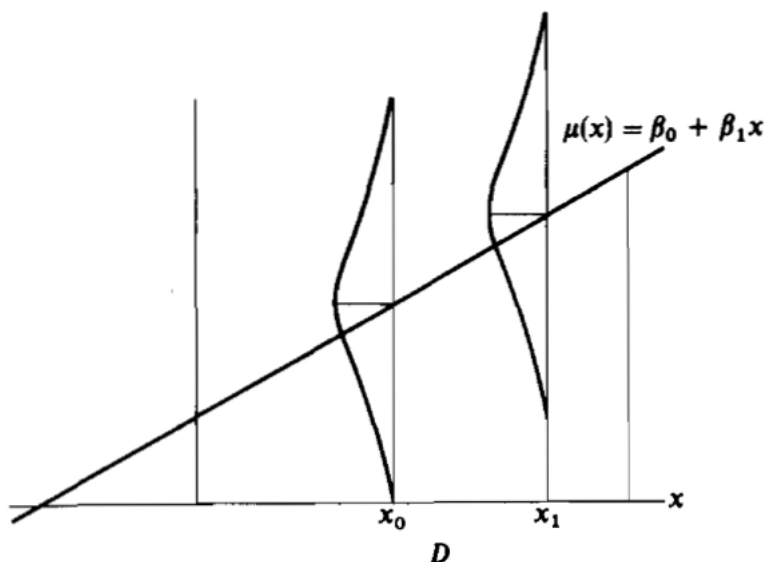
nebo také

$$H_w = \beta_0 + \beta_1 w + E$$

kde E je chyba, která sleduje normální rozdělení.

10.2 Definice regresního modelu

Uvažujme lineární funkci $\mu(\cdot)$ reálné proměnné x , tj. $\mu(x) = \beta_0 + \beta_1 x$, kde $x \in D$. Množina D je obvykle reprezentována celou reálnou přímkou popř. polopřímkou nebo intervalem. Předpokládejme, že existuje taková rodina kumulativních distribučních funkcí (pro každé x jedna), jejichž střední hodnota pro dané $x = x_0$ je rovna $\beta_0 + \beta_1 x_0$. To znamená, že střední hodnoty těchto kumulativních distribučních funkcí tvoří přímku definovanou jako $\mu(x) = \beta_0 + \beta_1 x$, jak je ilustrováno obrázkem (10.1).



Obrázek 10.1: Lineární regrese

Cílem je na základě náhodného výběru z několika těchto kumulativních distribučních funkcí získat odhad parametrů β_0 a β_1 . Toho je dosaženo následovně.

1. Soubor n pozorování hodnot x označme jako x_1, \dots, x_n . Soubor x_1, \dots, x_n není souborem náhodných veličin, nicméně může být vybrán jak náhodně tak na základě vědomé selekce.
2. Každé x_i určuje kumulativní distribuční funkci se střední hodnotou $\beta_0 + \beta_1 x_i$ a rozptylem σ^2 . Z této kumulativní distribuční funkce je náhodně vybrána jedna hodnota označovaná jako Y_i . Tímto způsobem získáme n uspořádaných dvojic $(Y_1, x_1), \dots, (Y_n, x_n)$. V předchozím textu jsme předpokládali

$$E[Y_i] = \beta_0 + \beta_1 x_i \quad D[Y_i] = \sigma^2$$

Proto jsme schopni definovat náhodné veličiny E_1, \dots, E_n jako

$$E_i = Y_i - \beta_0 - \beta_1 x_i, \quad i = 1, \dots, n$$

kde $E[E_i] = 0$ a $D[E] = \sigma^2$.

Výše uvedený vztah lze převést do tvaru

$$Y_i = \beta_0 + \beta_1 x_i + E_i, \quad i = 1, \dots, n$$

kde $E[E_i] = 0$ a $D[E_i] = \sigma^2$, což je definice jednoduché lineární regrese.

Definice 10.1 (Lineární regrese) Uvažujme funkci $\mu(x) = \beta_0 + \beta_1 x$ pro všechna $x \in D$. Pro každé x z D uvažujme kumulativní distribuční funkci $F_{Y_x}(\cdot)$ se střední hodnotou $\mu(x) = \beta_0 + \beta_1 x$ a rozptylem σ^2 . Necht x_1, \dots, x_n představuje soubor n (nikoliv nutně náhodných) pozorování z D . Pro x_i uvažujme jednorázový náhodný výběr Y_i z $F_{Y_{x_i}}(\cdot)$, kde $i = 1, \dots, n$. Pak $(Y_1, x_1), \dots, (Y_n, x_n)$ představuje soubor n uspořádaných dvojic, které jsou “provázány” vztahem

$$E[Y_i] = \beta_0 + \beta_1 x_i$$

a

$$D[Y_i] = \sigma^2$$

kde $i = 1, \dots, n$. Výše uvedené rovnice je možné vyjádřit také ve tvaru

$$Y_i = \beta_0 + \beta_1 x_i + E_i$$

$$E[E_i] = 0$$

$$D[E_i] = \sigma^2$$

kde $i = 1, \dots, n$.

Tyto vztahy a podmínky definují jednoduchý regresní model.

Slovo “lineární” v pojmu “lineární regrese” odpovídá skutečnosti, že funkce $\mu(\cdot)$ je lineární v parametrech β_0 a β_1 . Proto je $Y = \mu(x) + E$, kde $\mu(x) = \beta_0 + \beta_1 x$, také lineární regresí. Na kumulativní distribuční funkci $F_{Y_x}(\cdot)$ jsou zpravidla kladeny další požadavky jako např. normalita. Dále je způsob náhodného výběru takový, že Y_i jsou buďto sdruženě nezávislé nebo párově nezávislé. Tímto se dostáváme k definici dvou základních případů, na které se budeme odkazovat v následujících kapitolách.

- **Případ A** - V tomto případě je všech n náhodných veličin Y_i sdruženě nezávislých a každé Y_i sleduje normální rozdělení.
- **Případ B** - V tomto případě jsou Y_i párově nezávislé, tj. $\text{cov}(Y_i, Y_j) = 0$ pro všechna $i \neq j = 1, \dots, n$.

Pro případ A budeme diskutovat

1. bodový odhad $\beta_0, \beta_1, \sigma^2$ a $\mu(x)$ pro libovolné $x \in D$
2. interval spolehlivosti pro $\beta_0, \beta_1, \sigma^2$ a $\mu(x)$ pro libovolné $x \in D$
3. testování hypotéz pro β_0, β_1 a σ^2

Pro případ B budeme diskutovat

1. bodový odhad $\beta_0, \beta_1, \sigma^2$ a $\mu(x)$ pro libovolné $x \in D$

10.3 Bodový odhad - případ A

V tomto případě jsou Y_1, \dots, Y_n nezávislé normální veličiny se středními hodnotami $\beta_0 + \beta_1 x_1, \dots, \beta_0 + \beta_1 x_n$ a rozptylem σ^2 . Pro účely bodového odhadu použijeme metodu maximální věrohodnosti. Věrohodnostní funkce má tvar

$$L(\beta_0, \beta_1, \sigma^2) = L(\beta_0, \beta_1, \sigma^2; y_1, \dots, y_n) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma} \right)^2} \right) \quad (10.1)$$

a

$$\ln(L(\beta_0, \beta_1, \sigma^2)) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

V dalším kroku jsou parciální derivace $\ln(L(\beta_0, \beta_1, \sigma^2))$ vzhledem k β_0, β_1 a σ^2 položeny rovno nule. Řešení těchto tří rovnic označme jako $\hat{\beta}_0, \hat{\beta}_1$ a $\hat{\sigma}^2$. Tyto tři rovnice mají po určitých zjednodušení tvar

$$\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (10.2)$$

$$\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0 \quad (10.3)$$

$$\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = n\hat{\sigma}^2 \quad (10.4)$$

První dvě rovnice jsou lineární v parametrech β_0 a β_1 , a proto z nich lze tyto parametry relativně jednoduše určit. Dosazením takto vypočtených hodnot do třetí rovnice lze vypočíst σ^2 . Řešení má pak konkrétně podobu

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (10.5)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (10.6)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \quad (10.7)$$

Protože

$$\begin{aligned} f_{Y_i}(y_i; \beta_0, \beta_1, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma} \right)^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2} (\beta_0 + \beta_1 x_i)^2} e^{-\frac{1}{2\sigma^2} y_i^2 + \frac{\beta_0}{\sigma^2} y_i + \frac{\beta_1}{\sigma^2} x_i y_i} \end{aligned}$$

je $f_{Y_i}(y_i; \beta_0, \beta_1, \sigma)$ členem tří parametrické exponenciální rodiny pravděpodobnostních rozdělení. Pro na základě zobecněné věty (7.19) představují

$$\sum_{i=1}^n Y_i^2 \quad \sum_{i=1}^n Y_i \quad \sum_{i=1}^n x_i Y_i \quad (10.8)$$

soubor minimálních dostatečných a sdruženě kompletních statistik. Navíc (10.8) je transformací jedna ku jedné statistik (10.2), (10.3) a (10.4), a proto jsou taktéž minimální dostatečné a sdruženě kompletní.

Pokusme se nalézt sdružené pravděpodobnostní rozdělení statistik odpovídajících $\hat{\beta}_0, \hat{\beta}_1$ a $\hat{\sigma}^2$. Nejprve nalezneme momentové funkce náhodných veličin $\hat{\Theta}_1, \hat{\Theta}_2$ a $\hat{\Theta}_3$, jejichž hodnoty jsou definovány jako

$$\hat{\theta}_1 = \frac{\hat{\beta}_0 - \beta_0}{\sigma} \quad \hat{\theta}_2 = \frac{\hat{\beta}_1 - \beta_1}{\sigma} \quad \hat{\theta}_3 = \frac{n\hat{\sigma}^2}{\sigma^2}$$

Z definice (4.25) vyplývá, že sdružená momentová funkce náhodných veličin $\hat{\Theta}_1, \hat{\Theta}_2, \hat{\Theta}_3$ je dána

$$m(t_1, t_2, t_3) = E[e^{t_1\hat{\Theta}_1 + t_2\hat{\Theta}_2 + t_3\hat{\Theta}_3}]$$

jestliže odpovídající střední hodnota existuje pro $-h < t_i < h$ a určité $h > 0$. Tímto se dostáváme k

$$m(t_1, t_2, t_3) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{t_1\hat{\theta}_1 + t_2\hat{\theta}_2 + t_3\hat{\theta}_3} \frac{e^{-1/2\sigma^2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2}}{(2\pi\sigma^2)^{n/2}} dy_1 \dots dy_n$$

kde jsme $\hat{\theta}_1, \hat{\theta}_2$ a $\hat{\theta}_3$ vyjádřili pomocí y_i a x_i . Řešením tohoto integrálu je

$$m(t_1, t_2, t_3) = e^{\frac{1}{2} \left(t_1^2 \frac{\sum_{i=1}^n x_i^2/n}{\sum_{i=1}^n (x_i - \bar{x})^2} - 2t_1 t_2 \frac{\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) + t_2^2 \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}} (1 - 2t_3)^{-\frac{n-2}{2}}$$

pro $t_3 < \frac{1}{2}$. Z této momentové funkce vyplývá několik věcí.

1. Tuto momentovou funkci lze rozložit na součin dvou funkcí, z nichž první obsahuje t_1 a t_2 zatímco druhá pouze t_3 . To znamená, že $m(t_1, t_2, t_3) = m_1(t_1, t_2)m_2(t_3)$. Z věty (4.15) vyplývá, že náhodné veličiny asociované s t_1 a t_2 jsou nezávislé na náhodných veličinách asociovaných s t_3 . To znamená, že $\hat{\Theta}_1$ a $\hat{\Theta}_2$ jsou nezávislé na $\hat{\Theta}_3$, což implikuje nezávislost odhadů β_0 a β_1 na odhadu σ^2 .
2. Připomeňme, že každé pravděpodobnostní rozdělení má jedinečnou momentovou funkci. Z věty (4.17) víme, že $m_1(t_1, t_2)$ je momentovou funkcí dvourozměrného normálního rozdělení. Náhodné veličiny \hat{B}_0 a \hat{B}_1 asociované s $\hat{\beta}_0$ a $\hat{\beta}_1$ tak sledují dvourozměrné normální rozdělení se středními hodnotami (β_0, β_1) a kovarianční maticí

$$\begin{bmatrix} \frac{\sigma^2 \sum_{i=1}^n x_i}{n \sum_{i=1}^n (x_i - \bar{x})^2} & -\frac{\sigma^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ -\frac{\sigma^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{bmatrix} \quad (10.9)$$

3. $m_2(t_3)$ je momentová funkce chi-kvadrát rozdělení s $n - 2$ stupni volnosti. Proto

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \hat{B}_0 - \hat{B}_1 x_i)^2$$

sleduje chi-kvadrát rozdělení s $n - 2$ stupni volnosti¹. Z tvrzení (6.3) víme, že $E \left[\frac{n\hat{\sigma}^2}{\sigma^2} \right] = n - 2$, a proto definujeme $\hat{\sigma}^2$ jako

$$\hat{\sigma}^2 = \frac{n}{n-2} \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

¹V následujícím textu označují $\hat{\sigma}^2$ resp. $\tilde{\sigma}^2$ náhodné veličiny s hodnotami $\hat{\sigma}^2$ resp. $\tilde{\sigma}^2$.

Shrňme výše uvedené závěry do následující věty.

Věta 10.1 *Uvažujme případ A jednoduché lineární regrese daný definicí (10.1). Funkce odhadu založené na metodě maximální věrohodnosti pro parametry β_0, β_1 a σ^2 (upravené o případné zkreslení) jsou*

$$\begin{aligned}\hat{B}_1 &= \frac{\sum_{i=1}^n (Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{B}_0 &= \bar{Y} - \hat{B}_1 \bar{x} \\ \hat{\sigma}^2 &= \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{B}_0 - \hat{B}_1 x_i)^2\end{aligned}$$

Tyto funkce odhadu splňují následující.

1. Jedná se o sdružené kompletní a dostatečné statistiky.
2. Jedná se o nezkreslené funkce odhadu příslušných parametrů.
3. (\hat{B}_0, \hat{B}_1) je nezávislé na $\hat{\sigma}^2$.
4. (\hat{B}_0, \hat{B}_1) sleduje dvourozměrné normální rozdělení se střední hodnotou (β_0, β_1) a kovarianční maticí (10.9).
5. $\frac{(n-2)\hat{\sigma}^2}{\sigma^2}$ sleduje chi-kvadrát rozdělení s $n-2$ stupni volnosti.

Z předchozího textu víme, že funkce odhadu založené na metodě maximální věrohodnosti mají řadu žádoucích vlastností, nicméně se nemusí jednat o nezkreslenou funkci odhadu s nejmenším rozptylem. S využitím zobecněné věty (7.20) a věty (10.1) definujme optimální vlastnosti funkcí odhadu $\hat{B}_0, \hat{B}_1, \hat{\sigma}^2$ parametrů $\beta_0, \beta_1, \sigma^2$.

Věta 10.2 *Uvažujme jednoduchý regresní model definovaný větou (10.1). Nechť $\tau(\beta_0, \beta_1, \sigma^2)$ je známá funkce parametrů β_0, β_1 a σ^2 , pro které existují nezkreslené funkce odhadu. Pak existuje nezkreslená funkce odhadu pro $\tau(\beta_0, \beta_1, \sigma^2)$, která je funkcí \hat{B}_0, \hat{B}_1 a $\hat{\sigma}^2$. Označme tuto funkci odhadu jako $t(\hat{B}_0, \hat{B}_1, \hat{\sigma}^2)$. $t(\hat{B}_0, \hat{B}_1, \hat{\sigma}^2)$ je UMVUE pro $\tau(\beta_0, \beta_1, \sigma^2)$.*

Důkaz 10.1 *Výše uvedená věta vyplývá ze zobecnění věty (7.20), protože $\hat{B}_0, \hat{B}_1, \hat{\sigma}^2$ je souborem dostatečných kompletních statistik.*

Tvrzení 10.1 *UMVUE pro každý z parametrů β_0, β_1 a σ^2 je, dle věty (10.1), dán \hat{B}_0, \hat{B}_1 a $\hat{\sigma}^2$.*

Tvrzení 10.2 *UMVUE pro $\mu(x) = \beta_0 + \beta_1 x$ pro libovolné $x \in D$ je $\hat{\mu}(x) = \hat{B}_0 + \hat{B}_1 x$. $\hat{\mu}(x)$ je náhodná veličina s hodnotami $\mu(x) = \beta_0 + \beta_1 x$.*

Tvrzení 10.3 *UMVUE pro $c_1 \beta_0 + c_2 \beta_1$, kde c_1 a c_2 jsou libovolné konstanty, je $c_1 \hat{B}_0 + c_2 \hat{B}_1$.*

10.4 Intervaly spolehlivosti - případ A

10.4.1 Interval spolehlivosti pro σ^2

Pro získání intervalu spolehlivosti pro σ^2 , si stačí uvědomit, že dle věty (10.1) sleduje

$$U = \frac{(n-2)\hat{\sigma}^2}{\sigma^2}$$

chi-kvadrát rozdělení s $n-2$ stupni volnosti. Proto je U centrální veličinou a platí

$$P[\chi_{(1-\gamma)/2}^2(n-2) \leq U \leq \chi_{(1+\gamma)/2}^2(n-2)] = \gamma$$

Jestliže dosadíme za U a zjednodušíme, získáme

$$P\left[\frac{(n-2)\hat{\sigma}^2}{\chi_{(1+\gamma)/2}^2(n-2)} \leq \sigma^2 \leq \frac{(n-2)\hat{\sigma}^2}{\chi_{(1-\gamma)/2}^2(n-2)}\right] = \gamma$$

což je 100γ procentní interval spolehlivosti pro σ^2 .

10.4.2 Interval spolehlivosti pro β_0

Připomeňme, že dle věty (10.1) platí

1. $Z = (\hat{B}_0 - \beta_0) / \sqrt{\sum (x_i - \bar{x})^2 n / \sigma^2 \sum x_i^2}$ sleduje normované normální rozdělení.
2. $U = (n-2)\hat{\sigma}^2 / \sigma^2$ sleduje chi-kvadrát rozdělení s $n-2$ stupni volnosti.
3. Z a U jsou vzájemně nezávislé.

Proto, dle věty (6.12), sleduje

$$T = \frac{\hat{B}_0 - \beta_0}{\hat{\sigma}} \sqrt{\frac{n \sum (x_i - \bar{x})^2}{\sum x_i^2}}$$

studentovo rozdělení s $n-2$ stupni volnosti. T je tedy centrální veličinou a platí

$$P[-t_{(1+\gamma)/2}(n-2) \leq T \leq t_{(1+\gamma)/2}(n-2)] = \gamma$$

Dosazením za T a následnými úpravami získáme

$$P\left[\hat{B}_0 - t_{(1+\gamma)/2}(n-2)\hat{\sigma} \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}} \leq \beta_0 \leq \hat{B}_0 + t_{(1+\gamma)/2}(n-2)\hat{\sigma} \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}}\right] = \gamma$$

Na základě věty (10.1) platí

$$D[\hat{B}_0] = \sigma^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}$$

a proto odhadovaný rozptyl \hat{B}_0 , který označíme jako $\hat{D}[\hat{B}_0]$, je tak roven

$$\hat{D}[\hat{B}_0] = \hat{\sigma}^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}$$

Interval spolehlivosti pak má tvar

$$P[\hat{B}_0 - t_{(1+\gamma)/2}(n-2)\sqrt{\hat{D}[\hat{B}_0]} \leq \beta_0 \leq \hat{B}_0 + t_{(1+\gamma)/2}(n-2)\sqrt{\hat{D}[\hat{B}_0]}] = \gamma$$

10.4.3 Interval spolehlivosti pro β_1

Připomeňme, že dle věty (10.1) platí

1. $Z = (\hat{B}_1 - \beta_1) \sqrt{\sum (x_i - \bar{x})^2 / \sigma^2}$ sleduje normované normální rozdělení.
2. $U = (n - 2) \hat{\sigma}^2 / \sigma^2$ sleduje chi-kvadrát rozdělení s $n - 2$ stupni volnosti.
3. Z a U jsou nezávislé.

Proto, dle věty (6.12), sleduje

$$T = \frac{\hat{B}_1 - \beta_1}{\hat{\sigma}^2} \sqrt{\frac{\sum (x_i - \bar{x})^2}{\hat{\sigma}^2}}$$

studentovo rozdělení s $n - 2$ stupni volnosti. T je tedy centrální veličinou a platí

$$P[-t_{(1+\gamma)/2}(n-2) \leq T \leq t_{(1+\gamma)/2}(n-2)] = \gamma$$

Dosazením za T a následnými úpravami získáme

$$P \left[-t_{(1+\gamma)/2}(n-2) \leq (\hat{B}_1 - \beta_1) \sqrt{\frac{\sum (x_i - \bar{x})^2}{\hat{\sigma}^2}} \leq t_{(1+\gamma)/2}(n-2) \right] = \gamma \quad (10.10)$$

neboli

$$P \left[\hat{B}_1 - t_{(1+\gamma)/2}(n-2) \sqrt{\frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}} \leq \beta_1 \leq \hat{B}_1 + t_{(1+\gamma)/2}(n-2) \sqrt{\frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}} \right] = \gamma$$

což je 100γ procentní interval spolehlivosti pro β_1 .

Připomeňme, že dle věty (10.1) platí

$$D[\hat{B}_1] = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

a že odhadovaný rozptyl \hat{B}_1 , který označíme jako $\hat{D}[\hat{B}_1]$, je roven

$$\hat{D}[\hat{B}_1] = \frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}$$

Dosazením do výše odvozeného intervalu spolehlivosti tak získáme

$$P[\hat{B}_1 - t_{(1+\gamma)/2}(n-2) \sqrt{\hat{D}[\hat{B}_1]} \leq \beta_1 \leq \hat{B}_1 + t_{(1+\gamma)/2}(n-2) \sqrt{\hat{D}[\hat{B}_1]}] = \gamma \quad (10.11)$$

10.4.4 Interval spolehlivosti pro $\mu(x)$

Z předchozího textu vyplývá

1. $\mu(x) = \beta_0 + \beta_1 x$
2. $\hat{\mu}(x) = \hat{B}_0 + \hat{B}_1 x$
3. $E[\hat{\mu}(x)] = \mu(x)$

4.

$$\begin{aligned}
D[\hat{\mu}(x)] &= D[\hat{B}_0] + \hat{B}_1 x = D[\hat{B}_0 + 2x \operatorname{cov}(\hat{B}_0, \hat{B}_1) + x^2 D[\hat{B}_1]] \\
&= \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \left(\frac{\sum x_i^2}{n} - 2x\bar{x} + x^2 \right) \\
&= \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \left((\bar{x} - x)^2 + \frac{1}{n} \sum (x_i - \bar{x})^2 \right) \\
&= \sigma^2 \left(\frac{1}{n} + \frac{(\bar{x} - x)^2}{\sum (x_i - \bar{x})^2} \right)
\end{aligned}$$

5. $Z = [\hat{\mu}(x) - \mu(x)] / \sqrt{D[\hat{\mu}(x)]}$ sleduje normované normální rozdělení.6. $U = (n-2)\hat{\mu}^2/\sigma^2$ sleduje chi-kvadrát rozdělení s $n-2$ stupni volnosti.7. U a Z jsou vzájemně nezávislé.8. $T = \frac{\hat{\mu}(x) - \mu(x)}{\sqrt{\hat{D}[\hat{\mu}(x)]}} = \frac{\hat{B}_0 + \hat{B}_1 x - \beta_0 - \beta_1 x}{\sqrt{\hat{\mu}^2 [1/n + (\bar{x} - x)^2 / \sum (x_i - \bar{x})^2]}}$ sleduje studentovo rozdělení s $n-2$ stupni volnosti. T je tak centrální veličinou a platí

$$P[-t_{(1+\gamma)/2}(n-2) \leq T \leq t_{(1+\gamma)/2}(n-2)] = \gamma$$

Dosažením za T a následnými úpravami získáme

$$\begin{aligned}
P \left[\hat{B}_0 + \hat{B}_1 x - t_{(1+\gamma)/2}(n-2) \sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{(\bar{x} - x)^2}{\sum (x_i - \bar{x})^2} \right]} \leq \beta_0 + \beta_1 x \right. \\
\left. \leq \hat{B}_0 + \hat{B}_1 x + t_{(1+\gamma)/2}(n-2) \sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{(\bar{x} - x)^2}{\sum (x_i - \bar{x})^2} \right]} \right]
\end{aligned}$$

což je 100γ procentní interval spolehlivosti pro $\mu(x) = \beta_0 + \beta_1 x$.

10.5 Testování hypotéz - případ A

V rámci lineárního regresního modelu existuje řada testů praktického významu, jako např.

1. test, zda-li regresní přímka prochází počátkem, tj. průsečík je nulový
2. test, zda-li je průsečík s osou y kladný
3. test, zda-li je průsečík s osou y záporný
4. test, zda-li je sklon regresní přímky nulový
5. test, zda-li je sklon regresní přímky kladný
6. test, zda-li je sklon regresní přímky záporný

což lze vyjádřit jako

1. $\mathcal{H}_0 : \beta_0 = 0$ vs. $\mathcal{H}_1 : \beta_0 \neq 0$

2. $\mathcal{H}_0 : \beta_0 \geq 0$ vs. $\mathcal{H}_1 : \beta_0 < 0$
3. $\mathcal{H}_0 : \beta_0 \leq 0$ vs. $\mathcal{H}_1 : \beta_0 > 0$
4. $\mathcal{H}_0 : \beta_1 = 0$ vs. $\mathcal{H}_1 : \beta_1 \neq 0$
5. $\mathcal{H}_0 : \beta_1 \geq 0$ vs. $\mathcal{H}_1 : \beta_1 < 0$
6. $\mathcal{H}_0 : \beta_1 \leq 0$ vs. $\mathcal{H}_1 : \beta_1 > 0$

10.5.1 $\mathcal{H}_0 : \beta_1 = 0$ vs. $\mathcal{H}_1 : \beta_1 \neq 0$

Pro testování $\mathcal{H}_0 : \beta_1 = 0$ vs. $\mathcal{H}_1 : \beta_1 \neq 0$ použijme statistiku

$$T = \frac{\hat{B}_1}{\sqrt{\hat{D}[\hat{B}_1]}}$$

Je-li \mathcal{H}_0 platné, sleduje T studentovo rozdělení s $n - 2$ stupni volnosti. Proto zamítneme \mathcal{H}_0 , jestliže $|T| > t_{1-\alpha/2}(n - 2)$, kde α je síla testu. Tento test je ekvivalentní s postupem, kdy pro $100(1 - \alpha)$ procentní interval spolehlivosti parametru β_1 zamítneme \mathcal{H}_0 , jestliže tento interval neobsahuje nulu.

Lze dokázat, že tento test je zobecněným věrohodnostním poměrovým testem. Pro $\mathcal{H}_0 : \beta_1 = 0$ vs. $\mathcal{H}_1 : \beta_1 \neq 0$ uvažujme parametrické prostory

$$\begin{aligned}\bar{\Theta} &= \{(\beta_0, \beta_1, \sigma^2) : -\infty < \beta_0 < \infty; -\infty < \beta_1 < \infty; \sigma^2 > 0\} \\ \bar{\Theta}_0 &= \{(\beta_0, \beta_1, \sigma^2) : -\infty < \beta_0 < \infty; \beta_1 = 0; \sigma^2 > 0\} \\ \bar{\Theta}_1 &= \bar{\Theta} - \bar{\Theta}_0\end{aligned}$$

Musíme určit λ , kde

$$\lambda = \frac{\sup_{\theta \in \bar{\Theta}_0} L(\theta; y_1, \dots, y_n)}{\sup_{\theta \in \bar{\Theta}} L(\theta; y_1, \dots, y_n)}$$

kde $\theta = (\beta_0, \beta_1, \sigma^2)$ a

$$L(\theta; y_1, \dots, y_n) = L(\beta_0, \beta_1, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum (y_i - \beta_0 - \beta_1 x_i)^2} \quad (10.12)$$

Hodnoty β_0, β_1 a σ^2 , které maximalizují tento výraz pro $\theta \in \bar{\Theta}$, jsou odhady založené na metodě maximální věrohodnosti a jsou dány rovnicemi (10.5) až (10.7). Tím dostáváme

$$\sup_{\theta \in \bar{\Theta}} L(\theta; y_1, \dots, y_n) = \frac{1}{(2\pi\tilde{\sigma}^2)^{n/2}} e^{-\frac{\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{2\tilde{\sigma}^2}} = (2\pi\tilde{\sigma}^2)^{-n/2} e^{-n/2}$$

kde $\tilde{\sigma}^2 = \frac{1}{n} \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$. Abychom získali $\sup_{\theta \in \bar{\Theta}} L(\theta; y_1, \dots, y_n)$ dosadíme $\beta_1 = 0$ do rovnice (10.12), čímž získáme

$$L(\beta_0, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum (y_i - \beta_0)^2}$$

Tato rovnice je však zároveň věrohodnostní funkcí náhodného výběru velikosti n z normálního rozdělení s střední hodnotou β_0 a rozptylem σ^2 . Hodnoty parametrů β_0 a σ^2 , které maximalizují věrohodnostní funkci $L(\beta_0, \sigma^2)$ jsou

$$\beta_0^* = \bar{y} \quad \hat{\sigma}^{*2} = \frac{1}{n} \sum (y_i - \bar{y})^2$$

což jsou funkce odhadu založené na metodě maximální věrohodnosti. Proto

$$\sup_{\theta \in \Theta} L(\theta; y_1, \dots, y_n) = (2\pi\hat{\sigma}^{*2})^{-n/2} e^{-\frac{1}{2\hat{\sigma}^{*2}} \sum (y_i - \beta_0^*)^2} = (2\pi\hat{\sigma}^{*2})^{-n/2} e^{-n/2}$$

a věrohodnostní poměr je tak roven

$$\lambda = \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^{*2}} \right)^2$$

Namísto λ se však zaměříme na $(n-2)(\lambda^{-2/n} - 1)$, což je monotónní funkce v λ , a proto poskytne ekvivalentní testovací funkci.

$$\lambda^{-2/n} - 1 = \frac{\hat{\sigma}^{*2} - \tilde{\sigma}^2}{\tilde{\sigma}^2} = \frac{\sum (y_i - \bar{y})^2 - \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}$$

Jestliže v čitateli použijeme vztah $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, dostaneme

$$\lambda^{-2/n} - 1 = \frac{\sum (y_i - \bar{y})^2 - \sum [(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})]^2}{\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}$$

a proto

$$(n-2)(\lambda^{-2/n} - 1) = \frac{\hat{\beta}_1^2 \sum (x_i - \bar{x})^2}{\hat{\sigma}^2} = \frac{\hat{\beta}_1^2 \sum (x_i - \bar{x})^2 / \sigma^2}{\hat{\sigma}^2 / \sigma^2}$$

což je poměr hodnot dvou nezávislých veličin, které sledují chi-kvadrát rozdělení s jedním stupněm volnosti v čitateli a $n-2$ stupni volnosti ve jmenovateli. Pouze připomeňme, že toto tvrzení je pravdivé za předpokladu platnosti $\mathcal{H}_0 : \beta_1 = 0$. Proto $(n-2)(\lambda^{-2/n} - 1)$ sleduje F rozdělení s jedním a $n-2$ stupni volnosti pro \mathcal{H}_0 . Obecný věrohodnostní test zamítne \mathcal{H}_0 , jestliže $\lambda \leq \lambda_0$, neboli jestliže

$$(n-2)(\lambda^{-2/n} - 1) \geq (n-2)(\lambda_0^{-2/n} - 1) = \lambda_0^*$$

což lze dále upravit do tvaru

$$\frac{\hat{\beta}_1^2 \sum (x_i - \bar{x})^2}{\hat{\sigma}^2} \geq \lambda_0^*$$

kde λ_0^* je zvoleno s ohledem na požadovanou velikost chyby typu I.

Všimněme si, že $(n-2)(\lambda^{-2/n} - 1)$ je kvadrát

$$\frac{\hat{B}_1}{\sqrt{\hat{D}[\hat{B}_1]}}$$

a připomeňme, že kvadrát studentova rozdělení s $n-2$ stupni volnosti odpovídá F rozdělení s jedním a $n-2$ stupni volnosti. Tímto jsme ověřili, že použijeme-li rovnici (10.11) pro testování $\mathcal{H}_0 : \beta_1 = 0$ vs. $\mathcal{H}_1 : \beta_1 \neq 0$, použijeme obecný věrohodnostní test.

10.5.2 Zobecnění pro $\mathcal{H}_0 : \beta_0 = b_0$ a $\mathcal{H}_0 : \beta_1 = b_1$

Výše uvedené závěry shrňme v následující větě.

Věta 10.3 Pro model lineární regrese dle definice (10.1) je obecný věrohodnostní poměrový test velikosti α pro $\mathcal{H}_0 : \beta_1 = b_1$ vs. $\mathcal{H}_1 : \beta_1 \neq b_1$, kde b_1 je daná konstanta, definován následovně. S pomocí rovnice (10.11) stanovíme $1 - \alpha$ procentní interval spolehlivosti parametru β_1 a zamítneme \mathcal{H}_0 , jestliže tento interval nezahrnuje b_1 .

Analogicky lze formulovat větu pro parametr β_0 .

Věta 10.4 Pro model lineární regrese dle definice (10.1) je obecný věrohodnostní poměrový test velikosti α pro $\mathcal{H}_0 : \beta_0 = b_0$ vs. $\mathcal{H}_1 : \beta_0 \neq b_0$, kde b_0 je daná konstanta, definován následovně. S pomocí rovnice (10.10) stanovíme $1 - \alpha$ procentní interval spolehlivosti parametru β_0 a zamítneme \mathcal{H}_0 , jestliže tento interval nezahrnuje b_0 .

10.6 Bodový odhad - případ B

V případě B jsou Y_1, \dots, Y_n párově nekorelované náhodné veličiny s středními hodnotami $\beta_0 + \beta_1 x_1, \dots, \beta_0 + \beta_1 x_n$ a rozptylem σ^2 . Protože sdružená pravděpodobnostní funkce není specifikována, nelze pro parametry β_0, β_1 a σ^2 získat funkce odhadu založené na metodě maximální věrohodnosti. V těchto případech však lze použít tzv. metodu nejmenších čtverců.

Definice 10.2 (Metoda nejmenších čtverců) Necht' $(Y_i, x_i), i = 1, \dots, n$ představuje n dvojic pozorování, které splňují lineární regresi danou definicí (10.1). Hodnoty parametrů β_0 a β_1 , které minimalizují sumu čtverců

$$\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$$

jsou odhady dle metody nejmenších čtverců.

Abychom našli funkce odhadu dle metody nejmenších čtverců pro parametry β_0 a β_1 , musíme nalézt hodnoty, které minimalizují

$$L(\beta_0, \beta_1) = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$$

což jsou současně hodnoty, které maximalizují funkci maximální věrohodnosti (10.1). Tímto se dostáváme k následující větě.

Věta 10.5 V případě B jednoduché lineární regrese dle definice (10.1) jsou funkce odhadu parametrů β_0 a β_1 dle metody nejmenších čtverců dány

$$\begin{aligned} \hat{B}_1 &= \frac{\sum (Y_i - \bar{Y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \\ \hat{B}_0 &= \bar{Y} - \hat{B}_1 \bar{x} \end{aligned}$$

Metoda nejmenších čtverců nedává funkci odhadu pro σ^2 , nicméně funkce odhadu tohoto parametru založená na funkcích odhadu β_0 a β_1 dle metody nejmenších čtverců je

$$\hat{\sigma}^2 = \frac{1}{n-2} \left(\sum_{i=1}^n (Y_i - \hat{B}_0 - \hat{B}_1 x_i)^2 \right)$$

V případě A měly funkce odhadu pro β_0, β_1 a σ^2 určité žádoucí optimální vlastnosti. Dle tvrzení (10.1) jsou \hat{B}_0 a \hat{B}_1 UMVUE funkce odhadu. To znamená, že v rámci všech nezkreslených funkcí odhadů parametrů β_0 a β_1 mají \hat{B}_0 a \hat{B}_1 definované větou (10.1) nejmenší rozptyl. V rámci případu B žádná taková žádoucí vlastnost garantována není - v případě A předpokládáme normální rozdělení Y_i , a proto nelze od případu B očekávat stejně silnou optimalitu.

V případě B omezíme množinu všech možných funkcí odhadu a na této omezené množině určíme funkce odhadu, které splňují stanovené podmínky optimality. Protože $E[Y_i] = \beta_0 + \beta_1 x_i$, lze β_0 a β_1 určit na základě očekávané hodnoty lineární funkce Y_i . V rámci této množiny lineárních funkcí definujeme nezkreslené funkce odhadu s minimálním rozptylem.

Definice 10.3 *Nechť jsou Y_1, \dots, Y_n pozorovatelné náhodné veličiny takové, že $E[Y_i] = \tau_i(\theta)$, kde $\tau_i(\cdot)$ jsou známé funkce, které obsahují neznámý parametr θ^2 . Pro odhad $\theta_j \in \theta$ uvažujeme pouze množinu funkcí odhadu, které jsou lineární v Y_i . V rámci tohoto souboru funkcí navíc uvažujeme pouze funkce, které jsou nezkresleným odhadem θ_j . Jestliže v této omezené množině funkcí odhadu existuje funkce odhadu pro θ_j , která má menší rozptyl než libovolná alternativní funkce, pak se jedná o tzv. nejlepší nezkreslenou lineární funkci odhadu θ_j .*

Uvedme větu, která uvádí optimální vlastnosti funkcí odhadu parametrů β_0 a β_1 odvozených pomocí metody nejmenších čtverců pro případ B.

Věta 10.6 (Gauss-Markova) *Uvažujme lineární model daný definicí (10.1) a předpokládejme, že platí předpoklady pro případ B. Pak jsou, pro metodu nejmenších čtverců, funkce odhadu parametrů β_0 a β_1 dané větou (10.5) nejlepší nezkreslenou lineární funkcí odhadu.*

Důkaz 10.2 *Důkaz provedeme pouze pro parametr β_0 . Důkaz pro β_1 je analogický.*

Protože uvažujeme pouze lineární funkce, má hledaná funkce odhadu tvar $\hat{B}_0 = \sum a_j Y_j$. Musíme určit konstantu a_j tak, že

1. $E[\hat{B}_0] = \beta_0$, tj. \hat{B}_0 je nezkreslenou funkcí odhadu parametru β_0
2. $D[\hat{B}_0]$ je minimální mezi všemi funkcemi odhadu, které splňují podmínku (1)

Z podmínky (1) vyplývá

$$\beta_0 = E[\hat{B}_0] = \sum a_j E[Y_j] = \sum a_j (\beta_0 + \beta_1 x_j)$$

čímž získáváme následující dvě rovnice, které musí být splněny.

$$\sum a_j = 1 \quad \sum a_j x_j = 0 \quad (10.13)$$

² θ může být i vektorem neznámých parametrů.

Pro rozptyl platí

$$\begin{aligned} D[\hat{B}_0] &= E[(\hat{B}_0 - \beta_0)^2] = E[(a_j Y_j - \beta_0)^2] \\ &= E\left[\left(\sum a_j(\beta_0 + \beta_1 x_j + E_j) - \beta_0\right)^2\right] \\ &= E\left[(\beta_0 \sum a_j + \beta_1 \sum a_j x_j + \sum a_j E_j - \beta_0)^2\right] \end{aligned}$$

S využitím omezení (10.13) pak získáme

$$D[\hat{B}_0] = E\left[\left(\sum a_j E_j\right)^2\right] = E\left[\sum_j a_j^2 E_j^2 + \sum_i \sum_{j, i \neq j} a_j a_i E_j E_i\right]$$

$E[E_i E_j]$ je pro $i \neq j$ rovno nule, protože předpokládáme, že E_i jsou párově nekorelované a jejich střední hodnota je rovna nule. Proto

$$D[\hat{B}_0] = \sigma^2 \sum a_j^2$$

Protože je σ^2 konstanta, musíme minimalizovat $\sum a_j^2$, abychom minimalizovali $D[\hat{B}_0]$. Proto je třeba nalézt takové konstanty a_j , které minimalizují $\sum a_j^2$ s ohledem na omezení (10.13). Aplikací Lagrangeových multiplikátorů dostáváme výraz

$$L = \sum a_j^2 - \lambda_1 \left(\sum a_j - 1\right) - \lambda_2 \sum a_j x_j$$

který je třeba minimalizovat. Odpovídající první derivace jsou

$$\frac{\partial L}{\partial a_t} = 2a_t - \lambda_1 - \lambda_2 x_t = 0, \quad t = 1, 2, \dots, n \quad (10.14)$$

$$\begin{aligned} \frac{\partial L}{\partial \lambda_1} &= -\sum a_j + 1 = 0 \\ \frac{\partial L}{\partial \lambda_2} &= -\sum a_j x_j = 0 \end{aligned}$$

Součtem přes prvních n rovnic a s využitím $\sum a_t = 1$ získáme

$$2 = n\lambda_1 + \lambda_2 \sum x_i \quad (10.15)$$

Jestliže j -tou rovnicí v (10.14) vynásobíme x_j a sečteme, získáme

$$2 \sum x_j a_j = \lambda_1 \sum x_j + \lambda_2 \sum x_j^2$$

což je vzhledem k $\sum a_j x_j = 0$ ekvivalentní s

$$\lambda_1 = -\lambda_2 \frac{\sum x_i^2}{\sum x_i} \quad (10.16)$$

Jestliže dosadíme (10.16) do (10.15), dostaneme

$$\lambda_2 = -\frac{2 \sum x_i / n}{\sum x_i^2 - n \bar{x}^2} = -\frac{2\bar{x}}{\sum (x_i - \bar{x})^2}$$

a

$$\lambda_1 = \frac{2 \sum x_i^2 / n}{\sum (x_i - \bar{x})^2}$$

Dosazením λ_1 a λ_2 do t -té rovnice v (10.14) a řešením pro a_t získáme

$$a_t = \frac{(\sum x_i^2 / n) - \bar{x}x_t}{\sum (x_i - \bar{x})^2}$$

Nejlepší nezkreslená lineární funkce odhadu parametru β_0 je tak

$$\hat{B}_0 = \sum a_t Y_t = \frac{\bar{Y} \sum x_i^2 - \bar{x} \sum Y_t x_t}{\sum (x_i - \bar{x})^2} = \bar{Y} - \hat{B}_1 \bar{x}$$

čímž ukončujeme důkaz.

Kapitola 11

Neparametrické metody

Při aplikaci obecného věrohodnostního poměru jsme pracovali s parametrickým prostorem, který definoval určitou rodinu pravděpodobnostních rozdělení. Při testování hypotéz jsme tak znali funkční předpis pravděpodobnostního rozdělení. Jediné, co jsme neznali, byly parametry tohoto rozdělení. Metody, které představíme v následujícím textu, se neopírají o znalost funkčního předpisu ani o znalost parametrů a jsou tak aplikovatelné pro široké spektrum pravděpodobnostních rozdělení¹. Význam těchto metod pro praxi je zřejmý, protože pravděpodobnostní rozdělení populace, ze které provádíme náhodný výběr, obvykle neznáme.

11.1 Kumulativní distribuční funkce

11.1.1 Výběrová kumulativní distribuční funkce

V kapitole (6.4.3) jsme výběrovou kumulativní distribuční funkci definovali jako

$$F_n(x) = \frac{1}{n}(\text{počet } X_i \text{ menších nebo rovných } x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i)$$

kde X_1, \dots, X_n představuje náhodný výběr z určité kumulativní distribuční funkce $F(\cdot)$. Dle věty (6.19) platí

$$P\left[F_n(x) = \frac{k}{n}\right] = \binom{n}{k} [F(x)]^k [1 - F(x)]^{n-k}, \quad k = 0, 1, \dots, n \quad (11.1)$$

kde $F_n(\cdot)$ je výběrovou kumulativní distribuční funkcí odpovídající $F(\cdot)$. Z rovnice (11.2) vyplývá

$$E[F_n(x)] = \sum_{k=0}^n \frac{k}{n} \binom{n}{k} [F(x)]^k [1 - F(x)]^{n-k} = F(x) \quad (11.2)$$

a podobně

$$D[F_n(x)] = \frac{1}{n} F(x) [1 - F(x)] \quad (11.3)$$

¹To znamená, že pro svou aplikaci nevyžadují určité konkrétní pravděpodobnostní rozdělení jako např. normální rozdělení.

Protože je $F_n(x)$ výběrovou střední hodnotou náhodných veličin $I_{(-\infty, x]}, \dots, I_{(-\infty, x]}(X_n)$, sleduje $F_n(x)$ dle centrálního limitního teorému asymptoticky normální rozdělení se střední hodnotou $F(x)$ a rozptylem $\frac{1}{n}F(x)[1 - F(x)]$.

Z rovnic (11.3) a (11.4) vyplývá, že pro fixní x je $F_n(x)$ nezkreslenou a MSE konzistentní funkcí odhadu pro $F(x)$ bez ohledu na její konkrétní funkční předpis. Protože nás však zajímá odhad $F(x)$ pro všechny hodnoty x , potřebujeme vědět, jak blízko je $F_n(x)$ k $F(x)$ sdruženě přes všechny hodnoty x . To znamená, že předmětem našeho zájmu je

$$P\left[\sup_{-\infty < x < \infty} |F_n(x) - F(x)| \xrightarrow{n \rightarrow \infty} 0\right] = 1$$

Tato rovnice, nazývaná Glivenko-Canteliovou větou, říká, že $F_n(x)$ konverguje s pravděpodobností k $F(x)$ pro libovolné x . Definujeme

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|$$

kde D_n náhodná veličina, která měří vzdálenost mezi $F_n(\cdot)$ a $F(\cdot)$. Rovnice (11.5) tedy říká, že $P[\lim_{n \rightarrow \infty} D_n = 0] = 1$. To znamená, že kumulativní distribuční funkce $F_{D_n}(\cdot)$ náhodné proměnné D_n konverguje k diskrétní kumulativní distribuční funkce s jediným bodem koncentrace v 0.

Namísto bodového odhadu $F(x) = P[X \leq x]$ může být někdy žádoucí bodový odhad $F(y) - F(x) = P[x < X < y]$ pro fixní $x < y$. Následující tvrzení ukazuje, že $F_n(y) - F_n(x)$ je nezkreslenou MSE konsistentní funkcí odhadu pro $F(y) - F(x)$.

Tvrzení 11.1

$$\text{cov}[F_n(x), F_n(y)] = \frac{1}{n}F(x)[1 - F(y)] \quad \text{pro } y \geq x \quad (11.4)$$

Důkaz 11.1

$$\begin{aligned} \text{cov}[F_n(x), F_n(y)] &= \text{cov}\left[\frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i), \frac{1}{n} \sum_{j=1}^n I_{(-\infty, y]}(X_j)\right] \\ &= \left(\frac{1}{n}\right)^2 \text{cov}\left[\sum_{i=1}^n I_{(-\infty, x]}(X_i), \sum_{j=1}^n I_{(-\infty, y]}(X_j)\right] = \left(\frac{1}{n}\right)^2 \sum_{i=1}^n \sum_{j=1}^n \text{cov}[I_{(-\infty, x]}(X_i), I_{(-\infty, y]}(X_j)] \\ &= \frac{1}{n} \text{cov}[I_{(-\infty, x]}(X_1), I_{(-\infty, y]}(X_1)] = \frac{1}{n} (E[I_{(-\infty, x]}(X_1)I_{(-\infty, y]}(X_1)] - E[I_{(-\infty, x]}(X_1)]E[I_{(-\infty, y]}(X_1)]) \\ &= \frac{1}{n}[F(x) - F(x)F(y)] = \frac{1}{n}F(x)[1 - F(y)] \end{aligned}$$

Z tvrzení (11.1) je zřejmé, že

$$D[F_n(y) - F_n(x)] = D[F_n(y)] - 2\text{cov}[F_n(x), F_n(y)] + D[F_n(x)] = \frac{1}{n}[F(y) - F(x)][1 - F(y) + F(x)]$$

z čehož vyplývá konzistence $F_n(y) - F_n(x)$ s $F(y) - F(x)$ ve smyslu MSE.

Namísto odhadu $P[x < X \leq y]$ uvažujme obecnější $P[X \in B]$, kde B představuje určitou množinu. Lze dokázat, že $\frac{1}{n} \sum_{i=1}^n I_B(X_i)$ je nezkreslenou funkcí odhadu pro $P[X \in B]$ a že

$$D \left[\frac{1}{n} \sum_{i=1}^n I_B(X_i) \right] = \frac{1}{n} P[X \in B](1 - P[X \in B])$$

a proto je tato funkce odhadu také MSE konzistentní.

11.1.2 Kolmogorov-Smirnovův test dobré shody

V předchozím textu jsme uvedli, že $F_n(x)$ asymptoticky sleduje normální rozdělení. Analogicky pak $\sqrt{n}[F_n(x) - F(x)]$ limitně sleduje normální rozdělení se střední hodnotou 0 a rozptylem $F(x)[1 - F(x)]$. Následující věta, kterou nebudeme dokazovat, uvádí limitní rozdělení

$$\sqrt{n}D_n = \sqrt{n} \sup_{-\infty < x < \infty} |F_n(x) - F(x)|$$

Věta 11.1 *Uvažujme nezávislé náhodné veličiny X_1, \dots, X_n pocházející ze stejné kumulativní distribuční funkce $F(\cdot)$. Definujme*

$$D_n = \mathfrak{d}_n(X_1, \dots, X_n) = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|$$

kde $F_n(x)$ představuje výběrovou kumulativní distribuční funkci. Pak

$$\begin{aligned} \lim_{n \rightarrow \infty} F_{\sqrt{n}D_n}(x) &= \lim_{n \rightarrow \infty} P[\sqrt{n}D_n \leq x] \\ &= \left[1 - 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 x^2} \right] I_{(0, \infty)}(x) = H(x) \end{aligned} \quad (11.5)$$

Kumulativní distribuční funkce (11.5) náhodné veličiny $\sqrt{n}D_n$ nezávisí na konkrétním funkčním předpisu $F(\cdot)$ - jedinou podmínkou je, že se jedná o spojitou funkci. To umožňuje použít D_n jako statistiku pro test dobré shody. Jako příklad předpokládejme, že testujeme $\mathcal{H}_0 : X_i \sim F_0(\cdot)$, kde $F_0(\cdot)$ je určitá známá spojitá kumulativní distribuční funkce. Jestliže je \mathcal{H}_0 pravdivé, pak

$$K_n = \mathfrak{k}_n(X_1, \dots, X_n) = \sqrt{n} \sup_{-\infty < x < \infty} |F_n(x) - F_0(x)| \quad (11.6)$$

sleduje přibližně stejné rozdělení jako $H(\cdot)$. Jestliže je \mathcal{H}_0 nepravdivé, pak má $F_n(\cdot)$ tendenci být blízko své skutečné kumulativní distribuční funkce, a proto $\sup_{-\infty < x < \infty} |F_n(x) - F_0(x)|$ má tendenci dosahovat relativně vysokých hodnot. Protože, je-li \mathcal{H}_0 pravdivé, shoduje se pravděpodobnostní rozdělení náhodné veličiny $K_n = \sqrt{n} \sup_{-\infty < x < \infty} |F_n(x) - F_0(x)|$ přibližně s $H(\cdot)$. Věta (11.1) dává přibližné rozdělení náhodné veličiny D_n , popř. pro $H(\cdot)$ existují tabulky pro různá n , s jejichž pomocí lze určit $k_{1-\alpha}$ takové, aby $1 - H(k_{1-\alpha}) = \alpha$, tj. $P[K_n > k_{1-\alpha}] \approx \alpha$. Hypotézu \mathcal{H}_0 pak zamítneme, jestliže $K_n > k_{1-\alpha}$. Kolmogor-Smirnovův test nám tedy říká, jak dobře náhodný výběr odpovídá určité specifikované kumulativní distribuční funkci $F_0(\cdot)$.

Příklad 11.1 V porodnici zaznamenali v průběhu několika dní následujících 37 časů narození dítěte.

19:02, 23:08, 03:56, 08:12, 8:40, 00:25, 01:24, 08:25, 14:02, 23:46, 10:07, 13:53, 18:45, 09:06, 15:57, 07:40, 03:02, 10:45, 15:06, 06:26, 16:44, 12:26, 14:17, 23:45, 05:08, 05:49, 06:32, 00:40, 13:30, 00:55, 15:22, 16:09, 19:46, 02:28, 10:06, 11:19 a 16:31

Pokusme se otestovat, zda-li čas narození dítěte sleduje uniformní rozdělení.

Hypotetická i výběrová kumulativní distribuční funkce jsou ilustrovány obrázkem (11.1)

OBRÁZEK

Kolmogorova statistika je rovna

$$\sqrt{n} \sup_x |F_n(x) - F(x)| = \sqrt{37} \left| \frac{31}{37} - \frac{1004}{1440} \right| \approx 0.85$$

Kritická hodnota pro $\alpha = 0.10$ je 1.22, a proto nemůžeme \mathcal{H}_0 zamítnout.

Kolmogor-Smirnovův test dobré shody předpokládá plně specifikovanou nulovou hypotézu, tj. hypotézu, která neobsahuje žádné neznámé parametry. Přírozenou otázkou je, zda-li je možné tento test rozšířit také na případy, kdy kumulativní distribuční funkce nulové hypotézy patří mezi určitou rodinu pravděpodobnostních rozdělení, tj. $\{F(\cdot; \theta) : \theta \in \Theta\}$. Pro takovéto hypotézy však již není $\sup_x |F_n(x) - F(x; \theta)|$ statistikou, protože závisí na neznámém parametru popř. parametrech θ . Jedním z možných řešení se zdá být nahradit θ jeho odhadem $\hat{\theta}$ podobně jako v případě klasického chi-kvadrátu testu dobré shody. Testovací statistika by pak měla podobu $\sup_x |F_n(x) - F(x; \hat{\theta})|$. Pravděpodobnostní rozdělení této statistiky závisí na hypotetické rodině pravděpodobnostních rozdělení uvažovaných v nulové hypotéze, a proto není známo. Ačkoliv existuje řada studií, které se tento problém snaží řešit, Kolmogor-Smirnovův test dobré shody nelze v praxi použít pro testování obecnějších hypotéz.

11.1.3 Hranice spolehlivosti pro kumulativní distribuční funkci

Větu (11.1) lze mimojiné použít také pro stanovení hranice spolehlivosti pro kumulativní distribuční funkci $F(\cdot)$, na které realizujeme náhodný výběr. Definujme k_γ takové, že $H(k_\gamma) = \gamma$, kde $H(\cdot)$ je kumulativní distribuční funkce definovaná rovnicí (11.5). Následující tabulka obsahuje hodnoty pro vybraná γ .

γ	0.99	0.95	0.90	0.85	0.80
k_γ	1.63	1.36	1.22	1.14	1.07

Z výše uvedeného vyplývá

$$P[\sqrt{n} \sup_x |F_n(x) - F(x)| \leq k_\gamma] \approx \gamma$$

nicméně

$$P[\sqrt{n}]$$

Kapitola 12

Příloha A

12.1 Řady a kombinatorika

12.1.1 Značení řady součtu a součinu

Definice 12.1 (Součet řady) *Součet členů $a_1 + a_2 + \dots + a_n$ se označuje jako $\sum_{i=1}^n a_i$.*

Příklad 12.1

$$\sum_{i=2}^5 (-1)^{i-2} i x^{2i} = 2x^4 - 3x^6 + 4x^8 - 5x^{10}$$

Definice 12.2 (Součin řady) *Analogicky součin členů $a_1 \cdot a_2 \cdots a_n$ značíme jako $\prod_{i=1}^n a_i$.*

Příklad 12.2

$$\prod_{i=1}^5 \left[a + (-1)^i \frac{i}{b} \right] = \left(a - \frac{1}{b} \right) \left(a + \frac{2}{b} \right) \left(a - \frac{3}{b} \right) \left(a + \frac{4}{b} \right) \left(a - \frac{5}{b} \right)$$

Věta 12.1 *V této větě předkládáme čtenáři seznam nejběžněji používaných řad součtu. Jednotlivé vztahy nebudeme dokazovat. Případný důkaz lze provést indukcí.*

$$\begin{aligned} \sum_{i=1}^n i &= \frac{n(n+1)}{2} \\ \sum_{i=1}^n i^2 &= \frac{n(n+1)(2n+1)}{6} \\ \sum_{i=1}^n i^3 &= \left(\frac{n(n+1)}{2} \right)^2 \\ \sum_{i=1}^n i^4 &= \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30} \end{aligned}$$

Věta 12.2 (Součet aritmetické posloupnosti) *První z výše uvedených rovnic lze použít k odvození součtu aritmetické posloupnosti.*

$$\sum_{i=1}^n [a + (i-1)d] = na + \frac{d}{2}n(n-1)$$

Věta 12.3 (Součet geometrické posloupnosti) *Součet geometrické posloupnosti je dán vztahem*

$$\sum_{i=0}^{n-1} ar^i = a \frac{1-r^n}{1-r}$$

12.1.2 Faktoriál a značení v kombinatorice

Definice 12.3 (Faktoriál) *Součin řady $1 \cdot 2 \cdot 3 \cdots n$ značíme $n!$ a nazýváme faktoriálem.*

$$n! = \prod_{i=0}^{n-1} (n-i)$$

Z definice platí $0! = 1$.

Definice 12.4 (Permutační koeficient) *Součin řady $(n-k+1) \cdot (n-k+2) \cdots n$ značíme jako $(n)_k$. Výraz $(n)_k$ tak vyjadřuje počet uspořádaných k -tic, které lze vybrat z n prvků.*

$$(n)_k = \prod_{i=1}^k (n-i+1)$$

Platí $(n)_k = \frac{n!}{(n-k)!}$ a $(n)_n = \frac{n!}{0!}$

Definice 12.5 (Binomický koeficient) *Binomický koeficient vyjadřuje počet neuspořádaných k -tic, které lze vybrat z n prvků.*

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{(n-k)!k!}$$

Z definice platí

$$\binom{n}{k} = 0 \quad \text{pro } k < 0 \text{ nebo } k > n$$

Věta 12.4

$$\begin{aligned} \binom{n}{0} &= \binom{n}{n} = 1 \\ \binom{n}{k} &= \binom{n}{n-k} \\ \binom{n+1}{k} &= \binom{n}{k} + \binom{n}{k-1} \quad \text{pro } n = 1, 2, \dots \text{ a } k = 0, \pm 1, \pm 2, \dots \end{aligned}$$

Jak permutační tak binomický koeficient mohou být zobecněny z libovolného přirozeného čísla n na libovolné reálné číslo t tak, že definujeme

$$\begin{aligned}(t)_k &= t(t-1) \cdots (t-k+1) \\ \binom{t}{k} &= \frac{t(t-1) \cdots (t-k+1)}{k!} \quad \text{pro } k = 1, 2, \dots \\ \binom{t}{k} &= 1 \quad \text{pro } k = 0\end{aligned}$$

Věta 12.5

$$\binom{-n}{k} = \frac{(-n)(-n-1) \cdots (-n-k+1)}{k!} = (-1)^k \frac{n(n+1) \cdots (n+k-1)}{k!} = (-1)^k \binom{n+k-1}{k}$$

12.1.3 Stirlingova aproximace

Definice 12.6 (Stirlingova aproximace) *Stirlingovu aproximaci lze použít pro přibližný výpočet faktoriálu. Podle této aproximace platí*

$$n! \approx (2\pi)^{\frac{1}{2}} e^{-n} n^{n+\frac{1}{2}}$$

nebo také

$$n! = (2\pi)^{\frac{1}{2}} e^{-n} n^{n+\frac{1}{2}} e^{r(n)/12n} \quad \text{kde } 1 - \frac{1}{12n+1} < r(n) < 1$$

Přesnost lze ilustrovat na příkladu $10!$, které je rovno 3 628 800, zatímco Stirlingovy aproximace dává 3 599 000. Celková chyba je tedy méně než jedno procento a tato odchylka se zmenšuje s tím, jak roste n .

12.1.4 Binomická věta

Definice 12.7 (Binomická věta) *Binomická věta má tvar*

$$(a+b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}$$

kde n je přirozené číslo.

Věta 12.6

$$\begin{aligned}(1+t)^n &= \sum_{i=0}^n \binom{n}{i} t^i \\ (1-t)^n &= \sum_{i=0}^n \binom{n}{i} (-1)^i t^i \\ 2^n &= \sum_{i=0}^n \binom{n}{i} \\ 0 &= \sum_{i=0}^n (-1)^i \binom{n}{i}\end{aligned}$$

Rozvojem obou stran vztahu $(1+x)^a(1+x)^b = (1+x)^{a+b}$ a následnými úpravami získáme

$$\sum_{i=0}^n \binom{a}{i} \binom{b}{n-i} = \binom{a+b}{n}$$

Věta 12.7 Binomickou větu lze zobecnit do podoby

$$\left(\sum_{j=1}^k\right)^n = \sum \frac{n!}{\prod_{i=1}^k n_i!} \prod_{i=1}^k a_i^{n_i}$$

kde součet je přes všechna přirozená čísla n_1, n_2, \dots, n_k jejichž součet je n .

Věta 12.8

$$\left(\sum_{j=1}^k a_j\right)^2 = \left(\sum_{i=1}^k a_i\right) \left(\sum_{j=1}^k a_j\right) = \sum_{i=1}^k \sum_{j=1}^k a_i a_j$$

12.2 Integrální a diferenciální počet

12.2.1 Úvod

Předpokládá se, že čtenář je obeznámen s konceptem limity, spojitosti, derivace, integrace a nekonečnými řadami. V této knize se velmi často odvoláváme na jednu specifickou limitu a to konkrétně limitu pro základ přirozeného logaritmu e .

$$\lim_{x \rightarrow \infty} (1+x)^{1/x} = e$$

Výše uvedený vztah lze dokázat pomocí logaritmu a následnou aplikací l'Hospitalova pravidla, které ve stručnosti představíme níže. S tímto vztahem je možné se setkat také v několika dalších variantách a to např.

$$\lim_{x \rightarrow \infty} (1+x^{-1})^x = e$$

$$\lim_{x \rightarrow 0} (1+\lambda x)^{1/x} = e^\lambda \quad \text{pro konstantu } \lambda$$

Definice 12.8 (l'Hospitalovo pravidlo) l'Hospitalovo pravidlo je velmi užitečné při odvozování limit. Jestliže $f(\cdot)$ a $g(\cdot)$ jsou funkce pro které $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow b} g(x) = 0$ a zároveň $\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}$ existuje, pak existuje také $\lim_{x \rightarrow a} \frac{f(x)}{g(x)}$ a platí

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}$$

Příklad 12.3 Nalezněme $\lim_{x \rightarrow 0} [(1/x)\ln(1+x)]$. Nechť $f(x) = \ln(1+x)$ a $g(x) = x$. Pak

$$\lim_{x \rightarrow 0} \frac{f'(x)}{g'(x)} = \lim_{x \rightarrow 0} \frac{1}{1+x} = 1 = \lim_{x \rightarrow 0} \left[\frac{1}{x} \ln(1+x) \right]$$

Definice 12.9 (Leibnitzovo pravidlo) Nechť $I(t) = \int_{g(t)}^{h(t)} f(x;t)dx$, kde $f(\cdot, \cdot), g(\cdot)$ a $h(\cdot)$ jsou derivovatelné. Pak

$$\frac{dI}{dt} = \int_{g(t)}^{h(t)} \frac{\partial f}{\partial t} dx + f(h(t);t) \frac{dh}{dt} - f(g(t);t) \frac{dg}{dt}$$

Věta 12.9 Mezi nejznámější aplikace Leibnitzova pravidla patří

$$\frac{d}{dt} \left[\int_{g(t)}^{h(t)} f(x) dx \right] = f(h(t)) \frac{dh}{dt} - f(g(t)) \frac{dg}{dt}$$

$$\frac{d}{dt} \left[\int_c^t f(x) dx \right] = f(t)$$

12.2.2 Taylorovy řady

Definice 12.10 (Taylorův rozvoj funkce jedné proměnné) Taylorův rozvoj funkce $f(x)$ kolem bodu $x = a$ je definován jako

$$f(x) = f(a) + f^{(1)}(a)(x-a) + \frac{f^{(2)}(a)(x-a)^2}{2!} + \dots + \frac{f^{(n)}(a)(x-a)^n}{n!} + R_n$$

kde

$$f^{(i)}(a) = \left. \frac{d^i f(x)}{dx^i} \right|_{x=a}$$

$$R_n = \frac{f^{(n+1)}(c)(x-a)^{n+1}}{(n+1)!}$$

$$a \leq c \leq x$$

R_n se nazývá zbytkem. O funkci $f(x)$ předpokládáme, že je derivovatelná alespoň do $n+1$ řádu. Jestliže není R_n příliš velké, pak je výše uvedená rovnice polynomickou aproximací n -tého stupně funkce $f(x)$. Odpovídající nekonečná řada konverguje v určitém intervalu, jestliže $\lim_{n \rightarrow \infty} R_n = 0$ v tomto intervalu.

Věta 12.10 Mezi nejznámější aplikace Taylorova rozvoje patří

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{j=0}^{\infty} \frac{x^j}{j!} \quad \text{pro } a = 0$$

$$(1-x)^t = \sum_{j=0}^{\infty} (-1)^j \binom{t}{j} \frac{x^j}{j!} = \sum_{j=0}^{\infty} \binom{t}{j} (-x)^j \quad \text{pro } a = 0 \quad -1 < x < 1$$

$$(1-x)^{-1} = \sum_{j=0}^{\infty} x^j \quad \text{pro } a = 0$$

$$(1-x)^{-2} = \sum_{j=0}^{\infty} (j+1)x^j \quad \text{pro } a = 0$$

$$(1-x)^{-t} = \sum_{j=0}^{\infty} \binom{-t}{j} (-x)^j = \sum_{j=0}^{\infty} \binom{n+j-1}{j} x^j \quad \text{pro } a = 0 \quad -1 < x < 1$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots = \sum_{j=0}^{\infty} (-1)^j \frac{x^{j+1}}{j+1} \quad \text{pro } a = 0 \quad -1 < x \leq 1$$

Definice 12.11 (Taylorův rozvoj funkce dvou proměnných) Taylorovy řady lze zobecnit také na funkce více proměnných. Např. Taylorův rozvoj funkce

$f(x, y)$ kolem bodů $x = a$ a $y = b$ lze vyjádřit jako

$$f(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) + \frac{1}{2!} [f_{xx}(a, b)(x - a)^2 + 2f_{xy}(a, b)(x - a)(y - b) + f_{yy}(a, b)(y - b)^2] + \dots$$

$$f_x(a, b) = \left. \frac{\partial f}{\partial x} \right|_{x=a, y=b}$$

$$f_{xy}(a, b) = \left. \frac{\partial^2 f}{\partial y \partial x} \right|_{x=a, y=b}$$

12.2.3 Gamma a beta funkce

Definice 12.12 (Gamma funkce) Gamma funkce $\Gamma(\cdot)$ je definována jako

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx \quad \text{pro } t > 0$$

Integrací per partes lze dokázat

$$\Gamma(t+1) = t\Gamma(t)$$

Proto, je-li $t = n$, platí

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

Je-li n celé číslo, pak

$$\Gamma\left(n + \frac{1}{2}\right) = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} \sqrt{\pi}$$

Definice 12.13 (Beta funkce) Funkce beta $B(\cdot, \cdot)$ je definována jako

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx \quad \text{pro } a > 0, b > 0$$

Pouhým prohozením proměnných ve výše uvedené rovnici lze dokázat $B(a, b) = B(b, a)$.

Mezi gamma a beta funkcí existuje následující vztah

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$