Bakalářské studium na MFF UK v Praze
Obecná matematika
Zaměření: Stochastika

Podrobnější rozpis okruhů otázek pro třetí část SZZ

Verze: 18. června 2014

---

1 Úvodní poznámky

- Smyslem SZZ by nemělo být toliko ukazovat schopnost „papouškovat“ definice či věty, ale spíše prokázat schopnost samostatně myslet a propojovat jednotlivé znalosti, resp. ukázat schopnost aktivně používat vybraný matematický aparát, resp. probrané metody. Otázka zadávaná studentovi v rámci SZZ tedy ne nutně obsahuje přímo pojem či okruh, který je explicitně vyjmenován na seznamu požadavků pro SZZ. V rámci otázek zadávaných u SZZ se typicky jedná o jeden z následujících požadavků, případně o výběr z následujících požadavků:

  – Matematicky korektní formulace věty nebo tvrzení a její důkaz, jestliže tento nevyžaduje znalost „triku“, ale spíše vzájemné propojení jednotlivých znalostí. Jedná se např. o důkaz nestrannosti vybraných odhadů nebo použití ZVČ při důkazu konzistence vybraných odhadů, důkaz Čebyševovy nerovnosti, použití CLV při odvození některých testů apod.

  – Vyřešení jednoduššího početního příkladu souvisejícího se zadanou otázkou (např. použití věty o transformaci pro odvození hustoty χ^2_1 rozdělení).

  – Návrh pravděpodobnostního modelu, formulace testovaných hypotéz a odvození (resp. náznak odvození) příslušného testu pro řešení „praktického“ problému.

  Zadávané problémy, resp. požadavky jsou vždy řešitelné pomocí aparátu, který se skrývá pod pojmy vyjmenovanými v oficiálním seznamu požadavků pro SZZ.

- V přehledu níže přesněji specifikujeme obsah jednotlivých okruhů zveřejněných v Karolínce v oficiálním seznamu požadavků pro SZZ. Dále se snažíme uvádět, jakým způsobem a do jaké hloubky je ověřována znalost jednotlivých termínů, pojmů, atp.

2 Oficiální požadavky dle Karolínky

1. **Základy teorie pravděpodobnosti**

   (1) Pravděpodobnostní prostor, podmíněná pravděpodobnost, Bayesova věta, nezávislost systému náhodných jevů, 0–1 zákony.

   (2) Náhodná veličina, náhodný vektor a jejich rozdělení, charakteristiky (střední hodnota, rozptyl, varianční matice, korelace atd.).

   (3) Charakteristická funkce a její použití, nezávislost náhodných veličin a vektorů, základní jedno- i mnohorozměrná diskrétní a spojitá rozdělení.

   (4) Transformace náhodné veličiny a náhodného vektoru.

   (5) Podmíněné rozdělení a podmíněná střední hodnota.

   (6) Typy konvergence náhodných veličin a vztahy mezi nimi, Čebyševova nerovnost, slabý a silný zákon velkých čísel, centrální limitní věta pro součet nezávislých stejně rozdělených náhodných veličin.

   (7) Cramérova-Sluckého věta.

2. **Základy matematické statistiky**

   (1) Náhodný výběr, uspořádaný náhodný výběr.

   (2) Bodové a intervalové odhady, nestrannost a konsistence odhadů.

   (3) Empirická distribuční funkce.

   (4) Principy testování hypotéz, Neymanovo-Pearsonovo lemma.

   (5) Fisherova informace, Rao-Cramérova věta, odhady metodou maximální věrohodnosti, asymptotické testy založené na maximální věrohodnosti.

   (6) Jednovýběrový, dvouvýběrový, párový t-test.

   (7) Jednovýběrové a dvouvýběrové testy pro vybrané parametrické problémy, test dobré shody na multinomické rozdělení, testy nezávislosti v dvourozměrných kontingenčních tabulkách.

3 Podrobnější rozpis požadavků

3.1 Základy teorie pravděpodobnosti

1. Pravděpodobnostní prostor, podmíněná pravděpodobnost, Bayesova věta, nezávislost systému náhodných jevů, 0–1 zákony.
   (a) Definice pravděpodobnostního prostoru, co je to σ-algebra, co je to pravděpodobnost a jaké jsou její vlastnosti. Nezávislost systému náhodných jevů.
   (b) Definice podmíněné pravděpodobnosti, věta o násobení pravděpodobností, věta o úplné pravděpodobnosti, Bayesova věta.
   (c) Cantelliho věta, Borelova věta.
2. Náhodná veličina, náhodný vektor a jejich rozdělení, charakteristiky (střední hodnota, rozptyl, varianční matice, korelace atd.).
   (a) Definice náhodné veličiny. Proč se vyžaduje měřitelnost? σ-algebra indukovaná náhodnou veličinou.
   (b) Distribuční funkce náhodné veličiny a její vlastnosti. Hustota (spojité, diskrétní) náhodné veličiny, její vztah k distribuční funkci a její využití pro výpočet $\mathrm{P}(X \in B)$.
   (c) Střední hodnota a rozptyl náhodné veličiny a jejich vlastnosti.
   (d) Momenty náhodné veličiny (absolutní, centrální). Momentová vytvořující funkce.
   (e) Definice náhodného vektoru a jeho distribuční funkce. Vlastnosti distribuční funkce dvourozměrného náhodného vektoru.
   (f) Sdružená versus marginální distribuční funkce (hustota) náhodných vektorů. Podmíněná hustota.
   (g) Kovariance, koeficient korelace, varianční matice a její vlastnosti.
   (h) Rozdělení a momenty měřitelné funkce náhodné veličiny (vektoru).
3. Charakteristická funkce a její použití, nezávislost náhodných veličin a vektorů.
   (a) Definice charakteristické funkce náhodné veličiny a náhodného vektoru, jejich vlastnosti.
   (b) Nezávislost náhodných veličin a vektorů.
   (c) Charakteristická funkce a nezávislost náhodných veličin, charakteristická funkce a transformace náhodných veličin, charakteristická funkce a součet náhodných veličin.
   (d) Charakteristická funkce a konvergence v distribuci.

   Poznámka. Součástí otázek z výše uvedených okruhů bude zpravidla i jednoduchý příklad, který prověří, zda student umí výše uvedené používat.
4. Základní jedno- i mnohorozměrná diskrétní a spojitá rozdělení.
   (a) Alternativní a binomické rozdělení: hustota, odvození střední hodnoty a rozptylu. Použití alternativního, resp. binomického rozdělení jakožto základního modelu pro dichotomickou odezvu, resp. součet nezávislých dichotomických znaků.
   (b) Multinomické rozdělení: definice, odvození střední hodnoty a varianční matice. Vztah mezi multinomickým a alternativním rozdělením. Představa o oblastech využití (základní pravděpodobnostní model pro kategoriální data).

(c) Poissonovo rozdělení: hustota, odvození střední hodnoty a rozptylu. Použití Poissonova rozdělení jakožto základního modelu pro počet (událostí, ...).

(d) Exponenciální rozdělení: hustota, odvození střední hodnoty a rozptylu, „bezpaměťovost“ exponenciálního rozdělení včetně důkazu. Použití exponenciálního rozdělení jakožto základního modelu pro čas do události, mezi událostmi, ...

(e) Jednorozměrné normální rozdělení: hustota, odvození střední hodnoty a rozptylu, informativně: hodnota šikmosti a špičatosti, charakteristická funkce. Vztah mezi $\mathcal{N}(0,\,1)$ a $\mathcal{N}(\mu,\,\sigma^2)$, symetrie hustoty, použití vztahu $\Phi(-x) + \Phi(x) = 1$ ve výpočtech. Odvození charakteristické funkce rozdělení $\mathcal{N}(\mu,\,\sigma^2)$, je-li známa charakteristická funkce rozdělení $\mathcal{N}(0,\,1)$. Představa o oblastech využití normálního rozdělení: zejména asymptotická inference (CLV, maximální věrohodnost, ...).

(f) Vícerozměrné normální rozdělení: definice. Informativně základní vlastnosti: vztah mezi nezávislostí a nekorelovaností, rozdělení lineární transformace, normalita marginálních a podmíněných rozdělení.

(g) Rozdělení χ^2: definice, odvození střední hodnoty a rozptylu (bez integrování na základě znalosti momentů normálního rozdělení). Odvození hustoty rozdělení χ^2_1. Odvození hustoty χ^2_2 pomocí konvoluce, je-li zadána hustota rozdělení χ^2_1. Konvergence rozdělení χ^2_n v distribuci pro $n \to \infty$ (s odůvodněním). Představa o oblastech využití rozdělení χ^2: intervaly spolehlivosti pro a testy o rozptylu ve výběru z normálního rozdělení, intervaly spolehlivosti a testy založené na asymptotické teorii maximální věrohodnosti, testy dobré shody a testy v kontingenčních tabulkách.

(h) Studentovo t rozdělení: definice. Vztah mezi rozdělením t_1 a Cauchyho rozdělením, informativně střední hodnota rozdělení t_1. Symetrie hustoty, použití této symetrie ve výpočtech (zejména při odvozování intervalů spolehlivosti a testů). Odvození (náznak odvození) hustoty rozdělení t_ν, je-li znám předpis hustoty rozdělení χ^2_ν. Konvergence rozdělení t_n v distribuci pro $n \to \infty$ (s odůvodněním). Představa o oblastech využití Studentova t rozdělení: intervaly spolehlivosti pro a testy o střední hodnotě ve výběru z normálního rozdělení, intervaly spolehlivosti pro a testy o rozdílu středních hodnot v nezávislých výběrech z normálního rozdělení.

(i) Fisherovo-Snedecorovo F rozdělení: definice. Odvození (náznak odvození) hustoty rozdělení F_{ν_1,ν_2}, je-li znám předpis hustoty rozdělení χ^2_ν. Představa o oblastech využití Fisherova-Snedecorova F rozdělení: srovnání rozptylů dvou nezávislých výběrů z normálního rozdělení, analýza rozptylu.

Poznámka. V žádném případě se neočekává znalost hustoty rozdělení χ^2_ν, t_ν nebo F_{ν_1,ν_2} zpaměti!

5. Transformace náhodné veličiny a náhodného vektoru.

(a) Formulace věty o hustotě transformace náhodné veličiny (též vícerozměrné).

(b) Postup výpočtu rozdělení po transformaci $\mathbb{R}^p \longrightarrow \mathbb{R}$.

(c) Formulace věty o konvoluci a její důkaz (lze ukázat, že se jedná o speciální případ použití postupu transformace $\mathbb{R}^2 \longrightarrow \mathbb{R}$).

(d) Praktické použití věty o transformaci na zadaný problém (jednodušší početní příklad vyžadující výpočet rozdělení po transformaci).

6. Podmíněné rozdělení a podmíněná střední hodnota.

   (a) Podmíněná střední hodnota náhodné veličiny, její definice a vlastnosti, podmíněná pravděpodobnost jako podmíněná střední hodnota.
   (b) Definice podmíněného rozdělení náhodné veličiny, věta o podmíněné hustotě, užití podmíněné hustoty při určení podmíněné střední hodnoty.
   (c) Určení podmíněné střední hodnoty pro zadaný problém (jednodušší příklady).

7. Typy konvergence náhodných veličin a vztahy mezi nimi, Čebyševova nerovnost, slabý a silný zákon velkých čísel, centrální limitní věta pro součet nezávislých stejně rozdělených náhodných veličin.

   (a) Konvergence skoro jistě a konvergence v pravděpodobnosti a jejich vztah. Metrizovatelnost těchto konvergencí.
   (b) Konvergence v distribuci, její vztah ke konvergenci v pravděpodobnosti.
   (c) Definice silného (slabého) zákona velkých čísel.
   (d) Slabý zákon velký čísel a jeho důkaz pomocí Čebyševovy nerovnosti.
   (e) Silný zákon velkých čísel pro stejně a nestejně rozdělené náhodné veličiny, porovnání předpokladů.
   (f) Centrální limitní věta (CLV) pro nezávislé stejně rozdělené náhodné veličiny (bez důkazu). Použití CLV na binomické rozdělení a v dalších příkladech.
   (g) Použití zákona velkých čísel k důkazu konzistence zadaného statistického odhadu.

   Poznámka. Součástí otázek z výše uvedených okruhů může být některý z následujících úkolů: (a) vyšetřit, zda daná konkrétní posloupnost náhodných veličin splňuje silný (slabý) zákon velkých čísel a co nám tento zákon velkých čísel dává; (b) využití CLV na řešení různých slovních úloh.

8. Cramérova-Sluckého věta.

   (a) Formulace věty (bez důkazu).
   (b) Použití při konstrukci asymptotických intervalů spolehlivosti a asymptotických testů založených na CLV a konzistentním odhadu neznámého rozptylu, resp. neznámých rozptylů (asymptotická varianta t-testu, . . .).

3.2 Základy matematické statistiky

1. Náhodný výběr, uspořádaný náhodný výběr.

   (a) Definice náhodného výběru.
   (b) Použití náhodného výběru jakožto pravděpodobnostního modelu pro data.
   (c) Definice uspořádaného náhodného výběru, odvození rozdělení (distribuční funkce, případně hustota) první a poslední pořádkové statistiky (minimum, maximum).
   (d) Výběrové kvantily.

2. Bodové a intervalové odhady, nestrannost a konsistence odhadů.

(a) Definice parametrické třídy rozdělení.

(b) Definice bodového a intervalového odhadu, jednostranné intervaly spolehlivosti.

(c) Schopnost srozumitelně vysvětlit, co si představit pod pojmem „interval spolehlivosti se spolehlivostí $(1-\alpha)$ %", chápat náhodnost mezí intervalu spolehlivosti, umět vysvětlit, odkud se tato náhodnost bere.

(d) Definice nestrannosti, konzistence odhadu a vysvětlení smyslu těchto definic.

(e) Definice asymptotické nestrannosti odhadu, resp. asymptotického intervalu spolehlivosti.

(f) Důkaz nestrannosti zadaného odhadu.

(g) Důkaz konzistence zadaného odhadu, zejména je-li možné využít zákona velkých čísel.

(h) Směrodatná odchylka (směrodatná chyba) odhadu a její význam a použití při konstrukci asymptotických intervalů spolehlivosti.

(i) Vlastnosti výběrového průměru a výběrového rozptylu v náhodném výběru z normálního rozdělení: nestrannost a konzistence včetně důkazu, normální rozdělení výběrového průměru včetně důkazu. Informativně: nezávislost výběrového průměru a výběrového rozptylu, χ^2 rozdělení výběrového rozptylu (po přenásobení vhodnou konstantou).

(j) Bodový a intervalový odhad střední hodnoty a rozptylu v náhodném výběru z normálního rozdělení $\mathcal{N}(\mu, \sigma^2)$ včetně odvození. Schopnost vysvětlit, z čeho plyne Studentovo t-rozdělení statistiky $\dfrac{\overline{X}_n - \mu}{\frac{S_n}{\sqrt{n}}}$. Podmínky, za jakých lze odvozený intervalový odhad střední hodnoty použít v případě, že je porušena normalita výběru.

(k) Bodový a intervalový odhad rozdílu středních hodnot dvou nezávislých náhodných výběrů z normálních rozdělení s neznámým shodným rozptylem včetně odvození. Podmínky, za jakých lze intervalový odhad rozdílu středních hodnot použít v případě, že je porušena normalita jednoho nebo druhého výběru.

(l) Bodový a intervalový odhad podílu rozptylů, resp. směrodatných odchylek dvou nezávislých náhodných výběrů z normálních rozdělení včetně odvození.

(m) Bodový a intervalový odhad střední hodnoty (pravděpodobnosti „úspěchu") v náhodném výběru z alternativního rozdělení založený na CLV a Cramérově-Sluckého větě včetně odvození.

(n) Bodový a intervalový odhad rozdílu středních hodnot (pravděpodobností „úspěchu") dvou nezávislých náhodných výběrů z alternativních rozdělení založený na CLV a Cramérově-Sluckého větě včetně odvození.

(o) Bodový a intervalový odhad střední hodnoty ve výběru z Poissonova rozdělení založený na CLV a Cramérově-Sluckého větě včetně odvození.

(p) Bodový a intervalový odhad střední hodnoty ve výběru z exponenciálního rozdělení založený na CLV a Cramérově-Sluckého větě včetně odvození.

3. Empirická distribuční funkce.

(a) Empirická distribuční funkce (edf) a její vlastnosti: nestrannost a konzistence včetně důkazu.

(b) Idea využití edf při konstrukci Kolmogorovova-Smirnovova testu.

(c) Použití Kolmogorovova-Smirnovova testu pro řešení prakticky zadaného problému.

4. Principy testování hypotéz, Neymanovo-Pearsonovo lemma.

(a) Formulace problému testování hypotéz, vysvětlení základních pojmů: nulová a alternativní hypotéza, chyba prvního a druhého druhu, hladina testu, síla testu, p-hodnota.

(b) Vysvětlení, proč nelze statistickým testem prokázat platnost nulové hypotézy. Obvyklý postup formulace hypotéz v praktické situaci: tvrzení, jehož platnost chceme prokázat $\longrightarrow$ alternativní hypotéza, následně formulace nulové hypotézy, výjimka: testy dobré shody.

(c) Vysvětlení principu odvození testu: a) testová statistika, která nějakým způsobem odráží (ne)shodu dat s nulovou hypotézou, resp. (ne)podporu alternativní hypotézy, b) rozdělení testové statistiky při nulové hypotéze (proč je potřeba ho znát při nulové a nikoliv při alternativní hypotéze), vysvětlení, odkud se bere náhodnost testové statistiky, když data jsou v konkrétní situaci pouze jedna, c) kritický obor, d) p-hodnota, e) souvislost kritického oboru s výpočtem p-hodnoty.

(d) Souvislost mezi testováním hypotéz a intervaly spolehlivosti (v případě parametrických testů). Odvození intervalu spolehlivosti „invertováním“ testové statistiky souvisejícího testu.

(e) Formulace Neymanova-Pearsonova lemmatu v případě jednoduché hypotézy i alternativy. Praktické použití při odvození testu v zadaném příkladě (test o pravděpodobnosti úspěchu v alternativním rozdělení, test o střední hodnotě v Poissonově rozdělení, test o střední hodnotě v exponenciálním rozdělení, test o rozptylu v normálním rozdělení, . . .).

5. Fisherova informace, Rao-Cramérova věta, odhady metodou maximální věrohodnosti, asymptotické testy založené na maximální věrohodnosti.

(a) Věrohodnost náhodného výběru z rozdělení z parametrické třídy rozdělení. Logaritmická věrohodnost, skórová funkce, Fisherova informace (pozorovaná a očekávaná).

(b) Princip odhadu metodou maximální věrohodnosti (ML).

(c) Formulace vět o konzistenci ML odhadu, o asymptotické normalitě skórové funkce a ML odhadu (bez důkazu). Není nutné znát do detailu všechny podmínky regularity.

(d) Vysvětlení principů, na jejichž základě fungují testy založené na asymptotické teorii maximální věrohodnosti (Waldův test, skórový test, test poměrem věrohodností).

(e) Princip odvození asymptotických intervalů spolehlivosti založených na Waldově a skórovém testu, resp. testu poměrem věrohodností. Význam směrodatné odchylky (směrodatné chyby) odhadu pro konstrukci Waldova intervalu spolehlivosti.

(f) Odvození maximálně věrohodného odhadu, jeho asymptotického rozptylu, souvisejících asymptotických testů a intervalů spolehlivosti pro zadaný problém (např. odhad parametru λ ve výběru z Poissonova rozdělení se střední hodnotou λ).

(g) Formulace Rao-Cramérovy věty (bez důkazu). Vysvětlení vztahu mezi Rao-Cramérovou větou a teorií maximální věrohodnosti.

6. Jednovýběrový, dvouvýběrový, párový t-test.

(a) Podmínky, za jakých je možné použít jednotlivé testy i bez platnosti předpokládaného normálního rozdělení.

(b) Význam párování při eliminaci výběrové variability.

(c) Viz též společné poznámky pro okruhy **6** a **7** níže.

7. Jednovýběrové a dvouvýběrové testy pro vybrané parametrické problémy, test dobré shody na multinomické rozdělení, testy nezávislosti v dvourozměrných kontingenčních tabulkách.

(a) Jednovýběrový test o rozptylu normálního rozdělení.
(b) Dvouvýběrový test o shodě rozptylů dvou nezávislých výběrů z normálního rozdělení.
(c) Test o střední hodnotě (pravděpodobnosti „úspěchu") alternativního rozdělení založený na CLV.
(d) Test o střední hodnotě rozdělení splňujícího podmínky pro platnost CLV (Poissonovo, exponenciální, ...) založený na CLV a případně Cramérově-Sluckého větě.
(e) Test o shodě středních hodnot (pravděpodobností „úspěchu") dvou nezávislých výběrů z alternativního rozdělení založený na CLV a Cramérově-Sluckého větě. Souvislost s χ^2-testem nezávislosti v kontingenční tabulce 2×2.
(f) χ^2-test dobré shody na multinomické rozdělení, umět napsat tvar testové statistky a vysvětlit její princip, znát asymptotické rozdělení této statistiky za nulové hypotézy (χ^2 včetně stupňů volnosti), není však nutné toto umět dokázat.
(g) χ^2-test nezávislosti, umět napsat tvar testové statistky a vysvětlit její princip, znát asymptotické rozdělení této statistiky za nulové hypotézy (χ^2 včetně stupňů volnosti), není však nutné toto umět dokázat. Souvislost χ^2-test nezávislosti s testem homogenity několika multinomických rozdělení.
(h) Viz též společné poznámky pro okruhy **6** a **7** níže.

3.2.1 Společné poznámky pro okruhy 6 a 7

Zadávané otázky odvozené z těchto okruhů budou vesměs spočívat v popisu „reálného" problému a požadavku navrhnout řešení tohoto problému.

> **Příklad.** *Mezi náhodně oslovenými návštěvníky obchodního domu Máj bylo zaznamenáno jejich pohlaví a odpověď na otázku, zda preferují na chlebu namazaný margarín nebo máslo. Zformulujte pravděpodobnostní model a na jeho základě odvoďte statistický test, pomocí něhož lze zjistit, zda preference másla/margarínu mezi návštěvníky obchodního domu Máj závisí na pohlaví.*

V rámci odpovědi se očekává následující (případně podmnožina následujícího):

- Formulace vhodného pravděpodobnostního modelu.
- Formulace testovaných hypotéz.
- Návrh testové statistiky, odvození (důkaz) jejího (asymptotického) rozdělení za nulové hypotézy. Podrobné odůvodnění rozdělení testové statistiky se nevyžaduje u χ^2 testů dobré shody a nezávislosti (viz výše).
- Návrh kritického oboru a důkaz, že odpovídá testu na zadané hladině α.
- Způsob výpočtu p-hodnoty testu.
- Jedná-li se o parametrický test, odvození souvisejícího intervalu spolehlivosti.

Obdobným způsobem mohou být zadány taktéž některé otázky směřující k vysvětlení či konstrukci intervalových odhadů (okruhy **2**j, **2**k, **2**l, **2**m, **2**n, **2**o, **2**p).