evropský sociální fond v ČR

EVROPSKÁ UNIE

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Jednofázové odhady pro populace kontinua

Odhady úhrnů, středních hektarových hodnot a podílů

Radim Adolt

Ústav pro hospodářskou úpravu lesů Brandýs nad Labem (ÚHÚL), pobočka Kroměříž, Analyticko-metodické Centrum Národní Inventarizace Lesů (ACNIL)

Obsah prezentace

Používaná terminologie I

Jednofázové odhady

Výběrové šetření probíhá pouze v jedné fázi. Tento typ odhadů používá **jediný zdroj dat - typicky data pozemního šetření**, která jsou považována za **absolutně bezchybná**. Žádné pomocné zdroje dat nevstupují do procesu odhadu parametrů.

Dvoufázové odhady

Dvoufázové odhady používají:

1. **data první fáze šetření - pomocná data, pomocný zdroj informací**, typicky se jedná o jednu nebo více GISových vrstev, tato data nemusí být a v praxi nebývají bezchybná ani je za taková nepovažujeme, nemusí mít ani homogenní kvalitu napříč zájmovou oblastí
2. **data druhé fáze šetření - bezchybná (absolutně přesná data, pozemní pravda)**, typicky data pozemního šetření

Různé varianty dvoufázových odhadů jsou dány nejen povahou dat první fáze, ale také detaily vlastních výpočtů.

Používaná terminologie II

Geografická doména D

Pokud není řečeno jinak, je D zájmovou oblastí, geografickou doménou.
Cílem je odhadnout hodnoty vybraných parametrů populace v D.
Rozloha D je známa a máme k dispozici též absolutně přesnou mapu D.

Přímé versus nepřímé odhady

Rozhoduje původ do odhadu vstupujících dat:

- **přímé odhady** používají pouze **data pořízená v nebo vztahující se k zájmové oblasti** D
- **nepřímé odhady** čerpají **informace též z oblastí mimo** D, pro kterou je odhad počítán

Všechny odhady prezentované v rámci kurzu jsou **striktně vzato přímé**.

Používaná terminologie III

Design-based versus model-based koncept

Odlišné pojetí nejistoty odhadu:

1. **design-based - nejistota vlivem výběru**
   - **odhad je založen pouze na části populace** - na výběru
   - **populace je neměnná** - např. stromy nemohou změnit svoji polohu ani atributy
   - **výběr se při opakování šetření s největší pravděpodobností změní, což vede k odlišným hodnotám odhadů**
   - **robustnost** - odhady nejsou ovlivněny žádnými nebo jen minimem předpokladů
   - **snažší přijetí výsledků šetření koncovými uživateli**
2. **model-based - nejistota vyplývá ze stochastické povahy procesu, jehož výsledkem je předmětná populace**
   - populace je považována za jeden z prvků tzv. **superpopulace**
   - superpopulace je množina všech možných realizací stochastického procesu
   - pro konkrétní výběr je **přesnost odhadu dána variabilitou superpopulace**
   - **vlastnosti odhadů silně závisí na kvalitě modelu** popisujícího stochastický proces, jehož realizací je studovaná populace
   - **koncového uživatele je třeba přesvědčit** o správnosti modelu

Používaná terminologie IV

Geografická aditivita odhadu úhrnu

Geografická aditivita odhadu úhrnu v D je definována

$$\hat{Y}_{\bigcup_{i=1}^{k} D_i} = \sum_{i=1}^{k} \hat{Y}_{D_i}. \tag{1}$$

Aditivní odhad vypočtený pro libovolnou doménu danou sjednocením $\bigcup_{i=1}^{k} D_i$ subdomén $D_1, D_2 \ldots D_k$ **je přesně roven součtu odhadů téhož typu** vypočtených jednotlivě **pro každou z k subdomén**.

Používaná terminologie V

Nestrannost nepodmíněná a podmíněná velikostí výběru

Pro výběrové designy, při jejichž opakované realizaci může dojít k umístění různých počtů inventarizačních bodů do zájmové oblasti D - **designy s nekonstantní velikostí výběru v** D, můžeme **nestrannost odhadů** definovat dvěma odlišnými způsoby:

1. **očekávaná hodnota (vyhodnocená skrze všechny možné výběry bez ohledu na jejich velikost n_D)** odhadu $\hat{\theta}$ **nestranného nepodmíněně na realizované velikosti výběru** v D je rovna skutečné hodnotě odhadovaného parametru předmětné populace v D

$$\mathbb{E}\left[\hat{\theta}\right] = \theta \tag{2}$$

2. **očekávaná hodnota (vyhodnocená skrze všechny možné výběry dané velikosti, $n_D = konst.$)** odhadu $\hat{\theta}$ **nestranného podmíněně na realizované velikosti výběru** v D je rovna skutečné hodnotě odhadovaného parametru předmětné populace v D

$$\mathbb{E}\left[\hat{\theta} \mid n_D\right] = \theta \tag{3}$$

Považujeme-li odhady nestranné podmíněně na všech možných velikostech výběru za tentýž odhad, platí, že **odhady nestranné podmíněně na realizované velikosti výběru n_D, jsou nestranné též nepodmíněně na n_D.**

Používaná terminologie VI

Konzervativnost a anti-konzervativnost odhadu rozptylu

Dle vztahu mezi očekávanou hodnotou odhadu rozptylu a skutečnou hodnotou rozptylu rozlišujeme následující varianty:

1. očekávaná hodnota **konzervativního odhadu rozptylu** $\hat{\mathbb{V}}_{ktv}(\hat{\theta})$ **je vyšší** než skutečná hodnota rozptylu

$$\mathbb{E}\left[\hat{\mathbb{V}}_{ktv}(\hat{\theta})\right] > \mathbb{V}(\hat{Y}) \tag{4}$$

2. očekávaná hodnota **anti-konzervativního odhadu rozptylu** $\hat{\mathbb{V}}_{atv}(\hat{\theta})$ **je nižší** než skutečná hodnota rozptylu

$$\mathbb{E}\left[\hat{\mathbb{V}}_{atv}(\hat{\theta})\right] < \mathbb{V}(\hat{Y}) \tag{5}$$

Stejně jako nestrannost odhadů, také konzervativnost může být posuzována **podmíněně nebo nepodmíněně na** n_D - velikosti výběru (počtu inventarizačních bodů umístěných v D).

Obecná teorie HTC, úhrn v D [Cordy, 1993] I

Nechť Y je **úhrn** (suma) požadovaného atributu populace v oblasti D tj. v **nekonečně velké množině bodů** x **kontinua**.

Definice úhrnu v D

$$Y = \int_D Y(x)dx = \lambda(D)\bar{Y} \quad (6)$$

Nestranný odhad $\hat{Y}$ **úhrnu** Y je dán (7), kde symbol $Y(x)$ je **lokální hustota** veličiny zjištěná na bodě x výběru s **pevné velikosti** n, $\pi(x)$ je hodnota **funkce hustoty výběru** na bodě x - zjednodušeně řečeno, očekávaný počet inventarizačních bodů vybraných na jednotku plochy vztažený k danému místu v D.

Odhad úhrnu v D

$$\hat{Y} = \sum_{x \in s} \frac{Y(x)}{\pi(x)} \quad (7)$$

Odhad (7) **je nestranný**, pokud je funkce $Y(x)$ kladná nebo omezená v D a platí $\pi(x) > 0 \ \forall x \in D$ (kterýkoli bod $x \in D$ může být vybrán).

Obecná teorie HTC, úhrn v D [Cordy, 1993] II

Pokud je funkce lokální hustoty $Y(x)$ omezená, platí $\pi(x) > 0 \ \forall x \in D$ a zároveň platí $\int_D \frac{1}{\pi(x)} dx < \infty$, lze rozptyl odhadu úhrnu vyjádřit pro populace kontinua upraveným **Horwitz-Thomsonovým (HTC)** nebo **Sen-Yates-Grundyho (SYG)** vzorcem.

HTC rozptyl odhadu úhrnu v D

$$\mathbb{V}_{HTC}(\hat{Y}) = \int_D \frac{Y^2(x)}{\pi(x)} dx + \int_D \int_D Y(x_i) Y(x_j) \left[\frac{\pi(x_i, x_j) - \pi(x_i)\pi(x_j)}{\pi(x_i)\pi(x_j)} \right] dx_i dx_j \qquad (8)$$

SYG rozptyl odhadu úhrnu v D

$$\mathbb{V}_{SYG}(\hat{Y}) = \frac{1}{2} \int_D \int_D \left[\pi(x_i)\pi(x_j) - \pi(x_i, x_j) \right] \left[\frac{Y(x_i)}{\pi(x_i)} - \frac{Y(x_j)}{\pi(x_j)} \right]^2 dx_i dx_j \qquad (9)$$

Existence **design-based, nestranných odhadů rozptylů HTC a SYG** je závislá na splnění podmínky $\pi(x_i, x_j) > 0 \ \forall x_i, x_j \in D$, která požaduje, aby libovolné dva body x_i, x_j zájmové oblasti D (kontinua) mohly být daným designem vybrány současně.

Horwitz-Thompsonův odhad rozptylu úhrnu upravený pro **výběr z populace kontinua (HTC)** je dán dvojicí ekvivalentních vzorců.

HTC odhady rozptylu úhrnu v D

$$\hat{\mathbb{V}}_{HTC}\left(\hat{Y}\right) = \sum_{x_i \in s} \left[\frac{Y(x_i)}{\pi(x_i)}\right]^2 + \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_i \neq x_j}} Y(x_i) Y(x_j) \left[\frac{\pi(x_i, x_j) - \pi(x_i)\pi(x_j)}{\pi(x_i, x_j)\pi(x_i)\pi(x_j)}\right] \tag{10}$$

$$\hat{\mathbb{V}}_{HTC}\left(\hat{Y}\right) = \sum_{x_i \in s} \left[\frac{Y(x_i)}{\pi(x_i)}\right]^2 + \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_i \neq x_j}} \frac{Y(x_i) Y(x_j)}{\pi(x_i)\pi(x_j)} - \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_i \neq x_j}} \frac{Y(x_i) Y(x_j)}{\pi(x_i, x_j)} \tag{11}$$

Obecná teorie HTC, úhrn v D [Cordy, 1993] IV

Sen-Yates-Grundyho odhad rozptylu je použitelný **pouze pro výběry konstantní velikosti**. Jeho podoba pro konečné populace i populace bodů kontinua je shodná [Cochran, 1977, teorém 9A.5, str. 260], [Särndal et al., 2003, důkaz na str. 45].

SYG odhad rozptylu úhrnu v D

$$\hat{\mathbb{V}}_{SYG}(\hat{Y}) = \frac{1}{2} \sum_{x_i \in s} \sum_{\substack{x_j \in s \\ x_j \neq x_i}} \left[\frac{\pi(x_i)\pi(x_j) - \pi(x_i, x_j)}{\pi(x_i, x_j)} \right] \left[\frac{Y(x_i)}{\pi(x_i)} - \frac{Y(x_j)}{\pi(x_j)} \right]^2 \qquad (12)$$

SYG odhad rozptylu úhrnu v D, výpočtový tvar

$$\hat{\mathbb{V}}_{SYG}(\hat{Y}) = \sum_{i}^{n-1} \sum_{j>i}^{n} \left[\frac{\pi(x_i)\pi(x_j) - \pi(x_i, x_j)}{\pi(x_i, x_j)} \right] \left[\frac{Y(x_i)}{\pi(x_i)} - \frac{Y(x_j)}{\pi(x_j)} \right]^2 \qquad (13)$$

Vzorec (12) **můžeme zapsat ve výpočetně efektivnějším tvaru** (13) zavedením libovolného pořadí $i = (1 \dots n)$ inventarizačních bodů $x_i \in s$, kde n je konstantní velikost výběru s (počet inventarizačních bodů v D).

Obecná teorie HTC, úhrn v D [Cordy, 1993] V

Porovnání HTC a SYG odhadů rozptylu úhrnu v D

- **odhady rozptylu HTC a SYG se pro konkrétní výběr zpravidla liší**
- **oba jsou však nestranné** při splnění podmínky $\pi(x_i, x_j) > 0 \ \forall x_i, x_j \in D$
- HTC i SYG odhady **mohou za určitých okolností nabývat záporných hodnot** [Cochran, 1977, kapitola strana 261], což nelze z teoretického hlediska ani prakticky (konstrukce intervalových odhadů) akceptovat
- **SYG odhad je vždy kladný při splnění** (14) [Särndal et al., 2003, poznámka 2.8.7 na straně 47 a 48]

$$\pi(x_i)\pi(x_j) - \pi(x_i, x_j) \ >= \ 0 \ \forall \ x_i \ \neq \ x_j \ \in \ D \tag{14}$$

- **SYG odhad rozptylu vykazuje větší stabilitu** - má nižší rozptyl [Cochran, 1977, kapitola 9A.7, strana 261]

Střední hektarová hodnota v D I

Kromě úhrnu Y potřebujeme často získat odhad **střední (hektarové) hodnoty** $\bar{Y}$ v D definované (15). Typickým příkladem je **odhad lesnatosti** - relativního podílu kategorie pozemků les na rozloze zájmového území D.

Definice střední hodnoty v D

$$\bar{Y} = \frac{Y}{\lambda(D)} \tag{15}$$

Poněvadž je v rámci (15) **provedeno pouze dělení konstantou** $\lambda(D)$**, jejíž hodnota je zcela přesně známa**, je odhad $\hat{\bar{Y}}$ v D dán analogicky.

Odhad střední hodnoty v D

$$\hat{\bar{Y}} = \frac{\hat{Y}}{\lambda(D)} \tag{16}$$

Symbol $\hat{Y}$, jako obvykle, značí odhad úhrnu v D a $\lambda(D)$ je zcela přesně známá rozloha zájmové oblasti D (v hektarech).

Střední hektarová hodnota v D II

Rozptyl střední hodnoty $\hat{\hat{Y}}$ v D je dán (17) viz Anděl [1978, věta 2 na straně 14], kde $\hat{Y}$ značí příslušný odhad úhrnu.

Rozptyl střední hodnoty v D

$$\mathbb{V}(\hat{\hat{Y}}) = \frac{\mathbb{V}(\hat{Y})}{\lambda^2(D)} \tag{17}$$

Odhad rozptylu střední hodnoty $\hat{\hat{Y}}$ v D je s pomocí odhadu rozptylu úhrnu $\hat{\mathbb{V}}(\hat{Y})$ vyjádřen analogicky

Odhad rozptylu střední hodnoty v D

$$\hat{\mathbb{V}}(\hat{\hat{Y}}) = \frac{\hat{\mathbb{V}}(\hat{Y})}{\lambda^2(D)} \tag{18}$$

Mezi úhrnem a střední hodnotou a mezi rozptyly jejich odhadů existuje **lineární vztah**, což v plném rozsahu platí i pro odhady úhrnů a středních hodnot a též pro odhady rozptylů. **Vlastnosti odhadu (nestrannost, konzervativnost apod.) střední hodnoty a odhadu jejího rozptylu se proto shodují s vlastnostmi úhrnu**.

Aplikace HTC pro URS

Pro (zcela) **náhodný výběr (URS z angl. Uniform Random Sampling) pevného počtu** n **inventarizačních bodů** v D přechází obecné vzorce (7) (odhad úhrnu) a (10), (11) (HTC odhady rozptylu) respektive (12), (13) (SYG odhady rozptylu) do podoby (19) a (20), (21).

Odhad úhrnu v D, design URS

$$\hat{Y} = \sum_{x \in s} \frac{Y(x)}{\pi(x)} = \lambda(D) \frac{\sum_{x \in s} Y(x)}{n} = \lambda(D) \hat{\bar{Y}} \tag{19}$$

Odhad rozptylu úhrnu v D, design URS

$$\hat{\mathbb{V}}_{URS}(\hat{Y}) = \frac{\lambda^2(D)}{n(n-1)} \sum_{x \in s} \left[Y(x) - \hat{\bar{Y}} \right]^2 \tag{20}$$

$$\hat{\mathbb{V}}_{URS}(\hat{Y}) = \frac{\lambda^2(D)}{n-1} \left[\widehat{\overline{Y^2}} - \hat{\bar{Y}}^2 \right] \tag{21}$$

Člen $\widehat{\overline{Y^2}}$ v rámci (21) představuje **aritmetický průměr čtverců lokální hustoty** $Y(x)$ na n inventarizačních bodech - **odhad střední hodnoty** $Y^2(x)$ **v** D. Symbol $\hat{\bar{Y}}^2$ představuje **čtverec průměru lokální hustoty** $Y(x)$ **na** n **inventarizačních bodech v** D **- odhad střední hodnoty lokální hustoty** $Y(x)$ **v** D, viz. (16) a (19).

Odhad podílu v D, obecná teorie I

Mnoho cílových parametrů IL a IK může být vyjádřeno jako podíl $R_{1,2}$ **úhrnů** Y_1 a Y_2 nebo ekvivalentně jako **podíl středních hodnot** $\bar{Y}_1$ a $\bar{Y}_1$ v zájmové oblasti D.

Definice podílu v D

$$R_{1,2} = \frac{Y_1}{Y_2} = \frac{\bar{Y}_1}{\bar{Y}_2} \tag{22}$$

Odhad $\hat{R}_{1,2}$ **podílu** $R_{1,2}$ **je počítán jako podíl odhadů úhrnu** $\hat{Y}_1$ a $\hat{Y}_2$ nebo shodně jako **podíl odhadů středních hodnot** $\hat{\bar{Y}}_1$ a $\hat{\bar{Y}}_1$.

Odhad podílu v D (asymptoticky nestranný)

$$\hat{R}_{1,2} = \frac{\hat{Y}_1}{\hat{Y}_2} = \frac{\hat{\bar{Y}}_1}{\hat{\bar{Y}}_2} \tag{23}$$

Typickým příkladem je **hektarová střední zásoba hroubí** tj. podíl celkové zásoby hroubí a rozlohy *porostní půdy* (jeden ze tří druhů pozemku v rámci kategorie pozemků les).

Odhad podílu v D, obecná teorie II

Odhad rozptylu $\hat{\mathbb{V}}(\hat{R}_{1,2})$ **podílu** $\hat{R}_{1,2}$ můžeme obecně vyjádřit pomocí vzorce (24) odvozeného na základě **Taylorovy aproximace**.

Aproximativní odhad rozptylu podílu v D

$$\hat{\mathbb{V}}(\hat{R}_{1,2}) \approx \frac{1}{\hat{Y}_2^2}\hat{\mathbb{V}}(\hat{Z}_\circ). \tag{24}$$

Vzorec (24) je odhadem rozptylu úhrnu $\hat{Z}_\circ$ **reziduální proměnné** $Z_\circ(x)$ v D.

Odhad úhrnu reziduální proměnné v D

$$\hat{Z}_\circ = \sum_{x \in s} \frac{Z_\circ(x)}{\pi(x)} = \lambda(D)\hat{\hat{Z}}_\circ, \tag{25}$$

$$Z_\circ(x) = Y_1(x) - \hat{R}_{1,2}Y_2(x). \tag{26}$$

Detaily odvození odhadu rozptylu podílu s využitím Taylorovy aproximace (delta metody) lze dohledat v monografii Särndal et al. [2003, sekce 5.5 a 5.6 od str. 172].

Odhad podílu v D, obecná teorie III

Ekvivalentně k (24) **lze aproximativní odhad rozptylu** $\hat{\mathbb{V}}\big(\hat{R}_{1,\,2}\big)$ **podílu** $\hat{R}_{1,\,2}$ **vyjádřit vzorcem** (27).

Alternativní vyjádření odhadu rozptylu podílu v D

$$\hat{\mathbb{V}}\big(\hat{R}_{1,\,2}\big) \approx \frac{1}{\hat{Y}_2^2}\Big[\hat{\mathbb{V}}\big(\hat{Y}_1\big) + \hat{R}_{1,\,2}^2\hat{\mathbb{V}}\big(\hat{Y}_2\big) - 2\hat{R}_{1,\,2}^2\hat{\mathbb{C}}\big(\hat{Y}_1,\,\hat{Y}_2\big)\Big] \tag{27}$$

Členy $\hat{\mathbb{V}}\big(\hat{Y}_1\big)$, $\hat{\mathbb{V}}\big(\hat{Y}_2\big)$ jsou odhady rozptylů jednofázových úhrnů v čitateli ($\hat{Y}_1$) a jmenovateli ($\hat{Y}_2$) podílu $\hat{R}_{1,\,2}$, $\hat{\mathbb{C}}\big(\hat{Y}_1,\,\hat{Y}_2\big)$ je **kovariance jednofázových odhadů** [Särndal et al., 2003, podrobnosti odvození v sekci 5.6, str. 176].

Dostatečně silná, pozitivní kovariance $\hat{\mathbb{C}}\big(\hat{Y}_1,\,\hat{Y}_2\big)$ je nutným předpokladem pro dosažení **vyšší přesnosti odhadu podílu** v porovnání s jednofázovými odhady v čitateli a jmenovateli $\hat{R}_{1,\,2}$.

Odhad podílu v D, aplikace pro design URS

Pro **design URS (zcela náhodný výběr pevného počtu n inventarizačních bodů v D)** odpovídá **bodový odhad podílu** v D obecnému vzorci (23).

Odhad rozptylu podílu $\hat{R}_{1,2}$ viz vzorce (28) a (29) vychází z obecného vzorce (24) **upraveného pro design URS**.

Aproximativní odhady rozptylu podílu v D, design URS

$$\hat{\mathbb{V}}_{URS}\left(\hat{R}_{1,2}\right) \approx \frac{\lambda^2(D)}{n(n-1)\,\hat{Y}_2^2} \sum_{x \in s} Z_\circ^2(x) \tag{28}$$

$$\hat{\mathbb{V}}_{URS}\left(\hat{R}_{1,2}\right) \approx \frac{\sum_{x \in s} \left[Y_1(x) - \hat{R}_{1,2}\, Y_2(x)\right]^2}{n(n-1)\,\hat{\bar{Y}}_2^2} \tag{29}$$

Odhad rozptylu $\hat{\mathbb{V}}_{URS}\left(\hat{R}_{1,2}\right)$ lze získat též **dosazením URS specifických odhadů rozptylů úhrnů a kovariance do obecného, aproximativního vzorce** (27).

Odhad rozptylu, (quasi-)systematické výběry I

Neexistence design-based odhadu rozptylu

- většina inventarizací používá nějakou variantu **systematického** nebo **prostorově stratifikovaného výběrového designu**
- **podmínka** $\pi(x_i, x_j) > 0 \ \forall x_i, x_j \in D$ **není splněna** pro **systematický** (CSS - Centric Systematic Sampling, jiným označením ASS - Aligned Systematic Sampling) ani pro **prostorově stratifikované designy s pevným počátkem** (např. TSS - Tesselated Stratified Sampling)

Existuje řada způsobů **jak aproximovat a odhadnout design-based rozptyl výběrových schémat nevyhovujících uvedené podmínce** viz např. Wolter M. [1985], Cordy and Thompson [1995], Heikkinen [2006, sekce 10, od str. 155] and Cooper [2006].

Odhad rozptylu, (quasi-)systematické výběry II

Odhad rozptylu pomocí URS aproximace

- použití **aproximace** $\pi(x_i, x_j) = n(n-1)/\lambda^2(D)$ **odpovídající URS s pevnou velikostí výběru** n **v** D
- **odhad rozptylu s touto aproximací je většinou nikoli však zákonitě konzervativní** [Heikkinen, 2006, sekce 10, od str. 155], [Mandallaz, 2007, sekce 4.1, od str. 53]
- **podmínky konzervativnosti** odhadu rozptylu
  1. **dostatečná velikost výběru**
  2. **pozitivní prostorová korelace** lokální hustoty
  3. **absence periodicity** lokální hustoty
- **odhad rozptylu je aproximován** vzorci (20), (21), (18) (úhrny a střední hodnoty) a (28), (29) (podíly)

Odhad rozptylu, (quasi-)systematické výběry III

Odhad rozptylu na principu kontrastů

Méně konzervativní odhad rozptylu lze pro (quasi-)systematické a prostorově stratifikované inventarizační sítě získat dle vzorce

$$\hat{\mathbb{V}}_{ST}\left(\hat{\theta}\right) = \frac{\lambda^2(D)}{2kn} \sum_{\substack{x \in s \\ x \in D}} \sum_{\substack{\dot{x} \in s \\ \dot{x} \in D}} \left[Y(x) - Y(\dot{x})\right]^2. \qquad (30)$$

Jednotlivé členy sumy tvoří **diference lokální hustoty na bodě** x **a lokální hustoty sousedního bodu** $\dot{x}$. Uvažují se sousedé směrem na sever, východ, jih a západ. Člen k ve jmenovateli zlomku je celkový počet diferencí - dvojic sousedů, které bylo možno v D utvořit.

Vlastnosti odhadu (30) pro různé výběrové designy **testovali Cordy and Thompson [1995] s velmi dobrými výsledky**.

Nekonstantní velikost výběru v D I

Důvody nekonstantní velikosti výběru v D

- většina IL a IK používá **čtvercové, méně často obdélníkové, trojúhelníkové nebo šestiúhelníkové inventarizační sítě s náhodným počátkem**
- inventarizační sítě dělí **suport**[a] $\mathcal{S}$ na navazující a nepřekrývající se buňky shodného tvaru a velikosti - **inventarizační bloky**
- **oblast** D **nemůže být bezezbytku rozdělena na podoblasti odpovídající celým inventarizačním blokům**
- **počet inventarizačních bodů umístěných do** D **se může výběr od výběru lišit** z důvodu **náhodného umístění počátku inventarizační sítě**, v případě některých designů též a nebo výhradně z důvodu **náhodného generování pozic bodů uvnitř inventarizačních bloků**

[a] Oblast používaná pro implementaci algoritmu generování pozic inventarizačních bodů - typicky obdélník obsahující D.

Nekonstantní velikost výběru v D II

Nestrannost odhadů při nekonstantní velikosti výběru v D

Cordy [1993] popisuje dvě varianty:

1. **nahrazení funkce hustoty výběru inventarizačních bodů a párových hustot jejich očekávanými hodnotami** přes všechny možné velikosti výběru
   - odhady jsou nestranné **nepodmíněně na velikosti výběru** n_D (unconditional inference)
   - odhady úhrnů **jsou aditivní**
2. **vyjádření hustot výběru inventarizačních bodů vzhledem k právě realizované velikosti výběru** n_D **v** D
   - odhady jsou nestranné **podmíněně na velikosti výběru** n_D (conditional inference)
   - odhady úhrnů **nejsou aditivní**.

Je také možné **provést odhady na úrovni suportu** $\mathcal{S}$ namísto D. Specifická topologie (ring-topology) [Stevens, 1997, sekce 3.1, od str. 172], [Mandallaz, 2007, sekce 5.6, od str. 92] zajišťuje konstantní velikost výběru v $\mathcal{S}$. Přitom je třeba **upravit definici lokální hustoty tak, aby mimo** D **nabývala nulových hodnot**. Řešení není vhodné pro výběrové designy používající URS aproximaci odhadu rozptylu - **další navýšení konzervativnosti odhadu rozptylu**.

Literatura I

J. Anděl. *Matematická statistika*. SNTL - Státní nakladatelství technické literatury, Praha, 1978.

W. G. Cochran. *Sampling Techniques*. Willey series in probability and mathematical sttaistics - applied. John Willey & Sons, 1977.

C. Cooper. Sampling and variance estimation on continuous domains. *Environmetrics*, 17:539–553, 2006.

B. Cordy, C. and M. Thompson, C. An application of the deterministic variogram to design-based variance estimation. *Mathematical Geology*, 27:173–205, 1995.

C. B. Cordy. An extension of the horwitz-thompson theorem to point sampling from a continuous universe. *Statistics and Probability Letters*, 18:353–362, 1993.

J. (eds. Kangas A. & Maltamo M.) Heikkinen. *Forest Inventory, Methodology and Applications*, chapter Assessment of Uncetainty in Spatially Systematic Sampling, pages 155–176. Springer Verlag, 2006.

D. Mandallaz. *Sampling Techniques For Forest Inventories*. Chapman and Hall/CRC, 2007.

D. L. Jr. Stevens. Variable density grid-based sampling designs for continuous spatial populations. *Environmetrics*, 8:167–195, 1997.

Literatura II

C. E. Särndal, B. Swensson, and J. Wretman. *Model Assisted Survey Sampling*. Springer, 2003.

K. Wolter M. *Introduction to Variance Estimation*. Springer, 1985.

Poděkování

Vytvořeno s podporou projektu „*Průřezová inovace studijních programů Lesnické a dřevařské fakulty MENDELU v Brně (LDF) s ohledem na discipliny společného základu*" (reg. č. CZ.1.07/2.2.00/28.0021) za přispění finančních prostředků EU a státního rozpočtu České republiky.

Děkuji za Vaši pozornost!

- Co Vás na přednášce zaujalo?
- Čím by jste přednášku doplnili?
- Další dotazy a připomínky k tématu NIL?