

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Giniho index

Jaroslav Zlatohlávek

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Doc. RNDr. Karel Zvára, CSc.

Studijní program: Ekonometrie

Chtěl bych touto cestou poděkovat své rodině za podporu a panu docentu Karlu Zvárovi, vedoucímu diplomové práce, za užitečné rady.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.
V Praze dne 9. srpna 2006

Jaroslav Zlatohlávek

Obsah

1	Konstrukce indexů chudoby	4
2	Giniho index nerovnoměrnosti	6
2.1	Axiomatické vlastnosti Giniho indexu	13
2.2	Zobecněné Giniho indexy nerovnoměrnosti	15
2.3	Druhy používaných dat	15
2.4	Rozdělení příjmů	16
2.5	Způsoby odhadování Giniho indexu a intervaly odhadu při použití seskupených dat	17
2.6	Statistické vlastnosti odhadu Giniho indexu	21
2.7	Gini a regrese	24
3	Alternativní míry chudoby	26
4	Dekompozice Giniho indexu	27
4.1	Dekompozice podle podskupin	28
4.2	Rozklad indexu světové nerovnoměrnosti v hrubém domácím pro- duktu	35
4.3	Dekompozice podle zdrojů příjmů	37
4.4	Modelování příjmů lineárním modelem a souvislost pro rozklad Gi- niho indexu	40
4.5	Vícerozměrná dekompozice	41
5	Dodatky	42

Název práce: Giniho index
Autor: Jaroslav Zlatohlávek
Katedra (ústav): Katedra pravděpodobnosti a matematické statistiky
Vedoucí diplomové práce: Doc. RNDr. Karel Zvára, CSc.
e-mail vedoucího: karel.zvara@mff.cuni.cz

Abstrakt: Pro zachycení míry nerovnosti v úrovni příjmů populace se osvědčil Giniho index, který je z podobných měř jednoznačně nejpoužívanější. Lze ho definovat jako charakteristiku náhodné veličiny nebo jako popisnou statistiku v původním historickém smyslu vyjadřující průměrnou absolutní vzdálenost dvou pozorování dělené dvojnásobkem výběrového průměru. Uvedeny jsou jeho různá vyjádření a souvislosti mezi nimi. Vedle Giniho indexu jsou představeny i alternativní indexy chudoby. Následně jsou probrány přístupy pro odhad Giniho indexu v případech, kdy nejsou k dispozici úplná data. V kapitole 4 se věnují možnostem rozkladu hodnoty Giniho indexu pomocí nerovnosti uvnitř podskupin populace. Druhým možným přístupem k dekompozici nerovnoměrnosti rozklad celkového příjmu jedince na komponenty a následná analýza příspěvků jednotlivých zdrojů příjmů na celkové nerovnoměrnosti. Rozkladu Giniho indexu pomocí nerovností uvnitř jednotlivých druhů příjmů se zabýváme v téže kapitole. Práce obsahuje též rozklad nerovnosti v úrovni hrubého domácího produktu na obyvatele podle příslušnosti státu ke kontinentu. Zjistíme, že nejvýraznější efekt na nerovnost HDP na obyvatele mají vnitřní nerovnosti uvnitř jednotlivých kontinentů.

Klíčová slova: Giniho index, Lorenzova křivka, rozdělení příjmů, dekompozice Giniho indexu

Title: Gini index
Author: Jaroslav Zlatohlávek
Department: Department of Probability and Mathematical Statistics
Supervisor: Doc. RNDr. Karel Zvára, CSc.
Supervisor's e-mail address: karel.zvara@mff.cuni.cz

Abstract: Gini index has proved a viable concept in income inequality measurement and therefore has become pre-dominantly used for this purpose a long time ago. There are two different perspectives of it: either as a random variable characteristic or population statistics, which in its original historic sense equals mean difference of two incomes divided by twice the mean. In the first part the Gini index is presented equivalently via different expressions and its axiomatic properties are listed. Besides the Gini index, there are alternative ways of measuring poverty mentioned. The following section presents different methods for estimation of the index based on only a few data points known (typically income deciles) and the possible bounds on computed index. Section 4 introduces decomposition of the Gini index in different ways – several types of subgroup decomposition and source decomposition is presented as well. At some stages presented concepts are used on real data sets. I.e. we find out that the main reason of world inequality in Gross Domestic Product are inequalities within the continents.

Keywords: Gini index, Lorenz curve, income distribution, Gini index decomposition

Úvod

V této práci se budu zabývat tzv. Giniho indexem nerovnoměrnosti¹. Jedná se o míru nerovnoměrnosti rozdělení neboli nakolik se dané rozdělení odchyluje od rovnoměrného rozprostření určitého zdroje mezi členy společnosti. Nejčastěji bývá použit pro vyjádření nerovnoměrnosti velikosti příjmů obyvatelstva. Ve formě indexu udává nakolik se dané rozdělení liší od případu, kdy by všichni jedinci měli stejný příjem. Samozřejmě lze ho použít i v jiných případech – rozdílnost ve využívání zdrojů, přístupnost k pitné vodě, úmrtnost novorozenců; ať uvedeme pár případů možného nasazení. Obecněji jej lze samozřejmě použít kdykoliv jsme schopni popsat proces pro porovnávání užiteků.

Pokud se uvádí nějaká charakteristika příjmů obyvatelstva, bývá použit průměr. Je čím dál tím více zřetelnější, že samotný údaj o průměrnému příjmu (popř. platu) může být dosti zavádějící. Připomeňme například, že za aktuálního stavu dvě třetiny populace nedosáhnou na průměrný příjem). Pověšinou rozdělení příjmů není symetrické (ať už je důvodem přirozená nezápornost příjmů nebo často těžké chvosty rozdělení). V takovém případě se již dlouho ukazuje jako výhodná míra nestejnorodosti rozdělení příjmů ve společnosti Giniho index, který má několik výhod. Za prvé je nejčastěji používaným indexem svého druhu, tudíž existuje velká základna dat pro komparaci ať už vzhledem k času nebo mezi oblastmi (státy apod.)². Navíc má mezi ostatními netriviálními indexy nerovnoměrnosti pro ekonomickou a sociologickou obec nejnázornější – geometrickou – interpretaci.

Nejprve si v kapitole 2 představíme samotný Giniho index a jeho základní definici. Budou ukázány i jiné způsoby jak jej nahlížet z jiných perspektiv.

Následně si zsumarizujeme požadavky, které se kladou na „dobrý“ index chudoby, jejich opodstatněnost a vlastnosti Giniho koeficientu. Představíme si i jiné používané míry nerovnoměrnosti.

V kapitole 4 se budeme zabývat strukturou Giniho indexu, resp. možnostmi analýzy nerovnosti ve společnosti z perspektivy rozdělení společnosti podle zvolených charakteristik a vliv těchto charakteristik na příjmovou nerovnoměrnost celé společnosti. Bude proveden i rozklad nerovnosti v hrubém domácího produktu na obyvatele pro 179 zemí.

1 Konstrukce indexů chudoby

Snaha o vyjádření souhrnného stavu společnosti z hlediska její chudoby, resp. diferenciací příjmů jejích členů je obsáhlým tématem. Už způsob podchycení rozdílnosti bohatství ve společnosti je zásadní otázkou – jak vůbec určovat úroveň bohatství. Je zřejmé, že prosté srovnání výše bohatství je sice vysoce korelované s životní úrovní; nemusí být však pro účely analýzy postačující. Dostáváme se do problémů spojených se strukturou (jiné množství může uspokojit potřeby jedinců ve městě nebo na venkově, rozdíl bude i při rozlišování podle věku).

Obecně můžeme zapojit užitekovou funkci. Zjišťování jejího tvaru je pro každého jedince značně složitým, ne-li neřešitelným problémem. Proto je přirozeným požadavkem minimalizace požadavků na její vlastnosti. Extrémním případem –

¹ v anglicky psané literatuře se vyskytuje pod názvy Gini index of inequality, Gini concentration ratio apod.

² ucelnou databázi udržuje například World Bank

poskytující ovšem užitečné informace – je užitková funkce definovaná jako příjem jedince. Pro účely této práce je v podstatě nedůležitý způsob určování blahobytu jednotlivců a tak se budu zmiňovat o příjmech, což je i nejčastější oblast použití Giniho indexu.

Velká pozornost byla věnována požadovaným vlastnostem měr nerovnoměrnosti. Většinou však bývala jejich formulace a volba těch potřebných pevně svázána s novým druhem indexu. Přehled používaných axiomů a jejich souvislosti můžeme najít například v [16]. Zde také Zheng identifikoval množinu základních axiomů, které společně implikují i další vhodné požadavky. Představme si vhodné vlastnosti indexů chudoby resp. nerovnosti.

Nadále budeme označovat *vektor příjmů* jedinců v populaci $x \in D = \bigcup_{n \in \mathbb{N}} \mathbb{R}_{0+}^n$ a dále $z \in \mathbb{R}_{0+}^n$ hodnotu hranici chudoby. Index chudoby je funkcí definovanou na kartézském součinu $D \times \mathbb{R}_{0+}^n - P : D \times \mathbb{R} \rightarrow \mathbb{R}$

Regresivním transferem se rozumí realokace určité částky (části bohatství) od relativně chudšího k relativně bohatšímu jedinci takovým způsobem, že pořadí jedinců podle výše příjmu zůstane stejné jako v původní populaci. Průměrný příjem zůstává vzhledem ke konstrukci stejný jako v původní populaci.

Progressivním transferem se rozumí realokace části bohatství od relativně bohatšího k relativně chudšímu jedinci, taktéž při zachování pořadí.

Cílená citlivost (Focus axiom) Zvýšení příjmu jedince nacházejícího se nad hranicí chudoby nesmí ovlivnit hodnotu míry chudoby. Tento požadavek na index chudoby ho automaticky specifikuje jako míru, která odráží jen hloubku chudoby nižší části spektra a nikdy jako míru nerovnoměrnosti v celé populaci.

Symetrie Míra chudoby je funkce invariantní vůči permutacím. Požadujeme tedy $P(x, z) = P(y, z)$, kdykoliv vektor y vznikne permutací složek vektoru x .

Invariance vůči replikaci Jestliže vznikne vektor y několikerým spojením vektoru x ; tedy $y = (x, x, \dots, x)$, pak se požaduje $P(x, z) = P(y, z)$. Tento požadavek je adekvátní z pohledu, že ač společnost s daným rozdělením příjmů popíšeme větším počtem pozorování, nemá velikost vzorku vliv na hodnotu indexu.

Spojitost $P(x, z)$ je při pevném z spojitou funkcí složek vektoru x .

Axiom regresivního transferu (Regressive transfer axiom) je splněn, pokud $P(x, z) > P(y, z)$ kdykoliv x vznikne z y regresivním transferem mezi jedinci pod hranicí chudoby.

Slabý axiom transferu (Weak transfer axiom) Varianta předchozího, která neklade požadavek na zařazení akceptora. Požadujeme tedy zvýšení hodnoty indexu kdykoliv dojde k realokaci částky k blahobytnějšímu jedinci.

Klesající efekt transferu (Diminishing Transfer Axiom) Změna v nerovnosti při progresivním transferu má být při pevné volbě obdarovaného jedince výraznější pro chudšího z dvojice potenciálních donátorů (donátorem může být pouze jedinec s vyšším příjmem než obdarovaný)

Princip Pigou-Daltonův Pokud rozdělení y vznikne konečnou sérií progresivních transferů z rozdělení x – a tedy zůstává zachováno pořadí jedinců vzhledem k příjmu – pak má platit $P(y, z) < P(x, z)$.

Monotonicita Logickým požadavkem je, aby se index chudoby zvýšil při každém snížení příjmu jedince pod hranicí chudoby.

Konzistence vzhledem k podskupinám (Sub-section nebo také Sub-group consistency) Tento požadavek je formulován jako potřeba, aby při změně koeficientu nerovnosti v nějaké skupině (při zachovaných podílech skupin na celkových příjmech a nezměněné velikosti jednotlivých skupin) se zvýšení (snížení) stejným způsobem promítlo do celkového indexu. Tento požadavek byl však časem rozpoznán jako neopodstatněný, protože vede k nežádoucím konsekvencím pro rozdělení příjmů ve skupinách. Dlužno dodat, že Giniho index tento axiom nesplňuje.

Normalizace Index nabývá pouze hodnot z intervalu $[0, 1]$.

Nezávislost na měřítku (Ratio Scale Invariance) neboli index je homogenní funkce řádu 0 (neboli $P(\lambda x) = P(x)$ pro každé $\lambda > 0$). Základ požadavku je tvořen potřebou zachytit nerovnoměrnost, ať už jsou podkladová data vyjádřena v jakýchkoliv jednotkách.

Netýká se ale posunu; tedy obecně index chudoby splňující tento axiom může být (a zpravidla bude) citlivý na fakt, že všem jedincům přidáme vždy stejnou sumu.

2 Giniho index nerovnoměrnosti

V této kapitole si představíme několik možných vyjádření Giniho indexu a prozkoumáme jej i z hlediska právě představených axiomů. Představme si nejprve základní koncept, který byl představen roku 1912³ v Giniho knize jako průměrná vzdálenost dvou pozorování. Tato kniha však vyšla v italštině a teprve v roce 1921 jej představil v anglicky psaném komentáři k článku Daltona. Tím byl tento pojem představen široké komunitě. Tuto průměrnou vzdálenost dnes označujeme jako *absolutní* Giniho index. Objektem našeho zájmu bude převážně relativní varianta, která je definována jako střední vzdálenost dvou náhodných veličin normovaná dvojnásobkem střední hodnoty náhodné veličiny. Označení Giniho index bude vždy náležet této relativní variantě.

³Nejranější zmínky o relativní vzdálenosti pocházejí zřejmě z Německa 70. let 19. století. Gini poprvé navrhnul průměrnou vzdálenost pozorování ve své italské monografii roku 1912. Jako datum zrození concentration ratio lze brát 29.3.1914, kdy Corrado Gini prezentoval článek *Sulla misura della concentrazione e della variabilità dei caratteri*.

Zajímavé je, že až poté byla v publikaci definována Lorenzova křivka pro spojitě rozdělení.

V následujícím budeme používat označení X_i pro náhodnou veličinu řídící příjem jedince, pozorovanou hodnotu realizace x_i a velikost sledované populace N . Příjmy se budou, jako je obvyklé, považovat za nezávislé a stejně rozdělené (i.i.d.) nezáporné náhodné veličiny s absolutně spojitou distribuční funkcí F . Náhodná veličina může být definována na celé kladné poloose nebo na nějakém intervalu. Mohli bychom připustit i volnější podmínky na n.v., ale tím bychom ztratili v budoucnu některé možnosti práce s GI. Dále požadujeme konečnou střední hodnotu.

Za základní definici považujeme původní Giniho návrh:

$$G = \frac{1}{2\mu_X} E |X_i - X_j| \quad (1)$$

jež vzešel z pojmu střední vzdálenosti (označované někdy jako *absolutní Giniho koeficient*) $\Delta = E |X_i - X_j|$. Zřejmě lze Giniho index jako charakteristiku náhodné veličiny zapisovat pomocí distribuční funkce n.v. ve tvaru

$$G = \frac{1}{2\mu_X} \int_{D_F} \int_{D_F} |x - y| dF(x) dF(y)$$

Seznámili jsme se s původním vyjádřením GI (občas budeme tímto způsobem v textu zkracovat pojem Giniho index) . K přednostem Giniho indexu patří i mnoho alternativních způsobů zápisu. Odlišné pohledy nám mohou přinést lepší pochopení podstaty Giniho koeficientu. Pro začátek si snadno všimneme, že jej lze zapsat jako

$$G = \frac{1}{2\mu_X} [E \max(X_i, X_j) - E \min(X_i, X_j)]$$

Pro takové tvrzení stačí nahlédnout, že díky linearitě střední hodnoty a vyjádření absolutní hodnoty pro nezáporná čísla platí rovnost $E |X_i - X_j| = E \max(X_i, X_j) - E \min(X_i, X_j)$.

Uvedené vyjádření je jiným zápisem GI jakožto relativního očekávaného rozsahu dvou náhodných veličin. Odsud drobnou úpravou získáváme ekvivalentní zápis; stačí použít zřejmý vztah $[E \max(X_i, X_j) + E \min(X_i, X_j)] = 2\mu_X$:

$$G = 1 - \mu^{-1} E \min(X_i, X_j) \quad (2)$$

nebo ekvivalentně pro náhodné veličiny definované na $[0, b]$ (pro odvození odkazují na kapitolu 5)

$$G = 1 - \frac{1}{\mu_X} \int_0^b (1 - F(y))^2 dy \quad (3)$$

Pozorování Mějme dvě spojitá rozdělení náhodných veličin se shodnou střední hodnotou a necht' nabývají hodnot od 0. Necht' rozdělení A stochasticky dominuje v prvním stupni řádu⁴ rozdělení B. Potom Giniho index příslušný rozdělení A je menší než GI příslušný rozdělení B.

Pozorování je zřejmé z definice stochastické dominance a rovnice (3).

Opak právě ukázané vlastnosti Giniho indexu bývá často brán jako potřebný pro funkce porovnávající blahobyt populací.

⁴stochastická dominance prvního stupně A nad B: $F_A(y) \leq F_B(y) \forall y$ a minimálně v jednom bodě platí ostrá nerovnost

Prozatím jsme používali vyjádření, která předpokládala známé rozdělení spojitě náhodné veličiny. Pojem Giniho indexu však vznikl pro vyjádření nerovnosti množiny hodnot; oblastí jeho prvotního a dodnes nejčastějšího nasazení je zjišťování nerovnosti mezi příjmy jednotlivců. Pro náhodný výběr lze očekávat, že tvar Giniho indexu jako statistiky bude vypadat např. jako původní Giniho návrh

$$G = \frac{1}{2N^2\bar{x}} \sum_{i=1}^N \sum_{j=1}^N |x_i - x_j| \quad (4)$$

$$= \frac{1}{N^2\bar{x}} \sum_{i=1}^N (2i - N - 1) x_{(i)} \quad (5)$$

$$= 1 - \frac{1}{N^2\bar{x}} \sum_{i=1}^N (2N - 2i + 1) x_{(i)} \quad (6)$$

který zřejmě není nestranným odhadem Giniho indexu řídicího rozdělení, avšak nestranný odhad⁵ se téměř nepoužívá. Pro důkaz druhého řádku budeme upravovat první řádek následujícím způsobem:

$$\begin{aligned} \frac{1}{2N^2\bar{x}} \sum_{i=1}^N \sum_{j=1}^N |x_i - x_j| &= \frac{1}{2N^2\bar{x}} \sum_{i=1}^N \sum_{j=1}^N |x_{(i)} - x_{(j)}| \\ &= \frac{2}{2N^2\bar{x}} \sum_{i=1}^N \sum_{j=1}^i x_{(i)} - x_{(j)} \\ &= \frac{1}{N^2\bar{x}} \left(\sum_{i=1}^N \sum_{j=1}^i x_{(i)} - \frac{1}{N^2\bar{x}} \sum_{i=1}^N \sum_{j=1}^i x_{(j)} \right) \\ &= \frac{1}{N^2\bar{x}} \left(\sum_{i=1}^N i x_{(i)} - \sum_{i=1}^N (N - i + 1) x_{(i)} \right) \\ &= \frac{1}{N^2\bar{x}} \sum (2i - N - 1) x_{(i)} \quad (7) \end{aligned}$$

Následně lze řádek (5) rychle upravit na tvar uvedený v řádku (6) rozpisem $(2N - 2i + 1) = [N - (2i - N - 1)]$. Dále zřejmě platí $(N^2\bar{x})^{-1} \sum N x_{(i)} = 1$. Tím je přechod proveden.

Lorenzova křivka a její souvislost s Giniho indexem Giniho index významnou měrou vděčí za svou oblíbenost úzkou spojitostí s Lorenzovou křivkou. Tento z ekonomického a sociologického hlediska velmi důležitý pojem, který nám dává možnost názorné prezentace odklonu rozdělení od hypotetického případu rovnoměrného rozložení bohatství, si představme důkladněji. Lorenzova křivka (Max Otto Lorenz, 1907) je grafem funkce $p \rightarrow L(p)$, kde $L(p)$ vyjadřuje podíl $100 \times p\%$ nejchudších jedinců na celkovém bohatství společnosti. Historicky se první objevila diskrétní verze pro konečnou populaci; ta je definována bodově

$$L(i/N) = N^{-1}\bar{x}^{-1} \sum_{j=1}^i x_{(j)} \quad (8)$$

⁵nestranný odhad vznikne např. záměnou členu před sumou v rovnici (4) za $(N(N-1)\bar{x})^{-1}$

v ekvidistantně rozdělených bodech $0, 1/N, 2/N \dots 1$. Nutným důsledkem definice je, že pro Lorenzovu křivku vždy platí $L(0) = 0$ a $L(1) = 1$.

Při známém rozdělení náhodné veličiny může být Lorenzova křivka definována například pomocí hustoty rozdělení

$$L(p) = \mu_X^{-1} \int_0^{F^{-1}(p)} xf(x) dx$$

respektive ji lze přepsat (pomocí substituce $m = F(x)$, kterou můžeme provést vzhledem k předpokládaným vlastnostem distribuční funkce)

$$L(p) = \mu_X^{-1} \int_0^p F^{-1}(m) dm \quad (9)$$

Z podstaty se tedy jedná o konvexní křivku, která je stejně fixovaná v krajních bodech.

Lorenzova křivka se nachází mezi dvěma hypotetickými extrémy, které nemohou být generovány spojitou n.v. se spojitou distribuční funkcí: degenerované jednobodové rozdělení (nenulový bod) by implikovalo Lorenzovu křivku $L(p) = p$, zatímco pro posloupnost rozdělení konvergující k alternativnímu rozdělení s pravděpodobnostmi $p_1 \rightarrow 0$ by konvergovala k $L(p) = 0$ pro $p \in [0, 1)$ a $L(1) = 1$.

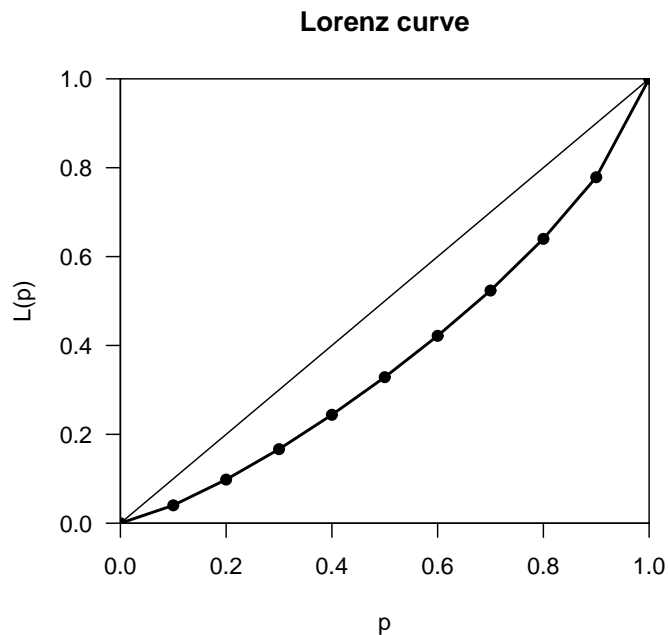
Z Lorenzovy křivky lze zjistit mnoho informací. Z definice vyplývá, že $L'(p)$ je relativní příjem jedince s pořadím Np ve srovnání s očekávanou hodnotou rozdělení. Bod v němž je tato derivace rovna jedné tedy udává podíl populace pod hranicí průměrného příjmu.

V grafu (1) máme zanesen odhad Lorenzovy křivky České republiky z roku 1996. Dostupná data byla ve formě podílů 10 %, 20 % ... atd. nejchudších jednotlivců na celkových příjmech společnosti. Tyto body empirické Lorenzovy křivky jsou v grafu zvýrazněny a pro lepší znázornění nerovnosti v příjmech byly tyto body linárně pospojovány.

Zabýváme se nyní souvislostí tohoto pojmu a Giniho indexu. Je obecně užívaným faktem, že hodnota GI odpovídá dvojnásobku plochy mezi Lorenzovou křivkou příslušnou řídicímu rozdělení a přímkou $x = y$. Zmíněný fakt můžeme zapsat ve tvaru

$$\begin{aligned} G &= 2 \int_0^1 [p - L(p)] dp \\ &= 1 - 2 \int_0^1 L(p) dp \end{aligned} \quad (10)$$

Empiricky získanou Lorenzovu křivku lze užít pro výpočet výběrového Giniho indexu pomocí součtu ploch mezi ní a přímkou úplné rovnoměrnosti vycházející (evidentně se odkazujeme na grafický významu GI). Zavedeme označení $F_i = i/N$



Obrázek 1: Spojité rozšíření výběrové Lorenzovy křivky spočtené z decilů

a dodefinujeme $F_0 = L_0 = 0$. Pak následující rovnost

$$\begin{aligned}
 G &= 1 - 2 \sum_{i=0}^{N-1} (F_{i+1} - F_i) \frac{(L_{i+1} + L_i)}{2} \\
 &= 1 - \sum_{i=0}^{N-1} (F_{i+1} - F_i) (L_{i+1} + L_i) \\
 &= 1 - \left[\sum_{i=0}^{N-1} F_{i+1} L_{i+1} - F_i L_i \right] - \left[\sum_{i=0}^{N-1} F_i L_{i+1} - F_{i+1} L_i \right] \\
 &= 1 - F_n L_n + F_0 L_0 - \left[\sum_{i=0}^{N-1} (F_i L_{i+1} - F_{i+1} L_i) \right] \\
 &= \sum_{i=0}^{N-1} (F_i L_{i+1} - F_{i+1} L_i)
 \end{aligned}$$

jasně vyjadřuje ideu odečíst od obsahu celé oblasti Lorenzovy křivky dvojnásobek plochy pod Lorenzovou křivkou (v prvním řádku). Poslední řádek poslouží k důkazu ekvivalence této definice s původní (viz (5)). Použijeme definice výběrové Lorenzovy křivky (8) a po označení

$$\Lambda \equiv \sum_{i=0}^{N-1} (F_i L_{i+1} - F_{i+1} L_i)$$

postupujeme následovně

$$\begin{aligned}
\Lambda &= \sum_{i=0}^{N-1} \left[\frac{i}{N} \times \frac{1}{N\bar{x}} \left(\sum_{j=1}^{i+1} x_{(j)} \right) - \frac{i+1}{N} \times \frac{1}{N\bar{x}} \left(\sum_{j=1}^i x_{(j)} \right) \right] \\
&= \frac{1}{N^2\bar{x}} \sum_{i=0}^{N-1} \left[i \left(\sum_{j=1}^i x_{(j)} \right) + ix_{(i+1)} - i \left(\sum_{j=1}^i x_{(j)} \right) - \left(\sum_{j=1}^i x_{(j)} \right) \right] \\
&= \frac{1}{N^2\bar{x}} \sum_{i=0}^{N-1} \left[ix_{(i+1)} - \left(\sum_{j=1}^i x_{(j)} \right) \right] \\
&= \frac{1}{N^2\bar{x}} \left[\left(\sum_{i=1}^N (i-1) x_{(i)} \right) - \left(\sum_{i=1}^N \sum_{j=0}^{i-1} x_{(j)} \right) \right] \\
&= \frac{1}{N^2\bar{x}} \sum_{i=1}^N [(i-1) - (N-i)] x_{(i)} \\
&= \frac{1}{N^2\bar{x}} \sum_{i=1}^N (2i - N - 1) x_{(i)}
\end{aligned}$$

kde předposlední přechod vyplývá z úpravy sumy „trojúhelníkové struktury“. Tímto je ekvivalence dokázána.

Kovarianční vyjádření Předchozí vyjádření indexu podle Lorenzovy křivky nám poskytlo cennou zkušenost, nicméně zde již byla mírně zastřena souvislost s řídicí náhodnou veličinou. Nyní si odvodíme vztah Giniho indexu na základě kovariance náhodné veličiny X popisující příjem a jejího relativního pořadí. Tento vztah využijeme např. pro dekompozici Giniho indexu v kapitole 4.

Předpokládejme náhodnou veličinu X nabývající hodnot z $[0, b]$. Vycházíme z vyjádření, které lze získat přímočarou úpravou z rovnice (3):

$$G = (\mu_X)^{-1} \int_0^b F(y) [1 - F(y)] dy \quad (11)$$

Pokračovat budeme jednak integrací *per partes* (při obvyklém značení funkcí určených pro přechod: za u celý vnitřek a $v' = 1$) a následně transformací $z = F(y)$ (tedy $dy = (f(y))^{-1} dz$). Dále si uvědomíme, že relativní pořadí n.v. X je také náhodná veličina $Z = F(X)$ s rovnoměrným rozdělením – a střední hodnotou 0,5. Vzhledem k vzájemně jednoznačnému vztahu náhodných veličin X a Z můžeme použitím rovnice (11)

$$\begin{aligned}
\mu_X G &= [yf(y) \{F(y)(1 - F(y))\}]_0^b + \int_0^b yf(y) [2F(y) - 1] dy \\
&= 0 + 2 \int_0^1 F^{-1}(z) [z - 0.5] dz \\
&= 2 \int_0^1 [F^{-1}(z) - EX + EX] [z - 0.5] dz \\
&= 2cov(X, F_X) + E(F^{-1}(z)) \int_0^1 [z - 0.5] dz \\
&= 2cov(X, F_X)
\end{aligned}$$

odvodit vyjádření Giniho indexu pomocí kovariance příjmu a jeho relativního zařazení:

$$G = \frac{2cov(X, F_X(X))}{\mu_X} \quad (12)$$

Vyjádření pomocí kovariance nám usnadní ukázat souvislost GI s regresním koeficientem speciálního modelu zmíněného v odstavci (2.7).

Vraťme se nyní k výběrovému vyjádření. Giniho index jsme schopni vyjádřit několika způsoby. Nyní si nastíníme význam Giniho indexu, tak jak jej publikovali Berrebi a Silber v [2]. Značíme m medián⁶ náhodného výběru a vektor x se výjimečně předpokládá sestupně seřazený. Z vyjádření platného pro sudé N (pro lichá N by došlo k drobné úpravě koeficientů)

$$G = \frac{1}{2\mu_X} \sum_{i=1}^N |x_i - m| \left| \frac{N - 2i + 1}{N^2/4} \right|$$

$$G = \frac{1}{4\mu_X} \left(\sum_{i=1}^{N/2} s_i - \sum_{i=N/2+1}^N d_i \right)$$

pro které je zavedeno $s_i = \frac{(y_i - m)}{n^2/4}((N - i) - (i - 1))$ a $d_i = \frac{(m - y_i)}{n^2/4}((i - 1) - (N - 1))$, vidíme, že Giniho index je váženým vyjádřením rozdílu spokojenosti blahobytnější poloviny populace a propadu hůře ohodnocené poloviny populace. Giniho index se ukazuje být speciálním vyjádřením rozdílu spokojenosti versus nespokojenosti jedinců s jejich postavením v příjmové stupnici (míra nespokojenosti jedince by zde byla součet rozdílů jeho příjmu od příjmů bohatších jedinců).

S jiným přístupem přišel Cheong v [3]. Ukázal, že Giniho index je vyjádřitelný pomocí střední hodnoty speciálního rozdělení. Lorenzovu křivku lze totiž vzhledem k jejím vlastnostem (neklesající, zleva spojitá, hodnoty z intervalu $[0, 1]$) nahlížet jako distribuční funkci náhodné veličiny (veličina nabývá hodnot z $[0, 1]$). Tato veličina P je vlastně relativní umístění (pořadí) jedince v populaci – 0 pro jedince s nejnižším příjmem, naopak 1 znamená nejvyšší příjem. Těžiště tohoto rozdělení určuje bod (relativní pořadí příslušné jednotky při vzestupném řazení podle výše příjmu) v němž dochází naplnění poloviny celkových příjmů společnosti. Pak z definice střední hodnoty n.v. $E(p) = \int_0^1 p dL(p)$ a dále si integrací per partes

$$E(p) = [pL(p)]_{p=0}^1 - \int_0^1 L(p) dp = 1 - \int_0^1 L(p) dp = \frac{1}{2} (1 + G)$$

odvodíme vztah Giniho indexu a střední hodnoty relativního pořadí jedince (poslední rovnost vychází jednoduchou úpravou z (10)). Tento zápis také názorně odhaluje způsob, jímž je GI citlivý na změnu v rozdělení příjmů.

Pro odhad střední hodnoty Giniho indexu využijeme odhadu střední hodnoty n.v. P

$$\bar{p} = \sum_{i=1}^n \frac{i}{n} \frac{X_{(i)}}{\sum X_i}$$

⁶pro sudý počet pozorování definujeme medián jako obvykle průměrem hodnot $x_{(N/2)}$ a $x_{((N+2)/2)}$

Zcela jiný přístup zvolil Pyatt v [13]. Je založen na teoretické hře a již ze zásady používá rozdělení do subpopulací. Vycházíme-li z rovnosti $\sum_i \sum_j |x_i - x_j| = 2 \sum_i \sum_j \max(0, x_i - x_j)$, lze zmíněnou rovnost chápat jako očekávaný zisk společnosti při následující hře: náhodně se vybere příjem z daného rozdělení a hráč má možnost ponechat si svůj příjem nebo přistoupit na onen náhodně vybraný jiný příjem. Předpokládá se, že hráč se chová racionálně a volí tedy větší ze dvou hodnot. Pak má Giniho index maticový zápis

$$G = (m'p)^{-1} p' Q p$$

Podíl jednotlivých skupin na populaci je zachycen ve vektoru p , kdežto průměrný příjem v jednotlivých skupinách je zachycen ve vektoru m . Původní návrh se objevuje v matici Q , která charakterizuje podmíněné očekávané zisky jednotlivců při příslušnosti k skupině i a možnosti přestupu do množiny j . Pak tedy v zápisu vidíme normalizaci průměrným příjmem ve společnosti $m'p$ a očekávaný zisk ze hry, kdy pravděpodobnost, že na tahu bude nabídnut hráči ze skupiny i , je rovna relativní velikosti skupiny na celé populaci a pravděpodobnost nabídky příjmu jedince z jiné skupiny j je nezávislý s obdobně definovanou pravděpodobností.

Struktura matice Q odráží, že čím víc je daná podskupina posazena ve svých příjmech výš oproti jiným, tím obsahuje příslušný řádek menší hodnoty; v extrémním případě obsahuje nuly mimo diagonálních prvků q_{ii} – pokud mají její příslušníci naprosto nejvyšší příjmy takové, že se nepřekrývají s jinými skupinami. Dodejme, že i prvek na diagonále q_{ii} může být nulový za předpokladu, že všichni jedinci dané skupiny mají stejný příjem.

Ze stejného základu vlastně vychází využití relativního propadu jednotky i vůči bohatším

$$\delta_i = N^{-1} \sum_{j=1}^N (X_i - X_j)_+ \quad (13)$$

Veličina tak zachycuje i postavení v příjmové škále a zohledňuje ho (dělení pomocí N ačkoliv tolik nenulových členů v sumě nebude). Po zavedení δ_i můžeme Giniho index jednoduše zapsat ve tvaru

$$G = \frac{\sum_{i=1}^N \delta_i}{N\mu_X} \quad (14)$$

2.1 Axiomatické vlastnosti Giniho indexu

Ukažme si, které požadavky na míru chudoby Giniho index splňuje.

- Axiom symetrie bude splněn, což lze nahlédnout pomocí jedné z ekvivalentních definic (4), která je založena čistě na pořadových statistikách.
- Replikační axiom můžeme dokázat následovně: označme počet replikací ve

vektoru k a Giniho index původního a „prodlouženého“ po řadě G_1, G_k .

$$\begin{aligned} G_k &= (\mu k^2 N^2)^{-1} \sum_{i=1}^{kN} \sum_{j=1}^{kN} |X_i - X_j| \\ &= (\mu k^2 N^2)^{-1} k^2 \sum_{i=1}^N \sum_{j=1}^N |X_i - X_j| \\ &= G_1 \end{aligned}$$

- Axiom cílené citlivosti nebude splněn, protože Giniho index zřejmě zohledňuje i změny v příjmech jedinců nad hranicí chudoby. Výjimka nastane pouze nastavením hranice chudoby dosti vysoko, což je ovšem proti duchu takové hranice.
- Spojitost. Vzhledem ke složkám příjmového vektoru vyplývá ze spojitosti sumy.
- Pigou-Daltonův axiom. Uvažujme jednotlivý progresivního transfer – došlo tedy k realokaci částky δ od jedince s pořadím r k chudšímu jedinci s pořadím $p < r$. Platí nejen $X_{(r)} - \delta > X_{(p)} + \delta$, ale i ostatní uspořádání zůstává nezměněno. Jak si můžeme všimnout z rovnice (5), dojde ke snížení nerovnosti, protože pro $r > p$ díky rostoucím koeficientům platí $(X_{(r)} - \delta)(2r - N - 1) + (X_{(p)} + \delta)(2p - N - 1) < X_{(r)}(2r - N - 1) + X_{(p)}(2p - N - 1)$. Ostatní členy ve zmíněné rovnici zůstávají, což implikuje snížení Giniho indexu.
Poněvadž při každém progresivním transferu dojde ke snížení Giniho indexu, jejich konečná posloupnost bude mít tentýž efekt.
- Axiom regresivního transferu. Tentokrát dojde k transferu od jedince s pořadím p k bohatšímu jedinci s pořadím $r > p$. Obměnou argumentace z předchozího bodu nahlédneme, že GI splňuje i tento axiom.
- Axiom klesajícího efektu transferu. Pro osvětlení tohoto faktu použijme rovnici (5), jež je založena na pořadí jednotlivých pozorování. Jelikož je v předpokladech zmíněného axiomu zakotveno, že pořadí po provedení transferu zůstává zachována, lze nahlédnout, že výrazněji snižující efekt nastane, pokud daná částka bude pocházet od výše zařazeného jedince.
- Normalizace. Poněvadž je GI založen na průměrné vzdálenosti dvou pozorování, jedná se za předpokladu kladnosti střední hodnoty n.v. (což pro naše účely bude) o nezápornou veličinu. Nuly nabývá evidentně v případě degenerovaného jednobodového rozdělení. Omezenost shora lze ukázat pomocí vyjádření (2), z čehož vyplývá, že omezenost shora jedničkou je ekvivalentní $\mu^{-1} E \min(X_i, X_j) \geq 0$. Jelikož uvažujeme nezáporné náhodné veličiny s nenulovou střední hodnotou, je omezenost shora evidentní.
- Nezávislost na měřítku. Při přechodu $Z = \lambda Y$ ($\lambda > 0$) se efekt na nerovnost vyruší proporcionalní změnou střední hodnoty n.v. Z .
Naproti tomu eventuální posun vyvolaný přísunem pevné částky všem jednotkám bude v důsledku znamenat snížení Giniho koeficientu. Střední absolutní vzdálenost zůstane zachována, ale změnou očekávané hodnoty pozorování naroste jmenovatel.

2.2 Zobecněné Giniho indexy nerovnoměrnosti

Právě popsaný Giniho index je sumární charakteristikou a snadno nahlédneme, že dvě rozdělení mohou být spjaty se stejnou hodnotou Giniho indexu. Tento jev byl předmětem častých diskuzí o rozděleních z pohledu GI ekvivalentních, ale ze sociologického hlediska různě přijatelných. Pokud je tedy na základě zadání jasné, jak by měly být upraveny váhy jednotlivým pořadovým statistikám (resp. částem příjmového spektra), lze je řídit pomocí parametrů zobecněných Giniho koeficientů. Pro úplnost uveďme nejvíce používané rodiny *S-Gini* (Single Parameter Gini, rovnice (15)) respektive vyjádření pomocí distribuční funkce (16)) a *E-Gini* (Extended Gini, rovnice (17)), tak jak jsou uvedeny např. v Giorgi a kol. [6]

$$SG_\delta = 1 - \frac{1}{N^\delta \mu_X} \sum_{i=1}^N \left((N-i+1)^\delta - (N-i)^\delta \right) X_{(i)} \quad (15)$$

$$SG_\delta = \frac{\mu_X - a}{2\mu_X} - \frac{1}{2\mu_X} \int_a^b (1 - F(x))^\delta dx \quad (16)$$

$$EG_\nu = 2 \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{i}{N} - \frac{1}{N\mu_X} \sum_{j=1}^i X_{(j)} \right)^\nu \right]^{1/\nu} \quad (17)$$

Při použití těchto indexů je potřeba zvolit parametry, které mění relativní významnost bodů příjmového spektra, a to za podmínky $\delta \geq 1$ respektive $\nu \geq 1$. Porovnáním právě uvedených tvarů s vyjádřeními původního indexu zjistíme, že pro volbu $\delta = 2$ respektive $\nu = 1$ dochází k ekvivalenci obou indexů a jedná se o klasický Giniho index.

Z vyjádření (15) sledujeme, že volbou většího parametru δ přesunujeme váhy na nižší příjmy. V případě indexu E-Gini vidíme, že definice přímo používá Lorenzovu křivku a její transformace (v případech $\nu \neq 1$). Je zde zvyrazňována odchylka aktuální Lorenzovy křivky a 45° přímkou (Lorenzova křivka stejných příjmů). Zajímavou souvislostí (viz Yitzhaki a Schechtman [1]) je možnost vyjádřit E-Gini pro celočíselné parametry ν jako $I_{R,\delta} = 1 - \mu_X^{-1} \min(X_1, X_2, \dots, X_\delta)$.

2.3 Druhy používaných dat

Giniho index lze aplikovat na mnoho druhů veličin, které ze zásady nabývají kladných hodnot. Nejčastější oblastí bývá výpočet nerovnosti v úrovni příjmů jedinců nebo domácností, méně často pak nerovnost v úrovni spotřeby.

Využití získaných Giniho koeficientů pro srovnání úrovně nerovnosti (např. v různých státech) je obvykle ztíženo rozdílností metodik získávání zdrojových dat. Za příklad vezměme rozdíl v použití hrubých mezd nebo hrubých příjmů jedinců, v čemž je značný, avšak často opomíjený, rozdíl. Jiná situace nastává použitím mezd (příjmů) čistých nebo zkreslení pocházející z faktu, že v některých státech pochází data od jednotlivců a v jiných se data pro jedince získávají úpravou dat získaných z domácností. Uvedené případy by měly podtrhnout potřebu obezřetnosti při srovnávání.

Ucelenou databázi lze nalézt např. v údajích skupiny Luxembourg Income Study, kde jsou uvedeny i hodnoty České republiky pro roky 1992 a 1996 (ty

pochází ze statistického šetření na domácnostech a následně byly standardizovány na obyvatele). Hodnoty Giniho indexu jsou $G_{LIS,CZ,92} = 20,71\%$ resp. $G_{LIS,CZ,96} = 25,81\%$.

2.4 Rozdělení příjmů

Představili jsme si různá vyjádření Giniho indexu. Pro odhad jeho hodnoty lze použít jeho klasický výběrový tvar vycházející z Giniho návrhu (popřípadě jej odhadnout jiným způsobem; např. pro seskupená data z odhadu Lorenzovy křivky). Pro hlubší analýzu odhadů je však žádoucí znát rozdělení příjmů ve společnosti. Tento předpoklad je mnohdy v reálných situacích obtížně splnitelný, protože podobnost s jakýmkoliv běžně používaným rozdělením může být nedostatečná. Jak bývá často zmiňováno a jak si lze všimnout i v dodatcích na grafu celosvětového rozdělení příjmů, mnohdy je empirické rozdělení multimodální.

Na druhou stranu, pokud se nám podaří nalézt explicitní vyjádření tvaru distribuční funkce, otevírají se nám možnosti pro analýzu dalších aspektů (rozdělení GI, parametrické vyjádření jeho rozptylu atd.). Obecnými požadavky na náhodnou veličinu použitou pro popis příjmů je nezápornost, neomezenost zprava (ohraničenost lze ospravedlnit cílem analýzy – např. nerovnost v populaci bez uvažování extrémních příjmů). Uveďme některé používané rodiny rozdělení:

Paretovo rozdělení⁷ Jeho hustota a distribuční funkce jsou dány následovně:

$$f_{k,m}^{Pareto}(y) = k \frac{m^k}{y^{k+1}}$$

$$F_{k,m}^{Pareto}(y) = 1 - \left(\frac{m}{y}\right)^k$$

pro kladný parametr k a pro hodnoty $y \geq m > 0$.

V budoucnu Paretovo rozdělení použijeme, provedme proto na tomto místě výpočet střední hodnoty

$$\begin{aligned} EX &= \int_m^\infty yk \frac{m^k}{y^{k+1}} dy = km^k \int_m^\infty y^{-k} \\ &= km^k \left[\frac{1}{-k+1} y^{-k+1} \right]_m^\infty = km^k \frac{1}{k-1} m^{-k+1} \\ &= \frac{km}{k-1} \end{aligned} \tag{18}$$

Logaritmicko-normální rozdělení Mějme kladné parametry m , θ a σ . Pak je hustota logaritmicko-normálního rozdělení definována pro $y \geq \theta$ rovnicí

$$f_{m,\theta,\sigma}(y) = \frac{\exp\left(-\left[\frac{\ln(y-\theta)}{m}\right]^2 / 2\sigma^2\right)}{\sqrt{2\pi}\sigma(y-\theta)}$$

⁷ pojmenováno po Vilfredo Federico Damaso Paretovi (1848-1923). Ten v letech 1896 a 1897 publikoval návrh rodiny rozdělení náhodných veličin s účelem aproximovat rozdělení příjmů.

Kromě těchto obecně rozšířených rozdělení se pro příjmy jedinců často používají dva speciální případy zobecněného beta rozdělení druhého typu. Jedná se o Dagumovo rozdělení a Singh-Maddalovo rozdělení.

Dagumovo rozdělení vzniká pevnou volbou parametru $q \equiv 1$. Jeho hustota je

$$f_{\beta,\delta}(x) = \beta\delta x^{-(\delta+1)} (1+x^{-\delta})^{-(\beta+1)}$$

pro kladné parametry a kladné y .

Singh-Maddalovo rozdělení s hustotou

$$f_{a,b,q}^{S-M}(y) = aq \frac{y^{a-1}}{b^a [1+(y/b)^a]^{1+q}}$$

vznikne fixací parametru $p \equiv 1$.

Předpoklad o rozdělení dat lze pak využít a např. pomocí rovnice (1) (vyjadřuje souvislost Giniho indexu a distribuční funkce) odvodit vzorec pro GI v závislosti na parametrech rozdělení, jak tomu bude např. v následující podkapitole pro Paretovo rozdělení (jeho Lorenzova křivka je odvozena v kapitole 5).

Znalost rozdělení příjmů bude mít vliv na schopnost vypočítat hodnotu GI pro danou společnost. Bohužel až na výjimky nám nepomůže v určení exaktního rozdělení GI jakožto náhodné veličiny. Tomuto tématu se budeme věnovat v podkapitole (2.6).

V grafu (2) ilustruji průběh Lorenzovy křivky. Použito bylo Dagumovo, logaritmicko-normální a Paretovo rozdělení⁸, vždy s parametry odpovídajícími Giniho indexu na úrovni údaje z roku 1996 pro Českou republiku (25,4 %).

Zeleně vyznačená Lorenzova křivka Paretova rozdělení s odhadnutým parametrem se jasně odlišuje od zbylých dvou strmějším nárůstem křivky v horní oblasti; tedy hustota příjmu v horním spektru je vyšší než u zbylých rozdělení. Pozorovatelný je též rovnoměrnější nárůst ve spodních hodnotách.

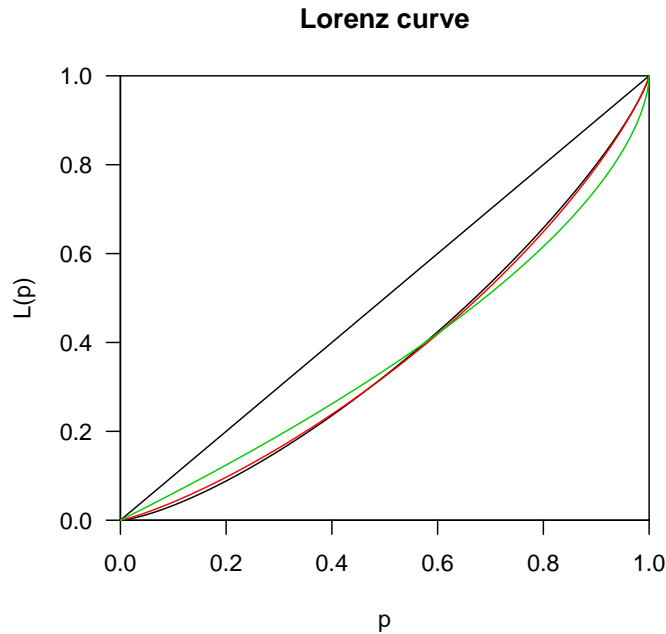
2.5 Způsoby odhadování Giniho indexu a intervaly odhadu při použití seskupených dat

Pro odhad Giniho indexu lze běžně použít jednotlivá pozorování. Mnohdy ovšem takto detailní popis populace není k dispozici a dostupná jsou pouze v seskupené podobě. Typicky jsou zveřejňovány decilové údaje – podíly vzestupně seřazených desetin na celkových příjmech populace. V následujícím se pokusíme představit odhadové procedury pro seskupená data. Používat při tom budeme data z *Luxembourg Income Study* popisující Českou republiku v roce 1996⁹ pomocí podílů deseti skupin (neprolínající se dělení podle výše příjmů) na celkových příjmech.

Uveřejněný index je $G_{CZ,1996} = 25,80946\%$. Kromě této hodnoty máme navíc k dispozici kumulované podíly na deset částí rozdělené populace.

⁸výpočet Lorenzových křivek byl proveden pomocí knihovny *ineq* pro statistické prostředí R

⁹Tyto údaje pocházejí z průzkumu na domácnostech. Výsledná nerovnost je ale uváděna ve vztahu k jednotlivcům. Obvykle se tento přechod provádí vydělením hodnoty pro celou domácnost počtem jejích členů nebo počtem standardizovaných členů (např. dítě může mít váhu 70 % dospělého)



Obrázek 2: Lorenzovy křivky různých rozdělení, Giniho index všech je shodný

0 %	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	100 %
0	4,028	9,825	16,667	24,397	32,861	42,168	52,361	63,988	77,859	100

Právě taková anonymizace a sumarizace dat je často jedinou dostupnou datovou základnou. Nyní si ukážeme, jak pomocí takových údajů přibližně spočítat hodnotu Giniho indexu. Popíšeme si několik možných způsobů:

- Lichoběžníkové pravidlo.

Nejpřirozenější postup spočívá v nahrazení diskrétní Lorenzovy křivky po částech lineární funkcí vzniklou prostou interpolací mezi známými body

$$\tilde{L}(a_i + d) = L(a_i) + d \times \frac{L(a_{i+1}) - L(a_i)}{a_{i+1} - a_i} \quad 0 < d < a_{i+1} - a_i$$

Takový odhad Lorenzovy křivky vlastně popisuje zjednodušenou populaci, kde všichni příslušníci jednotlivých skupin mají příjem rovný skupinovému průměru.

Jak bylo uvedeno, Lorenzova křivka je z principu konvexní. Pokud ji nahradíme lineární lomenou funkcí (přirozeným „spojením“ známých bodů), pak plocha mezi ní a kladnou horizontální poloosou bude větší (nebo výjimečně rovna) než jak by tomu bylo pro křivku příslušící přímo k rozdělení příjmů. Tedy takto odhadnutá hodnota bude menší než opravdové GI (lze podrobněji nahlédnout v kapitole 4).

Zkusme metodu aplikovat na výše zmíněná data. Potom dostaneme odhad

25,17 %, což srovnáme s reálným Giniho indexem 25,81 %. Rozdílnost hodnot tedy při použití decilů (a stavu ČR 1996) není významná, ale rozhodně nemůžeme považovat odchylku zhruba 0,64 procentního bodu za zanedbatelnou.

- Pro zjištění horní hranice hodnoty GI ze seskupených dat se nám naskýtá několik možností. Triviální přístup by byl založen na konstrukci dolní hranice pro Lorenzovu křivku – schodovité funkce zdefinované pomocí známých bodů empirické Lorenzovy křivky ($\rho(x) = L(a_i)$ pro $x \in [a_i, a_{i+1})$). Ta by byla následně použita k výpočtu dvojnásobku plochy mezi křivkou ρ a LC absolutní rovnosti.

Těsnější horní hranici možné hodnoty GI populace zjistíme, pokud využijeme dalších vlastností Lorenzových křivek. Pokročilou metodu představuje Mehran v [9], kde využívá konvexnosti Lorenzovy křivky a z toho plynoucího faktu, že tečna v každém bodě a_i (samozřejmě vyjma a_0) má větší sklon než spojnice $L(a_{i-1})$ a $L(a_i)$ respektive je menší než u spojnice $L(a_i)$ a $L(a_{i+1})$. Takto zkonstruované tečny jistě „podbíhají“ Lorenzovu křivku příslušící původním datům. Použijeme-li její části tak, aby tvořily po částech lineární spojnici bodů $(0, 0)$ a $(1, 1)$, jsme s to určit horní hranici pro hodnotu GI původní populace. Otázka sklonu tečen se řeší úlohou maximalizace plochy mezi Lorenzovou křivkou perfektní rovnosti a zmíněnými tečnami utvořenou hranicí vzhledem k omezením popsáním v předchozí větě.

Aplikací Mehranova postupu můžeme spočítat¹⁰ velmi těsnou horní hranici 26,23 %.

Kromě tohoto postupu je zde také uveden zjednodušený, ale velmi hrubý horní odhad¹¹. Ten vychází z téhož principu, ale bez aplikace omezení na optimalizované sklony. Pro naše data určuje značně nadsazenou horní hranici 81,62 %.

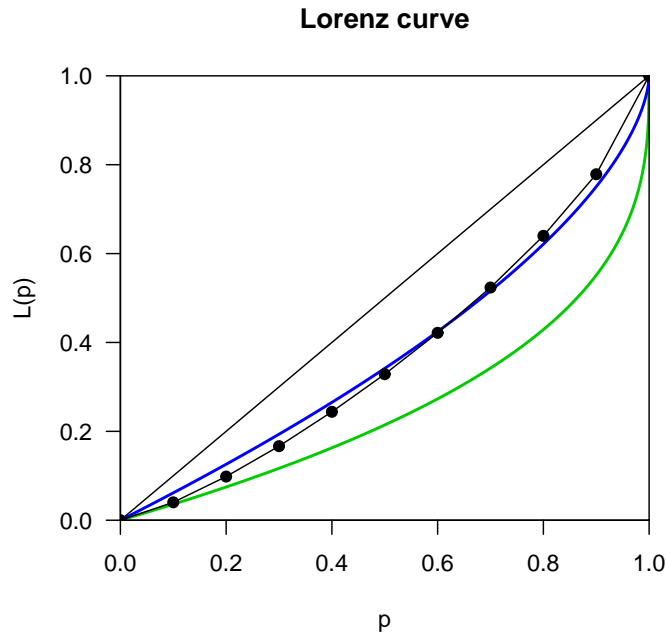
Z provedených výpočtů můžeme usuzovat, že odhady Giniho indexu pro podobné případy (zkoumání nerovnoměrnosti v příjmech jednotlivců, data ve formě decilových podílů) jsou přijatelně blízko reálné hodnotě Giniho indexu.

- Volba distribuce příjmů.
Pokud máme nějaký oprávněný předpoklad o typu rozdělení příjmů, lze odhadnout jeho parametry a následně vyjádřit GI tohoto rozdělení (např. podle vztahu (3)).
Ilustrujme tento postup pro Paretovo rozdělení. Životní minimum¹² v ČR v roce 1996 bylo na úrovni 1800 Kč (zjednodušení, bez nákladů na domácnost a pro zaopatřené osoby nad 15 let). Zvolíme tedy první parametr rozdělení $m = 1800$. Druhý parametr můžeme odhadnout ze závislosti střední hodnoty na parametrech $EX = \frac{km}{k-1}$, z čehož úpravou $k = \frac{EX}{EX-m}$. Uveřejněný výběrový průměr činil 5171 Kč; tedy při jeho užití jako odhadu střední hodnoty by odpovídající odhad parametru k byl $\hat{k} = 1,534$. Použijeme rovnost

¹⁰ použijeme metodu implementovanou v knihovně *ineq* (pro R).

¹¹ $1 - 4 \sum_{i=0}^{k+1} (p_i - p_i^*) (L(p_i) - L(p_i^*))$, body p_i^* jsou dány následovně: $p_{k+1}^* = 1$, ostatní $p_i^* = 2p_i - p_{i+1}^*$

¹² Tento ukazatel byl zvolen kvůli charakteru dat jako nejpříhodnější. V průzkumu by se neměl vyskytovat jedinec s menším příjmem. Naproti tomu se v průzkumu mohou vyskytovat i jedinci bez zaměstnání, tedy pod úrovní minimální mzdy.



Obrázek 3: Lorenzovy křivky pro Paretova rozdělení s odhadnutými parametry a výchozí Lorenzova křivka

$G_{Pareto} = \frac{1}{2^{\hat{k}-1}}$ (důkaz viz(31)) a pro takové rozdělení pak dostáváme

$$G_{CZ96,Pareto}^{[1]} = \frac{1}{2_1^{\hat{k}-1}} = 48,36 \%$$

Zde je vidět, že tento způsob odhadu parametru nebo předpoklad o typu rozdělení není vhodný.

Jinou možností je odhadnout druhý parametr k z jedenácti známých bodů Lorenzovy křivky, která je nezávislá na parametru m Paretova rozdělení. V dodatcích lze vyhledat vyjádření hodnot LC, které je stejně jako vyjádření GI nezávislé na parametru m .

Jelikož krajní body Lorenzovy křivky jsou fixovány, zůstává devět relevantních bodů. Jako kritérium shody zvolme součet druhých mocnin odchylek a numerickou aproximací (krok 0,01) zvolíme hodnotu $2_1^{\hat{k}} = 2,52$. Čemuž by odpovídala hodnota indexu

$$G_{CZ96,Pareto}^{[2]} = 24,75 \%$$

Na grafu (3) je vidět, že zeleně vyznačená Lorenzova křivka odpovídající rozdělení s parametrem $1_1^{\hat{k}}$ se značně odchyluje od empiricky získaného průběhu. Z tohoto hlediska bylo možno již na počátku očekávat lepší výsledek od druhé metody. Její výsledek se poměrně dobře přimyká.

- Dílčí předpoklady o typu rozdělení v jednotlivých skupinách
Znamá data o vlastnostech jednotlivých disjunktních příjmových skupin (se známými hraničními body) mohou posloužit pro odhad rozdělení celé populace. Pro každou skupinu se obecně může použít jiné rozdělení. Získáváme tak zobecnění předešlé metody.
Jelikož je poslední skupina přirozeně otevřená směrem k ∞ , je vhodné použít rozdělení definované až do ∞ .
- Aproximace Lorenzovy křivky pomocí funkce je dalším možným přístupem. Volbou typu této funkce se v podstatě omezujeme na nějakou množinu odpovídajících rozdělení.
Pro ilustraci opět použijme již představená data. Velmi jednoduchou křivkou splňující základní podmínky kladené na Lorenzovu křivku může být např. $L(p) = p^k$. Pokud zkusíme vyhledat ideální k pro naše data, opět podle stejného kritéria, jako optimální volba se jeví ${}_3\hat{k} = 1,66$. Kvadratická odchylka od jedenácti (resp. devíti relevantních) bodů empirické LC je o něco méně než trojnásobek hodnoty v případě Paretova rozdělení (2. postup) : 0,0101 versus 0,0037.
Samozřejmě volba vyššího řádu polynomu zvýší přesnost. Stejně tak zajisté existují i jiné typy funkcí, jejichž použití by mělo stejný důsledek. Je ovšem nutné mít na zřeteli, že by v takovém případě už odhadnutá křivka nemusela splňovat vlastnosti Lorenzovy křivky.

Kromě těchto možných cest můžeme zvolit i jiné. Pionýrem této oblasti byl Gastwirth v [4], následován mnohými dalšími.

2.6 Statistické vlastnosti odhadu Giniho indexu

Doposud jsme se zabývali odhadem Giniho indexu. Navazující otázkou je rozdělení této statistiky. Jelikož není obecně známo řídicí rozdělení ani jeho střední hodnota, jedná se v zásadě o otázku rozdělení poměru absolutního Giniho indexu a odhadu střední hodnoty (typicky výběrový průměr).

Na téma exaktního odhadu rozdělení Giniho indexu bylo publikováno mnoho článků. Bohužel jsou aplikovatelné pouze na několik typů rozdělení příjmů. Za všechny jmenujme výsledky pro normální, trojúhelníkové a exponenciální řídicí rozdělení odvozené v Nair [12].

Obecně je ale exaktní rozdělení Giniho indexu komplikovaná záležitost; představme zprvu alespoň asymptotické vlastnosti. Asymptotické rozdělení pro celou rodinu S-Gini (viz rovnost (15)) indexů lze například najít v Giorgi a kol. [6]. Zmiňují, že limitní rozdělení je normální

$$\sqrt{n} \frac{SG_{n,\delta} - SG_{F,\delta}}{\sigma_{F,\delta}} \sim^{asy} N(0, 1)$$

Při aplikaci zmíněného výsledku se dostaneme k problému určení členu $\sigma_{F,\nu}$, který vyjadřují jako

$$\sigma_{F,\delta}^2 = \int_0^1 \int_0^1 (\min(x, y) - xy) J(x) J(y) dF^{-1}(x) dF^{-1}(y)$$

$$J(q) = (1 - q)^{\delta-1} - \delta^{-1} (1 - G_\delta)$$

Následně pro klasický Giniho index volíme $\delta = 2$. Je zřejmé, že komplexnost tohoto výrazu omezuje možné využití na známé případy v nichž je známé rozdělení řídicí náhodné veličiny. Problémem tohoto přístupu může být také fakt, že rozdělení příjmů se málokdy přibližuje k některému z běžně používaných rodin rozdělení¹³. Nicméně stále platí, že Giniho koeficient je asymptoticky normální, ačkoliv, jak je zmíněno ve uvedeném článku, se nejedná o konvergenci uniformní. Dále Giorgi a kol. [6] uvádějí, že konfidenční intervaly získané pomocí asymptotického chování jsou široké a ostatní výsledky se dostatečně neshodují se simulačními. Zmíněná fakta hovoří v neprospěch asymptotiky pro empirické studie. Pro praxi se tedy většinou odhadnou parametry z populace a ty se použijí (při vědomí, že odhad samotný je zatížen chybou).

Důsledek: *Odhad zobecněného Giniho indexu je konzistentní. Tím pádem je konzistentní i odhad klasického Giniho indexu.*

Bohužel je v uvedeném článku konstatováno, že konvergence není uniformní.

Rozptyl odhadu absolutního Giniho indexu

V tomto odstavci se budeme zabývat absolutní variantou Giniho indexu. Protože odhad relativního Giniho indexu vzniká podílem dvou statistik (průměrná vzdálenost pozorování a výběrový průměr), jedná se o složitější problém, kterým se budeme zabývat později. Věnujme se rozptylu absolutního Giniho indexu (bude značen Δ). Nestranným odhadem absolutního Giniho indexu je

$$\hat{\Delta} = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N |x_i - x_j| \quad (19)$$

Abychom získali rozptyl odhadu absolutního GI, použijeme uvedený odhad a notoricky známou rovnost $\text{var}(Y) = E(Y^2) - (EY)^2$. Vyjádříme si nyní člen $E(\hat{\Delta}^2)$ využitím rovnice (19):

$$\begin{aligned} E(\hat{\Delta}^2) &= E \left[\left(\frac{1}{N(N-1)} \sum_{j \neq i} |X_i - X_j| \right)^2 \right] \\ &= \frac{1}{N^2(N-1)^2} \sum_{i=1}^N \sum_{j \neq i}^N \sum_{k=1}^N \sum_{l \neq k}^N E(|X_i - X_j| |X_k - X_l|) \end{aligned}$$

Poslední rovnost se čtyřmi úrovněmi sumace obsahuje

- případy souladu prvních dvou a druhých dvou sumačních indexů $\{i, j\} = \{l, k\}$ v počtu $2N(N-1)$. Jedná se o členy $E(|X_i - X_j|^2)$, které nabývají zřejmě hodnoty $2\sigma_X^2$ a následně zjišťujeme, že celkový příspěvek těchto členů do hodnoty $E(\hat{\Delta}^2)$ dává $4[N(N-1)]^{-1} \sigma_X^2$.

¹³Pro ilustrativní a účely výzkumu se používají rozdělení normální, log-normální, ... nebo také často speciálně kvůli modelování příjmů vyvinuté Dagumovo rozdělení.

V dodatcích si je možno prohlédnout celkové rozdělení příjmů na světě z roku 1993.

- případy $i \neq l$ a zároveň $j \neq k$, pro obvykle velké rozsahy výběrů bude zřejmě nejčastější, což je v důsledku Δ^2 a to celkem v $N(N-1)(N-2)(N-3)$ případech.
- nakonec zbývá $4N(N-1)(N-2)$ členů, které připadají na částečné „překrytí“ sumačních indexů ($i = l$ a zároveň $j \neq k$, resp. převráceně). Tento člen značme $K = E(|X_1 - X_2| |X_1 - X_3|)$.

Můžeme tedy ekvivalentně zapsat

$$E(\hat{\Delta}^2) = \frac{1}{N(N-1)} [4\sigma_X^2 + 4(N-2)K + (N-2)(N-3)\Delta^2] \quad (20)$$

a následně použitím zmíněné rovnosti pro rozptyl a vzhledem k nestrannosti odhadu absolutního Giniho indexu (19)

$$\text{var}(\hat{\Delta}) = \frac{1}{N(N-1)} [4\sigma_X^2 + 4(N-2)K - 2(2N-3)\Delta^2]$$

Zmíněná rovnost nám tedy ukazuje závislost rozptylu odhadu absolutního indexu na řídicí náhodné veličině. Pokud známe řídicí náhodnou veličinu, lze aplikovat zmíněný výsledek přímo na rozptyl výběrového relativního Giniho indexu (poněvadž známe i střední hodnotu). Nicméně pro neznámé řídicí rozdělení nezbývá než použít odhadů pro neznámé charakteristiky n.v.

Nestranný odhad rozptylu odhadové statistiky $\hat{\Delta}$ vychází z nestranného odhadu rozptylu $S^2 = (N-1)^{-1} \sum (x_i - \bar{x})^2$ a odhadu hodnoty $E(|X_i - X_j| |X_i - X_k|)$ ve tvaru

$$\hat{K} = \frac{1}{N(N-1)(N-2)} \sum_i^N \sum_j^N \sum_k^N |x_i - x_j| |x_i - x_k| - \frac{2S^2}{N-2}$$

Finálně si tedy můžeme zapsat odhad rozptylu absolutního Giniho indexu

$$\hat{\text{var}}(\hat{\Delta}) = \frac{4}{N(N-1)} \left[S^2 + (N-2)\hat{K} - \frac{2N-3}{2}\hat{\Delta}^2 \right]$$

Druhý možný odhad rozptylu vyplývá z rovnice (20) použitím odhadu pro rozptyl řídicí náhodné veličiny $\tilde{\sigma}_X^2 = N^{-1} \sum_{i=1}^N (x_i - \bar{x})^2$ a pomocí odhadu $\tilde{K} = \frac{1}{N^3} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N |x_i - x_j| |x_i - x_k|$ resp. $\tilde{\Delta} = N^{-1} \sum^N \sum^N |x_i - x_j|$.

$$\tilde{\text{var}}(\hat{\Delta}) = \frac{4}{N(N-1)} \left[\tilde{\sigma}^2 + (N-2)\tilde{K} - \frac{2N-3}{2}\tilde{\Delta}^2 \right]$$

Z obou finálních vyjádření lze pozorovat jakým způsobem se projevuje rozptyl příjmů, a že samotná vyšší úroveň GI implikuje menší vlastní rozptyl. Za povšimnutí stojí fakt, že v poslední rovnici používáme pro vyjádření odhadu rozptylu $\hat{\Delta}$ hodnotu $\tilde{\Delta}$. Dále s rostoucím rozsahem výběru se rozptyl snižuje úměrně druhé mocnině velikosti populace. Přičemž nejvíce se na poklesu bude projevovat rozptyl výchozí náhodné veličiny. Vliv veličiny K je řádově stejně silný jako vliv absolutního Giniho indexu.

Numerické výsledky V minulém odstavci jsme se zabývali rozptylem výběrového absolutního Giniho indexu. Přejichod od absolutní k relativní verzi však znamená v obvyklém případě (neznámá střední hodnota rozdělení) dělení výběrovým průměrem. Účelem tohoto odstavce je ukázat simulační výsledky rozptylu Giniho indexu v závislosti na velikosti vzorku.

Předpokládejme, že používáme Paretovo rozdělení nebo logaritnicko-normální model příjmů. Můžeme si tak nastíinit, jaký vliv může mít volba rozdělení na rozptyl odhadu Giniho indexu. Budeme postupovat takto: pro každé rozdělení provedem 1000× odhad hodnoty GI z náhodného vzorku, který bude obsahovat jistý počet realizací. V následující tabulce předkládám získané charakteristiky v případech, kdy náhodný vzorek byl generován s objemem 50, 100, 1 000 nebo 10 000 realizací. Použito bylo Paretovo a logaritnicko-normální rozdělení (viz graf (2)). V tabulce (1) si můžeme všimnout, že volba logaritnicko-normálního modelu příjmů s sebou přinesla až řádově lepší hodnoty výběrového rozptylu.

	50	100	1 000	10 000
Paretovo výběrový průměr	24,17 %	24,71 %	25,34 %	25,40 %
výběrový rozptyl	$3,73 \times 10^{-3}$	$2,17 \times 10^{-4}$	$3,05 \times 10^{-4}$	$3,41 \times 10^{-5}$
logaritnicko-normální výběrový průměr	24,87 %	25,13 %	25,36 %	25,39 %
výběrový rozptyl	$6,94 \times 10^{-4}$	$3,51 \times 10^{-4}$	$3,48 \times 10^{-5}$	$3,55 \times 10^{-6}$

Tabulka 1: Simulační výsledky pro Paretovo a logaritnicko-normální rozdělení

Kromě simulačních pokusů se pro odhad rozptylu výběrového Gini indexu také často používají bootstrapové metody¹⁴. Tento přístup je intuitivní a získal si své místo.

2.7 Gini a regrese

Podívejme se na možné využití kovariančního vysvětlení Giniho indexu, které nás dovádí k souvislosti s metodou nejmenších čtverců a odhadem parametrů lineárního modelu $X_{(i)} = \alpha + \beta i + \epsilon_i$. Rozeberme si podrobněji. Jak je obecně známo, odhad parametru β metodou nejmenších čtverců bude mít pro uvedený model tvar

$$\hat{\beta} = \frac{N \sum i x_i - \sum i \sum x_i}{N \sum i^2 - (\sum i)^2}$$

¹⁴nevýhodou při použití bootstrapu oproti simulačnímu přístupu se může stát citlivost Giniho indexu na extrémní hodnoty (viz např. vyjádření (5))

a použijeme-li vztah $\sum_{i=1}^N i^2 = \frac{1}{6}N(N+1)(2N+1)$, můžeme jednoduchými úpravami psát

$$\begin{aligned}
\hat{\beta} &= \frac{N \sum_{i=1}^N i x_{(i)} - N\bar{x} \sum_{i=1}^N i}{\frac{1}{6}N^2(N+1)(2N+1) - N^2\left(\frac{N+1}{2}\right)^2} \\
&= \frac{N \sum_{i=1}^N i (x_{(i)} - \bar{x})}{\frac{1}{6}N^2(N+1)(2N+1) - N^2\left(\frac{N+1}{2}\right)^2} \\
&= \frac{\sum_{i=1}^N \left(i - \frac{N+1}{2}\right) (x_{(i)} - \bar{x}) + \frac{N+1}{2} \sum_{i=1}^N (x_{(i)} - \bar{x})}{\frac{1}{6}N(N+1)(2N+1) - \frac{1}{4}N(N+1)^2} \\
&= \frac{\sum_{i=1}^N \left(i - \frac{N+1}{2}\right) (x_{(i)} - \bar{x})}{\frac{1}{6}N(N+1)(2N+1) - \frac{1}{4}N(N+1)^2} \\
&= \frac{N \hat{cov}_{X, F_X}}{\frac{1}{6}(N+1)(2N+1) - \frac{1}{4}(N+1)^2}
\end{aligned}$$

Pak docházíme použitím vzorce (12) a následujícím výpočtem

$$\begin{aligned}
G &= \frac{2cov(X, F_X)}{\bar{x}} \\
&= \frac{2\hat{\beta}}{N\bar{x}} \left[\frac{1}{6}(N+1)(2N+1) - \frac{1}{4}(N+1)^2 \right] \\
&= \frac{\hat{\beta}}{N\bar{x}} \left[\frac{1}{6}(N+1)(4N+2-3N-3) \right] \\
&= \frac{\hat{\beta}}{6N\bar{x}} [(N+1)(N-1)]
\end{aligned}$$

k jinému vyjádření odhadu Giniho indexu pomocí lineární regrese

$$G = \frac{N^2 - 1}{6N} \frac{\hat{\beta}}{\bar{x}}$$

Tato rovnost ukazuje další možný postup pro odhad Giniho indexu. A dále za předpokladu nezávislých chybových členů $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ je rozdělení odhadu parametru $\hat{\beta} \sim N\left(\beta, \sigma_\epsilon^2 \left(\frac{1}{3}N(N-1)(N+1)\right)^{-1}\right)$.

Kritickým pohledem ovšem odhalíme, že vycházíme ze znalosti rozdělení dat. Tento předpoklad je v této chvíli nepodložený, navíc data obecně nebudou vykazovat potřebné kvality (obzvláště normalita bude u příjmů stěží dosažitelná). Giles [5] se věnoval i srovnání odhadů rozptylu GI získaných výše zmíněným postupem a jack-knife metodou. Použity byly data z Penn World Table pro 133 států v několika časových bodech¹⁵. Rozdíly rozptylů získaných oběmi metodami se ukázali vcelku přijatelnými – cca do 15 %. V kontextu minimální numerické náročnosti lze tento postup chápat jako dobrou metodu prvotního zjištění přibližné hodnoty rozptylu GI.

Druhou možnou cestou, kterou je možno zmínit, je použití metody vážených nejmenších čtverců (Weighted Least Squares) pro lineární model $i = \theta + \nu_i$, kde ν_i

¹⁵jedná se data o spotřebě na jednoho obyvatele

jsou chybové členy s rozptylem σ^2/x_i . Potom si bez odvození uveďme, že Giniho index lze spočítat i následujícím způsobem uvedeným v Giles [5]

$$G = \frac{2\hat{\theta}}{N} - 1 - \frac{1}{N}$$

připomene-li si $\hat{\theta} = \frac{\sum ix_{(i)}}{\sum x_i}$. Výhodou použití metody vážených nejmenších čtverců je, že Giniho index je jednoduchou lineární transformací pouze odhadnutého parametru modelu a tím pro rozptyl Giniho indexu zřejmě platí $\text{var}(G) = 4N^{-2}\text{var}(\hat{\theta})$.

Komparace Giniho indexů

Častou praktickou úlohou je porovnání hodnot Giniho indexu, ať už vzhledem k časové, místní či jiné doméně. Závěry takových analýz by však měly být podpořeny statistickými výsledky. Zejména se jedná o hypotézy o rovnosti indexu pro dva různé výběry, případně jednostranné testy. Uplatnění zde může nalézt srovnání Lorenzových křivek

Vrátíme-li se zpět k Giniho indexu, pak před námi leží v zásadě dvě možné cesty. Buď aplikací bootstrapu nebo přes asymptotické rozdělení GI, tak jak bylo uvedeno v kapitole (2.6).

Pro testování významnosti rozdílnosti dvou hodnot indexu je možno využít například rovnou bootstrap hodnoty $G_1 - G_2$.

3 Alternativní míry chudoby

Diskuze o potřebných vlastnostech dobrého indexu chudoby přinesla mnoho návrhů nových měr. Nicméně nadále jsou v praxi nejvíce používány jednoduché statistiky. Pokročilejší přístupy vycházejí obvykle z axiomatického přístupu motivovaného teorií společenského blahobytu; v čemž odkazují na kapitolu 1.

Giniho index by dle rigidního členění patřil mezi míry nerovnoměrnosti, bývá však zmiňován a používán i v kontextu indexů chudoby. Budeme-li ho používat pro společnosti, ve kterých je vedle blahobytu přítomna i chudoba, pak se i on stává zrcadlem chudoby.

Podíl obyvatelstva pod hranicí chudoby je jednoduše uchopitelná veličina. Bohužel její vypovídací hodnota je extrémně závislá na nastavení hranice chudoby. K zohlednění rozdílů v množstvích prostředků potřebných pro pokrytí nutných výdajů v jednotlivých zemích se může použít parita kupní síly.

Kumulovaný propad jedinců pod hranicí chudoby. V tomto případě spočítáme kumulativní rozdíly příjmů jedinců pod hranicí chudoby od této hranice. Na jednu stranu získáváme informativnější charakteristiku, na straně druhé ztrácíme výhodu indexu, jež by nám dovozoval porovnávat různé populace. Navíc stále zůstává problém nastavení hranice chudoby.

Axiomatický přístup k tvorbě indexů, které by měli zvolené vlastnosti, vedl k mnoha novým návrhům. Ty se ovšem tolik neujaly.

Zobecněný Shorrocksův index. Jedná se o třídu indexů, které jsou diferencovatelné, striktně S-konkávni¹⁶ a jsou definovány následovně:

$$S_c = \frac{1}{Nc(c-1)} \sum \left[\left(\frac{X_i}{\mu} \right)^c - 1 \right] \quad c \neq 0, c \neq 1$$

Jako speciální případ (rozšíření pro $c = 1$) obsahuje i *Theilův index*, který vychází z Shannonovy míry informační entropie. Jeho výhodou je přímočará dekompozice dle subpopulací. Definice Theilova indexu

$$T = S_1 = \frac{1}{N} \sum_{i=1}^N \left(\frac{X_i}{\mu} \right) \ln \left(\frac{X_i}{\mu} \right)$$

Naposledy, pro $c = 0$ je zadefinován jako průměrná odchylka logaritmu střední hodnoty od logaritmu náhodné veličiny

$$S_0 = N^{-1} \sum_{i=1}^N \ln \left(\frac{\mu}{X_i} \right)$$

Shorrocksova varianta Senova indexu Předpokládáme hranici chudoby na úrovni z . Jeho definice je

$$P_{SST}(X, z) = \frac{1}{N^2} \sum_{i: X_{(i)} < z} \frac{X_{(i)} - z}{z} (2N - 2i + 1)$$

Jedná se o index, který splňuje axiom symetrie a je invariantní vůči replikaci.

Na závěr učinme ještě poznámku o měřích polarizace. Ty jsou konstruovány za účelem vyjádřit rozsah roztržetosti společnosti. Tématicky do jisté míry souvisí s měřami nerovnoměrnosti, pokoušejí se však nahlížet na společnost jako více kompaktních skupin, jejichž oddělenost je brána jako nežádoucí a rozsah této oddělenosti se snaží numericky zachytit.

4 Dekompozice Giniho indexu

Následující kapitola se bude zabývat strukturou Giniho indexu. Hned vedle samotné hodnoty GI se do centra zájmu dostávají důvody nerovnoměrnosti. Struktura této veličiny má významnou informační hodnotu a její pochopení může pomoci, kde je to žádoucí, k zvolení efektivních kroků pro snížení nerovnosti. Rozlišujeme dva různé pohledy. Zaprvé se běžně zkoumá souvislost celkového GI s hodnotami indexu spočtenými pro jednotlivé skupiny disjunktčního rozkladu a nebo způsob jímž se jednotlivé složky příjmů (nebo spotřeby) odrážejí v celkové hodnotě GI.

V zásadě budeme přeskupovat členy dvojice sumy

$$\sum_{i=1}^N \sum_{j=1}^N |X_i - X_j|$$

¹⁶Funkci W nazveme S-konkávni (Schur-konkavní), jestliže pro každou biastickou matici B platí $W(By) \geq W(y)$.

Jsou to funkce z principu splňující axiom symetrie.

do tématicky souvisejících bloků. Předpokládáme, že obecně v každé skupině nebo každý zdroj příjmů se řídí jiným rozdělením. Náhodné veličiny uvnitř dané skupiny nebo popisující jistý zdroj příjmů však budeme nadále předpokládat stejně jako u definice Giniho indexu za nezávislé. Uvnitř dané skupiny nebo pro daný druh příjmu pak i za stejně rozdělené.

Kromě tohoto analytického postupu bychom mohli významnost faktorů zkoumat i jinými způsoby. Jak by se změnila nerovnost v populaci, kde by byl jednotlivý faktor eliminován (např. nějaká skupina by se nezapočítala do vzorku)? Kdyby daný faktor zůstal zachován, ale všechny ostatní bychom rozdělili rovnoměrně? Kdybychom daný faktor rozdělili rovnoměrně, ostatní bychom nezměnili? Jaký podíl má daný faktor na celkové nerovnoměrnosti?

4.1 Dekompozice podle podskupin

Zprvu si ukažme přirozený rozklad podle subpopulací. Uvažujme rozklad jednotek populace (jedinec, společná domácnost) do skupin podle zvolené charakteristiky. Nyní si zavedme označení pro charakteristiky skupin

- jednotlivé skupiny budeme indexovat čísly $k = 1 \dots K$ a výše příjmů jednotlivců ve skupině k jako X_{ik} , $i = 1 \dots N_k$
- N_k je kardinalita množiny k a bez rizika zmatení může označovat i skupinu samu
- μ_k označuj střední hodnotu rozdělení příjmů skupiny k
- $\phi_k = \frac{N_k}{N}$ označuj relativní početní zastoupení skupiny k v populaci
- $\psi_k = \frac{N_k \mu_k}{N \mu}$ označuj podíl skupiny k na příjmech společnosti

V práci [7] je popsán postup, jímž se index rozkládá na 3 části vyjadřující

1. celkovou nerovnost uvnitř skupin pomocí nerovností jednotlivých subpopulací¹⁷
2. nerovnost mezi skupinami¹⁸
3. efekt překryvu skupin

a celkově je Giniho index rozepsán jako

$$G = \sum_{k=1}^K a_k G_k + G_b + R$$

s doposud neznámými koeficienty a_k , jejichž vyjádření lze získat z následujícího postupu:

¹⁷ Při praktických úlohách je nutno očekávat, že většina z celkové hodnoty GI zůstává v této části. Tento fakt lze nahlížet tak, že skupiny jsou jen částečnou příčinou nerovnoměrnosti. Stejný výsledek bude mít i dekompozice nerovnoměrnosti relativních HDP v odstavci (4.2).

¹⁸ Právě vliv dělení populace je pevně spjat s roztržitostí společnosti. Tedy jedna část dekompozice Giniho indexu (resp. její podíl na celkové hodnotě) se bude vlastně projevovat jako míra polarizace.

$$\begin{aligned}
G &= (2N^2\mu_X)^{-1} \sum_{k=1}^K \left(\sum_{i=1}^{n_k} \sum_{j=1}^{n_k} |X_{ik} - X_{jk}| + \sum_{i=1}^{n_k} \sum_{j \notin N_k} |X_{ik} - X_j| \right) \\
&= (2N^2\mu_X)^{-1} \sum_{k=1}^K 2\mu_k n_k^2 G_k \\
&\quad + \frac{1}{2} \sum_{k=1}^K \sum_{h=1}^K \frac{N_k}{N} \frac{N_h}{N} \left| \frac{\mu_k - \mu_h}{\mu_X} \right| + R \\
&= \sum_k \phi_k \psi_k G_k + G_b + R \tag{21}
\end{aligned}$$

Zmíněné členy vyjadřují: G_k jsou Giniho indexy čistě jednotlivých skupin s koeficienty $\frac{N_k^2 \mu_k}{N^2 \mu}$, které jsou součiny relativních velikostí skupin ϕ_k a relativního podílu skupiny na příjmech společnosti ψ_k .

Pozorování *Koeficienty uvnitř sumy označované jako efekt uvnitř skupin se (vyjma patologicky zvolené struktury skupin) nenasčítávají na jedničku: $\sum \phi_k \psi_k < 1$.*

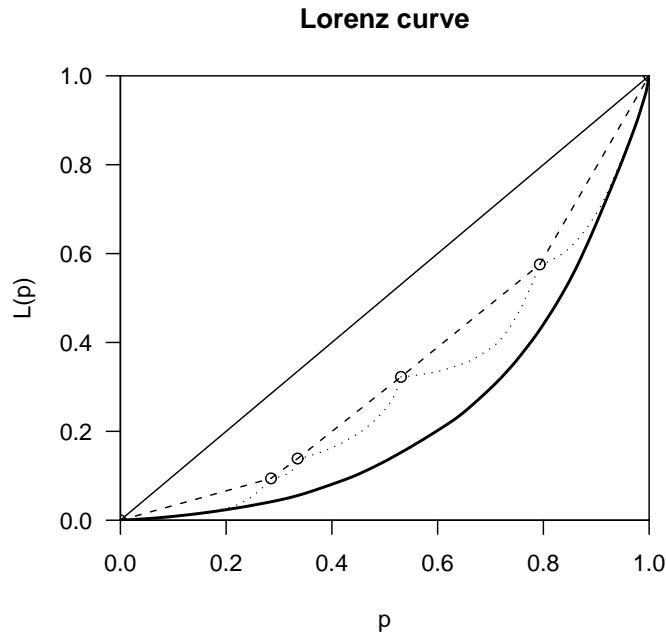
Důsledky této vlastnosti můžeme ilustrovat na situaci, kdy je populace rozdělena rovnoměrně (vzhledem k velikosti) na K_q částí (předpokládejme pro jednoduchost, že to lze), přičemž ve všech skupinách se příjmy řídí stejným rozdělením – tedy i Giniho index je všude stejný. Nicméně při rozkladu GI celé výchozí populace případně na efekt uvnitř skupin $G_w = \sum_{i=1}^{K_q} K_q^{-1} K_q^{-1} G = K_q^{-1} G$ celkové hodnoty GI populace. Jelikož střední hodnoty příjmů jsou pro všechny skupiny totožné, zbylá část $1 - \frac{K_q - 1}{K_q}$ – celá případně na reziduální člen R .

Krom toho takto definovaný meziskupinový člen z definice nezachycuje rozdílnost rozdělení příjmů mezi skupinami, ale je pouze jejím přiblížením přes očekávané hodnoty v jednotlivých skupinách.

Výraz G_b je nerovnoměrnost mezi skupinami¹⁹; vlastně odpovídá populaci složené z jedinců, jejichž příjem je roven příslušnému skupinovému průměru. Posledním členem rozkladu je R , který je třeba nahlížet jako součást potřebnou pro zachování rovnosti. Je spjat s efektem překryvu, spíše pak je vyjádřením blízkosti průměrů příjmů jednotlivých skupin. Pokud se skupiny nepřekrývají (ve smyslu hodnot x), pak je nulový. Naopak maxima dosahuje pokud jsou si průměry všech skupin rovny.

V uvedené publikaci je přístup motivován průběžnou změnou Lorenzovy křivky (od zjednodušené LC zahrnutím dalších efektů k pravdivé Lorenzově křivce). Úvodním krokem je spočítat G_b jako nerovnost v hypotetické populaci zkonstruované tak, jako bychom nahradili jednotlivé veličiny x_i degenerovaným rozdělením nabývajícím jen hodnoty průměru hodnot v příslušných podskupinách. Bez ztráty obecnosti má nový vektor \tilde{x} tvar $(\mu_1, \dots, \mu_1, \mu_2, \dots, \mu_2, \dots, \mu_K, \dots, \mu_K)$ a platí $\mu_1 \leq \mu_2 \leq \dots \leq \mu_K$. Takový stav by byl znázorněn na následujícím grafu

¹⁹všimněme si, že při dekompozici nerovnosti jedné a též populace podle různých dělení nezávisí tento člen jen na způsobu dělení, ale i na počtu tříd takového dělení. V patologickém případě, kdy každý jedinec definuje svoji vlastní třídu nebo v každé třídě mají všichni tentýž příjem, případně všechna nerovnoměrnost na tento člen.



Obrázek 4: Znázornění klasické dekompozice podle skupin

jako $L_B(q)$ (přerušovaně značená). Dále pak zavedeme nerovnost ve skupinách, avšak seřazení x_i nadále respektuje „soudržnost skupin“ – výsledkem je po částech monotonně seřazený vektor $x = (x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{K1}, \dots, x_{Kn_K})$. K tomu je příslušná po částech konvexní křivka $C(q)$ (tečkovaně značená; zlom nastává pokud je nejnižší příjem následující skupiny nižší než nejvyšší příjem skupiny předešlé).

Graf (4) znázorňuje zmíněnou dekompozici Giniho indexu v případě hrubého domácího produktu na obyvatele – datech podrobněji popsanych a použitých i v odstavci (4.2). Diagonála znázorňuje stav, v němž všichni jedinci mají tentýž příjem. Plocha $\int_0^1 [p - L_B(p)] dp = \frac{1}{2}G_b$ je spjata se situací, kdy by příslušníci každé jedné skupiny měli stejné příjmy (v tomto případě 5 skupin; příjmy mohou být mezi skupinami různé). Dále $\int_0^1 [L_B(p) - C(p)] dp = \frac{1}{2} \sum a_k G_k$ nám ukazuje příspěvek nerovnosti ve skupinách; následně zbývající $\int_0^1 [C(p) - L(p)] dp = \frac{1}{2}R$ je efekt překryvu skupin²⁰. Jak je možné nahlédnout ze struktury rozkladu na obrázku (4), nulovost této části je ekvivalentní se stavem, kdy nedochází k překryvu skupin. Názorně je tedy ukázáno, jak lze provést rozklad indexu na základě jejího geometrického významu.

Kolem reziduálního členu se rozpoutala široká diskuze, která vedla samozřejmě k různým pohledům na něj i celkově jiným návrhům rozkladu. Ty měly za účel

²⁰ z ekonomického hlediska tento člen vyjadřuje nestratifikovanost jednotlivých skupin (tj. překryv příjmových oblastí skupin)

přinést lépe interpretovatelné komponenty. Sastry a Kelkar navrhli v [14] upravený rozklad zachovávající váženou sumu nerovností uvnitř skupin. Zbylou část rozdělují způsobem, který je zřejmý z celé dekompozice. Použitím odhadu Giniho indexu podle kovariančního vyjádření (12) a rozpisem $x_{ij} - \bar{x} = x_{ij} - \bar{x}_i + \bar{x}_i - \bar{x}$ a $R_{ki} - R = R_{ki} - R_{i\cdot} + R_{i\cdot} - R$ docházíme k rozkladu ve tvaru

$$G = \sum_{k=1}^K \phi_k \psi_k G_k + 2(N\mu_X)^{-1} \sum_{k=1}^K N_k (\bar{x}_k - \bar{x}) (R_{k\cdot} - R) \\ + 2(N\bar{x})^{-1} \sum_{k=1}^K \sum_{i=1}^{N_k} (X_{ik} - \mu_k) (R_{ki} - \tilde{R}_{ki})$$

kde \tilde{R}_{ki} označuje pořadí veličiny X_{ki} vzhledem k ostatním pozorováním ve skupině k . Za použití průměrného pořadí jednotlivých n.v. $R = \frac{1}{2}(N+1)$, resp. průměrných pořadí jednotlivých skupin $R_{k\cdot} = N_k^{-1} \sum_{i=1}^{N_k} R_{ki}$ můžeme napsat rozklad ve tvaru

$$G = \sum_{k=1}^K \phi_k \psi_k G_k + \frac{2}{N\bar{x}} \sum_{k=1}^K \phi_k (\bar{x}_k - \bar{x}) (R_{k\cdot} - R)$$

Při porovnání s dříve představeným – nazývejme ho klasickým – rozkladem (21) si všímáme některých rozdílů. Zatímco původní způsob využíval meziskupinové nerovnosti, zde se objevuje jinak podložená veličina. Lze ji chápat jako míru zachycující společně efekt nerovnosti z hlediska průměrného příjmu skupiny a jejich stratifikovanosti pomocí rozdílu průměrného pořadí jedinců skupiny v celé populaci a celkového průměrného pořadí R , které zřejmě nabývá hodnoty $\frac{N+1}{2}$.

Využití relativního propadu jedince pro dekompozici Giniho indexu V kapitole 2 jsme si mezi jinými uvedli i rovnici vyjadřující GI pomocí relativního propadu příjmu jedince. Zavedeme-li si pro současné účely charakteristiky střední propad příjmu jedince i vůči členům jeho vlastní skupiny k :

$$\delta_i(\bar{k}) = N_k^{-1} \sum_{j \in N_k} (X_i - X_j)_+$$

a podobnou charakteristiku vzhledem k členům všech ostatních skupin jako

$$\delta'_i(k) = N^{-1} \sum_{j \notin N_k} (X_i - X_j)_+$$

tak vzhledem k možnému rozpisu středního propadu jedince (rovnice (13)) pomocí právě definovaných pojmů $\delta_{ik} = \phi_k \delta_i(\bar{k}) + \delta'_i(k)$ a dosazením do rovnice (14) můžeme Giniho index zapsat

$$G = (N\mu_X)^{-1} \sum_{k=1}^K \sum_{i=1}^{N_k} [\phi_k \delta_i(\bar{k}) + \delta'_i(k)] \\ = (N\mu_X)^{-1} \sum_{k=1}^K \phi_k N_k \mu_k G_k + (N\mu_X)^{-1} \sum_{k=1}^K \sum_{i=1}^{N_k} \delta'_i(k) \\ = \sum_{k=1}^K \phi_k \psi_k G_k + G_{gb}$$

Při tomto rozkladu vidíme, že první část je identická. Zbylá část Giniho indexu je zde zastoupena členem G_{gb} , který má evidentně také trochu jiný význam než byl doposud používaný meziskupinový člen. Hrubou meziskupinovou nerovnost jsme si představili jako očekávaný průměrný propad vůči jedincům ostatních skupin, měřený relativně vůči střední hodnotě příjmů.

Rozklad dle Daguma Stále se vyskytuje problém s interpretací členů (vyjma efektu nerovností uvnitř skupin) a špatná ospravedlnitelnost výše definovaného G_b jako míry nerovnoměrnosti. Tyto důvody pak vedly Daguma k návrhu jiné dekompozice. Pokud sumy pravé strany rovnosti (4) rozdělíme na členy „setrvávající“ ve skupinách a „jdoucí skrz hranice“, máme první krok dekompozice. Zaveďme si ještě pro dvojici množin zobecnění Giniho indexu – Giniho index mezi dvěmi množinami

$$G_{ij} = \left(2N_i N_j \frac{\mu_i + \mu_j}{2} \right)^{-1} \sum_{s=1}^{N_s} \sum_{t=1}^{N_t} |X_{is} - X_{jt}| \quad (22)$$

Pro volbu $i = j$ se evidentně shoduje s již známým Giniho indexem G_i . Použití nově definovaného meziskupinového Giniho indexu přináší

$$\begin{aligned} G &= \frac{1}{2N^2 \mu_X} \sum_{i=1}^N \sum_{j=1}^N |X_i - X_j| = G_w + G_{gb} \\ &= \frac{1}{2N^2 \mu_X} \left(\sum_{k=1}^K \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} |X_{ki} - X_{kj}| + 2 \sum_{k=2}^K \sum_{l=1}^{k-1} \sum_{i=1}^{N_k} \sum_{j=1}^{N_l} |X_{ki} - X_{lj}| \right) \\ &= \frac{1}{2N^2 \mu_X} \left(\sum_{k=1}^K 2\mu_k N_k^2 G_k + 2 \sum_{k=2}^K \sum_{l=1}^{k-1} \frac{(\mu_k + \mu_l)}{2} N_k N_l G_{kl} \right) \\ &= \sum_{k=1}^K \phi_k \psi_k G_k + \sum_{j=2}^K \sum_{k=1}^{j-1} G_{jk} [\phi_j \psi_k + \phi_k \psi_j] \end{aligned} \quad (23)$$

21

Pozorování Na začátku odstavce definovaný pojem meziskupinového Giniho indexu se jakožto rovnicí (22) definovaná statistika – stejně jako Giniho index pro populaci – nabývá hodnot rozmezí hodnot 0 a 1.

Nezápornost je zřejmá. Omezenost jedničkou snadno dokážeme použitím klasické nerovnosti $|a + b| \leq |a| + |b|$ na rozpis $|X_{is} - X_{jt}| = |X_{is} - \mu_i + \mu_i - X_{jt} + \mu_j - \mu_j|$. Pak se dostáváme po krátké úpravě ($|X_{is} - \mu_j| \geq 0$ a stejně pro skupinu j) k nerovnosti

$$\left(2N_i N_j \frac{\mu_i + \mu_j}{2} \right) G_{ij} \leq \sum_{s=1}^{N_i} \sum_{t=1}^{N_j} |\mu_i - \mu_j| = N_i N_j |\mu_i - \mu_j|$$

z čehož nakonec vzhledem k nezápornosti středních hodnot příjmů vyplývá

$$G_{ij} \leq \frac{|\mu_i - \mu_j|}{\mu_i + \mu_j} \leq 1$$

²¹ Měli bychom si uvědomit, že první člen zastupuje $\sum N_j^2$ prvků, zatímco druhý $\sum_{j=2}^k \sum_{i=1}^{j-1} N_j N_i$

První člen rozkladu vyjadřuje, stejně jako u předchozích typů rozkladů, jen vážený součet Giniho indexů jednotlivých skupin. Interpretace samotné hodnoty nezáporné (nezápornost lze nahlédnout např. z klasického rozkladu podle podskupin) části G_{gb} je však obtížnější, protože krom vlivu rozestupu průměrů v jednotlivých skupinách (byl představen na začátku kapitoly) zahrnuje i míru překryvu rozdělení příjmů jednotlivých skupin. Zmíněná forma rozkladu zřejmě nepřináší nic nového. Budeme tedy pokračovat v rozkladu této části pomocí koeficientů

$$D_{jk} = \frac{d_{jk} - p_{jk}}{d_{jk} + p_{jk}}$$

definovaných pomocí hrubého blahobytu $d_{jk} = \int_0^\infty \int_0^x (x - y) dF_k(y) dF_j(x)$ a transvariačního koeficientu prvního řádu $p_{jk} = \int_0^\infty \int_0^x (x - y) dF_j(y) dF_k(x)$. Rozdíl mezi definicemi spočívá v záměně pořadí intergrace dle distribučních funkcí příjmů v jednotlivých skupinách. Právě definované charakteristiky udávají v jaké míře jsou příjmy ve skupině j vyšší než ve skupině k (v uvedeném článku označován jako ekonomický blahobyt – *Relative Economic Affluence*), resp. p_{jk} udává nakolik jsou příjmy ve skupině j níže než příjmy ve skupině k . Veličina D_{jk} nabývá hodnot z intervalu $[0, 1]$ – minima za předpokladu, že rozdělení příjmů je v obou skupinách identické; naopak pokud rozdělení příjmů daných dvou skupin se překrývá na množině pravděpodobnosti 0 – pak je její hodnota 1 a v důsledku pak bude 100 % koeficientu nerovnosti mezi těmito dvěma skupinami přiřazeno k meziskupinové nerovnosti.

Veličinu G_{gb} (druhý člen pravé strany rovnice 23) rozdělíme jednoduchou úpravou na

$$G_{nb} = \sum_{j=2}^K \sum_{k=1}^{j-1} G_{jk} D_{jk} [\phi_j \psi_k + \phi_k \psi_j]$$

což lze nahlížet jako čistý příspěvek meziskupinové nerovnosti a příspěvek překryvu rozdělení mezi skupinami

$$G_t = \sum_{j=2}^K \sum_{k=1}^{j-1} G_{jk} (1 - D_{jk}) [\phi_j \psi_k + \phi_k \psi_j]$$

Dostáváme tak dle potřeby buď dvousložkový rozklad Giniho indexu nebo podrobnější tříslložkový.

Dekompozice indexu rovnoměrnosti Zvolme poněkud odlišný způsob a za základ nezvykle vezměme Giniho index *rovnoměrnosti*, což je doplněk k indexu nerovnoměrnosti: $G_e = 1 - G$. V zásadě nám tento přístup uvedený např. v [11] nezakrývá žádnou z informací, která by nám byla dostupná z rozkladu indexu nerovnoměrnosti.

Alternativně budeme používat i výběrové vyjádření

$$G_e = 1 - G = \frac{1}{2N^2\bar{x}} \left(\sum_{i=1}^N \sum_{j=1}^N (x_i + x_j) - \sum_{i=1}^N \sum_{j=1}^N |x_i - x_j| \right) \quad (24)$$

V případě rozkladu rovnoměrnosti pak lze pomocí Giniho indexů rovnoměrnosti uvnitř skupiny k a mezi skupinami k, l definovanými následujícím způsobem

$$G_{ek} = (2\bar{x}_k N_k^2)^{-1} \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} [(x_{ki} + x_{kj}) - |x_{ki} - x_{kj}|]$$

$$G_{ekl} = ((\bar{x}_k + \bar{x}_l) N_k N_l)^{-1} \sum_{i=1}^{N_k} \sum_{j=1}^{N_l} [(x_{ki} + x_{lj}) - |x_{ki} - x_{lj}|]$$

rozložit celkovou rovnoměrnost bez reziduálního členu ve smyslu $G_e = G_{ew} + G_{egb}$; podrobněji píšeme

$$G_e = \sum_{k=1}^K \phi_k \psi_k G_{ek} + \sum_{i=2}^K \sum_{j=1}^{i-1} (\phi_i \psi_j + \phi_j \psi_i) G_{eij} \quad (25)$$

Jak si ukážeme, jsou rozklady Giniho indexu rovnoměrnosti nadmnožinou rozkladu nerovnoměrnosti²² – zjistíme, že dekompozice Giniho indexu rovnoměrnosti při výpočtu obsahuje rozklad indexu nerovnoměrnosti. Porovnejme jednotlivé části rozkladu indexů ne- a rovnoměrnosti na část uvnitř a mezi skupinami:

$$G_{ew} = \sum_{k=1}^K \phi_k \psi_k G_{ek} \quad G_w = \sum_{k=1}^K \phi_k \psi_k G_{kk}$$

$$G_{egb} = \sum_{i=2}^K \sum_{j=1}^{i-1} (\phi_i \psi_j + \phi_j \psi_i) G_{eij} \quad G_{gb} = \sum_{i=2}^K \sum_{j=1}^{i-1} (\phi_i \psi_j + \phi_j \psi_i) G_{ij}$$

Lerman-Yitzhakiho dekompozice²³ Využitím kovariančního ekvivalentu v zápisu Giniho indexu $-G = 2\mu_X^{-1} \text{cov}(X, F_X(X))$ – přinesli Lerman a Yitzhaki v [15] možnost jiného způsobu rozkladu Giniho indexu. Vychází z rozkladu distribuční funkce příjmů na váženou sumu distribučních funkcí jednotlivých skupin – váhy odpovídají zastoupení skupin v populaci. Předpokládáme, že příjmy ve skupinách se řídí různými rozděleními s distribučními funkcemi F_k a distribuční funkci F celé populace lze poté zapsat jako součet těchto funkcí s váhami danými relativními velikostmi skupin: $F(y) = \sum_{k=1}^K \phi_k F_k(y)$. Při fixaci skupiny k a definici $F_{ni} = -F_i + \sum_{k=1}^K \phi_k F_k$ distribuční funkce příjmů mimo skupinu i můžeme psát

$$F(y) = \phi_i F_i(y) + (1 - \phi_i) F_{ni}(y)$$

$$= F_i(y) + (1 - \phi_i) (F_{ni}(y) - F_i(y)) \quad (26)$$

²² V tom smyslu, že při výpočtu hodnot rovnoměrnosti se vyskytuje nerovnost adekvátní sekce populace

²³ lze se setkat s modifikovanou variantou tohoto rozkladu nazývanou Analysis of Gini; čímž se prezentuje myšlenková podobnost postupu s ANOVA (Analysis Of Variance).

Zavedme pro náhodnou veličinu řídící rozdělení příjmů ve skupině k označení $X^{[k]}$. Předchozím rozpisem Lerman a Yitzhaki dochází k vyjádření kovariance

$$\begin{aligned} cov(X, F_X) &= \sum_{k=1}^K \phi_k cov(X^{[k]}, F_k) \\ &+ \sum_{k=1}^K (\phi_k - 1) \phi_k cov(X^{[k]}, F_k - F_{nk}) \\ &+ cov(\mu_i, EF^{-1}(X^{[i]})) \end{aligned}$$

Poslední člen je kovariancí středních hodnot příjmů v jednotlivých skupinách a středních zařazení příjmů z dané skupiny v celé populaci. Finálně pak kovarianční definicí Giniho indexu (12) získávají rozklad podle skupin ve tvaru

$$\begin{aligned} G &= \sum_{k=1}^K \psi_k G_k \\ &+ \sum_{k=1}^K \psi_k G_k (\phi_k - 1) \frac{cov(X^{[k]}, F_k - F_{nk})}{cov(X^{[k]}, F_k)} \\ &2 \frac{cov(\mu_k, EF^{-1}(X^{[i]}))}{\mu} \end{aligned} \quad (27)$$

Hlavní myšlenka spočívá v rozdělení celkového efektu nerovnosti uvnitř skupin do dvou částí, tak jak je zřejmé z uvedené rovnice. Poslední člen je založen na kovarianci středních příjmů jednotlivých skupin a středních relativních pořadí náhodné veličiny $X^{[k]}$ v celé populaci a zastupuje tak meziskupinový efekt na nerovnost.

Všimněme si zlomku ve druhé sumě: ten může nabývat hodnot od -1 do 1 a je vyjádřením stratifikovanosti skupiny i . Tento člen by měl být obvykle pozitivní, což v kombinaci s nutnou záporností členu $\phi_k - 1$ znamená, že celkově bude druhá suma obvykle záporná.

Věnujme nyní trochu pozornosti výrazu $cov(X^{[i]}, F_i - F_{ni})$. Jedná se o kovarianci příjmů členů skupiny i s rozdílem $F_i - F_{ni}$, což je odlišnost relativního pořadí jednotlivce vzhledem ke své skupině i a vzhledem ke zbylé populaci (populace bez jedinců skupiny do níž pozorování náleží). V této části je zachycena stratifikovanost.

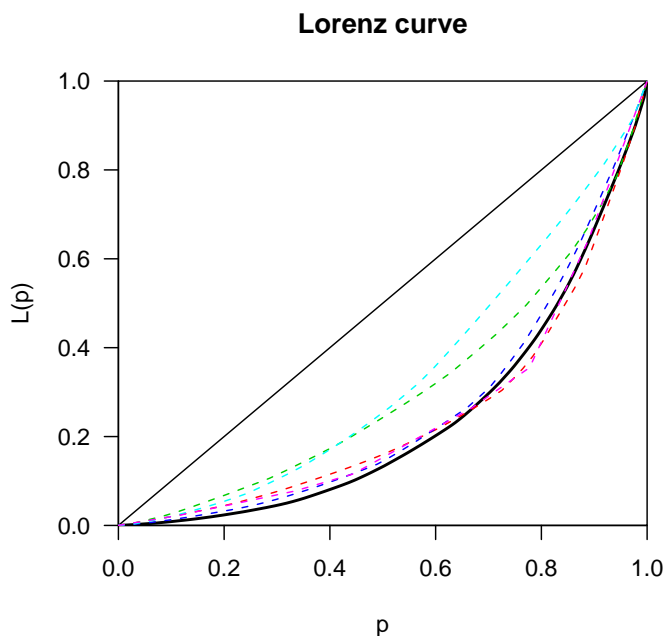
4.2 Rozklad indexu světové nerovnoměrnosti v hrubém domácím produktu

Giniho index nemusíme používat jen pro nerovnost mezi příjmy jednotlivců. Tato statistika je široce využitelná pro různé druhy dat. V tomto odstavci použijí data Mezinárodního měnového fondu²⁴ a pomocí Lerman-Yitzhakiho metody se pokusíme vhlédnout do struktury nerovnosti v hrubém domácím produktu na jednoho obyvatele.

²⁴IMF World Economic Outlook Database, April 2006

Popis dat Data jsou pro jednotlivé státy – členy IMF – a skupiny zvolíme dle kontinentů. Použijeme HDP na obyvatele přepočtené metodou *Purchasing Power Parity*²⁵. Tento ukazatel má oproti prostému nominálnímu vyjádření lepší vypovídací hodnotu v ohledu na životní úroveň jedinců v dané zemi. Jsou to tedy data, která nám ukáží nerovnoměrnost v relativním výkonu jednotlivých ekonomik a v důsledku jsou dobrou aproximací životní úrovně obyvatel.

Pro dokreslení některé výběrové statistiky (v jednotkách PPP \$): minimum 596, maximum 69 800, průměr 10 801, medián 6 681. Směrodatná odchylka 11 352. Výrazná je hodnota maxima, která přísluší Lucembursku; hranici 40 000 překročili ještě tři státy: Irsko (40 610), USA (41 399) a Norsko (42 364). V případě Lucemburska se jedná o významně odlehle pozorování; nicméně vliv jeho vyloučení na Giniho index celé skupiny je pouze na úrovni jednoho procentního bodu a směrodatná odchylka s jeho vyloučením 10 480.



Obrázek 5: Lorenzova křivka hrubého domácího produktu pro celý svět (plně černě) a jednotlivé kontinenty (Afrika ●, Amerika ●, Evropa ●, Asie ●, Oceánie ●)

Použitá data pochází z roku 2005 a celkový Giniho index členů fondu (celkem 179 dostupných hodnot) spočteme $G_{HDP} = 53,17\%$ ²⁶. Graf 5 ukazuje Lorenzovu křivku celé množiny a její případy pro jednotlivé kontinenty.

Rozklad nerovnoměrnosti ve výši hrubého domácího produktu S rozkladem těchto dat jsme se již setkali v grafu 4, která ilustrovala klasickou dekom-

²⁵Parita kupní síly – převod nominální hodnoty dolarového ukazatele podle cenové úrovně dané země

²⁶pro srovnání v roce 2004 na hodnotě 53,034 %, v roce 2003 měl hodnotu 53,219 %

pozici Giniho indexu. V tomto odstavci se ale spočteme členy Lerman-Yitzhakiho rozkladu (rovnice (27)). Tabulka 2 přehledně shrnuje dekompozici.

Její první část je svou hodnotou

$$\sum_{k=1}^5 \psi_k G_k = 40,855 \%$$

nejvýraznějším z příspěvků²⁷. Jak vidíme, druhý člen dekompozice je záporný a v absolutní hodnotě ze všech tří nejmenší. Celkově je z dekompozice vidět, že nejvýrazněji se celkovém Giniho indexu projevuje vnitřní nerovnoměrnost skupin, zhruba poloviční efekt má meziskupinová nerovnost. Důvody můžeme hledat, jak je vidět v tabulce, v kombinaci relativní blízkosti průměrů Asie, Ameriky a Oceánie na straně jedné a ne úplně kompaktně sdruženými hodnotami průměrných relativních pořadí, obzvláště Oceánie, na straně druhé. Mírné snížení pochází ze malé stratifikovanosti skupin – dochází mezi nimi k překryvu hodnot. Důvod nevýraznosti tohoto prvku lze tušit ve výrazném odloučení Afriky a Oceánie od zbylých kontinentů.

Z celkového pohledu stojí za povšimnutí, že Evropa svými příspěvky hodně ovlivňuje celkovou dekompozici.

4.3 Dekompozice podle zdrojů příjmů

Důvody nerovnosti mezi příjmy populace mohou být i jiné než příslušnost jedince k určité skupině. Lze je nahlížet i tak, že nerovnost je způsobena nerovností mezi složkami příjmů jedinců. Účelem dekompozice podle zdrojů příjmů je zjištění vlivu jednotlivých složek na nerovnost uvnitř populace a tak například rozpoznat míru významu různých sociálních dávek nebo jiných benefitů (a nakolik dobře splňují svůj úkol snižovat nerovnost).

Uvažujme tedy, že příjem jednotlivce je složen z více částí (plat, kapitálové příjmy, sociální příspěvky apod.). Přičemž tyto kategorie jsou samozřejmě jednotně dány pro celou populaci; tedy celkový příjem jedince rozepisujeme jako součet všech příjmů:

$$X_i = \sum_{m=1}^M X_i^m$$

kde index m používáme pro označení typů příjmů. Zavedme v této podkapitole úmluvu

- střední hodnota m . komponenty bude značena μ_m , její výběrový průměr \bar{x}_m
- podíl druhu příjmu m na celkovém příjmu populace značíme $s_m = \frac{\mu_m}{\mu_X}$

²⁷ Graficky je znázorněna i v grafu 4. Celková Lorenzova křivka pro tyto údaje je znázorněna tlustou čarou a zmíněný příspěvek je dvojnásobkem plochy mezi slabou čarou a přímkou $x = y$. Další příspěvky si bohužel takto znázornit nemůžeme. Krom toho Lerman-Yitzhakiho dekompozice neumožňuje přímé porovnání výraznosti jednotlivých efektů jako v případě metody představené na začátku kapitoly.

		Afrika	Amerika	Evropa	Asie	Oceánie	Σ
počet států (podíl)	ϕ_g	51 (28,5 %)	35 (19,6 %)	47 (26,3 %)	37 (20,7 %)	9 (5 %)	179
Gini kontinentu	G_g	0,5150323	0,3898	0,4944019	0,3314408	0,5078476	
podíl HDP kontinentu	ψ_g	0,09407998	0,18383109	0,25301399	0,42441671	0,04465822	
efekt skupin	$G_g\psi_g$	0,04845423	0,07165735	0,12509060	0,14066901	0,02267957	0,4085508
	$\psi_i G_i (\phi_i - 1)$	-0,03464883	-0,05764614	-0,09224558	-0,11159218	-0,02153926	
	$\frac{cov_i(y_i, F_i - F_{ni})}{cov_i(y_i, F_i)}$	0,005481201	0,021371246	0,010417965	0,059182936	0,002931954	
							-0,0993853
průměrné HDP (v PPP\$)	\bar{x}_g	3566,366	10154,294	10407,478	22176,326	9593,073	
relativně k celkovému průměru	\bar{x}_g/\bar{x}	0,3302023	0,9401647	0,9636065	2,0532592	0,888202	
průměrné pořadí států (relativní R_k)		46,72549 (0,255)	100,82857 (0,673)	92 (0,514)	139 (0,777)	86,88889 (0,485)	
X	$2 \frac{cov(\bar{x}_k, R_k)}{\bar{x}}$						0,2225702
							0,5317357

Tabulka 2: Giniho index hrubého národního produktu (HDP v jednotkách parity kupní síly na 1 obyvatele) a jeho dekompozice Lemman-Yitzhakého metodou.

Dekompozice indexu rovnoměrnosti Stejně jako v případě rozkladu podle podskupin, i zde můžeme využít přechodu k indexu rovnoměrnosti²⁸. Ukážeme si, že lze provést rozklad na příspěvky jednotlivých zdrojů k celkové rovnoměrnosti ve tvaru $G_e = \sum_{i=1}^m V_e^m$, kde V_e^m je příspěvek m . zdroje příjmů k celkové míře rovnoměrnosti. Abychom dokázali uvedenou rovnost, zavedeme si nový operátor

$$X_{ij}^{*m} = \begin{cases} X_i^m & X_i < X_j \\ X_j^m & X_i \geq X_j \end{cases} \quad (28)$$

Jedná se o složku m menšího z dvojice příjmů X_i a X_j – to znamená, že může nastat případ $X_i^m > X_j^m$, ačkoliv $X_j > X_i$. Následně si pro rozklad indexu rovnoměrnosti (rovnice (24)) vypůjčíme vztah $G_e \equiv 1 - G$, právě zavedený operátor (28) a rovnost $|a - b| = a + b - 2 \min(a, b)$. Můžeme tedy postupovat následovně pro rozklad statistiky G_e

$$\begin{aligned} G_e &= 1 - G = \frac{\sum_i \sum_j \sum_{m=1}^M (x_i^m + x_j^m)}{2\bar{x}N^2} - \frac{\sum_i \sum_j \left| \sum_{m=1}^M (x_i^m - x_j^m) \right|}{2\bar{x}N^2} \\ &= (2\bar{x}N^2)^{-1} \sum_{i=1}^N \sum_{j=1}^N 2 \min \left(\sum_{m=1}^M x_i^m, \sum_{m=1}^M x_j^m \right) \\ &= (2\bar{x}N^2)^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{m=1}^M 2x_{ij}^{*m} \\ &= \sum_{m=1}^M (2\bar{x}N^2)^{-1} \sum_{i=1}^N \sum_{j=1}^N 2x_{ij}^{*m} \end{aligned}$$

Z tohoto zápisu změnou pořadí sumace obdržíme rozklad na efekty jednotlivých zdrojů

$$G_e = \sum_{m=1}^M \frac{\sum_i \sum_j 2x_{ij}^{*m}}{2N^2\bar{x}} = \sum_{m=1}^M V_{em} \quad (29)$$

Z daného je vidět, že při rozkladu Giniho indexu rovnoměrnosti nás přivádí k jinému druhu dekompozice. Získáme tak rozklad $G_e = 1 - G$ na příspěvky jednotlivých zdrojů příjmů k celkové rovnoměrnosti.

Lerman-Yitzhakiho rozklad Jak bylo ukázáno v kapitole (2), existuje pro vyjádření GI několik ekvivalentních přístupů – stejně tak i pro jeho rozklad podle zdrojů příjmů vycházejme nyní z kovarianční interpretace. Lerman a Yitzhaki (autoři jednoho ze zmíněných rozkladů podle skupin) v [8] využili znovu kovarianční význam Giniho indexu. Budeme-li stále uvažovat celkový příjem jako prostý součet jeho komponent, pak přepíšeme rovnici (12) jako

$$G = 2\mu_X^{-1} \text{cov} \left(\sum_{m=1}^M X^m, F_X(X) \right)$$

a využitím vlastností kovariance a drobnými úpravami

²⁸ představený postup lze najít v Mussard [10]

$$\begin{aligned}
G &= 2\mu_X^{-1} \sum_{m=1}^M \text{cov}(X^m, F_X(X)) \\
&= \sum_{m=1}^M \frac{\text{cov}(X^m, F_X(X))}{\text{cov}(X^m, F_{X^m}(X^m))} \times \frac{2\text{cov}(X^m, F_{X^m}(Y_m))}{\mu_m} \times \frac{\mu_m}{\mu_X} \\
&= \sum_{m=1}^M G_{\text{corr}_m} \times G^m \times s_m
\end{aligned}$$

kde zlomek vzniklý rozšířením výrazu uvnitř sumy (za předpokladu nenulovosti G^m) je rozšířením pojmu korelace (tzv. Giniho korelace složky i a celkového příjmu), který lze nahlížet jako míru rozlišnosti uspořádání hodnot celkového příjmu jedince na straně jedné a uspořádání hodnot typu zdroje j na straně druhé.

Při této dekompozici používáme pouze členy přímo spjaté s charakteristikami jednotlivých příjmových komponent. Díky faktu, že při rozkladu podle zdrojů příjmů, kdy je při této separaci definována hodnota pro každého jedince, v důsledku znamená absenci reziduálního členu, který se vyskytoval při rozkladu dle skupin.

Na závěr podkapitoly si všimněme odlišnosti rozkladu indexu rovnoměrnosti a Lerman-Yitzhakiho rozkladu indexu nerovnoměrnosti. Ačkoliv je přechod mezi indexy ne- a rovnoměrnosti přímočarý a jednoduchý, celkově se oba rozkladové výsledky liší v interpretační rovině. Zatímco rozklad indexu rovnoměrnosti zahrnuje pouze příspěvky jednotlivých zdrojů příjmů Lerman-Yitzhakiho přístupu je navíc zakomponována i míra podobnosti rozdělení jednotlivého zdroje s rozdělením celých příjmů.

4.4 Modelování příjmů lineárním modelem a souvislost pro rozklad Giniho indexu

Dosud zmíněné druhy dekompozice poskytují cenné informace o struktuře nerovnosti. Z jejich výstupů lze například ve formě dvourozměrné tabulky zjistit příspěvky jednotlivých dvojic „podskupina-druh příjmu“ nebo „podskupina-podskupina jiného dělení“ k celkovému GI. Na základě těchto údajů lze následně interferovat o důvodech nerovnosti v celé populaci.

Často se příjmy modelují jako lineární kombinace charakteristik daného jednotlivce (takovými charakteristikami mohou být délka vzdělání, pohlaví, věk apod.). Získáme tak – za splnění předpokladů – statisticky lépe podložené výsledky analýzy.

Využijme ekonometrického modelu příjmu jedince ve formě lineárního modelu²⁹ s jeho výhodami pomocí vysvětlujících charakteristik jako

$$X_i = \alpha_0 + \sum_{m=1}^M \alpha_m X_i^m + \epsilon_i \quad (30)$$

veličiny X^m jsou atributy, které by podle našich domněnek měly mít vliv na velikost příjmu a tím pádem být příčinou nerovnosti.

²⁹Zde nejsme a priori omezení na volbu lineárního modelu. V zásadě si můžeme vybrat jakýkoliv dobrý model, pro který známe chování jeho koeficientů.

Limitování jsme našimi schopnostmi najít pravdivý model tak, abychom neztráceli informace v chybových členech.

Vedle toho zahrnuje rovnice konstantní člen α_0 a náhodný člen ϵ_i , který v podstatě zastupuje neznámé efekty. V následujícím bude vlastně chápán jako $(m+1)$. vstup. Odhady parametry modelu (30) mohou sloužit pro porovnání významnosti jednotlivých druhů příjmů.

Dále předpokládáme, že máme populaci rozdělenou na skupiny (ženy/muži, velkoměsta/města/malé obce apod.). Tímto přístupem budeme moci nahlédnout, zda je nerovnost způsobena čistě nějakou charakteristikou nebo zda se jedná o nerovnost vzniklou rozdílnou strukturou skupin³⁰. Pokud bychom totiž dělali závěry na základě hodnot meziskupinových nerovností, mohli bychom se dopustit chyb. Meritum nerovnosti mezi skupinami, jak jsme se už zmínili, totiž nemusí být ve faktické nerovnosti mezi skupinami, ale v rozdílné struktuře jednotlivých skupin. Tato parametrizace příjmu nám dovolí odhadnout relativní příspěvky jednotlivých atributů na celkové rovnoměrnosti ve skupině k (pro atribut m značeno C_{ek}^m) a mezi skupinami (C_{ekh}^m označuje vliv atributu m na rovnoměrnost mezi skupinami k a h), tak jak je odvozeno v [10]

$$C_{ek}^m = \hat{\alpha}_m \frac{2}{G_e 2\mu_X N^2} \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} X_{ki,kj}^{*m}$$

$$C_{ekh}^m = \hat{\alpha}_m \frac{2}{G_e 2\mu_X N^2} \sum_{i=1}^{N_k} \sum_{j=1}^{N_h} 2X_{ki,hj}^{*m}$$

Kromě samotných příspěvků jednotlivých zdrojů příjmů k rovnoměrnosti uvnitř skupiny a rovnoměrnosti mezi skupinou k a h odvozuje i odhady směrodatných odhylek těchto veličin jako

$$\sigma_{ek}^m \equiv \sigma_{C_{ew}^m} = \sigma_{\hat{\alpha}_m} \frac{1}{G_e} \frac{1}{2N^2\mu_X} \left[\sum_{i=1}^{N_k} \sum_{j=1}^{N_k} 2X_{ki,kj}^{m*} \right]$$

$$\sigma_{ekh}^m \equiv \sigma_{C_{egb}^m} = \sigma_{\hat{\alpha}_m} \frac{1}{G_e} \frac{1}{2N^2\mu_X} \left[2 \sum_{i=1}^{N_k} \sum_{j=1}^{N_h} 2X_{ki,hj}^{m*} \right]$$

4.5 Vícerozměrná dekompozice³¹

Zatím jsme se zabývali rozkladem podle jediného atributu. Pro některé analýzy mohou být doposud představené způsoby rozkladu nerovnoměrnosti omezující a nedostatečně detailní. V této podkapitole si tedy ukážeme, jak lze hlouběji vhlédnout do struktury nerovnoměrnosti.

Prvotní ideou se může stát sloučení obou (nebo i více) dimenzí dělení a jejich kartézským součinem získat nové jednorozměrné dělení. Nevýhodou se může

³⁰ Při dělení populace a následné analýze rozdílnosti rozdělení příjmů je zpravidla od počátku zřejmé, že difference nevznikají pouze z důvodů příslušnosti k určité skupině, ale také rozdílností vnitřních struktur skupin. Například při dělení podle velikosti sídla v němž jedinci žijí se bude projevat vyšší úroveň kvalifikace v určité skupině vyššími příjmy. Tedy rozdílnost pochází už z průměrné vzdělanosti v oblasti.

³¹ Při použití vícerozměrné dekompozice (ve smyslu, že jedinec je charakterizován příslušností k několika subpopulacím (např. muž/žena \times venkov/městská oblast) získáváme lepší náhled na nerovnost mezi skupinami. Důvodem je fakt, že u jednorozměrné dekompozice velká část z indexu spadne do efektu nerovnosti uvnitř skupin, který obsahuje nevysvětlenou část – nerovnosti uvnitř skupin i nerovnosti skrz jiné dimenze dělení.

ukázat vysoký (a případným zjemňováním jednoho z dělení rychle rostoucí) počet skupin a následně zhoršená možnost interpretace vlivu jednotlivých atributů, protože ten je potřeba zkoumat simultánně v několika skupinách do nichž se tímto postupem „rozpusť“. Obecně totiž platí, že zvyšování počtu skupin vede k nárůstu podílu nerovnosti mezi skupinami na úkor nerovnosti uvnitř skupin. Tedy např. schéma, kdy bychom provedli rozklady podle jednotlivých dimenzí a následně se snažili o porovnání vzhledem k obsáhlému (ve smyslu množství podskupin) „součinnému“ dělení, dopouštěli bychom se nepřesností a s velkým rizikem následně i mylných závěrů.

Předpokládejme, že druhý typ dělení má celkem Q tříd. Podíl průniku skupiny j druhého dělení a skupiny k dělení prvního na celkové velikosti (celkových příjmech) skupiny k budeme značit ϕ_j^k (resp. ψ_j^k)³². Rozklad podle více druhů dělení populace založíme na již představeném Dagumově rozkladu (viz strana 32) Giniho indexu na dvě části. Rozšíření o jeden rozměr rozkladu bude mít za důsledek, že Giniho indexy skupin prvního dělení G_k rozložíme ve stejném smyslu podle skupin druhého dělení. Zavedeme-li značení $G_{k,h}$ pro Giniho index jednotlivců patřících do skupiny k dělení prvního a zároveň do skupiny h dělení druhého, získáváme první část dekompozice, která vyjadřuje efekt nerovností uvnitř skupin vzniklých dvojstupňovým dělením

$$G_{wd} = \sum_{k=1}^K \phi_k \psi_k \sum_{j=1}^Q G_{k,j} \phi_j^k \psi_j^k$$

Druhá část nerovnoměrnosti uvnitř skupin prvního dělení vzniklá aplikací druhého dělení ($G_{k,ij}$ značí meziskupinový Giniho index mezi dvěma podmožinami prvního dělení)

$$G_{wb} = \sum_{k=1}^K \phi_k \psi_k \sum_{i=2}^Q \sum_{j=1}^{i-1} G_{k,ij} (\phi_j^k \psi_i^k + \phi_i^k \psi_j^k)$$

Třetí část zůstává v původním tvaru zohledňujícím pouze první dělení

$$G_b = \sum_{k=2}^K \sum_{l=1}^{k-2} G_{kl} (\phi_k \psi_l + \phi_l \psi_k)$$

Celkový Giniho index tak můžeme zapsat ve tvaru $G = G_{wd} + G_{wb} + G_b$ a vliv dodatečného druhého dělení je tak vyjádřen poměrem G_{wd} a G_{wb} . Pokud druhé dělení dobře vystihuje důvody nerovnosti příjmů, pak zřejmě G_{wb} bude mít tendenci nabývat větších hodnotu na úkor G_{wd} .

5 Dodatky

Odvození $G = 1 + \mu_X^{-1} a - \frac{1}{\mu_X} \int (1 - F(y))^2 dy$. Předpokládáme spojitou náhodnou veličinu definovanou na intervalu \mathbb{R}_0^+ s počátkem v $a \geq 0$ a koncem $b < \infty$. Využijeme rovnost $G = 1 - \mu_X^{-1} E \min(X_i, X_j)$, která byla již v textu prezentována na straně 7.

³²tedy $\forall k$ bude platit $\sum_{j=1}^Q \phi_j^k = \sum_{j=1}^Q \psi_j^k = 1$

Nyní si připomeneme obecně užívaný fakt

$$\begin{aligned} F_{\min(X_i, X_j)}(y) &= P(\min(X_i, X_j) \leq y) \\ &= 1 - P(X_i > y)P(X_j > y) \\ &= 1 - [1 - F(y)]^2 \end{aligned}$$

který použijeme v následujícím postupu

$$\begin{aligned} G &= 1 - \mu_X^{-1} \int_a^b y dF_{\min(X_i, X_j)}(y) \\ &= 1 + \mu_X^{-1} \int_a^b y d[1 - F(y)]^2 \\ &\quad \left[\begin{array}{l} u = y \quad dv = d[1 - F(y)]^2 \\ du = 1 \quad v = [1 - F(y)]^2 \end{array} \right] \\ &= 1 + \mu_X^{-1} \left[y[1 - F(y)]^2 \right]_a^b - \mu_X^{-1} \int_a^b [1 - F(y)]^2 dy \\ &= 1 + \mu_X^{-1} a - \mu_X^{-1} \int_0^b [1 - F(y)]^2 dy \end{aligned}$$

Integrací *per partes* jsme obdrželi dva určité integrály, z nichž první zřejmě bude nabývat hodnoty $\mu_X^{-1}a$ a zbývající členy odpovídají tomu, co jsme chtěli dokázat. Pro náhodných veličin nabývajících hodnot z $[0, b]$ se výsledek triviální úpravou zjednodušuje na

$$G = 1 - \mu_X^{-1} \int_0^b [1 - F(y)]^2 dy$$

Odvození Lorenzovy křivky a Giniho indexu pro Paretovo rozdělení

Vyjdeme z definice $L(p) = \mu^{-1} \int_0^p F^{-1}(x) dx$ a střední hodnoty Paretova rozdělení (viz rovnice (18)) $EX = \frac{km}{k-1}$. Potřebujeme si ještě vyjádřit funkci inverzní k distribuční funkci Paretova rozdělení $F(x) = 1 - \left(\frac{m}{x}\right)^k$. Inverzní funkce, jak si lze dosazením ověřit, má tvar $F^{-1}(q) = m(1 - q)^{-1/k}$. Následně si parametricky vyjádříme Lorenzovu křivku ($p \in [0, 1]$)

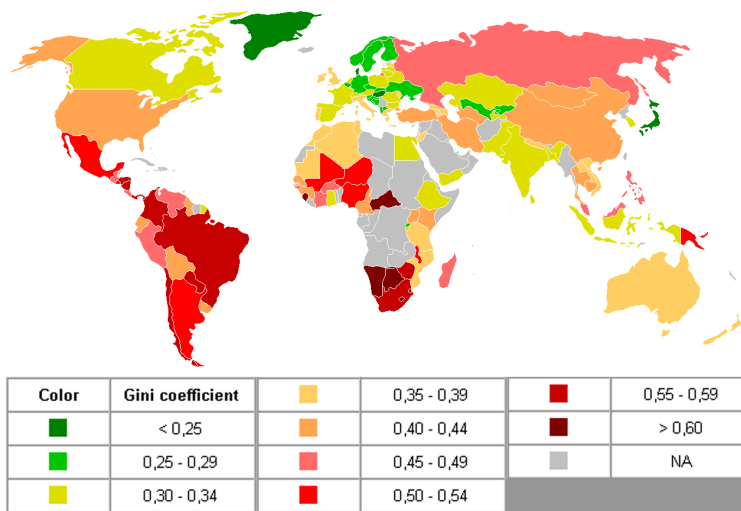
$$\begin{aligned} L(p) &= \frac{k-1}{km} \int_0^p m(1-q)^{-1/k} dq \\ &= -\frac{k-1}{k} \int_1^{1-p} s^{-1/k} ds = \frac{k-1}{k} \int_{1-p}^1 s^{-1/k} ds \\ &= \frac{k-1}{k} \left[\frac{k}{k-1} s^{\frac{k-1}{k}} \right]_{s=1-p}^1 = \left[s^{\frac{k-1}{k}} \right]_{s=1-p}^1 \\ &= 1 - (1-p)^{\frac{k-1}{k}} \end{aligned}$$

Pro vztah Giniho indexu k parametru k (jak se ukáže, parametr m nemá vliv)

poslouží následující výpočet založený na vyjádření Giniho indexu(10)

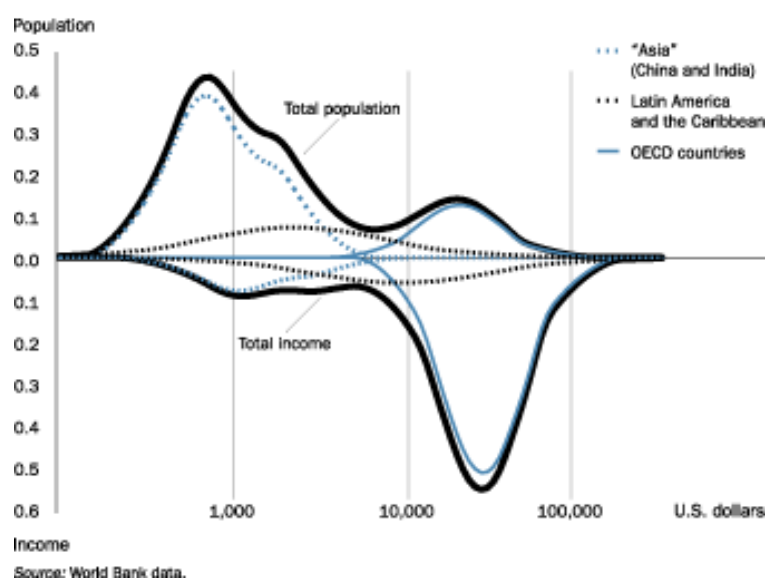
$$\begin{aligned}
 G &= 2 \int_0^1 p - L(p) dp = 2 \int_0^1 p - 1 + (1-p)^{\frac{k-1}{k}} \\
 &= 2 \int_0^1 p - 1 dp + 2 \int_0^1 (1-p)^{\frac{k-1}{k}} \\
 &= -1 + 2 \int_0^1 s^{\frac{k-1}{k}} ds = -1 + 2 \left[\frac{k}{2k-1} s^{\frac{2k-1}{k}} \right]_{s=0}^1 \\
 &= -1 + 2 \frac{k}{2k-1} \\
 &= \frac{1}{2k-1}
 \end{aligned} \tag{31}$$

Grafické přílohy Pro ilustraci úrovně nerovnoměrnosti na světě je přikloženo jeho znázornění v atlasu, resp. graf reálného rozdělení příjmů ve světové populaci.



Obrázek 6: Giniho index nerovnoměrnosti ve světě.
zdroj Wikipedia

FIGURE 1
International income distribution, 1993



Obrázek 7: Rozdělení příjmů ve světové populaci (stav roku 1993) dle World Bank.

Reference

- [1] Shlomo Yitzhaki and Edna Schechtman. The properties of the extended gini measures of variability and inequality. *Metron - International Journal of Statistics*, 0(3):401–433, 2005.
- [2] Z. M. Berrebi and Jacques G. Silber. Dispersion, asymmetry and the gini index of inequality. *International Economic Review*, 28(2):331–38, June 1987.
- [3] Kwang Soo Cheong. A note on the interpretation and application of the gini coefficient. Technical report, 1999.
- [4] Joseph L. Gastwirth. The estimation of the lorenz curve and gini index. *The Review of Economics and Statistics*, 54(3):306–16, 1972.
- [5] David E. A. Giles. Calculating a standard error for the gini coefficient: Some further results. Econometrics working papers, Department of Economics, University of Victoria, April 2002.
- [6] G. M. Giorgi, T. Palmitesta, and C. Provasi. Asymptotic and bootstrap inference for the generalized gini indices. *Metron*, 64(1), April 2006.
- [7] J. Richard Lambert, Peter J. Aronson. Inequality decomposition analysis and the gini coefficient revisited. *Economic Journal*, 103(420):1221–27, 1993.
- [8] Shlomo Lerman, Robert I. Yitzhaki. Income inequality effects by income. *The Review of Economics and Statistics*, 67(1):151–56, February 1985.
- [9] Farhad Mehran. Bounds of the gini index based on observed points of the lorenz curve. *Journal of the American Statistical Association*, 70(349):64–66, March 1975.
- [10] Stephane Mussard. On decomposition of the gini index of equality. Technical report, 2005.
- [11] Stephane Mussard, Pi Alperin Maria Noel, Francoise Seyte, and Michel Terraza. Extensions of dagums gini decomposition. Technical report, 2005.
- [12] U. S. Nair. The standard error of gini’s mean difference. *Biometrika*, 28(3-4):428–436.
- [13] Graham Pyatt. On the interpretation and disaggregation of gini coefficients. *Economic Journal*, 86(342):243–55, 1976.
- [14] Ujwala R. Sastry, D. V. S. Kelkar. Note on the decomposition of gini inequality. *The Review of Economics and Statistics*, 76(3):584–86, August 1994.
- [15] Robert I. Yitzhaki, Shlomo Lerman. Income stratification and income inequality. *Review of Income and Wealth*, 37(3):313–29, September 1991.
- [16] Buhong Zheng. Aggregate poverty measures. *Journal of Economic Surveys*, 11(2):123–62, 1997.