

Statistische Modellierung III

-Propensity Score-

Dr. Martin Scharpenberg

MSc Medical Biometry/Biostatistics

WiSe 2019/2020

Hintergrund

- Der Behandlungseffekt neuer Therapien sollte in randomisierten Studien geprüft werden
- Nur durch die Randomisierung ist sichergestellt, dass bekannte und unbekannte Störgrößen gleichmäßig zwischen den Therapiegruppen verteilt sind
- Damit ist automatisch für potentiell Confounding kontrolliert
- **Problem:** Randomisierte Studien sind aus verschiedenen Gründen nicht immer durchführbar
- Außerdem haben randomisierte Studien häufig eine geringe externe Validität

Hintergrund

- Wenn randomisierte Studien nicht durchführbar oder gewollt sind, bieten nichtrandomisierte Studien eine Alternative
- Die Ergebnisse nichtrandomisierter Studien können bessere externe Validität haben
- **Problem:** Bei nichtrandomisierten Studien können sich die Behandlungsgruppen systematisch voneinander unterscheiden
- Diese Unterschiede können bezüglich bekannter und unbekannter Störgrößen vorliegen (Confounding)
- Analysen müssen für die Confounder adjustieren
- Dies kann mittels herkömmlicher Regressionsverfahren oder Propensity-Score-Methoden geschehen

Potential Outcome Modell und kausale Effekte

Rubin's potential outcome model

- Wir bezeichnen mit T die Behandlungsvariable ($T = 1$ entspricht experimenteller Behandlung, $T = 0$ entspricht Kontrolle)
- Wir definieren die *potential outcomes* (auch counterfactuals):

Y_1 : Individuelles Outcome unter $T = 1$

Y_0 : Individuelles Outcome unter $T = 0$

- Der *Average Treatment Effect* (*ATE*) ist dann definiert als

$$ATE = E(Y_1 - Y_0) = E(Y_1) - E(Y_0)$$

- Der *Average Treatment Effect on the Treated* (*ATT*) ist dann definiert als

$$ATT = E(Y_1 - Y_0 | T = 1) = E(Y_1 | T = 1) - E(Y_0 | T = 1)$$

ATE vs. ATT

- ATE und ATT sind populationsabhängig
- ATE und ATT werden als kausale Effekte bezeichnet
- In randomisierten Studien gilt $ATE=ATT$, da wegen der Randomisierung
$$E(Y_1|T=1) = E(Y_1|T=0) = E(Y_1) \text{ und}$$
$$E(Y_0|T=1) = E(Y_0|T=0) = E(Y_0)$$
- In nicht randomisierten Studien unterscheiden sich ATE und ATT in der Regel
- Welcher der beiden Effekte betrachtet werden sollte, hängt von der Fragestellung ab

ATE vs. ATT

- Der ATE beschreibt den Effekt wenn man die ganze Population von der Kontrollbehandlung auf die experimentelle Behandlung umstellen würde
- Der ATT beschreibt den Effekt der Behandlung in den Patienten die tatsächlich behandelt wurden (welchen Benefit hatten die behandelten Patienten von der Behandlung)
- Der ATE ist von größerem Interesse, wenn die Intervention von außen entschieden wird (z.B. Entscheidung über Art einer Operation)
- In diesem Fall stellen experimentelle und Kontrollbehandlung ähnliche Anforderungen an die Patienten

ATE vs. ATT

- Der ATT ist von größerem Interesse, wenn es darum geht den Effekt der Intervention in Individuen zu beurteilen, die der Intervention zugeordnet wurden bzw. sich selbst der Intervention zuordnen würden
- Wenn die Intervention höhere Ansprüche an die Patienten stellt, als die Kontrollbehandlung könnte der ATT von größerem Interesse sein

Schätzung der kausalen Effekte

- Es sind niemals Y_0 und Y_1 am gleichen Individuum beobachtbar
- In einer randomisierten Studie gilt für die Kovariablen \mathbf{X} :

$$\mathcal{L}(\mathbf{X}|T=1) = \mathcal{L}(\mathbf{X}|T=0) = \mathcal{L}(\mathbf{X})$$

- Zudem hat man dann:

$$\begin{aligned}\bar{Y}_1 &\xrightarrow{n_1 \rightarrow \infty} E(Y_1) \\ \bar{Y}_0 &\xrightarrow{n_0 \rightarrow \infty} E(Y_0),\end{aligned}$$

wobei \bar{Y}_1 und \bar{Y}_0 die beobachteten Mittelwerte in den Behandlungsgruppen und n_1 und n_0 die zugehörigen Gruppengrößen sind

- Wir können den ATE (hier gleich dem ATT) also über die Differenz der Mittelwerte in den Behandlungsgruppen konsistent Schätzen

Schätzung der kausalen Effekte

- In nichtrandomisierten Studien gilt dieses Resultat aufgrund möglicher Confounding-Effekte nicht
- Wir müssen also anders auf mögliches Confounding reagieren

Der Propensity Score

Der Propensity Score

Definition: Propensity Score

Sei \mathbf{X} ein Vektor von Kovariablen, die vor der Behandlung erhoben wurden. Der *Propensity Score* $e(\mathbf{x})$ ist die bedingte Wahrscheinlichkeit die Behandlung $T = 1$ zu erhalten, gegeben $\mathbf{X} = \mathbf{x}$:

$$e(\mathbf{x}) = P(T = 1 | \mathbf{X} = \mathbf{x})$$

Eigenschaften des Propensity Scores

- Mit dem Propensity Score lassen sich die in \mathbf{X} eingeschlossen Kovariablen zwischen den Behandlungsgruppen balancieren (siehe später)
- Kann der Propensity Score bei der Schätzung des kausalen Effektes helfen?

SITA-Bedingung (Strong Ignorable Treatment Assignment)

1. $(Y_0, Y_1) \perp\!\!\!\perp T | \mathbf{X}$, das heißt \mathbf{X} enthält alle Y und T beeinflussenden Faktoren
 2. $0 < e(\mathbf{x}) < 1$
- Unter der SITA-Bedingung kann man den kausalen Effekt Δ in einer Population aus einer Beobachtungsstudie valide schätzen, indem man nach dem Propensity Score matcht oder stratifiziert. Weitere Verwendungsmöglichkeiten sind Adjustieren und Gewichten. (siehe später)

Balancierungseigenschaft des Propensity Scores

Theorem

\mathbf{X} und T sind, bedingt auf $e(\mathbf{x})$, stochastisch unabhängig:

$$P(t, \mathbf{x} | e(\mathbf{x})) = P(t | e(\mathbf{x})) P(\mathbf{x} | e(\mathbf{x}))$$

Beweis: Es ist

$$P(t, \mathbf{x} | e(\mathbf{x})) = P(t | \mathbf{x}, e(\mathbf{x})) P(\mathbf{x} | e(\mathbf{x})).$$

Weil $e(\mathbf{x})$ eine Funktion von \mathbf{x} ist, gilt zudem $P(t | \mathbf{x}, e(\mathbf{x})) = P(t | \mathbf{x})$. Es bleibt also zu zeigen:

$$P(t | \mathbf{x}) = P(t | e(\mathbf{x})).$$

Balancierungseigenschaft des Propensity Scores

Beweis (Fortsetzung): Aufgrund der Definition des Propensity Scores gilt für die linke Seite der letzten Gleichung:

$$P(T = 1|\mathbf{x}) = e(\mathbf{x}).$$

Für die rechte Seite der Gleichung folgt:

$$P(T = 1|e(\mathbf{x})) = E(T|e(\mathbf{x})) = E(E(T|\mathbf{X})|e(\mathbf{x})) = E(e(\mathbf{x})|e(\mathbf{x})) = e(\mathbf{x}).$$

Damit ist die Behauptung gezeigt.

Balancierungseigenschaft des Propensity Scores

- Allgemein gilt die folgende Balancierungseigenschaft:

Balancierungseigenschaft

Gegeben ein Wert e des Propensity Scores, sind für diese, durch e charakterisierte Gruppe die Kovariablen \mathbf{X} in der behandelten Gruppe so verteilt wie in der nichtbehandelten Gruppe:

$$\mathcal{L}(\mathbf{X} | T = 1, e) = \mathcal{L}(\mathbf{X} | T = 0, e)$$

- Zu beachten ist, dass diese Balancierungseigenschaft nur für die in \mathbf{X} enthaltenen Kovariablen gilt
- Für nicht gemessene oder bei der Bestimmung des Propensity Scores nicht berücksichtigte Variablen gilt die Eigenschaft nicht!

Schätzen des Propensity Scores

Schätzen des Propensity Scores

- Der Propensity Score ist unbekannt und muss daher geschätzt werden
- Ist T binär (nur 2 Behandlungen werden verglichen) wird der Propensity Score in der Regel mit Hilfe der logistischen Regression geschätzt
- Zielvariable: Behandlung T (damit wird die Wahrscheinlichkeit für $T = 1$, also der Propensity Score, modelliert und geschätzt)
- Welche Kovariablen sollen ins Modell aufgenommen werden?
 - Viele! Generell ein reichhaltiges Modell (auch mit Interaktionen und nicht-linearen Termen)
 - Vor allem Variablen, die späteren Therapieerfolg vorhersagen
 - Pre-Treatment Variablen, die Behandlung und Outcome beeinflussen
 - z.B. Alter, Geschlecht, bekannte Confounder
 - „Nimm alles aus der Datenbank, was du kriegen kannst“

Schätzen des Propensity Scores

- Wichtig: Es ist wichtig ein gut passendes Propensity Score Modell zu finden
- Die Ergebnisse der weiteren Analysen hängen in der Regel stark von den geschätzten Scores ab

Beurteilung des Propensity Score Modells

Zur Beurteilung des Modells zur Schätzung des Propensity Scores gibt es verschiedene Methoden, hier eine Möglichkeit:

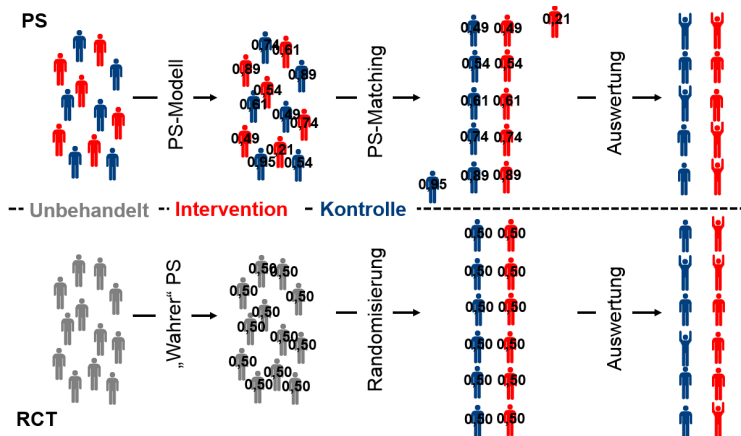
- Berechne für jedes Individuum aus dem angepassten Modell den geschätzten Propensity Score $\hat{e}(\mathbf{x})$
- Ordne die Individuen der Studie nach $\hat{e}(\mathbf{x})$ und bilde „schmale Streifen“ , z.B. Dezile oder Quintile
- Vergleiche in diesen „Streifen“ $\mathcal{L}(\mathbf{X})$ zwischen $T = 0$ und $T = 1$ für alle Kovariablen in \mathbf{X} (später mehr)
- Bei einem gut passenden Modell müssten die Verteilungen in beiden Behandlungsgruppen jeweils gleich sein
- Falls nicht, verbessere das Modell und beginne von vorn

Propensity Score Matching

Idee des Propensity Score Matching

- Ordne jedem Patienten in der Behandlungsgruppe einen (oder mehrere) Patienten in der Kontrollgruppe zu, die denselben (bzw. einen ähnlichen) Propensity Score haben
- Damit ist der Propensity Score in beiden Gruppen balanciert
- Damit sollten nach der Balancierungseigenschaft des Propensity Scores auch die Verteilungen der (berücksichtigten) Kovariablen in beiden Behandlungsgruppen gleich sein
- Auf den Daten der gematchten Patienten können wir dann herkömmliche Analysen durchführen um kausale Effekte zu schätzen

Propensity Score Matching vs. RCT



aus Kuß & Scherag (Vortragsfolien)

Propensity Score Matching – Vorteile

- Analysen nach dem Matching können kausale Interpretation haben
- Im Gegensatz zu anderen Propensity Score Methoden (wie Regressionsadjustierung oder inverse probability weighting) wird der Propensity Score nicht direkt in der Analyse verwendet. Daher erwartet man höhere Robustheit gegenüber Fehlspezifikation des Modells zum Schätzen des Propensity Scores.
- Das Matching wird als beste Methode betrachtet, um Unbalanciertheiten zwischen den Behandlungsgruppen auszugleichen (vgl. Austin 2009)
- Beim Propensity Score Matching lassen sich die Eigenschaften von behandelten und unbehandelten Patienten explizit darstellen (z.B. demographische Variablen). Es wird klar, für welche Population die Aussagen getroffen werden.

Propensity Score Matching – Nachteile

- Nicht immer können für alle Individuen Matching-Partner gefunden werden. Die Stichprobe verkleinert sich also in der Regel.
- Damit einher geht ein Verlust an Power
- Wie bei allen Propensity Score Methoden kann nur auf beobachtete Confounder adjustiert werden.
- Unbeobachtete Confounder können nach wie vor das Ergebnis verzerren.
- Dieses Problem wird schwächer, wenn die unbeobachteten Confounder unabhängig von den betrachteten Kovariablen sind. Ansonsten werden sie „mitgematcht“.
- Mit dem Propensity Score Matching kann nur der ATT geschätzt werden, nicht der ATE

Propensity Score Matching – allgemein

- In der Regel unterscheidet sich die Kovariablen Verteilung in der Ausgangskohorte von der Verteilung in der gematchten Kohorte
- Es wird also eine andere Population repräsentiert
- Der kausale Effekt ist populationsabhängig
- Die Population muss also immer beschrieben und die Verallgemeinerungsfähigkeit der Ergebnisse diskutiert werden

Propensity Score Matching – allgemein

- Die Matchpartner werden ausschließlich aufgrund ihres ähnlichen Propensity Scores zusammengeführt.
- Das heißt nicht, dass auch die Kovariablen zwischen den Matchpartnern ähnlich sein müssen!
- Die Balance der Kovariablen bezieht sich nur auf die **Verteilung** innerhalb der **Behandlungsgruppen**.

Methoden zum Propensity Score Matching

- Es existieren viele verschiedene Methoden die Matchingpartner zu bestimmen
- Werden häufig verwendete Methoden hier vorstellen
- Häufig wird nicht der geschätzte Propensity Score $\hat{e}(\mathbf{x})$ sondern sein logit

$$\hat{q}(\mathbf{x}) = \log \frac{\hat{e}(\mathbf{x})}{1 - \hat{e}(\mathbf{x})}$$

zum Matchen benutzt

- Die Verfahren können genauso auf $\hat{e}(\mathbf{x})$ wie auf $\hat{q}(\mathbf{x})$ angewendet werden
- In der Praxis werden zum Teil beide Ansätze ausprobiert und der Ansatz der zu besserer Kovariablen-Balance führt wird eingesetzt

Mahalanobis-Metrik Matching

1. Ordne die Studienteilnehmer zufällig
2. Berechne den „Abstand“ zwischen dem ersten behandelten Studienteilnehmer und allen Kontrollen via:

$$d(i, j) = (\mathbf{u} - \mathbf{v})^T \mathbf{C}^{-1} (\mathbf{u} - \mathbf{v}),$$

wobei \mathbf{u} und \mathbf{v} die Vektoren mit den Matchingvariablen für Individuum i bzw. j sind und \mathbf{C} die Stichprobenkovarianz der Matchingvariablen auf der Stichprobe aller nicht behandelten Patienten

3. Der nicht behandelte Patient mit dem kleinsten Abstand zum betrachteten behandelten Patient wird als Matching-Partner gewählt und beide Patienten werden für das Matching nicht weiter berücksichtigt
4. Dieses Vorgehen wird wiederholt bis zu jedem behandelten Patienten ein Kontrollpatient gematcht wurde

Mahalanobis-Metrik Matching

- Das Mahalanobis-Metrik Matching wurde bereits vor dem Propensity Score entwickelt und berücksichtigt diesen dementsprechend nicht
- Je mehr Variablen in die Berechnung des Mahalanobis-Distanz eingebunden werden, desto größer wird die durchschnittliche Mahalanobis-Distanz
- Gerade wenn viele Variablen in der Berechnung berücksichtigt werden kann es also schwer sein Matching-Partner zu finden, die „nah“ beieinander liegen
- Verschiedene Autoren definieren die Matrix **C** unterschiedlich (mal auf allen Patienten, mal nur Kontrollpatienten). Die Methode wird dort dann trotzdem als Mahalanobis Matching benannt

Mahalanobis-Metrik Matching mit dem Propensity Score

- Diese Methode wird genauso durchgeführt wie das Mahalanobis Matching, außer dass als zusätzliche Kovariable der geschätzte Propensity Score ($\hat{e}(\mathbf{x})$ bzw. $\hat{q}(\mathbf{x})$) in die Berechnung der Distanz mit aufgenommen wird

Nearest Neighbor Matching

1. Ordne die Studienteilnehmer zufällig
2. Betrachte den ersten behandelten Studienteilnehmer und finde den Patienten aus der Kontrollgruppe, dessen Propensity Score am nächsten zum Propensity Score des behandelten Patienten ist
3. Beide Patienten werden für das weitere Matching nicht berücksichtigt
4. Dieses Vorgehen wird wiederholt bis zu jedem behandelten Patienten ein Kontrollpatient gematcht wurde

Nearest Neighbor Matching

- Beim Nearest Neighbor Matching gibt es keine Anforderung, wie nah sich die Propensity Scores der gematchten Patienten sein müssen
- Es ist also möglich, dass Patienten mit sehr unterschiedlichen Propensity Scores gematcht werden, nur weil es zu dem betrachteten behandelten Patienten keinen „gut passenden“ Kontrollpatienten gibt
- Es kann also sein, dass durch dieses Matching nach wie vor große Unterschiede im Propensity Score zwischen den gematchten Gruppen vorliegen
- Alternative: Caliper Matching

Caliper Matching

- Das Vorgehen beim Caliper Matching entspricht dem des Nearest Neighbor Matching
- Einschränkung: Zwei Patienten werden nur dann als „gut passend“ angesehen und dementsprechend gematcht, wenn sich ihre Propensity Scores um nicht mehr als einen vordefinierten Wert ϵ (Caliper) unterscheiden
- Findet man zu einem Patienten der Behandlungsgruppe so keinen Matching-Partner, so wird dieser Patient entfernt
- In der Literatur wird ein Wert von $\epsilon \leq 0.25\sigma_P$ vorgeschlagen, wobei σ_P die Standardabweichung der geschätzten Propensity Scores in der Stichprobe ist (Rosenbaum & Rubin (1985))

Nearest Available Mahalanobis-Metrik Matching in Calipern definiert durch den Propensity Score

1. Ordne die Studienteilnehmer zufällig und betrachte den ersten behandelten Patienten
2. Betrachte alle nicht behandelten Studienteilnehmer, deren Propensity Score sich um nicht mehr als einen vordefinierten Wert ϵ (Caliper) vom Propensity Score des betrachteten Patienten der Behandlungsgruppe unterscheiden
3. Findet man zu einem Patienten der Behandlungsgruppe so keinen Matching-Partner, so wird dieser Patient entfernt
4. Berechne für all diese Patienten die Mahalanobis-Distanz zum betrachteten Patienten der Behandlungsgruppe basierend auf einer kleinen Anzahl an Kovariablen (ohne den Propensity Score)
5. Wähle den Patienten der Kontrollgruppe mit dem kleinsten Abstand zum betrachteten behandelten Patient Matching-Partner und berücksichtige die beiden Patienten für das Matching nicht weiter
6. Wiederhole dieses Vorgehen für alle Patienten aus der Behandlungsgruppe

Nearest Available Mahalanobis-Metrik Matching in Calipern definiert durch den Propensity Score

- Dieses Vorgehen ist also eine Kombination aus Mahalanobis-Metrik Matching, und Caliper Matching
- Rosenbaum und Rubin (1985) argumentieren, dass diese Methode die beste Balance der Kovariablen in den Behandlungsgruppen liefert

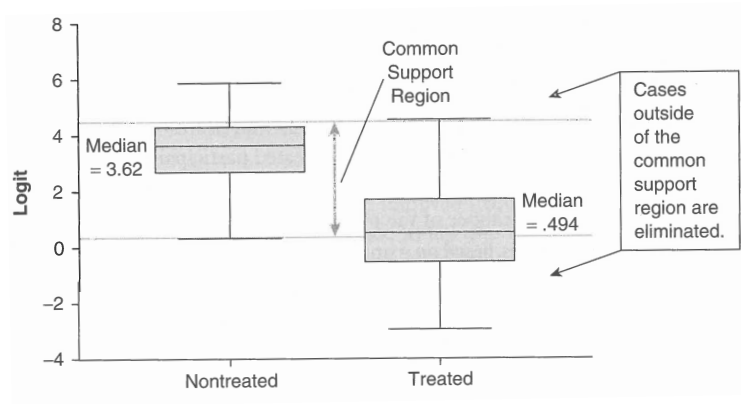
Greedy Matching

- Alle bisher vorgestellten Matching Prozeduren werden unter dem Begriff *Greedy Matching* zusammengefasst
- Sie werden dafür kritisiert, dass die Reihenfolge des Matchings die Qualität des Matchings und damit die Analyseergebnisse stark beeinflussen kann (Alternative: Optimal Matching)
- Außerdem wird kritisiert, dass nicht für jeden behandelten Patienten auch immer ein passender Kontrollpatient gefunden wird (incomplete matching)
- Wenn viele behandelte Patienten ausgeschlossen werden, kann es schwer sein, die Ergebnisse der Analysen zu interpretieren
- Hier muss man zwischen ungenauem und unvollständigem Matchen abwägen

Greedy Matching

- Ein weiterer Kritikpunkt ist, dass Greedy Matching Prozeduren nur auf dem gemeinsamen Träger der Propensity Scores beider Behandlungsgruppen matchen
- Individuen beider Gruppen außerhalb dieses Bereiches wird kein Matching-Partner zugeordnet
- Der gemeinsame Träger der Propensity Scores beider Behandlungsgruppen ist jedoch sehr stark abhängig von dem Modell mit dem der Propensity Score geschätzt wird (Wahl der Kovariablen und Wahl des funktionalen Zusammenhangs)
- Damit hängt auch das Analyseresultat stark vom verwendeten Modell ab
- In der Praxis wird manchmal empfohlen verschiedene Modelle zum Schätzen des Propensity Scores zu betrachten und Sensitivitätsanalysen durchzuführen, in denen die Größe des gemeinsamen Trägers variiert

Greedy Matching



aus Guo & Fraser (2010)

Optimal Matching – Rechenbeispiel

- Wir nehmen an, dass wir 2 gematchte Paare aus 4 Individuen bilden müssen
- Beobachtete Propensity Scores: 0.1, 0.5, 0.6, 0.9
- Ein Greedy Matching Verfahren würde das zweite und dritte Individuum als erstes Paar auswählen, weil hier die Differenz der Propensity Scores am geringsten ist (das zweite Paar bilden entsprechend Individuum 1 und 4)
- Insgesamt liegt die (summierte) Differenz der Propensity Scores bei:
 $|0.5-0.6|+|0.1-0.9|=0.9$
- Matcht man Individuum 1 und 2 und Individuum 3 und 4, so liegt die (summierte) Differenz der Propensity Scores bei: $|0.1-0.5|+|0.6-0.9|=0.7$

Optimal Matching – Rechenbeispiel

- Im zweiten Fall hat jedes der Paare eine höhere Propensity Score Differenz als das erste Paar im ersten Fall
- Global gesehen führt der zweite Ansatz jedoch zu einer besseren Balance

Optimal Matching

- Greedy Matching Algorithmen versuchen lediglich in jedem Schritt ein „bestes“ Matching zu erzielen
- Optimal Matching Algorithmen versuchen die Paare so zu bilden, dass z.B. eine globale Differenz der Scores zwischen den Behandlungsgruppen minimiert wird
- Es existieren verschiedene Optimal Matching Algorithmen, die hier nicht im Detail besprochen werden
- In SAS und R gibt es Makros bzw. Packages, in denen Optimal Matching Prozeduren implementiert sind

Prüfen der Balancierung nach dem Matching

- Nach erfolgreichem Matching kann/sollte man die Balance der beiden Behandlungsgruppen hinsichtlich der Kovariablenverteilung prüfen
- Es existieren verschiedene Maße um die Balance zwischen den Gruppen zu beschreiben
- Eines ist die z-Differenz, die ursprünglich für stetige Variablen entwickelt und dann auf ordinale und nominale Variablen erweitert wurde (Kuß (2013))
- Unter der Hypothese homogener Kovariablenverteilung ist die z-Differenz standardnormalverteilt

Balancierung nach dem Matching – Datenbeispiel

- Publiizierte PS-Analyse aus der koronaren Bypasschirurgie (Börgermann (2012))
- 1282 Patienten, die zwischen Juli 2009 und November 2010 am Herz- und Diabeteszentrum NRW in Bad Oeynhausen isoliert koronarchirurgisch versorgt wurden
- Vergleich konventionelle Technik (cCABG, $n=887$, 69,2%) und Clampless-off-pump (Less-OPCAB, $n=395$, 30,8%)-Technik
- Entscheidung für Operationsmethode durch den jeweiligen Operateur

Balancierung nach dem Matching – Datenbeispiel

- Schätzung des PS-Modell durch logistisches Regressionsmodell: Alter, Geschlecht, BMI, Hauptstammstenose, LVEF, Präoperativer Myokardinfarkt, Hypertonie, Diabetes mellitus, COPD, Niereninsuffizienz, Schlaganfall, pAVK, Voroperationen, Dringlichkeit, Präoperative IABP
- 1:1-Matching mit logit-transformierten PS, Optimal-Matching-Algorithmus, Caliper-Weite von 0.2 Standardabweichungen des logit-transformierten PS

Balancierung nach dem Matching – Datenbeispiel

	Alle Patienten (n=1282)			PS gematchte Patienten (n=788)		
	Less-OPCAB (n=395)	cCABG (n=887)	z-Differenz	Less-OPCAB (n=394)	cCABG (n=394)	z-Differenz
Alter (Jahre)	69.3 ± 9.1	67.5 ± 9.4	3.24	69.3 ± 9.1	69.0 ± 8.9	0.49
Männlich (%)	78.2	77.9	0.13	78.2	77.9	0.09
Prä-OP MI (%)	27.1	35.7	-3.14	27.2	26.7	0.16

Kovariablenbalance vor und nach dem PS-Matching

Propensity-Score-stratifizierte Analyse

Stratifizierte Propensity Score Analyse

- Bei der stratifizierten Propensity Score Analyse teilen wir die Patienten basierend auf ihrem Propensity Score in M Strata
- Ein häufig angewendeter Fall: Bilde Patientengruppen anhand der Qunitile der Verteilung der geschätzten Propensity Scores ($M = 5$)
- Laut Rubin wird so 90% des Bias der durch gemessene Confounder induziert ist beseitigt
- Innerhalb jedes Stratums wird der Behandlungseffekt über die Mittelwertdifferenz Δ_m des Outcomes in diesem Stratum geschätzt

Stratifizierte Propensity Score Analyse

- Der average treatment effect über alle Strata hinweg wird dann als gewichtetes Mittel der startumsspezifischen Effekten geschätzt:

$$\widehat{ATE}^S = \sum_{m=1}^M \Delta_m \frac{n_{0,m} + n_{1,m}}{n},$$

wobei n die Gesamtfallzahl ist, $n_{1,m}$ und $n_{0,m}$ die Zahlen der behandelten und Kontrollpatienten im Stratum m sind und Δ_m die Mittelwertdifferenz des Outcomes in Stratum m ist

- Wenn die Strata als Quintile der Verteilung der geschätzten Propensity Scores gebildet werden, ist $\frac{n_{0,m} + n_{1,m}}{n} \approx 1/5$

Stratifizierte Propensity Score Analyse

- Die stratifizierte Propensity Score Analyse bietet auch die Möglichkeit den ATT zu schätzen
- Dazu wird beim zusammenführen der stratumsspezifischen Mittelwertdifferenzen eine andere Gewichtung vorgenommen:

$$\widehat{ATT}^S = \sum_{m=1}^M \Delta_m \frac{n_{1,m}}{n_1},$$

wobei n_1 die Gesamtzahl behandelter Patienten ist, $n_{1,m}$ die Zahl der behandelten Patienten in Stratum m ist und Δ_m die Mittelwertdifferenz des Outcomes in Stratum m ist

- Hier werden also die stratumsspezifischen Effekte nach der Zahl der behandelten Patienten in diesem Stratum gewichtet

Stratifizierte Propensity Score Analyse – Vorteile

- Alle Daten werden genutzt
- Relativ einfach umzusetzen
- Schätzung von ATE und ATT möglich
- Relativ robust gegenüber Misspezifikation des Propensity Score Modells

Stratifizierte Propensity Score Analyse – Nachteile

- Die Reduktion des Bias ist in der Regel geringer als beim Propensity Score Matching oder Weighting
- Liefert unzureichende Resultate bei Verwendung an Survivaldaten (Deb et al. (2016))

Propensity Score Weighting

Propensity Score Weighting – ATE

- Wir wissen, dass die Mittelwertdifferenz zwischen den Behandlungsgruppen $\hat{\Delta}$ aufgrund möglicher Confounder kein unverzerrter Schätzer für den ATE oder den ATT ist:

$$\hat{\Delta} = \frac{\sum_{i=1}^n T_i Y_i}{\sum_{i=1}^n T_i} - \frac{\sum_{i=1}^n (1 - T_i) Y_i}{\sum_{i=1}^n 1 - T_i}$$

- Die Idee des Propensity Score Weighting ist nun eine andere Gewichtung bei der Berechnung der Mittelwerte zu wählen (basierend auf den geschätzten Propensity Scores)
- Um den ATE zu schätzen, benutzen wir den Schätzer:

$$\widehat{ATE}^w = \sum_{i=1}^n \frac{T_i Y_i}{\hat{e}(\mathbf{x}_i)} - \sum_{i=1}^n \frac{(1 - T_i) Y_i}{1 - \hat{e}(\mathbf{x}_i)}$$

Propensity Score Weighting – ATE

- So wird also einem Individuum aus der Behandlungsgruppe das Gewicht $1/\hat{e}(\mathbf{x}_i)$ und einem Individuum aus der Kontrollgruppe das Gewicht $1/(1 - \hat{e}(\mathbf{x}_i))$ zugeordnet
- Nun ist $\hat{e}(\mathbf{x}_i)$ ein Schätzer für die Wahrscheinlichkeit die Behandlung zu erhalten und $1 - \hat{e}(\mathbf{x}_i)$ ein Schätzer dafür die Kontrollbehandlung zu erhalten
- Das Gewicht entspricht also dem Inversen der Wahrscheinlichkeit die jeweilige Behandlung zu erhalten
- Man spricht daher auch von *Inverse Probability of Treatment Weighting (IPTW)*

Propensity Score Weighting – ATT

- Das Propensity Score Weighting bietet auch die Möglichkeit, den ATT zu schätzen
- Dazu verwenden wir modifizierte Gewichte:

$$\widehat{ATT}^w = \sum_{i=1}^n T_i Y_i - \sum_{i=1}^n \frac{(1 - T_i) Y_i \hat{e}(\mathbf{x}_i)}{1 - \hat{e}(\mathbf{x}_i)}$$

- Individuen der Behandlungsgruppe erhalten also das Gewicht 1, Individuen der Kontrollgruppe erhalten das Gewicht $\hat{e}(\mathbf{x}_i)/(1 - \hat{e}(\mathbf{x}_i))$, was also den odds für die Behandlung entspricht

Propensity Score Weighting – allgemein

- Für beide Methoden soll an dieser Stelle nicht gezeigt werden, warum die jeweilige Gewichtung das gewünschte Resultat in der Schätzung verursacht
- Geeignete Literatur ist in Kapitel 5.6 von Guo & Fraser (2010) zu finden
- Die Gewichte in beiden Methoden summieren sich nicht auf 1. Es gibt Methoden, die diese Gewichte modifizieren, sodass sie sich auf 1 summieren

Propensity Score Weighting – Vorteile

- Einfach umzusetzen
- Alle Daten werden genutzt
- Bessere Biaskorrektur als bei Stratifizierung oder Kovariablenadjustierung

Propensity Score Weighting – Nachteile

- Anfällig für Fehlspezifizierung des Propensity Score Modells
- Liefert nur unter relativ starken Bedingungen optimale Ergebnisse

Kovarialbenadjustierung mit dem Propensity Score

Kovariablenadjustierung mit dem Propensity Score

- Eine Möglichkeit um Imbalancen zwischen den Behandlungsgruppen zu berücksichtigen ist ein Regressionsmodell mit Zielvariable Y und allen relevanten Kovariablen zu formulieren
- Gerade wenn viele Kovariablen betrachtet werden sollen, birgt dieser Ansatz viele Nachteile, insbesondere die Dimensionsproblematik
- Ein Ansatz diesen Problemen zu begegnen ist, den Propensity Score als Kovariable im Modell zu berücksichtigen
- Dafür wird der Propensity Score auf Basis der relevanten Kovariablen geschätzt und anstelle dieser Variablen als Kovariable ins Regressionsmodell aufgenommen

Kovariablenadjustierung mit dem Propensity Score – Vorteile

- Benutzt alle Daten
- Relativ einfach umzusetzen

Kovariablenadjustierung mit dem Propensity Score – Nachteile





- Der funktionale Zusammenhang zwischen Outcome und Propensity Score muss spezifiziert werden
- Weniger Robust gegen Fehlspezifikationen des Propensity Score Modells, da der Score direkt in der Analyse verwendet wird
- Keine kausale Interpretation. Weder ATE noch ATT werden standardmäßig geschätzt

Fazit

Fazit

- Propensity Score Matching scheint häufig die favorisierte Auswertungsmethode basieren auf dem Propensity Score zu sein
- Propensity Score Analysen können, wenn sie richtig verwendet werden, auf gemessene Confounder adjustieren
- Ungemessenes Confounding ist nach wie vor ein Problem
- Dieses Problem lässt sich nur durch randomisierte Studien lösen
- Sind diese nicht möglich oder umsetzbar, bieten Propensity Score Methoden gute Alternativen

Literatur I

-  P.R. Rosenbaum und D.B. Rubin.
The Central Role of the Propensity Score in Observational Studies for Causal Effects.
Biometrika, 70:41–55, 1983.
-  P.R. Rosenbaum und D.B. Rubin.
Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score.
The American Statistician, 39:33–38, 1985.
-  O. Kuß und A. Scherag.
Die Propensity Score-Methode -Eine anwendungsorientierte Einführung.
Vortragsfolien, http://www.oliverkuss.de/science/publications/Kuss_Scherag_GMDS2015_Der_Propensity_Score_20150909.pdf.
-  S. Guo und M.W. Fraser.
Propensity Score Analysis – Statistical Methods and Applications.
SAGE Publications, Thousand Oaks, 2010.

Literatur II



S. Deb et al.

A Review of Propensity-Score Methods and Their Use in Cardiovascular Research.

Canadian Journal of Cardiology, 32:259–265, 2016.



P.C. Austin.

The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies.

Med Decis Making, 29:661-77, 2009.