Causal Inference for Survival Analysis

Yossi Levy
November 24, 2019

Causal inference on a shoestring

The potential outcomes framework (Rubin):

- Y : outcome of interest
- A : treatment/intervention, assumed to be binary (0/1)
- $Y_j^{A=0}, Y_j^{A=1}$: potential outcomes for patient j
- Problem: for each patient we observe only one of the potential outcomes
- Yet, we still want to estimate the average treatment effect:

$$ATE = E[Y^1 - Y^0]$$

Potential outcomes framework assumptions

- One version of treatment
- Subjects do not interfere with each other
- Consistency: for each subject, one of the counterfactuals outcomes is actually factual.
- Exchangeability: the conditional probability of receiving every value of treatment depends only on the measured covariates
- Positivity: the conditional probability of receiving every value of treatment is greater than zero
- Nice to have: causal mechanism (Pearl)

ATE estimation: weighting

$$\widehat{ATE} = \frac{1}{n_1}\sum_{j:A_j=1} w_j y_j - \frac{1}{n_0}\sum_{k:A_k=0} w_k y_k$$

For example: IPTW = Propensity scores weighting

$$PS_j = P(A_j = 1|X_j)$$

$$w_j = \frac{1}{A_j \cdot PS_j + (1 - A_j) \cdot (1 - PS_j)}$$

ATE estimation: parametric g-formula

Let g be a model for Y^A given A and X:

$$E(Y^A|X) = g(X, A)$$

Then we can estimate the potential outcomes for patient j by

$$\widehat{Y_j^{A=a}} = g(X_j, A = a)$$

And therefore

$$\widehat{ATE} = \frac{1}{n}\sum_j \left[\widehat{Y_j^{A=1}} - \widehat{Y_j^{A=0}}\right]$$

Survival analysis basics

Let T be the time until an event of interest is happening.

T is a random variable with probability distribution F where F is "nice".

Example: event of interest is an occurrence of relapse

$$F(t) = P(Patient\ relapsed\ before\ time\ t)$$

$$F(t) = P(T \le t)$$

The survival function:

$$S(t) = P(Patient\ did\ not\ relapse\ by\ time\ t)$$

$$S(t) = 1 - F(t)$$

Survival analysis – the hazard function

If a patient survived so far, what is the probability that he will survive a little more?

$$P(\text{Patient did not die by time } t \text{ but died shortly after that})$$
$$= P(\text{Patient died shortly after } t \mid \text{Patient did not die by time } t)$$
$$= P(t < T < t + \Delta t \mid T > t)$$

$$\text{Mean hazard over time interval} = \frac{P(t < T < t + \Delta t \mid T > t)}{\Delta t}$$

$$h(t) = \text{Hazard at time } t = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t \mid T > t)}{\Delta t}$$

$$h(t) = \frac{F'(t)}{S(t)} \Leftrightarrow S(t) = \exp\left\{-\int_0^t h(x)dx\right\}$$

7

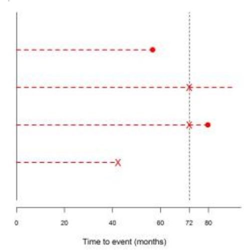
Survival analysis: outcomes of interest

- Mean survival time: $\mu = \int_0^\infty S(x)dx$
- Restricted mean survival time: $\mu^* = \int_0^{t^*} S(x)dx$
- Median survival time: $M = S^{-1}(0.5)$
- Some other quantile of survival function: $Q_p = S^{-1}(1-p)$
- And more… (no spoilers)

8

Survival analysis - censoring

id	actual_time2event	time2event	event
1	56.8	56.8	1
2	NA	72.0	0
3	79.8	79.8	0
4	NA	42.3	0



9

Example: Rotterdam data

- Patients diagnosed with breast cancer
- Some patients received chemo, some didn't
- Outcome of interest: time to relapse
- Follow up time: up to 231 months

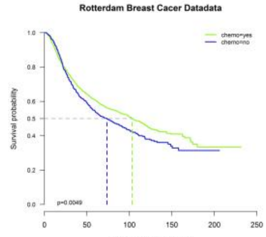
id	time2relapse	relapse	chemo	age	meno	size	grade	pr	er	nodes
1623	56.542095	1	yes	44	pre	>20-50mmm	2	500	115	2
2508	8.410678	1	no	80	post	>50mm	3	7	62	9
1631	13.174538	1	no	49	pre	<=20mm	3	0	0	0
2402	21.519506	1	yes	36	pre	>20-50mmm	3	55	36	3
2283	74.546204	1	no	73	post	<=20mm	3	23	890	2
515	86.669403	0	no	65	post	<=20mm	3	0	0	0

10

Survival – Kaplan-Meyer

	records	events	*rmean	*se(rmean)	median
chemo=no	2402	1181	115.6916	2.324118	103.22793
chemo=yes	580	337	105.1970	3.981588	73.69199

Rotterdam Breast Cacer Datadata



11

Estimate propensity scores

```
pats$chemo.numeric=2-as.numeric(pats$chemo)
ps.formula=formula(chemo.numeric ~  age + meno + size + grade +
                   pr + er + I(exp(-0.12*nodes)))
library(CBPS)
cbps.fit=CBPS(formula=ps.formula, data=pats, ATT=0)
pats$ps=cbps.fit$fitted.values
pats$iptw.weight=pats$chemo.numeric/pats$ps+(1-pats$chemo.numeric)/(1-pats$ps)
```

id	chemo	ps	iptw.weight
1623	yes	0.368968581	2.710258
2508	no	0.046982771	1.049299
1631	no	0.137912691	1.159975
2402	yes	0.547013536	1.828108
2283	no	0.004546575	1.004567
515	no	0.020889972	1.021336

12

Potential outcomes using Cox PH model

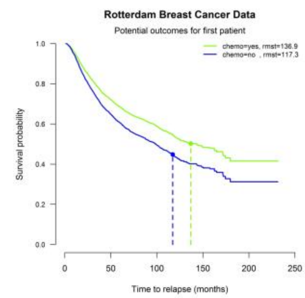
```
# estimate cox ph model
cox.formula=formula(surv.relapse~ chemo.numeric + age + meno + size + grade +
                      I(exp(-0.12*nodes)) + pr + er)
cox.model=coxph(cox.formula, data=pats, ties="breslow")

# estimate potential survival functions for first patient
pat=pats[1,]
pat.cf=pat
pat.cf$chemo.numeric=1-pat$chemo.numeric

sv.yes=survfit(cox.model, newdata=pat)
sv.no=survfit(cox.model, newdata=pat.cf)


             *rmean *se(rmean)   median
chemo=yes 136.8925   2.745241 143.4086
chemo=no  117.3350   2.174864  99.0883
```


Potential outcomes for first patient



Potential outcomes all patients

```
# calculate factual and counterfactual restricted mean for each patient
cox.fit.all=survfit(cox.model, newdata=pats)
tb=summary(cox.fit.all)$table
pats$rmean.fact=tb[,5]

pats.cf=pats
pats.cf$chemo.numeric=1-pats.cf$chemo.numeric
cox.fit.cf=survfit(cox.model, newdata=pats.cf)
tb=summary(cox.fit.cf)
tb=tb$table
pats$rmean.cf=tb[,5]
```


ATE of RMST using IPTW

```
iptw.model=lm(rmean.fact~chemo.numeric, data=pats, weights=pats$iptw.weight)
coeff=summary(iptw.model)$coefficients
conf=confint(iptw.model)
tb=cbind(coeff[,-3], conf)
print(tb)


                Estimate Std. Error     Pr(>|t|)      2.5 %     97.5 %
(Intercept)    112.92249  0.9655889 0.000000e+00 111.029202 114.81578
chemo.numeric   11.94045  1.4093394 3.723112e-17   9.177076  14.70383
```


ATE of RMST using G-formula

```
pats$Y1=0
pats$Y0=0

w=which(pats$chemo.numeric==1)
pats$Y1[w]=pats[w, "rmean.fact"]
pats$Y1[-w]=pats[-w, "rmean.cf"]
pats$Y0[w]=pats[w, "rmean.cf"]
pats$Y0[-w]=pats[-w, "rmean.fact"]
pats$diff=pats$Y1-pats$Y0


   id chemo rmean.fact  rmean.cf        Y1        Y0     diff
 1623   yes  136.89253 117.33500 136.89253 117.33500 19.55753
 2508    no   51.25906  68.10959  68.10959  51.25906 16.85053
 1631    no  142.24213 159.32847 159.32847 142.24213 17.08634
 2402   yes   87.94699  68.56066  87.94699  68.56066 19.38633
 2283    no  135.85747 153.70169 153.70169 135.85747 17.84422
  515    no  150.00561 166.06680 166.06680 150.00561 16.06119
```


ATE of RMST using G-formula

```
lm.diff=lm(diff~1, data=pats)
coeff=summary(lm.diff)$coefficients
conf.int=confint(lm.diff)
tb=as.data.frame(cbind(coeff, conf.int))
tb=tb[,-3]
names(tb)=c("effect", "std.err", "p.value", "lower.ci", "upper.ci")
             effect    std.err p.value lower.ci upper.ci
(Intercept) 17.2687 0.05127172       0 17.16817 17.36923

# alternatice calculation
t.test(x=pats$diff)
        One Sample t-test

 data:  pats$diff
 t = 336.81, df = 2981, p-value < 2.2e-16
 alternative hypothesis: true mean is not equal to 0
 95 percent confidence interval:
  17.16817 17.36923
 sample estimates:
 mean of x
   17.2687
```

AF: Attributable Fraction

- AF, the Attributable Fraction, is the proportion of incidents in the population that are attributable to the risk factor
- In survival analysis context, we look at AF(t) which is the proportion of incidents in the population that occurred before time t and that are attributable to the risk factor

$$AF(t) = 1 - \frac{P[Y^{A=0}(t) = 1]}{P[Y(t) = 1]} = 1 - \frac{P[Y^{A=0}(t) = 1]}{1 - S(t)}$$

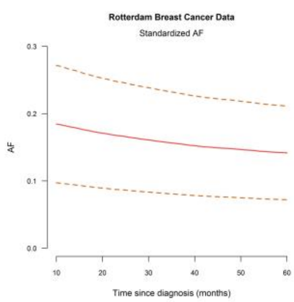
AF(t) estimation

```
# in this case, the risk factor is "no chemotherapy"
pats$nochemo=as.numeric(pats$chemo=="no")
AF.formula=formula(surv.relapse ~ nochemo + age + meno + size + grade +
                     I(exp(-0.12*nodes)) + pr + er)
AF.fit=coxph(AF.formula, data=pats, method="breslow")
AF.std=stdCoxph(fit=AF.fit, data=pats, X="nochemo", t=10:60, x=c(NA, 0))

# AF function
AF=function(est){
  p=1-est[,1]
  p0=1-est[,2]
  af=1-p0/p
  return(af)
}

AF.estimates=AF(AF.std$est)
AF.ci=confint(object=AF.std, fun=AF)
```


AF(t)



NNT: Numbers Needed to Treat

- NNT, the Number Needed to Treat, is the average number of patients who need to be treated in order to prevent one additional bad outcome
- In survival analysis context, we look at NNT(t), which the average number of patients who need to be treated by time t in order to prevent one additional bad outcome

$$CRD(t) = P[Y(t) = 1|A = 0] - P[Y^{A=1}(t) = 1|A = 0]$$

$$= 1 - S_0(t) - P[Y^{A=1}(t) = 1|A = 0]$$

$$NNT(t) = 1/CRD(t)$$

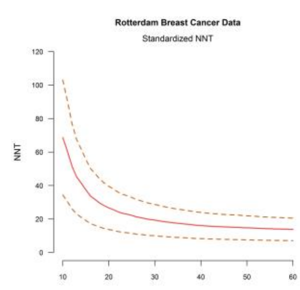
NNT(t) estimation

```
NNT.formula=formula(surv.relapse ~ chemo + age + meno + size + grade +
                      I(exp(-0.12*nodes)) + pr + er)
NNT.fit=coxph(NNT.formula, data=pats, method="breslow")
NNT.std=stdCoxph(fit=NNT.fit, data=pats, X="chemo", t=10:60,
                 x=c(NA, "yes"), subsetnew=(chemo=="no"))

# NNT function
NNT=function(est){
  p=1-est[, 1]
  p1=1-est[, 2]
  nnt=1/(p-p1)
  return(nnt)
}

NNT.estimates=NNT(NNT.std$est)
NNT.ci=confint(object=NNT.std, fun=NNT)
```


NNT(t)



Pseudo observations

- Let θ be a survival parameter, and let $\hat{\theta}$ be a consistent estimate for θ based on a standard survival model (no covariates)
- For each patient j, let $\hat{\theta}_{-j}$ be the parameter estimated from the same model, but patient j is omitted from the population.
- Then $\hat{\theta}_j$, the pseudo observation of patient *i* is then defined as
- $\hat{\theta}_j = n \cdot \hat{\theta} - (n-1) \cdot \hat{\theta}_{-j}$
- Once the pseudo observations are estimated, standard approaches, such as the g-formula and IPTW can be applied to the pseudo observations.

25

Pseudo-observations ATE estimation using IPTW

```
# Create pseudo observations using 231-month RMST
pats$pseudo.rmean=pseudomean(pats$time2relapse, pats$relapse, 231)

# model formula for pseudo observations
pseudo.formula=formula(pseudo.rmean ~ chemo.numeric + age + meno + size + grade
+ pr + er + I(exp(-0.12*nodes)))

# ATE of restricted mean using IPTW
iptw.model=lm(pseudo.formula, data=pats, weights=pats$iptw.weight)
```

	Estimate	Std. Error	Pr(>|t|)	2.5 %	97.5 %
(Intercept)	-1.75630005	21.734571222	9.356011e-01	-44.372632498	40.86003240
chemo.numeric	11.95829357	4.745812240	1.179560e-02	2.652882849	21.26370430
age	0.94783056	0.310634987	2.298905e-03	0.338749127	1.55691200
menopre	13.36326978	7.178686071	6.276939e-02	-0.712428743	27.43896831
size>20-50mmm	-26.81483196	4.898955535	4.777386e-08	-36.420520315	-17.20914360
size>50mm	-22.31646373	9.251164860	1.591301e-02	-40.455800973	-4.17712649
grade3	-36.03630746	5.221153278	6.243283e-12	-46.273749073	-25.79886585
pr	0.01832063	0.008398712	2.923479e-02	0.001852748	0.03478851
er	-0.03860403	0.007609177	4.148963e-07	-0.053523822	-0.02368424
I(exp(-0.12 * nodes))	127.64446659	9.901952405	4.890972e-37	108.229089529	147.05984366

26

Pseudo-observations ATE estimation using G-formula

```
pseudo.fit=lm(pseudo.formula, data=pats)
pats$pseudo.rmean.factual=predict(pseudo.fit, newdata=pats)
pats.cf=pats
pats.cf$chemo.numeric=1-pats.cf$chemo.numeric
pats$pseudo.rmean.cf=predict(pseudo.fit, newdata=pats.cf)
pats$pseudo.Y1=0
pats$pseudo.Y0=0
w=which(pats$chemo.numeric==1)
pats$pseudo.Y1[w]=pats[w, "rmean.fact"]
pats$pseudo.Y1[-w]=pats[-w, "rmean.cf"]
pats$pseudo.Y0[w]=pats[w, "rmean.cf"]
pats$pseudo.Y0[-w]=pats[-w, "rmean.fact"]
pats$pseudo.diff=pats$pseudo.Y1-pats$pseudo.Y0
```

id	chemo	pseudo.rmean.factual	pseudo.rmean.cf	pseudo.Y1	pseudo.Y0	pseudo.diff
1623	yes	137.21445	121.55757	136.89253	117.33500	19.55753
2508	no	58.56850	74.22538	68.10959	51.25906	16.85053
1631	no	141.93275	157.58963	159.32847	142.24213	17.08634
2402	yes	94.41474	78.75785	87.94699	68.56066	19.38633
2283	no	120.88074	136.53762	153.70169	135.85747	17.84422
515	no	147.87365	163.53053	166.06680	150.00561	16.06119

27

Pseudo-observations ATE estimation using G-formula

```
lm.pseudo.diff=lm(pseudo.diff~1, data=pats)
coeff=summary(lm.pseudo.diff)$coefficients
conf.int=confint(lm.pseudo.diff)
tb=as.data.frame(cbind(coeff, conf.int))
tb=tb[,-3]
names(tb)=c("effect", "std.err", "p.value", "lower.ci", "upper.ci")
```

	effect	std.err	p.value	lower.ci	upper.ci
(Intercept)	17.2687	0.05127172	0	17.16817	17.36923

28

29