COMPETING RISKS

INTRODUCTION THROUGH EXAMPLES

Medical Example

- Cancer researcher finds a medicine that cures cancer.
- Run a medical study where you follow 100 patients for 5 years after giving them cancer cure to see how many die.
- In year 4, 7 of these patients travel together to Iceland and die in a volcano accident.
- The other 93 patients made it to the end of five years without passing away.

Medical Example

- Cancer researcher finds a medicine that cures cancer.
- Run a medical study where you follow 100 patients for 5 years after giving them cancer cure to see how many die.
- In year 4, 7 of these patients travel together to Iceland and die in a volcano accident.
- The other 93 patients made it to the end of five years without passing away.

WHAT IS THE MORTALITY RATE?

Medical Example

- Cancer researcher finds a medicine that cures cancer.
- Run a medical study where you follow 100 patients for 5 years after giving them cancer cure to see how many die.
- In year 4, 7 of these patients travel together to Iceland and die in a volcano accident.
- The other 93 patients made it to the end of five years without passing away.

WHAT IS THE MORTALITY RATE?

DOES 7% FEEL RIGHT?

Customer Example

- Observe customers over the past year to try and analyze **voluntary** churn.
- Of the 1000 customers in the data set, 240 left voluntarily, while 60 left involuntarily.

WHAT IS THE CUSTOMER CHURN RATE?

Customer Example

- Observe customers over the past year to try and analyze **voluntary** churn.
- Of the 1000 customers in the data set, 240 left voluntarily, while 60 left involuntarily.

WHAT IS THE CUSTOMER CHURN RATE?

DOES 30% FEEL RIGHT?

Fixed vs. Random Censoring

- **Fixed censoring** – censoring only occurs at the end of the study ($C_i = c$ is known in advance).
  - Recidivism data: Not arrested in 52 weeks is censored by design because that is when study ended.

- **Random censoring** – C_i may vary between subjects for reasons beyond the investigator's control.
  - Recidivism data: No arrest within first 30 weeks, but lose contact with subject for whatever reason.
  - Recidivism data: Study done only for one year, but people can have delayed entry into the study (as they were released).

COMPETING RISKS

Multiple Event Types

- All of the models used so far have been for studying the time until **one** event occurs.
- All of the models used so far can be extended to studying **multiple** events or **multiple types** of events.

Competing Risks

- Examples:
  - Death from cancer in medical study vs. other causes of death.
  - Leaving job due to retirement, injury, or being fired.
  - Pump failure due to jamming, flooding, motor failure, or surge.
- In all of the above cases there are multiple, **mutually exclusive** causes of failure.
- These are examples of a **competing risks** problem, where each subject can experience only one of several possible events.

Independence Again…

- Assume T_i and C_i are independent – subjects censored at time *t* were randomly selected to be censored from all subjects still in the risk set at *t*.
- **IF** this is true, then fixed vs. random censoring is mathematically equivalent.
- What does independence "mean" here?
  - In competing risks, independence implies that a censored observation and an uncensored observation have the **same risk of the event**, **regardless of the reason for censoring**.

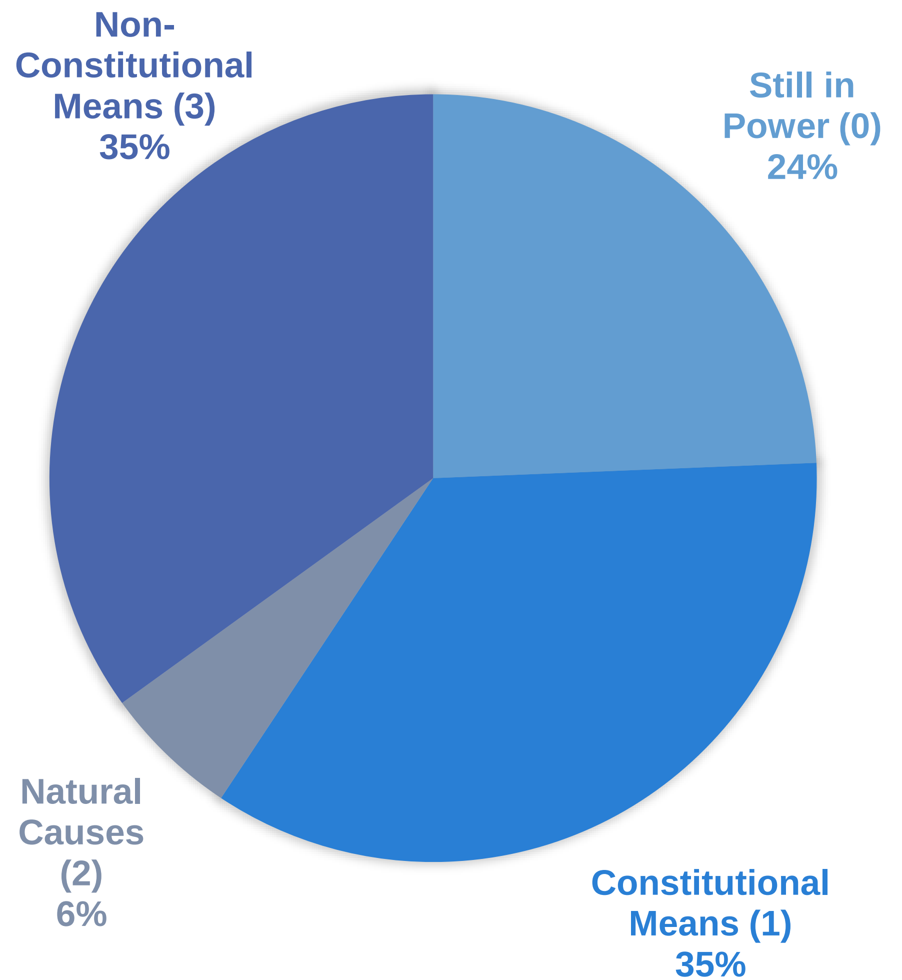
Independence Again…

- Example:
  - By treating other failure types as censored, we're essentially implying once a pump fails due to jamming, we still don't know when it would fail due to flooding – we assume that the event types are independent.
- **NO TEST FOR THIS!**
- Decide independent or not based on context of problem.
- In other words, are observations with a high risk of one event equally likely to experience the other events?

ESTIMATION

World Leaders Data Set

- Compiled by Bienen and van de Walle in 1991.
- Primary leaders of all countries between 1960 and 1987.
- Number of years the leader was in power and the manner they lost power.



World Leaders Data Set

- Manner – how the leader reached power (0: constitutional, 1: non-constitutional)
- Start – year of entry to power
- Military – background of leader (1: military, 0: civilian)
- Age – age at time of entry
- Conflict – level of ethnic conflict (1: medium/high, 0:low)
- LogInc – log of GNP per capita
- Growth – avg. annual growth rate of GNP
- Pop – population in millions
- Land – land area in 1000 km^2
- Literacy – literacy rate (unknown year)
- Region – 0: Middle East, 1: Africa, 2: Asia, 3: Latin America
- Years – length of time leader was in power (in years)

Review

- Two major functions in survival analysis:
- Survival Function – probability of surviving beyond time *t:*

$$S(t) = P(T > t) = 1 - F(t)$$

- Hazard Function – conditional failure rate in an interval:

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

Cause-Specific Hazard Function

- When there are multiple event types, the hazard function contains two variables – *T* and *J (time til event occurs or is censored and which event type it belongs to)*.
- The cause/type specific hazard function is as follows:

$$h_{i,j}(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T_i < t + \Delta t, J_i = j \mid T_i \geq t)}{\Delta t}$$

$$h_i(t) = \sum_j h_{i,j}(t)$$

- The interpretation stays the same, just type specific.

Cause-Specific Hazard Function

- When there are multiple event types, the hazard function contains two variables – *T* and *J*.
- The cause/type specific hazard function is as follows:

$$h_{i,j}(t) = \lim_{\Delta t \to 0} \frac{P(t \le T_i < t + \Delta t, \boxed{J_i = j} \mid T_i \ge t)}{\Delta t}$$

$$h_i(t) = \sum_j h_{i,j}(t)$$

- The interpretation stays the same, just type specific.

CAUSE-SPECIFIC HAZARD MODEL

Modeling Type-Specific Events

- Type-Specific events can be modeled with both proportional hazard models …

$$\log h_k(t) = \log h_{0,k}(t) + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}$$

- … and accelerated failure time (AFT) models :

$$\log T_{i,k} = \beta_0 + \beta_1 x_{i,1} + \cdots + \sigma e_i$$

Cox Regression Competing Risks

- Typical modeling approach for competing risks is to use separate Cox regression models for **each** cause, treating all other events as censored.
- Essentially, modeling the effects of predictors on the cause-specific hazard:

$$\log h_k(t) = \log h_{0,k}(t) + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}$$

Cox Competing Risks – R

```
cox_nat <- coxph(Surv(years, lost == "Natural") ~ manner + start
                 + military + age + conflict + loginc + growth
                 + pop + land + literacy + factor(region),
                 data = leaders)
summary(cox_nat)
```

Cox Competing Risks – R

```
## Call:
## coxph(formula = Surv(years, lost == "Natural") ~ manner + start +
##     military + age + conflict + loginc + growth + pop + land +
##     literacy + factor(region), data = leaders)
##
##   n= 438, number of events= 27
##    (34 observations deleted due to missingness)
##
##                       coef  exp(coef)   se(coef)      z Pr(>|z|)
## manner           3.747e-01  1.455e+00  6.633e-01  0.565    0.572
## start           -5.403e-02  9.474e-01  3.386e-02 -1.596    0.111
## military        -3.646e-01  6.945e-01  7.409e-01 -0.492    0.623
## age              7.386e-02  1.077e+00  1.840e-02  4.015 5.95e-05 ***
## conflict        -2.609e-01  7.704e-01  4.720e-01 -0.553    0.580
## loginc           3.285e-01  1.389e+00  2.673e-01  1.229    0.219
## growth           8.817e-02  1.092e+00  8.518e-02  1.035    0.301
## pop              1.991e-03  1.002e+00  2.138e-03  0.931    0.352
## land            -3.969e-05  1.000e+00  1.781e-04 -0.223    0.824
## literacy        -8.796e-03  9.912e-01  1.260e-02 -0.698    0.485
## factor(region)1 -6.427e-01  5.259e-01  8.360e-01 -0.769    0.442
## factor(region)2 -7.776e-01  4.595e-01  9.031e-01 -0.861    0.389
## factor(region)3  6.591e-01  1.933e+00  7.852e-01  0.839    0.401
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cox Competing Risks – R

```
##                   exp(coef) exp(-coef) lower .95 upper .95
## manner               1.4546     0.6875   0.39644     5.337
## start                0.9474     1.0555   0.88657     1.012
## military             0.6945     1.4400   0.16255     2.967
## age                  1.0767     0.9288   1.03853     1.116
## conflict             0.7704     1.2980   0.30548     1.943
## loginc               1.3889     0.7200   0.82251     2.345
## growth               1.0922     0.9156   0.92423     1.291
## pop                  1.0020     0.9980   0.99780     1.006
## land                 1.0000     1.0000   0.99961     1.000
## literacy             0.9912     1.0088   0.96707     1.016
## factor(region)1      0.5259     1.9015   0.10217     2.707
## factor(region)2      0.4595     2.1763   0.07827     2.698
## factor(region)3      1.9330     0.5173   0.41484     9.007
##
## Concordance= 0.819  (se = 0.046 )
## Likelihood ratio test= 32.42  on 13 df,   p=0.002
## Wald test            = 29.47  on 13 df,   p=0.006
## Score (logrank) test = 33.21  on 13 df,   p=0.002
```

AFT Models with Competing Risks

- Accelerated Failure Time models have a similar structure to Cox regression models when dealing with competing risks.
- With AFT Models, distributions need to be evaluated for all types of failure!

CONDITIONAL PROCESSES

Independent Events?

- The cause-specific hazard method for competing risks presumes that each event type has its own hazard that governs **both** the occurrence and timing of events of that type.
- They are assumed to be independent processes acting in parallel with each other.
- Example:
  - Death due to natural causes vs. forcible removal from power.

Conditional Processes

- What if independence **DOES NOT** seem reasonable?
- **Conditional processes** occur when these events are NOT independent of each other – conditional on each other.
- Fine-Gray Model

FINE-GRAY MODEL

Cumulative Incidence Function

- The **cumulative incidence function** (CIF) is marginal probability for each competing risk
- The CIF is the product of two estimates

Hazard at time t_f:

$$\hat{h}_c(t_f) = \frac{m_{cf}}{n_f}$$

Where m_{cf} denotes the number of events for risk c at time t_f and n_f is the number of subjects at that time

$$\hat{S}(t_{f-1})$$

Where S(t) denotes the OVERALL survival function (not cause specific survival function)

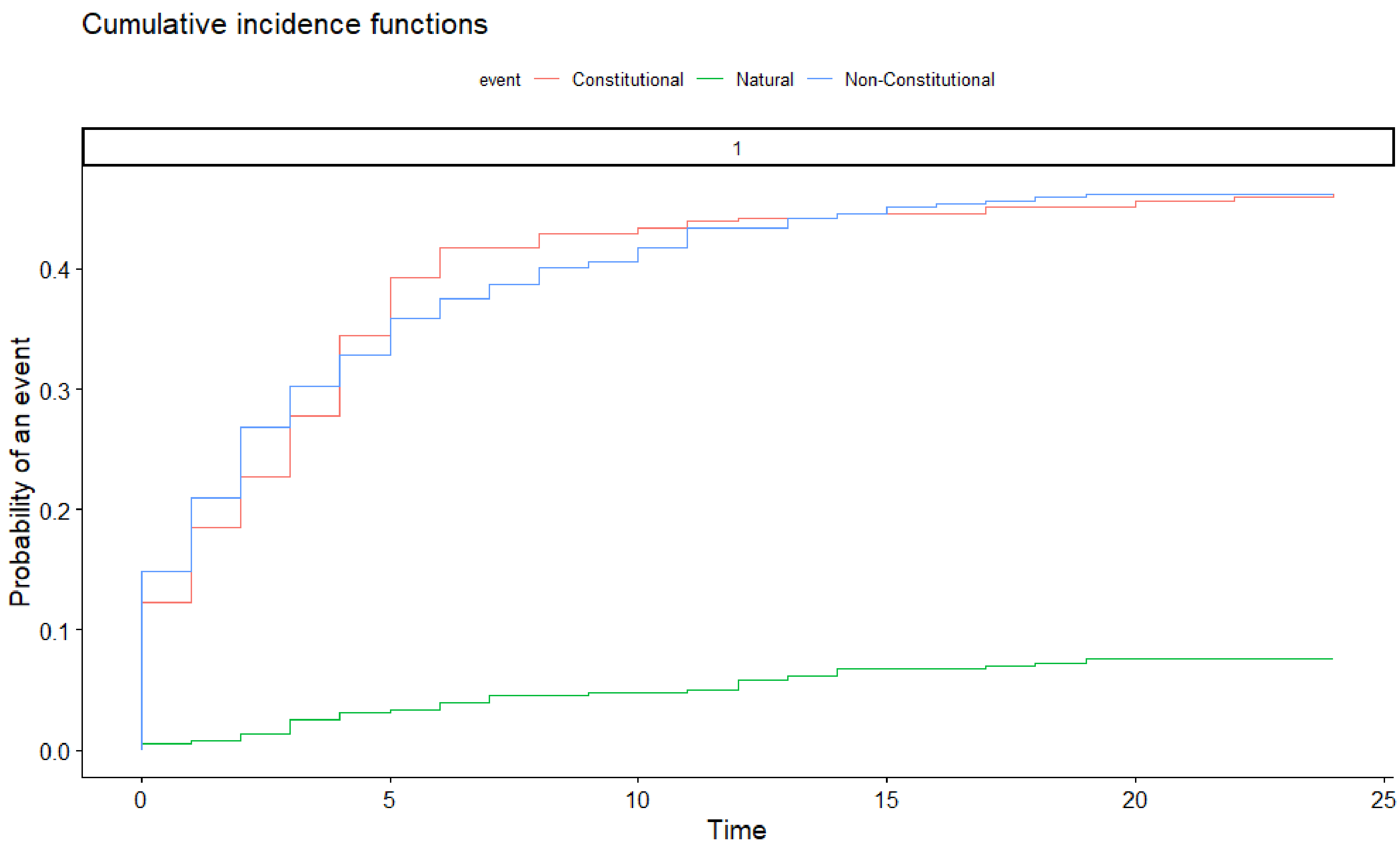
CIF

- In other words, the product of surviving the previous time periods and the cause specific hazard at time t_f

$$\hat{I}_c(t_f) = \hat{S}(t_{f-1})\hat{h}_c(t_f)$$

- Fine and Gray proposed a proportional hazards model for the CIF with covariates (censoring times and event times no longer need to be independent)

Estimating the CIF's – R



Fine-Gray Model – R

```
gray.natural=crr(tenure,status.leaders,x,failcode="Natural")
 summary(gray.natural)
```

Fine-Gray Model – R

Competing Risks Regression

Call:
crr(ftime = tenure, fstatus = status.leaders, cov1 = x, failcode = "Natural")

	coef	exp(coef)	se(coef)	z	p-value
manner	-7.32e-02	0.929	0.562617	-0.13007	0.9000
start	-8.33e-02	0.920	0.026453	-3.14757	0.0016
Military	-2.51e-01	0.778	0.549101	-0.45674	0.6500
age	4.75e-02	1.049	0.018127	2.62190	0.0087
conflict	-2.13e-03	0.998	0.440003	-0.00484	1.0000
loginc	5.55e-01	1.741	0.261591	2.11995	0.0340
growth	9.80e-02	1.103	0.128218	0.76408	0.4400
pop	2.41e-03	1.002	0.002784	0.86500	0.3900
land	-8.76e-05	1.000	0.000189	-0.46445	0.6400
Literacy	-6.71e-03	0.993	0.011000	-0.60956	0.5400
africa	4.44e-01	1.559	0.710063	0.62578	0.5300
asia	-6.97e-01	0.498	0.821035	-0.84936	0.4000
latin	1.22e-01	1.129	0.625890	0.19434	0.8500