Modele długości trwania

- Pierwotne zastosowania:
  - przemysłowe (trwałość produktów)
  - aktuarialne (długość trwania życia)
- Zastosowania ekonomiczne:
  - długości bezrobocia
  - długości czasu między zakupami dóbr trwałego użytku
  - długości trwania strajków, itd.
- Dane długości trwania $t_1, \ldots, t_n$ od początku zdarzenia do jego końca lub momentem badania T

– problem ocenzurowania (n.p. przeżywalność ofiar Czarnobyla,wszytkie $t_i < T$, gdzie T jest czasem, który upłyną od katastrofy)

Funkcja przeżywalności

- Prawdopodobieństwo, że czas trwania będzie krótszy niż t:

$$F(t) = \int_0^t f(s)\,ds = \Pr(T \leq t)$$

- Funkcję przeżywalności (*survival function),* prawdopodobieństwo, że czas trwania conajmniej równy t

$$S(t) = 1 - F(t) = \Pr(T > t)$$

Funkcja hazardu

- Prawdopodobieństwo, że po dotrwaniu do t, zjawisko zakończy się w okresie $(t, t+h)$.

$$l(t, h) = \Pr\left(t < T \leq t + h \middle| T > t\right)$$

- Dla granicznie małych Δ otrzymujemy funkcję hazardu (*hazard rate*)

$$\boldsymbol{\lambda}(t) = \lim_{h \to 0} \frac{\Pr\left(t < T \leq t + h \middle| T > t\right)}{h} = \lim_{h \to 0} \frac{F(t+h) - F(t)}{hS(t)} = \frac{f(t)}{S(t)}$$

- Funkcja hazardu

– może przyjmować wartości spoza przedziału $[0, 1]$.
– jest miarą ryzyka, na które wystawiony jest dany obiekt

- **Przykład**: rosnąca funkcja hazardu w przypadku osób starszych odbija prawidłowość, że są one bardziej wystawieni na ryzyko śmierci niż osoby młode

Zależności między funkcjami hazardu, gęstości i przeżywalności

- Z definicji

$$\boldsymbol{\lambda}(t) = -\frac{\partial \ln S(t)}{\partial t}$$

$$f(t) = \boldsymbol{\lambda}(t) S(t)$$

- Zcałkowana funkcję hazardu

$$\Lambda(t) = \int_0^t \boldsymbol{\lambda}(s)\, ds$$

wtedy

$$S\left(t\right)=e^{-\Lambda\left(t\right)}$$

a więc

$$\Lambda\left(t\right)=-\ln S\left(t\right)$$

Funkcja wiarygodności

- Funkcja wiarygodności dla modeli długości trwania może być zapisana jako

$$\ell\left(\boldsymbol{\theta}\right)=\ln\left[\prod_{i=1}^{n}f_{\boldsymbol{\theta}}\left(t_i\right)\right]=\sum_{i=1}^{n}\ln\left[\boldsymbol{\lambda}_{\boldsymbol{\theta}}\left(t_i\right)S_{\boldsymbol{\theta}}\left(t_i\right)\right]=\sum_{i=1}^{n}\ln\boldsymbol{\lambda}_{\boldsymbol{\theta}}\left(t\right)+\sum_{i=1}^{n}\ln S_{\boldsymbol{\theta}}\left(t\right)$$

- Przypadek ocenzurowane zmiennych.
  - dla t_i nieocenzurowanych wstawiamy w funkcji wiarygodności $f\left(t_i\right)$ a dla t_i ocenzurowanych $S\left(t_i\right)$.

– funkcja wiarygodności:

$$\ell(\boldsymbol{\theta}) = \ln\left[\prod_{\text{nieocenzurowane}} f_{\boldsymbol{\theta}}(t_i)\right] \ln\left[\prod_{\text{ocenzurowane}} S_{\boldsymbol{\theta}}(t_i)\right]$$

$$= \sum_{\text{nieocenzurowane}} \ln \boldsymbol{\lambda}_{\boldsymbol{\theta}}(t) + \sum_{\text{wszystkie}} \ln S_{\boldsymbol{\theta}}(t)$$

Formy funkcyjne

- Modele długości trwania: założenia na temat rozkładów:
  - monotoniczne funkcje hazardu
  - niemotoniczne funkcje hazardu
- Monotoniczne funkcji hazardu:

Hazard	Funkcja hazardu $\boldsymbol{\lambda}(t)$	Funkcja przeżywalności $S(t)$
Stały	$\boldsymbol{\lambda}$	$S(t) = e^{-\boldsymbol{\lambda} t}$
Liniowy	$\boldsymbol{\lambda}(t) = \boldsymbol{\alpha} + \boldsymbol{\beta} t$	$S(t) = e^{-\left(\boldsymbol{\alpha} t + \frac{1}{2}\boldsymbol{\beta} t^2\right)}$
Wykładniczy	$\boldsymbol{\lambda}(t) = e^{a+\boldsymbol{\beta} t}$	$S(t) = e^{-\frac{1}{\boldsymbol{\beta}}\exp(a+\boldsymbol{\beta} t)}$

Table 1: Warianty rozkładu wykładniczego

- Niemonotoniczne funkcje hazardu:

Rozkład	Funkcja hazardu $\boldsymbol{\lambda}(t)$	Funkcja przeżywalności $S(t)$
Weibull	$\boldsymbol{\lambda}(t) = \boldsymbol{\lambda} p (\boldsymbol{\lambda} t)^{p-1}$	$S(t) = e^{-(\boldsymbol{\lambda} t)^p}$
Lognormalny	$f(t) = \left(\frac{p}{t}\right) \boldsymbol{\phi}[p \ln(\boldsymbol{\lambda} t)]$	$S(t) = \Phi[-p \ln(\boldsymbol{\lambda} t)]$
Loglogistyczny	$\boldsymbol{\lambda}(t) = \frac{\boldsymbol{\lambda} p (\boldsymbol{\lambda} t)^{p-1}}{1+(\boldsymbol{\lambda} t)^p}$	$S(t) = \frac{1}{1+(\boldsymbol{\lambda} t)^p}$

Table 2: Funkcje hazardu i przeżywalności

- **Uwagi**:
  - dla $p = 1$ rozkład Weibulla sprowadza się do rozkładu wykładniczeg
  - dla rozkładu Lognormalnego, $\ln(t)$ powinno mieć rozkład normalny $N\left(-p \ln \boldsymbol{\lambda}, p^{-2}\right)$ a
  - dla rozkładu loglogistycznego $\ln(t)$ ma rozkład logistyczny o wartości oczekiwanej $\mathrm{E}[\ln(t)] = -p \ln \boldsymbol{\lambda}$.

- Zmienne objaśniające wprowadzamy do modelu, zakładając, że

$$\ln \boldsymbol{\lambda}_i = -\boldsymbol{x}_i \boldsymbol{\beta}$$

Interpretacja β

- Dla rosnącego $\boldsymbol{\lambda}$ maleje wartość oczekiwana $\mathrm{E}\,(t)$.
  - interpretowalne znaki współczynników.

- Interpretacja wielkości parametrów zależy od modelu.
  - wartość oczekiwana t w rozkładzie Weibulla wynosi $\mathrm{E}\,(t|\,\boldsymbol{x}_i) = \exp\left(\boldsymbol{x}_{it}\boldsymbol{\beta}\right)\Gamma\left(\frac{1}{p}+1\right)$.
  - dla rozkładu lognormalnego i loglogistycznego $\mathrm{E}\left[\ln\left(t\right)|\,\boldsymbol{x}_i\right] = \boldsymbol{x}_i\boldsymbol{\beta}$.

Model proporcjonalego hazardu Coxa

- Zakładamy, że funkcja hazardu ma postać

$$\boldsymbol{\lambda}\left(\boldsymbol{x}_i, t_i\right) = \exp\left(-\boldsymbol{x}_i \boldsymbol{\beta}\right) \boldsymbol{\lambda}_0\left(t\right)$$

- $\boldsymbol{\lambda}_0\left(t\right)$ jest ryzykiem bazowym

- Funkcję hazrdu bazowego jest trudno oszacować, ponieważ może mieć dowolną postać.

- Zamiast szacować $\boldsymbol{\lambda}_0\left(t\right)$ dla każdej obserwacji i badamy prawdopodobieństwa warunkowe

- Załóżmy, że

  - każda jednostka opuszcza próbę w różnym momencie
  - jednostka i opuszcza próbę w czasie t_i

- Zdefiniujmy zbiór ryzyka R_{t_i} jako zbiór zawierający jednostki, które dotrwały do t_i i w związku z tym mogą wypaść z próby w czasie t_i

- Zakładamy, że w czasie t_i wypadła z próby tylko jedna jednostka

- Prawdopodobieństwo warunkowe, że spośród jednostek w R_{t_i}, to jednostka i opuściła próbę w czasie t_i, przy znanym zbiorze ryzyka R_{t_i}

jest równe prawdopodobieństwu:

$$\begin{aligned}
\Pr\left(T_i = t_i \middle| R_{t_i}\right) &= \frac{\Pr\left(T_i = t_i \middle| T_i \geq t_i\right)}{\sum_{j \in R_{t_i}} \Pr\left(T_j = t_i \middle| T_j \geq t_i\right)} \\
&= \frac{\boldsymbol{\lambda}\left(\boldsymbol{x}_j, t_i\right) \boldsymbol{\lambda}_0\left(t\right)}{\sum_{j \in R_{t_i}} \boldsymbol{\lambda}\left(\boldsymbol{x}_j, t_i\right) \boldsymbol{\lambda}_0\left(t\right)} = \frac{\exp\left(-\boldsymbol{x}_j \boldsymbol{\beta}\right)}{\sum_{i \in R_{t_i}} \exp\left(-\boldsymbol{x}_j \boldsymbol{\beta}\right)}
\end{aligned}$$

- W rezultacie okazuje się, że prawdopodobieństwo warunkowe $\Pr\left(T_i = t_i \middle| R_{t_i}\right)$ nie zależy od funkcji hazardu bazowego.

- Parametr $\boldsymbol{\beta}$ można oszacować budując estymując powyższy model za pomocą pewnego wariantu Metody Największej Wiarygodności.