

# Survival models in banking

Een bank geeft leningen uit, zoals hypotheek. Van deze leningen willen we weten welke kasstromen we kunnen verwachten. In dit artikel beschrijven we hoe deze kasstromen bepaald worden met behulp van een techniek die zijn oorsprong heeft in een heel andere hoek: de biostatistiek.

## ACHTERGROND

Een hypotheek heeft contractuele kasstromen, zoals rentebetalingen en reguliere aflossingen. Er kunnen ook afwijkingen zijn van het contract. Als de klant minder betaalt dan afgesproken, spreken we van kredietrisico: een klant kan niet aan zijn betalingsverplichtingen voldoen. Als de klant meer betaalt dan afgesproken noemen we het prepayment: de klant wil de lening vervroegd aflossen. Dat komt bijvoorbeeld voor als de klant wil verhuizen of de lening wil oversluiten naar een andere bank.

Het inschatten van de kasstromen vanwege prepayment is belangrijk voor het beheren van de balans van de bank. We hebben met een klant een contract voor een bepaald rentepercentage. Als een hypotheek vervroegd aflost, dan ontvangen we die rente niet langer. We kunnen

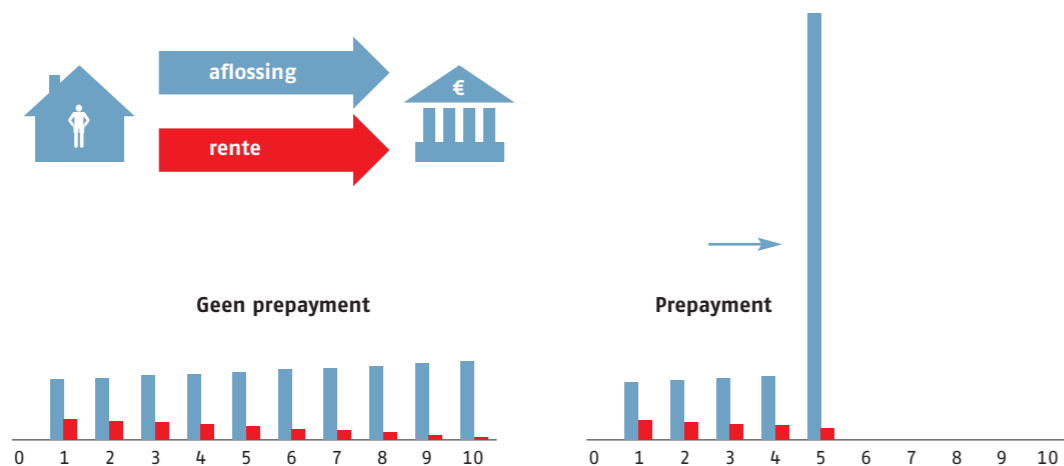
wel proberen een nieuwe hypotheek te verstrekken, maar als de rente intussen gewijzigd is, dan ontvangen we meer of minder rente dan verwacht. Dat is het renterisico dat de bank loopt. Daarnaast moet de bank zijn liquiditeit in de gaten houden: is er genoeg geld in de kas om aan de vraag (van bijvoorbeeld spaarders) te voldoen? Het inschatten van de verwachte inkomsten en uitgaven is dan van groot belang.

## BIOSTATISTIEK (I)

Vergelijkbare inschattingen moeten ook gemaakt worden in de geneeskunde. Als een patiënt met leukemie (of een andere vorm van kanker) gediagnosticeerd wordt, dan krijgt deze van de arts te horen wat de drie- en vijfjaars overlevingskans is. Deze kans hangt onder andere af van het type leukemie. Het bepalen van de overlevingskansen gebeurt met survival modellen. Deze vorm van statistiek gebruiken we als we geïnteresseerd zijn in de tijd tot een gebeurtenis, zoals de tijd tot overlijden. In dit geval zijn we natuurlijk vooral geïnteresseerd in de tijd dat deze gebeurtenis nog niet gebeurd is, de overlevingsduur.

## HYPOTHEKEN (I)

Bij hypotheekleningen zijn we geïnteresseerd in de prepayment gebeurtenis. We bekijken de versimpelde situatie waarin er geen kredietrisico is: alle klanten kunnen aan hun betalingsverplichtingen voldoen. Ook nemen we aan dat als een klant vervroegd wil aflossen, deze dan de volledige hypotheek aflost. Als we het tijdstip van prepayment kennen, dan zijn alle kasstromen bekend. De klant betaalt de contractuele kasstromen tot het moment van prepayment, waarop de gehele resterende hoofdsom wordt terugbetaald (zie Figuur 1).



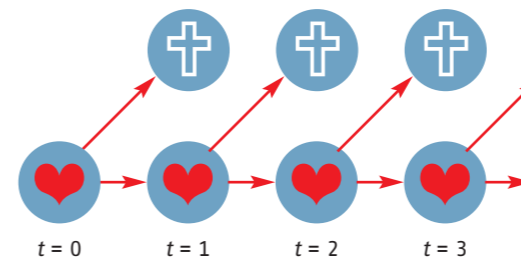
Figuur 1: Kasstromen van een hypotheek.

A. Holst MSc is Senior Risk Modeller bij De Volksbank.



## BIOSTATISTIEK (II)

Met survivalmodellen beschrijven we de kans dat een individu nog leeft na een bepaalde tijd, zoals de driejaars overlevingskans van de leukemiepatiënt. Om deze kans te vinden, gebruiken we een *hazard rate*. Dit is de conditionele kans op overlijden in het komende jaar, gegeven dat de patiënt nog leeft. Figuur 2 illustreert dit model. De patiënt kan bijvoorbeeld in het eerste jaar een conditionele kans hebben van 1/2 om te sterven, in het tweede jaar een kans van 1/3, en in het derde jaar een kans van 1/4. De overlevingskans na drie jaar is dan gelijk aan drie keer achtereenvolgende kans op niet dood gaan in het komende jaar. In dit voorbeeld is de driejaars overlevingskans dus gelijk aan  $1/2 \times 2/3 \times 3/4 = 1/4$ .



Figuur 2: Illustratie survival model.

## HYPOTHEKEN (II)

Terug naar de bank: we waren geïnteresseerd in de kasstromen van een hypotheek. Als we op elk toekomstig moment weten wat de overlevingskans is, dan kunnen we ook de verwachte kasstromen bepalen. Stel we beginnen met een hypotheek van 100 euro, en de overlevingskans na één jaar is 90%, dan verwachten we dat er (gemiddeld) nog 90 euro over is na één jaar. Dan moet er dus gemiddeld een kasstroom van 10 euro geweest zijn in dat jaar. Op deze manier kunnen we met een survivalmodel de kasstromen van een hypotheek voorspellen.

## BIOSTATISTIEK (III)

Om de hazard rates in te kunnen schatten gebruikt het survivalmodel historische gegevens van patiënten die gediagnosticeerd zijn met leukemie. In een onderzoek worden deze mensen gevolgd en wordt het vastgelegd als iemand sterft. Zo kunnen we empirisch meten wat de driejaars overlevingskans is. Er worden ook andere gegevens verzameld, bijvoorbeeld welke vorm van leukemie de patiënten hebben, en hoe oud ze zijn. Deze risicofactoren kunnen de overlevingskans verhogen of verlagen.

## HYPOTHEKEN (III)

Ook bij de bank gebruiken we risicofactoren om de uitstroom van geld door prepayment beter in te kunnen schatten. Van alle hypotheekleningen die we de afgelopen jaren hebben verstrekt hebben we allerlei informatie verzameld. Zo weten we bijvoorbeeld dat klanten die in een appartement wonen vaker verhuizen dan gemiddeld. Of dat de rente voor klanten een belangrijke motivatie is om de hypotheek over te sluiten naar een andere bank.

## BIOSTATISTIEK (IV)

Survival modelling is zeer geschikt voor overlevingsvraagstukken omdat het model goed om kan gaan met censoring in de data. Censoring is de term voor gedeeltelijke informatie, je weet bijvoorbeeld dat een patiënt minstens drie jaar overleefd heeft, maar niet of deze op een later moment alsnog gestorven is. Dat is bijvoorbeeld het geval als het onderzoek geëindigd is terwijl de patiënt nog leeft. Normaal gesproken is het onduidelijk hoe we deze patiënt mee moeten nemen in de vijfjaar overlevingskans: we weten immers niet of deze patiënt vijf jaar overleeft.

Hoe zouden we dat op kunnen lossen?

- We kunnen aannemen dat deze patiënt stierf op het moment dat het onderzoek eindigde, maar dan onderschatten we waarschijnlijk de overlevingskans, want het kan zijn dat deze patiënt wel vijf jaar overleeft.
- We kunnen ook aannemen dat deze patiënt voor altijd blijft leven, maar dan overschatten we de overlevingskans waarschijnlijk weer.
- We kunnen de patiënt ook helemaal weglaten, maar dan creëren we een *bias* in het model, want patiënten die langer leven zijn vaker gecensord en worden dan dus vaker weggelaten uit het onderzoek. We zullen dan de overlevingskans onderschatten.

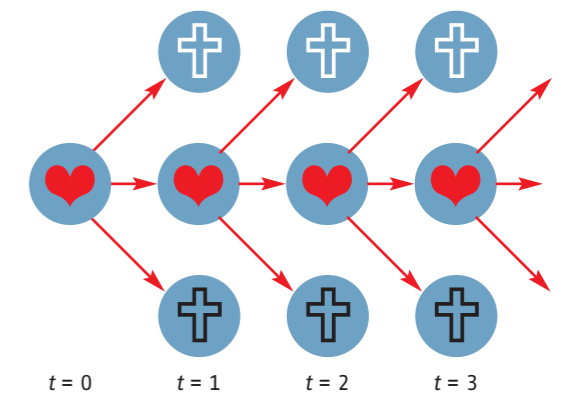
Alle drie deze methoden zijn dus ongeschikt. We hebben namelijk de belangrijke informatie dat deze patiënt *langer* dan drie jaar leeft niet gebruikt. Door het gebruik van de hazard rate in het survival model kijken we steeds slechts één jaar vooruit. Op die manier kan de patiënt in kwestie de eerste drie jaar bijdragen aan het model en gebruiken we alle informatie die beschikbaar is.

## HYPOTHEKEN (IV)

Ook in de bank hebben we te maken met censoring. Van elke nog lopende hypotheek weten we namelijk niet of en zo ja, wanneer deze vervroegd gaat aflossen. Door een survival model te gebruiken kunnen we deze hypotheekleningen toch op een correcte manier meenemen in het schatten van het prepayment model.

## BIOSTATISTIEK (V)

Het is natuurlijk mogelijk dat de patiënt overlijdt aan iets anders dan de leukemie. Als we dit in kaart willen brengen gebruiken we een competing risk survival model. Dit model is geïllustreerd in Figuur 3. De kans om één jaar te overleven is nu gelijk aan één min de som van de kansen op de twee manieren om te overlijden.



Figuur 3: Illustratie competing risks survival model.

## HYPOTHEKEN (V)

Niet elke vervroegde aflossing is voor de bank gelijk. Als de klant oversluit naar een andere bank moet deze namelijk onder bepaalde voorwaarden een vergoeding voor gemiste inkomsten aan de bank betalen. Als de klant verhuist dan hoeft er in het algemeen, volgens de voorwaarden, geen vergoeding betaald te worden. Als er compensatie betaald wordt, dan is dat een extra kasstroom. Er zijn dus twee soorten prepayment: met compensatie en zonder compensatie. Daarom maken we gebruik van een competing risk survival model.

## TOT SLOT

Survivalmodellen worden veel gebruikt bij banken (kredietrisico en prepaymentrisico) en verzekeraars (langlevensrisico). We hebben deze techniek beschreven aan de hand van een voorbeeld uit zijn oorsprong: de biostatistiek. Merk op dat we in dit artikel discrete survivalmodellen hebben bekeken, waarbij de tijd steeds met stappen van één jaar vooruit gaat. De meeste literatuur beschrijft continue survivalmodellen. Meer informatie kan bijvoorbeeld gevonden worden in het zeer toegankelijke boek 'Survival Analysis' van Kleinbaum & Klein (2012). ■