Prediction

The number of patients \( n \) at mid-year \( j, j=2013,\ldots, 2035 \), can be calculated by adding the number of surviving patients from the last year to the newly diagnosed patients, that have also survived until time \( j \). If the German population stayed constant, this could be done by simply adding the number of surviving patients from the last year to the newly diagnosed patients that have also survived until mid-year. However, this does not seem to be realistic. Thus, the number of patients consisted of four subgroups: These were the number of surviving patients that were already in the same group one year earlier, the number of surviving patients that were in an age group below one year earlier, the number of newly diagnosed and surviving patients that were already in the same age group one year earlier, and the number of newly diagnosed and surviving patients that were in the age group below one year earlier. For every mid-year all four subgroups were calculated for men and women and each age group and summed up.

We have to keep in mind that the newly diagnosed patients are at risk for only half a year on average, assuming that new diagnoses of CML occur at a constant rate over time.

\[
n(j) = n(j - 1)S + e(j - 1, j) \left( 1 - \frac{1}{2} (1 - S) \right) \tag{1}
\]

Here \( S \) denotes the one-year survival probability and \( e(j-1,j) \) denotes the number of incident cases in the time between \( j-1 \) and \( j \). Please note that the survival probability \( S \) is assumed to be constant over time and therefore does not have an argument. This simple approach is useful under the requirement that the German population will stay constant in its age composition over the complete prediction time, which is an unrealistic assumption. In the next step, we will therefore introduce age groups to account for the anticipated changes in the population. The number of patients at time \( j \) in age-group \( g \) (\( g=1,\ldots, 8 \)) should further be denoted \( n_g(j) \). Please keep in mind that every model has to be run twice, for male and for female patients. But for simplicity reasons we dispense with an additional index for sex. We calculated \( n_g(j) \) by adding four groups of patients:

- the number of surviving patients that were already in group \( g \) one year earlier
- the number of surviving patients that were in group \( g-1 \) one year earlier
- the number of newly diagnosed and surviving patients that were already in group \( g \) one year earlier
- the number of newly diagnosed and surviving patients that were in group \( g-1 \) one year earlier

Note that for the youngest patients (\( g=1 \)) the second and fourth term do not exist. We have to introduce \( a_{g-1,g}(j-1,j) \) for the number of cases that change age groups from \( g-1 \) to \( g \) between mid-year
$j-1$ and $j$. Additionally, $e_{g-1,g}(j-1,j)$ is the number of newly diagnosed cases that have been in group $g-1$ at time $j-1$ and are in group $g$ at $j$.

$$n_g(j) = (n_g(j-1) - a_{g,g+1}(j-1,j))S_g + a_{g-1,g}(j-1,j)\frac{1}{2}(S_{g-1} + S_g) + e_{g,g}(j-1,j)\left(1 - \frac{1}{2}(1 - S_g)\right) + e_{g-1,g}(j-1,j)\left[\frac{1}{3}\left(1 - \frac{1}{2}(1 - S_{g-1})\right) + \frac{2}{3}\left(1 - \frac{1}{2}(1 - S_g)\right)\right]$$  \hspace{1cm} (2)

While the first and third line concern the patients that did not change age groups and are almost equivalent to model (1) (with the exception that we have to subtract the number of cases that change age groups from $g$ to $g+1$), the second and fourth line need some explanation. The second line describes the already existing cases that were in the younger group one year ago. For these patients, on average the survival probability for age group $g-1$ as well as the one for age group $g$ were valid for half a year each. The newly diagnosed patients that changed age group from $g-1$ to $g$ are described in the fourth line. On average these patients were at risk of death for half a year again. It can be shown that they were in age group $g-1$ for one third of this time and in age group $g$ for two thirds.

In the next step, these parameters had to be estimated. For the starting year 2012, the number of cases was estimated using the age group-specific prevalence and the German population:

$$\hat{n}_g(2012) = \frac{Population_g(2012)}{100,000} * Prevalence_g(2012)$$ \hspace{1cm} (3)

With $n_g(j-1)$, the number of patients changing to an older age group can be estimated by

$$a_{g,g+1}(j-1,j) = \hat{n}_g(j-1)p_{g,g+1}(j-1,j)$$ \hspace{1cm} (4)

where $p_{g-1,g}(j-1,j)$ denotes the ratio of patients changing age group compared to all patients in age group $g-1$, derived from a kernel density estimator.

With the population data and the incidence, $e_{g,g}(j-1,j)$ and $e_{g-1,g}(j-1,j)$ can be estimated:

$$e_{g,g}(j-1,j) = \frac{Population_{g,g}(j-1,j)}{100,000} * Incidence_g$$ \hspace{1cm} (5)

$$e_{g-1,g}(j-1,j) = \frac{Population_{g-1,g}(j-1,j)}{100,000} * \frac{1}{2}(Incidence_{g-1} + Incidence_g)$$ \hspace{1cm} (6)
with \( \text{Population}_{g,g}(j-1,j) \) as the population size that does not change to an older age group and \( \text{Population}_{g-1,g}(j-1,j) \) as the population size that changes to an older age group.

For \( g=1 \) naturally no patients in a younger age group exist, thus \( a_{0,1}(j-1,j) \) and \( e_{0,1}(j-1,j) \) are defined to be zero. Analogously for the oldest age group, \( a_{8,9}(j-1,j) \) is set to zero as well.