0). Three persons wait in line at a bank which has two ATM machines. Suppose the time lengths
needed to do business for the three persons are random variables \((X_1, X_2, X_3)\) with exponential \((\lambda)\)
distribution and are independent.
What is the probability that the 3rd person in line will finish the second?

1). Use a discrete distribution (with 4-point mass) to verify the two discrete formula connecting the
CDF \((F)\) and cumulative hazard function \((H)\).
Define
\[
H(t) = \sum_{s \leq t} \frac{\Delta F(s)}{1 - F(s-)}
\]
then we must have
\[
1 - F(t) = \prod_{s \leq t} (1 - \Delta H(s))
\]

2). If \(X\) has hazard \(h_1(t)\); \(Y\) has hazard \(h_2(t)\) and are independent. Show \(\min(X,Y)\) has hazard
\(h_1(t) + h_2(t)\).

Homework 2

3).
(a) for the data set Recidivism, fit the weibull regression model of \(\log(\text{week})\) to covariates \(\text{FIN, AGE, RACE, WEXP, MAR, PARO, PRIO}\).

The data set can be downloaded from SAS company site:
http://ftp.sas.com/samples/A55233

Use both R and SAS to carry out the fit and compare the results.

(b) obtain the 90% confidence interval for the regression coefficient for \(\text{AGE}\) by way of log likelihood
ratio.

Compare the confidence interval with the Wald confidence interval.

(4) Suppose \(X_1, \ldots, X_n\) are iid random variables from a piecewise exponential distribution with 4
pieces.

Please compute the (observed) Fisher information matrix for the 4 parameters.

Also compute the inverse of the above matrix.

What change to the above results are needed if we instead have censored sample \(T_1, \delta_1, \ldots, T_n, \delta_n\)
with \(T_i = \min(X_i, C_i)\) and \(\delta_i = I_{[X_i \leq C_i]}\)?
(5) Simplify the Kaplan-Meier estimator and Greenwood formular for estimating $F(t)$ for a fixed $t$ when the $n$ observations are all uncensored.

Identify with the familiar estimation of binomial probability.

(6) Plot the three cumulative hazard function estimators (Nelson-Aalen) and three survival function estimators (Kaplan-Meier) with the data set colon. One for each treatment type. Use only the data with etype=2. You need time and status; ignor age and sex etc.

The data set colon is available from within R. load survival package first.