The Nelson-Aalen, Kaplan-Meier estimator is MLE:

Let $(x_1, \delta_1), (x_2, \delta_2), \cdots, (x_n, \delta_n)$ be the i.i.d. right censored observations as defined in class.

The likelihood pertaining to $F(\cdot)$ based on censored samples is

$$L(F) = \prod_{\delta_i=1} \Delta F(x_i) \prod_{\delta_i=0} (1 - F(x_i)).$$

Because the survival function $(1-F)$ and hazard function (Λ) are mathematically equivalent, inference for one can be obtained though a transformation of the other.

For discrete CDF, we have

$$\begin{cases} \Delta F(x_i) & = \Delta\Lambda(x_i) \prod_{j:x_i<x_j} (1 - \Delta\Lambda(x_j)) \\ 1 - F(x_i) & = \prod_{j:x_j \le x_i} (1 - \Delta\Lambda(x_j)) \end{cases},$$

therefore, while assuming F is discrete,

$$L = \prod_{i=1}^{n} \left\{ (\Delta\Lambda(x_i))^{\delta_i} (\prod_{j:x_j<x_i} (1 - \Delta\Lambda(x_j)))^{\delta_i} (\prod_{j:x_j \le x_i} (1 - \Delta\Lambda(x_j)))^{1-\delta_i} \right\}.$$

Let $x_{k,n}$ be the m distinctive and ordered values of the x-sample above, where $k = 1 \cdots m$ and $m \le n$. Let us denote

$$N_k = \sum_{i:x_i=x_{k,n}} \delta_i \qquad \text{and} \qquad Y_k = \sum_{i=1}^{n} I(x_i \ge x_{k,n}).$$

Then, L becomes

$$L = \prod_{k=1}^{m} (\Delta\Lambda(x_{k,n}))^{N_k} (1 - \Delta\Lambda(x_{k,n}))^{(Y_k - N_k)}. \tag{1}$$

From this, we can easily get the log likelihood function. Maximize the log likelihood function wrt $\Delta\Lambda(x_{k,n})$ gives the maximum likelihood estimate of $\Delta\Lambda(x_{k,n})$ which is the Nelson Aalen estimator,

$$\Delta\hat{\Lambda}(x_{k,n}) = \frac{N_k}{Y_k}.$$

By the invariance property of the MLE this imply that the Kaplan-Meier estimator

$$1 - \hat{F}(t) = \prod_{x_{k,n} \le t} (1 - \frac{N_k}{Y_k})$$

is MLE of $1 - F(t)$ (i.e. it maximizes the likelihood $L(F)$ above.)

If we use a 'wrong' formula connecting the CDF and cumulative hazard: $1 - F(t) = \exp(-\Lambda(t))$, then the likelihood would be

$$AL = \prod_{i=1}^{n} (\Delta\Lambda(x_i))^{\delta_i} \exp\{-\Lambda(x_i)\} = \prod_{i=1}^{n} (\Delta\Lambda(x_i))^{\delta_i} \exp\{-\sum_{j:x_j \le x_i} \Delta\Lambda(x_j)\}.$$

If we let $w_i = \Delta\Lambda(x_i)$ then

$$\log AL = \sum_{i=1}^{n} \delta_i \log w_i - \sum_{i=1}^{n} \sum_{j:x_j \le x_i} w_j \ .$$

It is worth noting that among all the cumulative hazard functions, the Nelson-Aalen estimator also maximizes the $\log AL$, (this can be verified easily by taking derivative).

Self-Consistent.

Given censored observations $(x_1, \delta_1), \cdots, (x_n, \delta_n)$ where $x_j = \min(t_j, c_j)$, and $\delta_j = I_{[x_j = t_j]}$. An estimator of the CDF of t_j's based on censored observations is self-consistent if

$$F(t) = \frac{1}{n} \sum_{j=1}^{n} E_F\{I_{[t_j \le s]} | x_j, \delta_j\} = \frac{1}{n} \sum_{j=1}^{n} \left[\delta_j I_{[x_j \le t]} + (1 - \delta_j) E_F\{I_{[t_j \le t]} | x_j, \delta_j = 0\} \right]$$

It states that if we replace the censored observations by its conditional expectation (computed with F), then the empirical distribution based on those should also be F.

Fact: the Kaplan-Meier estimator for right censored observations is self-consistent.

Notice F appears on both side of the equation, this leads to iterative equation – plug the old estimator of F on RHS and you get a new estimator on the LHS.

The Kaplan-Meier estimator can be computed by this iteration, starting from an empirical distribution based on all uncensored observations in the sample as F_0.

Our interest in this property is largely due to its nice interpretation and potential to apply to other similar cases.