Written problems to be handed in Friday, March 9, 2012.

1. Suppose $X_1, X_2, \ldots$ are independent Poisson random variables with $X_i \sim \mathcal{P}(\lambda)$. Verify that this is a "regular" problem for inference about λ, including that the score statistic is consistent for testing $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda \neq \lambda_0$.

Ans: We have $\theta = \lambda$ and parameter space $\Theta = (0, \infty)$. Thus we consider the necessary assumptions:

- *Density:* By supposition, $X_1, X_2, \ldots$ are independent and identically distributed, where for $\lambda > 0$

$$f(x;\lambda) \equiv Pr(X_i = x \,|\, \lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & x = 0, 1, 2, 3, \ldots, \\ 0 & \text{else} \end{cases}$$

  is a density with respect to the σ-finite counting measure, with $f(x;\lambda)$ satisfying $Pr(X_i = x; \lambda) \geq 0$ and $\sum_{x=0}^{\infty} Pr(X_i = x; \lambda) = 1$.

- *Common support:* From the above density it is clear that $Pr(X_i = x \,|\, \lambda) > 0)$ if and only if $x \in \{0, 1, 2, 3, \ldots\}$, the whole numbers, irrespective of the specific choice of $\lambda > 0$.

- *Identifiability:* Suppose that for $\lambda_1 > 0$ and $\lambda_2 > 0$ we have $\forall x \in (-\infty, \infty)$ $Pr(X_i = x \,|\, \lambda_1) = Pr(X_i = x \,|\, \lambda_2)$. Then,

$$Pr(X_i = 0 \,|\, \lambda_1) = Pr(X_i = 0 \,|\, \lambda_2) \quad \Rightarrow \quad e^{-\lambda_1} = e^{-\lambda_2} \quad \Rightarrow \quad \lambda_1 = \lambda_2.$$

  Furthermore, for $\theta_1 \neq \theta_2$, $Pr(X_i = 0 \,|\, \lambda_1) = e^{-\lambda_1} \neq e^{-\lambda_2} = Pr(X_i = 0 \,|\, \lambda_2)$, so the probability distribution $P_{\lambda_1} \neq P_{\lambda_2}$.

- *Continuous differentiability of log likelihood:* We have likelihood function $L_n(\lambda \,|\, \vec{X})$ defined for $\lambda > 0$ as

$$L_n(\lambda \,|\, \vec{X}) = \prod_{i=1}^{n} f(X_i;\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{X_i}}{X_i!} = \frac{e^{-n\lambda}\lambda^{n\overline{X}_n i}}{\prod_{i=1}^{n} X_i!},$$

  where $\overline{X}_n = \sum_{i=1}^{n} X_i/n$. We thus derive log likelihood function and its first three derivatives as

$$\mathcal{L}_n(\lambda \,|\, \vec{X}) = \log\left(L_n(\lambda \,|\, \vec{X})\right) = -n\lambda + n\overline{X}_n \log(\lambda) - \sum_{i=1}^{n} \log(X_i!)$$

$$\mathcal{U}_n(\lambda \,|\, \vec{X}) = \frac{\partial}{\partial\lambda}\mathcal{L}_n(\lambda \,|\, \vec{X}) = -n + \frac{n\overline{X}_n}{\lambda}$$

$$\frac{\partial^2}{\partial\lambda^2}\mathcal{L}_n(\lambda \,|\, \vec{X}) = \frac{\partial}{\partial\lambda}\mathcal{U}_n(\lambda \,|\, \vec{X}) = -\frac{n\overline{X}_n}{\lambda^2}$$

  From the above, by inspection we have that the log likelihood function $\mathcal{L}_n(\lambda \,|\, \vec{X})$ is twice continuously differentiable with respect to $\theta = \lambda$ for all $\lambda \in \Theta$, an open set.

- *Distribution of score function:* Noting that by independence and identical distribution of the X_i's, $E[\overline{X}_n] = \lambda$, $Var(\overline{X}_n) = \lambda/n$, and $E[(\overline{X}_n)^2] = Var(\overline{X}_n) + E^2[\overline{X}_n] = \lambda/n + \lambda^2$, we thus have

$$E\left[\mathcal{U}_n(\lambda \,|\, \vec{X})\right] = E\left[-n + \frac{n\overline{X}_n}{\lambda}\right] = 0$$

$$Var\left(\mathcal{U}_n(\lambda \,|\, \vec{X})\right) = Var\left(-n + \frac{n\overline{X}_n}{\lambda}\right) = \frac{n^2}{\lambda^2}\frac{\lambda}{n} = \frac{n}{\lambda}$$

$$-E\left[\frac{\partial^2}{\partial\lambda^2}\mathcal{L}_n(\lambda \,|\, \vec{X})\right] = -E\left[-\frac{n\overline{X}_n}{\lambda^2}\right] = \frac{n}{\lambda}$$

so the score function is distributed with mean 0 and finite variance, and the negative expectation of the derivative of the score function is positive definite (and equal to the score function's variance).

- *Boundedness of the third derivatives of the log density:* The third derivative of the log density is

$$\frac{\partial^3}{\partial\lambda^3}\mathcal{L}_1(\lambda \mid X_i) = \frac{2X_i}{\lambda^3}.$$

Letting $\theta_0 = \lambda_0 \in \Theta = (0, \infty)$ be the true value of λ, and defining $\Theta_0 = (\lambda_0 - \epsilon, \lambda_0 + \epsilon) \subset \Theta$ for some fixed $\epsilon > 0$, we have

$$\forall\lambda \in \Theta_0 \quad \left|\frac{\partial^3}{\partial\lambda^3}\mathcal{L}_1(\lambda \mid X_i)\right| = \frac{2X_i}{\lambda^3} \leq M(X_i) = \frac{2X_i}{(\lambda_0 - \epsilon)^3}.$$

Furthermore, under $\lambda = \lambda_0$,

$$E_0\left[M(X_i)\right] = \frac{2\lambda_0}{(\lambda_0 - \epsilon)^3} < \infty.$$

2. Suppose $X_1, X_2, \ldots$ are independent Poisson random variables with $X_i \sim \mathcal{P}(\lambda t_i)$ for known scalars t_i and $Y_1, Y_2, \ldots$ are independent Poisson random variables with $Y_i \sim \mathcal{P}(\omega u_i)$ for known scalars u_i. We are interested in testing the rate ratio $\theta = \lambda/\omega$ using samples X_i for $i = 1, \ldots, n$ and Y_j for $j = 1, \ldots, m$ where $m/n \to r > 0$ as $n \to \infty$.

a. Show that likelihood inference about θ can be based on the statistics $S_n = \sum_{i=1}^n X_i$ and $T_m = \sum_{j=1}^m Y_j$.

Ans: Solution to problem 2 is made easier by defining $t_n^* = \sum_{i=1}^n t_i$ and $u_m^* = \sum_{i=1}^m u_i$ for notational convenience and considering inference for $\vec{\phi} = (\theta \quad \omega)$ where $\lambda = \theta\omega$. The likelihood and log likelihood functions for $\vec{\phi}$ are then

$$\begin{aligned}
L_{n,m}(\vec{\phi} \mid \vec{X}, \vec{Y}) &= \prod_{i=1}^n \frac{e^{-\theta\omega t_i}(\theta\omega t_i)^{X_i}}{X_i!} \prod_{j=1}^m \frac{e^{-\omega u_i}(\omega u_i)^{Y_i}}{Y_i!} \\
&= e^{-\theta\omega t_n^*}(\theta\omega)^{S_n} e^{-\omega u_m^*}\lambda^{T_m} h(\vec{X}, \vec{t}, \vec{Y}, \vec{u}) \\
\mathcal{L}_{n,m}(\vec{\phi} \mid \vec{X}, \vec{Y}) &= \log(L_{n,m}(\vec{\phi} \mid \vec{X}, \vec{Y})) \\
&= -\theta\omega t_n^* + S_n \log(\theta\omega) - \omega u_m^* + T_m \log(\omega) + \log(h(\vec{X}, \vec{t}, \vec{Y}, \vec{u})) \\
&= -\omega(\theta t_n^* + u_m^*) + S_n \log(\theta) + (S_n + T_m)\log(\omega) + \log(h(\vec{X}, \vec{t}, \vec{Y}, \vec{u}))
\end{aligned}$$

where we define . Hence, by factorization, (S_n, T_m) is sufficient for $\vec{\phi}$, and the score function and likelihood ratio statistics involve $\vec{X}$ and $\vec{Y}$ through (S_n, T_m):

$$\begin{aligned}
\mathcal{U}(\vec{\phi}) &= \frac{\partial}{\partial\vec{\phi}}\left[-\omega(\theta t_n^* + u_m^*) + S_n \log(\theta) + (S_n + T_m)\log(\omega) + \log(h(\vec{X}, \vec{t}, \vec{Y}, \vec{u}))\right] \\
&= \frac{\partial}{\partial\vec{\phi}}\left[-\omega(\theta t_n^* + u_m^*) + S_n \log(\theta) + (S_n + T_m)\log(\omega)\right] \\
2\left(\mathcal{L}_{n,m}(\tilde{\vec{\phi}}) - \mathcal{L}_{n,m}(\vec{\phi}_0)\right) &= 2\Big[\Big(-\tilde{\omega}(\theta t_n^* + u_m^*) + S_n \log(\tilde{\theta}) + (S_n + T_m)\log(\tilde{\omega}) + \log(h(\vec{X}, \vec{t}, \vec{Y}, \vec{u}))\Big) - \\
&\quad \Big(-\omega_0(\theta_0 t_n^* + u_m^*) + S_n \log(\theta_0) + (S_n + T_m)\log(\omega_0) + \log(h(\vec{X}, \vec{t}, \vec{Y}, \vec{u}))\Big)\Big] \\
&= 2\Big[\Big(-\tilde{\omega}(\theta t_n^* + u_m^*) + S_n \log(\tilde{\theta}) + (S_n + T_m)\log(\tilde{\omega})\Big) - \\
&\quad (-\omega_0(\theta_0 t_n^* + u_m^*) + S_n \log(\theta_0) + (S_n + T_m)\log(\omega_0))\Big]
\end{aligned}$$

and as the information matrix is just the negative expectation of the partial derivative of the score function, it similarly will depend on $\vec{X}$ and $\vec{Y}$ only through (S_n, T_m).

b. Derive general expressions for the score, Wald, and likelihood ratio statistics for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$.

Ans: Using the expression for the score function given above, we derive score functions as

$$\begin{aligned}
\mathcal{U}_1(\vec{\phi}) &= \frac{\partial}{\partial \theta} \left[-\omega(\theta t_n^* + u_m^*) + S_n \log(\theta) + (S_n + T_m)\log(\omega) \right] \\
&= -\omega t_n^* + \frac{S_n}{\theta} \\
\mathcal{U}_2(\vec{\phi}) &= \frac{\partial}{\partial \omega} \left[-\omega(\theta t_n^* + u_m^*) + S_n \log(\theta) + (S_n + T_m)\log(\omega) \right] \\
&= -(\theta t_n^* + u_m^*) + \frac{S_n + T_m}{\omega}
\end{aligned}$$

which lead to efficient likelihood estimates for the unrestricted case of

$$\begin{aligned}
-\tilde{\omega}_{m,n} t_n^* + \frac{S_n}{\tilde{\theta}_{m,n}} \equiv 0 \qquad &\Rightarrow \qquad \tilde{\theta}_{m,n} = \frac{S_n}{t_n^* \tilde{\omega}_{m,n}} \\
-(\tilde{\theta}_{m,n} t_n^* + u_m^*) + \frac{S_n + T_m}{\tilde{\omega}_{m,n}} \equiv 0 \qquad &\Rightarrow \qquad \tilde{\omega}_{m,n} = \frac{T_m}{u_m^*} \qquad \Rightarrow \qquad \tilde{\theta}_{m,n} = \frac{S_n u_m^*}{T_m t_n^*}.
\end{aligned}$$

Under the null hypothesis $H_0 : \theta = \theta_0$, the efficient likelihood estimate $\tilde{\omega}_{m,n}^{(0)}$ of ω is found from the likelihood equations as

$$-(\theta_0 t_n^* + u_m^*) + \frac{S_n + T_m}{\tilde{\omega}_{m,n}^{(0)}} \equiv 0 \qquad \Rightarrow \qquad \tilde{\omega}_{m,n}^{(0)} = \frac{S_n + T_m}{\theta_0 t_n^* + u_m^*}.$$

The information matrix is derived as

$$\begin{aligned}
\mathcal{I}_{11}(\vec{\phi}) &= -E\left[\frac{\partial}{\partial \theta}\mathcal{U}_1(\vec{\phi})\right] = \frac{E[S_n]}{\theta^2} = \frac{t_n^* \omega}{\theta} \\
\mathcal{I}_{12}(\vec{\phi}) &= -E\left[\frac{\partial}{\partial \theta}\mathcal{U}_2(\vec{\phi})\right] = t_n^* \\
\mathcal{I}_{21}(\vec{\phi}) &= -E\left[\frac{\partial}{\partial \omega}\mathcal{U}_1(\vec{\phi})\right] = t_n^* \\
\mathcal{I}_{22}(\vec{\phi}) &= -E\left[\frac{\partial}{\partial \omega}\mathcal{U}_2(\vec{\phi})\right] = \frac{E[S_n + T_m]}{\omega^2} = \frac{t_n^* \theta + u_m^*}{\omega}
\end{aligned}$$

The inverse information matrix is thus

$$\mathcal{I}^{-1}(\vec{\phi}) = \begin{pmatrix} \frac{\theta(\theta t_n^* + u_m^*)}{\omega t_n^* u_m^*} & -\frac{\theta}{u_m^*} \\ -\frac{\theta}{u_m^*} & \frac{\omega}{u_m^*} \end{pmatrix}.$$

Using $\hat{\mathcal{I}}(\vec{\phi}) = \mathcal{I}(\tilde{\vec{\phi}})$ as a consistent estimator of the information we find the Wald statistic is

$$W_n = (\tilde{\theta}_n - \theta_0)^2 \frac{t_n^* u_m^*}{\tilde{\theta}(\tilde{\theta} t_n^* + u_m^*)},$$

the score statistic is

$$R_n = \mathcal{U}^T(\tilde{\vec{\phi}}^{(0)}) \mathcal{I}^{-1}(\tilde{\vec{\phi}}^{(0)}) \mathcal{U}(\tilde{\vec{\phi}}^{(0)}),$$

and the likelihood ratio statistic is

$$2(\mathcal{L}_{n,m}(\tilde{\vec{\phi}} \mid \vec{X}, \vec{Y}) - \mathcal{L}_{n,m}(\tilde{\vec{\phi}}^{(0)} \mid \vec{X}, \vec{Y})).$$

c. Provide expressions for the score, Wald, and likelihood ratio statistics (and their critical values) for a level α test of H_0 for the special case of $\theta_0 = 1$.

Ans: The distribution of the test statistics is χ^2_1, though the distributional theory derived in class has to be modified to be able to argue this. Essentially,

- The score contribution from each observation still has mean 0 and a variance that depends on the information for the respective groups.
- Providing the Lindeberg condition holds, the Lindeberg-Feller CLT can then be used to show that

$$\mathcal{I}^{-1/2}(\tilde{\vec{\phi}}^{(0)})\mathcal{U}(\tilde{\vec{\phi}}^{(0)} \to_d \mathcal{N}(0,1).$$

- Establishing the Lindeberg condition will involve constraints on the sampling distribution similar to those used in problem 4b of Homework 2.

Under these conditions, the critical value will then be $\chi^2_{1,1-\alpha}$.

d. Suppose we observe $S_{100} = 335$ for $\sum_{i=1}^{100} t_i = 52$ and $T_{50} = 160$ for $\sum_{i=1}^{50} u_i = 28$. Derive the p value for testing null hypotheses with the score, Wald, and likelihood ratio statisics for the cases that $\theta_0 = 1$, $\theta_0 = 1.3$, and $\theta_0 = 0.8$. (Be sure to provide efficient likelihood estimates of θ overall and under the null hypothesis in each case).

3. Let $X_1, X_2, \ldots$ be independent random variables distributed according to a mixture of normals in which $X_i \sim f_X(x)$ where density f_X is derived from a mixture having probability $1-p$ of a $\mathcal{N}(\mu, \sigma^2)$ distribution and probability p of a $\mathcal{N}(\mu+\delta, \sigma^2)$ distribution, where σ^2 is known and μ and δ are unknown.

a. Suppose p is known. Derive the score, Wald, and likelihood ratio statistics for testing $H_0 : \delta = 0$ versus $H_1 : \delta \neq 0$. What are the asymptotic distributions of these statistics? Justify your answer.

Ans. Presuming that $p > 0$, the likelihood and log likelihood functions are

$$L(\theta|\vec{X}) = \prod_{i=1}^n \left[(1-p)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)}{2\sigma^2}^2\right) + p\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu-\delta)^2}{2\sigma^2}\right)\right]$$
$$\mathcal{L}(\theta|\vec{X}) = \sum_{i=1}^n \log\left[(1-p)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)}{2\sigma^2}^2\right) + p\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu-\delta)^2}{2\sigma^2}\right)\right].$$

Note that maximizing the likelihood under the null hypothesis is easily seen to correspond to the choice $\tilde{\mu}^{(0)} = \overline{X}_n$. Maximization of the likelihood more generally is a bit more involved. The score functions are

$$\mathcal{U}_1(\theta) = \frac{\partial}{\partial\delta}\mathcal{L}(\theta|\vec{X}) = \sum_{i=1}^n \frac{p\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(X_i-\mu-\delta)^2}{2\sigma^2}\right)\frac{(X_i-\mu-\delta)}{\sigma^2}}{\left[(1-p)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)}{2\sigma^2}^2\right) + p\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(X_i-\mu-\delta)^2}{2\sigma^2}\right)\right]}$$
$$\mathcal{U}_2(\theta) = \frac{\partial}{\partial\mu}\mathcal{L}(\theta|\vec{X}) = \sum_{i=1}^n \frac{p\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(X_i-\mu-\delta)^2}{2\sigma^2}\right)\frac{(X_i-\mu-\delta)}{\sigma^2} + (1-p)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right)\frac{(X_i-\mu)}{\sigma^2}}{\left[(1-p)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(X_i-\mu)}{2\sigma^2}^2\right) + p\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(X_i-\mu-\delta)^2}{2\sigma^2}\right)\right]}$$

These are not immediately solvable in closed form, though Newton-Raphson can be used. FInding the information matrix via the second derivative would be a similarly complicated equation, though having found an expression for the score, the information can be estimated using the outer product of the score vectors:

$$\hat{\mathcal{I}}(\theta) = \sum_{i=1}^n \frac{\partial}{\partial\theta}\mathcal{L}(\theta|X_i)\frac{\partial}{\partial\theta}^T\mathcal{L}(\theta|X_i).$$

This approach will tend to yield less stable estimates of the information than using the sum of the second partials, but is easily done in this setting. Note that for $\tilde{\theta}^{(0)} = (0 \quad \overline{X}_n)^T$,

$$\hat{\mathcal{I}}_{11}(\tilde{\theta}^{(0)}) = \sum_{i=1}^n p^2 \frac{(X_i - \overline{X}_n)^2}{\sigma^4}$$
$$\hat{\mathcal{I}}_{12}(\tilde{\theta}^{(0)}) = \sum_{i=1}^n \frac{(X_i - \overline{X}_n)^2}{\sigma^4}$$
$$\hat{\mathcal{I}}_{21}(\tilde{\theta}^{(0)}) = \sum_{i=1}^n \frac{(X_i - \overline{X}_n)^2}{\sigma^4}$$
$$\hat{\mathcal{I}}_{22}(\tilde{\theta}^{(0)}) = \sum_{i=1}^n \frac{(X_i - \overline{X}_n)^2}{\sigma^4}.$$

The Wald, score, and likelihood ratio statistics can then be expressed using the numerical results. (It is not an uncommon situation that

b. Suppose p is unknown. Derive the score, Wald, and likelihood ratio statistics for testing $H_0 : \delta = 0$ versus $H_1 : \delta \neq 0$. What are the asymptotic distributions of these statistics? Justify your answer.

Ans: The key point of this problem is that when p is unknown, then δ is nonestimable when $p = 0$, and p is nonestimable when $\delta = 0$. This is therefore not a regular problem, because our parameter vector is not identifiable. We could still compute a likelihood ratio statistic, because the likelihood is the same under $\delta = 0$ for all p. The distribution of the likelihood ratio statistic is not expected to be well-behaved, however, owing to the poor behavior of the estimates for p and δ with distributions near the null.

4. Consider the censored data setting in which independent, identically distributed random variables $T_i \sim F$ for $i = 1, \ldots, n$ are subject to censoring by known C_i which are independent of the T_i's. Hence, we observe only $Y_i = \min(T_i, C_i)$ and $\delta_i = 1_{[T_i = Y_i]}$. Find the nonparametric maximum likelihood estimator of F. Does the theory about regular asymptotic theory apply? (Hint: Argue that the likelihood of the data will be maximized by a discrete distribution function, and consider estimates of the hazard function

$$\lambda(t) = \lim_{\epsilon \to 0} \frac{P(T > t + \epsilon \,|\, T \geq t)}{\epsilon},$$

using the fact that the T_i's are independent of the C_i's.)

Ans: A solution in the more general case where the C_i's are random, but independent of the T_i's.

Let $T_i \sim F$ and $C_i \sim G$ be totally independent for $i = 1, \ldots, n$. Define (Y_i, δ_i) as above, and let

$$\begin{aligned} Y_i \sim H(y) \equiv Pr(Y_i \leq y) &= Pr(Y_i \leq y, \delta_i = 1) + Pr(Y_i \leq y, \delta_i = 0) \\ &= Pr(T_i \leq y, T_i \leq C_i) + Pr(C_i \leq y, C_i < T_i) \\ &= \int_0^y \left(1 - G(y^-)\right) dF(y) + \int_0^y \left(1 - F(y)\right) dG(y) \\ &\equiv \tilde{F}(y) + \tilde{G}(y). \end{aligned}$$

Now by independence, $1 - H(y) = (1 - F(y))\,(1 - G(y))$ and we can define the cumulative hazard function for T as

$$\Lambda(t) \equiv \int_0^t \frac{dF(u)}{1 - F(u^-)} = \int_0^t \frac{(1 - G(u^-))\, dF(u)}{(1 - G(u^-))\,(1 - F(u^-))} = \int_0^t \frac{d\tilde{F}(u)}{1 - H(u^-)}.$$

The nonparametric MLE (NPMLE) is then

$$\widehat{\Lambda_n}(t) \equiv \int_0^t \frac{d\widehat{\tilde{F}_n}(u)}{1 - \widehat{H}_n(u^-)},$$

and noting that we have the greatest likelihood by placing all mass at the observed data, we use the empirical distribution functions

$$\widehat{H}_n(t) \equiv \frac{1}{n}\sum_{i=1}^{n} 1_{[Y_i \le t]} \quad \text{and} \quad \widehat{\tilde{F}}_n(t) \equiv \frac{1}{n}\sum_{i=1}^{n} 1_{[Y_i \le t]}\delta_i.$$

Inversion of $\Lambda(t)$ yields

$$1 - F(t) = \exp\left(-\Lambda_c(t)\right) \prod_{u \le t} \left(1 - \Delta\Lambda(u)\right),$$

where

$$\begin{aligned} \Delta\Lambda(u) &\equiv \Lambda(u) - \Lambda(u^-) \\ \Lambda_c(t) &\equiv \Lambda(t) - \sum_{u \le t} \Delta\Lambda(u). \end{aligned}$$

Now

$$\Delta\Lambda(u) = \frac{\Delta F(u)}{1 - F(u^-)} = \frac{F(u) - F(u^-)}{1 - F(u^-)} \Rightarrow 1 - \Delta\Lambda(u) = \frac{1 - F(u)}{1 - F(u^-)}.$$

Because $\widehat{\Lambda_n}$ is purely a jump function, and for ordered observations $(Y_{n:i}, \delta_{n:i})$ such that $Y_{n:1} \le \cdots \le Y_{n:n}$,

$$\begin{aligned} \Delta\widehat{\Lambda_n}(u) &= \frac{\Delta\widehat{\tilde{F}_n}(u)}{1 - \widehat{H_n}(u^-)} \\ &= \frac{\delta_{n:i}/n}{1 - (i-1)/n} \quad \text{for } u = Y_{n:i}. \end{aligned}$$

So

$$\begin{aligned} 1 - \widehat{\tilde{F}_n}(t) &= \prod_{u \le t} \left(1 - \Delta\widehat{\Lambda_n}(u)\right) \\ &= \prod_{i: Y_{n:i} \le t} \left(1 - \frac{\delta_{n:i}}{n - i + 1}\right) \\ &= \prod_{i: Y_{n:i} \le t} \left(\frac{n-i}{n-i+1}\right)^{\delta_{n:i}}. \end{aligned}$$