Bêta Poisson.

$I(x) = \frac{\lambda'}{G} e^{-\lambda' x}$ $\qquad I(\ell) = \lambda e^{-\lambda \ell}$ $\qquad \left| \begin{array}{l} \lambda = \frac{\lambda'}{G} \\ \cancel{x} = \frac{\ell}{G} \end{array} \right.$

Densité de distribution de longueurs de lectures pdf $I(\ell) = \lambda e^{-\lambda \ell}$

Si on suppose que le séquenceur a une probabilité p de s'arrêter pour chaque nucléotide, la longueur du fragment ℓ traité par la machine est $(1-p)^{\ell'} p \qquad \ell' \leq \ell$

$$f_{L'}(\ell') = \int_0^G (1-p)^{\ell'} p \; \lambda e^{-\lambda \ell} d\ell \qquad \begin{cases} (1-p)^{\ell'} p & \text{si } \ell' \leq \ell \\ 0 & \text{autrement} \end{cases}$$

$$f_{L'}(\ell') = \int_0^G (1-p)^{\ell'} p \; \mathbb{1}\{0 \leq \ell' \leq \ell\} \; \lambda e^{-\lambda \ell} d\ell$$

Condition $\mathbb{1}\{0 \leq \ell' \leq \ell\}$ implique $\ell \geq \ell'$.

$$f_{L'}(\ell') = \int_{\ell'}^G (1-p)^{\ell'} p \; \lambda e^{-\lambda \ell} d\ell = \lambda p (1-p)^{\ell'} \left[-\frac{1}{\lambda} e^{-\lambda \ell} \right]_{\ell'}^G$$

$$= p(1-p)^{\ell'} \left[e^{-\lambda \ell'} - e^{-\lambda G} \right]$$

Si on supprime les fragments ayant une longueur $\leq L_{min}$

$$f_{L'}(\ell') = \int_0^G (1-p)^{\ell'} p \; \mathbb{1}\{0 \leq \ell' \leq \ell \text{ et } \ell \geq L_{min}\} \; \lambda e^{-\lambda \ell} d\ell$$

$\mathbb{1}\{\ell \geq \ell' \text{ et } \ell \geq L_{min}\}$

$\mathbb{1}\{\ell \geq \max(L_{min}, \ell')\}$

$$f_{L'}(\ell') = \int_{\max(\ell', L_{min})}^G (1-p)^{\ell'} p \; \lambda e^{-\lambda \ell} d\ell$$

$$f_{L'}(\ell') = p(1-p)^{\ell'} \left[e^{-\lambda \ell} \right]_G^{\max(\ell', L_{min})}$$

$$f_{L'}(\ell') = \begin{cases} \left[e^{-\lambda L_{min}} - e^{-\lambda G} \right] p(1-p)^{\ell'} & 0 \leq \ell' \leq L_{min} \\ \left[e^{-\lambda \ell'} - e^{-\lambda G} \right] p(1-p)^{\ell'} & L_{min} < \ell' \end{cases}$$

On n'obtient pas de courbe croissante pour longueurs ℓ' petites comme PacBio

Two parameters Weibull

$$\begin{cases} f_W(t) = \dfrac{\gamma}{\alpha}\left(\dfrac{t}{\alpha}\right)^{\gamma-1} \exp\left(-\left(\dfrac{t}{\alpha}\right)^{\gamma}\right) & t \geq 0 \\ F_W(t) = 1 - e^{-(t/\alpha)^{\gamma}} \end{cases}$$

α scale parameter
characteristic life par

γ shape parameter

Often used in failure analysis

the value of the shape parameter γ determines the failure rate

- if $\gamma < 1$ the failure rate decreases with time
- if $\gamma = 1$ the failure rate is constant and Weibull $\rightarrow$ exponential
- if $\gamma > 1$ the failure rate increases with time (the distribution models wear-out failures which tend to increase with time)

Weibull extreme value distribution of type $\underline{III}$

Weibull reliability function : $R(t) = e^{-\left(\frac{t}{\alpha}\right)^{\gamma}}$ = survival function

Weibull failure rate function ; $\lambda(t) = \dfrac{f(t)}{R(t)} = \dfrac{\gamma}{\alpha}\left(\dfrac{t}{\alpha}\right)^{\gamma-1}$

Weibull mean life (MTTF) : $\langle t \rangle = \alpha\, \Gamma\left(\dfrac{1}{\gamma} + 1\right)$ $\Gamma(*)$ gamma function

$\Gamma(n) = \int_0^{\infty} e^{-x} x^{n-1}\, dx$

If we use Weibull function with $\gamma > 1$ instead of geometric function

$$f_{L'}(\ell') = \int_0^G \frac{\gamma}{\alpha}\left(\frac{\ell'}{\alpha}\right)^{\gamma-1} e^{-\left(\frac{\ell'}{\alpha}\right)^{\gamma}} \mathbb{I}\{\ell \geq \max(L_{min}, \ell')\}\, \lambda e^{-\lambda \ell}\, d\ell$$

$$f_{L'}(\ell'; \alpha, \gamma, \lambda) = \frac{\gamma}{\alpha}\left(\frac{\ell'}{\alpha}\right)^{\gamma-1} e^{-\left(\frac{\ell'}{\alpha}\right)^{\gamma}} \left[-e^{-\lambda \ell}\right]_{\max(L_{min}, \ell')}^{G}$$

$$f_{L'}(\ell'; \alpha, \gamma, \lambda) = \begin{cases} e^{-\lambda L_{min}} \dfrac{\gamma}{\alpha}\left(\dfrac{\ell'}{\alpha}\right)^{\gamma-1} e^{-\left(\frac{\ell'}{\alpha}\right)^{\gamma}} & \ell' \leq L_{min} \\ \dfrac{\gamma}{\alpha}\left(\dfrac{\ell'}{\alpha}\right)^{\gamma-1} e^{-\left(\frac{\ell'}{\alpha}\right)^{\gamma}} e^{-\lambda \ell'} & \ell > L_{min} \end{cases}$$

$e^{-\lambda G}$ is negligible

Maximum likelihood estimate of parameters α, γ, λ.