Weak Law for Random Functions.

- $X_1\ X_2\ X_3 \ldots$ iid K — compact set in $\mathbb{R}^p$.

  $W_i(t) = h(t, X_i)$ $t \in K$ $h(t.x)$ — continuous function of t for all x.

  $W_1, W_2, \cdots$ iid random functions taking values in $C(K)$.

- $w \in C(K)$: $\|w\|_\infty = \sup_{t\in K} |w(t)|$.

  w_n ~~converges~~ to w in this norm if $\|w_n - w\|_\infty \to 0$.

  $C(K)$ is complete (all Cauchy sequences converge)

  $C(K)$ is separable (dense. countable dense subset)

Lemma: W — a random function in $C(K)$. define $\mu(t) = EW(t)$ $t \in K$. if $E\|W\|_\infty < \infty$. then μ is continuous. Also $\sup_{t\in K} E \sup_{s:\|t-s\|<\varepsilon} |W(s) - W(t)| \to 0$ as $\varepsilon \downarrow 0$.

pf. $t_n \in K$. $t_n \to t$. $W(t_n) \to W(t)$

They are dominated by $\|W\|_\infty$. and $E\|W\|_\infty < \infty$

$\mu(t_n) = EW(t_n) \to EW(t) = \mu(t)$.

Define $M_\varepsilon(t) = \sup_{s:\|s-t\|<\varepsilon} |W(s) - W(t)|$

let λ_ε be the mean of M_ε.

W - continuous $\Rightarrow$ M_ε is continuous

$|M_\varepsilon(t)| \leq 2\|W\|_\infty$. $\quad E\|M_\varepsilon\|_\infty < \infty$. $\quad \lambda_\varepsilon$ is continuous.

$M_\varepsilon(t) \to 0$ as $\varepsilon \to 0$. $\quad \lambda_\varepsilon(t) \to 0$ as $\varepsilon \to 0$.

Since λ_ε's are decreasing as $\varepsilon \downarrow 0$. by Dini's theorem.

$\sup_{t\in K} \lambda_\varepsilon(t) \to 0$ as $\varepsilon \downarrow 0$.

Thm 9.2. $W_1\ W_2 \cdots$ iid Random functions in $C(K)$. K is compact, with mean μ and $E\|W\|_\infty < \infty$. let $\overline{W}_n$ be $\frac{1}{n}(W_1 + \cdots + W_n)$. then $\|\overline{W}_n - \mu\|_\infty \xrightarrow{P} 0$. as $n \to \infty$

Thm. 9.4. random $G_n \in C(K)$. K compact. $\|G_n - g\|_\infty \xrightarrow{P} 0$ with g a non random function in $C(K)$.

1. if t_n $n \geq 1$. are random variables converging in probability to a constant $t^* \in K$, $t_n \xrightarrow{P} t^*$, then $G_n(t_n) \xrightarrow{P} g(t^*)$
2. if g achieves its maximum at a unique t^*. if t_n are random varables maximizing G_n. $G_n(t_n) = \sup_{t\in K} G_n(t)$ then $t_n \xrightarrow{P} t^*$.
3. if $K \subset \mathbb{R}$. $g(t)=0$ has a unque solution t^*. $\quad G_n(t_n) = 0$ then $t_n \xrightarrow{P} t^*$.

Consistency of the MLE

- $X_1 \cdots X_n \overset{iid}{\sim} f_\theta \quad \theta \in \Omega.$ θ — truth

$$\ell_n(\omega) = \log \prod_{i=1}^n f_\omega(X_i) = \sum_{i=1}^n \log f_\omega(X_i).$$

MLE: $\hat{\theta}_n = \hat{\theta}_n(X_1 \cdots X_n)$

- Kullback-leibler information.

$$I(\theta, \omega) = E_\theta \log \frac{f_\theta(X)}{f_\omega(X)}$$

$\longrightarrow$ as a measure of the information discriminating between θ and ω

Lemma: if $P_\theta \neq P_\omega$ then $I(\theta, \omega) > 0$

pf.
$$
\begin{aligned}
-I(\theta, \omega) &= E_\theta \log \frac{f_\omega(X)}{f_\theta(X)} \\
&\leq \log E_\theta \frac{f_\omega(X)}{f_\theta(X)} \qquad \text{Jensen's inequality} \\
&= \log \int_{f_\theta > 0} \frac{f_\omega(x)}{f_\theta(x)} f_\theta(x) \, d\mu(x) \\
&\leq \log 1 \\
&= 0
\end{aligned}
$$

strict equality will occur if $\frac{f_\omega(x)}{f_\theta(x)}$ is a constant.

- Define $W(\omega) = \log \frac{f_\omega(x)}{f_\theta(x)}$

Thm: If Ω is compact. $E_\theta \|W\|_\infty < \infty$. $f_\omega(x)$ is a continuous function of ω for a.e. x, and $P_\omega \neq P_\theta$ for all $\omega \neq \theta$.

then $\hat{\theta}_n \xrightarrow{P} \theta$ under P_θ.

Pf. $W_i(\omega) = \log \frac{f_\omega(X_i)}{f_\theta(X_i)}$

Under P_θ. $W_1, W_2, \ldots$ are iid random functions in $C(\Omega)$ with mean $\mu(\omega) = -I(\theta, \omega)$. $\mu(\theta) = 0$. $\mu(\omega) < 0$ for $\omega \neq \theta$.

So μ has a unique maximum at θ.

$$\overline{W_n}(\omega) = \frac{1}{n} \sum_{i=1}^{n} W_i(\omega) = \frac{l_n(\omega) - l_n(\theta)}{n}$$

$\hat{\theta}_n$ maximizes $\overline{W}_n$. $\|\overline{W}_n - \mu\|_\infty \to 0$

Thm: $\Omega = \mathbb{R}^p$. $f_\omega(x)$ is a continuous function of ω for a.e. x. $P_\omega \neq P_\theta$ for all $\omega \neq \theta$. and $f_\omega(x) \to 0$ as $\omega \to \infty$. If $E_\theta \| 1_K W\|_\infty < \infty$ for any compact set $K \subset \mathbb{R}^p$, and if $E_\theta \sup_{\|\omega\| > a} W(\omega) < \infty$ for some $a > 0$ then under P_θ.

$\hat{\theta}_n \xrightarrow{P} \theta$.

Pf: Since $f_\omega(x) \to 0$ as $\omega \to \infty$ & $f_\theta(x) > 0$

$$\sup_{\|\omega\| > b} W(\omega) \longrightarrow -\infty \quad \text{as } b \to +\infty$$

By a dominated convergence argument, we can choose b so that

$$E_\theta \sup_{\|\omega\|>b} W(\omega) < 0 .$$

Note that b must exceed $\|\theta\|$ because $W(\theta) = 0$. Since

$$\sup_{\|\omega\|>b} \overline{W}_n(\omega) \le \frac{1}{n} \sum_{j=1}^{n} \sup_{\|\omega\|>b} W_i(\omega) \xrightarrow{P} E_\theta \sup_{\|\omega\|>b} W(\omega)$$

$$P_\theta\left(\sup_{\|\omega\|>b} \overline{W}_n(\omega) \ge 0 \right) \to 0 \quad \text{as } n \to +\infty$$

let K be the closed ball of radius b, and let $\tilde{\theta}_n$ be variables maximizing $\overline{W}_n$ over K. $\tilde{\theta}_n \xrightarrow{P} \theta$.

Since $\hat{\theta}_n$ must lie in K whenever

$$\sup_{\|\omega\|>b} \overline{W}_n(\omega) < \overline{W}_n(\theta) = 0$$

$P_\theta(\hat{\theta}_n = \tilde{\theta}_n) \to 1$. $\therefore \hat{\theta}_n \xrightarrow{P} \theta$.

ex: Location family $f_\theta(x) = g(x-\theta)$ $\theta \in \mathbb{R}$.

1. g is continuous and bounded. $\sup_{x\in\mathbb{R}} g(x) = K < \infty$
2. $g(x) \to 0$ as $x \to \pm\infty$
3. $\int |\log g(x)|\, g(x) < \infty$. (sastified by cauchy t)

Then

$$E_\theta \sup_{\omega\in\mathbb{R}} W(\omega) = E_\theta \sup_{\omega\in\mathbb{R}} \log \frac{g(X-\omega)}{g(X-\theta)}$$

$$= \log K - E_\theta \log g(X-\theta)$$

$$= \log K - \int \log g(x)\, g(x)\, dx$$

$$< \infty$$

Hence $\hat{\theta}_n$ is consistent by the one-sided adaptation of our consistency theorems.

Limiting distribution for the MLE.

Thm 9.14. 1. $X, X_1, X_2, \dots$ iid f_θ. $\theta \in \Omega \subset \mathbb{R}$.

2. $A = \{x : f_\theta(x) > 0\}$ is independent of θ.

3. $\frac{\partial^2 f_\theta(x)}{\partial \theta^2}$ exists for $x \in A$ and is continuous in θ

4. $W(\theta) = \log f_\theta(x)$. Fisher $I(\theta)$ exists finite.

$I(\theta) = E_\theta W'(\theta)^2$ or $I(\theta) = -E_\theta W''(\theta)$. $E_\theta W'(\theta) = 0$

5. For every θ in the interior of Ω $\exists\, \varepsilon > 0$

$E_\theta \| I_{[\theta-\varepsilon, \theta+\varepsilon]} W'' \|_\infty < \infty$

6. MLE. $\hat{\theta}_n \xrightarrow{P} \theta$.

Then $\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N\left(0, \frac{1}{I(\theta)}\right)$. under P_θ. as $n \to \infty$.

Lemma: $Y_n \Rightarrow Y$. $P(B_n) \to 1$ as $n \to \infty$. Then for any RV Z_n, $n \geq 1$. $Y_n 1_{B_n} + Z_n 1_{B_n^c} \Rightarrow Y$ as. $n \to \infty$

pf. $P(|Z_n 1_{B_n^c}| > \varepsilon) \leq P(B_n^c) = 1 - P(B_n) \to 0$ $\quad Z_n 1_{B_n^c} \xrightarrow{P} 0$.

$P(|1_{B_n} - 1| \geq \varepsilon) \leq P(B_n^c) = 1 - P(B_n) \to 0$. $\quad 1_{B_n} \xrightarrow{P} 1$

②

Pf: $E_\theta \| 1_{[\theta-\varepsilon, \theta+\varepsilon]} W'' \|_\infty < \infty$.

B_n — event $\hat{\theta}_n \in [\theta-\varepsilon, \theta+\varepsilon]$. $P_\theta(B_n) \to 1$.

$\overline{W}_n'(\hat{\theta}_n) = 0$.

$0 = \overline{W}_n'(\hat{\theta}_n) = \overline{W}_n'(\theta) + \overline{W}_n''(\tilde{\theta}_n)(\hat{\theta}_n - \theta)$.

$\tilde{\theta}_n$ — an intermediate value between $\hat{\theta}_n$ and θ.

On B_n, $\sqrt{n}(\hat{\theta}_n - \theta) = \dfrac{\sqrt{n}\, \overline{W}_n'(\theta)}{-\overline{W}''(\tilde{\theta}_n)}$

$\overline{W}_n'(\theta)$ — average, mean 0, Var $I(\theta)$

$\sqrt{n}\, \overline{W}'(\theta) \Rightarrow Z \sim N(0, I(\theta))$.

$|\tilde{\theta}_n - \theta| \le |\hat{\theta}_n - \theta|$. on B_n $\tilde{\theta}_n \xrightarrow{P} \theta$.

Thm 9.2. $\| 1_{[\theta-\varepsilon, \theta+\varepsilon]} (\overline{W}_n'' - \mu) \|_\infty \xrightarrow{P} 0$ $\mu(\omega) = E_\theta W''(\omega)$.

Thm 9.4. $\overline{W}_n''(\tilde{\theta}_n) \xrightarrow{P} \mu(\theta) = -I(\theta)$.

Confidence Intervals

Def: let δ_0, δ_1 be two statistics. the random interval (δ_0, δ_1) is called a $1-\alpha$ confidence interval for $g(\theta)$ if

$$P_\theta\Big(g(\theta) \in (\delta_0, \delta_1)\Big) \geq 1-\alpha \quad \text{for all } \theta \in \Omega.$$

A random set $S = S(X)$ is called a $1-\alpha$ confidence region for $g(\theta)$ if $P_\theta\big(g(\theta) \in S\big) \geq 1-\alpha$. for all $\theta \in \Omega$.

- converge probabilities equal $1-\alpha$ for all $\theta \in \Omega$ — exact CI.
exact confidence region.

ex. $X_1 \cdots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. $\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$ $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$. $V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. — Pivot

not statistic

$Z \perp V$. $T = \frac{Z}{\sqrt{V/n-1}} \sim t_{n-1}$ pivot.

$$P\left(-t_{\frac{\alpha}{2}, n-1} < T < t_{\frac{\alpha}{2}, n-1}\right) = 1-\alpha \qquad T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

$$P\left(\mu \in \left(\bar{X} - t_{\frac{\alpha}{2}, n-1}\frac{S}{\sqrt{n}},\ \bar{X} + t_{\frac{\alpha}{2}, n-1}\frac{S}{\sqrt{n}}\right)\right) = 1-\alpha.$$

$$\underset{\delta_0}{\shortparallel} \qquad \underset{\delta_1}{\shortparallel}$$

(δ_0, δ_1) — $(1-\alpha)$ - CI for μ.

$$1-\alpha = P_\theta\left(\chi^2_{1-\frac{\alpha}{2}, n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}, n-1}\right) = P_\theta\left(\sigma^2 \in \left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\right)\right)$$

- Asymptotic CIs.

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N\left(0, \frac{1}{I(\theta)}\right).$$

$$\sqrt{nI(\theta)}(\hat{\theta}_n - \theta) \Rightarrow N(0,1).$$

$\sqrt{nI(\theta)}(\hat{\theta} - \theta)$ is called an <u>approximate pivot</u>.

$$P_\theta\left(\sqrt{nI(\theta)}\,|\hat{\theta}_n - \theta| < z_{\frac{\alpha}{2}}\right) \to 1-\alpha.$$

$$S = \left\{\theta \in \Omega : \sqrt{nI(\theta)}\,|\hat{\theta}_n - \theta| < z_{\frac{\alpha}{2}}\right\} \qquad P_\theta(\theta \in S) \to 1-\alpha.$$

S: asymptotic confidence region for θ

if $I(\cdot)$ is continuous, $\sqrt{\frac{I(\hat{\theta}_n)}{I(\theta)}} \xrightarrow{P_\theta} 1.$

$$\sqrt{nI(\hat{\theta}_n)}(\hat{\theta}_n - \theta) = \sqrt{\frac{I(\hat{\theta}_n)}{I(\theta)}}\sqrt{nI(\theta)}(\hat{\theta}_n - \theta) \Rightarrow N(0,1).$$

$$P\left(\sqrt{nI(\hat{\theta}_n)}\,\left|\hat{\theta}_n-\theta\right| < z_{\frac{\alpha}{2}}\right) = P_\theta\left[\theta \in \left(\hat{\theta}_n - \frac{z_{\frac{\alpha}{2}}}{\sqrt{nI(\hat{\theta}_n)}},\ \hat{\theta}_n + \frac{z_{\frac{\alpha}{2}}}{\sqrt{nI(\hat{\theta}_n)}}\right)\right]$$

$$\longrightarrow 1-\alpha$$

asymptotic CI: $\hat{\theta}_n \pm \dfrac{z_{\frac{\alpha}{2}}}{\sqrt{nI(\hat{\theta}_n)}}$.

$$-l_n''(\hat{\theta}_n)/n \xrightarrow{P_\theta} I(\theta). \qquad \frac{\sqrt{-l_n''(\hat{\theta}_n)}}{\sqrt{nI(\theta)}} \xrightarrow{P_\theta} 1$$

$$\sqrt{-l_n''(\hat{\theta}_n)}\left(\hat{\theta}_n-\theta\right) \Rightarrow N(0,1)$$

$-l_n''(\hat{\theta}_n)$

$\hat{\theta}_n \pm \dfrac{z_{\frac{\alpha}{2}}}{\sqrt{-l''(\hat{\theta}_n)}}$: <u>observed or sample fisher information</u>

$$2l_n(\hat{\theta}_n) - 2l_n(\theta) = \left[\sqrt{-l_n''(\theta_n^*)}\,(\theta-\hat{\theta}_n)\right]^2.$$

$$2l_n(\hat{\theta}_n) - 2l_n(\theta) \Rightarrow z^2 \sim \chi_1^2.$$

$$P_\theta\left(2l_n(\hat{\theta}_n) - 2l_n(\theta) < z_{\frac{\alpha}{2}}^2\right) \to 1-\alpha.$$

$$S = \left\{\theta \in \Omega :\ 2l_n(\hat{\theta}_n) - 2l_n(\theta) < z_{\frac{\alpha}{2}}^2\right\}.$$

ex: $X_1 \cdots X_n \overset{iid}{\sim}$ Poisson (θ)

$$l_n(\theta) = n\bar{x}\log\theta - n\theta - \log\left(\prod_{i=1}^{n} X_i!\right)$$

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n) \qquad l_n'(\theta) = \frac{n\bar{x}}{\theta} - n \qquad \hat{\theta} = \bar{X}$$

$$I(\theta) = \frac{1}{\theta} \qquad S = \left\{\theta > 0 : \sqrt{\frac{n}{\theta}}\,|\hat{\theta} - \theta| < z_{\frac{\alpha}{2}}\right\}$$

$$= \left\{\theta > 0 : \hat{\theta}^2 - 2\hat{\theta}\theta + \hat{\theta}^2 < z_{\frac{\alpha}{2}}^2 \theta / n\right\} = (\hat{\theta}^-, \hat{\theta}^+)$$

$$\hat{\theta}^{\pm} = \hat{\theta} + \frac{z_{\frac{\alpha}{2}}^2}{2n} \pm \sqrt{\left(\hat{\theta} + \frac{z_{\frac{\alpha}{2}}^2}{2n}\right)^2 - \hat{\theta}^2}$$

$$I(\hat{\theta}) = \frac{1}{\bar{X}} \qquad \left(\bar{X} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{X}}{n}}\right)$$

$$-l''(\hat{\theta}) = \frac{n\bar{X}}{\hat{\theta}} = \frac{n}{\bar{X}}$$

profile confidence interval $\left\{\theta > 0 : \theta - \bar{X}\log\frac{\theta}{\bar{X}} - \bar{X} < \frac{z_{\frac{\alpha}{2}}^2}{2n}\right\}$

ex: X — either 1 or 2. according to the toss of a fair coin.

$$Y \mid X = x \sim N(\theta, x)$$

joint: $$f_\theta(x, y) = \frac{1}{2\sqrt{2\pi x}} e^{-\frac{(y-\theta)^2}{2x}}$$

$$I(\theta) = -E_\theta \frac{\partial}{\partial\theta^2}\log f_\theta(x, y) = E_\theta\left[\frac{1}{X}\right] = \frac{3}{4}$$

EM Algorithm.

- The EM algorithm is a recursive method to calculate MLEs from incomplete data.
- The "full data" X has density from an exponential family.
- observed data Y. $Y = g(X)$. g many-to-one.
- X — density $h(x)\, e^{\eta T(x) - A(\eta)}$ $\quad \eta \in \Omega \subset \mathbb{R}$.

ex. $X_1 \cdots X_n \overset{iid}{\sim} f(x)$. Y_i — X_i rounded to the nearest integer

- $\mathcal{X}(y) = \{ x \in \mathcal{X} : g(x) = y \}$. $\quad \mathcal{X}$ — sample space for X
  $\mathcal{Y}$ — sample space for Y.

$Y = y$ iff $X \in \mathcal{X}(y)$.

Prop 9.22. The joint density of X and Y is

$$1_{\mathcal{X}(y)}(x)\, h(x)\, e^{\eta T(x) - A(\eta)}$$

nonnegative $f(X,Y) = \sum_{y \in \mathcal{Y}} f(X,y) I\{Y=y\}$

pf. $E f(X,Y) = \sum_{y \in \mathcal{Y}} E f(X,y) I\{g(X)=y\} = \sum_{y \in \mathcal{Y}} \int_{\mathcal{X}} f(x,y) I\{g(x)=y\}\, h(x)\, e^{\eta T(x) - A(\eta)}\, d\mu(x)$

- Define $e(y,\eta) = E_\eta[T(X) \mid Y=y]$

① initial guess $\hat{\eta}_0$. ~~T(x) is impute~~

② $T(x)$ is inputed by $T_1 = e(Y, \hat{\eta}_0)$ <u>E-step</u>

③ $\hat{\eta}_1 = \psi(T_1)$ <u>M-step</u>.

if the EM algorithms converges to $\tilde{\eta}$. then $\tilde{\eta}$ will satisfy

$$\tilde{\eta} = \psi(e(Y,\tilde{\eta})). \quad A'(\tilde{\eta}) = e(Y,\tilde{\eta})$$

$$f_\eta(Y) = P_\eta(Y=y) = P_\eta(X \in \mathcal{X}(y)) = \int_{\mathcal{X}(y)} h(x) e^{\eta T(x) - A(\eta)} d\mu(x)$$

$$\frac{\partial}{\partial \eta} \lg f_\eta(Y) = \frac{\frac{\partial}{\partial \eta}\int_{\mathcal{X}(y)} h(x) e^{\eta T(x) - A(\eta)} d\mu(x)}{f_\eta(Y)}$$

$$= \frac{\int_{\mathcal{X}(Y)} [T(x) - A'(\eta)] h(x) e^{\eta T(x) - A(\eta)} d\mu(x)}{f_\eta(Y)}$$

$$= e(Y,\eta) - A'(\eta)$$

the likelihood has zero slope when $\eta = \tilde{\eta}$.

Ex. (rounding) $X_1 \cdots X_n \overset{iid}{\sim} \exp(\eta)$. $f_1(x) = \eta e^{-\eta x}$. $x > 0$

$Y_i = \lfloor X_i \rfloor$

$X_1 \cdots X_n \longrightarrow$ complete suff. $T = -\sum_{i=1}^{n} X_i$.

MLE: $\psi(T) = -\frac{n}{T}$.

$$E_\eta(X_i \mid Y_i = y_i] = E_\eta(X_i \mid y_i \le X_i < y_i + 1)$$

$$= \frac{\int_{y_i}^{y_i+1} x \,\eta e^{-\eta x} dx}{\int_{y_i}^{y_i+1} \eta e^{-\eta x} dx} = y_i + \frac{e^\eta - 1 - \eta}{\eta(e^\eta - 1)}$$

$$E_\eta(X_i \mid Y_1 = y_1, \cdots Y_n = y_n) = E_\eta(X_i \mid Y_i = y_i)$$

$$e(y, \eta) = E_\eta(T \mid Y_1 = y_1 \cdots Y_n = y_n) = -\sum_{i=1}^{n} E_\eta(X_i \mid Y_i = y_i)$$

$$= -n\left[\bar{Y} + \frac{e^\eta - 1 - \eta}{\eta(e^\eta - 1)}\right]$$

EM: $\hat{\eta}_j = -\frac{n}{T_j}$. $\qquad T_{j+1} = -n\left[\bar{Y} + \frac{e^{\hat{\eta}_j} - 1 - \hat{\eta}_j}{\hat{\eta}_j(e^{\hat{\eta}_j} - 1)}\right]$

The mass of Y_i

$$P_\eta(Y_i = y) = P_\eta(y \le X_i < y+1) = (1 - e^{-\eta})(e^{-\eta})^y \qquad y = 0, 1, \cdots$$

$\sim$ geometric with $p = 1 - e^{-\eta}$.

$$\hat{p} = \frac{1}{1 + \bar{Y}} \qquad \hat{\eta} = -\log(1 - \hat{p}) = \log\left(1 + \frac{1}{\bar{Y}}\right)$$