Corrigé examen 24 janvier 2013

Exercice 1

1) $\int_{\mathbb{R}} f_\theta(x)dx = 1 \Rightarrow 1 = \int_0^\infty \sqrt{\frac{2}{\pi}}\, c \cdot \exp\left(-\frac{x^2}{2\theta}\right) dx$

$\frac{x}{\sqrt{\theta}} = y \Rightarrow dx = \sqrt{\theta}\, dy$

$1 = c\sqrt{\frac{2}{\pi}} \int_0^\infty \exp\left(-\frac{y^2}{2}\right) \cdot \sqrt{\theta}\, dy \Rightarrow 1 = c\sqrt{\frac{2}{\pi}} \cdot \sqrt{\theta} \cdot \frac{\sqrt{2\pi}}{2} \Rightarrow \boxed{c = \frac{1}{\sqrt{\theta}}}$

2) $\mathbb{E}(X) = \sqrt{\frac{2}{\pi \cdot \theta}} \cdot \int_0^\infty x \cdot \exp\left(-\frac{x^2}{2\theta}\right) dx = -\sqrt{\frac{2}{\pi \cdot \theta}} \cdot \theta \int_0^\infty \left(\exp\left(-\frac{x^2}{2\theta}\right)\right)' dx$

$= \sqrt{\frac{2}{\pi}} \cdot \sqrt{\theta} \Rightarrow \boxed{\mathbb{E}(X) = \sqrt{\frac{2}{\pi}} \cdot \sqrt{\theta}}$

$\mathbb{E}(X^2) = \sqrt{\frac{2}{\pi \cdot \theta}} \cdot \int_0^\infty x^2 \exp\left(-\frac{x^2}{2\theta}\right) dx = -\sqrt{\frac{2}{\pi \cdot \theta}} \cdot \theta \int_0^\infty x \left(\exp\left(-\frac{x^2}{2\theta}\right)\right)' dx$

$= \sqrt{\frac{2\theta}{\pi}} \cdot \int_0^\infty \exp\left(-\frac{x^2}{2\theta}\right) dx = \sqrt{\frac{2\theta}{\pi}} \cdot \sqrt{\frac{\pi\theta}{2}} = \theta$

$\Rightarrow \boxed{\mathbb{E}(X^2) = \theta} \Rightarrow Var(X) = \theta - \frac{2\theta}{\pi} = \theta\left(1 - \frac{2}{\pi}\right) > 0$

3) $\overline{X}_n = \mathbb{E}(X) \Rightarrow \overline{X}_n = \sqrt{\frac{2}{\pi}}\sqrt{\theta} \Rightarrow \boxed{\hat{\theta}_n = \left(\frac{\pi}{2}\right) \cdot \left(\overline{X}_n\right)^2}$

4) Par LFGN:

$\overline{X}_n \xrightarrow[n \to \infty]{p.s} \mathbb{E}(X) \Rightarrow \overline{X}_n \xrightarrow{p.s} \sqrt{\frac{2}{\pi}} \cdot \theta \Rightarrow \frac{\pi}{2}\left(\overline{X}_n\right)^2 \xrightarrow{p.s} \theta$

$\Rightarrow \hat{\theta}_n$ fortement convergent.

$\mathbb{E}\left(\overline{X}_n^2\right) = Var\left(\overline{X}_n\right) + \mathbb{E}^2\left(\overline{X}_n\right) = \frac{Var(X)}{n} + \mathbb{E}^2(X)$

$= \frac{\theta}{n}\left(1 - \frac{2}{\pi}\right) + \frac{2}{\pi}\theta$

$\mathbb{E}\left(\hat{\theta}_n\right) = \frac{\pi}{2} \cdot \mathbb{E}\left(\overline{X}_n^2\right) = \frac{\theta \cdot \pi}{2n}\left(1 - \frac{2}{\pi}\right) + \theta \xrightarrow[n \to \infty]{} \theta$

Donc $\hat{\theta}_n$ est biaisé, mais asymptotiquement sans biais.

L'information de Fisher existe parce que l'ensemble :

$A = \{x \,;\, f_\theta(x) > 0\} =]0, \infty[$ ne dépende pas de θ.

$\forall x \in A,\ \forall \theta > 0,\ \frac{\partial}{\partial \theta} \log f_\theta(x) = -\frac{1}{2\theta} + \frac{x^2}{2\theta^2}$

$$\frac{\partial}{\partial\theta} f_\theta(x) = \sqrt{\frac{2}{\pi\theta}} \cdot \frac{x^2}{2\theta^2} \exp\left(-\frac{x^2}{2\theta}\right) \mathbb{1}_{x>0} - \sqrt{\frac{2}{\pi}} \cdot \frac{1}{2} \cdot \theta^{-3/2} \cdot \exp\left(-\frac{x^2}{2\theta}\right) \mathbb{1}_{x>0}$$

$$= \frac{1}{\sqrt{\pi}} \cdot \theta^{-3/2} \cdot \left(\frac{x^2}{\theta} - 1\right) \exp\left(-\frac{x^2}{2\theta}\right) \mathbb{1}_{x>0}$$

$$\int_{\mathbb{R}} \frac{\partial}{\partial\theta} f_\theta(x)\,dx = \frac{1}{\sqrt{\pi}} \cdot \theta^{-3/2} \int_0^\infty \left(\frac{x^2}{\theta} - 1\right) \exp\left(-\frac{x^2}{2\theta}\right) dx$$

On a vu à la question 2) que : $\int_0^\infty x^2 \exp\left(-\frac{x^2}{2\theta}\right) dx = \theta \int_0^\infty \exp\left(-\frac{x^2}{2\theta}\right) dx$

Donc : $\int_{\mathbb{R}} \frac{\partial}{\partial\theta} f_\theta(x)\,dx = 0 = \frac{\partial}{\partial\theta}\int_{\mathbb{R}}$

Par définition : $\frac{\partial}{\partial\theta} \int_{\mathbb{R}} f_\theta(x)\,dx = 0$ $\Big\} \Rightarrow$ la 2ème condition d'existence

de l'information de Fisher est satisfaite.

$$\frac{\partial^2}{\partial\theta^2} \log f_\theta(x) = \frac{1}{2\theta^2} - \frac{x^2}{\theta^3}$$

Alors l'information de Fisher de X est :

$$I_1(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \log f_\theta(X)\right] = -\frac{1}{2\theta^2} + \frac{1}{\theta^3}\mathbb{E}(X^2)$$

$$= -\frac{1}{2\theta^2} + \frac{\theta}{\theta^3} = \frac{1}{2\theta^2}$$

L'information de Fisher pour $(X_1, \dots, X_n)$ est :

$$I_n(\theta) = n \cdot I_1(\theta) = \frac{n}{2\theta^2}$$

5) La log-vraisemblance est :

$$\log L_n(\theta) = \sum_{i=1}^n \log f_\theta(X_i) = \frac{n}{2} \log\left(\frac{2}{\pi}\right) - \frac{1}{2} \sum_{i=1}^n \left(\log\theta + \frac{X_i^2}{\theta}\right)$$

$$\frac{\partial}{\partial\theta} \log L_n(\theta) = 0 \Rightarrow + \frac{n}{\theta} - \frac{1}{\theta^2} \sum_{i=1}^n X_i^2 = 0 \Rightarrow \boxed{\tilde{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i^2}$$

$$\frac{\partial^2}{\partial\theta^2} \log L_n(\theta) = + \frac{n}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^n X_i^2$$

$$\frac{\partial^2}{\partial\theta^2}\log L_n(\tilde{\theta}_n) = +\frac{n}{\tilde{\theta}_n^2} - 2\frac{n\tilde{\theta}_n}{\tilde{\theta}_n^3} = -n\frac{1}{\tilde{\theta}_n^2} < 0$$

$\Rightarrow \tilde{\theta}_n$ point de max pour $L_n(\tilde{\theta}_n)$

Donc : $\boxed{\tilde{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2}$ est l'EMV.

6) On écrit la densité sous la forme :

$$f_\theta(x) = \exp\left(-\frac{x^2}{2\theta} - \frac{1}{2}\log\theta + \frac{1}{2}\log\left(\frac{2}{\pi}\right)\right)\cdot \mathbb{1}_{x>0}$$

Si on considère : $c(\theta) = -\frac{1}{2\theta}$ (mesurable)

mesurables $\begin{cases} T(x) = x^2 \\ D(\theta) = -\frac{1}{2}\log\theta + \frac{1}{2}\log\left(\frac{2}{\pi}\right) \\ \tilde{S}(x) = \mathbb{1}_{x>0} \end{cases}$

On peut écrire la densité sous la forme :

$$f_\theta(x) = \exp(c(\theta)\cdot T(x) + D(\theta))\cdot S(x)$$

donc, de type exponentiel.

7) $\tilde{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = U_2^n = \overline{X_n^2}$

Par LFGN : $\tilde{\theta}_n \xrightarrow[n\to\infty]{p.s} \mathbb{E}(X^2) = \theta \Rightarrow \tilde{\theta}_n$ fortement convergent

Biais : $\mathbb{E}(\tilde{\theta}_n) = \mathbb{E}(X^2) = \theta \Rightarrow \tilde{\theta}_n$ sans biais.

Efficacité et exhaustivité : on utilise que c'est une loi de type exponentiel.

$\frac{1}{n}\sum_{i=1}^{n} T(X_i) = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ exhaustive et sans biais pour $\mathbb{E}[T(X)] = \mathbb{E}(X^2) = \theta$

8) $\mathbb{E}(X^4) = \sqrt{\frac{2}{\pi\cdot\theta}}\int_0^\infty x^4\exp\left(-\frac{x^2}{2\theta}\right)dx = -\sqrt{\frac{2}{\pi\cdot\theta}}\int_0^\infty \theta x^3\left(\exp\left(-\frac{x^2}{2\theta}\right)\right)' dx$

$$= 3\sqrt{\frac{2\theta}{\pi}}\int_0^\infty x^2\exp\left(-\frac{x^2}{2\theta}\right)dx = 3\cdot\sqrt{\frac{2\theta}{\pi}}\cdot\sqrt{\frac{\pi\theta}{2}}\cdot\theta = 3\theta^2$$

(9) Un estimateur sans biais de θ :

$$\tilde{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2 > 0,$$

Par le TCL : $$\sqrt{n}\ \frac{\tilde{\theta}_n - \mathbb{E}(X^2)}{\sqrt{Var(X^2)}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

$$\mathbb{E}(X^2)=\theta,\quad Var(X^2) = \mathbb{E}(X^4) - \mathbb{E}^2(X^2) = 3\theta^2 - \theta^2 = 2\theta^2$$

Donc : $$\sqrt{n}\ \frac{\tilde{\theta}_n - \theta}{\sqrt{2}\cdot\theta} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1) \leftarrow \text{loi symétrique (en 0)}$$

On considère alors, la variable aléatoire :

$$Z_n = \sqrt{n}\ \frac{\tilde{\theta}_n - \theta}{\sqrt{2}\cdot\theta}$$

L'intervalle de confiance de niveau $1-\alpha$ est par déf :

$$1-\alpha = \mathbb{P}\left[A_n \leq \theta \leq B_n\right]$$

D'autre part : $$1-\alpha = \mathbb{P}\left[a_1 \leq \tilde{\theta}_n \leq b_1\right]$$
$$= \mathbb{P}\left[a_2 \leq Z_n \leq b_2\right]$$
$$\overset{n\to\infty}{=} \mathbb{P}\left[-u_{1-\frac{\alpha}{2}} \leq Z_n \leq u_{1-\frac{\alpha}{2}}\right]$$

avec $u_{1-\frac{\alpha}{2}}$ le fractile d'ordre $1-\frac{\alpha}{2}$ de la loi $\mathcal{N}(0,1)$. En remplaçant :

$$1-\alpha = \mathbb{P}\left[-u_{1-\frac{\alpha}{2}} \leq \sqrt{n}\ \frac{\tilde{\theta}_n}{\sqrt{2}\cdot\theta} - \frac{1}{\sqrt{2}} \leq u_{1-\frac{\alpha}{2}}\right]$$

$$= \mathbb{P}\left[\frac{1}{\sqrt{2}} - u_{1-\frac{\alpha}{2}} \leq \sqrt{n}\ \frac{\tilde{\theta}_n}{\sqrt{2}\,\theta} \leq \frac{1}{\sqrt{2}} + u_{1-\frac{\alpha}{2}}\right]$$

$$= \mathbb{P}\left[\frac{\sqrt{2}}{\sqrt{n}\cdot\tilde{\theta}_n}\cdot\Big(\underbrace{\frac{1}{\sqrt{2}} - u_{1-\frac{\alpha}{2}}}_{<0}\Big) \leq \frac{1}{\theta} \leq \frac{\sqrt{2}}{\sqrt{n}\cdot\tilde{\theta}_n}\Big(\frac{1}{\sqrt{2}} + u_{1-\frac{\alpha}{2}}\Big)\right]$$

$$= \mathbb{P}\left[\frac{\sqrt{n}}{\sqrt{2}}\tilde{\theta}_n\Big(\frac{1}{\sqrt{2}} + u_{1-\frac{\alpha}{2}}\Big)^{-1} \leq \theta\right]$$

$$\Rightarrow A_n = \frac{\sqrt{n}}{\sqrt{2}}\tilde{\theta}_n\Big(\frac{1}{\sqrt{2}} + u_{1-\frac{\alpha}{2}}\Big)^{-1} \quad \text{et} \quad B_n = \infty.$$

10) $H_0 : \theta = \theta_0$, $H_1 : \theta > \theta_0$.

Puisque H_0 est une hypothèse simple, on applique le Lemme de Neyman-Pearson.

$\alpha = \mathbb{P}[L_1 > k L_0 | H_0]$, avec L_1 la vraisemblance sous H_1

L_0 —— || —— H_0 .

$$L_1 = \prod_{i=1}^{n} f_\theta(x_i) = \left(\frac{2}{\pi\theta}\right)^{n/2} \exp\left(-\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2\right) \cdot \mathbb{1}_{\min x_i > 0} \text{ , si } \theta > \theta_0 .$$

Alors :

$$\alpha = \mathbb{P}\left[\left(\frac{\theta_0}{\theta}\right)^{n/2} \cdot \exp\left(\frac{1}{2}\sum_{i=1}^{n} x_i^2 \left(\frac{1}{\theta_0} - \frac{1}{\theta}\right)\right) > k \;\middle|\; \theta = \theta_0\right]$$

$$= \mathbb{P}\left[\exp\left(\frac{1}{2}\sum_{i=1}^{n} x_i \left(\frac{1}{\theta_0} - \frac{1}{\theta}\right)\right) > k_1 \;\middle|\; \theta = \theta_0\right]$$

$$= \mathbb{P}\left[\exp\left(\frac{1}{2}\sum x_i^2 \underbrace{(\theta - \theta_0)}_{\geq 0 \text{ sous } H_1}\right) > k_2 \;\middle|\; \theta = \theta_0\right]$$

$$= \mathbb{P}\left[\sum_{i=1}^{n} x_i^2 > k_3 \;\middle|\; \theta = \theta_0\right]$$

$$= \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} x_i^2 > k_4 \;\middle|\; \theta = \theta_0\right]$$

$$= \mathbb{P}\left[\tilde{\theta}_n > k_4 \;\middle|\; \theta = \theta_0\right]$$

$$= \mathbb{P}\left[\underbrace{\sqrt{n}\,\frac{\tilde{\theta}_n - \theta_0}{\sqrt{2}\,\theta_0}}_{\xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1) \text{, sous } H_0} > k_5\right]$$

Donc $k_5 = u_{1-\alpha}$ le fractile d'ordre $1-\alpha$ de la loi $\mathcal{N}(0,1)$

Donc, la statistique de test est : $Z_n = \sqrt{n}\,\frac{\tilde{\theta}_n - \theta_0}{\sqrt{2} \cdot \theta_0}$, et la zone de rejet : $R = \{Z_n > u_{1-\alpha}\}$. Ce qui nous donne un test UPP.

⑥

Exercice 2 $X \sim \mathcal{N}(m,1)$

C'est un test sur la moyenne d'une loi Normale, de variance connue (fait en cours).

$H_0 : m \leq 1 \qquad H_1 : m > 1$.

On démontre d'abord que X a un rapport de vraisemblance monotone (MVR) par rapport à une statistique T_n.

La vraisemblance est :

$$L_n(m) = (2\pi)^{-n/2} \cdot \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - m)^2\right)$$

Si $m_1 < m_2$:

$$\frac{L_n(m_2)}{L_n(m_1)} = \exp\left(-\frac{1}{2}n \cdot m_2^2 + m_2\sum_{i=1}^{n} x_i + \frac{1}{2}n \cdot m_1^2 - m_1\sum_{i=1}^{n} x_i\right)$$

$$= \exp\left(\frac{n}{2}(m_1^2 - m_2^2) + \underbrace{(m_2 - m_1)}_{>0}\sum_{i=1}^{n} x_i\right)$$

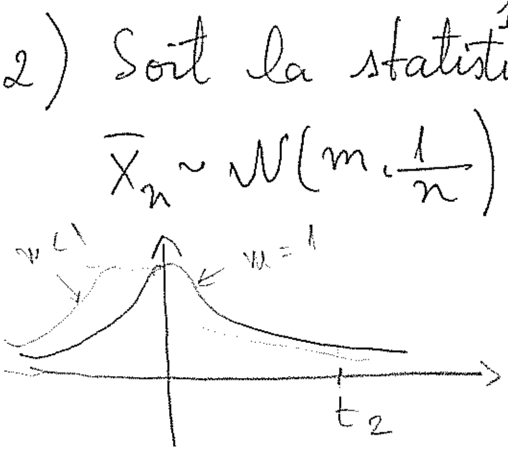
Ce rapport est croissant par rapport à : $T_n = \sum_{i=1}^{n} x_i$

1) Le risque de première espèce :

$$\alpha(m) = \mathbb{P}[(X_1, \dots, X_n) \in R \mid m \leq 1]$$
$$= \mathbb{P}\left[\sum_{i=1}^{n} X_i > t_0 \mid m \leq 1\right]$$
$$= \mathbb{P}\left[\overline{X}_n > t_1 \mid m \leq 1\right]$$
$$= \mathbb{P}\left[\sqrt{n}\,\frac{\overline{X}_n - 1}{1} > t_2 \mid m \leq 1\right]$$

2) Soit la statistique : $Z = \sqrt{n}(\overline{X}_n - 1)$

$\overline{X}_n \sim \mathcal{N}\left(m, \frac{1}{n}\right) \Rightarrow$ pour $m \leq 1$, $Z \sim \mathcal{N}(\underbrace{m-1}_{\leq 0}, 1)$



$$\alpha = \sup_{m \leq 1} \alpha(m) = \alpha(1) \Rightarrow t_2 = u_{1-\alpha}$$

le fractile d'ordre $1-\alpha$

de la loi $\mathcal{N}(0,1)$

Exercice 3

$$X_i : \begin{cases} 0 & \text{pas satisfait} \\ 1 & \text{satisfait} \end{cases} \sim \mathcal{B}(p)$$

$p = \mathbb{P}(X_i = 1)$ proba qu'une personne soit satisfaite.

$p_0 = 0{,}50$ $\quad 1-\alpha = 0{,}95$

la confiance a baissé : $p < p_0$.

$H_0 : p \geq p_0$ (la confiance cte ou elle a augmenté)

$H_1 : p < p_0$ (confiance en baisse).

C'est un test sur la proportion d'une loi de Bernoulli.

Stat de test : $\quad Z_n = \sqrt{n}\,\dfrac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)}}$ $\qquad \bar{x}_n = 0{,}48$

La zone de rejet :

$R = \{Z_n < u_\alpha\}$, u_α le fractile d'ordre α de la loi $\mathcal{N}(0,1)$

$= \cancel{\{Z_n < -2{,}5758\}} = \{Z_n < -1{,}6449\}$

La valeur de la statistique de test :

$$z = \sqrt{1300} \cdot \frac{\frac{625}{1300} - 0{,}5}{0{,}5} = -1{,}38675 \notin R$$

$\Rightarrow H_0$ acceptée et H_1 rejetée. Donc la côte de confiance n'a pas baissé.

Exercice 4

1) Y = count, $H_0 : Y \sim \mathcal{N}$, $H_1 : Y \not\sim \mathcal{N}$

On utilise le test de Shapiro. p-value $= 2.10^{-4} \Rightarrow H_0$ rejetée

Donc, on essaie une transf, par exemple la racine carée :

$\tilde{Y} = \log(Y)$ $\qquad H_0 : \tilde{Y} \sim \mathcal{N}$ $\quad H_1 : \tilde{Y} \not\sim \mathcal{N}$.

Par Shapiro : p-value $= 0{,}06 \Rightarrow H_0$ acceptée $\Rightarrow \tilde{Y} \sim \mathcal{N}$.

2) Les degrés de liberté du total : $n - 1 = 5 + 66 \Rightarrow n = 72$.

3) Puisque la variable "spray" est declarée avec "factor" elle est de type caractère. Ses valeurs

sont : A, B, C, D, E, F $\Rightarrow$ Spray, variable qualitative à 6 niveaux (valeurs). Donc on a un modèle d'analyse de variance à 1 facteur :

(1) $\tilde{Y}_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ $i = 1, \dots, 6$
$j = 1, \dots, n_i$

μ : l'effet moyen. ~~de A~~
α_1 : l'effet du spray A sur $\tilde{Y}$
$\vdots$
α_6 : — " — F sur $\tilde{y}$

La contrainte est $\alpha_1 = 0$. Donc le spray A sert de référenciel.

4) H_0 : (1) non significatif $\Leftrightarrow \forall \alpha_i = 0 \Leftrightarrow$ le type d'insecticide n'influe pas le nb. d'insectes.
Modèle réduit : (2) $\tilde{y}_{ij} = \mu + \varepsilon_{ij}$

H_1 : (1) signif $\Leftrightarrow \exists \alpha_i \neq 0 \Leftrightarrow$ le nb d'insectes est influencé par le type d'insecticide.
Modèle complet : (1).

Statistique de test : $Z = \dfrac{SM/5}{SR/66} \underset{H_0}{\sim} F(5, 66)$

Valeur stat test : $z = 44.8 \Rightarrow$ p-value $= 10^{-16} \Rightarrow H_0$ rejetée

5) $\hat{\alpha}_2 = 0{,}11$, $\hat{\alpha}_3 = -2{,}5158$, $\hat{\alpha}_4 = -1{,}5963$ $\hat{\alpha}_5 = -1{,}95$
$\hat{\alpha}_6 = 0{,}2579$ $\hat{\alpha}_1 = 0$ $\hat{\sigma} = 0{,}6283$

Par le spray B on a 0,11 d'insectes de plus qu'en utilisant A. Le moins d'insectes il y a si on utilise B et le plus par F ($+0{,}2579$).

6) $H_0 : \alpha_2 = 0 \Leftrightarrow$ les spray A et B ont la même action sur $\tilde{y}$
Stat test : $Z = \dfrac{\hat{\alpha}_2}{\sqrt{\widehat{Var}(\alpha)}} \underset{H_0}{\sim} t(66)$

$z = \frac{0{,}1160}{0{,}2565} = 0{,}452$ $\Rightarrow$ p-value $= 0{,}653 \Rightarrow H_0$ acceptée.

On fait la même analyse ~~en titi~~ pour les autres types de sprays. En conclusion : A, B et F ont la même action

7) $R^2 = 0{,}77$ R^2 ajusté $= 0{,}75 \Rightarrow$ modèle de assez ~~bonne~~ qualité

8) $\hat{\tilde{Y}} = \hat{\mu} + \hat{\alpha}_6 = 3{,}7607 + 0{,}2579$

$\hat{Y} = \log(\hat{\tilde{Y}})$