Examen Mercredi 14 Janvier 2009

Exercice 1 : Estimation

Partie 1 : variance σ^2 connue

1) La vraisemblance de $(y_1, ..., y_n)$ est définie par

$$\begin{aligned} L(y_1, ..., y_n; a) &= \prod_{i=1}^{n} f(y_i; a), \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - ax_i)^2}{2\sigma^2}\right], \\ &= \left(2\pi\sigma^2\right)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - ax_i)^2\right]. \end{aligned}$$

On a alors

$$\ln L(y_1, ..., y_n; a) = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - ax_i)^2.$$

En étudiant le signe de $\frac{\partial \ln L(x_1,...,x_n;\lambda)}{\partial a}$, on obtient :

$$\begin{aligned} \frac{\partial \ln L(y_1, ..., y_n; a)}{\partial a} &\geq 0 \iff \frac{1}{\sigma^2}\left[\sum_{i=1}^{n} y_i x_i - a\sum_{i=1}^{n} x_i^2\right] \geq 0, \\ &\iff a \leq \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}. \end{aligned}$$

La vraisemblance possède donc un maximum global unique obtenu pour $a = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}$ d'où

$$\widehat{a}_{\mathrm{MV}} = \frac{\sum_{i=1}^{n} Y_i x_i}{\|x\|^2}.$$

2) Le biais de l'estimateur $\widehat{a}_{\mathrm{MV}}$ est défini par

$$\text{Biais}(\widehat{a}_{\mathrm{MV}}) = E\left[\widehat{a}_{\mathrm{MV}}\right] - a,$$

avec

$$\begin{aligned} E\left[\widehat{a}_{\mathrm{MV}}\right] &= \frac{1}{\|x\|^2}\sum_{i=1}^{n} E\left[Y_i\right] x_i, \\ &= \frac{1}{\|x\|^2}\sum_{i=1}^{n} (ax_i)\, x_i, \\ &= a. \end{aligned}$$

L'estimateur $\widehat{a}_{\mathrm{MV}}$ est donc un estimateur non biaisé de a.
3) La borne de Cramer-Rao pour les estimateurs non biaisés du paramètre a est définie par

$$\begin{aligned} BCR(a) &= \frac{-1}{E\left[\frac{\partial^2 \ln L(Y_1,...,Y_n;a)}{\partial a^2}\right]}, \\ &= \frac{-1}{E\left[\frac{-1}{\sigma^2}\sum_{i=1}^n x_i^2\right]}, \\ &= \frac{\sigma^2}{\|x\|^2}, \end{aligned}$$

tandis que la variance de l'estimateur $\widehat{a}_{\mathrm{MV}}$ est

$$\begin{aligned} \mathrm{var}\left[\widehat{a}_{\mathrm{MV}}\right] &= \frac{1}{\|x\|^4}\mathrm{var}\left[\sum_{i=1}^n Y_i x_i\right], \\ &= \frac{1}{\|x\|^4}\sum_{i=1}^n x_i^2 \mathrm{var}(Y_i), \\ &= \frac{\sigma^2}{\|x\|^2}. \end{aligned}$$

L'estimateur $\widehat{a}_{\mathrm{MV}}$ est donc l'estimateur efficace de a.
4) En vertu de l'indépendance entre les variables aléatoires $e_1, ..., e_n$, le vecteur $(y_1, ..., y_n)$ est un vecteur Gaussien. Donc par transformation affine d'un vecteur Gaussien, $\widehat{a}_{\mathrm{MV}}$ est une variable aléatoire normale $\mathcal{N}\left(a, \frac{\sigma^2}{\|x\|^2}\right)$. On a alors

$$\frac{\widehat{a}_{\mathrm{MV}} - a}{\frac{\sigma}{\|x\|}} \sim \mathcal{N}(0,1).$$

On a alors

$$\begin{aligned} P\left[\left|\frac{\widehat{a}_{\mathrm{MV}} - a}{\frac{\sigma}{\|x\|}}\right| < u\right] &= \int_{-u}^{u} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du, \\ &= 1 - 2G(u). \end{aligned}$$

Si on se fixe

$$\alpha = P\left[\left|\frac{\widehat{a}_{\mathrm{MV}} - a}{\frac{\sigma}{\|x\|}}\right| < u\right],$$

on a

$$u = G^{-1}\left(\frac{1-\alpha}{2}\right).$$

Un intervalle de confiance pour $\widehat{a}_{\mathrm{MAP}}$ avec le coefficient de confiance α est donc défini par

$$\left|\frac{\widehat{a}_{\mathrm{MV}} - a}{\frac{\sigma}{\|x\|}}\right| < G^{-1}\left(\frac{1-\alpha}{2}\right),$$

c'est-à-dire
$$-\frac{\sigma}{\|x\|}G^{-1}\left(\frac{1-\alpha}{2}\right) < \widehat{a}_{\text{MV}} - a < \frac{\sigma}{\|x\|}G^{-1}\left(\frac{1-\alpha}{2}\right),$$
et finalement
$$\widehat{a}_{\text{MV}} - \frac{\sigma}{\|x\|}G^{-1}\left(\frac{1-\alpha}{2}\right) < a < \widehat{a}_{\text{MV}} + \frac{\sigma}{\|x\|}G^{-1}\left(\frac{1-\alpha}{2}\right).$$
L'intervalle de confiance pour le paramètre a avec un degré de confiance α est donc
$$\left[\widehat{a}_{\text{MV}} - \frac{\sigma}{\|x\|}G^{-1}\left(\frac{1-\alpha}{2}\right), \widehat{a}_{\text{MV}} + \frac{\sigma}{\|x\|}G^{-1}\left(\frac{1-\alpha}{2}\right)\right].$$
5) La loi a posteriori de $a|\, y_1, ..., y_n$ est définie par
$$\begin{aligned} f(a|\, y_1, ..., y_n) &\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - ax_i)^2\right]\exp\left[-\frac{1}{2\sigma_a^2}(a-m_a)^2\right], \\ &\propto \exp\left[-\frac{(a-\mu)^2}{2v^2}\right] \end{aligned}$$
qui est loi normale $\mathcal{N}(\mu, v^2)$. Le calcul de μ et de v^2 a été fait en cours. En identifiant les coefficients de a^2 et de a, on obtient
$$\begin{aligned} \frac{1}{v^2} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}x_i^2 + \frac{1}{\sigma_a^2}, \\ \frac{\mu}{v^2} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}y_i x_i + \frac{m_a}{\sigma_a^2}, \end{aligned}$$
c'est-à-dire
$$\begin{aligned} v^2 &= \frac{\sigma^2\sigma_a^2}{\|x\|^2\sigma_a^2 + \sigma^2}, \\ \mu &= \frac{\sigma_a^2}{\|x\|^2\sigma_a^2 + \sigma^2}\sum_{i=1}^{n}y_i x_i + \frac{\sigma^2}{\|x\|^2\sigma_a^2 + \sigma^2}m_a. \end{aligned}$$
L'estimateur MAP est la valeur qui maximise cette loi a posteriori, d'où
$$\widehat{a}_{\text{MAP}} = \frac{\|x\|^2\sigma_a^2}{\|x\|^2\sigma_a^2 + \sigma^2}\left(\frac{1}{\|x\|^2}\sum_{i=1}^{n}Y_i x_i\right) + \frac{\sigma^2}{\|x\|^2\sigma_a^2 + \sigma^2}m_a.$$
On voit que cet estimateur est une combinaison linéaire de $\widehat{a}_{\text{MV}}$ et de la moyenne de la loi a priori m_a. De plus

- quand $\sigma_a^2 \to 0$, on a
$$\frac{\|x\|^2\sigma_a^2}{\|x\|^2\sigma_a^2 + \sigma^2} \simeq 0 \text{ et } \frac{\sigma^2}{\|x\|^2\sigma_a^2 + \sigma^2} \simeq 1,$$
ce qui montre que $\widehat{a}_{\text{MAP}}$ se comporte comme la moyenne de la loi a posteriori m_a (la loi a priori est très informative et donc on lui fait confiance),

- quand $\sigma_a^2 \to \infty$, on a
$$\frac{\|x\|^2 \sigma_a^2}{\|x\|^2 \sigma_a^2 + \sigma^2} \simeq 1 \text{ et } \frac{\sigma^2}{\|x\|^2 \sigma_a^2 + \sigma^2} \simeq 0,$$
ce qui montre que $\widehat{a}_{\text{MAP}}$ se comporte comme $\widehat{a}_{\text{MV}}$ (la loi a priori est très peu informative et donc on fait confiance aux données).

Partie 2 : variance σ^2 inconnue

1) L'estimateur du maximum de vraisemblance du vecteur $\theta = (a, \sigma^2)$ s'obtient en annulant les dérivées partielles de la vraisemblance par rapport à a et σ^2. On obtient alors

$$\begin{aligned}
\frac{\partial \ln L(y_1, ..., y_n; a)}{\partial a} &= 0 \Longrightarrow a = \frac{\sum_{i=1}^n y_i x_i}{\|x\|^2}, \\
\frac{\partial \ln L(y_1, ..., y_n; a)}{\partial \sigma^2} &= 0 \Longrightarrow -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (y_i - a x_i)^2 = 0,
\end{aligned}$$

d'où

$$\begin{aligned}
a &= \frac{\sum_{i=1}^n y_i x_i}{\|x\|^2}, \\
\sigma^2 &= \frac{1}{n}\sum_{i=1}^n (y_i - a x_i)^2 .
\end{aligned}$$

On en conclut que l'estimateur du maximum de vraisemblance de $\theta = (a, \sigma^2)$ est $\widehat{\theta}_{\text{MV}} = \left(\widehat{a}_{\text{MV}}, \widehat{\sigma^2}_{\text{MV}}\right)$ avec

$$\begin{aligned}
\widehat{a}_{\text{MV}} &= \frac{\sum_{i=1}^n Y_i x_i}{\|x\|^2}, \\
\widehat{\sigma^2}_{\text{MV}} &= \frac{1}{n}\sum_{i=1}^n (Y_i - \widehat{a}_{\text{MV}} x_i)^2 .
\end{aligned}$$

2) On sait que
$$E\left[\widehat{a}_{\text{MV}}\right] = E\left[\frac{\sum_{i=1}^n Y_i x_i}{\|x\|^2}\right] = a.$$

Il reste donc à déterminer $E\left[\widehat{\sigma^2}_{\text{MV}}\right]$ comme suit
$$E\left[\widehat{\sigma^2}_{\text{MV}}\right] = \frac{1}{n}\sum_{i=1}^n E\left[(Y_i - \widehat{a}_{\text{MV}} x_i)^2\right].$$

Mais

$$\begin{aligned}
E\left[(Y_i - \widehat{a}_{\text{MV}} x_i)^2\right] &= E\left[Y_i^2\right] - 2E\left[x_i Y_i \widehat{a}_{\text{MV}}\right] + x_i^2 E\left[\widehat{a}_{\text{MV}}^2\right] \\
&= \sigma^2 + a^2 x_i^2 - 2E\left[(x_i Y_i)\, \widehat{a}_{\text{MV}}\right] + x_i^2 E\left[\widehat{a}_{\text{MV}}^2\right],
\end{aligned}$$

d'où

$$
\begin{aligned}
E\left[\widehat{\sigma^2}_{\mathrm{MV}}\right] &= \sigma^2 + a^2\frac{1}{n}\sum_{i=1}^n x_i^2 - \frac{2}{n}E\left[\left(\sum_{i=1}^n x_iY_i\right)\widehat{a}_{\mathrm{MV}}\right] + \left(\frac{1}{n}\sum_{i=1}^n x_i^2\right)E\left[\widehat{a}^2_{\mathrm{MV}}\right], \\
&= \sigma^2 + \left(\frac{1}{n}\sum_{i=1}^n x_i^2\right)\left(a^2 + E\left[\widehat{a}^2_{\mathrm{MV}}\right]\right) - \frac{2}{n}E\left[\|x\|^2\,\widehat{a}^2_{\mathrm{MV}}\right], \\
&= \sigma^2 + \left(\frac{1}{n}\|x\|^2\right)\left(a^2 + a^2 + \frac{\sigma^2}{\|x\|^2}\right) - \frac{2}{n}\|x\|^2\left(a^2 + \frac{\sigma^2}{\|x\|^2}\right), \\
&= \sigma^2\left(1 + \frac{1}{n} - \frac{2}{n}\right), \\
&= \frac{n-1}{n}\sigma^2.
\end{aligned}
$$

On en déduit qu'un estimateur non biaisé de σ^2 est

$$
\widetilde{\sigma^2} = \frac{n}{n-1}\widehat{\sigma^2}_{\mathrm{MV}} = \frac{1}{n-1}\sum_{i=1}^n \left(Y_i - \widehat{a}_{\mathrm{MV}}x_i\right)^2.
$$

3) Pour déterminer les variances minimales d'estimateurs non biaisés de a et σ^2, il faut déterminer l'inverse de la matrice d'information de Fisher

$$
\begin{pmatrix}
E\left[\frac{\partial^2 \ln L}{\partial a^2}\right] & E\left[\frac{\partial^2 \ln L}{\partial a \partial \sigma^2}\right] \\
E\left[\frac{\partial^2 \ln L}{\partial a \partial \sigma^2}\right] & E\left[\frac{\partial^2 \ln L}{\partial (\sigma^2)^2}\right]
\end{pmatrix}
$$

dont les termes se calculent comme suit

$$
\begin{aligned}
E\left[\frac{\partial^2 \ln L}{\partial a^2}\right] &= E\left[\frac{-1}{\sigma^2}\sum_{i=1}^n x_i^2\right] = \frac{-1}{\sigma^2}\sum_{i=1}^n x_i^2, \\
E\left[\frac{\partial^2 \ln L}{\partial a \partial \sigma^2}\right] &= E\left[\frac{-1}{\sigma^4}\left(\sum_{i=1}^n Y_ix_i - a\sum_{i=1}^n x_i^2\right)\right] = \frac{-1}{\sigma^4}\sum_{i=1}^n \left[x_i\left(ax_i\right) - ax_i^2\right] = 0, \\
E\left[\frac{\partial^2 \ln L}{\partial \left(\sigma^2\right)^2}\right] &= E\left[\frac{n}{2}\frac{1}{\sigma^4} + \frac{-2}{2}\sum_{i=1}^n \left(y_i - ax_i\right)^2\right] = \frac{n}{2\sigma^4} - \frac{n\sigma^2}{\sigma^6} = \frac{-n}{2\sigma^4}.
\end{aligned}
$$

La matrice d'information de Fisher est donc diagonale et s'écrit

$$
\begin{pmatrix}
E\left[\frac{\partial^2 \ln L}{\partial a^2}\right] & E\left[\frac{\partial^2 \ln L}{\partial a \partial \sigma^2}\right] \\
E\left[\frac{\partial^2 \ln L}{\partial a \partial \sigma^2}\right] & E\left[\frac{\partial^2 \ln L}{\partial (\sigma^2)^2}\right]
\end{pmatrix}
=
\begin{pmatrix}
\frac{-1}{\sigma^2}\sum_{i=1}^n x_i^2 & 0 \\
0 & \frac{-n}{2\sigma^4}
\end{pmatrix}.
$$

Son inverse vérifie

$$
\begin{pmatrix}
E\left[\frac{\partial^2 \ln L}{\partial a^2}\right] & E\left[\frac{\partial^2 \ln L}{\partial a \partial \sigma^2}\right] \\
E\left[\frac{\partial^2 \ln L}{\partial a \partial \sigma^2}\right] & E\left[\frac{\partial^2 \ln L}{\partial (\sigma^2)^2}\right]
\end{pmatrix}^{-1}
=
\begin{pmatrix}
\frac{-\sigma^2}{\|x\|^2} & 0 \\
0 & \frac{-2\sigma^4}{n}
\end{pmatrix}.
$$

On en déduit les inégalités

$$
\operatorname{var}\left[\widehat{a}\right] \geq \frac{\sigma^2}{\|x\|^2} \text{ et } \operatorname{var}\left[\widehat{\sigma^2}\right] \geq \frac{2\sigma^4}{n}.
$$

$$\boxed{\textbf{Exercice } 2 : \textbf{ Test de Neyman-Pearson}}$$

1) Le test de Neyman-Pearson est défini par

$$\text{Rejet de } H_0 \text{ si} \quad \frac{L(x_1, ..., x_n; \theta_1)}{L(x_1, ..., x_n; \theta_0)} > S_\alpha.$$

Mais

$$\begin{aligned}
\frac{L(x_1, ..., x_n; \theta_1)}{L(x_1, ..., x_n; \theta_0)} &> S_\alpha \Leftrightarrow \frac{(2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - a_1 x_i)^2\right]}{(2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - a_0 x_i)^2\right]} > S_\alpha, \\
&\Leftrightarrow \exp\left[\frac{1}{2\sigma^2}\left(\sum_{i=1}^n (y_i - a_0 x_i)^2 - \sum_{i=1}^n (y_i - a_1 x_i)^2\right)\right] > S_\alpha, \\
&\Leftrightarrow 2(a_1 - a_0)\sum_{i=1}^n y_i x_i + \left(a_0^2 - a_1^2\right)\sum_{i=1}^n x_i^2 > \nu_\alpha.
\end{aligned}$$

Pour $a_1 > a_0$, on en déduit le test équivalent

$$\text{Rejet de } H_0 \text{ si} \quad T_n = \sum_{i=1}^n Y_i x_i > K_\alpha.$$

La statistique du test de Neyman-Pearson est donc

$$T_n = \sum_{i=1}^n x_i Y_i.$$

La région critique de ce test est définie par

$$\left\{ y \in \mathbb{R}^n | \sum_{i=1}^n x_i Y_i > K_\alpha \right\}.$$

2) Le risque de première espèce α est défini par

$$\begin{aligned}
\alpha &= P\left[\text{Rejeter } H_0 \mid H_0 \text{ vraie}\right], \\
&= P\left[T_n > K_\alpha \mid a = a_0\right].
\end{aligned}$$

Puisque les variables aléatoires $Y_i, i = 1, ..., n$ sont indépendantes, le vecteur $(Y_1, ..., Y_n)$ est un vecteur Gaussien et donc dans la mesure où $x = (x_1, ..., x_n)$ est non nul, la transformation affine $\sum_{i=1}^n x_i Y_i$ est distribuée suivant une loi normale. La moyenne et la variance de cettte loi normale sont

$$\begin{aligned}
E\left[\sum_{i=1}^n x_i Y_i\right] &= \sum_{i=1}^n x_i (a x_i) = a \|x\|^2, \\
\text{var}\left[\sum_{i=1}^n x_i Y_i\right] &= \sigma^2 \left(\sum_{i=1}^n x_i^2\right) = \sigma^2 \|x\|^2.
\end{aligned}$$

On en déduit

$$\begin{aligned}
\alpha &= P\left[T_n > K_\alpha \mid a = a_0\right], \\
&= P\left[T_n > K_\alpha \mid T_n \sim \mathcal{N}\left(a_0 \|x\|^2, \sigma^2 \|x\|^2\right)\right], \\
&= P\left[U_n = \frac{T_n - a_0 \|x\|^2}{\sigma \|x\|} > \frac{K_\alpha - a_0 \|x\|^2}{\sigma \|x\|} \mid U_n \sim \mathcal{N}(0,1)\right], \\
&= G\left(\frac{K_\alpha - a_0 \|x\|^2}{\sigma \|x\|}\right).
\end{aligned}$$

De la même façon

$$\begin{aligned}
\beta &= P\left[\text{Rejeter } H_1 \mid H_1 \text{ vraie}\right], \\
&= P\left[T_n \leq K_\alpha \mid a = a_1\right], \\
&= 1 - P\left[T_n > K_\alpha \mid a = a_1\right], \\
&= 1 - P\left[T_n > K_\alpha \mid T_n \sim \mathcal{N}\left(a_1 \|x\|^2, \sigma^2 \|x\|^2\right)\right], \\
&= 1 - G\left(\frac{K_\alpha - a_1 \|x\|^2}{\sigma \|x\|}\right).
\end{aligned}$$

3) Les courbes caractéristiques opérationnelles du récepteur (courbes COR) expriment la puissance du test $\pi = 1 - \beta$ en fonction de α. Dans le cas présent, on a

$$\frac{K_\alpha - a_0 \|x\|^2}{\sigma \|x\|} = G^{-1}(\alpha).$$

Donc

$$K_\alpha = \sigma \|x\| G^{-1}(\alpha) + a_0 \|x\|^2.$$

En remplaçant cette valeur de K_α dans l'expression de la puissance, on obtient

$$\begin{aligned}
\pi &= G\left(\frac{K_\alpha - a_1 \|x\|^2}{\sigma \|x\|}\right), \\
&= G\left[G^{-1}(\alpha) - (a_1 - a_0)\frac{\|x\|}{\sigma}\right].
\end{aligned}$$

On voit donc que la performance du test dépend de la quantité

$$(a_1 - a_0)\frac{\|x\|}{\sigma}.$$

En particulier, plus $\|x\|$ est grand, plus la puissance du test est élevée.

$$\boxed{\textbf{Exercice } 3 : \textbf{Tests d'ajustement}}$$

1) La variable aléatoire U_i définie par

$$U_i = \frac{Y_i - ax_i}{\sigma}.$$

est clairement distribuée suivant une loi normale $\mathcal{N}(0,1)$. Pour effectuer le test du χ^2, il faut tout d'abord déterminer les quatre classes équiprobables. En vertu de la symétrie de la loi normale, ces classes sont nécessairement de la forme

$$\begin{array}{lcl} C_1 & : &]-\infty, -b[\,, \\ C_2 & : & [-b, 0[\,, \\ C_3 & : & [0, b[\,, \\ C_4 & : & [b, \infty[\,. \end{array}$$

On peut déterminer b de la façon suivante

$$\int_b^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du = G(b) = \frac{1}{4},$$

d'où

$$b = G^{-1}\left(\frac{1}{4}\right).$$

Ensuite, on détermine les nombres d'observation $u_1, ..., u_n$ appartenant à chaque classe notés $N_1, ..., N_4$. On calcule ensuite la statistique de test

$$\phi = \sum_{i=1}^{4} \frac{(N_i - np_i)^2}{np_i} = \frac{4}{n} \sum_{i=1}^{4} \left(N_i - \frac{n}{4}\right)^2.$$

On sait que ϕ est distribuée suivant une loi du χ^2_{K-1} sous l'hypothèse H_0, où $K = 4$ est le nombre classes. On rejette donc l'hypothèse H_0 si

$$\phi > s_{\alpha},$$

avec

$$\begin{array}{rcl} \alpha & = & P\left[\text{Rejeter } H_0 \mid H_0 \text{ vraie}\right], \\ & = & P\left[\phi > s_\alpha \mid \phi \sim \chi^2_3\right]. \end{array}$$

Si $g_n(u)$ est la densité d'une loi du χ^2_n, on pose

$$\int_x^{\infty} g_n(u) du = G_n(x),$$

et on obtient

$$\alpha = G_3\left(s_\alpha\right) \Longrightarrow s_\alpha = G_3^{-1}\left(\alpha\right).$$

2) Si Z_i est la variable aléatoire définie par

$$Z_i = \begin{cases} 1 \text{ si } \left|\frac{Y_i}{x_i}\right| > 2a \\ 0 \text{ sinon} \end{cases},$$

on a

$$\begin{aligned}
p_i &= P\left[Z_i = 1 \middle| H_0 \text{ vraie}\right], \\
&= P\left[\left|\frac{Y_i}{x_i}\right| > 2a \middle| Y_i \sim \mathcal{N}(a_0 x_i, \sigma^2)\right], \\
&= P\left[|Y_i| > 2a_0 x_i \middle| Y_i \sim \mathcal{N}(a_0 x_i, \sigma^2)\right], \\
&= P\left[Y_i > 2a_0 x_i \middle| Y_i \sim \mathcal{N}(a_0 x_i, \sigma^2)\right] \\
&\quad + P\left[Y_i < -2a_0 x_i \middle| Y_i \sim \mathcal{N}(a_0 x_i, \sigma^2)\right], \\
&= P\left[U_i = \frac{Y_i - a_0 x_i}{\sigma} > \frac{a_0 x_i}{\sigma} \middle| U_i \sim \mathcal{N}(0,1)\right] \\
&\quad + P\left[U_i = \frac{Y_i - a_0 x_i}{\sigma} < \frac{-3a_0 x_i}{\sigma} \middle| U_i \sim \mathcal{N}(0,1)\right], \\
&= G\left(\frac{a_0 x_i}{\sigma}\right) + G\left(\frac{3a_0 x_i}{\sigma}\right).
\end{aligned}$$

La moyenne et la variance de la variable aléatoire $Z = \sum_{i=1}^{n} Z_i$ sont alors définies par

$$\begin{aligned}
m_Z &= E[Z] = \sum_{i=1}^{n} p_i, \\
\sigma_Z^2 &= \text{var}[Z] = \sum_{i=1}^{n} p_i q_i.
\end{aligned}$$

Le risque α associé à la stratégie

$$\text{Rejet de } H_0 \text{ si } Z > \frac{n}{10}.$$

se détermine comme suit

$$\begin{aligned}
\alpha &= P\left[Z > \frac{n}{10} \middle| H_0 \text{ vraie}\right], \\
&= P\left[U = \frac{Z - m_Z}{\sigma_Z} > \frac{\frac{n}{10} - m_Z}{\sigma_Z} \middle| U \sim \mathcal{N}(m_Z, \sigma_Z^2)\right], \\
&= G\left(\frac{\frac{n}{10} - m_Z}{\sigma_Z}\right).
\end{aligned}$$