Proof of formula in *Gaussian Process Regression with Censored Data Using Expectation Propagation, Groot, Lucas* [GL12]

12 octobre 2021

1 Formula (14)

1.1 Expression de $\hat{Z}_i$

$\hat{Z}_i$ est défini par

$$\hat{Z}_i = \int q_{\setminus i} t_i df_i \tag{1}$$

avec $q_{\setminus i} = \mathcal{N}\left(\mu_{\setminus i}, \sigma^2_{\setminus i}\right)$ et $t_i = p(y_i|f_i)$.

Dans le cas $y_i = l$, on a

$$p(y_i = l|f_i) = 1 - \Phi\left(\frac{f_i - l}{\sigma}\right) \tag{2}$$

où Φ est la fonction de répartition de la loi normal centrée réduite.

On a alors

$$\begin{aligned}
\hat{Z}_i &= \int \left(1 - \Phi\left(\frac{f_i - l}{\sigma}\right)\right) \mathcal{N}\left(f_i; \mu_{\setminus i}, \sigma^2_{\setminus i}\right) df_i & (3)\\
&= 1 - \int \Phi\left(\frac{f_i - l}{\sigma}\right) \mathcal{N}\left(f_i; \mu_{\setminus i}, \sigma^2_{\setminus i}\right) df_i & (4)\\
&= 1 - E_{q_{\setminus i}}\left(\Phi\left(\frac{f_i - l}{\sigma}\right)\right) & (5)\\
&= 1 - E_{q_{\setminus i}}\left(P\left(X \leq \frac{f_i - l}{\sigma}\right)\right) & (6)
\end{aligned}$$

où X est une VA suivant une loi normale centrée réduite. X et f_i étant des VA indépendantes, on a alors simplement

$$\begin{aligned}
\hat{Z}_i &= 1 - P\left(X \leq \frac{f_i - l}{\sigma}\right) & (7)\\
&= 1 - P\left(X - \frac{f_i}{\sigma} \leq \frac{-l}{\sigma}\right) & (8)\\
&= 1 - P\left(Z \leq \frac{-l}{\sigma}\right) & (9)
\end{aligned}$$

où $Z = X - \frac{f_i}{\sigma} \sim \mathcal{N}\left(\frac{-\mu_{\setminus i}}{\sigma}, 1 + \frac{\sigma^2_{\setminus i}}{\sigma^2}\right)$. On en déduit, avec Z' une loi normale centrée réduite telle que

$Z = \frac{-\mu_{\setminus i}}{\sigma} + \sqrt{1+\frac{\sigma^2_{\setminus i}}{\sigma^2}}Z'$,

$$\begin{aligned}\hat{Z}_i &= 1 - P\left(Z' \leq \frac{\mu_{\setminus i} - l}{\sigma\sqrt{1+\frac{\sigma^2_{\setminus i}}{\sigma^2}}}\right) && (10)\\ &= 1 - P\left(Z' \leq \frac{\mu_{\setminus i} - l}{\sqrt{\sigma^2+\sigma^2_{\setminus i}}}\right) && (11)\\ &= 1 - \Phi\left(\frac{\mu_{\setminus i} - l}{\sqrt{\sigma^2+\sigma^2_{\setminus i}}}\right) && (12)\end{aligned}$$

2 Moyenne de la prédiction $p(y|\mathcal{D})$

La formule n'est pas présente dans le papier mais dans le code source Matlab à la fonction *lik_tobit1_predy*. Seule la moyenne est donnée par

$$\begin{aligned}E(y|\mathcal{D}) &= \left(\Phi\left(\frac{u-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}}\right) - \Phi\left(\frac{l-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}}\right)\right)\left(\mu_f + \sqrt{\sigma^2+\sigma_f^2}\frac{\phi\left(\frac{l-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}}\right) - \phi\left(\frac{u-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}}\right)}{\Phi\left(\frac{u-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}}\right) - \Phi\left(\frac{l-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}}\right)}\right)\\ &\quad + l\Phi\left(\frac{l-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}}\right)\mathbb{1}_{l>-\infty} + u\left(1-\Phi\left(\frac{u-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}}\right)\right)\mathbb{1}_{u<+\infty} && (13)\end{aligned}$$

Démonstration. Tout d'abord, pour simplifier l'écriture on pose :

$$\begin{aligned}a &= \frac{l-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}},\ b = \frac{u-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}}, s = \sqrt{\sigma^2+\sigma_f^2} && (14)\\ \Phi_a &= \Phi\left(\frac{l-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}}\right),\ \phi_a = \phi\left(\frac{l-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}}\right) && (15)\\ \Phi_b &= \Phi\left(\frac{u-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}}\right),\ \phi_b = \phi\left(\frac{u-\mu_f}{\sqrt{\sigma^2+\sigma_f^2}}\right) && (16)\end{aligned}$$

On a

$$p(y|\mathcal{D}) = \int p(y|f,\mathcal{D})p(f|\mathcal{D})df \qquad (17)$$

L'espérance est donc donnée par les trois termes de la vraisemblance Tobit

$$\begin{aligned}E(y|\mathcal{D}) &= l\ p(y=l|\mathcal{D}) + u\ p(y=u|\mathcal{D}) + \int_{y=l}^{u} y\ p(y|\mathcal{D})dy && (18)\\ &= l\int_f p(y=l|f,\mathcal{D})p(f|\mathcal{D})df + u\int_f p(y=u|f,\mathcal{D})p(f|\mathcal{D})df\\ &\quad + \int_{y=l}^{u} y\int_f p(y|f,\mathcal{D})p(f|\mathcal{D})df dy && (19)\end{aligned}$$

Pour le premier terme, on a en utilisant la même technique qu'à la section précédente pour le calcul de $\hat{Z}_i$

$$\begin{aligned}l\int_f p(y=l|f,\mathcal{D})p(f|\mathcal{D})df &= l\int_f\left(1-\Phi\left(\frac{f-l}{\sigma}\right)\right)p(f|\mathcal{D})df && (20)\\ &= l\Phi_a && (21)\end{aligned}$$

De même, on a :

$$u \int_f p(y=u|f,\mathcal{D})p(f|\mathcal{D})df = u\left(1-\Phi_b\right) \tag{22}$$

Pour le dernier terme, on utilise la formule sur le produit de deux pdf gaussiennes :

$$\begin{aligned}
\int_{y=l}^{u} y \int_f p(y|f,\mathcal{D})p(f|\mathcal{D})df dy &= \int_{y=l}^{u} y\mathcal{N}\left(y|\mu_f, \sigma^2+\sigma_f^2\right) dy &(23)\\
&= \int_{y=l}^{u} \frac{y}{s}\phi\left(\frac{y-\mu_f}{s}\right) dy &(24)\\
&= \int_{y=a}^{b} \left(sy+\mu_f\right)\phi(y)dy &(25)\\
&= \mu_f\left(\Phi_b-\Phi_a\right) + s\int_{y=a}^{b} y\phi(y)dy &(26)
\end{aligned}$$

Sachant que la primitive de Φ est donnée par $\int \Phi(x)dx = x\Phi(x)+\phi(x)+C$, on a par IPP

$$\begin{aligned}
\int_{y=a}^{b} y\phi(y)dy &= \left[y\Phi(y)\right]_a^b - \int_a^b \Phi(y)dy &(27)\\
&= b\Phi_b - a\Phi_a - \left(b\Phi_b - a\Phi_a + \phi_a - \phi_b\right) &(28)\\
&= \phi_a - \phi_b &(29)
\end{aligned}$$

ce qui donne

$$\int_{y=l}^{u} y \int_f p(y|f,\mathcal{D})p(f|\mathcal{D})df dy = \mu_f\left(\Phi_b-\Phi_a\right) + s\left(\phi_a-\phi_b\right) \tag{30}$$

On obtient donc le résultat final

$$E(y|\mathcal{D}) \quad = \quad l\Phi_a + u\left(1-\Phi_b\right) + \mu_f\left(\Phi_b-\Phi_a\right) + s\left(\phi_a-\phi_b\right) \tag{31}$$

$\square$

3 Variance de la prédiction $p(y|\mathcal{D})$

La variance de la prédiction n'est pas implémentée dans le code source du papier mais est facilement calculable. En reprenant les même calculs que précédemment, on obtient

$$Var(Y|\mathcal{D}) = E(Y^2|\mathcal{D}) - E(Y|\mathcal{D})^2 \tag{32}$$

Le deuxième terme est directement obtenu à la section précédente. Pour le premier terme, on calcule

$$\int_a^b y^2\phi(y)dy = \left[y^2\Phi(y)\right]_a^b - 2\int_a^b y\Phi(y)dy \tag{33}$$

et avec encore une IPP

$$\int_a^b y\Phi(y)dy = \frac{1}{2}\left(b(b\Phi_b+\phi_b) - a(a\Phi_a+\phi_a)\right) \tag{34}$$

On obtient finalement

$$E(Y^2|\mathcal{D}) = l^2\Phi_a + u^2\left(1-\Phi_b\right) + \mu_f^2\left(\Phi_b-\Phi_a\right) + s^2\left(a\phi_a+\Phi_a-b\phi_b-\Phi_b\right) + 2s\mu_f\left(\phi_a-\phi_b\right) \tag{35}$$

4 Dérivée seconde de la log-vraisemblance

La dérivée seconde de la log-vraisemblance se calcule facilement et le résultat est donné par

$$\frac{\partial^2 \log L}{\partial f_i^2} = \frac{1}{\sigma^2}\left[\phi(z_l)\frac{z_l\left(1-\Phi(z_l)\right)-\phi(z_l)}{\left(1-\Phi(z_l)\right)^2}\mathbf{1}_{[y_i=l]} - \phi(z_u)\frac{z_u\Phi(z_u)+\phi(z_u)}{\Phi(z_u)^2}\mathbf{1}_{[y_i=u]} - 1\mathbf{1}_{[l<y_i<u]}\right] \tag{36}$$

avec $z_l = \frac{f_i - l}{\sigma}, z_u = \frac{f_i - u}{\sigma}$. Afin d'exploiter l'approximation du ratio $\frac{\phi}{\Phi}$ pour des petites valeurs de z donnée dans [GL] par

$$\frac{\phi(z)}{\Phi(z)} \sim \sqrt{\frac{1}{4}z^2 + 1} - \frac{1}{2}z \tag{37}$$

et en exploitant le fait que [GL12]

$$\frac{\phi(z)}{1 - \Phi(z)} = \frac{\phi(-z)}{\Phi(-z)} \tag{38}$$

on peut réécrire la dérivée seconde comme suit

$$\frac{\partial^2 \log L}{\partial f_i^2} = \frac{1}{\sigma^2}\left[\frac{\phi(-z_l)}{\Phi(-z_l)}\left(z_l - \frac{\phi(-z_l)}{\Phi(-z_l)}\right)\mathbf{1}_{[y_i=l]} - \frac{\phi(z_u)}{\Phi(z_u)}\left(z_u + \frac{\phi(z_u)}{\Phi(z_u)}\right)\mathbf{1}_{[y_i=u]} - 1\mathbf{1}_{[l<y_i<u]}\right] \tag{39}$$

Références

[GL] Perry Groot and Peter Lucas. Supplementary material for 'gaussian process regression with censored data using expectation propagation'.

[GL12] Perry Groot and Peter JF Lucas. Gaussian process regression with censored data using expectation propagation. 2012.