PMSE Approximation

November 17, 2022

1   Approximation

To investigate the unconditional prediction square error distribution of $(\hat{y} - y^*)^2$, we start with moments of $(\hat{y} - y^*)$, proven by Richard Sawyer: Let M be a positive integer. Then

$$\mathbb{E}\left[(\hat{y} - y^*)^{2M}\right] = \frac{\sigma^{2M} \frac{(2M)!}{M!} \left(\frac{n+1}{2n}\right)^M \prod_{j=1}^{M}(n-2j)}{\prod_{j=1}^{M}(n-p-2j)}$$

when $2M \leq n - p - 1$.

If we consider the distribution $\hat{y} - y^*$ follows asymptotically normal distribution, then $(\hat{y} - y^*)^2$ is likely to be approximate Gamma distribution, by the proposition above,

$$\mathbb{E}\left[(\hat{y} - y^*)^2\right] = \sigma^2 \frac{(n+1)(n-2)}{n(n-p-2)} \tag{1}$$

$$\mathbb{E}\left[(\hat{y} - y^*)^4\right] = 3\sigma^4 \frac{(n+1)^2(n-2)(n-4)}{n^2(n-p-2)(n-p-4)} \tag{2}$$

$$\text{Var}\left[(\hat{y} - y^*)^2\right] = \sigma^4 \frac{(n+1)^2(n-2)}{n^2(n-p-2)} \left(\frac{3n-12}{n-p-4} - \frac{n-2}{n-p-2}\right) \tag{3}$$

Applying method of moments, for $Gamma(\alpha, \beta)$ with shape-scale parameters

$$\mathbb{E}(X^n) = \frac{\beta^n(n+\alpha-1)!}{(\alpha-1)!}$$

$$\mathbb{E}X = \alpha\beta, \mathbb{E}X^2 = \alpha(\alpha+1)\beta^2$$

let

$$\begin{cases} \mathbb{E}\left[(\hat{y} - y^*)^2\right] = \alpha\beta \\ \mathbb{E}\left[(\hat{y} - y^*)^2\right] = \alpha(\alpha+1)\beta^2 \end{cases} \Rightarrow \begin{cases} \alpha = \frac{(n-2)(n-p-4)}{3(n-4)(n-p-2)-(n-2)(n-p-4)} \\ \beta = \sigma^2 \frac{n+1}{n}\left(\frac{3(n-4)}{n-p-4} - \frac{n-2}{n-p-2}\right) \end{cases}$$

Based on normal assumption, we approximate the prediction square error as

$$Gamma\left(\frac{(n-2)(n-p-4)}{3(n-4)(n-p-2)-(n-2)(n-p-4)}, \sigma^2\frac{n+1}{n}\left(\frac{3(n-4)}{n-p-4}-\frac{n-2}{n-p-2}\right)\right)$$

$$(4)$$

For $n-p \geq 5$, the approximation using Gram-Charlier is

$$P(\hat{y}-y^* \leq t) \doteq \Phi\left(\frac{t}{\sigma'}\right) + \frac{p}{4(n-2)(n-p-4)}\Phi^{(4)}\left(\frac{t}{\sigma'}\right) \qquad (5)$$

where Φ is the standard normal distribution function, $\Phi^{(4)}$ is its fourth derivative,

$$\sigma' = \sqrt{MSE} = \sigma\sqrt{\frac{(n+1)(n-2)}{n(n-p-2)}}$$

Though adding second term may not even improve the approximation by Sawyer's Appendix, we examine the approximation for $P((\hat{y}-y^*)^2 \leq t)$ distribution.

Denote $X = \hat{y}-y^*, Y = X^2 = (\hat{y}-y^*)^2$

$$\begin{aligned}
F_Y(t) &= P(Y \leq t) = P(|\hat{y}-y^*| \leq \sqrt{t}) \\
&= P(\hat{y}-y^* \leq \sqrt{t}) - P(\hat{y}-y^* \leq -\sqrt{t}) \\
&= \Phi(\frac{\sqrt{t}}{\sigma'}) - \Phi(-\frac{\sqrt{t}}{\sigma'}) = 2\Phi(\frac{\sqrt{t}}{\sigma'}) - 1 \quad (first \quad term)
\end{aligned}$$

$$or$$

$$\begin{aligned}
&= \Phi(\frac{\sqrt{t}}{\sigma'}) + \frac{p}{4(n-2)(n-p-4)}\Phi^{(4)}(\frac{\sqrt{t}}{\sigma'}) \\
&\quad - \Phi(-\frac{\sqrt{t}}{\sigma'}) - \frac{p}{4(n-2)(n-p-4)}\Phi^{(4)}(-\frac{\sqrt{t}}{\sigma'}) \quad (first\&second \quad term)
\end{aligned}$$

where $\Phi^{(4)}(t) = \frac{1}{\sigma^4}(t^4 - 6t^2 + 3)\Phi(t)$ for standard normal distribution by Gram-Charlier definition. Thus, the approximation using first two terms can be written as

$$\begin{aligned}
&2\Phi(\frac{\sqrt{t}}{\sigma'}) - 1 + \frac{p}{4(n-2)(n-p-4)}\left\{\frac{1}{\sigma^4}\left[(\frac{\sqrt{t}}{\sigma'})^4 - 6(\frac{\sqrt{t}}{\sigma'})^2 + 3\right]\left(2\Phi(\frac{\sqrt{t}}{\sigma'}) - 1\right)\right\} \\
&= \left(2\Phi(\frac{\sqrt{t}}{\sigma'}) - 1\right)\left(1 + \frac{p}{4\sigma^4(n-2)(n-p-4)}\left[\frac{t^2}{MSE^2} - 6\frac{t}{MSE} + 3\right]\right)
\end{aligned}$$

$$(6)$$

Plug in that $MSE = \sigma^2\frac{(n+1)(n-2)}{n(n-p-2)}$

By the result of MAE approximation, normal distribution is good enough with no reason

adding the second term of Gram-Charlier. Thus, Chi-square distribution

$$\sigma^2 \chi_1^2 \tag{7}$$

may be a good choice for MSE distribution approximation.

2   Simulation

We keep the parameter settings from MAE approximation with number of predictor p 1, 2, 3, 5, 8, 10 and sample size n 10, 25, 50, 75. To satisfy the requirement from variance calculation $n - p \geq 5$, we removed sample size 10 with $p = 8, 10$ and added the sample size 100. The mean of predictors $\boldsymbol{\mu}$ and the covariance between each pair of predictors are set as constant, so as the intercept α of linear model and the error variance σ^2.

The unconditional mean square error $(\hat{y} - y^*)^2$ is generated by 10,000 random samples for each of several values of n and p. In the simulation, the response variable is generated by predictors with linear relationship and the new data is drawn from the normal distribution with same mean and variance as predictors. We calculated the empirical cumulative distribution function as standard function and compared with the values predicted by Gram-charlier approximations, gamma approximation and chi-square approximation.

The results are displayed in MSE approximation Table. In this table $\hat{F}$ denoted the empirical distribution function of $(\hat{y} - y^*)^2$ corresponding to the unknown distribution F as true distribution of MSE. $\hat{F}$ is computed from the 10,000 random samples in each category defined by n and p. F_1 and F_2 are the computed approximation based on the first 5 and second 6 Gram-Charlier approximations, respectively to F. F_3 and F_4 are the gamma distribution 4 and chi-square distribution 7 as we discussed above. MSE_{est} and $Var(SE)_{est}$ are the estimated mean and variance under assumption for square error using equation 1 and 3. Several random seeds are provided to eliminate randomness.

For different random seeds, the performance of each approximation is evaluated by the absolute value of maximum difference between estimated distribution function and empirical distribution function. Although we can find the best-performing approximation for each category, the differences between the 4 estimates is actually not significant and not greater than 0.015 most of the time except one case, $(p, n) = (5, 10)$. The variance of square error estimation varies from empirical square error variance for (5,10), corresponding larger inflation factor K may lead to poor accuracy of approximation. Considering (5,10), the chi-square approximation is the best-performing approximation.