We have some latent variable Z, and some noisy estimator X of Z. Suppose the amount of noise depends on the value of Z. Can we produce a better variance estimator for Z?

1 Toy example

Suppose that Z has two classes which are known. Suppose for the observations in class 0, the measurement error is normal with mean 0 and variance ${\sigma_1}^2$, and for observations in class 1, the measurement error is normal with mean 0 and variance ${\sigma_2}^2$. Let C be the indicator variable for class 1. In particular, we let $X = Z + E$, where E is uncorrelated with Z and follows a normal distribution with mean 0 and variance ${\sigma_1}^2$ if Z is in class 0 and ${\sigma_2}^2$ if Z is in class 1. We have

$$\begin{aligned}\operatorname{Var}(Z) = P(C=1)\operatorname{Var}(Z|C=1) + P(C=0)\operatorname{Var}(Z|C=0) \\ + 2P(C=0)P(C=1)\left(\mathbb{E}(Z|C=0) - \mathbb{E}(Z|C=1)\right)^2\end{aligned}$$

Furthermore, we have that $\operatorname{Var}(Z|C=1) = \operatorname{Var}(X|C=1) - {\sigma_2}^2$ and $\operatorname{Var}(Z|C=0) = \operatorname{Var}(X|C=0) - {\sigma_1}^2$, so the natual estimator

$$\widehat{\operatorname{Var}(Z)} = \widehat{\operatorname{Var}(X)} - P(C=1){\sigma_2}^2 - P(C=0){\sigma_1}^2$$

can be rewritten as

$$\begin{aligned}\widehat{\operatorname{Var}(Z)} = P(C=1)\left(\widehat{\operatorname{Var}(X|C=1)} - {\sigma_2}^2\right) + P(C=0)\left(\widehat{\operatorname{Var}(X|C=1)} - {\sigma_1}^2\right) \\ + 2P(C=0)P(C=1)\left(\widehat{\mathbb{E}(X|C=0)} - \widehat{\mathbb{E}(X|C=1)}\right)^2\end{aligned}$$

Now recall that

$$\widehat{\operatorname{Var}(X|C=1)} = \frac{1}{n_1 - 1}\sum_{C_i=1} {X_i}^2 - \frac{n_1}{n_1-1}\left(\widehat{\mathbb{E}(X|C=1)}\right)^2$$

has mean $\operatorname{Var}(Z|C=1) + {\sigma_2}^2$ and variance

$$\widehat{\operatorname{Var}(X|C=1)} = \frac{1}{n_1(n_1-1)}\sum_{i,j}\left(X_i - X_j\right)^2$$

has mean $\operatorname{Var}(Z|C=1) + {\sigma_2}^2$ and raw second moment

$$\frac{1}{{n_1}^2(n_1-1)^2}\mathbb{E}\left(\sum_{i,j}\left(X_i - X_j\right)^4 + 2\sum_{i,j,k}\left(X_i - X_j\right)^2\left(X_i - X_k\right)^2 + \sum_{i,j,k,l}\left(X_i - X_j\right)^2\left(X_k - X_l\right)^2\right)$$

We also have that $X_i - X_j = Z_i - Z_j + E_i - E_j$, and we know that $E_i - E_j$ is normal with mean 0 and variance $2{\sigma_2}^2$. We can therefore expand this raw second moment by noting

$$\begin{aligned}
\mathbb{E}\left((X_i - X_j)^4\right) &= \mathbb{E}\left((Z_i - Z_j)^4 + 6\,(Z_i - Z_j)^2\,(E_i - E_j)^2 + (E_i - E_j)^4\right) \\
&= \mathbb{E}\left((Z_i - Z_j)^4\right) + 6{\sigma_2}^2\mathbb{E}\left((Z_i - Z_j)^2\right) + 3{\sigma_2}^4 \\
\mathbb{E}\left((X_i - X_j)^2\,(X_i - X_k)^2\right) &= \mathbb{E}\Big((Z_i - Z_j)^2\,(Z_i - Z_k)^2 + (Z_i - Z_j)^2\,(E_i - E_k)^2 + (Z_i - Z_k)^2\,(E_i - E_j)^2 \\
&\qquad + (Z_i - Z_j)\,(E_i - E_j)\,(Z_i - Z_k)\,(E_i - E_k) + (E_i - E_j)^2(E_i - E_k)^2\Big) \\
&= \mathbb{E}\left((Z_i - Z_j)^2\,(Z_i - Z_k)^2\right) + {\sigma_2}^2\mathbb{E}\left((Z_i - Z_j)^2 + (Z_i - Z_k)^2\right) \\
&\qquad + \mathbb{E}\left((E_i - E_j)(E_i - E_k)\right)\mathbb{E}\left((Z_i - Z_j)\,(Z_i - Z_k)\right) + \mathbb{E}\left((E_i - E_j)^2(E_i - E\right. \\
\mathbb{E}\left((X_i - X_j)^2\,(X_k - X_l)^2\right) &= \mathbb{E}\Big((Z_i - Z_j)^2\,(Z_k - Z_l)^2 + (Z_i - Z_j)^2\,(E_k - E_l)^2 + (Z_k - Z_l)^2\,(E_i - E_j)^2 \\
&\qquad + (E_i - E_j)^2(E_k - E_l)^2\Big) \\
&= \mathbb{E}\left((Z_i - Z_j)^2\,(Z_k - Z_l)^2\right) + {\sigma_2}^2\mathbb{E}\left((Z_i - Z_j)^2 + (Z_k - Z_l)^2\right) + {\sigma_2}^4
\end{aligned}$$

We also have

$$\begin{aligned}
\mathbb{E}\left((E_i - E_j)(E_i - E_k)\right) &= \mathbb{E}\left({E_i}^2 - E_i(E_j + E_k) + E_jE_k\right) \\
&= {\sigma_2}^2 \\
\mathbb{E}\left((E_i - E_j)^2(E_i - E_k)^2\right) &= \mathbb{E}\left({E_i}^4 - {E_i}^2({E_j}^2 + {E_k}^2 + 4E_jE_k) + {E_j}^2{E_k}^2\right) \\
&= 3{\sigma_2}^4 - 2{\sigma_2}^4 + {\sigma_2}^4 \\
&= 2{\sigma_2}^4
\end{aligned}$$

If we let ${\rho_2}^2$ be the variance of the corresponding estimator for $\mathrm{Var}(Z|C=1)$ based on the true values of Z_i, then we see that the variance of $\widehat{\mathrm{Var}(X|C=1)}$ is

$${\rho_2}^2 + {\sigma_2}^2(6n(n-1) + 2n(n-1)(n-2) + 2n(n-1)(n-2)(n-3))\mathbb{E}\left((Z_i - Z_j)^2\right)$$

Note that terms of odd degree vanish.

Recall that our objective is to obtain an estimator for

$$\mathrm{Var}(Z) = P(C=1)\,\mathrm{Var}(Z|C=1) + P(C=0)\,\mathrm{Var}(Z|C=0) + P(C=0)P(C=1)\left(\mathbb{E}(Z|C=1) - \mathbb{E}(Z|C=0)\right)^2$$

We have that

$$\begin{aligned}
\mathrm{Var}(Z|C=0) &= \mathrm{Var}(X|C=0) - {\sigma_1}^2 \\
\mathrm{Var}(Z|C=1) &= \mathrm{Var}(X|C=1) - {\sigma_2}^2 \\
\mathbb{E}(Z|C=1) - \mathbb{E}(Z|C=0) &= \mathbb{E}(X|C=1) - \mathbb{E}(X|C=0)
\end{aligned}$$

The natural choice for estimator is therefore the unbiassed

$$\begin{aligned}
\widehat{\mathrm{Var}(Z)} = P(C=1)\left(\mathrm{Var}(\widehat{X|C}=1) - {\sigma_2}^2\right) + P(C=0)\left(\mathrm{Var}(\widehat{X|C}=0) - {\sigma_1}^2\right) \\
+ P(C=0)P(C=1)\left(\mathbb{E}(\widehat{X|C}=1) - \mathbb{E}(\widehat{X|C}=0)\right)^2
\end{aligned}$$

However, if ${\sigma_2}^2$ is large, then the variance of $\mathrm{Var}(\widehat{X|C}=1)$ is large, so we may be able to improve the accuracy of the estimation by reducing the weight of $\mathrm{Var}(\widehat{X|C}=1)$. This approach is particularly effective if we have reason to believe that $\mathrm{Var}(Z|C=1) \approx \mathrm{Var}(Z|C=1)$.

2 Variance of Log-transformed Λ