

Variance Estimation using Proposal Priors

November 14, 2017

We fix our prior $p(\mathbf{x})$ with variance σ^2 and our proposal prior $\tilde{p}(\mathbf{x})$. Sampling independent samples \mathbf{x}_n from $\tilde{p}(\mathbf{x})$, the MLE estimate for the variance of $p(\mathbf{x})$ is given by

$$\frac{\sum_{n=1}^N w_n (\mathbf{x}_n - \boldsymbol{\mu})^2}{\sum_n w_n} \quad (1)$$

where the are the importance weights. We write

$$\sum_{n=1}^N w_n = N + \sum_{n=1}^N (w_n - 1) =: N + \sum_{n=1}^N \varepsilon_n =: N + \varepsilon$$

Here, ε is a random variable with variance $\approx N$. For large enough N , the ratio $\frac{\varepsilon}{N}$ (the average deviation around the mean) is likely to be small and we can perform a Taylor expansion of Eq. 1 with respect to ε :

$$\frac{\sum_{n=1}^N w_n (\mathbf{x}_n - \boldsymbol{\mu})^2}{N + \varepsilon} = \frac{1}{N} \left(\sum_{n=1}^N w_n (\mathbf{x}_n - \boldsymbol{\mu})^2 \right) \left(1 - \frac{\varepsilon}{N} + \frac{\varepsilon^2}{N^2} - \frac{\varepsilon^3}{N^3} + \dots \right)$$

We now focus on the linear term. A simple calculation yields:

$$\begin{aligned} \frac{1}{N} \left(\sum_{n=1}^N w_n (\mathbf{x}_n - \boldsymbol{\mu})^2 \right) \frac{\varepsilon}{N} &= \frac{1}{N^2} \left(\sum_{n=1}^N w_n (\mathbf{x}_n - \boldsymbol{\mu})^2 \right) \left(\sum_{n=1}^N (w_n - 1) \right) \\ &= \frac{1}{N^2} \sum_{n \neq m=1}^N w_n w_m (\mathbf{x}_n - \boldsymbol{\mu})^2 + \frac{1}{N^2} \sum_{n=1}^N w_n^2 (\mathbf{x}_n - \boldsymbol{\mu})^2 - \frac{1}{N} \sum_{n=1}^N w_n (\mathbf{x}_n - \boldsymbol{\mu})^2 \end{aligned}$$

Thus we get

$$\frac{1}{N} \left(\sum_{n=1}^N w_n (\mathbf{x}_n - \boldsymbol{\mu})^2 \right) \left(1 - \frac{\varepsilon}{N} \right) = \frac{2}{N} \sum_{n=1}^N w_n (\mathbf{x}_n - \boldsymbol{\mu})^2 - \frac{1}{N^2} \sum_{n \neq m=1}^N w_n w_m (\mathbf{x}_n - \boldsymbol{\mu})^2 - \frac{1}{N^2} \sum_{n=1}^N w_n^2 (\mathbf{x}_n - \boldsymbol{\mu})^2$$

This is a first-order (in $\frac{\varepsilon}{N}$) approximation of Eq. 1. Keeping in mind that the weights are functions of the (independent) random variables \mathbf{x}_n , the expected value is equal to

$$\begin{aligned} 2\mathbb{E}_{p(\mathbf{x})}((\mathbf{x} - \boldsymbol{\mu})^2) - \frac{N-1}{N} \mathbb{E}_{p(\mathbf{x})}((\mathbf{x} - \boldsymbol{\mu})^2) - \frac{1}{N} \mathbb{E}_{p(\mathbf{x})}(w(\mathbf{x} - \boldsymbol{\mu})^2) \\ = \mathbb{E}_{p(\mathbf{x})}((\mathbf{x} - \boldsymbol{\mu})^2) + \frac{1}{N} (E_{p(\mathbf{x})}((\mathbf{x} - \boldsymbol{\mu})^2) - \mathbb{E}_{p(\mathbf{x})}(w(\mathbf{x} - \boldsymbol{\mu})^2)) \end{aligned}$$

where $\mathbb{E}_{p(\mathbf{x})}$ denotes the expected value with respect to the prior $p(\mathbf{x})$. We read off from this that our estimator has a bias of approximately

$$\frac{1}{N} (E_{p(\mathbf{x})}(\mathbf{x}^2) - E_{p(\mathbf{x})}(w\mathbf{x}^2))$$

This is consistent with the fact that the bias should disappear as $N \rightarrow \infty$, or if $w \equiv 1$, ie. $p(\mathbf{x}) = \tilde{p}(\mathbf{x})$.

A note on the validity of the Taylor expansion: ε has variance $\frac{1}{N}$. By the Chebyshev inequality, ε will be very small with an overwhelming probability, so the Taylor expansion can be applied. In any case, this is just an asymptotic estimate of the bias.

We apply this to the model where $p(\mathbf{x}) \sim \mathcal{N}(0, \sigma^2)$ and $\tilde{p}(\mathbf{x}) \sim \mathcal{N}(0, \eta^2)$, with $\eta < \sigma$. Then we have

$$w(\mathbf{x}) = \frac{\eta}{\sigma} e^{\frac{d}{2} \mathbf{x}^2}$$

where we let $d = \frac{1}{\eta^2} - \frac{1}{\sigma^2} > 0$. From the above expression we get that our estimator is biased towards underestimating the correct variance if $\mathbb{E}_{p(\mathbf{x})}(w\mathbf{x}^2) > \mathbb{E}_{p(\mathbf{x})}(\mathbf{x}^2)$, which is the case for our toy problem.

We now perform the full Taylor expansion, considering all higher-order terms. For this it will prove convenient to define the shifted weights $\tilde{w}_n = w_n - 1$. It follows that

$$\begin{aligned} \varepsilon^k &= \sum_{n_1, \dots, n_k=1}^N \tilde{w}_{n_1} \dots \tilde{w}_{n_k} \\ \sum_{n=1}^N w_n (\mathbf{x}_n - \boldsymbol{\mu})^2 &= \sum_{n=1}^N \tilde{w}_n (\mathbf{x}_n - \boldsymbol{\mu})^2 + \sum_{n=1}^N (\mathbf{x}_n - \boldsymbol{\mu})^2 \end{aligned} \quad (2)$$

Attempting to calculate the expected values of the terms, we remark that the \tilde{w}_n are independent random variables, that \tilde{w}_n is independent of \mathbf{x}_m for $n \neq m$, and that the \tilde{w}_n have mean 0. Using the formula $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ for independent random variables X, Y we get the following, where a hat over a variable indicates that it should be omitted:

$$\begin{aligned} \mathbb{E} \left[(\mathbf{x}_n - \boldsymbol{\mu})^2 \left(\sum_{n_1, \dots, n_k=1}^N \tilde{w}_{n_1} \dots \tilde{w}_{n_k} \right) \right] &= \mathbb{E} \left[(\mathbf{x}_n - \boldsymbol{\mu})^2 \left(\sum_{e_1 + \dots + e_N = k} \binom{k}{e_1, \dots, e_N} \tilde{w}_1^{e_1} \dots \tilde{w}_N^{e_N} \right) \right] \\ &= \mathbb{E} \left[(\mathbf{x}_n - \boldsymbol{\mu})^2 \left(\sum_{j=0}^k \sum_{\substack{e_1 + \dots + e_N = k \\ e_n = j}} \binom{k}{j} \binom{k-j}{e_1, \dots, \widehat{e_n}, \dots, e_N} \tilde{w}_1^{e_1} \dots \tilde{w}_N^{e_N} \right) \right] \\ &= \sum_{j=0}^k \binom{k}{j} \mathbb{E} [\tilde{w}_n^j (\mathbf{x}_n - \boldsymbol{\mu})^2] \\ &\quad \left(\sum_{\substack{e_1 + \dots + e_N = k \\ e_n = j}} \binom{k-j}{e_1, \dots, \widehat{e_n}, \dots, e_N} \mathbb{E}(\tilde{w}_1^{e_1}) \dots \mathbb{E}(\widehat{\tilde{w}_n^{e_n}}) \dots \mathbb{E}(\tilde{w}_N^{e_N}) \right) \\ &= \sum_{j=0}^k \binom{k}{j} \mathbb{E} [\tilde{w}_n^j (\mathbf{x}_n - \boldsymbol{\mu})^2] \\ &\quad \left(\sum_{\substack{e_1 + \dots + e_N = k \\ e_n = j}} \binom{k-j}{e_1, \dots, \widehat{e_n}, \dots, e_N} \mathbf{M}_{\tilde{w}}^{e_1} \dots \widehat{\mathbf{M}_{\tilde{w}}^{e_n}} \dots \mathbf{M}_{\tilde{w}}^{e_N} \right) \end{aligned}$$

Here $\mathbf{M}_{\tilde{w}}^a := \mathbb{E}[\tilde{w}^a]$ denotes the a -th moment of \tilde{w} . To simplify this expression we will make use of the characteristic function $\phi_{\tilde{w}}(t) := \mathbb{E}(e^{t\tilde{w}})$. Recall that

$$\mathbf{M}_{\tilde{w}}^a = (-i)^a \phi_{\tilde{w}}^{(a)}(0)$$

This yields

$$\begin{aligned}
\ldots &= \sum_{j=0}^k \binom{k}{j} \mathbb{E} [\tilde{w}_n^j (\mathbf{x}_n - \boldsymbol{\mu})^2] \left(\sum_{\substack{e_1 + \dots + e_N = k \\ e_n = j}} \binom{k-j}{e_1, \dots, \widehat{e_n}, \dots, e_N} (-i)^{e_1} \phi_{\tilde{w}}^{(e_1)}(0) \dots (-i)^{e_n} \widehat{\phi_{\tilde{w}}^{(e_n)}}(0) \dots (-i)^{e_N} \phi_{\tilde{w}}^{(e_N)}(0) \right) \\
&= \sum_{j=0}^k \binom{k}{j} \mathbb{E} [\tilde{w}_n^j (\mathbf{x}_n - \boldsymbol{\mu})^2] (-i)^{k-j} (\phi_{\tilde{w}}^{N-1})^{(k-j)}
\end{aligned}$$

Let $\mathbf{M}_{\tilde{w}_1 + \dots + \tilde{w}_{N-1}}^a$ denote the a -th moment of the sum of $N-1$ independent \tilde{w} -distributed random variables. The above can be written as:

$$\mathbb{E} \left[(\mathbf{x}_n - \boldsymbol{\mu})^2 \left(\sum_{n_1, \dots, n_k=1}^N \tilde{w}_{n_1} \dots \tilde{w}_{n_k} \right) \right] = \sum_{j=0}^k \binom{k}{j} \mathbb{E} [\tilde{w}^j (\mathbf{x} - \boldsymbol{\mu})^2] \mathbf{M}_{\tilde{w}_1 + \dots + \tilde{w}_{N-1}}^{k-j}$$

Doing the same with the other term in 2 gives

$$\mathbb{E} \left[\tilde{w}_n (\mathbf{x}_n - \boldsymbol{\mu})^2 \left(\sum_{n_1, \dots, n_k=1}^N \tilde{w}_{n_1} \dots \tilde{w}_{n_k} \right) \right] = \sum_{j=0}^k \binom{k}{j} \mathbb{E} [\tilde{w}^{j+1} (\mathbf{x} - \boldsymbol{\mu})^2] \mathbf{M}_{\tilde{w}_1 + \dots + \tilde{w}_{N-1}}^{k-j}$$

Adding the two we get:

$$\mathbb{E} \left[w_n (\mathbf{x}_n - \boldsymbol{\mu})^2 \left(\sum_{n_1, \dots, n_k=1}^N \tilde{w}_{n_1} \dots \tilde{w}_{n_k} \right) \right] = \sum_{j=0}^k \binom{k}{j} \mathbb{E}_{p(\mathbf{x})} [\tilde{w}^j (\mathbf{x} - \boldsymbol{\mu})^2] \mathbf{M}_{\tilde{w}_1 + \dots + \tilde{w}_{N-1}}^{k-j}$$

Now recall that $\sigma^2 < \infty$. Define the probability distribution $q(x) = \sigma^{-2}(x - \mu)^2 p(\mathbf{x})$. The above expression equals:

$$\mathbb{E} \left[w_n (\mathbf{x}_n - \boldsymbol{\mu})^2 \left(\sum_{n_1, \dots, n_k=1}^N \tilde{w}_{n_1} \dots \tilde{w}_{n_k} \right) \right] = \sigma^2 \sum_{j=0}^k \binom{k}{j} \mathbf{M}_{\tilde{w}, q(\mathbf{x})}^j \mathbf{M}_{\tilde{w}_1 + \dots + \tilde{w}_{N-1}}^{k-j}$$

This is the k -th moment of a random variable Y which is the sum of $N-1$ \tilde{w} -distributed random variables and the random variable $\tilde{w}(X)$, with $X \sim \sigma^{-2}(\mathbf{x} - \mu)^2 p(\mathbf{x})$. We thus get an asymptotic expansion of the expectation using the higher moments of Y . For $k \leq 2$ the moments are easy to calculate because we have no cross terms; the first two we have computed above:

$$\begin{aligned}
T_0 &= \mathbb{E}_{p(\mathbf{x})} ((\mathbf{x} - \mu)^2) \\
T_1 &= -\frac{1}{N} \mathbb{E}_{p(\mathbf{x})} (\tilde{w}(\mathbf{x} - \mu)^2) \\
T_2 &= \frac{N-1}{N^2} \mathbb{E}(\tilde{w}^2) + \mathbb{E}_{p(\mathbf{x})} (\tilde{w}^2 (\mathbf{x} - \mu)^2)
\end{aligned}$$

The second-order term hints at the fact that, if the \tilde{w} are not bounded, the series we derived will generally not converge for any finite N , with the higher-order terms blowing up. An example is our Gaussian toy model, where for $\eta < \sigma$ we will observe a blowup after finitely many terms.

My theory: unbounded importance weights can make all sorts of things go wrong.