Variance Estimation using Proposal Priors

November 10, 2017

We fix our prior $p(\mathbf{x})$ 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

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

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:

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

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

Thus we get

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

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

$$2\mathbb{E}_{p(\mathbf{x})}((\mathbf{x} - \boldsymbol{\mu})^2) - \frac{N-1}{N} \mathbb{E}_{p(\mathbf{x})}((\mathbf{x} - \boldsymbol{\mu})^2) - \frac{1}{N} \mathbb{E}_{p(\mathbf{x})}(w(\mathbf{x} - \boldsymbol{\mu})^2)$$

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

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

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

This is consistent with the fact that the bias should disappear as $N \to \infty$, 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

$$\varepsilon^k = \sum_{n_1, \ldots, n_k = 1}^{N} \tilde{w}_{n_1} \ldots \tilde{w}_{n_k}$$

$$\sum_{n=1}^{N} w_n (\mathbf{x}_n - \boldsymbol{\mu})^2 = \sum_{n=1}^{N} \tilde{w}_n (\mathbf{x}_n - \boldsymbol{\mu})^2 + \sum_{n=1}^{N} (\mathbf{x}_n - \boldsymbol{\mu})^2$$

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 $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) = 0$ for independent random variables X, Y with $\mathbb{E}(Y) = 0$ we get:

$$\mathbb{E}\left[\frac{1}{N}\left(\sum_{n=1}^{N} \tilde{w}_n(\mathbf{x}_n - \boldsymbol{\mu})^2\right)\left(\sum_{n_1,\ldots,n_k=1}^{N} \tilde{w}_{n_1}\ldots\tilde{w}_{n_k}\right)\right] = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left(\tilde{w}^{k+1}(\mathbf{x}-\boldsymbol{\mu})^2\right) \quad = \mathbb{E}\left(\tilde{w}^{k+1}(\mathbf{x}-\boldsymbol{\mu})^2\right)$$

$$\mathbb{E}\left[\frac{1}{N}\left(\sum_{n=1}^{N} (\mathbf{x}_n - \boldsymbol{\mu})^2\right)\left(\sum_{n_1,\ldots,n_k=1}^{N} \tilde{w}_{n_1}\ldots\tilde{w}_{n_k}\right)\right] = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left(\tilde{w}^{k}(\mathbf{x}-\boldsymbol{\mu})^2\right) \quad = \mathbb{E}\left(\tilde{w}^{k}(\mathbf{x}-\boldsymbol{\mu})^2\right)$$

$$\mathbb{E}\left[\frac{1}{N}\left(\sum_{n=1}^{N} w_n (\mathbf{x}_n - \boldsymbol{\mu})^2\right)\frac{\varepsilon^k}{N^k}\right] = \frac{1}{N^k}\mathbb{E}\left(\tilde{w}^{k} w(\mathbf{x}-\boldsymbol{\mu})^2\right) \quad = \mathbb{E}_{p(\mathbf{x})}\left(\tilde{w}^{k}(\mathbf{x}-\boldsymbol{\mu})^2\right)$$

Summing the terms in the Taylor expansion yields

$$\sum_{k=0}^{\infty}(-1)^k \frac{1}{N^k}\mathbb{E}_{p(\mathbf{x})}\left(\tilde{w}^{k}(\mathbf{x}-\boldsymbol{\mu})^2\right) = \mathbb{E}_{p(\mathbf{x})}\left[\sum_{k=0}^{\infty}(-1)^k \frac{\tilde{w}^{k}}{N^k}(\mathbf{x}-\boldsymbol{\mu})^2\right] \quad = \mathbb{E}_{p(\mathbf{x})}\left[\frac{(\mathbf{x}-\boldsymbol{\mu})^2}{1 + \frac{\tilde{w}}{N}}\right]$$

This equals the true variance if either $N = \infty$ (asymptotically) or the weights are $\equiv 1$. As a sanity check, it also gives the correct value for $N = 1$.

Analysing the expression leads to the same result as before, just in a more precise way.