

---

# Variance Sensitivity Proofs

---

March 16, 2020

**Definition 1.** *Let variance be defined as*

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

## 1 NEIGHBORING DEFINITION: CHANGE ONE

### 1.1 $\ell_1$ -sensitivity

**Lemma 1.** *For arbitrary  $a$ ,*

$$\sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(a - \bar{x})^2.$$

*Proof.*

$$\begin{aligned} \sum_{i=1}^n (x_i - a)^2 &= \sum_{i=1}^n ((x_i - \bar{x}) - (a - \bar{x}))^2 \\ &= \sum_{i=1}^n ((x_i - \bar{x})^2 - 2(x_i - \bar{x})(a - \bar{x}) + (a - \bar{x})^2) \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 - 2 \sum_{i=1}^n (x_i a - x_i \bar{x} - \bar{x} a + \bar{x}^2) + \sum_{i=1}^n (a^2 - 2a\bar{x} + \bar{x}^2) \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 - 2a \sum_{i=1}^n x_i + 2\bar{x} \sum_{i=1}^n x_i + 2\bar{x} a n - 2\bar{x}^2 n + a^2 n - 2a\bar{x} n + \bar{x}^2 n \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + a^2 n - 2a\bar{x} n + \bar{x}^2 n \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(a - \bar{x})^2 \end{aligned}$$

□

**Theorem 1.** *Let*

$$f(\mathbf{x}) = \sum_{i=1}^n (x_i - \bar{x})^2.$$

*Then for  $\mathbf{x}$  bounded between  $m$  and  $M$ ,  $f$  has sensitivity bounded above by*

$$\frac{n-1}{n}(M-m)^2.$$

*Proof.* Consider databases  $\mathbf{x}'$  and  $\mathbf{x}''$  which differ in a single point. For notational ease, call  $\mathbf{x}$  the part of  $\mathbf{x}'$  and  $\mathbf{x}''$  that is the same, and say that  $\mathbf{x}$  contains  $n$  points. WLOG say that the last data point in the database is the one that differs. I.e.,  $\mathbf{x}' = \mathbf{x} \cup \{x_{n+1}\}$ , and  $\mathbf{x}'' = \mathbf{x} \cup \{x'_{n+1}\}$ . This proof assumes that a “neighboring database” is one that differs in a single data-point, so we will ultimately be comparing  $f(\mathbf{x}')$  and  $f(\mathbf{x}'')$ . However, it is useful to first write  $f(\mathbf{x}')$  in terms of  $f(\mathbf{x})$ . Note that

$$\begin{aligned} \bar{x}' &= \frac{1}{n+1} \sum_{i=1}^{n+1} x_i \\ &= \frac{n\bar{x} + x_{n+1}}{n+1}. \end{aligned} \tag{1.1}$$

Then,

$$\begin{aligned} f(\mathbf{x}') &= \sum_{i=1}^n (x_i - \bar{x}')^2 + (x_{n+1} - \bar{x}')^2 \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x}' - \bar{x})^2 + (x_{n+1} - \bar{x}')^2 && \text{(By Lemma 1)} \\ &= f(\mathbf{x}) + n \left( \frac{n\bar{x} + x_{n+1}}{n+1} - \bar{x} \right)^2 + \left( x_{n+1} - \frac{n\bar{x} + x_{n+1}}{n+1} \right)^2 && \text{(By Equation 1.1)} \\ &= f(\mathbf{x}) + n \left( \frac{x_{n+1} - \bar{x}}{n+1} \right)^2 + \left( \frac{nx_{n+1} - n\bar{x}}{n+1} \right)^2 \\ &= f(\mathbf{x}) + (x_{n+1} - \bar{x})^2 \frac{n + n^2}{(n+1)^2} \\ &= f(\mathbf{x}) + (x_{n+1} - \bar{x})^2 \frac{n}{n+1} \end{aligned} \tag{1.2}$$

Now, to bound the sensitivity of  $f$ , note that

$$\begin{aligned} |f(\mathbf{x}') - f(\mathbf{x}'')| &= \left| (x_{n+1} - \bar{x})^2 \frac{n}{n+1} - (x'_{n+1} - \bar{x})^2 \frac{n}{n+1} \right| \\ &\leq (M-m)^2 \frac{n}{n+1}. \end{aligned} \tag{1.3}$$

The bound in the final line follows from the case where  $x_{n+1} = M$  (resp.  $m$ ) and  $\bar{x} = x'_{n+1} = m$  (resp.  $M$ ).

So we have a bound on the sensitivity of  $f$  for a data set of size  $n+1$ , but traditionally we consider sensitivities on a data set of size  $n$ . Redefining  $n+1$  as  $n$  in the above equation gives

$$(M-m)^2 \frac{n-1}{n}.$$

□

**Corollary 1.** *Sample variance has sensitivity bounded above by*

$$\frac{(M - m)^2}{n}.$$

## 1.2 $\ell_2$ -sensitivity

**Theorem 2.** *Let  $X$  be a data set with  $n$  elements,  $x_1, \dots, x_n$  and*

$$f(X) = \sum_{i=1}^n (x_i - \bar{x})^2$$

*be the sample variance. For  $X$  bounded between  $m$  and  $M$ ,  $f$  has a global sensitivity of*

$$\left( \frac{n-1}{n} (M - m)^2 \right)^2$$

*Proof.* We can pick up from statement 1.3, switching from  $\ell_1$  to  $\ell_2$  norm and interpreting the data sets in question to be of size  $n$  rather than  $n + 1$ .

$$\begin{aligned} (f(x') - f(x''))^2 &= \left( (x_n - \bar{x})^2 \frac{n-1}{n} - (x'_n - \bar{x})^2 \frac{n-1}{n} \right)^2 \\ &= \left( \frac{n-1}{n} \right)^2 ((x_n - \bar{x})^2 - (x'_n - \bar{x})^2)^2 \\ &\leq \left( \frac{n-1}{n} \right)^2 ((M - m)^2)^2 \\ &= \left( \frac{n-1}{n} \right)^2 (M - m)^4 \\ &= \left( \frac{n-1}{n} (M - m)^2 \right)^2. \end{aligned}$$

□

## 2 NEIGHBORING DEFINITION: ADD/DROP ONE

### 2.1 $\ell_1$ -sensitivity

**Theorem 3.** *Let  $X$  be a data set with  $n$  elements,  $x_1, \dots, x_n$  and*

$$f(X) = \sum_{i=1}^n (x_i - \bar{X})^2$$

*be the sample variance. For  $X$  bounded between  $m$  and  $M$ ,  $f$  has a global sensitivity of*

*Proof.* We must consider both adding and removing an element from  $X$ .

Adding an element:

Let  $X' = X \cup x'_{n+1}$ . Recall from Eq. 1.2 that for

$$f(x) = \sum_{i=1}^n (x_i - \bar{x})^2,$$

$$f(x') = (x_n + 1 - \bar{x})^2 \frac{n}{n+1}.$$

So,

$$\begin{aligned} |f(x') - f(x)| &= \left| f(x) + (x_{n+1} - \bar{x})^2 \frac{n}{n+1} - f(x) \right| \\ &= \left| (x_{n+1} - \bar{x})^2 \frac{n}{n+1} \right| \\ &\leq (M - m)^2 \frac{n}{n+1} \end{aligned}$$

So the variance between  $x$  and  $x'$  is bounded by

$$(M - m)^2 \frac{n}{(n+1)(n-1)}.$$

□

## 2.2 $\ell_2$ -sensitivity