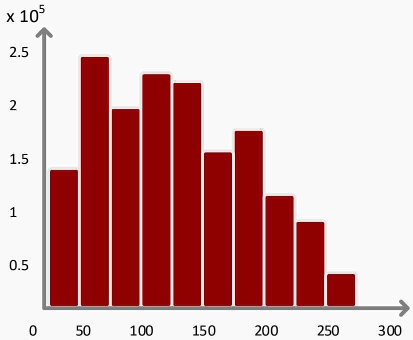
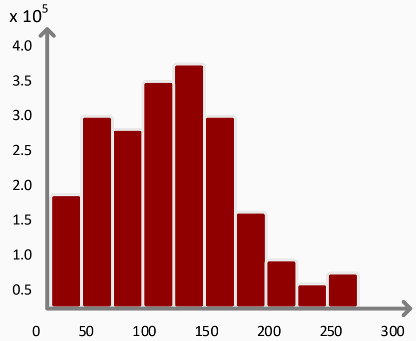
working with spatial data

Variograms – used to describe spatial correlations in a way that Histograms do not.

In the case of the two images below, histograms were computed on the grayscale values of each:

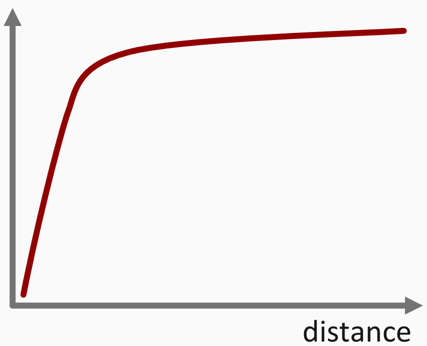


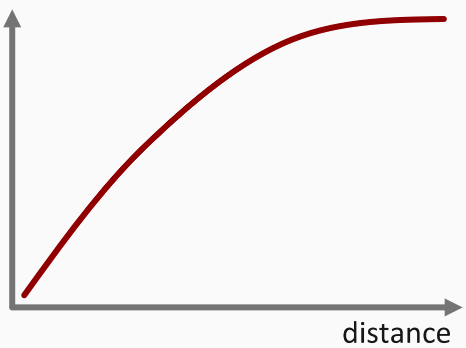


The Histograms ploted against the frequency of each images greyscale points are insignificantly different from each other; Additionally, the **Histograms cannot capture** any similiarities between nearby sprinkles as seen in the images themslevles.

Therefore, if the sprinkles in the pictures (or greyscale points) were randomly rearranged, there would be no actual effect on the histograms produced from the image data.

Variogram





In contrast, **Varigrams show *how*** nearby points are **correlated**. The left image illustrates how nearby points are equally correlated as far away points, representing the actual image. The right image curves smoothly demonstrating nearby points as more correlated than further points represented in the image.

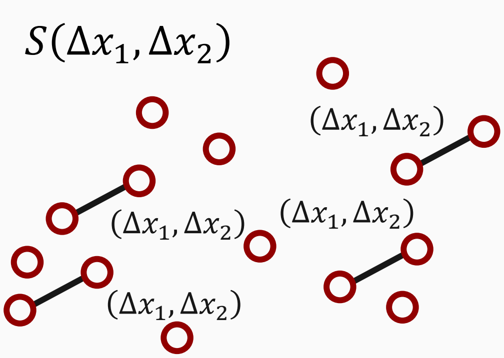
Formally Defining Variograms

(x_1, x_2) is a point in space

$y(x_1, x_2)$ is the label at point (x_1, x_2)

$y(x_1 + \Delta x_1, x_2 + \Delta x_2)$ is the label at point $(x_1 + \Delta x_1, x_2 + \Delta x_2)$

Δx_1 and Δx_2 are separations between points (lags)



Variogram:

$$\gamma(\Delta x_1, \Delta x_2) = \frac{1}{2} E[y(x_1 + \Delta x_1, x_2 + \Delta x_2) - y(x_1, x_2)]^2$$

The expression measures the question: On average, how much does the label $y(x_1, x_2)$ change at distance delta Δ away from the initialized point in space (x_1, x_2)?

The expression also includes an expectation E that needs to be estimated to compute:

For each change in distance Δx_1 and the change in distance Δx_2, find all pairs that are separated by those two distances Δx_1 and Δx_2 to denote $S(\Delta x_1, \Delta x_2)$.

$S(\Delta x_1, \Delta x_2)$ = the set (number count) of i, j pairs sharing the same distances:

$$x_1^i - x_1^j \approx \Delta x_1 \text{ and } x_2^i - x_2^j \approx \Delta x_2$$

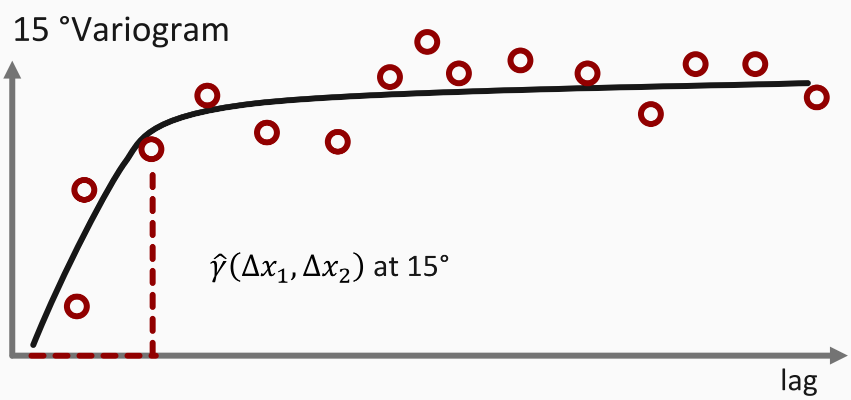
$$|S(\Delta x_1, \Delta x_2)| = \text{the number of pairs in S}$$

Therefore, the Varigram Expresion with Estimated Expectation is computed by taking the points in to approximate the Variogram:

$$\hat{\gamma}(\Delta x_1, \Delta x_2) = \frac{1}{2} \frac{1}{|S(\Delta x_1, \Delta x_2)|} \sum_{i,j \ in\ |S(\Delta x_1, \Delta x_2)|} \left(y^i - y^j\right)^2$$

Essentially, the expression above produces a variogram for each individually measured angle:

Assuming a $(\Delta x_1, \Delta x_2)$ of 15° (left illustration), all point pairs that are 15° apart will be included in the set $|S(\Delta x_1, \Delta x_2)|$. The values will then be squared and averaged over all the pairs in $|S(\Delta x_1, \Delta x_2)|$. providing a point on the 15° Variogram. The process used to compute the 15° lag distance is repeated at different lags and multiple directions to produce a comprehensive Variogram representive of the actual image.

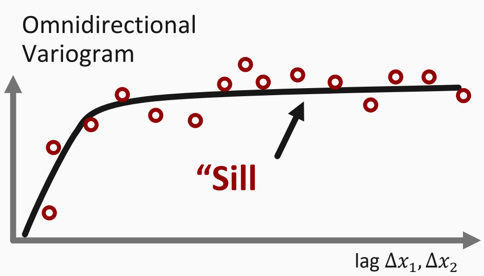


In some circumstances, the Variograms will be similar between multiple directions applied (15°, 45°, 60°, etc.). In this case, it is not necessary to show multiple Variograms and can instead use a singe **Omnidirectional Variogram** representative of all the measured lags.

Omnidirectional Variograms can be computed with ***any*** directions measured by simple **Euclidean Distance**.

An important note: A Variogram is everything needed to compute a Covariance Estimate between y and it's spatially lagged version; simply using the negative of the Variogram and flipping the Sill (the flat portion of the Variogram Curve):

$$\text{Covariance: } k((x_1 + \Delta x_1\ , x_2 + \Delta x_2) = \text{"}\mathbf{Sill}\text{"} - \hat{\gamma}(\Delta x_1, \Delta x_2)$$

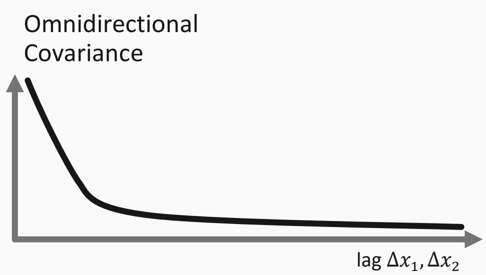


Additional items needed from the Variogram for Kriging or Gaussian Processes:

- A way to etimate pairwise covariances (the Variogram provides this)
- A full matrix of pairwise variances for training points in the dataset.
- Be able to estimate labels at a new point (variance of label at point x)

$$\begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ k(x_n, x_1) & \cdots & \cdots & k(x_n, x_n) \end{bmatrix} \qquad \begin{bmatrix} k(x, x^1) \\ k(x, x^2) \\ \vdots \\ k(x, x^n) \end{bmatrix} \qquad k(x, x)$$

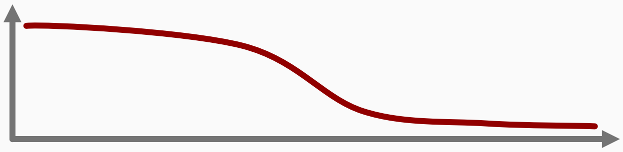
$$K^{\text{neighbors}} \qquad K^x$$



The machine learning community uses a simpler method:

Use the **Gaussian PDF** (Probability Density Function) formula to estimate covariance. The intuition is the covariance decreasing according to the Gaussian PDF with a particular bandwidth, in all directions:

$$k(x, x^1) = C\exp\left[\frac{-(x - x^1)^2}{2h^2}\right]$$



However, regardless of the method chosen, the matrix of numbers $K^{\text{neighbors}}$, vector of numbers K^x, and single number $k(x, x)$, are still needed in order to proceed to Kriging.

The final term for Variograms is the concept of a Semi-Variogram; removing the fraction $\frac{1}{2}$.

Variogram:

$$\hat{\gamma}(\Delta x_1, \Delta x_2) = \mathbf{\frac{1}{2}} \frac{1}{|S(\Delta x_1, \Delta x_2)|} \sum_{i,j \text{ } in \text{ } |S(\Delta x_1, \Delta x_2)|} (y^i - y^j)^2$$

SemiVariogram:

$$\hat{\gamma}(\Delta x_1, \Delta x_2) = \frac{1}{|S(\Delta x_1, \Delta x_2)|} \sum_{i,j \text{ } in \text{ } |S(\Delta x_1, \Delta x_2)|} (y^i - y^j)^2$$

A SemiVariogram is just twice (2x) the same Variogram.