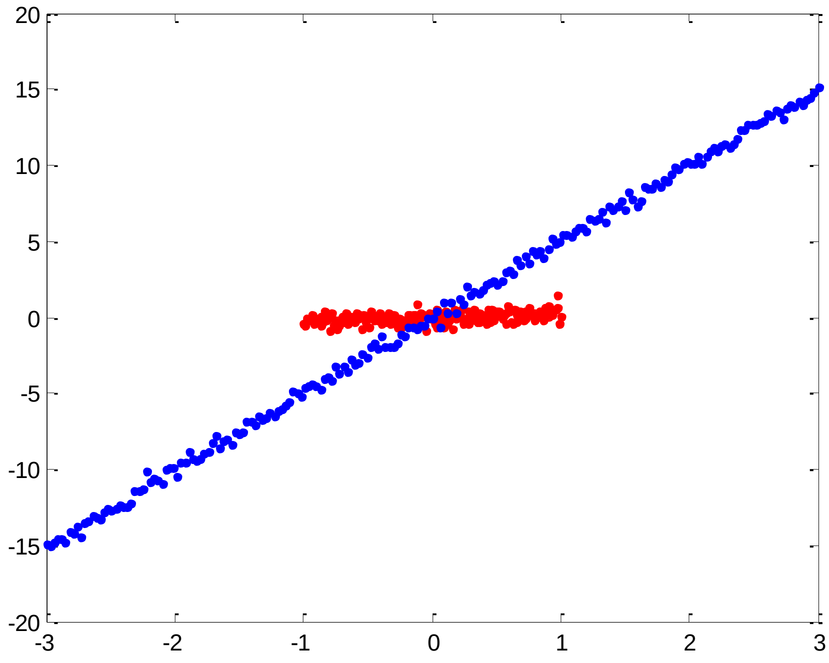
下一路径点的生成

思路一：以历史平均风速和平均方向角为基准

记历史的第i场台风的第j个路径点的位置为(x_{ij}, y_{ij})，设一共有N场台风，且第i场台风有n_i个路径点。

显然在每个历史路径点(x_{ij}, y_{ij})处，其实际的速度$v(x_{ij}, y_{ij})$和在该点计算出的历史平均速度$\bar{v}(x_{ij}, y_{ij})$一般有偏差$\Delta v_{ij} = v(x_{ij}, y_{ij}) - \bar{v}(x_{ij}, y_{ij})$。我们要考察一条路径（比如第$i$条）上，前后两个路径点偏差$\Delta v_{i,j-1}$和$\Delta v_{ij}$之间的相关性，即前者的信息量究竟多大程度地被后者继承。也就是说，我们可假设Δv_{ij}是由$\Delta v_{i,j-1}$和一些随机成分结合生成的。问题是怎么去权衡$\Delta v_{i,j-1}$和随机成分之间的比例。**这里为了去除尺度的影响，把所有路径点的自相关都平等地对待**，我们要把Δv_{ij}都标准化为$u_{ij} = \Delta v_{ij} / \delta_{ij}$，其中$\delta_{ij}$表示在$(x_{ij}, y_{ij})$处的历史标准差。如果不这么做，那么就会出现如图所示的问题：虽然都是同样比重相关性，但若不标准化，蓝色的影响会大于红色。



现在我们建立模型，首先暂把u_{ij}看成随机变量，假设标准化后的偏差有相关关系：$u_{ij}=au_{i,j-1}+b\varepsilon$，其中$\varepsilon$为标准差为 1，期望为 0 的随机数，$u_{i,j-1}$和$\varepsilon$是相互独立的。于是

$$1=\mathrm{var}(u_{ij})=E(u_{ij}^2)=a^2E(u_{i,j-1}^2)+2abE(u_{i,j-1})E(\varepsilon)+b^2E(\varepsilon^2)=a^2+b^2$$

a的绝对值和b的绝对值之间的关系是此消彼长的，且只要得到了a就确定了b的绝对值（注意b是正是负无关紧要）。又有u_{ij}和$u_{i,j-1}$之间的相关系数为：

$$Corr(u_{ij},u_{i,j-1})=\frac{E[(u_{ij}-E(u_{ij}))(u_{i,j-1}-E(u_{i,j-1}))]}{\mathrm{var}(u_{ij})\,\mathrm{var}(u_{i,j-1})}=E(u_{ij}u_{i,j-1})=E[(au_{i,j-1}+b\varepsilon)u_{i,j-1}]=a$$

我们假定这个系数a是完全由地理位置决定的（也可以把时间因素加进去）。但是在实际计算中，我们不可能有在(x_{ij},y_{ij})这个单点处有足够的样本去估计$Corr(u_{ij},u_{i,j-1})$，所以我们同样用核密度的思想去该点搜集附近的信息。

首先建立一个核密度函数f，它以(x,y)为中心，对每个历史路径点(x_{ij},y_{ij})都给出一个值$f(x_{ij},y_{ij};x,y)$，我们记它为w_{ij}。我们的意图是用w_{ij}作为权重来收集(x,y)周围偏差的相关性。计算方式如下：

首先定义加权平均

$$(\bar{u},\bar{u}')=\sum_{i=1}^{N}\sum_{j=1}^{n_i}w_{ij}(u_{ij},u_{ij-1})$$

再定义加权方差：

$$(\mathrm{var}(u),\mathrm{var}(u'))=\sum_{i=1}^{N}\sum_{j=1}^{n_i}w_{ij}(|u_{ij}-\bar{u}|^2,|u_{ij-1}-\bar{u}'|^2),$$

然后有加权相关系数：

$$a=\frac{\sum_{i=1}^{N}\sum_{j=1}^{n_i} w_{ij}(u_{ij}-\bar{u})(u_{ij-1}-\bar{u}')}{\sqrt{\operatorname{var}(u)\operatorname{var}(u')}}$$

对前进方向角度θ用一致的方式处理。

思路二：以历史平均风速变化和平均方向角变化为基准

现在我们控制台风前进的量变为了台风的相邻路径点之间速度变化w和角度的变化α，不过仍然需要每个初始点处的初速度v_0和初始角度θ_0。我们对w和α也分别按上述方式处理即可。

总结：上述对速度u做的处理方法，可以抽象出来，即可以是台风路径点处的任意特征量，比如强度。这个处理方法的一大关键就在那些核密度函数的形式和宽度的选择。

这里我们要区分两个跟“核密度”有关的概念。首先是在估计路径点的空间 p.d.f.时，我们采用的方法是真正意义上的统计学方法 kernel density estimation, 即用离散的样本分布去估计总体分布的方法。而算各种平均时，实际上是“核密度插值”这样一个概念。