公式

一、Granger因果关系

都是函数调用。

信息准则相关：函数调用

格兰杰：F、p、强度、是否存在，函数调用

```
import statsmodels.api as sm
result = sm.tsa.stattools.grangercausalitytests(data_to_analysis, maxlag=self.lag_order)
```

返回检验结果的的，F检验、卡方检验以及对应的p值，以及其它参数。F检验和卡方检验即可表示Granger因果关系，p值小于0.05即表示存在因果关系。强度即F检验或者卡方检验值，这里用卡方的，因为F检验你单独提出来了。(卡方检验和F检验的检验值存在一点差异，相对大小是一样的)

R2 是衡量回归模型拟合优度的。计算方式：

```
from sklearn.metrics import r2_score
```

> 衡量回归模型拟合优度的（与Granger无关）,即相关性系数(评估变量间的线性关系)。

BCP贝叶斯因果关系概率：

```
from pgmpy.models import BayesianNetwork
```

FEVD ：用于衡量各个变量对某个变量预测误差方差的贡献比例。程序假设预测未来10期。

```
 # 拟合VAR模型
            model = VAR(data)
            results = model.fit(maxlags=15)
            # 计算FEVD
            fevd = results.fevd(steps)
            # 获取FEVD结果
            fevd_results = fevd.decomp   # 获取分解结果
```

二、方向性条件概率

- 方向性条件概率：

1. **输入参数**：
    - `series_x` 和 `series_y` ：两个一维的数值序列，差分后的

2. **计算过程**：
    - `positive_x = series_x > 0`：创建一个布尔序列，表示 `series_x` 中大于 0 的元素。
    - `positive_y = series_y > 0`：创建一个布尔序列，表示 `series_y` 中大于 0 的元素。
3. **条件概率的计算**：
    - `p_y_given_x`：
        - 计算在 $X > 0$ 的情况下，$Y > 0$ 的概率。
        - 公式为：$P(Y > 0|X > 0) = \frac{\text{正的 } X \text{ 和正的 } Y \text{ 的数量}}{\text{正的 } X \text{ 的数量}}$。
    - `p_y_given_not_x`：
        - 计算在 $X \leq 0$ 的情况下，$Y > 0$ 的概率。
        - 公式为：$P(Y > 0|X \leq 0) = \frac{\text{负的 } X \text{ 和正的 } Y \text{ 的数量}}{\text{负的 } X \text{ 的数量}}$。
4. **返回值**：
    - 返回两个条件概率值：`p_y_given_x` 和 `p_y_given_not_x`。

🟢 `DR`（方向性比率）：

```
DR = p_y_given_x / p_y_given_not_x
```

🟢信息增益：

1. 熵 (Entropy)

对于给定的序列 (S) 的熵计算公式为：

$$H(S) = -\sum_{i} P(x_i)\log_2(P(x_i))$$

其中 (P(x_i)) 是类别 (x_i) 的概率。

2. 条件熵 (Conditional Entropy)

给定特征 (X) 的条件熵计算公式为：

$$H(Y|X) = \sum_{v \in V(X)} P(X = v)H(Y|X = v)$$

其中 (V(X)) 是特征 (X) 的所有唯一值。

3. 信息增益 (Information Gain)

信息增益的计算公式为：

$$IG(X) = H(Y) - H(Y|X)$$

其中：

- (H(Y)) 是目标变量的熵。
- (H(Y | X)) 是给定特征 (X) 的条件熵。

🟢方向性熵：

1. 联合分布

计算联合分布 (P(X, Y)):

- 使用二维直方图来估计 (X) 和 (Y) 的联合分布。

2. 边缘分布

边缘分布 (P(X)) 和 (P(Y)):

- $P(X) = \sum_y P(X, Y)$$
- $P(Y) = \sum_x P(X, Y)$

3. 条件熵

计算条件熵:

- **条件熵 (H(Y | X))**:

$$H(Y|X) = -\sum_x \sum_y P(X, Y) \log \left(\frac{P(X, Y)}{P(X)} \right)$$

- **条件熵 (H(X | Y))**:

$$H(X|Y) = -\sum_y \sum_x P(X, Y) \log \left(\frac{P(X, Y)}{P(Y)} \right)$$

4. 熵

计算单变量的熵:

- **熵 (H(X))**:

$$H(X) = -\sum_x P(X) \log(P(X))$$

- **熵 (H(Y))**:

$$H(Y) = -\sum_y P(Y) \log(P(Y))$$

5. 方向性熵

方向性熵的计算:

- **方向性熵 (DE_1)** (从 (Y) 到 (X)) :

$$DE_1 = H(Y) - H(Y|X)$$

- **方向性熵 (DE_2)** (从 (X) 到 (Y)) :

$$DE_2 = H(X) - H(X|Y)$$

三、交互延迟互信息

🟢条件熵：前面已经写过

🟢交叉熵:

$$textcross\_entropy = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\log(p_{ij})$$

● 互信息、方向性转移熵的计算：对不同延迟阶数的数据

1. 高斯核概率密度估计 (Kernel Density Estimation, KDE)

函数 `kde_estimation` 计算给定数据的高斯核概率密度估计。

2. 带宽估计

- **Silverman's Rule of Thumb**:

$$h = \max\left(1.06 \cdot \text{std\_dev} \cdot n^{-1/5}, 1e-3\right)$$

- **Empirical Rule of Thumb**:

$$h = \max\left(0.9 \cdot \min\left(\text{std\_dev}, \frac{\text{IQR}}{1.34}\right) \cdot n^{-1/5}, 1e-3\right)$$

3. 互信息 (Mutual Information)

$$I(X;Y) = \sum p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right)$$

4. 方向性转移熵 (Transfer Entropy)

$$TE(X \to Y) = \sum p(y_t, x_{t-\tau})\log\left(\frac{p(y_t|x_{t-\tau})}{p(y_t)}\right)$$

四、时间滞后相干性分析

● 一致性相关性系数:

$$\text{CCC} = \frac{2 \cdot \rho \cdot \sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}}{\text{Var}(X) + \text{Var}(Y) + (\text{Mean}(X) - \text{Mean}(Y))^2}$$

ρ 是皮尔逊相关系数。Var是方差、Mean是均值。

● 信息熵差异：熵的差值。单变量熵前面已有公式。

● 直接互信息：不带延迟的互信息，公式前面已有。

● 交叉互信息:

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} P(x,y)\log\left(\frac{P(x,y)}{P(y)}\right)$$

● 动量相关系数:

动量相关系数可以通过以下公式表示:

1. **动量计算**:
   对于时间序列 X 和 Y，动量在窗口大小为 N 的情况下计算为:

$$\text{Momentum}(X) = X_t - X_{t-N}$$

$$\text{Momentum}(Y) = Y_t - Y_{t-N}$$

2. **相关系数计算**：
   相关系数 r 可以通过以下公式计算：

$$r=\frac{\mathrm{Cov}(\mathrm{Momentum}(X),\mathrm{Momentum}(Y))}{\sigma_{\mathrm{Momentum}(X)}\cdot\sigma_{\mathrm{Momentum}(Y)}}$$

   其中：

   - $\mathrm{Cov}(\cdot)$ 是协方差；
   - $\sigma_{\mathrm{Momentum}(X)}$ 和 $\sigma_{\mathrm{Momentum}(Y)}$ 是动量序列的标准差。

五、递归量化分析

●方向性转移熵：已列过

●序列相似度：

1. **方向性序列计算**：

$$\mathrm{direction}_1=\mathrm{sign}(X_t-X_{t-1})$$

$$\mathrm{direction}_2=\mathrm{sign}(Y_t-Y_{t-1})$$

2. **皮尔逊相关系数计算**：

$$r=\frac{\mathrm{Cov}(\mathrm{direction}_1,\mathrm{direction}_2)}{\sigma_{\mathrm{direction}_1}\cdot\sigma_{\mathrm{direction}_2}}$$

其中，X 和 Y 是两个时间序列，r 是方向性序列的相似度（皮尔逊相关系数）。

●交叉关联熵：

$$H(P,Q)=-\sum_i P(i)\log(Q(i))$$