**第二章 随机信号分析**

**2.2.2 随机过程的数字特征**

http://jpkc.hdu.edu.cn/comm/txyl/webstandard/txbg/pinksquare.gif数学期望和方差

http://jpkc.hdu.edu.cn/comm/txyl/webstandard/txbg/pinksquare.gif自相关函数和自协方差函数

http://jpkc.hdu.edu.cn/comm/txyl/webstandard/txbg/pinksquare.gif两随机过程的联合分布函数和数字特征

http://jpkc.hdu.edu.cn/comm/txyl/webstandard/txbg/pinksquare.gif数学期望

随机过程http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ21/daxt.gif的数学期望定义为:

http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ21/equ2108.gif

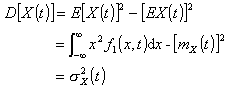
http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ24/mxt.gif是时间函数，它表示随机过程所有样本函数的统计平均函数。图2-2-2示出了随机过程http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ21/daxt.gif的http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ21/xiaon.gif个样本函数http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ24/xiaox1t.gif、http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ24/xiaox2t.gif、http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich6/txequ6/txequ64/3dot.gifhttp://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ24/xiaoxnt.gif和它的数学期望http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ24/mxt.gif。

http://jpkc.hdu.edu.cn/comm/txyl/webstandard/txbg/pinksquare.gif方差：

随机过程http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ21/daxt.gif的方差定义为:

http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ21/equ2109.gif

上式还可以表示为



http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ21/equ2111.gif称为随机过程http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ21/daxt.gif的方差或均方差。它表示随机过程在http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ21/xiaot.gif时刻对于均值http://jpkc.hdu.edu.cn/comm/txyl/webstandard/tximages/txich2/txequ2/txequ24/mxt.gif的偏离程度。

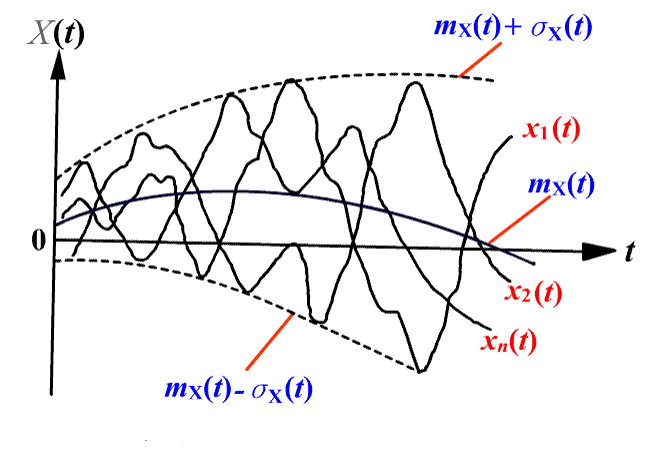
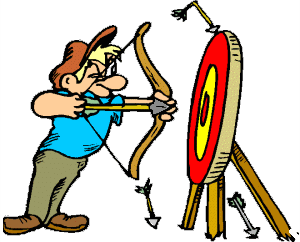
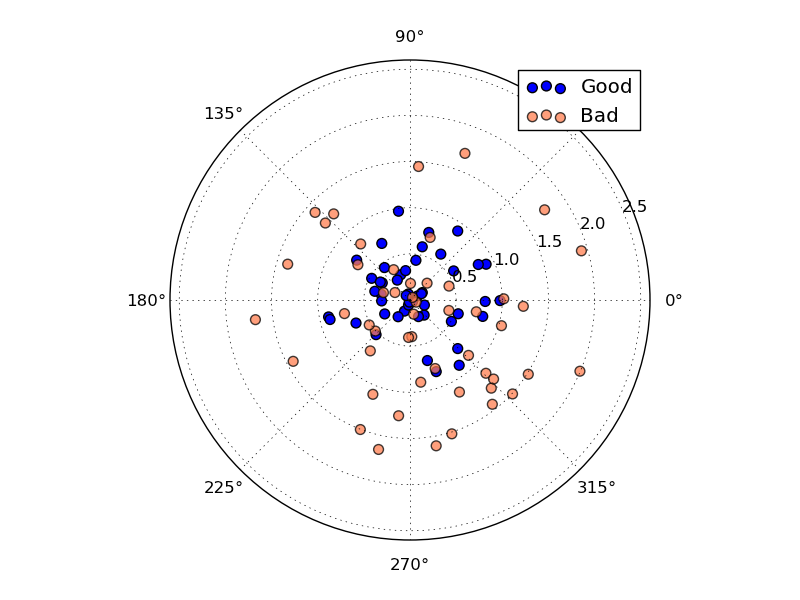


图2-2-2 随机过程的数学期望和均方差

除了[期望](http://www.cnblogs.com/vamei/p/3230753.html)，方差(variance)是另一个常见的分布描述量。如果说期望表示的是分布的中心位置，那么方差就是分布的离散程度。方差越大，说明随机变量取值越离散。



比如射箭时，一个优秀的选手能保持自己的弓箭集中于目标点附近，而一个经验不足的选手，他弓箭的落点会更容易散落许多地方。



上面的靶上有两套落点。尽管两套落点的平均中心位置都在原点 (即期望相同），但两套落点的离散程度明显有区别。蓝色的点离散程度更小。

数学上，我们用方差来代表一组数据或者某个概率分布的离散程度。可见，方差是独立于期望的另一个对分布的度量。两个分布，完全可能有相同的期望，而方差不同，正如我们上面的箭靶。

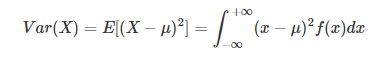
### 方差

对于一个随机变量XX来说，它的方差为:

Var(X)=E[(X−μ)2]

其中，μμ表示XX的期望值，即μ=E(X)。

我们可以代入[期望](http://www.cnblogs.com/vamei/p/3230753.html)的数学表达形式。比如连续随机变量：



方差概念背后的逻辑很简单。一个取值与期望值的“距离”用两者差的平方表示。该平方值表示取值与分布中心的偏差程度。平方的最小取值为0。当取值与期望值相同时，此时不离散，平方为0，即“距离”最小；当随机变量偏离期望值时，平方增大。由于取值是随机的，不同取值的概率不同，我们根据概率对该平方进行加权平均，也就获得整体的离散程度——方差。

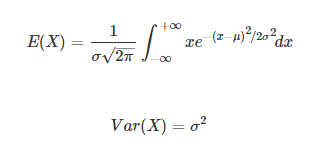
方差的平方根称为标准差(standard deviation, 简写std)。我们常用σ表示标准差

C:\Users\googol\AppData\Local\Temp\mx33E25.png

标准差也表示分布的离散程度。

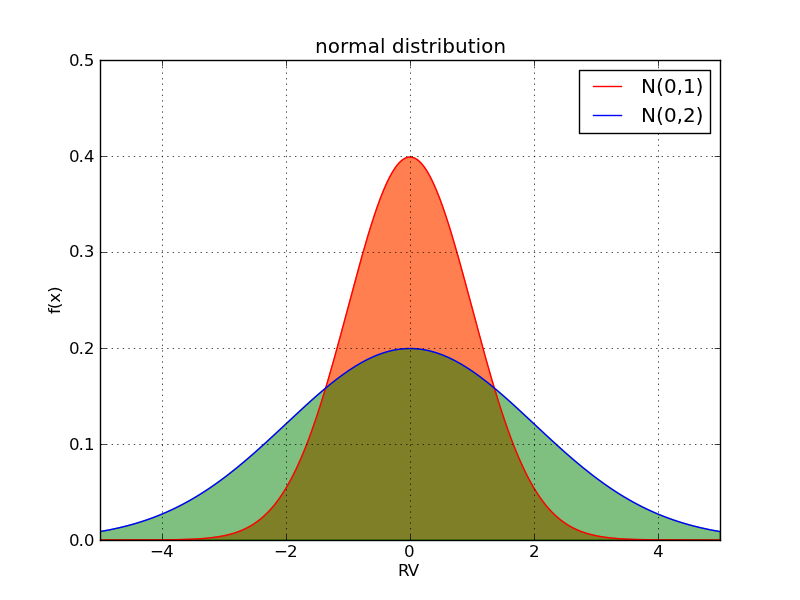
**正态分布的方差**

根据上面的定义，可以算出正态分布



正态分布的标准差正等于正态分布中的参数σ。这正是我们使用字母σ来表示标准差的原因！

可以预期到，正态分布的σ越大，分布离散越大，正如我们从下面的分布曲线中看到的:

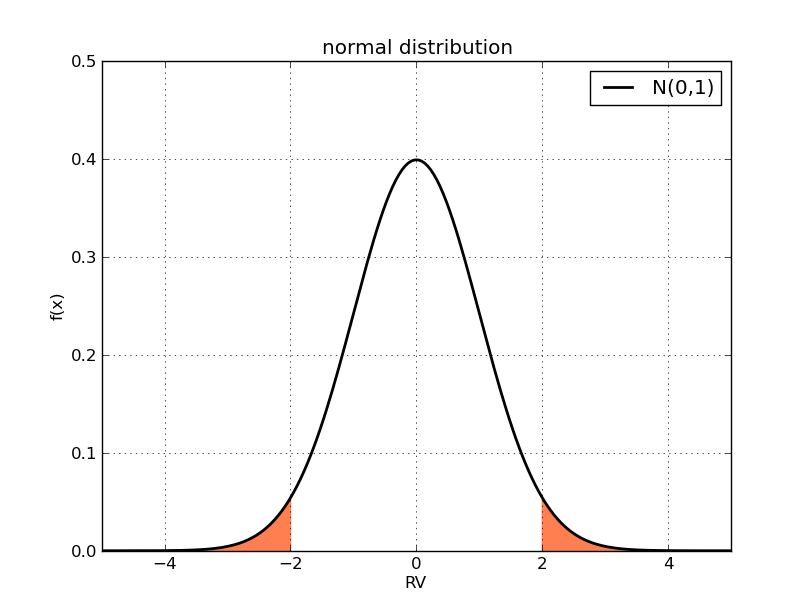
当方差小时，曲线下的面积更加集中于期望值0附近。当方差大时，随机变量更加离散。此时分布曲线的“尾部”很厚，即使在取值很偏离0时，比如x=4时，依然有很大的概率可以取到。

### Chebyshev不等式

我们一直在强调，标准差(和方差)表示分布的离散程度。标准差越大，随机变量取值偏离平均值的可能性越大。如何定量的说明这一点呢？我们可以计算一个随机变量与期望偏离超过某个量的可能性。比如偏离超过2个标准差的可能性。即

P(|X−μ|>2σ)

这个概率依赖于分布本身的类型。比如正态分布N(0,1)，这一概率即为x大于2，或者x小于-2的部分对应的曲线下面积：



实际上，无论μ和σ如何取值，对于正态分布来说，偏离期望超过两个标准差的概率都相同，约等于0.0455 (可以根据正态分布的表达式计算)。随机变量的取值有约95.545%的可能性落在正负两个标准差的区间内，即从-2到2。如果我们放大区间，比如正负三个标准差，这一概率超过99%。我们可以相当有把握的说，随机变量会落正负三个标准差之内。上面的论述并不依赖于标准差的具体值。这里可以看到标准差所衡量的“离散”的真正含义：如果取相同概率的极端值区间，比如上面的0.0455，标准差越大，该极端值区间距离中心值越远。

然而，上面的计算和表述依赖于分布的类型（正态分布）。如何将相似的方差含义套用在其它随机变量身上呢？

Chebyshev不等式让我们摆脱了对分布类型的依赖。它的叙述如下：

对于任意随机变量X，如果它的期望为μ，方差为σ2，那么对于任意t>0，

P(|X−μ|>t)≤σ2/t

无论X是什么分布，上述不等式成立。我们让t=2σ，那么

P(|X−μ|>2σ)≤0.25

也就是说，X的取值超过两个正负标准差的可能性最多为25%。换句话说，随机变量至少有75%的概率落在正负两个标准差的范围内。（显然这是最“坏”的情况下。正态分布显然不是”最坏“的）

前面介绍的分布描述量，比如期望和方差，都是基于单一随机变量的。现在考虑多个随机变量的情况。我们使用联合分布来表示定义在同一个样本空间的多个随机变量的概率分布。

联合分布中包含了相当丰富的信息。比如从联合分布中抽取某个随机变量的边缘分布，即获得该随机变量的分布，并可以据此，获得该随机变量的期望和方差。这样做是将视线限制在单一的一个随机变量上，我们损失了联合分布中包含的其他有用信息，比如不同随机变量之间的互动关系。为了了解不同随机变量之间的关系，需要求助其它的一些描述量。

### 协方差

协方差(covariance)表达了两个随机变量的协同变化关系。我们取一个样本空间，即学生的体检数据。学生的身高为随机变量X，学生的体重为随机变量Y。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 160cm | 170cm | 180cm |
| 60kg | 0.2 | 0.05 | 0.05 |
| 70kg | 0.05 | 0.3 | 0.05 |
| 80kg | 0.05 | 0.05 | 0.2 |

根据上表，大的身高(180cm)和大的体重(80kg)同时出现的概率较大(0.2)，小的身高值(160cm)和小的体重(60kg)的概率也较大(0.2)。偏大的身高往往伴随偏大的体重，偏小的身高常伴随偏小的体重。这种“大”伴随着“大”，“小”伴随着“小”的情形，叫做正相关。根据上面的数据，身高和体重两个随机变量正相关性很强。

另一方面，如果“大”配“小”，“小”配“大”的概率很高，那么两个随机变量负相关。“最萌身高差”是负相关的一个范例。（样本空间为情侣的身高信息。可以定义男生身高为一个随机变量，女生身高为另一个随机变量）



正如其他的分布描述量一样，协方差从概率分布中提取信息，让我们获知分布的“性能”。对于一个已知的联合分布来说，任意两个随机变量之间都可以计算出一个协方差，即一个数值。

### 定义

协方差的定义如下，如果X和Y是联合分布的随机变量，且分别有期望μX，μY，那么X和Y的协方差为

Cov(X,Y)=E[(X−μX)(Y−μY)]

协方差的定义基于期望。根据期望的定义，协方差可以直接用于离散随机变量和连续随机变量。

我们已经知道，期望是某个随机变量根据概率的加权平均。我们所要加权平均的目标是X−μX和Y−μY的乘积。随机变量和期望的差，代表了随机变量的取值和中心值的偏离程度，也就是我们上面所谓的“偏大”或者“偏小”的情况：正值的偏离表示“偏大”，负值的偏离表示“偏小”。如果是正相关，即大配大，小配小的情况，那么这一乘积为正；如果是负相关，乘积为负。所以，通过(X−μX)(Y−μY)这个量，我们表达了X和Y的相关性。

回到刚才的数据来计算相关性，

|  |  |  |  |
| --- | --- | --- | --- |
|  | 160cm | 170cm | 180cm |
| 60kg | 0.2 | 0.05 | 0.05 |
| 70kg | 0.05 | 0.3 | 0.05 |
| 80kg | 0.05 | 0.05 | 0.2 |

让身高为X，体重为Y。我们可以通过边缘分布，来分别获得X和Y的分布(回忆一下)。求得X和Y的期望，分别为170和70。计算各个格子中的(X−μX)(Y−μY)

|  |  |  |  |
| --- | --- | --- | --- |
|  | 160cm | 170cm | 180cm |
| 60kg | 100 | 0 | -100 |
| 70kg | 0 | 0 | 0 |
| 80kg | -100 | 0 | 100 |

上面的两个表，对应的格子相乘，并求和，就得到协方差:

Cov(X,Y)=0.2×100+0.2×100+0.05×(−100)+0.05×(−100)=30

在上面的计算中，正相关的项目都分配有比较大的概率值。最终的协方差也是一个正值。

根据期望的性质，我们可以改写协方差的表达形式:

Cov(X,Y)=E(XY−XμX−YμX+μXμY)=E(XY)−E(X)μX−E(Y)μY+μXμY=E(XY)−E(X)E(Y)

推导过程

假设c是常数，x,y是随机变量，那么数学期望E有以下的一些性质

E(c)=c

E(x+c)=E(x)+c

E(cx)=cE(x)

E(x+y)=E(x)+E(y)

所以

E{[X-E(X)]}{[Y-E(Y)]}=E{xy-xE(y)-yE(x)+E(x)E(y)}=E(xy)-E(x)E(y)-E(x)E(y)+E(x)E(y)=E(xy)-E(x)E(y)

当X和Y独立时，有E(XY)=E(X)E(Y) Cov(X,Y)=0。

(注意，Cov(X,Y)=0并不意味着X和Y独立)

### 相关系数

正的协方差表达了正相关性，负的协方差表达了负相关性。对于同样的两个随机变量来说，计算出的协方差越大，相关性越强。

但随后一个问题，身高和体重的协方差为30，这究竟是多大的一个量呢？如果我们又发现，身高与鞋号的协方差为5，是否说明，相对于鞋号，身高与体重的的相关性更强呢？

这样横向对比超出了协方差的能力范围。从日常生活经验来说，体重的上下浮动大约为20kg，而鞋号的上下浮动大约可能只是5个号码。所以，对于体重来说，5kg与中心的偏离并不算大，而5个号码的鞋号差距，就可能是最极端的情况了。假设身高和体重的相关强度，与身高和鞋码的相关强度类似，但由于体重本身的数值上下浮动更大，所计算出的协方差也会更大。另一个情况，依然是计算身高与体重的协方差。数据完全不变，而只更改单位。我们的体重用克而不是千克做单位，计算出的协防差是原来数值的1000倍！

为了能进行这样的横向对比，我们需要排除用统一的方式来定量某个随机变量的上下浮动。这时，我们计算相关系数(correlation coefficient)。相关系数是“归一化”的协方差。它的定义如下:

ρ=Cov(X,Y)/sqrt(Var(X)Var(Y))

相关系数是用协方差除以两个随机变量的标准差。相关系数的大小在-1和1之间变化。再也不会出现因为计量单位变化，而数值暴涨的情况了。

依然使用上面的身高和体重数据，可以计算出

Var(X)=0.3×(60−70)2+0.3×(80−70)2=60

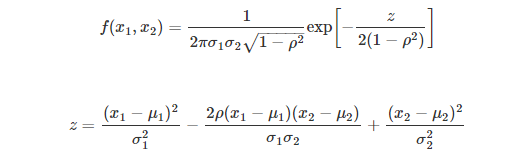
Var(Y)=0.3×(180−170)2+0.3×(160−170)2=60

ρ=30/60=0.5

这样一个“归一化”了的相关系数，更容易让人把握到相关性的强弱，也更容易在不同随机变量之间，做相关性的横向比较。

### 双变量正态分布

双变量正态分布是一种常见的联合分布。它描述了两个随机变量X1X1和X2X2的概率分布。概率密度的表达式如下：

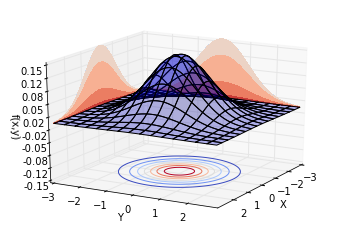
X1和X2的边缘密度分别为两个正态分布，即正态分布N(μ1,σ1), N(μ2,σ2)。

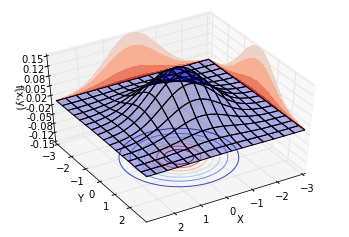
另一方面，除非ρ=0，否则联合分布也并不是两个正态分布的简单相乘。可以证明，ρ正是双变量正态分布中，两个变量的相关系数。

我们现在绘制该分布的图像。

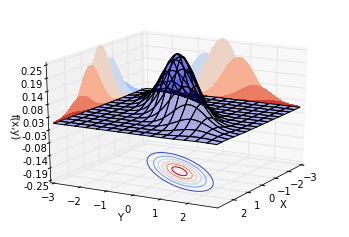
选取所要绘制的正态分布，为了简单起见，让μ1=0μ2=0, σ1=1σ2=1。

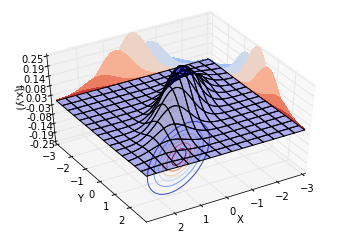
我们先让ρ=0，此时的联合分布相当于两个正态分布的乘积。绘制不同视角的同一分布，结果如下。可以看到，概率分布是中心对称的。





再让ρ=0.8，也就是说，两个随机变量的相关系数为0.8。绘制不同视角的同一分布，结果如下。可以看到，概率分布并不中心对称。沿着Y=X这条线，概率曲面隆起，概率明显比较高。而沿着Y=−X这条线，概率较低。这也就是我们所说的正相关。





现在，ρ对于我们来说，有了更具体的现实意义。:-)

总结如下

协方差

“归一化”的度量: 相关系数

以下来自维基百科

协方差[[编辑](http://zh.wikipedia.org/w/index.php?title=%E5%8D%8F%E6%96%B9%E5%B7%AE&action=edit&section=0)]

**协方差**（Covariance）在[概率论](http://zh.wikipedia.org/wiki/%E6%A6%82%E7%8E%87%E8%AB%96" \o "概率论)和[统计学](http://zh.wikipedia.org/wiki/%E7%B5%B1%E8%A8%88%E5%AD%B8)中用于衡量两个变量的总体[误差](http://zh.wikipedia.org/wiki/%E8%AF%AF%E5%B7%AE)。而[方差](http://zh.wikipedia.org/wiki/%E6%96%B9%E5%B7%AE)是协方差的一种特殊情况，即当两个变量是相同的情况。

[期望值](http://zh.wikipedia.org/wiki/%E6%9C%9F%E6%9C%9B%E5%80%BC)分别为E(X)=.mu与E(Y)=.nu的两个[实数](http://zh.wikipedia.org/wiki/%E5%AE%9E%E6%95%B0)[随机变量](http://zh.wikipedia.org/wiki/%E9%9A%8F%E6%9C%BA%E5%8F%98%E9%87%8F)*X* 与*Y* 之间的**协方差**定义为：

.operatorname{cov}(X, Y) = .operatorname{E}((X - .mu) (Y - .nu))，

其中E是期望值。它也可以表示为：

.operatorname{cov}(X, Y) = .operatorname{E}(X .cdot Y) - .mu .nu，

直观上来看，协方差表示的是两个变量的总体的[误差](http://zh.wikipedia.org/wiki/%E8%AF%AF%E5%B7%AE" \o "误差)，这与只表示一个变量误差的[*方差*](http://zh.wikipedia.org/wiki/%E6%96%B9%E5%B7%AE)不同。 如果两个[变量](http://zh.wikipedia.org/wiki/%E5%8F%98%E9%87%8F" \o "变量)的变化趋势一致，也就是说如果其中一个大于自身的期望值，另外一个也大于自身的期望值，那么两个变量之间的协方差就是正值。 如果两个变量的变化趋势相反，即其中一个大于自身的期望值，另外一个却小于自身的期望值，那么两个变量之间的协方差就是负值。

如果*X* 与*Y* 是[统计独立](http://zh.wikipedia.org/wiki/%E7%B5%B1%E8%A8%88%E7%8D%A8%E7%AB%8B%E6%80%A7" \o "统计独立性)的，那么二者之间的协方差就是0，这是因为

E(X .cdot Y)=E(X) .cdot E(Y)=.mu.nu,

但是反过来并不成立，即如果*X* 与*Y* 的协方差为0，二者并不一定是统计独立的。

取决于协方差的[相关性](http://zh.wikipedia.org/wiki/%E7%9B%B8%E5%85%B3)η

 .eta = .left| .dfrac{.operatorname{cov}(X, Y)}{.sqrt{.operatorname{var}(X) .cdot .operatorname{var}(Y)}} .right| ,

更准确地说是[线性相关性](http://zh.wikipedia.org/wiki/%E7%B7%9A%E6%80%A7%E7%9B%B8%E9%97%9C%E6%80%A7)，是一个衡量线性独立的[无量纲](http://zh.wikipedia.org/wiki/%E6%97%A0%E9%87%8F%E7%BA%B2)数，其取值在[0,+1]之间。相关性η = 1时称为“完全线性相关”，此时将Yi对Xi作Y-X [散点图](http://zh.wikipedia.org/w/index.php?title=%E6%95%A3%E7%82%B9%E5%9B%BE&action=edit&redlink=1)，将得到一组精确排列在直线上的点；相关性数值介于0到1之间时，其越接近1表明线性相关性越好，作散点图得到的点的排布越接近一条直线。

相关性为0（因而协方差也为0）的两个随机变量又被称为是[不相关](http://zh.wikipedia.org/wiki/%E7%9B%B8%E5%85%B3" \o "相关)的，或者更准确地说叫作“线性无关”、“线性不相关”，这仅仅表明*X* 与*Y* 两随机变量之间没有线性相关性，并非表示它们之间一定没有任何内在的（非线性）函数关系，和前面所说的“X、Y二者并不一定是统计独立的”说法一致。

属性[[编辑](http://zh.wikipedia.org/w/index.php?title=%E5%8D%8F%E6%96%B9%E5%B7%AE&action=edit&section=1)]

如果*X* 与*Y* 是实数随机变量，*a* 与*b* 不是随机变量，那么根据协方差的定义可以得到：

.operatorname{cov}(X, X) = .operatorname{var}(X)， .operatorname{cov}(X, Y) = .operatorname{cov}(Y, X)， .operatorname{cov}(aX, bY) = ab., .operatorname{cov}(X, Y)，

对于随机变量序列*X*1, ..., *Xn*与*Y*1, ..., *Ym*，有

.operatorname{cov}.left(.sum_{i=1}^n {X_i}, .sum_{j=1}^m{Y_j}.right) =  .sum_{i=1}^n{.sum_{j=1}^m{.operatorname{cov}.left(X_i, Y_j.right)}}，