期望、方差、协方差及相关系数的基本运算

作者 张洋 | 发布于 2013-06-05

概率 统计 数学

这篇文章总结了概率统计中期望、方差、协方差和相关系数的定义、性质和基本运算规则。

期望

定义

设$P(x)$是一个离散概率分布函数，自变量的取值范围为$\{x_1,x_2,\cdots,x_n\}$。其期望被定义为：

$$E(x)=\sum_{k=1}^{n}x_kP(x_k)$$

设$p(x)$是一个连续概率密度函数。其期望为：

$$E(x)=\int_{-\infty}^{+\infty}xp(x)dx$$

性质

1、线性运算规则

期望服从线性性质（可以很容易从期望的定义公式中导出）。因此线性运算的期望等于期望的线性运算：

$$E(ax+by+c)=aE(x)+bE(y)+c$$

这个性质可以推广到任意一般情况：

$$E(\sum_{k=1}^{n}a_ix_i+c)=\sum_{k=1}^{n}a_iE(x_i)+c$$

2、函数的期望

设$f(x)$为x的函数，则$f(x)$的期望为：

离散：

$$E(f(x))=\sum_{k=1}^{n}f(x_k)P(x_k)$$

连续：

$$E(f(x))=\int_{-\infty}^{+\infty}f(x)p(x)dx$$

一定要注意，**函数的期望不等于期望的函数**，即$E(f(x))\neq f(E(x))$！。

3、乘积的期望

一般来说，**乘积的期望不等于期望的乘积**，除非变量相互独立。因此，如果x和y相互独立，则$E(xy)=E(x)E(y)$。

期望的运算构成了统计量的运算基础，因为**方差、协方差等统计量本质上是一种特殊的期望。**

方差

定义

方差是一种特殊的期望，被定义为：

$$Var(x)=E((x-E(x))^2)$$

性质

1、展开表示

反复利用期望的线性性质，可以算出方差的另一种表示形式：

$$\begin{aligned}Var(x) &= E((x-E(x))^2)\\ &= E(x^2-2xE(x)+(E(x))^2)\\ &= E(x^2)-2E(x)E(x)+(E(x))^2\\ &= E(x^2)-2(E(x))^2+(E(x))^2\\ &= E(x^2)-(E(x))^2\end{aligned}$$

2、常数的方差

常数的方差为0，由方差的展开表示很容易推得。

3、线性组合的方差

方差不满足线性性质，两个变量的线性组合方差计算方法如下：

$$Var(ax+by)=a^2Var(x)+b^2Var(y)+2Cov(x,y)$$

其中$Cov(x,y)$为x和y的协方差，下一节讨论。

4、独立变量的方差

如果两个变量相互独立，则：

$$Var(ax+by)=a^2Var(x)+b^2Var(y)$$

作为推论，如果x和y相互独立：$Var(x+y)=Var(x)+Var(y)$。

协方差

定义

两个随机变量的协方差被定义为：

$$Cov(x,y)=E((x-E(x))(y-E(y)))$$

因此**方差是一种特殊的协方差**。当x=y时，$Cov(x,y)=Var(x)=Var(y)$。

性质

1、独立变量的协方差

独立变量的协方差为0，可以由协方差公式推导出。

2、线性组合的协方差

协方差最重要的性质如下：

$$Cov(\sum_{i=1}^{m}a_ix_i,\sum_{j=1}^{n}b_jy_j)=\sum_{i=1}^{m}\sum_{j=1}^{n}a_ib_jCov(x_i,y_j)$$

很多协方差的计算都是反复利用这个性质，而且可以导出一些列重要结论。

作为一种特殊情况：

$$Cov(a+bx,c+dy)=bdCov(x,y)$$

另外当x=y时，可以导出方差的一般线性组合求解公式：

$$Var(\sum_{k=1}^{n}a_ix_i)=\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_jCov(x_i,x_j)$$

相关系数

定义

相关系数通过方差和协方差定义。两个随机变量的相关系数被定义为：

$$Corr(x,y) = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}$$

性质

1、有界性

相关系数的取值范围为-1到1，其可以看成是无量纲的协方差。

2、统计意义

值越接近1，说明两个变量正相关性（线性）越强，越接近-1，说明负相关性越强，当为0时表示两个变量没有相关性。