# 第四章 随机变量的数字特征

由前面的讨论知道，分布函数完整地描述了随机变量的统计规律，然而在一些实际问题中要确定一个随机变量的分布函数却是非常困难的，而且有一些实际问题，并不要求全面考察随机变量的统计规律，而只需知道它的某些特征，因而并不需要求出它的分布函数，例如考察日光灯管的质量，常常关心的是日光灯管的平均寿命，即其平均寿命是一个重要指标.这就是说，随机变量的平均值，常常是一个重要的数量特征。在考察日光灯管的质量时还不能单就平均寿命来决定其质量，还必须要考察日光灯管的寿命与平均寿命的偏离程度，只有平均寿命较长同时偏离程度又较小的日光灯管才是质量较好的。随机变量与其平均值偏离的程度也是一个重要的数量特征。这些与随机变量有关的数量，虽不能完整地描述它的统计规律，但已反映出随机变量在某些方面的重要特征，它们在理论和实践上都具有重要的意义。本章将介绍常用的随机变量的数字特征：数学期望、方差、相关系数和矩。

## 4.1 随机变量的数学期望

### 4.1.1 离散型随机变量的数学期望

**例 1** 甲、乙两射手进行射击训练，已知在 100 次射击中命中环数与次数记录如下：

甲：

| 环数 | 8 | 9 | 10 |
|---|---|---|---|
| 次数 | 30 | 10 | 60 |

乙：

| 环数 | 8 | 9 | 10 |
|---|---|---|---|
| 次数 | 20 | 50 | 30 |

试问如何评定甲、乙射手的技术优劣？

**解** 从上面的成绩表很难立即看出结果，我们可以从其平均射中的环数来评定其技术优劣。

甲平均射中的环数为

$$(8\times30+9\times10+10\times60)\div100=8\times0.3+9\times0.1+10\times0.6=9.3\text{（环）},$$

乙平均射中的环数为

$$(8\times20+9\times50+10\times30)\div100=8\times0.2+9\times0.5+10\times0.3=9.1\text{（环）},$$

故从平均射中的环数看，甲的技术优于乙。

本例中 $\frac{30}{100}=0.3,\frac{60}{100}0.6,\frac{50}{100}=0.5$ 等是事件 $\{X=k\}$ 在 100 次试验中发生的频率（ $X$ 为命中环数），当射击次数相当大时，这个频率接近于事件 $\{X=k\}$ 一次试验中发生的概率 $p_k$，上述平均环数的计算可表示为 $\sum_{k=8}^{10}kp_k$，称之为随机变量 $X$ 的数学期望或均值.下面给出定义。

**定义 4.1** 设离散型随机变量 $X$ 的分布律为

$$P(X = x_k) = p_k, k = 1,2,\cdots,$$

若级数 $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛，则称其和为随机变量 $X$ 的**数学期望**或**平均值**，简称**期望**或**均值**,记为 $E(X)$，即 $$E(X) = \sum_{k=1}^{\infty} x_k p_k .$$

随机变量 $X$ 的数学期望 $E(X)$ 完全是由 $X$ 的分布律确定的，而不应受 $X$ 的可能取值的排列次序的影响，因此要求级数 $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛。若级数 $\sum_{k=1}^{\infty} x_k p_k$ 不绝对收敛，则称随机变量 $X$ 的数学期望不存在。

若把 $x_1, x_2, \cdots, x_k, \cdots$ 看成 $x$ 轴上质点的坐标，而 $p_1, p_2, \cdots, p_k, \cdots$ 看成相应质点的质量，质量总和 $\sum_{k=1}^{\infty} x_k p_k$ =1，则(4-1)式就表示质点系的重心坐标。

**例 2** 设随机变量 $X$ 的分布列为

| $X$ | $-1$ | $3$ |
|---|---|---|
| $p$ | $\frac{2}{3}$ | $\frac{1}{3}$ |

求 $E(X)$ 。

**解** 由(4-1)式有 $$E(X) = -1 \times \frac{2}{3} + 3 \times \frac{1}{3} = \frac{1}{3}.$$

若将此例视为甲、乙两人“赌博"，甲赢的概率为 $\frac{1}{3}$，输的概率为 $\frac{2}{3}$，但甲每赢一次可从乙处得 3 元，而每输一次，要给乙 1 元，则 $E(X) = \frac{1}{3}$ 是指甲平均每次可赢 $\frac{1}{3}$ 元 。每个“赌徒”在参加赌博时，心中首先要盘算这个数字。这正是称 $E(X)$ 为“期望“的原因。

例 3 按规定，某公交车每天 8 点至 9 点和 9 点至 10 点都恰有一辆公交车到站，各车到站的时刻是随机的，且各车到站的时间是相互独立的，其规律为

| 到站时刻 | 8:10<br>9:10 | 8:30<br>9:30 | 8:50<br>9:50 |
|---|---|---|---|
| 概率 | $\frac{1}{5}$ | $\frac{2}{5}$ | $\frac{2}{5}$ |

其乘客 8:20 到站，求他候车时间的数学期望。

**解** 设乘客的候车时间为 $X$(单位为分)，若该乘客 8:20 到车站，而 8 点到 9 点的一趟车已于 8:10 开走，第二趟车 9:10 开，则他候车的时间为 50 分钟，对应的概率为事件“第一趟车 8:10 开走，且第二趟 9:10 开”发生的概率，即

$$P(X=50)=\frac{1}{5}\times\frac{1}{5}=\frac{1}{25}.$$

该乘客其余候车时间对应的概率可类似得到，于是候车时间 $X$ 的分布律为

| $X$ | 10 | 30 | 50 | 70 | 90 |
|---|---|---|---|---|---|
| $p$ | $\frac{2}{5}$ | $\frac{2}{5}$ | $\frac{1}{5}\times\frac{1}{5}$ | $\frac{1}{5}\times\frac{2}{5}$ | $\frac{1}{5}\times\frac{2}{5}$ |

从而该乘客候车时间的数学期望为

$$E(X)=10\times\frac{2}{5}+30\times\frac{2}{5}+50\times\frac{1}{25}+70\times\frac{2}{25}+90\times\frac{2}{25}=30.8\text{（分）}.$$

**例 4** 从一个装有。$m$ 个白球和 $n$ 个红球的袋中取球，直到出现白球为止。若每次取出的球仍放回袋中，试求取出红球数的数学期望.

**解** 设取出的红球数为 $X$，则 $X$ 的分布律为

$$P(X=k)=\left(\frac{n}{m+n}\right)^k\left(\frac{m}{m+n}\right),k=0,1,2,\cdots.$$

令
$$p=\frac{n}{m+n},q=\frac{m}{m+n},$$

则
$$P(X=k)=p^k q,k=0,1,2,\cdots,$$

于是
$$E(X)=\sum_{k=0}^{\infty}kp^k q=pq\sum_{k=1}^{\infty}kp^{k-1}$$

$$=pq\left(\sum_{k=1}^{\infty}p^k\right)'=pq\left(\frac{p}{1-p}\right)'$$

$$=\frac{pq}{(1-p)^2}=\frac{p}{q}=\frac{n}{m}.$$

### 4.1. 2 连续型随机变量的数学期望

若 $X$ 为连续型随机变量，其密度函数为 $f(x)$，则 $X$ 落入 $(x_k,x_k+dx)$ 内的概率可近似地表为 $f(x_k)dx$，它与离散型随机变量的 $p_k$ 类似，下面给出定义。

**定义 4.2** 设连续型随机变量 $X$ 的密度函数为 $f(x)$。若积分 $\int_{-\infty}^{+\infty}xf(x)dx$ 绝对收敛，称该积分值为随机变量 $X$ 的**数学期望**或**平均值**，简称**期望**或**均值**，记为 $E(X)$，即

$E(X)=\int_{-\infty}^{+\infty}xf(x)dx.$

若积分 $\int_{-\infty}^{+\infty}xf(x)dx.$不绝对收敛，则称随机变量 X 的数学期望不存在。

设在 $x$ 轴上连续分布着质量，其线密度为 $f(x)$，则由

$$\int_{-\infty}^{+\infty}xf(x)dx.=\frac{\int_{-\infty}^{+\infty}xf(x)dx}{\int_{-\infty}^{+\infty}f(x)dx},$$

可知 $E(x)$ 的物理意义是质量密度为 $f(x)$ 的一维连续质点系的质量中心的坐标。

**例 5** 设随机变量 $X$ 服从柯西(Cauchy)分布,其密度函数为

$$f(x)=\frac{1}{\pi(1+x)^2},-\infty<x<+\infty$$

试证 $X$ 的数学期望不存在。

**证** 因为

$$\int_{-\infty}^{+\infty}|x|\ f(x)dx=\int_{-\infty}^{+\infty}\frac{|x|}{\pi(1+x^2)}dx$$

$$=2\int_{0}^{+\infty}\frac{x}{\pi(1+x^2)}dx$$

$$=\frac{1}{\pi}\ln(1+x^2)\ \Big|_{0}^{+\infty}$$

$$=+\infty,$$

即 $\int_{-\infty}^{+\infty}xf(x)dx$ 不绝对收敛，所以 $E(x)$ 不存在。

**例 6** 有 5 个相互独立工作的电子装置，其寿命 $X_k\ (k=1,2,3,4,5)$ 服从同一指数分布，分布函数为

$$F(x)=\begin{cases}1-e^{-\lambda x},x\geq 0,\lambda>0.\\ 0,x<0\end{cases}$$

若将这 5 个电子装置串联组成整机，求整机寿命 N 的数学期望；

(1) 若将这 5 个电子装置串联组成整机，求整机寿命 M 的数学期望。

(2) 若将这 5 个电子装置并联组成整机，求整机寿命 M 的数学期望。

**解** 由随机变量函数的分布可知

(1) $N=\min(X_1,X_2,X_3,X_4,X_5)$ 的分布函数为

$$F_N(x)=1-[1-F(x)]^5=\begin{cases}1-e^{-5\lambda x}, x\geq 0\\ 0, x<0\end{cases}$$

其密度函数为

$$f_N(x)=\begin{cases}5\lambda e^{-5\lambda x}, x\geq 0\\ 0, x<0\end{cases}$$

故

$$E(N)=\int_{-\infty}^{+\infty} xf_N(x)dx=\int_0^{+\infty} x\cdot 5\lambda e^{-5\lambda x}dx=\frac{1}{5\lambda}$$

(2) $M=\min(X_1,X_2,X_3,X_4,X_5)$ 的分布函数

$$F_M(x)=[F(x)]^5=\begin{cases}(1-e^{-\lambda x})^5, x\geq 0\\ 0, x<0\end{cases}$$

其密度函数为

$$f_M(x)=\begin{cases}5\lambda(1-e^{-\lambda x})^4, x\geq 0\\ 0, x<0\end{cases}$$

故

$$E(M)=\int_{-\infty}^{+\infty} xf_M(x)dx=\int_0^{+\infty} x\cdot 5\lambda(1-e^{-\lambda x})^4 e^{-\lambda x}dx=\frac{137}{60\lambda}$$

由

$$\frac{E(M)}{E(N)}=\frac{137/60\lambda}{1/5\lambda}\approx 11.4$$

可知，同样 5 个电子装置，并联组成整机的平均寿命是串联组成整机的平均寿命的 11.4 倍。

### 4.1.3 随机变量函数的数学期望

在实际问题中常常需要求出随机变量的函数的数学期望，例如 $Y=g(X)$，要求 $E(Y)$。我们可以不必求出 $Y$ 的密度函数，而直接由 $X$ 的密度函数来求 $E(Y)$。

**定理 4.1** 设随机变量 $Y$ 是随机变量 $X$ 的函数 $Y=g(X)$ ($g$ 为连续函数)。

(1) 设 $X$ 为离散型变量，其分布律为 $P(X=x_k)=p_k, k=1,2,\ldots$

若级数 $\sum_{k=1}^{\infty} g(x_k)p_k$ 绝对收敛，则有

$$E(Y)=E[g(X)]=\sum_{k=1}^{\infty} g(x_k)p_k.$$

(2) 设$X$为连续型变量,其密度函数为$f(x)$。若积分$\int_{-\infty}^{+\infty}g(x)f(x)dx$绝对收敛，则有
$$E(Y)=E[g(X)]=\int_{-\infty}^{+\infty}g(x)f(x)dx$$

这个定理说明，在求$Y=g(X)$的数学期望时，不必知道$Y$的分布而只需知道$X$的分布即可。定理的证明超出了本书的范围，此处从略。

这个定理还可以推广到两个或多个随机变量的函数的情况。

设$Z$是随机变量$X,Y$的函数$Z=g(X,Y)$($g$为连续函数)，则$Z$也是一个随机变量.

若$(X,Y)$为离散型随机变量，且其联合分布律为
$$P(X=x_i,Y=y_i)=p_{ij},i,j=1,2\ldots,$$
则有
$$E(Z)=E[g(X,Y)]=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}g(x_i,y_i)p_{ij}.$$

若$(X,Y)$为连续型随机变量，且其联合密度函数为$f(x,y)$，则有
$$E(Z)=E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)dxdy.$$
这里要求等式右端的级数或积分都是绝对收敛的。

**例 7** 设随机变量$X$服从参数为$n,p$的二项分布,$Y=e^{2X}$,求 E($Y$)。

**解** 因为$X\sim B(n,p)$，分布律为
$$P(X=k)=C_n^k p^k q^{n-k},k=0,1,2,\ldots,n,$$
所以
$$E(Y)=E(e^{2X})=\sum_{k=0}^{n}e^{2k}C_n^k p^k q^{n-k}=\sum_{k=0}^{n}C_n^k(p\ e^2)^k q^{n-k}=(pe^2+q)^n$$
其中$p+q=1$

**例 8** 设二维随机变量$(X,Y)$的密度函数为
$$f(x,y)=\begin{cases}x+y, & 0\le x\le 1,\ 0\le y\le 1\\ 0, & \text{其他}\end{cases}$$
求$E(XY)$.

**解** $E(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)dxdy=\int_0^1\int_0^1 xy(x+y)dxdy=\dfrac{1}{3}$

**例 9** 设二维随机变量$(X,Y)$的密度函数为

$$f(x,y)=\begin{cases}\frac{1}{4}x(1+3y^2), & 0<x<2,\ 0<y<1\\ 0, & \text{其他}\end{cases}$$

求$E(X)$、$E(Y)$、$E(XY)$和/ $E(\frac{Y}{X})$.

**解**

$$E(X)=\int_0^1\int_0^2 x\bullet\frac{1}{4}x(1+3y^2)dxdy=\frac{1}{4}\int_0^2 x^2dx\int_0^1(1+3y^2)dy=\frac{4}{3}.$$

$$E(Y)=\int_0^1\int_0^2 y\bullet\frac{1}{4}x(1+3y^2)dxdy=\frac{1}{4}\int_0^2 x\ dx\int_0^1 y(1+3y^2)dy=\frac{5}{8}.$$

$$E(XY)=\int_0^1\int_0^2 xy\bullet\frac{1}{4}x(1+3y^2)dxdy=\frac{1}{4}\int_0^2 x^2dx\int_0^1 y(1+3y^2)dy=\frac{5}{6}.$$

$$E(\frac{Y}{X})=\int_0^1\int_0^2\frac{y}{x}\bullet\frac{1}{4}x(1+3y^2)dxdy=\frac{1}{4}\int_0^2dx\int_0^1 y(1+3y^2)dy=\frac{5}{8}.$$

**例 10** 设国际市场上每年对我国某种出口商品的需求量是随机变量$X$(单位:吨)，它服从[2000，4000]上的均匀分布。若售出这种商品 1 吨，可赚 3 万元，但销售不出去，则每吨需付仓库保管费 1 万元，问该商品应出口多少吨才可的得到最大利益?

**解** 设每年出口该种商品 y 吨$(2000\le y\le 4000)$，则收益

$$Y=g(X)=\begin{cases}3y, & X\ge y\\ 3X-(y-X), & X<y\end{cases}=\begin{cases}3y, & X\ge y\\ 4X-y, & X<y\end{cases}$$

于是由 $$g(x)=\begin{cases}3y, & x\ge y\\ 4x-y, & x<y\end{cases}$$

$$f(x)=\begin{cases}\frac{1}{2000}, 2000\le x\le 4000\\ 0\end{cases},$$

得 $$E(Y)=\int_{2000}^{4000}g(x)\bullet\frac{1}{2000}dx=\frac{1}{2000}[\int_{2000}^{y}(4x-y)dx+\int_y^{4000}3ydx$$

$$=\frac{1}{2000}(-y^2+7000y-4\times 10^6).$$

当$y=3500$时，$E(Y)$取到最大值，故出口此种商品 3500 吨长可得到最大收益。

### 4.1.4 数学期望的性质

现在给出数学期望的几个常用性质。在下面的讨论中，所遇到的随机变量的数学期望均假设存在，且只对连续型随机变量给予证明，至于对离散型随机变量的证明只需将积分换为类似的求和即可。

**1.** 设 $C$ 为常数，则有 $E(C)=C$.

**证** 可将 $C$ 看成离散型随机变量，分布律为 $P(X=C)=1$. 故由定义即得 $E(C)=C$.

**2.** 设 $C$ 为常数，$X$ 为随机变量，则有 $E(CX)=CE(X)$.

**证** 设 $X$ 的密度函数为 $f(x)$，则有

$$E(CX)=\int_{-\infty}^{+\infty}Cxf(x)dx=C\int_{-\infty}^{+\infty}xf(x)dx$$
$$=CE(X)$$

**3.** 设 $X,Y$ 为任意两个随机变量，则有 $E(X+Y)=E(X)+E(Y)$.

**证** 设二维随机变量 $(X,Y)$ 的密度函数为 $f(x,y)$，边缘密度函数分别为 $f_X(x)$、$f_Y(y)$，,则

$$E(X+Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)f(x,y)\,dxdy$$
$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)\,dxdy+\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)\,dxdy$$
$$=\int_{-\infty}^{+\infty}xf_X(x)\,dx+\int_{-\infty}^{+\infty}yf_Y(y)\,dx$$
$$=E(X)+E(Y).$$

这一性质可以推广到任意有限多个随机变量之和的情形,即

$$E(X_1+X_2+\ldots+X_n)=E(X_1)+E(X_2)+\ldots+E(X_n).$$

一般地，随机变量线性组合的数学期望，等于随机变量数学期望的线性组合，即

$$E(a_1X_1+a_2X_2+\ldots+a_nX_n)=a_1E(X_1)+a_2E(X_2)+\ldots+a_nE(X_n).$$

其中 $a_1,a_2,\ldots a_n$ 为常数。

**4.** 设 $X,Y$ 为相互独立的随机变量，则有 $E(XY)=E(X)E(Y)$.

**证** 因为 $X$ 与 $Y$ 相互独立，其联合密度函数与边缘密度函数满足

$f(x,y)=f_X(x)f_Y(y)$，所以

$$E（XY）=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf（x，y）dxdy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf_X（x）f_Y(y)dxdy$$

$$=[\int_{-\infty}^{+\infty}xf_X（x） dx]\bullet[\int_{-\infty}^{+\infty}yf_Y（x） dx]=E(X)E(Y).$$

这一性质可以推广到任意有限多个相互独立的随机变量之积的情形，即若 $X_1,X_2,...,X_n$ 为相互独立的随机变量，则有

$$E(X_1X_2...X_n)=E(X_1)E(X_2)...E(X_n).$$

**例 11** 一民航机场的送客班车载有 20 位旅客，自机场开出，沿途旅客有 10 个车站可以下车。如到达一个车站没有旅客下车班车就不停。设每位旅客在各个车站下车是等可能的。且各旅客是否下车相互独立，以 $X$ 表示停车的次数，求 $E(X)$。

**解** 设随机变量

$$X_i=\begin{cases}0,\text{第 i 个车站无人下车}\\1,\text{第 i 个车站有人下车}\end{cases}\quad(i=1,2,...10)$$

则有 $X=X_1+X_2+...+X_{10}$

由题意，任一旅客在第 $i$ 个车站不下车的概率为 $\frac{9}{10}$, $X_i=0$ 表示第 $i$ 站没有旅客下车，故 20 位旅客都不在第 $i$ 站下车的概率为 $\left(\frac{9}{10}\right)^{20}$，在第 $i$ 站有人下车的概率为 $1-\left(\frac{9}{10}\right)^{20}$，于是得 $X_i$ 的分布律如下：

| $X_i$ | 0 | 1 |
|---|---|---|
| P | $(9/10)^{20}$ | $1-(9/10)^{20}$ |

因此 $E(X_i)=0\times\left(\frac{9}{10}\right)^{20}+1\times\left[1-\left(\frac{9}{10}\right)^{20}\right]=1-\left(\frac{9}{10}\right)^{20}, i=1,2,...10,$

从而

$$E(X)=E(X_1+X_2+...+X_{10})=E(X_1)+E(X_2)+...+E(X_{10}).$$

$$=10\times\left[1-\left(\frac{9}{10}\right)^{20}\right]\approx 8.8$$

这表明班车平均停车约 9 次。

类似本例将 $X$ 分解为若干个随机变量的和，然后利用数学期望的性质再求 $X$ 的数学期望的方法，具有一定的普遍意义，使用得当，可使复杂问题简单化。

**例 12** 设二维随机变量 $(X,Y)$ 的密度函数为 $f(x,y)=\begin{cases}\dfrac{1}{\pi},x^2+y^2\le 1\\ \quad 0,\end{cases}$

试验证 $E(XY)=E(X)E(Y)$, 但 $X$ 和 $Y$ 是不独立的.

**解** 因为 $E(XY)=\iint\limits_{x^2+y^2\le 1} xy\bullet\dfrac{1}{\pi}dxdy=\dfrac{1}{\pi}\int_{-1}^{1}xdx\int_{-\sqrt{1-x^2}}^{\sqrt{1+x^2}}ydy=0$,

$$E(X)=\iint\limits_{x^2+y^2\le 1} x\bullet\frac{1}{\pi}dxdy=0,$$

$$E(Y)=\iint\limits_{x^2+y^2\le 1} y\bullet\frac{1}{\pi}dxdy=0,$$

所以 $E(XY)=E(X)E(Y)$.

$X$ 的边缘密度函数

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dy=\frac{2}{\pi}\sqrt{1-x^2},-1\le x\le 1,$$

即 $f_X(x)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-x^2},-1\le x\le 1\\ \quad 0\end{cases}$

同理可得 $Y$ 的边缘密度函数 $f_Y(y)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-y^2},-1\le y\le 1\\ \quad 0\end{cases}$

因为，所以 $f(x,y)\ne f_X(x)f_Y(y)$, $X$ 和 $Y$ 是不独立的。

本例说明由 $E(XY)=E(X)E(Y)$. 是不能得出 $X$ 和 $Y$ 是相互独立的结论的.

# 4.2 随机变量的方差

## 4.2.1 方差概念

例如，有一批灯泡，其平均寿命 $E(X)$=1000 小时，但仅由这一指标还不能判断这一批灯泡质量的好坏，我们还需考察灯泡寿命 $X$ 与 $E(X)$ 的偏离程度，若偏离程度较小，则灯泡质量比较稳定。因此，研究随机变量与其平均值的偏离程度是十分重要的。

用什么量去表示随机变量 $X$ 与其数学期望的偏离程度呢？显然，可用随机变量 $|X-E(X)|$ 的平均值 $E[|X-E(X)|]$ 来表示 $X$ 与 $E(X)$ 的偏离程度，但为了数字上处理的方便，通常用 $E[X-E(X)]^2$ 来表示 $X$ 与 $E(X)$ 的偏离程度。

**定义 4.3** 设 $X$ 是一个随机变量。若 $E[X-E(X)]^2$ 存在，则称 $E[X-E(X)]^2$ 为 ***X*的方差**，记为 $D(X)$，即 $D(X)=E[X-E(X)]^2$。

与 $X$ 具有相同量纲的量 $\sqrt{D(X)}$ 称为 $X$ 的**均方差**或**标准差**，记为 $\sigma_x$。

由定义，随机变量 $X$ 的方差反映出 $X$ 的取值与其数学期望的偏离程度。若 $D(X)$ 较小，则 $X$ 取值比较集中，否则，$X$ 取值比较分散。因此，方差 $D(X)$ 是刻化 $X$ 取值分散程度的一个量。

方差实际上是随机变量 $X$ 的函数的数学期望。故若 $X$ 为离散型随机变量，其分布律为

$$P(X=x_k)=p_k, k=1,2,\cdots,$$

则

$$D(X)=\sum_{k=1}^{\infty}[x_k-E(X)]^2 p_k.$$

若 $X$ 为连续型随机变量，其密度函数为 $f(x)$，则

$$D(X)=\int_{-\infty}^{+\infty}[x-E(X)]^2 f(x)dx.$$

另外，还有一个常用的计算方差的重要公式：$D(X)=E(X^2)-[E(X)]^2$.

事实上，

$$\begin{aligned}D(X)&=E[X-E(X)]^2=E[X^2-2XE(X)+(E(X))^2]\\&=E(X^2)-2E(X)E(X)+[E(X)]^2=E(X^2)-[E(X)]^2.\end{aligned}$$

**例 1** 设随机变量 $X$ 服从（0-1）分布，分布律为

$P$（$X$=1）$=p$，$P$（$X$=0）=1-$p=q$，

求 $D(X)$。

**解** 因为 $E(X)=p, E(X^2)=1^2\times p+0^2\times q=p,$

所以 $D(X)=E(X^2)-[E(X)]^2=p-p^2=p(1-p)=pq.$

**例 2** 设随机变量 X 的密度函数为

$$f(x)=\begin{cases}1+x,-1\le x\le 0\\ 1-x,0<x\le 1\end{cases},$$

求 $D(X)$。

**解** 因为

$$E(X)=\int_{-1}^{0}x(1+x)dx+\int_{0}^{1}x(1-x)dx=0,$$

$$E(X^2)=\int_{-1}^{0}x^2(1+x)dx+\int_{0}^{1}x^2(1-x)dx=\frac{1}{6},$$

所以 $$D(X)=E(X^2)-[E(X)]^2=\frac{1}{6}.$$

### 4.2.2 方差的性质

假定下面所遇到的随机变量的方差均存在。

**1.** 设 $C$ 为常数，则 $D(C)=0$。

**证** $D(X)=E(C^2)-[E(C)]^2=C^2-C^2=0$.

**2.** 设 $X$ 为随机变量，$C$ 为常数，则有 $D(CX)=C^2D(X)$.

**证** $D(CX)=E(C^2X^2)-[E(CX)]^2=C^2[E(X^2)-(E(X))^2]=C^2D(X).$

**3.** 设随机变量 $X$ 与 $Y$ 相互独立，则有 $D(X+Y)=D(X)+D(Y)$。

**证**

$$\begin{aligned}D(X+Y)&=E[(X+Y)-E(X+Y)]^2=E[(X-E(X))+(Y-E(Y))]^2\\&=E[X-E(X)]^2+E[Y-E(Y)]^2+2E[X-E(X)][Y-E(Y)]\\&=D(X)+D(Y)+2E[X-E(X)][Y-E(Y)]\\&=D(X)+D(Y)+2E[X-E(X)][Y-E(Y)].\end{aligned}$$

因为

$$\begin{aligned}E[X-E(X)][Y-E(Y)]&=E[XY+E(X)E(Y)-XE(Y)-YE(Y)]\\&=E(XY)+E(X)E(Y)-E(X)E(Y)-E(Y)E(X)=E(XY)-E(X)E(Y),\end{aligned}$$

由于 $X$ 与 $Y$ 相互独立，$E(XY)=E(X)E(Y)$，故 $E[X-E(X)][Y-E(Y)]=0$,
所以 $D(X+Y)=D(X)+D(Y)$。

性质 3 可推广到有限多个相互独立的随机变量之和的情形。即若 $X_1,X_2,\cdots,X_n$ 相互独立，则有 $D(X_1+X_2+\cdots+X_n)=D(X_1)+D(X_2)+\cdots+D(X_n)$.

若 $X$ 与 $Y$ 是相互独立的随机变量，$C_1,C_2$ 为常数，则有

$$D(C_1X+C_2Y)=C_1^2D(X)+C_2^2D(Y).$$

特别地，$D(X-Y)=D(X)+D(Y)$。

**例 3** 设随机变量 $X$ 服从几何分布，其分布律为 $P(X=k)=p(1-p)^{k-1}, k=1,2,\cdots,$

其中 $0<p<1$ 为常数，求 $E(X)$ 和 $D(X)$。

**解**

$$E(X)=\sum_{k=1}^{\infty}kp(1-p)^{k-1}=\frac{p}{[1-(1-p)]^2}=\frac{1}{p},$$

$$E(X^2)=\sum_{k=1}^{\infty}k^2p(1-p)^{k-1}=p\cdot\frac{1+(1-p)}{[1-(1-p)]^3}=\frac{2-p}{p^2},$$

于是 $D(X)=E(X^2)-[E(X)]^2=\dfrac{1-p}{p^2}.$

## 4.2.3 几种重要随机变量的数学期望和方差

**1.二项分布**

设随机变量 $X$ 服从参数为 $n$、$p$ 的二项分布，即 $X\sim B(n，p)$，其分布律为

$P(X=k)=C_n^k p^k q^{n-k}, k=0,1,2,\cdots,n$,其中 $0<p<1$，$p+q=1$，则

$$E(X)=\sum_{k=0}^{n}kp_k=\sum_{k=0}^{n}kC_n^k p^k q^{n-k}=\sum_{k=0}^{n}k\cdot\frac{n!}{k!(n-k)!}p^k q^{n-k}$$

$$=np\sum_{k=0}^{n}\frac{(n-1)!}{(k-1)![(n-1)-(k-1)]!}p^{k-1}q^{(n-1)-(k-1)}$$

$$=np\sum_{k=0}^{n}C_{n-1}^{k-1}p^{k-1}q^{(n-1)-(k-1)}=np\sum_{k=0}^{n}C_{n-1}^{k}p^{k-1}q^{n-1-k}$$

$$=np(p+q)^{n-1}=np.$$

对于服从二项分布的随机变量，还可用更简单的方法来计算 $E(X)$与 $D(X)$.

在 $n$ 重贝努里试验中，每次试验事件 $A$ 发生的概率为 $p$，不发生的概率为 $q$=1-$p$，若引入随机变量

$X_i=\begin{cases}1,\text{第}i\text{次实验}A\text{发生}\\0，\text{第}i\text{次实验}A\text{不发生}\end{cases}, i=1,2,\cdots,n$ 则 $A$ 发生的次数为 $X=X_1+X_2+\cdots+X_n$

$X\sim B(n,p), X_i\sim(0-1)$ 分布，且 $X_1,X_2,\cdots,X_n$ 是相互独立的，于是由数学期望的性质可

得 $$\begin{aligned}E(X)&=E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n)=nE(X_i)=np,\\D(X)&=D(X_1+X_2+\cdots+X_n)=D(X_1)+D(X_2)+\cdots+D(X_n)=nD(X_i)=npq.\end{aligned}$$

**2.泊松分布**

设随机变量 $X$ 服从参数为 $\lambda$ 泊松分布，即 $X\sim P(\lambda)$，其分布律为

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},k=0,1,2,\cdots,\lambda>0,\text{则}$$

$$E(X)=\sum_{k=0}^{\infty}kp_k=\sum_{k=0}^{\infty}k\cdot\frac{\lambda^k}{k!}e^{-\lambda}=\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=\lambda e^{-\lambda}\cdot e^{\lambda}=\lambda,$$

$$D(X)=E(X^2)-[E(X)]^2=E[X(X-1)+X]-\lambda^2$$

$$=E[X(X-1)]+\lambda-\lambda^2=\sum_{k=1}^{\infty}k(k-1)\frac{\lambda^k}{k!}e^{-\lambda}+\lambda-\lambda^2$$

$$=\lambda^2e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}+\lambda-\lambda^2=\lambda^2e^{-\lambda}\cdot e^{\lambda}+\lambda-\lambda^2=\lambda.$$

**3.均匀分布**

设随机变量 $X$ 在区间[$a$，$b$]上服从均匀分布，即 $X$~U[$a$，$b$]，其密度函数

$$f(x)=\begin{cases}\dfrac{1}{b-a},a\le x\le b\\ 0,\text{其他}\end{cases},$$

则

$$E(X)=\int_a^b x\cdot\frac{1}{b-a}dx=\frac{a+b}{2},$$

$$D(X)=E(X^2)-[E(X)]^2$$

$$=\int_a^b x^2\cdot\frac{1}{b-a}dx-\left(\frac{a+b}{2}\right)^2=\frac{(b-a)^2}{12}.$$

**4. 指数分布**

设随机变量 $X$ 服从以 $\lambda$ 为参数的指数分布，即 $X\sim E(\lambda)$，其密度函数

$$f(x)=\begin{cases}\lambda e^{-\lambda},x\ge 0,\lambda>0\\ 0,x<0\end{cases},$$

则

$$E(X)=\int_0^{+\infty}x\cdot\lambda e^{-\lambda x}dx=-[xe^{-\lambda x}\mid_0^{+\infty}-\int_0^{+\infty}e^{-\lambda x}dx]=-\frac{1}{\lambda}e^{-\lambda x}\mid_0^{+\infty}=\frac{1}{\lambda},$$

$$D(X)=E(X^2)-[E(X)]^2=\int_0^{+\infty}x^2\cdot\lambda e^{-\lambda x}dx-\left(\frac{1}{\lambda}\right)^2=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}.$$

**5.正态分布**

设随机变量 $X$ 服从参数为 $\mu$、$\sigma$ 的正态分布，即 $X\sim N(\mu,\sigma^2)$，其密度函数

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x+<\infty,$$

则

$$E(X)=\int_{-\infty}^{+\infty}x\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(x-\mu)e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx+\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}\mu e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}te^{-\frac{t^2}{2\sigma^2}}dt+\mu\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\mu,$$

$$D(X)=E[X-E(x)]^2=\int_{-\infty}^{+\infty}(x-\mu)^2\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t^2e^{-\frac{t}{2}}dt\left(t=\frac{x-\mu}{\sigma}\right)$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\left[t(-e^{-\frac{t}{2}})\,|_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty}e^{-\frac{t}{2}}dt\right]$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\cdot\sqrt{2\pi}=\sigma^2.$$

从上面所述的各种随机变量可见，一些常用分布的数学期望和方差知道后，其分布中的参数也就知道了，从而分布也就唯一确定。由此可见随机变量数字特征的重要性。

在概率论里，有时需要将随机变量“标准化”，即对任意随机变量 $X$，若其数学期望 $E(X)$，方差 $D(X)$均存在，且 $D(X)>0$，则称

$$X^*=\frac{X-E(X)}{\sqrt{D(X)}}$$

为 $X$ 的**标准化随机变量.**

通过计算可知 $E(X^*)=0, D(X^*)=1$，这正是标准化随机变量所具有的特征。

# 4.3　协方差和相关系数　矩

## 1.协方差和相关系数

对于二维随机变量 $(X,Y)$，我们除了讨论随机变量 $X$ 与 $Y$ 的数学期望和方差之外，还要给出一个描述 $X$ 与 $Y$ 之间相互关系的数字特征。由（4.2.2）式我们知道，若 $X$ 与 $Y$ 相互独立，则有 $E[X-E(X)][Y-E(Y)]=0$；若 $E[X-E(X)][Y-E(Y)]\neq 0$，则说明 $X$ 与 $Y$ 不相互独立，而是有一定的关系的。因此给出以下定义。

**定义 4.4** 设 $(X,Y)$ 是一个二维随机变量。若 $E[X-E(X)][Y-E(Y)]$ 存在，则称它是随机变量 $X$ 与 $Y$ 的协方差，记为 $\operatorname{cov}(X,Y)$，即

$$\operatorname{cov}(X,\ Y)=E[X-E(X)][Y-E(Y)]$$

而当 $D(X)>0, D(Y)>0$ 时，

$$\rho_{XY}=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

称为随机变量 $X$ 与 $Y$ 的**相关系数。**

当 $\rho_{XY}=0$ 时，称随机变量 $X$ 与 $Y$ 是**不相关的**或**无关的**。

由定义及(4.2.2)式可知，对于任意两个随机变量 $X$ 与 $Y$ 下列等式成立：

$$D(X+Y)=D(X)+D(Y)+2\operatorname{cov}(X,Y),$$

$$\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y).$$

**例 1** 设二维随机变量 $(X,Y)$ 的联合分布律为

| Y \ X | 0 | 1 |
|---|---|---|
| 0 | q | 0 |
| 1 | 0 | p |

其中 $p+q=1$，求相关系数 $\rho_{XY}$。

**解**　由上面的分布律，可得随机变量 $X$ 与 $Y$ 的边缘分布律为

| | | |
|---|---|---|
| $X$ | 0 | 1 |
| $p$ | $q$ | $p$ |
| $Y$ | 0 | 1 |
| $p$ | $q$ | $p$ |

$X$与$Y$均服从(0-1)分布，故知

$$E(X)=p, D(X)=pq,$$

$$E(Y)=p, D(Y)=pq.$$

于是有$\mathrm{cov}(X,Y)=E(XY)-E(X)E(Y)$

$$=0\times0\times q+0\times1\times0+1\times0\times0+1\times1\times p-p\times p=p\times p^2=pq,$$

得$\rho_{XY}=\dfrac{\mathrm{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\dfrac{pq}{\sqrt{pq}\sqrt{pq}}=1.$

**例 2** 设二维随机变量$(X,Y)$的密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{(b-a)(d-c)}, a\le x\le b, c\le y\le d\\ 0,\text{其它}\end{cases},$$

求$\mathrm{cov}(X,Y)$.

**解** 因为$E(X)=\int_a^b\int_c^d x\cdot\dfrac{1}{(b-a)(d-c)}dxdy=\dfrac{a+b}{2},$

$$E(Y)=\int_a^b\int_c^d y\cdot\frac{1}{(b-a)(d-c)}dxdy=\frac{c+d}{2},$$

$$E(XY)=\int_a^b\int_c^d xy\cdot\frac{1}{(b-a)(d-c)}dxdy=\frac{(a+b)(c+d)}{4},$$

所以$\mathrm{cov}(X,Y)=E(XY)-E(X)E(Y)=0.$

协方差的性质:

**1.** $\mathrm{cov}(X,Y)=\mathrm{cov}(Y,X)$;

**2.** $\mathrm{cov}(aX,bY)=ab\,\mathrm{cov}(X,Y), a, b$为常数；

**3.** $\mathrm{cov}(X+Y, Z)=\mathrm{cov}(X,Z)+\mathrm{cov}(Y,Z).$

以上性质可由协方差的定义直接推得。

将随机变量$X$与$Y$标准化，得

$$X^*=\frac{X-E(X)}{\sqrt{D(X)}}, Y^*=\frac{Y-E(X)}{\sqrt{D(Y)}}.$$

由相关系数定义，显然有$\rho_{XY}=\mathrm{cov}(X^*,Y^*).$

下面我们来讨论相关系数的性质，进而说明相关系数反映了随机变量间的一种相互关系的本质。

相关系数的性质:

**1.** $|\rho_{XY}|\le 1.$

**证** 由 $D(X^*\pm Y^*)=D(X^*)+D(Y^*)\pm 2\operatorname{cov}(X^*,Y^*)$

$$=1+1\pm 2\operatorname{cov}(X^*,Y^*)=2(1\pm\rho_{XY})\ge 0,$$

即 $1\pm\rho_{XY}\ge 0,\qquad |\rho_{XY}|\le 1.$

**2.** $|\rho_{XY}|=1$ 的充分必要条件是 $X$ 与 $Y$ 依概率 1 线性相关，即 $P(Y=aX+b)=1$，其中 $a\ne 0, a, b$ 为常数。

**证** 由方差性质 4，知 $P(Y=aX+b)=P(Y-aX-b=0)=1$
成立的充分必要条件为

$$D(Y-aX-b)=E[(Y-aX-b)^2]-[E(Y-aX-b)]^2$$

$$=E[(Y-aX-b)^2]=0$$

而 $E[(Y-aX-b)^2]=E[(Y-E(Y))-a(X-E(X))+(E(Y)-aE(X)-b)]^2$

$$=E[(Y-E(Y)]^2+a^2E[X-E(X)]^2+E[E(Y)-aE(X)-b]^2$$

$$-2aE[Y-E(Y)][X-E(X)]+2E[Y-E(Y)][E(Y)-aE(X)-b]$$

$$-2aE[X-E(X)][E(Y)-aE(X)-b]$$

$$=D(Y)+a^2D(X)+[E(Y)-aE(X)-b]^2-2a\operatorname{cov}(X,Y)$$

$$=D(X)[a-\frac{\operatorname{cov}(X,Y)}{D(X)}]^2+D(Y)[1-(\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}})^2]+[E(Y)-aE(X)-b]^2,$$

上式右端 3 项均是非负的，故由 $E[(Y-aX-b)^2]=0,$

必有
$$1-\left(\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}\right)^2=0,$$

即 $1-\rho^2{}_{XY}=0,$ 从而 $|\rho_{XY}|=1.$

若考虑用 $(X,Y)\sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho)$

的任意线性函数$\alpha X+\beta$来近似表示 $Y$，则有误差$Y-\alpha X-\beta$。我们以均方误差

$$E[Y-\alpha X-\beta]^2=D(X)\left[\alpha-\frac{\operatorname{cov}(X,Y)}{D(X)}\right]^2+D(Y)(1-\beta^2 xy)+[E(Y)-\alpha E(X)-\beta]^2$$

来衡量用$\alpha X+\beta$近似表示$Y$的好坏程度，均方误差的值越小表示$\alpha X+\beta$与$Y$的近似程度越好。由于上式右端每项均是非负的，故若取第一、三项为零，则均方误差取得最小值$D(Y)(1-\beta^2 xy)$,即取$\hat{a}=\frac{\operatorname{cov}(X,Y)}{D(X)}=\rho_{XY}\sqrt{\frac{D(Y)}{D(X)}}$,

$$\hat{\beta}=E(Y)-\alpha E(X)=E(Y)-\rho_{XY}E(X)\sqrt{\frac{D(Y)}{D(X)}}$$

$$E[Y-\hat{\alpha}X-\hat{\beta}]^2=\min_{\alpha,\beta}[Y-\alpha X-\beta]^2=D(Y)(1-\rho^2{}_{XY}) \qquad (4.4.1)$$

（4.4.1）式表明，用线性函数$\hat{\alpha}X+\beta$表示$Y$，比用其他任何线性函数$\alpha X+\beta$表示$Y$都好。由（4.4.1）式知，均方误差是$|\rho_{XY}|$的严格单调减少函数，$|\rho_{XY}|$越大，均方误差越小，这时$Y$与$X$的线性关系就越密切，当$|\rho_{XY}|$=1 时，$Y$与$X$就有线性关系。反之，$|\rho_{XY}|$越小，均方误差就越大，说明$Y$与$X$的线性关系就越差，若$|\rho_{XY}|$=0，则均方误差最大，表明$Y$与$X$间无线性关系，故称$X$与$Y$是不相关的。可见，$|\rho_{XY}|$的大小确是$X$与$Y$间线性关系强弱的一种度量。

设随机变量$X$与$Y$的相关系数$\rho_{XY}$存在。若$X$与$Y$相互独立，则有$\operatorname{cov}(X,Y)=0$，从而$\rho_{XY}=0$，即若$X$与$Y$相互独立，则$X$与$Y$不相关。反之，若$X$与$Y$不相关，则$X$与$Y$却不一定是相互独立的。这说明“不相关”与“相互独立”是两个不同的概念，其含义是不同的，不相关只是就线性关系而言的，而相互独立是就一般关系而言的。

**例 3** 若$X\sim N(0,1)$，且$Y=X^2$，问$X$与$Y$是否不相关？

**解** 因为$X\sim N(0,1)$，密度函数$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$为偶函数，所以$E(X)=E(X^3)=0$。

于是由$\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)=E(X^3)-E(X)E(X^2)=0$

得$\rho_{XY}=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=0$.

这说明$X$与$Y$是不相关的，但$Y=X^2$，显然，$X$与$Y$是不相互独立的。

**例 4** 设$(X,Y)$服从二维正态分布，即$(X,Y)\sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho)$ ，求$\rho_{XY}$.

**解** 有 3.1 中的正态分布，有$X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)$即有

$$E(X)=\mu_1,D(X)=\sigma_1^2,E(Y)=\mu_2,D(Y)=\sigma_2^2.$$

而

$$\operatorname{cov}(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)f(x,y)dxdy$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)\cdot e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]}dxdy$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}(x-\mu_1)e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}dx\cdot\int_{-\infty}^{+\infty}(y-\mu_2)e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)}dy,$$

令$t=\dfrac{1}{\sqrt{1-\rho^2}}\left(\dfrac{y-\mu_2}{\sigma_2}-\rho\dfrac{x-\mu_1}{\sigma_1}\right),u=\dfrac{x-\mu_1}{\sigma_1}$,

则有

$$\operatorname{cov}(X,Y)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(\sigma_1\sigma_2\sqrt{1-\rho^2}tu+\rho\sigma_1\sigma_2e^{-\frac{u^2}{2}-\frac{t^2}{2}})dtdx=$$

$$=\frac{\rho\sigma_1\sigma_2}{2\pi}(\int_{-\infty}^{+\infty}\sigma_1\sigma_2e^{-\frac{u^2}{2}}du)(\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt)+\frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{2\pi}(\int_{-\infty}^{+\infty}ue^{-\frac{u^2}{2}}du)(\int_{-\infty}^{+\infty}te^{-\frac{t^2}{2}}dt)$$

$$=\frac{\rho\sigma_1\sigma_2}{2\pi}\sqrt{2\pi}\cdot\sqrt{2\pi}=\rho\sigma_1\sigma_2,$$

于是
$$\rho_{XY}=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\rho.$$

可见二维正态随机变量(X,Y)的密度函数中的参数$\rho$就是$X$与$Y$的相关系数，因此，二维正态随机变量的分布完全可由每个变量的数学期望$\mu_1,\mu_2$，方差$\sigma_1^2,\sigma_2^2$及相关系数$\rho$确定。

大家已经知道，对二维正态随机变量（$X$，$Y$）来说，$X$与$Y$相互独立的重要条件为$\rho=0$。现在又知$\rho_{XY}=\rho$，故对二维正态随机变量（X，Y）来说，$X$与$Y$不相关和$X$与$Y$相互独立是等价的。

## 4.4.2 矩

数学期望、方差与协方差都是随机变量常用的数字特征，实际上它们都是某种矩。下面给出矩的一般定义。

**定义 4.5** 设$X$与$Y$是随机变量。若$E(X^k)(k=1,2,\cdots)$存在，则称它为$X$的 ***k*** **阶原点矩**.

若 $E[X-E(X)]^k (k=1,2,\cdots)$ 存在，则称它为 $X$ 的 **$k$ 阶中心矩**.

若 $E(X^k Y^l)(k,l=1,2,\cdots)$ 存在，则称它为 $X$ 与 $Y$ 的 **$k$+l 阶混合矩**.

若 $E[X-E(X)]^k [Y-E(Y)]^l (k=1,2,\cdots)$ 存在，则称它为 $X$ 与 $Y$ 的 **$k$+l 阶混合中心矩**.

由定义可知，随机变量 $X$ 的数学期望 $E(X)$ 是 $X$ 的一阶原点矩，方差 $D(X)$ 是 $X$ 的二阶中心矩，协方差 cov ($X$，$Y$)是 $X$ 与 $Y$ 的 1+1 阶混合中心矩.

## 4.4 大数定律和中心极限定理

### 1. 切贝雪夫不等式

对于任何具有有限方差的随机变量 $X$ 都有

$$P\{|x-E(X)|\geq \varepsilon\} \leq \frac{D(X)}{\varepsilon^2}$$

其中 $\varepsilon$ 是任一正数。

**证** 设 $F(x)$ 是 $X$ 的分布函数,则显然有

$$P\{|x-E(X)|\geq \varepsilon = \int_{|x-E(X)|\geq\varepsilon} dF(x) \leq \int_{|x-E(X)|\geq\varepsilon} \frac{[x-E(X)]^2}{\varepsilon^2} dF(x)$$
$$\leq \frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}[x-E(X)]^2 dF(x) = \frac{D(X)}{\varepsilon^2}$$

切贝雪夫不等式也可以表示成 $P\{|X-E(X)|<\varepsilon\} > 1-\frac{D(X)}{\varepsilon^2}$。由于切贝雪夫不等式只利用随机变量的数学期望 $E(X)$ 及方差 $D(X)$ 就可对 $X$ 的概率分布进行估计，因此它在理论研究及实际应用中有价值。从切贝雪夫不等式还可以看出，当方差越小时，事件 $\{|X-E(X)|\geq\varepsilon\}$ 发生的概率也越小，从而可知，方差确实是一个描述随机变量与其期望值离散程度的一个量。

**例 1** 设电站供电网有 10000 盏电灯，夜晚每盏灯开灯的概率均为 0.7，假定灯的开、关是相互立的，使用切贝雪夫不等式估计夜晚同时开着的灯数在 6800 到 7200 盏之间的概率。

**解** 令 $X$ 表示在夜晚同时开着的灯数目，则 $X$ 服从 $n$=10000，$p$=0.7 的二项分布，这时 $E(X)=npq=2100$，由切贝雪夫不等式可得

$$P\{6800<X<7200\} = P\{|X-7000|<200\} \geq 1-\frac{2100}{200^2} \approx 0.95$$

事实上，这个概率的近似值表明，在 10000 盏灯中，开着的灯数在 6800 到 7200 的概率大于 0.95。而实际此概率的精确值可由贝努里公式求得为 0.99999。由此可知，切贝雪夫不等式虽可用来估计概率，但精度不够高，它的重要意义是在理论上的应用，在

大数定律的证明中，用切贝雪夫不等式可使证明非常简洁。

## 2.大数定律

## 贝努里大定理

设 $\mu_n$ 是 $n$ 重贝奴里试验中事件 $A$ 出现的次数，而 $p$ 是事件 $A$ 在每次试验中出现的概率，则对任意 $\varepsilon>0$，都有

$$\lim_{n\to\infty} P\{|\frac{\mu_n}{n}-p|<\varepsilon\}=1$$

$\mu_n$ 是 $n$ 重贝努里试验中 $A$ 出现的次数,则 $\frac{\mu_n}{n}$ 便是这 $n$ 次试验中 $A$ 出现的频率， 上式表明，当次数 $n$ 很大时，事件 $A$ 出现的频率与事件 $A$ 出现的概率 $p$ 的偏差超过任意正数ε的可能性很小，或者基本上说 ，是不可能的。也就是说，要从理论上证明：对于任意的ε>0，有 $\lim_{n\to\infty} P\{\frac{\mu_n}{n}-p|\ge\varepsilon\}=0$ 它等价于 $\lim_{n\to\infty} P\{|\frac{\mu_n}{n}-p|<\varepsilon\}=1$ 。

贝努里大数定律证明了在大量重复实验时，随机事件的频率在它的概率的附近摆动，若事件 $A$ 的概率很小，则正如贝努里定律所指出的，事件 $A$ 的频率也很小，或者说事件 $A$ 很少发生。

“概率很小的随机事件在个别试验中是几乎不能发生的”这一原理称为小概率事件的实际不可能性原理。它在国家经济建设中有着广泛的应用。至于“小概率”小到什么程度才能看作实际上不可能发生，则要视具体情况的要求和性质而定。例如，自动车床加工零件出现次品的概率为 0.01，若零件的重要性不大而价格又低，则完全可允许有 1%的次品率，即可忽视 100 个零件中出现一个次品的可能性。但如果制造一批降落伞出现的次品的概率为 0.01，显然在这种情况下，这 1%的忽视也是绝对不允许的，因为它可能危及这百分之一跳伞者的生命。

贝努里大数定律还提供了通过试验来确定事件概率的方法。既然频率 $\frac{\mu_n}{n}$ 与概率 $p$ 有较大偏差的可能性很小，那么我们就可以通过做试验确定某事件发生的频率并把它作为相应概率的估计，这种方法称为参数估计，它是数理统计中主要的研究课题之一。参数估计的一个重要理论基础就是大数定律。

设 $Y_1,Y_2,....,Y_n,....$ 是一个互相独立的随机变量序列，$a$ 是一个常数，若对于任意正数ε，有 $\lim_{n\to\infty} P\{|Y_{n-a}|<\varepsilon\}=1$，则称序列 $Y_1,Y_2,....,Y_n,....$ **依概率收敛于 *a***。

因此，由贝奴里大数定律可得：设 $\mu_n$ 是 $n$ 次独立试验中事件 $A$ 出现的次数，而 $p$ 是事件 $A$ 在每次试验中出现的概率,则频率 $\frac{\mu_n}{n}$ 依概率收敛于概率 $p$ 。

人们在事件中还发现，除了频率具有稳定性之外，大量观察值的平均值也具有稳定

性。这就是切贝雪夫大数定律。

## 切贝雪夫大数定律

设随机变量 $X_1, X_2, ...., X_n, ....$ 相互独立，每一随机变量都有数学期望 $E(X_1), E(X_2), ...., E(X_n), ....$ 和有限的方差 $D(X_1), D(X_2), ...., D(X_n), ....$，并且它们有公共上界 $c$,即 $D(X_1) \le c, D(X_2) \le c, ...., D(X_n) \le c, ....$，则对任意的$\varepsilon > 0$，皆有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)| < \varepsilon\} = 1$$

**证** 因 $X_1, X_2, ....$ 相互独立，所以 $D(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) < \frac{1}{n^2} nc = \frac{c}{n}$

又因 $E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$，由切贝雪夫不等式可得

$$P\{|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| < \varepsilon\} \ge 1 - \frac{D(\frac{1}{n}\sum_{i=1}^{n} X_i)}{\varepsilon^2} \ge 1 - \frac{c}{n\varepsilon^2},$$

所以 $$1 \ge P\{|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)| < \varepsilon\} \ge 1 - \frac{c}{n\varepsilon^2},$$

于是, $$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)| < \varepsilon\} = 1$$

在 1866 年由俄国数学家切贝雪夫证明的大数定律是关于大数定律大的一个相当普遍的结论。贝努里大数定律就是切贝雪夫大数定律的一个特例。

切贝雪夫大数定律表明相互独立的随机变量的算术平均值 $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 与数学期望的算术平均值的差在$n$充分大时是一个无穷小量，这也意味着在$n$从分大时，经算术平均后得到的随机变量$\overline{X_n}$的值将比较紧密地聚集在它的数学期望$E(\overline{X_n})$的附近。

有切贝雪夫大数定律还得益的下面的推论：

设独立随机变量 $X_1, X_2, ...., X_n, ....$服从同一分布，并且有数学期望 $a$ 及方差$\sigma^2$，则 $X_1, X_2, ...., X_n$ 的算术平均值$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 在$n \to \infty$时，依概率收敛与数学期望$a$，即对任意正数$\varepsilon$，有$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^{n} X_i - a| < \varepsilon\} = 1$.

上述推论，是我们关于算术平均值的法则有了理论上的依据。如我们要测量某一物理量$a$，在不便条件下重复进行$n$次，得$m$个测量值$X_1, X_2, ...., X_n$，显然它们可以看成是$n$个相互独立的随机变量，具有相同的分布，并且有数学期望$a$。由大数定理可知，当$n$充分大时，$n$次测量值得平均值可作为$a$得近似值：

$$a \approx \frac{X_1 + X_2 + .... + X_n}{n}$$

则由此所因发的误差是很小的。

## 3. 中心极限定理

由前几章的讨论可知，正态分布在随机变量的一切可能分布中占有特殊的地位。在客观世界中我们遇到的许多随机变量都是服从或近似服从正态分布的。为什么大量的随机变量都服从正态分布呢?李雅普诺夫证明了在某些一般的充分条件下，当随机变量的个数无限增加时，独立随机变量的和的分布是趋于正态分布的。在概率论中把研究大量独立随机变量和的分布以正态分布为极限的这一类定理统称为中心极限定理。在概率论发展史上，许多科学家建立了众多的中心极限定理。在本节中我们只给出两个常用的中心极限定理。因为定理的证明要用到较多的数学知识，所以下面我们只给出结论而不予证明。

### 独立同分布的中心极限定理

设随机变量$X_1, X_2, ...., X_n, ....$相互独立，服从同一分布，具有有限的数学期望和方差：$E(X_i) = \mu, D(X_i) = \sigma^2 \neq 0 (i = 1,2,...)$，则随机变量$Y_n = \dfrac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$的分布函数$F_n(x)$对任意的$x \in (-\infty, +\infty)$，都有

$$\lim_{n \to \infty} F_n(x) = \lim_{n \leftarrow \infty} P\left\{ \frac{\sum_{k=1}^{n} \left( X_k - \mu \right)}{\sqrt{n}\sigma} \leq x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$$

这个定理告诉我们，当$n$很大是，$Y_n$近似地服从标准正态分布$N(0,1)$。随机变量$\sum_{k=1}^{n} X_k = \sqrt{n}\sigma Y_n + n\mu$近似地服从正态分布$N(n\mu, n\sigma^2)$。由于期望$n\mu = E(\sum_{k=1}^{n} X_k)$，方差$\left(\sqrt{n}\sigma\right)^2 = D(\sum_{k=1}^{n} X_k)$，故$Y_n$实际上就是$\sum_{k=1}^{n} X_k$的标准化的随机变量。

中心极限定理可以解释如下：假设被研究的随机变量可以表示为大量独立的随变量

的和，其中每一个随机变量对于总和的作用都很微小，则可以认为这个随机变量实际上是服从正态分布的。在实际工作中，只要 $n$ 足够大，便可把独立同分布的随机变量之和近似当作正态变量。

**隶美弗—拉普拉斯(DeMoivre-Laplace)定理**

设随机变量 $\eta_n(n=1,2,\cdots)$ 服从参数为 $n,p(0<p<1)$ 的二项分布，则对于任一区间 $(a,b)$，恒有 $\lim\limits_{n\leftarrow\infty} P\left\{a<\dfrac{\eta_n-np}{\sqrt{np(1-p)}}\leq b\right\}=\int_a^b\dfrac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$。

**证** 由于服从二项分布的随机变量 $\eta_n$ 可视为 $n$ 个相互独立的、服从同一的参数 $p$ 的（0-1）分布的随机变量 $X_1,X_2,....,X_n$ 之和，即 $\eta_n=\sum\limits_{i=1}^n X_i$，其中 $E(X_k)=p,D(X_k)=pq,k=1,2,...,n,q=1-p$。由独立同分布中心极限定理可得

$$\lim_{n\leftarrow\infty} P\left\{\frac{\eta_n-np}{\sqrt{np(1-p)}}<x\right\}=\lim_{n\to\infty} P\left\{\frac{\sum\limits_{k=1}^n X_k-np}{\sqrt{npq}}<x\right\}=\int_a^b\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$

于是对于任意区间 $(a,b)$ 有 $\lim\limits_{n\leftarrow\infty} P\left\{a<\dfrac{\eta_n-np}{\sqrt{np(1-p)}}\leq b\right\}=\int_a^b\dfrac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$

此定理表明，正态分布是二项分布的极限分布。当 $n$ 充分大时，服从二项分布的随机变量 $\eta_n$ 的概率计算可以转化为正态随机变量的概率计算：

$$P\{\eta_n=k\}\approx\frac{1}{\sqrt{2\pi npq}}e^{-\frac{(K-np)^2}{2npq}};$$

$$P\{a<\eta_n\leq b\}=P\left\{\frac{a-np}{\sqrt{npq}}<\frac{\eta_n-np}{\sqrt{npq}}\leq\frac{b-np}{\sqrt{npq}}\right\}\approx\phi\left(\frac{b-np}{\sqrt{npq}}\right)-\phi\left(\frac{a-np}{\sqrt{npq}}\right)$$

由于当 $n$ 较大，$p$ 又较小时，二项式分布的计算十分麻烦，因此，若用上面的近似公式计算将是非常简洁的。

**例 2** 某运输公司有 500 辆汽车参加保险，在一年里汽车初事故的概率为 0.006，参加保险的汽车每年交 800 元的保险费。若出事故，保险公司最多赔偿 500000 元，试利用中心极限定理计算，保险公司一年赚钱不小于 200000 元的概率？

**解** 设 $x$ 表示 500 辆汽车中出事故的车辆数，则 $X$ 服从 $n=500,p=0.006$ 的二项分布，这时 $np=500\times0.006=3,npq=3\times0.994=2.982$。保险公司一年赚钱不小于 200000

元的事件为 $\{500\times800\geq500\times800-50000X>200000\}$，即事件 $\{0\leq X\leq4\}$，从而有

$$P\{0<X\leq4\}=P\left\{\frac{0-3}{\sqrt{2.982}}\leq\frac{X-3}{\sqrt{2.982}}\leq\frac{4-3}{\sqrt{2.982}}\right\}$$

$$\approx\phi\left(\frac{1}{\sqrt{2.982}}\right)-\phi\left(\frac{-3}{\sqrt{2.982}}\right)=\phi(0.579)-\phi(-1.737)$$

$$=0.7190+0.9591-1=0.7781$$

可见，保险公司在一年里赚钱不小于 200000 元的概率为 0.7781.

**例 3** 现有一大批种子，其中良种占 1/6，今从其中任意选 6000 粒，试问在这些种子中，良种所占的比例与 1/6 之差小于 1%的概率是多少?

**解** 选一粒良种看成是一次随机试验，因此选 6000 粒种子看作是 6000 重贝奴里试验。若令 $X$ 表示 6000 粒种子中的良种数，则 $X$ 服从 $n=6000, p=1/6$ 的二项分布，故由拉普拉斯积分定理可得

$$P\left\{\left|\frac{X}{6000}-\frac{1}{6}\right|<0.01\right\}=P\left\{\frac{\left|X-6000\times\frac{1}{6}\right|}{\sqrt{6000\times\frac{1}{6}\times\frac{5}{6}}}\leq\frac{0.01\times\sqrt{6000}}{\sqrt{\frac{1}{6}\times\frac{5}{6}}}\right\}$$

$$\approx\phi(2.078)-\phi(-2.078)=2\phi(2.078)-1=0.9624$$

本例是积分极限定理的应用之一。在用频率估计概率时对误差的一种估计，可用公式表示如下：

$$P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}=P\{|\mu_{n-np}|<n\varepsilon\}=P\left\{\frac{|\mu_n-p|}{\sqrt{npq}}\leq\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right\}$$

$$\underline{\underline{n充分大}}\phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)-\phi\left(-\varepsilon\sqrt{\frac{n}{pq}}\right)=2\phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)-1$$

此公式可用来计算许多概率问题。

**例 4** 设某集成电路出厂时一级品率为 0.7，装配一台仪器需要 100 只一级品集成电路，问购置多少只才能以 99.9%的概率保证装该仪器是够用(不能因一级品不够而影响工作)。

**解** 设购置 $n$ 只，并用随机变量 $X$ 表示 $n$ 只中非一级品的只数；现要求购置的 $n$ 之集成电路中一级数不少于 100 只，亦即非一级品数 $X\leq n-100$ 的概率 $P\{X\leq n-100\}\geq99.9\%$。由有题意知，非一级品率为 0.3，则

$$P\{X\leq n-100\}=\sum_{k=0}^{n-100}C_n^k0.3^k0.7^{n-k}\approx\phi\left(\frac{n-100-0.3n}{\sqrt{n\bullet0.3\times0.7}}\right)=\phi\left(\frac{0.7n-100}{\sqrt{0.21n}}\right)\geq0.999$$

查表得 $\frac{0.7n-100}{\sqrt{0.21n}}=3.090$，即 $0.49n^2141.89n+1000=0$，解之得 $n=168$，即至少要

购置 168 只集成电路。

**例 5** 独立地多次测量一个物理量，每次测量产生的随机误差，都服从(-1,1)内的均匀分布。

(1) 若取 $n$ 次测量的算术平均值作为测量结果，求它与真值得差小于正数$\varepsilon$ 的概率;

(2) 计算当 n=36, $\varepsilon=\dfrac{1}{6}$时的概率得近似值;

(3) 要使上述概率不小于$\alpha=0.95$，应进行多少次测量?

**解** (1) 设 $X_i$ 表示第 $i$ 次测量值, $\varepsilon_i$ 表示第 $i$ 次测量产生的随机误差$(i=1,2,\cdots,n)$, $\mu$ 表示所测物理量的真值，则 $X_i=\mu+\varepsilon_i$ 由题设，$\varepsilon_i$ 在(-1,1)内服从均匀分布，所以

$$E(\varepsilon_i)=\frac{-1+1}{2}=0,\qquad D(\varepsilon_i)=\frac{[(1-(-1)]}{12}=\frac{1}{3},\qquad E(\varepsilon_i)=E(\mu+\varepsilon_i)=\mu,$$

$$D(\varepsilon_i)=D(\mu+\varepsilon_i)=\frac{1}{3},i=1,2,...,n.$$

由题设条件知, $X_1,X_2,....,X_n$ 独立同分布，而 $E(\varepsilon_i)=\mu,\ D(\varepsilon_i)=\dfrac{1}{3}(i=1,2,...,n).$ 所以，当 $n$ 很大时，由独立同分布中心极限定理可知，随机变量 $\dfrac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{\dfrac{n}{3}}}$ 近似服从标准正态分布。于是，所求概率为

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|<\varepsilon\right\}=P\left\{\left|\sum_{i=1}^{n}X_i-n\mu\right|<n\varepsilon\right\}$$

$$=P\left\{\left|\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{\dfrac{n}{3}}}\right|<\varepsilon\bullet\sqrt{3n}\right\}\approx2\phi\left(\varepsilon\bullet\sqrt{3n}\right)-1.$$

(2) 当 $n=36,\varepsilon=\dfrac{1}{6}$时，所求概率为

$$P\left\{\left|\frac{1}{36}\sum_{i=1}^{36}X_i-\mu\right|<\frac{1}{6}\right\}\approx2\phi\left(\frac{1}{6}\sqrt{3\times36}\right)-1$$

$$=2\phi\left(\sqrt{3}\right)-1=2\phi(1.73)-1=0.92.$$

(3) 由题意可知，现在要求$n$的值，使概率

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|<\varepsilon\right\}=2\phi\left(\varepsilon\sqrt{3n}\right)-1\geq\alpha,$$

令 $\varepsilon\sqrt{3n}=x$，使 $\phi(x)\geq\dfrac{1+\alpha}{2}=\dfrac{1+0.95}{2}=0.975$，反查标准正态分布表得 $x=1.96$，因此

$\varepsilon\sqrt{3n} > 1.96$，即 $n > \dfrac{(1.96)^2}{3\varepsilon^2} = \dfrac{(1.96)^2}{3\times\dfrac{1}{36}} \approx 46$。由此可知，当 $\varepsilon = \dfrac{1}{6}$ 时，要使概率不小于 0.95，至少需增加 10 次测量。