第四章：随机变量的数字特征[1]

Xiao Yuan[1]

[1]Center on Frontiers of Computing Studies, Peking University, Beijing 100871, China
xiaoyuan@pku.edu.cn

March 22, 2024

[1]讲义内容基于三本参考教材以及网络素材

目录

在一些实际问题中，我们需要了解随机变量的分布函数外，更关心的是随机变量的某些特征

数学期望

在一些实际问题中，我们需要了解随机变量的分布函数外，更关心的是随机变量的某些特征

- 在评定某地区粮食产量的水平时，最关心的是平均产量;
- 在检查一批棉花的质量时，既需要注意纤维的平均长度，又需要注意纤维长度与平均长度的偏离程度;
- 考察杭州市区居民的家庭收入情况，我们既知家庭的年平均收入，又要研究贫富之间的差异程度。

离散型随机变量的数学期望

离散型随机变量的数学期望

离散型随机变量 X 的概率分布为 $p_i = P(X = x_i)$，若 x_i 取值有限或 $\sum_i x_i p_i$ 绝对收敛（满足 $\sum_{i=1}^{\infty} |x_i p_i| < \infty$），则随机变量 X 的数学期望或均值为 $E(X) = \sum_i x_i p_i$

离散型随机变量的数学期望

离散型随机变量 X 的概率分布为 $p_i = P(X = x_i)$，若 x_i 取值有限或 $\sum_i x_i p_i$ 绝对收敛（满足 $\sum_{i=1}^{\infty} |x_i p_i| < \infty$），则随机变量 X 的数学期望或均值为 $E(X) = \sum_i x_i p_i$

- 例：考虑 $p_i = P(X = x_i = 3^i/i) = 2/3^i, i = 1, 2, \ldots$，则 $\sum_{i=1}^{\infty} |x_i p_i| = \sum_{i=1}^{\infty} 2/i$ 发散，因此不存在期望

离散型随机变量的数学期望

离散型随机变量 X 的概率分布为 $p_i = P(X = x_i)$，若 x_i 取值有限或 $\sum_i x_i p_i$ 绝对收敛（满足 $\sum_{i=1}^{\infty} |x_i p_i| < \infty$），则随机变量 X 的数学期望或均值为 $E(X) = \sum_i x_i p_i$

- 例：考虑 $p_i = P(X = x_i = 3^i/i) = 2/3^i, i = 1, 2, \ldots,$ 则 $\sum_{i=1}^{\infty} |x_i p_i| = \sum_{i=1}^{\infty} 2/i$ 发散，因此不存在期望
- 已知 $X \sim B(n, p)$，则 $E(X) = \sum_{i=0}^{n} i\binom{n}{i} p^i (1-p)^{n-i} = np \sum_{i=1}^{n} \binom{n-1}{i-1} p^{i-1} (1 - p)^{n-i} = np \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} = np$
（这里用到了 $i\binom{n}{i} = n\binom{n-1}{i-1}$）

例子

- 已知 $X \sim \pi(\lambda)$，则
  $E(X) = \sum_{i=0}^{\infty} ie^{-\lambda}\lambda^i/i! = \lambda \sum_{i=1}^{\infty} e^{-\lambda}\lambda^{i-1}/(i-1)! = \lambda$

例子

- 已知 $X \sim \pi(\lambda)$，则 $E(X) = \sum_{i=0}^{\infty} ie^{-\lambda}\lambda^i/i! = \lambda \sum_{i=1}^{\infty} e^{-\lambda}\lambda^{i-1}/(i-1)! = \lambda$
- 已知 $X \sim NB(r,p)$，则

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} k \binom{k+r-1}{r-1} p^r (1-p)^k \\
&= \left[\sum_{k=0}^{\infty} (k+r) \binom{k+r-1}{r-1} p^r (1-p)^k\right] - r \\
&= r\left[\sum_{k=0}^{\infty} \binom{k+r}{r} p^r (1-p)^k\right] - r \\
&= r/p \left[\sum_{k=0}^{\infty} \binom{k+r}{r} p^{r+1} (1-p)^k\right] - r \\
&= r(1-p)/p
\end{aligned}
$$

例子

- 已知 X 是取值非负的离散型随机变量，则

$$E(X)=\sum_{n=1}^{+\infty}P(X\geq n)=\sum_{n=0}^{+\infty}P(X>n)$$

例子

- 已知 X 是取值非负的离散型随机变量，则

$$E(X) = \sum_{n=1}^{+\infty} P(X \geq n) = \sum_{n=0}^{+\infty} P(X > n)$$

设 $P(X = k) = p_k$，由定义：

$$E(X) = \sum_{k=1}^{+\infty} kp_k = \sum_{k=1}^{+\infty} \sum_{n=1}^{k} p_k = \sum_{n=1}^{+\infty} \sum_{k=n}^{+\infty} p_k$$

$$= \sum_{n=1}^{+\infty} P(X \geq n) = \sum_{n=0}^{+\infty} P(X > n).$$

- 已知 X 是取值非负的离散型随机变量，则

$$E(X) = \sum_{n=1}^{+\infty} P(X \geq n) = \sum_{n=0}^{+\infty} P(X > n)$$

设 $P(X = k) = p_k$，由定义：

$$E(X) = \sum_{k=1}^{+\infty} kp_k = \sum_{k=1}^{+\infty} \sum_{n=1}^{k} p_k = \sum_{n=1}^{+\infty} \sum_{k=n}^{+\infty} p_k$$

$$= \sum_{n=1}^{+\infty} P(X \geq n) = \sum_{n=0}^{+\infty} P(X > n).$$

- 期望和概率的关系：设 A 是一个事件，称随机变量 1_A 为事件 A 的示性函数，它在 A 发生时取 1，否则取 0，则 $E(1_A) = P(A)$.

连续型随机变量的数学期望

连续型随机变量的数学期望

连续型随机变量 X 的概率密度为 $f(x)$，若 $\int_{-\infty}^{\infty} xf(x)dx$ 绝对收敛（满足 $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$），则随机变量 X 的数学期望或均值为 $E(X) = \int_{-\infty}^{\infty} xf(x)dx$

连续型随机变量的数学期望

连续型随机变量 X 的概率密度为 $f(x)$，若 $\int_{-\infty}^{\infty} xf(x)dx$ 绝对收敛（满足 $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$），则随机变量 X 的数学期望或均值为 $E(X) = \int_{-\infty}^{\infty} xf(x)dx$

- 例：已知 $X \sim U(a,b)$，则
  $E(X) = \int_a^b x/(b-a)dx = (a+b)/2$

连续型随机变量的数学期望

连续型随机变量 X 的概率密度为 $f(x)$，若 $\int_{-\infty}^{\infty} xf(x)dx$ 绝对收敛（满足 $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$），则随机变量 X 的数学期望或均值为 $E(X) = \int_{-\infty}^{\infty} xf(x)dx$

- 例： 已知 $X \sim U(a, b)$，则
  $E(X) = \int_a^b x/(b-a)dx = (a+b)/2$
- 已知 $X \sim \mathrm{Exp}(\lambda)$，则 $E(X) = \int_0^{\infty} \lambda e^{-\lambda x} x dx = 1/\lambda$

连续型随机变量的数学期望

连续型随机变量 X 的概率密度为 $f(x)$，若 $\int_{-\infty}^{\infty} xf(x)dx$ 绝对收敛（满足 $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$），则随机变量 X 的数学期望或均值为 $E(X) = \int_{-\infty}^{\infty} xf(x)dx$

- 例：已知 $X \sim U(a,b)$，则 $E(X) = \int_a^b x/(b-a)dx = (a+b)/2$
- 已知 $X \sim \mathrm{Exp}(\lambda)$，则 $E(X) = \int_0^\infty \lambda e^{-\lambda x} x dx = 1/\lambda$
- 已知 $X \sim \Gamma(\alpha,\lambda)$，则 $E(X) = \int_0^\infty \frac{x^\alpha \lambda^\alpha e^{-\lambda x}}{\Gamma(\alpha)}dx = \alpha/\lambda$

连续型随机变量的数学期望

连续型随机变量 X 的概率密度为 $f(x)$，若 $\int_{-\infty}^{\infty} xf(x)dx$ 绝对收敛（满足 $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$），则随机变量 X 的数学期望或均值为 $E(X) = \int_{-\infty}^{\infty} xf(x)dx$

- 例：已知 $X \sim U(a, b)$，则 $E(X) = \int_a^b x/(b-a)dx = (a+b)/2$
- 已知 $X \sim \mathrm{Exp}(\lambda)$，则 $E(X) = \int_0^{\infty} \lambda e^{-\lambda x} x dx = 1/\lambda$
- 已知 $X \sim \Gamma(\alpha, \lambda)$，则 $E(X) = \int_0^{\infty} \frac{x^{\alpha}\lambda^{\alpha}e^{-\lambda x}}{\Gamma(\alpha)} dx = \alpha/\lambda$
- 已知 $X \sim N(\mu, \sigma^2)$，则 $E(X) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu$

连续型随机变量的数学期望

连续型随机变量的数学期望

连续型随机变量 X 的概率密度为 $f(x)$，若 $\int_{-\infty}^{\infty} xf(x)dx$ 绝对收敛（满足 $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$），则随机变量 X 的数学期望或均值为 $E(X) = \int_{-\infty}^{\infty} xf(x)dx$

若 $\int_{-\infty}^{\infty} xf(x)dx$ 不绝对收敛，则称 X 的期望不存在.

连续型随机变量的数学期望

连续型随机变量 X 的概率密度为 $f(x)$，若 $\int_{-\infty}^{\infty} xf(x)dx$ 绝对收敛（满足 $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$），则随机变量 X 的数学期望或均值为 $E(X) = \int_{-\infty}^{\infty} xf(x)dx$

若 $\int_{-\infty}^{\infty} xf(x)dx$ 不绝对收敛，则称 X 的期望不存在.

- 例：已知 X 的密度 $f(x) = \frac{1}{\pi(x^2+1)}$（柯西分布），则

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

X 的期望不存在.

期望和分布函数的关系

期望和分布函数的关系

已知随机变量 X 的分布函数为 $F(x)$，则有
$E(X) = \int_{-\infty}^{\infty} x dF(x) = \int_0^{\infty}(1 - F(x))dx - \int_{-\infty}^{0} F(x)dx.$
当 X 取值非负时，有 $E(X) = \int_0^{\infty}(1 - F(x))dx$

期望和分布函数的关系

已知随机变量 X 的分布函数为 $F(x)$，则有
$E(X)=\int_{-\infty}^{\infty}xdF(x)=\int_0^{\infty}(1-F(x))dx-\int_{-\infty}^{0}F(x)dx.$
当 X 取值非负时，有 $E(X)=\int_0^{\infty}(1-F(x))dx$

- $\int_0^{\infty}xdF(x)=\int_0^{\infty}f(x)dx\int_0^{x}dy=\int_0^{\infty}dy\int_y^{\infty}f(x)dx=\int_0^{\infty}(1-F(y))dy$

期望和分布函数的关系

已知随机变量 X 的分布函数为 $F(x)$，则有
$E(X) = \int_{-\infty}^{\infty} x dF(x) = \int_0^{\infty} (1 - F(x)) dx - \int_{-\infty}^{0} F(x) dx.$
当 X 取值非负时，有 $E(X) = \int_0^{\infty} (1 - F(x)) dx$

- $\int_0^{\infty} x dF(x) = \int_0^{\infty} f(x) dx \int_0^{x} dy = \int_0^{\infty} dy \int_y^{\infty} f(x) dx = \int_0^{\infty} (1 - F(y)) dy$
- 例：已知 $X \sim U(-1, 2)$，$Y = \max\{X, 0\}$，求 $E(Y)$

期望和分布函数的关系

已知随机变量 X 的分布函数为 $F(x)$，则有
$E(X) = \int_{-\infty}^{\infty} xdF(x) = \int_0^{\infty}(1-F(x))dx - \int_{-\infty}^{0} F(x)dx.$
当 X 取值非负时，有 $E(X) = \int_0^{\infty}(1-F(x))dx$

- $\int_0^{\infty} xdF(x) = \int_0^{\infty} f(x)dx \int_0^{x} dy = \int_0^{\infty} dy \int_y^{\infty} f(x)dx = \int_0^{\infty}(1-F(y))dy$
- 例：已知 $X \sim U(-1,2)$，$Y = \max\{X, 0\}$，求 $E(Y)$

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ (y+1)/3 & 0 < y < 2 \\ 1 & y \geq 2 \end{cases}$$

首先 Y 非离散型也非连续型。则
$E(Y) = \int_0^{\infty}(1-F(x))dx = \int_0^{1}[1-(y+1)/3]dx = 2/3$

离散型随机变量函数的期望

离散型随机变量 X 的概率分布为 $p_i = P(X = x_i)$，g 为实单指函数，若 x_i 有限或 $\sum_i |g(x_i)| p_i < \infty$，则随机变量 $Y = g(X)$ 的数学期望或均值为 $E(Y) = E(g(X)) = \sum_i g(x_i) p_i$

离散型随机变量函数的期望

离散型随机变量 X 的概率分布为 $p_i = P(X = x_i)$，g 为实单指函数，若 x_i 有限或 $\sum_i |g(x_i)| p_i < \infty$，则随机变量 $Y = g(X)$ 的数学期望或均值为 $E(Y) = E(g(X)) = \sum_i g(x_i) p_i$

- 定理的意义在于不用先求 Y 的分布再算均值

离散型随机变量函数的期望

离散型随机变量 X 的概率分布为 $p_i = P(X = x_i)$，g 为实单指函数，若 x_i 有限或 $\sum_i |g(x_i)|p_i < \infty$，则随机变量 $Y = g(X)$ 的数学期望或均值为 $E(Y) = E(g(X)) = \sum_i g(x_i)p_i$

- 定理的意义在于不用先求 Y 的分布再算均值

证明：$P(Y = y_j) = \sum_{g(x_i)=y_j} p_i$，则

$$
\begin{aligned}
E(Y) &= \sum_j P(Y = y_j)y_j = \sum_j \sum_{g(x_i)=y_j} p_i g(x_i) \\
&= \sum_i g(x_i)p_i
\end{aligned} \tag{1}
$$

连续型随机变量函数的期望

连续型随机变量 X 的概率密度为 $f_X(x)$，g 为实单指函数，若 $\int_{-\infty}^{\infty}|g(x)|f_X(x)dx<\infty$，则随机变量 $Y=g(X)$ 的数学期望或均值为 $E(Y)=E(g(X))=\int_{-\infty}^{\infty}g(x)f_X(x)dx$

连续型随机变量函数的期望

连续型随机变量函数的期望

连续型随机变量 X 的概率密度为 $f_X(x)$，g 为实单指函数，若 $\int_{-\infty}^{\infty}|g(x)|f_X(x)dx<\infty$，则随机变量 $Y=g(X)$ 的数学期望或均值为 $E(Y)=E(g(X))=\int_{-\infty}^{\infty}g(x)f_X(x)dx$

- 定理的意义在于不用先求 Y 的分布再算均值

连续型随机变量函数的期望

连续型随机变量 X 的概率密度为 $f_X(x)$，g 为实单指函数，若 $\int_{-\infty}^{\infty}|g(x)|f_X(x)dx<\infty$，则随机变量 $Y=g(X)$ 的数学期望或均值为 $E(Y)=E(g(X))=\int_{-\infty}^{\infty}g(x)f_X(x)dx$

- 定理的意义在于不用先求 Y 的分布再算均值

证明：

$$
\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} y dF_Y(y) \\
&= \int_{-\infty}^{\infty} y f_Y(y) dy \\
&= \int_{-\infty}^{\infty} g(x) f_X(x) dx
\end{aligned}
\tag{2}
$$

随机向量函数的期望

离散随机向量函数的期望

随机变量 $Z = g(X, Y)$ 是离散型随机变量 X, Y 的函数。X, Y 的概率分布为 $p_{ij} = P(X = x_i, Y = y_j)$，$g$ 为实单指函数，若 x_i, y_j 有限或 $\sum_{i,j} |g(x_i, y_j)| p_{i,j} < \infty$，则随机变量 $Z = g(X, Y)$ 的数学期望或均值为 $E(Z) = E(g(X, Y)) = \sum_{i,j} g(x_i, y_j) p_{i,j}$

随机向量函数的期望

PEKING UNIVERSITY

离散随机向量函数的期望

随机变量 $Z=g(X,Y)$ 是离散型随机变量 X,Y 的函数。X,Y 的概率分布为 $p_{ij}=P(X=x_i,Y=y_j)$，g 为实单指函数，若 x_i,y_j 有限或 $\sum_{i,j}|g(x_i,y_j)|p_{i,j}<\infty$，则随机变量 $Z=g(X,Y)$ 的数学期望或均值为 $E(Z)=E(g(X,Y))=\sum_{i,j}g(x_i,y_j)p_{i,j}$

连续随机向量函数的期望

随机变量 $Z=g(X,Y)$ 是连续型随机变量 X,Y 的函数。X,Y 的概率密度为 $f(x,y)$，g 为实单指函数，若 $\int_{-\infty}^{\infty}|g(x,y)|f(x,y)dxdy<\infty$，则随机变量 $Z=g(X,Y)$ 的数学期望或均值为 $E(Z)=E(g(X,Y))=\int_{-\infty}^{\infty}g(x,y)f(x,y)dxdy$

- 已知随机变量 X, Y 的概率分布为

X\Y	Y=0	Y=1	Y=2
X=0	0.1	0.25	0.15
X=1	0.15	0.2	0.15

求 $E(X)$, $E(Y)$, $E(Z)$, 其中 $Z = \sin\frac{\pi(X+Y)}{2}$

- 已知随机变量 X, Y 的概率分布为

X\Y	Y=0	Y=1	Y=2
X=0	0.1	0.25	0.15
X=1	0.15	0.2	0.15

求 $E(X)$, $E(Y)$, $E(Z)$, 其中 $Z = \sin\frac{\pi(X+Y)}{2}$

$E(X) = 0.5$, $E(Y) = 0.45 + 2 * 0.3 = 1.05$,

$E(Z) = \sin\frac{\pi(0+0)}{2} * 0.1 + \sin\frac{\pi(1+0)}{2} * 0.15 + \sin\frac{\pi(0+1)}{2} * 0.25 + \sin\frac{\pi(1+1)}{2} * 0.2 + \sin\frac{\pi(0+2)}{2} * 0.15 + \sin\frac{\pi(1+2)}{2} * 0.15 = 0.25$

- 已知随机变量 X, Y 的概率分布为

X\Y	Y=0	Y=1	Y=2
X=0	0.1	0.25	0.15
X=1	0.15	0.2	0.15

求 $E(X)$, $E(Y)$, $E(Z)$, 其中 $Z = \sin\frac{\pi(X+Y)}{2}$

$E(X) = 0.5,\ E(Y) = 0.45 + 2 * 0.3 = 1.05,$

$E(Z) = \sin\frac{\pi(0+0)}{2} * 0.1 + \sin\frac{\pi(1+0)}{2} * 0.15 + \sin\frac{\pi(0+1)}{2} * 0.25 + \sin\frac{\pi(1+1)}{2} * 0.2 + \sin\frac{\pi(0+2)}{2} * 0.15 + \sin\frac{\pi(1+2)}{2} * 0.15 = 0.25$

- 已知随机变量 X, Y 的概率密度为

$$f(x,y) = \begin{cases} \frac{3}{2x^3y^2} & x > 1, \frac{1}{x} < y < x \\ 0 & else \end{cases}$$ 求 $E(Y), E\left(\frac{1}{XY}\right)$

例子

- 已知随机变量 X, Y 的概率分布为

X\Y	Y=0	Y=1	Y=2
X=0	0.1	0.25	0.15
X=1	0.15	0.2	0.15

求 $E(X)$, $E(Y)$, $E(Z)$, 其中 $Z = \sin\frac{\pi(X+Y)}{2}$

$E(X) = 0.5,\ E(Y) = 0.45 + 2 * 0.3 = 1.05,$

$E(Z) = \sin\frac{\pi(0+0)}{2} * 0.1 + \sin\frac{\pi(1+0)}{2} * 0.15 + \sin\frac{\pi(0+1)}{2} * 0.25 + \sin\frac{\pi(1+1)}{2} * 0.2 + \sin\frac{\pi(0+2)}{2} * 0.15 + \sin\frac{\pi(1+2)}{2} * 0.15 = 0.25$

- 已知随机变量 X, Y 的概率密度为

$$f(x,y) = \begin{cases} \frac{3}{2x^3y^2} & x > 1, \frac{1}{x} < y < x \\ 0 & else \end{cases}$$ 求 $E(Y), E\left(\frac{1}{XY}\right)$

$E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy = 3/4$

$E\left(\frac{1}{XY}\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{xy}f(x,y)dxdy = 3/5$

期望的性质

- 对于常数 C，有 $E(C) = C$

期望的性质

期望的性质

- 对于常数 C，有 $E(C) = C$
- 对于常数 C，随机变量 X，有 $E(CX) = CE(X)$

期望的性质

PEKING UNIVERSITY

期望的性质

- 对于常数 C，有 $E(C) = C$
- 对于常数 C，随机变量 X，有 $E(CX) = CE(X)$
- 对于两个随机变量 X, Y，有 $E(X + Y) = E(X) + E(Y)$

期望的性质

- 对于常数 C，有 $E(C)=C$
- 对于常数 C，随机变量 X，有 $E(CX)=CE(X)$
- 对于两个随机变量 X,Y，有 $E(X+Y)=E(X)+E(Y)$
- 对于常数 C，对于随机变量 $\{X_i\}$，有 $E(C+\sum_i X_i)=C+\sum_i E(X_i)$

期望的性质

- 对于常数 C，有 $E(C) = C$
- 对于常数 C，随机变量 X，有 $E(CX) = CE(X)$
- 对于两个随机变量 X, Y，有 $E(X + Y) = E(X) + E(Y)$
- 对于常数 C，对于随机变量 $\{X_i\}$，有 $E(C + \sum_i X_i) = C + \sum_i E(X_i)$
- 假设随机变量 $\{X_i\}$ 相互独立，有 $E\left(\prod_i X_i\right) = \prod_i E(X_i)$

例子

- X_i 为 $1, 2, \dots, 5$ 的独立均匀分布（$i = 1, 2, \dots, n$），将 X_i 依次排列，得到十进制数 Y，求 $E(Y)$

- X_i 为 $1, 2, \ldots, 5$ 的独立均匀分布（$i = 1, 2, \ldots, n$），将 X_i 依次排列，得到十进制数 Y，求 $E(Y)$

  $Y = \sum_{i=1}^{n} 10^{i-1} X_i$，$E(X_i) = 3$，因此

  $E(Y) = \sum_{i=1}^{n} 10^{i-1} E(X_i) = \sum_{i=1}^{n} 10^{i-1} 3 = \frac{10^{n-1} - 1}{3}$

例子

PEKING UNIVERSITY

- X_i 为 $1, 2, \ldots, 5$ 的独立均匀分布（$i = 1, 2, \ldots, n$），将 X_i 依次排列，得到十进制数 Y，求 $E(Y)$
  $Y = \sum_{i=1}^{n} 10^{i-1} X_i$，$E(X_i) = 3$，因此
  $E(Y) = \sum_{i=1}^{n} 10^{i-1} E(X_i) = \sum_{i=1}^{n} 10^{i-1} 3 = \frac{10^{n-1} - 1}{3}$
- $X_i \sim U(0, 2i)$ 且相互独立，求 $E(Y)$，其中 $Y = \det(M)$ 和
  $$M = \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}$$

例子

- X_i 为 $1,2,\ldots,5$ 的独立均匀分布（$i=1,2,\ldots,n$），将 X_i 依次排列，得到十进制数 Y，求 $E(Y)$
  $Y=\sum_{i=1}^{n}10^{i-1}X_i$，$E(X_i)=3$，因此
  $E(Y)=\sum_{i=1}^{n}10^{i-1}E(X_i)=\sum_{i=1}^{n}10^{i-1}3=\frac{10^{n-1}-1}{3}$
- $X_i\sim U(0,2i)$ 且相互独立，求 $E(Y)$，其中 $Y=\det(M)$ 和
  $M=\begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}$
  $Y=X_1X_4-X_2X_3$，
  $E(Y)=E(X_1)E(X_4)-E(X_2)E(X_3)=1*4-2*3=-2$

例子

- 已知 $X \sim N(0,1)$，求 $E(X^2)$

PEKING UNIVERSITY

- 已知 $X \sim N(0,1)$，求 $E(X^2)$
  考虑另一个独立随机变量 $Y \sim N(0,1)$，则有
  $E(X^2+Y^2) = \frac{1}{2\pi}\int (x^2+y^2)e^{-(x^2+y^2)/2}dxdy = \frac{1}{2\pi}\int r^3 e^{-r^2/2} dr d\theta = 2$. 因此 $E(X^2) = E(X^2+Y^2)/2 = 1$

例子

- 已知 $X \sim N(0,1)$，求 $E(X^2)$
  考虑另一个独立随机变量 $Y \sim N(0,1)$，则有 $E(X^2+Y^2) = \frac{1}{2\pi}\int(x^2+y^2)e^{-(x^2+y^2)/2}dxdy = \frac{1}{2\pi}\int r^3 e^{-r^2/2}drd\theta = 2$. 因此 $E(X^2) = E(X^2+Y^2)/2 = 1$
- 某建筑商销售一吨水泥获利 a 元，每库存一吨水泥损失 b 元，设水泥的销量 Y 服从 $\mathrm{Exp}(\lambda)$，库存多少水泥的平均利益最大

- 已知 $X \sim N(0,1)$，求 $E(X^2)$
  考虑另一个独立随机变量 $Y \sim N(0,1)$，则有 $E(X^2+Y^2) = \frac{1}{2\pi}\int(x^2+y^2)e^{-(x^2+y^2)/2}dxdy = \frac{1}{2\pi}\int r^3 e^{-r^2/2} dr d\theta = 2$. 因此 $E(X^2) = E(X^2+Y^2)/2 = 1$
- 某建筑商销售一吨水泥获利 a 元，每库存一吨水泥损失 b 元，设水泥的销量 Y 服从 $\mathrm{Exp}(\lambda)$，库存多少水泥的平均利益最大
  设库存 x，则利润为 $Q(x,Y) = \begin{cases} aY - b(x-Y) & Y < x \\ ax & Y \geq x \end{cases}$
  则 $E(Q(x)) = \int_0^x [ay - b(x-y)]\lambda e^{-\lambda y} dy + ax\int_x^\infty \lambda e^{-\lambda y} dy$.
  因此 $dE(Q(x))/dx = (a+b)e^{-\lambda x} - b$，也即是 $x = \lambda^{-1}\ln[(a+b)/b]$ 取极值

例子

- 设袋子里有大小、质地相同的 r 个红球，b 个黑球，不放回地随机抽取 n 个球（$n < r + b$），设抽出的球中有 X 个红球，求 $E(X)$.

例子

- 设袋子里有大小、质地相同的 r 个红球，b 个黑球，不放回地随机抽取 n 个球（$n < r + b$），设抽出的球中有 X 个红球，求 $E(X)$.
  方法一：写出 X 的分布列，用公式计算期望.

例子

PEKING UNIVERSITY

- 设袋子里有大小、质地相同的 r 个红球，b 个黑球，不放回地随机抽取 n 个球（$n < r + b$），设抽出的球中有 X 个红球，求 $E(X)$.
  方法一：写出 X 的分布列，用公式计算期望.
  方法二：运用期望的性质.

- 设袋子里有大小、质地相同的 r 个红球，b 个黑球，不放回地随机抽取 n 个球（$n < r + b$），设抽出的球中有 X 个红球，求 $E(X)$.
  方法一：写出 X 的分布列，用公式计算期望.
  方法二：运用期望的性质.
  定义事件 $A_i = \{$第i次抽出红球$\}$，则
  $X = 1_{A_1} + \cdots + 1_{A_n}$.
  第一章中我们曾计算过 $P(A_i) = \frac{r}{r+b}$，所以 $E(1_{A_i}) = \frac{r}{r+b}$，
  所以 $E(X) = \sum_{i=1}^{n} E(1_{A_i}) = \frac{nr}{r+b}$.

例子

PEKING UNIVERSITY

- 有 n 个信封，每个信封里面分别有一封信．现在将信取出，随机打乱，放回信封，设有 X 个信封中装有原来的信，求 $E(X)$.

例子

- 有 n 个信封，每个信封里面分别有一封信. 现在将信取出，随机打乱，放回信封，设有 X 个信封中装有原来的信，求 $E(X)$.
  定义事件 $A_i = \{第i个信封中装有原来的信\}$，则 $X = 1_{A_1} + \cdots + 1_{A_n}$.
  因为装回信封时信是随机打乱的，所以 $P(A_i) = \frac{1}{n}$，所以 $E(1_{A_i}) = \frac{1}{n}$，所以 $E(X) = \sum_{i=1}^{n} E(1_{A_i}) = 1$.

例子

- 有 n 个信封，每个信封里面分别有一封信．现在将信取出，随机打乱，放回信封，设有 X 个信封中装有原来的信，求 $E(X)$.
  定义事件 $A_i = \{第i个信封中装有原来的信\}$，则 $X = 1_{A_1} + \cdots + 1_{A_n}$.
  因为装回信封时信是随机打乱的，所以 $P(A_i) = \frac{1}{n}$，所以 $E(1_{A_i}) = \frac{1}{n}$，所以 $E(X) = \sum_{i=1}^{n} E(1_{A_i}) = 1$.
- 一民航送客车载有 20 位旅客自机场出发，旅客有 10 个车站可以下车，如到达一个车站没有旅客下车就不停车，以 X 表示停车的次数，求 $E(X)$(设每位旅客在各个车站下车是等可能的，并设各旅客是否下车相互独立)

例子

- 有 n 个信封，每个信封里面分别有一封信. 现在将信取出，随机打乱，放回信封，设有 X 个信封中装有原来的信，求 $E(X)$.
  定义事件 $A_i=\{第i个信封中装有原来的信\}$，则 $X=1_{A_1}+\cdots+1_{A_n}$.
  因为装回信封时信是随机打乱的，所以 $P(A_i)=\frac{1}{n}$，所以 $E(1_{A_i})=\frac{1}{n}$，所以 $E(X)=\sum_{i=1}^{n}E(1_{A_i})=1$.
- 一民航送客车载有 20 位旅客自机场出发，旅客有 10 个车站可以下车，如到达一个车站没有旅客下车就不停车，以 X 表示停车的次数，求 $E(X)$(设每位旅客在各个车站下车是等可能的，并设各旅客是否下车相互独立)
  记事件 A_i 为第 i 站停车，则有 $X=\sum_i 1_{A_i}$，
  $P(A_i)=1-(9/10)^{20}$ 因此
  $E(\sum_i X_i)=\sum_i E(1_{A_i})=10P(A_i)\approx 8.8$

例子

- 给定随机变量 X，若希望寻找 $x \in \mathbb{R}$ 来对 X 的值进行预测，使得"误差"$f(x) = E[(X-x)^2]$ 最小，求 x 和 $f(x)_{\min}$.

- 给定随机变量 X，若希望寻找 $x \in \mathbb{R}$ 来对 X 的值进行预测，使得“误差”$f(x) = E[(X-x)^2]$ 最小，求 x 和 $f(x)_{\min}$.
  不妨设 $E(X) = \mu$，则

  $$f(x) = E[((X-\mu)+(\mu-x))^2] = E[(X-\mu)^2] + (\mu-x)^2.$$

  其中用到 $E[(X-\mu)(\mu-x)] = (\mu-x)E(X-\mu) = 0$.
  所以当 $x = E(X)$ 时 $f(x)$ 最小，且
  $f(x)_{\min} = E[(X-E(X))^2]$.
  可以看出，$x = E(X)$ 在某种意义上是我们对 X 能给出的最佳预测. 这里 $f(x) = E[(X-x)^2]$ 称为均方误差.

定义

方差和标准差

考虑随机变量 X, 已知 $E[(X-E(X))^2]$ 存在，则
$D(X)=\mathrm{Var}(X)=E[(X-E(X))^2]$ 为 X 的方差,
$\sigma(X)=\sqrt{D(X)}$ 为标准差

定义

方差和标准差

考虑随机变量 X, 已知 $E[(X-E(X))^2]$ 存在，则 $D(X)=\mathrm{Var}(X)=E[(X-E(X))^2]$ 为 X 的方差,
$\sigma(X)=\sqrt{D(X)}$ 为标准差

- 方差 $D(X)$ 刻画了随机变量的分散程度
- 离散型随机变量 $D(X)=\sum_i p_i(x_i-E(X))^2$
- 连续型随机变量 $D(X)=\int f(x)(x-E(X))^2dx$

方差公式

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

定义

方差和标准差

考虑随机变量 X，已知 $E[(X-E(X))^2]$ 存在，则
$D(X)=\mathrm{Var}(X)=E[(X-E(X))^2]$ 为 X 的方差，
$\sigma(X)=\sqrt{D(X)}$ 为标准差

- 方差 $D(X)$ 刻画了随机变量的分散程度
- 离散型随机变量 $D(X)=\sum_i p_i(x_i-E(X))^2$
- 连续型随机变量 $D(X)=\int f(x)(x-E(X))^2dx$

方差公式

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

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

常用离散型随机变量的方差

- $X \sim B(n,p)$, $E(X) = np$, $E(X^2) = n(n-1)p^2 + np$, $D(X) = npq$

常用离散型随机变量的方差

PEKING UNIVERSITY

常用离散型随机变量的方差

- $X \sim B(n, p)$, $E(X) = np$, $E(X^2) = n(n-1)p^2 + np$, $D(X) = npq$
- $X \sim \pi(\lambda)$, $E(X) = \lambda$, $E(X^2) = \lambda^2 + \lambda$, $D(X) = \lambda$

常用离散型随机变量的方差

PEKING UNIVERSITY

常用离散型随机变量的方差

- $X \sim B(n, p)$, $E(X) = np$, $E(X^2) = n(n-1)p^2 + np$, $D(X) = npq$
- $X \sim \pi(\lambda)$, $E(X) = \lambda$, $E(X^2) = \lambda^2 + \lambda$, $D(X) = \lambda$
- $X \sim NB(r, p)$, $E(X) = \frac{r(1-p)}{p}$, $D(X) = \frac{r(1-p)}{p^2}$

常用离散型随机变量的方差

常用离散型随机变量的方差

- $X \sim B(n,p)$, $E(X) = np$, $E(X^2) = n(n-1)p^2 + np$, $D(X) = npq$
- $X \sim \pi(\lambda)$, $E(X) = \lambda$, $E(X^2) = \lambda^2 + \lambda$, $D(X) = \lambda$
- $X \sim NB(r,p)$, $E(X) = \frac{r(1-p)}{p}$, $D(X) = \frac{r(1-p)}{p^2}$

考虑 $X \sim \pi(\lambda)$，我们有

$$
\begin{aligned}
E(X^2) &= E(X(X-1)) + E(X) \\
&= \sum_{i=0}^{\infty} i(i-1)e^{-\lambda}\lambda^i/i! + \lambda \\
&= \lambda^2 \sum_{i=2}^{\infty} e^{-\lambda}\lambda^{i-2}/(i-2)! + \lambda = \lambda^2 + \lambda
\end{aligned}
$$

推导

PEKING UNIVERSITY

考虑 $X \sim NB(r, p)$，我们有

$$
\begin{aligned}
E(X^2) &= E(X(X-1)) + E(X) \\
&= \sum_{k=0}^{\infty} k(k-1)\binom{k+r-1}{r-1} p^r q^k + \frac{r(1-p)}{p} \quad (q = 1-p) \\
&= p^r q^2 \left[\sum_{k=0}^{\infty} \binom{k+r-1}{r-1} q^k\right]'' + \frac{r(1-p)}{p} \\
&= p^r q^2 \left[\frac{1}{(1-q)^r}\right]'' + \frac{r(1-p)}{p} \\
&= r(r+1)\frac{q^2}{p^2} + \frac{r(1-p)}{p}
\end{aligned}
$$

思考如何推导二项分布

常用连续型随机变量的方差

- $X \sim U(a,b)$, $E(X)=(a+b)/2$, $E(X^2)=\frac{b^3-a^3}{3(b-a)}$,
  $D(X)=\frac{(b-a)^2}{12}$

常用连续型随机变量的方差

- $X \sim U(a,b)$, $E(X) = (a+b)/2$, $E(X^2) = \frac{b^3-a^3}{3(b-a)}$, $D(X) = \frac{(b-a)^2}{12}$
- $X \sim \Gamma(\alpha,\lambda)$, $E(X) = \alpha/\lambda$, $E(X^2) = (\alpha^2+\alpha)/\lambda^2$, $D(X) = \alpha/\lambda^2$

常用连续型随机变量的方差

- $X \sim U(a,b)$, $E(X) = (a+b)/2$, $E(X^2) = \frac{b^3-a^3}{3(b-a)}$, $D(X) = \frac{(b-a)^2}{12}$
- $X \sim \Gamma(\alpha,\lambda)$, $E(X) = \alpha/\lambda$, $E(X^2) = (\alpha^2+\alpha)/\lambda^2$, $D(X) = \alpha/\lambda^2$
- $X \sim N(\mu,\sigma^2)$, $E(X) = \mu^2$, $D(X) = \sigma^2$

常用连续型随机变量的方差

- $X \sim U(a,b)$, $E(X) = (a+b)/2$, $E(X^2) = \frac{b^3-a^3}{3(b-a)}$, $D(X) = \frac{(b-a)^2}{12}$
- $X \sim \Gamma(\alpha,\lambda)$, $E(X) = \alpha/\lambda$, $E(X^2) = (\alpha^2+\alpha)/\lambda^2$, $D(X) = \alpha/\lambda^2$
- $X \sim N(\mu,\sigma^2)$, $E(X) = \mu^2$, $D(X) = \sigma^2$

考虑 $X \sim N(\mu,\sigma^2)$，令 $Y = (X-\mu)/\sigma$，有 $E(Y^2) = 1$，因此

$$E(X^2) = E((\sigma Y + \mu)^2) = \sigma^2 + \mu^2$$

因此 $D(X) = E(X^2) - E(X)^2 = \sigma^2$

方差的性质

- 对于常数 C，有 $D(C) = 0$

方差的性质

PEKING UNIVERSITY

方差的性质

- 对于常数 C，有 $D(C) = 0$
- 对于常数 C，随机变量 X，有 $D(CX) = C^2 D(X)$

方差的性质

- 对于常数 C，有 $D(C)=0$
- 对于常数 C，随机变量 X，有 $D(CX)=C^2D(X)$
- 对于常数 B,C，随机变量 X，有 $D(B+CX)=C^2D(X)$

方差的性质

- 对于常数 C，有 $D(C)=0$
- 对于常数 C，随机变量 X，有 $D(CX)=C^2D(X)$
- 对于常数 B,C，随机变量 X，有 $D(B+CX)=C^2D(X)$
- 对于两个随机变量 X,Y，有 $D(X+Y)=D(X)+D(Y)+E[(X-E(X)(Y-E(Y)))]$

方差的性质

- 对于常数 C，有 $D(C)=0$
- 对于常数 C，随机变量 X，有 $D(CX)=C^2D(X)$
- 对于常数 B,C，随机变量 X，有 $D(B+CX)=C^2D(X)$
- 对于两个随机变量 X,Y，有
  $D(X+Y)=D(X)+D(Y)+E[(X-E(X)(Y-E(Y)))]$
- 假设随机变量 $\{X_i\}$ 相互独立，有 $D\left(\sum_i X_i\right)=\sum_i D(X_i)$

方差的性质

- 对于常数 C，有 $D(C)=0$
- 对于常数 C，随机变量 X，有 $D(CX)=C^2D(X)$
- 对于常数 B,C，随机变量 X，有 $D(B+CX)=C^2D(X)$
- 对于两个随机变量 X,Y，有
  $D(X+Y)=D(X)+D(Y)+E[(X-E(X)(Y-E(Y)))]$
- 假设随机变量 $\{X_i\}$ 相互独立，有 $D(\sum_i X_i)=\sum_i D(X_i)$
- $D(X)\leq E[(X-c)^2]$

方差的性质

- 对于常数 C，有 $D(C)=0$
- 对于常数 C，随机变量 X，有 $D(CX)=C^2D(X)$
- 对于常数 B,C，随机变量 X，有 $D(B+CX)=C^2D(X)$
- 对于两个随机变量 X,Y，有
  $D(X+Y)=D(X)+D(Y)+E[(X-E(X)(Y-E(Y)))]$
- 假设随机变量 $\{X_i\}$ 相互独立，有 $D(\sum_i X_i)=\sum_i D(X_i)$
- $D(X)\leq E[(X-c)^2]$
- $D(X)=0\equiv P(X=C)=1$

- 已知 $X\sim B(n,p)$，求 $D(X)$

方差的性质

- 对于常数 C，有 $D(C)=0$
- 对于常数 C，随机变量 X，有 $D(CX)=C^2D(X)$
- 对于常数 B,C，随机变量 X，有 $D(B+CX)=C^2D(X)$
- 对于两个随机变量 X,Y，有
  $D(X+Y)=D(X)+D(Y)+E[(X-E(X)(Y-E(Y)))]$
- 假设随机变量 $\{X_i\}$ 相互独立，有 $D(\sum_i X_i)=\sum_i D(X_i)$
- $D(X)\le E[(X-c)^2]$
- $D(X)=0\equiv P(X=C)=1$

- 已知 $X\sim B(n,p)$，求 $D(X)$ 考虑独立 0-1 分布
  $X_i\sim B(1,p)$，则 $X=\sum_{i=1}^n X_i$，且
  $D(X)=\sum_{i=1}^n D(X_i)=npq$

随机变量标准化

PEKING UNIVERSITY

随机变量标准化

已知随机变量 X 的期望 $E(X)$ 和方差 $D(X)$，$Y=\dfrac{X-E(X)}{\sqrt{D(X)}}$ 为 X 的标准化

随机变量标准化

PEKING UNIVERSITY

随机变量标准化

已知随机变量 X 的期望 $E(X)$ 和方差 $D(X)$，$Y = \frac{X - E(X)}{\sqrt{D(X)}}$ 为 X 的标准化

- 对于标准化的随机变量我们有 $E(Y) = 0,\ D(Y) = 1$

随机变量标准化

已知随机变量 X 的期望 $E(X)$ 和方差 $D(X)$，$Y = \frac{X-E(X)}{\sqrt{D(X)}}$ 为 X 的标准化

- 对于标准化的随机变量我们有 $E(Y) = 0, D(Y) = 1$
- 例如，对于 $X \sim N(\mu, \sigma^2)$，有 $Y = (X-\mu)/\sigma$，且 $E(Y) = 0$，$D(Y) = 1$

协方差和相关系数

随机变量 X, Y 的协方差和相关系数为
$\mathrm{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$ 和
$\rho_{XY} = \mathrm{Cov}(X, Y)/\sqrt{D(X)D(Y)}$

协方差和相关系数

随机变量 X, Y 的协方差和相关系数为
$\mathrm{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$ 和
$\rho_{XY} = \mathrm{Cov}(X, Y)/\sqrt{D(X)D(Y)}$

一些常用的结论有

- $E(XY) = E(X)E(Y) + \mathrm{Cov}(X, Y)$

协方差和相关系数

随机变量 X, Y 的协方差和相关系数为
$\mathrm{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$ 和
$\rho_{XY} = \mathrm{Cov}(X, Y)/\sqrt{D(X)D(Y)}$

一些常用的结论有

- $E(XY) = E(X)E(Y) + \mathrm{Cov}(X, Y)$
- $D(X + Y) = D(X) + D(Y) + 2\mathrm{Cov}(X, Y)$

协方差和相关系数

随机变量 X, Y 的协方差和相关系数为
$\mathrm{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$ 和
$\rho_{XY} = \mathrm{Cov}(X, Y)/\sqrt{D(X)D(Y)}$

一些常用的结论有

- $E(XY) = E(X)E(Y) + \mathrm{Cov}(X, Y)$
- $D(X + Y) = D(X) + D(Y) + 2\mathrm{Cov}(X, Y)$
  $D(X + Y) = E((X + Y)^2) - (E(X + Y))^2 =$
  $E(X^2) + E(Y^2) + 2E(XY) - (E(X))^2 - (E(Y))^2 - 2E(X)E(Y)$
- $D(\sum_i X_i) = \sum_i D(X_i) + 2\sum_{i,j} \mathrm{Cov}(X_i, X_j)$

协方差和相关系数

随机变量 X, Y 的协方差和相关系数为
$\mathrm{Cov}(X,Y) = E[(X-E(X))(Y-E(Y))]$ 和
$\rho_{XY} = \mathrm{Cov}(X,Y)/\sqrt{D(X)D(Y)}$

一些常用的结论有

- $E(XY) = E(X)E(Y) + \mathrm{Cov}(X,Y)$
- $D(X+Y) = D(X) + D(Y) + 2\mathrm{Cov}(X,Y)$
  $D(X+Y) = E((X+Y)^2) - (E(X+Y))^2 =$
  $E(X^2) + E(Y^2) + 2E(XY) - (E(X))^2 - (E(Y))^2 - 2E(X)E(Y)$
- $D(\sum_i X_i) = \sum_i D(X_i) + 2\sum_{i,j}\mathrm{Cov}(X_i, X_j)$
  $D(\sum_{i=1}^{n} X_i) = D(\sum_{i=1}^{n-1} X_i) + D(X_n) + 2\mathrm{Cov}(\sum_{i=1}^{n-1} X_i, X_n).$
  注意到 $\mathrm{Cov}(\sum_{i=1}^{n-1} X_i, X_n) = \sum_{i=1}^{n-1}\mathrm{Cov}(X_i, X_n)$

协方差和相关系数的性质

PEKING UNIVERSITY

协方差的性质

- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(X, X) = D(X)$
- $\mathrm{Cov}(aX, bY) = ab \cdot \mathrm{Cov}(X, Y)$
- $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$

协方差和相关系数的性质

协方差的性质

- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(X, X) = D(X)$
- $\mathrm{Cov}(aX, bY) = ab \cdot \mathrm{Cov}(X, Y)$
- $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$

我们注意到协方差的大小与随机变量的大小相关，考虑标准化的随机变量 $\tilde{X} = \frac{X - E(X)}{\sqrt{D(X)}}$，$\tilde{Y} = \frac{Y - E(Y)}{\sqrt{D(Y)}}$，我们有

协方差和相关系数的性质

协方差的性质

- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(X, X) = D(X)$
- $\mathrm{Cov}(aX, bY) = ab \cdot \mathrm{Cov}(X, Y)$
- $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$

我们注意到协方差的大小与随机变量的大小相关，考虑标准化的随机变量 $\tilde{X} = \frac{X-E(X)}{\sqrt{D(X)}}$, $\tilde{Y} = \frac{Y-E(Y)}{\sqrt{D(Y)}}$，我们有

相关系数—归一化的协方差

$\mathrm{Cov}(\tilde{X}, \tilde{Y}) = \rho_{XY}$

协方差和相关系数的性质

协方差的性质

- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(X, X) = D(X)$
- $\mathrm{Cov}(aX, bY) = ab \cdot \mathrm{Cov}(X, Y)$
- $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$

我们注意到协方差的大小与随机变量的大小相关，考虑标准化的随机变量 $\tilde{X} = \frac{X-E(X)}{\sqrt{D(X)}}$，$\tilde{Y} = \frac{Y-E(Y)}{\sqrt{D(Y)}}$，我们有

相关系数—归一化的协方差

$\mathrm{Cov}(\tilde{X}, \tilde{Y}) = \rho_{XY}$

$$\mathrm{Cov}(\tilde{X}, \tilde{Y}) = E[(\tilde{X} - E(\tilde{X}))(\tilde{Y} - E(\tilde{Y}))] = E[\tilde{X}\tilde{Y}] = E\left[\frac{X-E(X)}{\sqrt{D(X)}} \frac{Y-E(Y)}{\sqrt{D(Y)}}\right] = \rho_{XY}$$

例子

- 某条蚕的产卵数 X 服从泊松分布 $X \sim \pi(\lambda)$，每个卵变为成虫的概率为 p，各卵是否变为成虫彼此独立. 设成虫数为 Y，求 $\operatorname{cov}(X, Y)$.

例子

- 某条蚕的产卵数 X 服从泊松分布 $X \sim \pi(\lambda)$，每个卵变为成虫的概率为 p，各卵是否变为成虫彼此独立. 设成虫数为 Y，求 $\mathrm{cov}(X, Y)$.
  直接计算：$\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y)$.
  由已知条件得 $X \sim \pi(\lambda) \Rightarrow E(X) = \lambda$.
  经过计算可知 $Y \sim \pi(\lambda p) \Rightarrow E(Y) = \lambda p$.
  由离散型随机变量函数的期望公式:

$$E(XY) = \sum_{n=k}^{+\infty} \sum_{k=1}^{+\infty} nkP(X = n, Y = k) = \cdots = \lambda p + \lambda^2 p$$

$$\Rightarrow \mathrm{cov}(X, Y) = \lambda p.$$

是否有更简单的做法?

例子

- 某条蚕的产卵数 X 服从泊松分布 $X \sim \pi(\lambda)$，每个卵变为成虫的概率为 p，各卵是否变为成虫彼此独立. 设成虫数为 Y，求 $\mathrm{cov}(X, Y)$.

- 某条蚕的产卵数 X 服从泊松分布 $X \sim \pi(\lambda)$，每个卵变为成虫的概率为 p，各卵是否变为成虫彼此独立. 设成虫数为 Y，求 $\operatorname{cov}(X, Y)$.
  设 $Z = X - Y$（死卵数），则

$$
\begin{aligned}
P(Y = k, Z = l) &= P(Y = k, X = k + l) \\
&= P(X = k + l)P(Y = k \mid X = k + l) \\
&= \frac{\lambda^{k+l}}{(k+l)!} e^{-\lambda} \cdot \frac{(k+l)!}{k!\,l!} p^k (1-p)^l \\
&= \frac{(\lambda p)^k}{k!} e^{-\lambda p} \cdot \frac{(\lambda(1-p))^l}{l!} e^{-\lambda(1-p)}
\end{aligned}
$$

这表明 $Y \sim \pi(\lambda p), Z \sim \pi(\lambda(1-p))$，两者相互独立.

$\Rightarrow \operatorname{cov}(X, Y) = \operatorname{cov}(Y, Y) + \operatorname{cov}(Z, Y) = D(Y) + 0 = \lambda p.$

相关系数与线性相关

相关系数与线性相关

- $|\rho_{XY}| \leq 1$
- $|\rho_{XY}| = 1$ 等价于存在 a, b 使得 $P(Y = aX + b) = 1$(考虑 ρ_{XY} 的正负号与相关性的关系)

相关系数与线性相关

相关系数与线性相关

- $|\rho_{XY}| \leq 1$
- $|\rho_{XY}| = 1$ 等价于存在 a, b 使得 $P(Y = aX + b) = 1$(考虑 ρ_{XY} 的正负号与相关性的关系)

定义 $\tilde{X} = X - E(X), \tilde{Y} = Y - E(Y)$, 我们有（柯西不等式）

$$\rho_{XY}^2 = \frac{E(\tilde{X}\tilde{Y}')^2}{E(\tilde{X}^2)E(\tilde{Y}^2)} \leq 1$$

相关系数与线性相关

- $|\rho_{XY}| \le 1$
- $|\rho_{XY}| = 1$ 等价于存在 a, b 使得 $P(Y = aX + b) = 1$(考虑 ρ_{XY} 的正负号与相关性的关系)

定义 $\tilde{X} = X - E(X), \tilde{Y} = Y - E(Y)$，我们有（柯西不等式）

$$\rho_{XY}^2 = \frac{E(\tilde{X}\tilde{Y})^2}{E(\tilde{X}^2)E(\tilde{Y}^2)} \le 1$$

这里主要证明 $|\rho_{XY}| = 1 \Rightarrow P(Y = aX + b) = 1.$

$$E[(\tilde{Y} - a\tilde{X})^2] = a^2 E(\tilde{X}^2) - 2aE(\tilde{X}\tilde{Y}) + E(\tilde{Y}^2)$$

相关系数与线性相关

相关系数与线性相关

- $|\rho_{XY}| \leq 1$
- $|\rho_{XY}| = 1$ 等价于存在 a, b 使得 $P(Y = aX + b) = 1$(考虑 ρ_{XY} 的正负号与相关性的关系)

定义 $\tilde{X} = X - E(X), \tilde{Y} = Y - E(Y)$, 我们有（柯西不等式）

$$\rho_{XY}^2 = \frac{E(\tilde{X}\tilde{Y}')^2}{E(\tilde{X}^2)E(\tilde{Y}^2)} \leq 1$$

这里主要证明 $|\rho_{XY}| = 1 \Rightarrow P(Y = aX + b) = 1.$

$$E[(\tilde{Y} - a\tilde{X})^2] = a^2 E(\tilde{X}^2) - 2aE(\tilde{X}\tilde{Y}) + E(\tilde{Y}^2)$$

因为 $|\rho_{XY}| = 1$，我们有 $E(\tilde{X}\tilde{Y})^2 = E(\tilde{X}^2)E(\tilde{Y}^2)$，因此存在 a 使得 $E[(\tilde{Y} - a\tilde{X})^2] = 0$. 因为 $E(\tilde{Y} - a\tilde{X}) = 0$, 因此 $D(\tilde{Y} - a\tilde{X}) = 0$，也即是 $P(\tilde{Y} - a\tilde{X} = b') = 1$ 或 $P(Y = aX + b) = 1$

不（线性）相关

不（线性）相关

随机变量 X, Y 不（线性）相关，则有

- $\rho_{XY} = \mathrm{Cov}(X, Y) = 0$
- $E(XY) = E(X)E(Y)$
- $D(X + Y) = D(X) + D(Y)$

且以上三个条件等价

不（线性）相关

随机变量 X, Y 不（线性）相关，则有

- $\rho_{XY} = \mathrm{Cov}(X, Y) = 0$
- $E(XY) = E(X)E(Y)$
- $D(X+Y) = D(X) + D(Y)$

且以上三个条件等价

- 独立可以推出不相关，反之不然

不（线性）相关

随机变量 X, Y 不（线性）相关，则有

- $\rho_{XY} = \mathrm{Cov}(X, Y) = 0$
- $E(XY) = E(X)E(Y)$
- $D(X+Y) = D(X) + D(Y)$

且以上三个条件等价

- 独立可以推出不相关，反之不然
- $X \sim N(0,1), Y = |X|$，X, Y 是否独立、相关？

不（线性）相关

不（线性）相关

随机变量 X, Y 不（线性）相关，则有

- $\rho_{XY} = \mathrm{Cov}(X, Y) = 0$
- $E(XY) = E(X)E(Y)$
- $D(X+Y) = D(X) + D(Y)$

且以上三个条件等价

- 独立可以推出不相关，反之不然
- $X \sim N(0,1), Y = |X|$，X, Y 是否独立、相关？
  $E(X) = 0, E(XY) = 0$，因此 X, Y 不相关；

不（线性）相关

不（线性）相关

随机变量 X, Y 不（线性）相关，则有

- $\rho_{XY} = \mathrm{Cov}(X, Y) = 0$
- $E(XY) = E(X)E(Y)$
- $D(X + Y) = D(X) + D(Y)$

且以上三个条件等价

- 独立可以推出不相关，反之不然
- $X \sim N(0, 1), Y = |X|$，X, Y 是否独立、相关？
  $E(X) = 0, E(XY) = 0$，因此 X, Y 不相关；
  $P(X \leq c, Y \leq c) = P(Y \leq c)$ 与 $P(X \leq c)P(Y \leq c)$ 可能不同，因此不独立

不（线性）相关

PEKING UNIVERSITY

不（线性）相关

随机变量 X, Y 不（线性）相关，则有

- $\rho_{XY} = \mathrm{Cov}(X, Y) = 0$
- $E(XY) = E(X)E(Y)$
- $D(X+Y) = D(X) + D(Y)$

且以上三个条件等价

- 独立可以推出不相关，反之不然
- $X \sim N(0,1), Y = |X|$，X, Y 是否独立、相关？
  $E(X) = 0, E(XY) = 0$，因此 X, Y 不相关；
  $P(X \le c, Y \le c) = P(Y \le c)$ 与 $P(X \le c)P(Y \le c)$ 可能不同，因此不独立
- 因此这里的不相关更多指的是不线性相关，但是仍可能存在其它关联

例子

- 设随机变量 $U \sim U(0,2\pi), X=\cos U, Y=\cos(U+a)$. 求 ρ_{XY}.

$$E(X)=E(Y)=0, E(X^2)=E(Y^2).$$

$$E(X^2)=\int_0^{2\pi}\frac{1}{2\pi}\cos^2 u\mathrm{d}u=\frac{1}{2\pi}\int_0^{2\pi}\frac{1+\cos 2u}{2}u\mathrm{d}u=\frac{1}{2},$$

$$E(XY)=\int_0^{2\pi}\frac{1}{2\pi}\cos u\cos(u+a)\mathrm{d}u$$

$$=\frac{1}{2\pi}\int_0^{2\pi}\frac{\cos a+\cos(2u+a)}{2}\mathrm{d}u=\frac{\cos a}{2}.$$

$$\Rightarrow \rho_{XY}=\cos a.$$

- 当 $a=0$ 时，$\rho_{XY}=1$，$X=Y$（线性相关、正相关）
- 当 $a=\pi$ 时，$\rho_{XY}=-1$，$X=-Y$（线性相关、负相关）
- 当 $a=\pi/2$ 或 $3\pi/2$ 时，$\rho_{XY}=0$，X,Y 不线性相关，但也不相互独立（如何证明?）.

二元正态分布

- 考虑二元正态分布 (X, Y)，其中概率密度为
  $$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]},$$
  (X, Y) 独立等价于 (X, Y) 不相关

- 考虑二元正态分布 (X, Y)，其中概率密度为

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

(X, Y) 独立等价于 (X, Y) 不相关

$$\begin{aligned}\mathrm{Cov}(X,Y)&=E\{(X-\mu_1)(Y-\mu_2)\}\\
&=\int_{-\infty}^{+\infty}\frac{(x-\mu_1)}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}dx\int_{-\infty}^{+\infty}\frac{(y-\mu_2)}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\\
&\quad\times\exp\left\{-\frac{1}{2(1-\rho^2)\sigma_2^2}\left[y-\left(\mu_2+\frac{\rho\sigma_2}{\sigma_1}(x-\mu_1)\right)\right]^2\right\}dy\\
&=\int_{-\infty}^{+\infty}\frac{(x-\mu_1)}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\cdot\left[\mu_2+\frac{\rho\sigma_2}{\sigma_1}(x-\mu_1)-\mu_2\right]dx\\
&=\frac{\rho\sigma_2}{\sigma_1}\int_{-\infty}^{+\infty}\frac{(x-\mu_1)^2}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}dx\\
&=\frac{\rho\sigma_2}{\sigma_1}\cdot\sigma_1^2=\rho\sigma_1\sigma_2\end{aligned}$$

PEKING UNIVERSITY

- 考虑二元正态分布 (X,Y)，其中概率密度为 $f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]}$，$(X,Y)$ 独立等价于 (X,Y) 不相关.

二元正态分布

- 考虑二元正态分布 (X, Y)，其中概率密度为 $f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]}$,
  (X, Y) 独立等价于 (X, Y) 不相关.
  计算得 $\mathrm{Cov}(X, Y) = \rho\sigma_1\sigma_2$.
  第三章中已经证明 X, Y 独立 $\Leftrightarrow \rho = 0$，故命题得证.

二元正态分布

- 考虑二元正态分布 (X, Y)，其中概率密度为 $f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]}$,
  (X, Y) 独立等价于 (X, Y) 不相关.
  计算得 $\mathrm{Cov}(X, Y) = \rho\sigma_1\sigma_2$.
  第三章中已经证明 X, Y 独立 $\Leftrightarrow \rho = 0$，故命题得证.
- 二元正态分布 $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 由 5 个参数决定，这些参数的含义是：
  - μ_1, μ_2 分别为两个边缘的期望（均值）
  - σ_1^2, σ_2^2 分别为两个边缘的方差
  - ρ 为两个边缘的相关系数.

高阶矩

随机变量 X, Y 的 $k+l$ 混合矩和中心混合矩为 $E(X^kY^l)$ 和 $E[(X-E(X))^k(Y-E(Y))^l]$

高阶矩

随机变量 X, Y 的 $k+l$ 混合矩和中心混合矩为 $E(X^kY^l)$ 和 $E[(X-E(X))^k(Y-E(Y))^l]$

- 三阶矩-偏度：$E[(X-E(X))^3]$ 描述了峰往左或右偏

高阶矩

随机变量 X, Y 的 $k+l$ 混合矩和中心混合矩为 $E(X^kY^l)$ 和 $E[(X-E(X))^k(Y-E(Y))^l]$

- 三阶矩-偏度：$E[(X-E(X))^3]$ 描述了峰往左或右偏
- 四阶矩-峰度：$E[(X-E(X))^4]$ 描述了峰尖的大小

高阶矩

随机变量 X, Y 的 $k+l$ 混合矩和中心混合矩为 $E(X^kY^l)$ 和 $E[(X-E(X))^k(Y-E(Y))^l]$

- 三阶矩-偏度：$E[(X-E(X))^3]$ 描述了峰往左或右偏
- 四阶矩-峰度：$E[(X-E(X))^4]$ 描述了峰尖的大小
- $X \sim N(0,1), Y = |X|$, $E(X^2Y^2)$ 是否等于 $E(X^2)E(Y^2)$

高阶矩

随机变量 X, Y 的 $k+l$ 混合矩和中心混合矩为 $E(X^k Y^l)$ 和 $E[(X-E(X))^k(Y-E(Y))^l]$

- 三阶矩-偏度：$E[(X-E(X))^3]$ 描述了峰往左或右偏
- 四阶矩-峰度：$E[(X-E(X))^4]$ 描述了峰尖的大小
- $X \sim N(0,1), Y = |X|$, $E(X^2Y^2)$ 是否等于 $E(X^2)E(Y^2)$
  注意 $X^2 = Y^2$, 同时 $D(X^2) \neq 0$, 因此
  $E(X^2Y^2) \neq E(X^2)E(Y^2)$

协方差矩阵

考虑随机向量 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$，则协方差矩阵为 $B = E[(\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{X} - E(\boldsymbol{X}))^T]$.

协方差矩阵

考虑随机向量 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$，则协方差矩阵为
$B = E[(\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{X} - E(\boldsymbol{X}))^T]$.

- $B = B^T$

协方差矩阵

考虑随机向量 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$，则协方差矩阵为 $B = E[(\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{X} - E(\boldsymbol{X}))^T]$.

- $B = B^T$
- $B \geq 0$, 也即是对于任意向量 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^T$, 我们有 $\boldsymbol{\alpha}^T B \boldsymbol{\alpha} \geq 0$

协方差矩阵

考虑随机向量 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$，则协方差矩阵为 $B = E[(\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{X} - E(\boldsymbol{X}))^T]$.

- $B = B^T$
- $B \geq 0$, 也即是对于任意向量 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^T$, 我们有 $\boldsymbol{\alpha}^T B \boldsymbol{\alpha} \geq 0$
  $\boldsymbol{\alpha}^T B \boldsymbol{\alpha} = \boldsymbol{\alpha}^T E[(\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{X} - E(\boldsymbol{X}))^T]\boldsymbol{\alpha} = E[Y^2]$, 其中 $Y = \boldsymbol{\alpha}^T \cdot \boldsymbol{X}$
- 例如，n 元协方差矩阵为

$$\begin{pmatrix} D(X_1) & \mathrm{Cov}(X_1, X_2) & \ldots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & D(X_2) & \ldots & \mathrm{Cov}(X_2, X_n) \\ \ldots & \ldots & \ldots & \ldots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \ldots & D(X_n) \end{pmatrix}$$

考虑二元正态分布 (X_1, X_2)，其中概率密度为

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]},$$

协方差矩阵为 $B = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

二元正态分布

考虑二元正态分布 (X_1, X_2)，其中概率密度为

$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]}$，协

方差矩阵为 $B = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

行列式为 $|B| = \sigma_1^2\sigma_2^2\left(1-\rho^2\right)$

B 的逆矩阵为 $B^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}$

同时

$$(\boldsymbol{x}-\boldsymbol{\mu})^T B^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \frac{1}{1-\rho^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]$$

考虑二元正态分布 (X_1, X_2)，其中概率密度为

$f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]}$，协

方差矩阵为 $B = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

行列式为 $|B| = \sigma_1^2\sigma_2^2\left(1-\rho^2\right)$

B 的逆矩阵为 $B^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}$

同时

$$(\boldsymbol{x}-\boldsymbol{\mu})^T B^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \frac{1}{1-\rho^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]$$

二元正态分布

二元正态分布的概率密度可写成

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{2}{2}}|B|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T B^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

考虑二元正态分布 (X_1, X_2)，其中概率密度为

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]},$$ 协

方差矩阵为 $B = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

行列式为 $|B| = \sigma_1^2\sigma_2^2\left(1-\rho^2\right)$

B 的逆矩阵为 $B^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}$

同时

$$(\boldsymbol{x}-\boldsymbol{\mu})^T B^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \frac{1}{1-\rho^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]$$

二元正态分布

二元正态分布的概率密度可写成

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{2}{2}}|B|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T B^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

多元正态分布

多元正态分布的概率密度为
$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^T B^{-1}(\boldsymbol{x}-\boldsymbol{a})\right\}$，其中 B 为协方差矩阵

多元正态分布

多元正态分布

多元正态分布的概率密度为
$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{a})^T B^{-1}(\mathbf{x}-\mathbf{a})\right\}$，其中 B 为协方差矩阵

n 元正态分布由一个 n 维向量 $\mathbf{a}$ 和一个对称正定矩阵 B 决定，记作 $N(\mathbf{a}, B)$. 与二维情形类似：

- $\mathbf{a}$ 为分布的期望（均值）.
- B 为分布的协方差矩阵.

多元正态分布

多元正态分布的概率密度为
$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^T B^{-1}(\boldsymbol{x}-\boldsymbol{a})\right\}$，其中 B 为协方差矩阵

n 元正态分布由一个 n 维向量 $\boldsymbol{a}$ 和一个对称正定矩阵 B 决定，记作 $N(\boldsymbol{a}, B)$. 与二维情形类似：

- $\boldsymbol{a}$ 为分布的期望（均值）.
- B 为分布的协方差矩阵.

称满足 $\boldsymbol{a} = \boldsymbol{0}, B = I_d$ 的正态分布为 n 维标准正态分布，此时

$$f(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2}\right)$$

恰为 n 个相互独立的一维标准正态的联合密度.

多元正态分布

多元正态分布的概率密度为
$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^T B^{-1}(\boldsymbol{x}-\boldsymbol{a})\right\}$，其中 B 为协方差矩阵

多元正态分布

多元正态分布的概率密度为
$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^T B^{-1}(\boldsymbol{x}-\boldsymbol{a})\right\}$，其中 B 为协方差矩阵

- 多元正态分布的边缘分布也为多元正态分布

多元正态分布

多元正态分布

多元正态分布的概率密度为
$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^T B^{-1}(\boldsymbol{x}-\boldsymbol{a})\right\}$，其中 B 为协方差矩阵

- 多元正态分布的边缘分布也为多元正态分布
- 多元正态分布相互独立等价于随机变量 X_i, X_j 两两不相关，也等价于所有协方差为 0

多元正态分布

多元正态分布的概率密度为
$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^T B^{-1}(\boldsymbol{x}-\boldsymbol{a})\right\}$，其中 B 为协方差矩阵

- 多元正态分布的边缘分布也为多元正态分布
- 多元正态分布相互独立等价于随机变量 X_i, X_j 两两不相关，也等价于所有协方差为 0
- 多元正态分布的线性变换也为多元正态分布

多元正态分布

多元正态分布

多元正态分布的概率密度为
$f(\boldsymbol{x})=\frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^TB^{-1}(\boldsymbol{x}-\boldsymbol{a})\right\}$，其中 B 为协方差矩阵

- 多元正态分布的边缘分布也为多元正态分布
- 多元正态分布相互独立等价于随机变量 X_i, X_j 两两不相关，也等价于所有协方差为 0
- 多元正态分布的线性变换也为多元正态分布
- 对于 $\boldsymbol{X}\sim N(\boldsymbol{a},B)$，有 $A\boldsymbol{X}+\boldsymbol{b}\sim N(A\boldsymbol{a}+\boldsymbol{b},ABA^T)$

多元正态分布

多元正态分布的概率密度为
$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|B|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^T B^{-1}(\boldsymbol{x}-\boldsymbol{a})\right\}$，其中 B 为协方差矩阵

- 多元正态分布的边缘分布也为多元正态分布
- 多元正态分布相互独立等价于随机变量 X_i, X_j 两两不相关，也等价于所有协方差为 0
- 多元正态分布的线性变换也为多元正态分布
- 对于 $\boldsymbol{X} \sim N(\boldsymbol{a}, B)$, 有 $A\boldsymbol{X} + \boldsymbol{b} \sim N(A\boldsymbol{a} + \boldsymbol{b}, ABA^T)$
- 对于 $\boldsymbol{X} \sim N(\boldsymbol{a}, B)$，有 $B = AA^T$, 且 $\boldsymbol{X} = A\boldsymbol{Z} + \boldsymbol{a}$, 这里 $\boldsymbol{Z} \sim N(0, I)$

多元正态分布线性变换

对于 $\boldsymbol{X} \sim N(\boldsymbol{a}, B)$，有 $A\boldsymbol{X} + \boldsymbol{b} \sim N(A\boldsymbol{a} + \boldsymbol{b}, ABA^T)$

记 $\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b}$，根据密度变换公式，我们有

$$
\begin{aligned}
f_{\boldsymbol{Y}}(\boldsymbol{y}) &= f_{\boldsymbol{X}}(A^{-1}(\boldsymbol{y} - \boldsymbol{b}))|\partial \boldsymbol{y}/\partial \boldsymbol{x}|^{-1} \quad (\partial \boldsymbol{y}/\partial \boldsymbol{x} = A) \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}|A||B|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(A^{-1}(\boldsymbol{y} - \boldsymbol{b}) - \boldsymbol{a}\right)^T B^{-1}\left(A^{-1}(\boldsymbol{y} - \boldsymbol{b}) - \boldsymbol{a}\right)\right\} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}|A||B|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{b} - A\boldsymbol{a}\right)^T (A^T)^{-1} B^{-1}\left(A^{-1}(\boldsymbol{y} - \boldsymbol{b} - A\boldsymbol{a})\right)\right. \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}|\tilde{B}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{y} - \tilde{\boldsymbol{a}}\right)^T \tilde{B}^{-1}\left(\boldsymbol{y} - \tilde{\boldsymbol{a}}\right)\right\}
\end{aligned}
$$

其中 $\tilde{B} = ABA^T$，$\tilde{a} = A\boldsymbol{a} + \boldsymbol{b}$.

例

- 设 $(X_1, X_2) \sim N(0,0,1,1,1/2)$，即 (X_1, X_2) 是期望皆为 0，方差皆为 1，相关系数为 1/2 的二元正态分布（二维正态向量）. (1) 是否存在实数 a，使得 $X_1 - aX_2$ 与 X_2 相互独立？若存在，求出 a 的值；若不存在，请证明. (2) 求 $E(X_1^2 X_2^2)$.

- 设 $(X_1, X_2) \sim N(0, 0, 1, 1, 1/2)$，即 (X_1, X_2) 是期望皆为 0，方差皆为 1，相关系数为 1/2 的二元正态分布（二维正态向量）. (1) 是否存在实数 a，使得 $X_1 - aX_2$ 与 X_2 相互独立？若存在，求出 a 的值；若不存在，请证明. (2) 求 $E(X_1^2 X_2^2)$.

由正态向量的性质，$X_1 - aX_2$ 与 X_2 相互独立等价于二者不相关，即

$$0 = Cov(X_1 - aX_2, X_2) = Cov(X_1, X_2) - a \cdot Cov(X_2, X_2) = \frac{1}{2} - a \Rightarrow a = \frac{1}{2}.$$

例

PEKING UNIVERSITY

- 设 $(X_1, X_2) \sim N(0,0,1,1,1/2)$，即 (X_1, X_2) 是期望皆为 0，方差皆为 1，相关系数为 1/2 的二元正态分布（二维正态向量）. (1) 是否存在实数 a，使得 $X_1 - aX_2$ 与 X_2 相互独立？若存在，求出 a 的值；若不存在，请证明. (2) 求 $E(X_1^2 X_2^2)$.

由正态向量的性质，$X_1 - aX_2$ 与 X_2 相互独立等价于二者不相关，即

$$0 = Cov(X_1 - aX_2, X_2) = Cov(X_1, X_2) - a \cdot Cov(X_2, X_2) = \frac{1}{2} - a \Rightarrow a = \frac{1}{2}.$$

令 $Z = X_1 - X_2/2$，则 $D(Z) = D(X_1) + D(X_2)/4 - Cov(X_1, X_2) = 3/4$，$Z \sim N(0, 3/4)$ 且与 $X_2 \sim N(0,1)$ 相互独立.

计算如下：

$$\begin{aligned} E(X_1^2 X_2^2) &= E\left(\left(Z + \frac{X_2}{2}\right)^2 X_2^2\right) = E\left(Z^2 X_2^2 + Z X_2^3 + \frac{1}{4} X_2^4\right) \\ &= E(Z^2)E(X_2^2) + E(Z)E(X_2^3) + \frac{1}{4}E(X_2^4) = \frac{3}{4} \cdot 1 + 0 \cdot 0 + \frac{1}{4} \cdot 3 \\ &= \frac{3}{2}. \end{aligned}$$