《概率论》

2021年4月2日 21:36

https://zhuanlan.zhihu.com/p/31743949

概率论的基本概念

a. 样本空间，随机事件

- 样本空间
  - 集合 S
- 随机事件
  - 集合 $A \subseteq S$
- 基本事件
  - 集合 A 只有一个元素
- 不可能事件
  - 集合 $A = \emptyset$

b. 事件的相互关系及运算

- 事件的关系
  1. 包含 $A \subseteq B$
  2. 相等 $A = B$
  3. 和事件 $A + B$
  4. 积事件 $A \cap B, AB$
  5. 不相容事件，互斥事件 $AB = \emptyset$
  6. 差事件 $A - B$
  7. 逆事件 $\overline{A}$
- 事件关系满足交换律，结合律，德摩根率
- 基本的运算规律
  1. $A + \overline{A} = 1$
  2. $A\overline{A} = \emptyset$
  3. $A - B = A \cap \overline{B} = A - AB$

c. 概率

- 直观定义：随机事件发生的稳定值，记为 $P(A)=p$
- 概率的性质（前三条为概率的公理化定义）
  1. 非负性 $P(\emptyset)=0$
  2. 规范性 $P(A)=1-P(\overline{A})$
  3. 可列可加性
     - 若 A,B 两两互斥
     - $$P(\bigcup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)$$
  4. $$P(B-A)=P(B)-P(AB)$$
  5. 概率的加法公式
     - $$P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)-\sum_{1\le i<j\le n}P(A_iA_j)+\sum_{1\le i<j<k\le n}P(A_iA_jA_k)+\ldots+(-1)^{n-1}P(A_1A_2\ldots A_n)$$

d. 等可能概型（古典模型）

- 特点

  有限性、等可能性
- 组合数

$$C_N^n=\binom{N}{n}=\frac{N!}{n!(N-n)!}$$

- 放回抽样、不放回抽样
- 实际推断原理（概率小的事情在单次实验中几乎不会发生）

a. 条件概率

- 定义

$$P(B|A)=\frac{P(AB)}{P(A)},P(A)>0$$

- 乘法公式

$$P(AB)=P(A)P(B|A)$$

b. 全概率公式和贝叶斯定理

- 全概率公式
  - 若$B_1,B_2,B_3,\ldots,B_n$是S的划分（离散数学中的概念），则
  - $$P(A)=\sum_{j=1}^{n}P(B_j)P(A|B_j)$$
  - 关键在于能否构造一个合适的划分
  - 原理是分情况讨论

- 贝叶斯公式
  - $$P(B_i|A)=\frac{P(AB_i)}{P(A)}=\frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n}P(B_j)P(A|B_j)}$$
  - A是后验概率，B是先验概率。贝叶斯公式描述了先验概率已知的情况下，后验概率对先验概率的修正。
  - **直观理解：**癌症检查中，已知一个人有患癌症的可能，那么后验概率（检查结果）对先验概率（检查前患癌症的可能）的修正，可以增加或减少这个人患癌症的概率。也即医院检查可以（一定概率上）确诊。
  - **作者拓展：**贝叶斯公式在推荐算法上（如搜索引擎排序）具有重要应用，它可以通过用户的点击修正推荐排序结果

c. 事件的独立性

- 事件的独立性常常通过实际情况来判断
  相互独立≠不相关
- 公理化定义
  - 对事件组 $A_1,A_2,...,A_n$，若他们相互独立，则必有
  - $$\begin{aligned}&P(A_iA_j)=P(A_i)P(A_j)\\&P(A_iA_jA_k)=P(A_i)P(A_j)P(A_k)\\&...\\&P(A_1A_2...A_n)=P(A_1)P(A_2)...P(A_n)\end{aligned}$$
  - 注意，若三个事件两两独立，不能推出三个事件相互独立
- 性质
  - 若 A,B 相互独立，则 $\overline{A},B$，$A,\overline{B}$，$\overline{A},\overline{B}$ 也相互独立
- 小概率事件
  小概率事件在一次实验中几乎不发生
  但在大规模重复实验中，至少有一次发生的概率非常高

随机变量及其概率分布

a. 随机变量

- 定义
  - 随机变量 $X(e)$，X 是 $S\rightarrow R$ 的函数，e 是样本点
  - 自变量 $e\subset S$
  - 随机事件 $A=\{e|X(e)=I\}=\{X=I\}$
  - 如多次投掷骰子，随机事件 {6点在第3次出现} 可以记作 $X=3$，X 是随机变量
- 随机变量
  离散型随机变量，值的集合的基数小于等于阿列夫零（离散数学概念）
  连续型随机变量
- 分布律
X	x_1	x_2	$...$	x_k	$...$
P	p_1	p_2	$...$	p_k	$...$
  - $$P(X=x_k)=p_k\ (k=1,2,...)$$
- 几何分布 Geometric Distribution

- 多次投掷骰子，6 点第一次出现时投掷的次数
X	1	2	3	$\dots$	k	$\dots$
P	$\frac{1}{6}$	$\frac{5}{6}\cdot\frac{1}{6}$	$(\frac{5}{6})^2\cdot\frac{1}{6}$	$\dots$	$(\frac{5}{6})^{k-1}\cdot\frac{1}{6}$	$\dots$

b. 离散型随机变量

- 0-1分布
  - $$P(X=k)=p^k(1-p)^{n-k}$$
X	0	1
P	$1-p$	p
  - 若X服从两点分布，则单次试验称为伯努利（Bernoulli）试验
  - 记为$X\sim 0-1(p)$
  - 也记为 $X\sim B(1,p)$，B 是Binomial的意思，两点分布可以看作Binomial分布的特例
  - $\sim$ 读作服从于
- 二项分布 Binomial Distribution
  - $$P(X=k)=C_n^k\cdot P^k\cdot(1-p)^{n-k}$$
  - n 重Bernoulli实验，事件发生次数 k 的统计规律
  - 记为$X\sim B(n,p)$
- 泊松分布 Poisson Distribution
  - $$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}\ (k=0,1,2,\dots)$$
  - 记为 $X\sim\pi(\lambda)$ 或 $x\sim P(\lambda)$

  与二项分布的关系
  - 当 n 很大 p 很小的时候
  - $$C_n^k P^k(1-p)^{n-k}=\frac{\lambda^k e^{-\lambda}}{k!}\ \lambda=np$$
- 几何分布

  定义：在n次伯努利试验中，试验k次才得到第一次成功的机率。

  详细：前k-1次皆失败，第k次成功的概率。
  - $P(X=k)=p(1-p)^{k-1}$
  - 记为 $X\sim Geom(p)$
  - 实例：研究段誉多少次施展武功能成功的统计规律

c. 分布函数

- 定义
  - $F_X(x) = P(X \leq x)$
- 离散型的随机变量分布函数为阶梯函数
- 性质
  - $P(a < X \leq b) = F(b) - F(a)$
  - $P(a < X < b) = F(b-0) - F(a)$
  - $P(X = b) = F(b) - F(b-0)$
  - $F(x)$ 单调不减
  - $F(-\infty) = 0, F(+\infty) = 1$
  - $F(x)$ 右连续

d. 连续性随机变量及其概率密度

- 定义
  - $$F(x) = \int_{-\infty}^{x} f(t)dt$$
  - $F(x)$ 为连续型随机变量的分布函数
  - $f(t)$ 为连续型随机变量的概率密度函数
  - 若一个随机变量有概率密度函数则其一定为随机变量
- 性质
  1. $$f(x) \geq 0$$
  2. $$F(+\infty) = 1$$
  3. $$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(t)dt$$
  4. $$F'(x) = f(x)$$
  5. $$f(x)$$

     可以大于1
  6. 概率密度对 $>, \geq, <, \leq$ 不敏感，即对端点取值不敏感

e. 均匀分布和指数分布

- 均匀分布 Uniform Distribution
  - $$f(x) = \frac{1}{b-a} \quad a \leq x < b$$
  - $$F(x) = \frac{x-a}{b-a} \quad a \leq x < b$$
  - 记为 $X \sim U(a,b)$ 或 $X \sim Unif(a,b)$
- 指数分布 Exponential Distribution

- $$f(x) = \lambda e^{-\lambda x} \quad x > 0$$
- $$F(x) = 1 - e^{-\lambda x} \quad x > 0$$
- 记为 $X \sim E(\lambda)$ 或 $X \sim Emp(\lambda)$
- 指数分布具有无记忆性（Memoryless Property）且在连续性随机变量的分布中，只有指数分布具有无记忆性
- 实例：设旅客等待时间服从指数分布，则已知旅客已经等了20分钟，求旅客再等5分钟的概率，和旅客从头开始等5分钟的概率相同
- 即 $P(X > 25|X > 20) = P(X > 5)$
- 指数分布常用来表示独立随机事件发生的时间间隔，如中文维基百科新条目出现的时间间隔
- 在排队论中，一个顾客接受服务的时间长短也可以用指数分布来近似

f. 正态分布

- 正态分布 Normal Distribution
  - $$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
  - 记为 $X \sim N(\mu, \sigma^2)$
- 性质
  - 关于 $x = \mu$ 对称
  - $$f_{max} = f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$$
  - $$lim f(x) = 0$$
- 参数性质
  - 改变 μ，$f(x)$ 只沿 x 轴评议
  - σ 越大，$f(x)$ 越矮胖，σ 称为尺度参数
- 实例：身高，体重，测量误差，多个随机变量的和
- 标准正态分布
  - $$Z \sim N(0, 1)$$
  - $$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$
  - $$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$
  - $\Phi(z)$ 有标准正态分布函数表
- 一般正态分布转为标准正态分布
  - 当 $X \sim N(\mu, \sigma^2)$ 时，$(x - \mu)/\sigma \sim N(0, 1)$
  - $$F_x(a) = P(x \le a) = P(\frac{x-\mu}{\sigma} \le \frac{a-\mu}{\sigma}) = \Phi(\frac{a-\mu}{\sigma})$$
- 3δ标准
  - 当 x 落在 $(-3\sigma, 3\sigma)$ 的概率为 99.73%

g. 随机变量函数的分布

- 已知 X 的概率分布，已知 $Y = g(x)$，求 Y 的概率分布

- 已知 X 的概率分布，已知 $Y=g(x)$，求 Y 的概率分布
  - 先给出 Y 的可能分布，再利用等价事件来给出概率分布
  - 离散型随机变量，直接利用分布律求解即可
  - 连续型随机变量，先利用分布函数找到等价事件，再利用概率密度函数即可
- 定理
  - 若 $Y=g(x)$，$g'(x)>0$ 或 $g'(x)<0$
  - $$f_Y(y)=f_x(h(y))\cdot|h'(y)|\ \alpha<y<\beta$$
  - $h(y)$ 是 $g(x)$ 的概率密度函数的反函数
  - α 和 β 是根据 x 与 y 的对应关系求得的
- 一般的
  - 若 $X\sim N(\mu,\sigma^2)$，$Y=aX+b$，则 $Y\sim(a\mu+b,a^2\sigma)$

- **当前的所有分布**
  二项分布 Binomial Distribution
  泊松分布 Poisson Distribution
  几何分布 Geometric Distribution
  均匀分布 Uniform Distribution
  指数分布 Exponential Distribution
  正态分布 Normal Distribution

二元随机变量及其分布

a. **二元随机变量，离散型随机变量分布律**

- 二元随机变量
  同一个样本空间的两个随机变量构成的向量
- 离散型随机变量的分布律
  - $$P(X=x_i,Y=y_i)=p_y\ i,j=1,2,\ldots$$
  -

$x\\y$	y_1	y_2	$\ldots$	y_1	$\ldots$
x_1	p_{11}	p_{12}	$\ldots$	p_{ij}	$\ldots$
x_2	p_{21}	p_{22}	$\ldots$	p_{2j}	$\ldots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_i	p_{i1}	p_{i2}	$\ldots$	p_{ij}	$\ldots$
$\ldots$	$\ldots$		$\ldots$		$\ldots$

  - 实例：$P(X=0,Y=1)=P(Y=1|X=0)\cdot P(X=0)$

b. **二元离散型随机变量边际分布律与条件分布律**

- 边际分布律

- $$P(X+x_i)=P(X=x_i,\bigcup_{j=1}^{\infty}(Y=y_i))=\sum_{j=1}^{\infty}p_{ij}=p_{i\cdot}$$
- $$P(Y+y_i)=P_{\cdot y}$$
-

$x\\y$	y_1	y_2	$\dots$	y_1	$\dots$	$P(X=x_i)$
x_1	p_{11}	p_{12}	$\dots$	p_{ij}	$\dots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\dots$	p_{2j}	$\dots$	$p_{2\cdot}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_i	p_{i1}	p_{i2}	$\dots$	p_{ij}	$\dots$	$p_{i\cdot}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$P(Y=y_j)$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\dots$	$p_{\cdot j}$	$\dots$	1

- 已知条件分布律一定能求出边际分布律，但已知编辑分布律不一定能求出条件分布律

- 条件分布律
  - $$P(X=x_i)|Y=y_j)=\frac{p_{ij}}{p_{\cdot j}}\ i=1,2,\dots$$
  - 条件分布律不唯一

a. 二元随机变量分布函数、边际分布函数及条件分布函数

- 联合分布函数

$$F(x,y)=P\{(X\leq x)\cap(Y\leq y)\}=P(X\leq x,Y\leq y)$$

- 边际分布函数

$$F_x(x)=F(x,+\infty)=\lim_{y\to\infty}F(x,y)$$

- 条件分布函数
  - 若 $P(Y=y)>0$
  - $$F_{X|Y}(x|y)=P(X\leq x|Y=y)=\frac{P(X\leq x,Y=y)}{P(Y=y)}$$
  - 对于连续型随机变量也可以用如上记法，但注意此时的 $y\leq Y\leq \epsilon$

a. 二元连续型随机变量，联合概率密度

- 二元随机变量的联合概率密度
  - $$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv$$
  - 其中 $f(x,y)$ 为 (X,Y) 的概率密度
- 性质
  - $$P((x,y)\in D)=\iint_D f(x,y)dxdy$$

- $$P((x,y)\in D)=\iint_D f(x,y)dxdy$$
- $$\frac{\partial^2 F(x,y)}{\partial x\partial y}=f(x,y)$$

- 提示

  此处应具有求二重积分的能力

b. 二元连续型随机变量边际概率密度

- 二元连续型随机变量的边际概率密度

  $$f_x(x)=\int_{-\infty}^{+\infty} f(x,y)dy$$

- 二元连续型随机变量的边际概率函数

  $$F_x(x)=F(x,+\infty)=\int_{-\infty}^{x}[\int_{-\infty}^{+\infty} f(u,y)dy]du$$

c. 二元连续型随机变量条件概率密度

- 二元连续型随机变量的条件概率密度

  $$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$

- 变式

  $$f(x,y)=f_{X|Y}(x|y)\cdot f_Y(y)$$

- **汇总：二元离散型与连续型随机变量分布比较**
  - 离散型
  - 联合分布律
  - 边际分布律
  - 条件分布律
  - 连续型
  - 联合概率密度
  - 边际概率密度
  - 条件概率密度

d. 二元均匀分布，二元正态分布

- 二元均匀分布
- $f(x,y)=1/A,(x,y)\in D$
- 二元正态分布
  - $$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\times$$
    $$exp\{\frac{-1}{2(1-\rho^2)}[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}]\}$$
  - $\sigma_1,\sigma_2>0$，$-1<\rho<1$
  - 记为 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$

- 二元正态分布的边际概率密度

  $X \sim N(\mu_1, \sigma_2^2)$

- 二元正态分布的条件概率密度
  - $$Y|X \sim N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), (1-\rho^2)\sigma_2^2)$$
  - 即条件概率分布服从正态分布

e. 随机变量的独立性

- 随机变量的独立性
  - $$F(x,y) = F_X(x)F_Y(y)$$
  - 离散型随机变量的独立性
  - $$P(X=x_i, Y=y_j) = P(X=x_i)P(Y=y_j)$$
  - 连续型随机变量的独立性
  - $$f(x,y) = f_X(x)f_Y(y)$$
- n元随机变量的分布
  - 分布函数
  - 分布律
  - 概率密度函数
  - 边际分布
- 向量的独立性
  - $$F(x_1,x_2,\ldots,x_m,y_1,y_2,\ldots,y_n) = F_1(x_1,x_2,\ldots,x_m)F_2(y_1,y_2,\ldots y_n)$$
  - 性质
  - 若两向量独立
    1. X_i 与 Y_j 相互独立
    2. 若 $g(x_1,x_2,\ldots,x_m)$ 与 $h(y_1,y_2,\ldots y_n)$ 是连续函数，则 $g(X_1,X_2,\ldots,X_m)$ 与 $h(Y_1,Y_2,\ldots Y_n)$ 相互独立
  - 直观理解
    - 性质1表明，若 X_i 与 Y_j 相互独立，则 X_1 与 Y_1 相互独立，X_1 与 X_2 相互独立
    - 性质2表明，若 X_i 与 Y_j 相互独立，则 X_1+X_2 与 $Y_1 \times Y_2$ 相互独立

f. 二元随机变量函数的分布

- 二元随机变量函数的分布（如 $Z=X<Y$ 的分布）
  - 离散型
  - 用分布律，分析各种情况
  - 连续型
  - 先求 $F(x)$，再求导得到

g. Z=X+Y的分布

- 连续型

- $$F_z(z) = P(Z \le z) = \iint_{x+y\le z} f(x,y)dxdy$$
- $$f_Z(z) = \int_{-\infty}^{+\infty} f(z-y,y)dy$$
- 卷积公式
- 当 X 与 Y 相互独立时
- $$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)dy$$
- 拓展：知乎问题：如何通俗易懂地解释卷积？
- 关于正态分布的结论
- 若 X 与 Y 相互独立， $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$，则
  - $$(Z = X+Y) \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$
- 更一般的，若 X_i 服从线性分布，则其线性组合
  - $c_0 + c_1X_1 + c_2X_2 + \ldots + c_nX_n \sim N(\mu, \sigma^2)$
- 其中
  - $$\mu = c_0 + c_1\mu_1 + \ldots + c_n\mu_n,\ \sigma^2 = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \ldots + c_n^2\sigma_n^2$$
- Γ 分布 Gamma Distribution （非重点，可略过）

有关卷积的部分 https://www.zhihu.com/question/22298352

- 离散型
  - 若 $X_1, X_2, \ldots, X_n$ 独立且服从 $B(1,p)$ 则
  - $$X_1 + X_2 + \ldots + X_n \sim B(n,p)$$
  - 若 X 与 Y 相互独立，$X \sim B(n_1,p), Y \sim B(n_2,p)$ 则
  - $$X+Y \sim B(n_1+n_2, p)$$
  - 若 X 与 Y 相互独立，$X \sim \pi(\lambda_1), Y \sim \pi(\lambda_1+\lambda_2)$ 则
  - $$X+Y \sim \pi(\lambda_1+\lambda_2)$$

h. MAX(X,Y)和MIN(X,Y)的分布

- 若 X 与 Y 相互独立
  - $$\begin{aligned} F_{max}(z) &= P(M \le z) \\ &= P(X \le z, Y \le z) \\ &= P(X \le z)P(Y \le z) \end{aligned}$$
  - $$f_{max}(z) = f_X(z)f_Y(z)$$
  - 同理
  - $$f_{min}(z) = 1-(1-f_X(z))(1-f_Y(z))$$
- n 个相互独立的随机变量同理
- 若 X_n 相互独立且分布相同
  - $$f_{max}(z) = n[F(z)]^{n-1}f(z)$$

- 若 X_n 相互独立且分布相同
  - $$f_{max}(z) = n[F(z)]^{n-1} f(z)$$
  - $$f_{min}(z) = n[1 - F(z)]^{n-1} f(z)$$

随机变量的数字特征

a. **随机变量的数学期望**

- 离散型
  - $$P(X = x_k) = p_k \ k = 1, 2, \ldots$$
  - $$E(X) = \sum_{k=1}^{+\infty} x_k p_k$$
  - 要求 $E(X)$ 绝对收敛
  - 拓展：绝对收敛与条件收敛
    - 如果级数

      $$\sum U_n$$

      收敛，而

      $$\sum |U_n|$$

      发散，则称级数

      $$\sum U_n$$

      条件收敛
    - 如

      $$(-1)^k \cdot \frac{1}{x}$$

      条件收敛，但他的期望不存在
- 连续型
  - $$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$
  - 要求 $E(X)$ 绝对收敛
- 常见分布的期望

- $$X \sim 0-1(p), E(X)=p$$
- $$X \sim \pi(\lambda), E(X)=\lambda$$
- $$X \sim N(\mu, \sigma^2), E(X)=\mu$$
- $$X \sim E(\lambda), E(X)=\frac{1}{\mu}$$
- $$X \sim B(n,p), E(X)=np$$
- $$X \sim G(p), E(X)=(1/p)$$
- $$X \sim U(a,b), E(X)=\frac{a+b}{2}$$

b. 随机变量函数的数学期望

- 离散型
  - $$E(Y)=E[g(X)]=\sum_{k=1}^{\infty} g(x_k)p_k$$
  - 要求 $E(Y)$ 绝对收敛
- 连续型
  - $$E(Y)=E[g(X)]=\int_{-\infty}^{+\infty} g(x)f(x)dx$$
  - 要求 $E(Y)$ 绝对收敛
- 提示：已知一个分布，求他的函数的分布，利用以上公式，无需再求一次概率密度
- 二元离散型

$$E(Z)=E[h(X,Y)]=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} h(x_i,y_j)p_{ij}$$

- 二元连续型
  - $$E(Z)=E[h(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)f(x,y)dxdy$$
  - 特别的
  - $$E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)dxdy$$

c. 数学期望的性质

- 数学期望的性质
  - 设 c 是常数，则有 $E(c) = c$
  - 设 X 是一个随机变量，c 是常数，则有 $E(cX) = xE(X)$
  - 设 X, Y 是两个随机变量，则有 $E(X+Y) = E(X) + E(Y)$
  - 设 X, Y 是**相互独立**的两个随机变量，则有
  - $E(XY) = E(X)E(Y)$
  - 上述运算均可推广至 n 个随机变量运算

d. 方差定义和计算公式

- 方差

$$D(X) = Var(X) = E\{[X - E(X)^2]\}$$

- 标准差

$$\sigma(X) = \sqrt{D(X)}$$

- 离散型

$$D(X) = \sum_{i=1}^{+\infty}[x_i - E(X)]^2 p_2$$

- 连续型

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

- 计算公式

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

- 常见分布的方差（一）

$$X \sim 0-1(p), E(X) = p, D(X) = p(1-p)$$

$$X \sim \pi(\lambda), E(X) = \lambda, D(X) = \lambda$$

$$X \sim N(\mu, \sigma^2), E(X) = \mu, D(X) = \sigma^2$$

$$X \sim E(\lambda), E(X) = \frac{1}{\mu}, D(X) = \frac{1}{\lambda^2}$$

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

e. 方差的性质

- 方差的性质
  - 设 c 是常数，则有 $D(c) = 0$
  - 设 X 是一个随机变量，c 是常数，则有 $D(cX) = c^2 D(X)$
  - 设 X, Y 是两个随机变量，则有 $D(X+Y) = D(X) + D(Y) + 2 \cdot Cov$
  - $Cov = E\{[X - E(X)][Y - E(Y)]\}$
  - 设 X, Y 是**相互独立**的两个随机变量，则
  - $D(X+Y) = D(X) + D(Y)$
  - 上述运算均可推广至 n 个随机变量运算
- 此处可证明第二章第15讲的正态分布的性质

- 若 $X_i \sim N(\mu_i, \sigma_i^2)$ 且他们相互独立，则它们的线性组合
- $$c_0 + c_1X_1 + c_2X_2 + \ldots + c_nX_n \\ \sim N(c_0 + c_1\mu_1 + \ldots + c_n\mu_n, c_1^2\sigma_2^2 + c_2\sigma_2^2 + \ldots + c_n^2\sigma_n^2)$$

- 标准化变量
  - $$X^* = \frac{X-\mu}{\sigma}$$
  - 性质
  - $$E(X^*) = 0$$
  - $$D(X^*) = 1$$
  - 标准化的变量是没有量纲的

f. 协方差与相关系数

- 协方差 Covariance
  - $$Cov(X,Y) = E\{[X-E(X)][Y-E(Y)]\}$$
  - $Cov(X,Y) > 0$，X 与 Y 正相关
  - $Cov(X,Y) < 0$，X 与 Y 负相关
  - $Cov(X,Y) = 0$，X 与 Y 不相关
  - $$Cov(X,Y) = E(XY) - E(X)E(Y)$$
- 协方差的性质
-
  - $$Cov(X,Y) = Cov(Y,X)$$
  - $$Cov(X,X) = D(X)$$
  - $$Cov(aX,bY) = ab \cdot Cov(X,Y)$$
  - $$Cov(X_1+X_2,Y) = Cov(X_1,Y) + Cov(X_2,Y)$$
- 相关系数 Correlation coefficient
  - $$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)D(Y)}}$$
  - $$\rho_{XY} = Cov(X^*,Y^*)$$
  - 提示：X^* 与 Y^* 的概念在上一讲有提到
  - $$X^* = \frac{X-E(X)}{\sqrt{D(X)}}$$
  - 相关系数的性质
  - $$|\rho_{XY}| \le 1$$
  - $$|\rho_{XY}| = 1 \Leftrightarrow P(Y = a+bX) = 1$$
    - 即当相关系数为 1 时，表明 X, Y 之间有严格的线性关系

- 即当相关系数为 1 时，表明 X,Y 之间有严格的线性关系
- 当相关系数为 0 时，X,Y 不相关

g. 不相关与独立

- 不相关 Uncorrelated
  - $\rho_{XY}=0$
- 独立
  - 第三章已说明
  - $$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)$$
  - $$f(x,y)=f_X(x)f_Y(y)$$
- **独立一定不相关，但不相关不一定独立**
  不相关其实是描述线性不相关
- 正态分布的独立性和相关性是相同的

h. 矩，协方差矩阵，多元正态分布的性质

- 矩
  - $E(X^k)$ 存在，则称为 X 的 k 阶（原点）矩
  - $E\{[X-E(X)]^k\}$ 存在，则成为 X 的 k 阶中心矩
  - 之前提到的随机变量的期望和方差就是其 1 阶原点矩和 2 阶中心矩
- 协方差矩阵 Covariance Matrix
  - $$C=Cov(\tilde{x})=\begin{pmatrix} D(X_1) & Cov(X_1,X_2) & \dots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & D(X_2) & \dots & Cov(X_2,X_n) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(X_n,X_1) & Cov(X_n,X_2) & \cdots & D(X_n) \end{pmatrix}$$
- n元正态随机变量的联合概率密度的矩阵表示
  - 列向量 $x=(x_1,x_2,\cdots,x_n)^T,\tilde{\mu}=(E(X_1),E(X_2),\cdots,E(X_n))^T$
  - 它的协方差矩阵为 $C=(C_{ij})_{n\times n},c_{ij}=Cov(X_i,X_j)$
  - 则其联合概率密度为
  - $$f(x_1,x_2,\cdots,x_n)=\frac{1}{(2\pi)^{\frac{n}{2}}|C|^{\frac{1}{2}}}\cdot exp\{-\frac{1}{2}(x-\tilde{\mu})^TC^{-1}(x-\tilde{\mu})\}$$
- 以上略有难度，无需详细掌握
- n元正态随机变量的重要性质
  - 对于

$$\tilde{X}=(X_1,X_2,\dots,X_n)^T$$

  - 任意切片是正态随机变量
  - 任意线性组合服从一元正态分布
  - X_n 相互独立 $\Leftrightarrow$ X_n 两两不相关 $\Leftrightarrow$ $\tilde{X}$ 的协方差矩阵为对角矩阵## 第27讲 随机变量的数学期望
- 离散型
  - $$P(X=x_k)=p_k\ k=1,2,\dots$$

- $$P(X=x_k)=p_k\ k=1,2,\ldots$$
- $$E(X)=\sum_{k=1}^{+\infty}x_k p_k$$
- 要求 $E(X)$ 绝对收敛
- 拓展：绝对收敛与条件收敛
  - 如果级数

    $$\sum U_n$$

    收敛，而

    $$\sum |U_n|$$

    发散，则称级数

    $$\sum U_n$$

    条件收敛
  - 如

    $$(-1)^k\cdot\frac{1}{x}$$

    条件收敛，但他的期望不存在

- 连续型
  - $$E(X)=\int_{-\infty}^{+\infty}xf(x)dx$$
  - 要求 $E(X)$ 绝对收敛
- 常见分布的期望
  - $$X\sim 0-1(p),E(X)=p$$
  - $$X\sim\pi(\lambda),E(X)=\lambda$$
  - $$X\sim N(\mu,\sigma^2),E(X)=\mu$$
  - $$X\sim E(\lambda),E(X)=\frac{1}{\mu}$$
  - $$X\sim B(n,p),E(X)=np$$
  - $$X\sim G(p),E(X)=(1/p)$$
  - $$X\sim U(a,b),E(X)=\frac{a+b}{2}$$