概率论复习提纲（茆诗松版）

1 概率论基本概念与方法

- 样本空间;
  ◇ 随即试验所有可能结果的组成的集合
- 事件;
  ◇ 样本空间的子集
- 事件的发生;
  ◇ 这一子集中的样本点出现
- 事件的关系与运算;
  ◇ $\subset; =; \cup; \cap; -$
  ◇ 互不相容；对立
- 互不相容与相互独立, 互不相关与相互独立;
- 概率的性质;
  ◇ 定义中的三条基本性质$\Longrightarrow$概率的六条性质
- 古典概率计算公式;
  ◇ $P(A) = \frac{n_A}{n}$
- 条件概率$P(B|A)$的意义及计算;
  ◇ 事件A发生的条件下，事件B发生的概率
- 全概率公式与Bayes公式.
  ◇ 前者是将一复杂事件的概率转换成不同原因/情况下一系列简单事件的概率求和
  ◇ 后者是一事件已出现，考察引发该事件发生的各种原因的可能性大小
  ◇ **注意:两者的考察可能会在同一题目中出现.**

  例1：某产品的合格品率为99%. 已知一个合格产品使用10年以上的概率达到0.90, 而一个不合格产品使用10年以上的概率仅为0.60. 求:

  1. 任取一个该产品, 它能使用10年以上的概率;
  2. 已知一个产品已经使用了10年还能正常工作的条件下, 它是合格品的概率.

1答: 设A为“任取一产品是合格品”, B为“产品能使用10年以上”. 样本空间的完备事件组为A和$\overline{A}$. 则

$$\begin{aligned} P(B) &= P(A)P(B|A) + P(\overline{A})P(B|\overline{A}) \\ &= 0.99 \times 0.9 + 0.01 \times 0.6 \\ &= 0.8970 \end{aligned}$$

2答: 在一个产品已经使用了10年还能正常工作的条件下, 它是合格品的概率为:

$$\begin{aligned} P(A|B) &= \frac{P(A)P(B|A)}{P(B)} \\ &= \frac{0.99 \times 0.9}{0.8970} \\ &= \frac{0.8910}{0.8970} \\ &= 0.9933 \end{aligned}$$

2 一维随机变量及其分布

- 离散型随机变量与连续型随机变量的定义;
  ◇ 前者基于随机变量取值的数目
  ◇ 后者是：r.v 的分布函数定义为非负可积函数的变上限积分
- 分布函数的定义及性质;
  ◇ 定义：$F(x) = P(X \leq x)$
  ◇ 性质：单调非减；$F(-\infty) = 0, F(+\infty) = 1$;右连续性
- 离散型随机变量分布率的性质;
  ◇ $p_i > 0; \sum p_i = 1$
- 连续型随机变量密度函数的性质
  ◇ $p(x) > 0; \int_{-\infty}^{+\infty} p(x) = 1$
- 分布函数与密度函数的关系;
  ◇ 离散型：$F(x)$是分段函数，p_i是跳跃点处的跳跃高度
  ◇ 连续型：$p(x) = F'(x)$
- 概率计算:
  - $P(X \leq a) = F(a)$, $P\{a < X \leq b\} = F(b) - F(a)$, 特别$P\{X = c\} = F(c) - F(c^-)$;(适用于离散及连续r.v.)
  - $P\{X \in A\} = \sum\limits_{x_i \in A} p_i$;
  - $P\{X \in A\} = \int_A p(x)dx$.
- 6个常用分布的分布及其数字特征(了解它们的性质及相互关系);
- 求离散型随机变量的分布:
  1. X的取值范围;
  2. X取每个值的概率.
- 随机变量函数$Y = g(X)$的分布.
  ◇ 离散型
  ◇ 连续型：$F_Y(y) = \int_{\{x|g(x) \leq y\}} p_X(x)dx$，或者由$P_{56}$定理1：$f_Y(y) = p_X(h(y))|h'(y)|$，其中$x = h(y)$是$y = g(x)$的反函数.

例2：设随机变量X的概率密度函数为:

$$f(x)=\begin{cases}\sin(x) & 0\le x\le A\\ 0 & \text{其它}\end{cases}$$

试计算

1. (5分) 常数A;
2. (5分) X的分布函数$F(x)$;
3. (5分) $P\{|X|\le 1\}$;
4. (5分) $Y=\sin(X)$的概率密度.

1答: 根据概率密度函数的性质,

$$\int_0^A \sin(x)\mathrm{d}x = 1-\cos(A)=1$$

（+2）

得$A=\frac{\pi}{2}+k\pi$. （+1）
又因为在$[0,\ A]$上, $f(x)\ge 0$, （+1）
所以$A=\frac{\pi}{2}$. （+1）

2答: X的分布函数为:

$$F(x)=\int_{-\infty}^{x} f(u)\mathrm{d}u=\begin{cases}0, & x<0 & (+1)\\ 1-\cos(x), & 0\le x\le \frac{\pi}{2} & (+3)\\ 1, & x>\frac{\pi}{2} & (+1)\end{cases}$$

3答:

$$P\{-1\le X\le 1\}=F(1)-F(-1)=1-\cos(1)\approx 0.4597$$

(+2+2+1)

4答: $g(x)=\sin(x)$, 其反函数$h(y)=\arcsin(y)$. 对于$0\le y<1$,

$$\begin{aligned} f_Y(y) &= f(h(y))\,|h'(y)| && (+2)\\ &= \sin(\arcsin(y))\frac{1}{\sqrt{1-y^2}} && (+1)\\ &= \frac{y}{\sqrt{1-y^2}} && (+1)\end{aligned}$$

所以

$$f_Y(y)=\begin{cases}\frac{y}{\sqrt{1-y^2}} & 0\le y<1\\ 0 & \text{其它}\end{cases} \qquad (+1)$$

3 二维随机变量及其分布

- 二维联合分布函数的定义及性质;

◇ 定义：$F(x,y)=P(X\le x, Y\le y)$

◇ 性质：关于x 和y 均为单调非减；$F(-\infty,y)=F(x,-\infty)=F(-\infty,-\infty)=0, F(+\infty,+\infty)=1$; 关于$x$ 和y 均为右连续; $F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)\ge 0$

- 二维分布函数与密度函数的关系;

  ◇ 离散型：$F(x,y)=\sum_{x_i\le x, y_i\le y} p_{ij}$

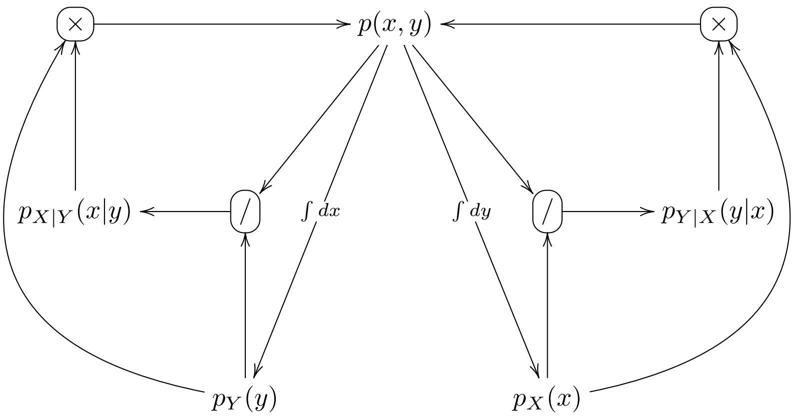
  ◇ 连续型：$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} p(u,v)dudv$

- 二维随机变量的概率计算:

  – $P\{(X,Y)\in A\}=\sum\limits_{(x_i,y_j)\in A} p_{ij}$;

  – $P\{(X,Y)\in A\}=\iint_A p(x,y)dxdy$.

- 联合分布, 边缘分布与条件分布的关系:



- 条件分布函数的定义$F_{Y|X}(y|x)=P\{Y\le y|X=x\}$;

- 条件密度计算(比如$p_{Y|X}(y|x)=\frac{p(x,y)}{p_X(x)}$, 用分布函数表示时就是$p_{Y|X}(y|x)=\frac{\partial^2 F(x,y)}{\partial x\partial y}/F'_X(x)$);

- 二维正态分布$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$的5个参数的意义及性质

  ◇ 边缘分布是正态

  ◇ 线性变换仍是正态(如，$Z=aX+bY\sim N(a\mu_1+b\mu_2,a^2\sigma_1^2+b^2\sigma_2^2+2ab\sigma_1\sigma_2\rho)$)

  ◇ 独立与不相关等价

- 如何验证两个随机变量的独立性;

  ◇ $F(x,y)=F_X(x)\cdot F_Y(y)$

  ◇ 离散型：$P(X=x_i,Y=y_i)=P(X=x_i)\cdot P(Y=y_i)$

  ◇ 连续型：$p(x,y)=p_X(x)\cdot p_Y(y)$, $p_{X|Y}(x|y)=p_X(x)$ 或者$p_{Y|X}(y|x)=p_Y(y)$

- $X+Y$, $\max(X,Y)$, $\min(X,Y)$分布的计算.

  ◇ $X+Y$的分布计算中，当X,Y 相互独立时有卷积公式

  ◇ 两个相互独立的且服从泊松分布的r.v.,其和仍然服从泊松分布，且参数为原参数之和. 两个相互独立的且服从正态分布的r.v.,其和仍然服从正态分布，且参数为原参数之和.

例3：设随机变量(X,Y)具有概率密度

$$f(x,y)=\begin{cases} Ce^{-x^2y}, & x\ge 1,\ y\ge 0, \\ 0, & \text{其他}. \end{cases}$$

1. 确定常数C;
2. 求概率$P\{X^2Y>1\}$;
3. 求条件密度$f_{Y|X}(y|x)$.

4 随机变量的数字特征

- $\mathrm{E}\,(g(X))$, $\mathrm{E}\,(g(X,Y))$的计算;

r.v.类型	$\mathrm{E}\,(g(X))$	$\mathrm{E}\,(g(X,Y))$
离散型	$\sum_i g(x_i)p_i$	$\sum_i \sum_j g(x_i,y_j)p_{ij}$
连续型	$\int g(x)p(x)dx$	$\int \int g(x,y)p(x,y)dxdy$

- 期望, 方差, 协方差以及相关系数的计算;
  ◇ 期望的计算由上述公式
  ◇ $var(X) = E[(X-EX)^2] = E(X^2) - (EX)^2$
  ◇ $cov(X,Y) = E[(X-EX)(Y-EY)]$
  ◇ $corr_{XY} = \frac{cov(X,Y)}{\sqrt{D(X)D(Y)}}$
- 期望, 方差, 协方差以及相关系数的性质;
  ◇ 期望性质P_{83}(3条)、P_{175}(2条)
  ◇ 方差性质P_{89}(3条)、P_{176}(1条)
  ◇ 协方差性质P_{178}(7条)
  ◇ 相关系数性质P_{183}(2条)

5 大数定律与中心极限定理

- Chebyshev不等式;
  ◇ $P(|X-EX| \geq \varepsilon) \leq \frac{VarX}{\varepsilon^2}$ 或者$P(|X-EX| < \varepsilon) \geq 1 - \frac{VarX}{\varepsilon^2}$
- 大数定律
  ◇ Chebyshev大数定律(条件和结论):设$X_1, \ldots, X_n, \ldots$ 两两不相关，若每个X_i的方差存在且有共同的上界，即$Var(X_i) \leq c$，对任给的$\varepsilon > 0$, $\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n E(X_i)| < \varepsilon) = 1$
  ◇ Bernoulli大数定律：样本$X_1, \ldots, X_n, \ldots$相互独立，服从两点分布，且$E(X_i) = p$, $Var(X_i) = p(1-p)$, 对任给的$\varepsilon > 0$, $\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^n X_i - p| < \varepsilon) = 1$
- 中心极限定理
  ◇ 近似计算: $\{X_i\}_1^n$独立同分布, $\mathrm{E}\,(X_i) = \mu$，$Var(X_i) = \sigma^2$, 当n很大时, $\sum\limits_{i=1}^n X_i \overset{\text{近似地}}{\sim} N(n\mu, n\sigma^2)$，即$\frac{\sum\limits_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} \overset{\text{近似地}}{\sim} N(0,1)$
  ◇ Demoivre-Laplace 定理：$\{X_i\}_1^n \overset{i.i.d}{\sim} b(1,p)$, 则$\frac{\sum\limits_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \overset{\text{近似地}}{\sim} N(0,1)$

例4：已知某地区每户居民拥有汽车辆数X的分布律为

X	0	1	2
p_k	0.3	0.6	0.1

现计划建造一拥有400户住户的居民小区, 在400户住户全部入住的假设下, 试利用中心极限定理解决如下问题:

1. 该公寓有车的住户将超过300户的概率是多少?
2. 至少需建多少车位, 才能保证每辆车都拥有车位的概率不低于0.9.

1解: 设Y是400户住户中有车的住户, 则$Y \sim b(400, 0.7)$,
$\mu = 400 \times 0.7 = 280$, $\sigma^2 = 400 \times 0.7 \times 0.3 = 84$. (1+1分)
根据De Moivre-Laplace中心极限定理, $Y \overset{\text{近似地}}{\sim} N(280, 84)$. (1分)

$$P\{Y > 300\} \approx 1 - \Phi\left(\frac{300-280}{\sqrt{84}}\right) \quad (1分)$$
$$= 1 - \Phi(2.1822) \quad (1分)$$
$$\approx 0.0146 \quad (1分)$$

或者

$$P\{Y > 300\} = P\{Y \geq 300 - 0.5\} \approx 1 - \Phi\left(\frac{300-0.5-280}{\sqrt{84}}\right) \quad (1分)$$
$$= 1 - \Phi(2.1276) \quad (1分)$$
$$\approx 0.0166 \quad (1分)$$

2解: 记$Z = \sum_{i=1}^{400} X_i$表示该小区住户拥有的车辆数. 设计话计划建造N个车位.

$$\mathrm{E}\left(X\right) = 0 \times 0.3 + 1 \times 0.6 + 2 \times 0.1 = 0.8 \quad (1分)$$
$$\mathrm{E}\left(X^2\right) = 0^2 \times 0.3 + 1^2 \times 0.6 + 2^2 \times 0.1 = 1$$
$$\mathrm{D}\left(X\right) = \mathrm{E}\left(X^2\right) - \mathrm{E}\left(X\right)^2 = 1 - 0.8^2 = 0.36 \quad (1分)$$
$$\mathrm{E}\left(Z\right) = 400 \times 0.8 = 320$$
$$\mathrm{D}\left(Z\right) = 400 \times 0.36 = 144$$
$$Z \overset{\text{近似地}}{\sim} N(320, 144) \quad (1分)$$
$$P\{Z \leq N\} \approx \Phi\left(\frac{N-320}{12}\right) \geq 0.9 \quad (1分)$$
$$\frac{N-320}{12} \geq 1.29, \left(\text{或}\frac{N-320}{12} \geq 1.28\right) \quad (1分)$$
$$N \geq 335.4800, \left(\text{或}N \geq 335.3600\right) \quad (1分)$$

需建336个车位.