概统未很好地掌握的知识点

第二章 随机变量及其分布

需要掌握的：各种常见分布的特点。

（1）若X是一个连续型随机变量，其取某一个点的概率为0；

（2）若X对任意实数都有可能取值，那么它一定是连续型的随机变量？

> - **参考资料1 奇异型随机变量**
>
>   离散型随机变量和连续型随机变量是概率论中最为常见的2种随机变量，除此之外，还存在着一种**既非离散也非连续型的随机变量**，常称之为奇异型的随机变量或**混合型随机变量**，其对应的分布称之为奇异型分布。
>
>   例子：假设随机变量 X 的绝对值不大于 1，$P\{X=-1\}=\frac{1}{8}$，$P\{X=1\}=\frac{1}{4}$，在事件 $\{|X|<1\}$ 出现的条件下，X 在区间 $(-1,1)$ 内任一子区间上取值的条件概率与该子区间长度成正比，则 X 的分布函数为 $F(x)=\begin{cases}0 & x<-1\\ \frac{5x+7}{16} & -1\le x<1\\ 1 & x\ge 1\end{cases}$。
>
> - **参考资料2 连续型随机变量的良定义**
>
>   连续型随机变量主要表现为其**分布函数是连续函数**。如果存在 $(-\infty,+\infty)$ 上的非负可积函数 $f(x)$，使 X 的分布函数 $F(x)=\int_{-\infty} f(t)\mathrm{d}t$，则称 X 为连续型随机变量，$f(x)$ 是 X 的概率密度函数。由于 $f(x)$ 是可积函数，则 $F(x)$ 一定是连续函数。只笼统地讲它的"值域为一个或若干个有限或无限区间"是不恰当的。

（3）离散型随机变量的分布函数。（从$P(X\le x)$和$P(X<x)$出发）

（4）负二项分布相关性质。

> - $P(X=r+k)=\mathrm{C}_{k+r-1}^{r-1}p^r(1-p)^k$，表示一个事件在伯努利试验中每次的出现概率是$p$，在一连串伯努利试验中，一件事件刚好在第$r+k$次试验出现第$r$次的概率。
> - $EX=\dfrac{rp}{1-p}$。

（5）几何分布相关性质。

> - $P(X=k)=p(1-p)^{k-1}$，$EX=\dfrac{1}{p}$，$DX=\dfrac{1-p}{p^2}$。

（6）二项分布$\mathcal{B}(n,p)$的参数含义。

（7）泊松分布的概念，即$P(X=k)=\dfrac{\lambda^k}{k!}e^{-\lambda}$。

（8）BO5、BO7等场景下，要注意"在四局比赛中就获胜"之类的说法，要考虑某一队不可能在一开始就连赢三场。

（9）泊松分布的可加性。

（10）分段函数用示性函数表示。

（11）设F_1,F_2为两个分布函数，其密度为f_1,f_2，则$f_1F_2+f_2F_1$也是概率密度。

> $f_1F_2+f_2F_1=(F_1F_2)'$，$\int_{-\infty}^{\infty}(F_1F_2)'dx=(F_1F_2)|_{-\infty}^{\infty}=1$，且是一个非负值函数，所以是概率密度。

（12）指数分布无后效性。即$P(X>s+t|X>t)=P(X>s)$，$P(X\le s+t|X>t)=P(X\le s)$。注意这里**指数分布隐含变量非负。**

> $P(X\le s+t|X>t)=P(X\le s),s,t>0$可以推出指数分布，记函数$g(x)=\ln(1-F(x)),x>0$，则题中式子变形为$g(t+s)=g(t)+g(s),t,s>0$，此函数方程有唯一解$g(x)=g(1)x$，令$\lambda=-g(1)$，则$F(x)=1-e^{-\lambda x}$。
>
> 离散情况的无后效性对应的是几何分布。

（13）标准正态分布的分布函数是$\Phi(x)$，密度函数是$\phi(x)$，分布函数是中心对称，密度函数是轴对称，即$\Phi(x)+\Phi(-x)=1,\phi(x)=\phi(-x)$。

（14）利用标准正态分布表来求非标准正态分布的概率。即若$X\sim N(\mu,s^2)$，则$Z=\frac{X-\mu}{s}\sim N(0,1)$，注意**分母是标准差，不是方差！**

（15）已知X的分布，求$Y=g(X)$的分布。

> 如果是单值函数，则$P(Y\le y)=P(X\sim g^{-1}(y)),\sim=\le,\ge$。
>
> - 离散情况：$P(X\sim g^{-1}(y))$是$F(g^{-1}(y))$或者$1-F(g^{-1}(y))$，再带入$y=g(x)$即可。
> - 连续情况：$P(X=g^{-1}(y))=f(g^{-1}(y))dx=f(g^{-1}(y))d[g^{-1}(y)]=h(y)dy$，密度函数$h(y)$。
>
> 如果是多值函数，比如$Y=|X|,Y=X^2$，可以拆分情况来讨论。

错题速览

> **例题1**：设某种昆虫单只每次产卵的数量服从参数为 λ 的Poisson分布，而每个虫卵能孵出幼虫的概率均为 $p(0<p<1)$ 且相互独立。分别以 Y 和 Z 记一只昆虫一次产卵后幼虫和未能孵出幼虫的虫卵的个数。试问 Y 和 Z 分别服从什么分布? 它们是否相互独立?
>
> 以 X 记一只昆虫产卵的个数，则对任意整数 m，$n\geq 0$，由
>
> $$\begin{aligned}\mathbb{P}(Y=m,Z=n)&=\mathbb{P}(Y=m,Z=n\mid X=m+n)\mathbb{P}(X=m+n)\\&=\binom{m+n}{m}p^m(1-p)^n\mathrm{e}^{-\lambda}\frac{\lambda^{m+n}}{(m+n)!}\\&=\mathrm{e}^{-\lambda p}\frac{(\lambda p)^m}{m!}\cdot\mathrm{e}^{-\lambda(1-p)}\frac{[\lambda(1-p)]^n}{n!}\end{aligned}$$
>
> 可知 Y 和 Z 分别服从参数为 λp 和 $\lambda(1-p)$ 的Poisson分布，且相互独立.

> **例题2**：一位篮球运动员练习投篮 100 次，且已知他前两次只投进了一次。从第 3 球开始，假设他每次投篮的命中率为其前面所投进球的比率（比如他前五次投进了四个球，则第六次他的投篮命中率为 $\frac{4}{5}$）。求他最终在这 100 次投篮中投进次数的分布律。
>
> 递推：
>
> $$P_n(k)=P_{n-1}(k)P(n,0)+P_{n-1}(k-1)P(n,1)$$
>
> $P_n(k)$表示投n次球，中k次的概率；$P(n,I[A])$表示第n次实验事件A是否发生的概率。
>
> 结论：服从$\{1,2,\ldots,99\}$的均匀分布，也可以用数学归纳法证明。

> **例题3**：设随机变量 X 只在区间 $(0,1)$ 内取值，且其分布函数 $F(x)$ 满足：对任意 $0\leq a<b\leq 1$，$F(b)-F(a)$ 的值仅与差 $b-a$ 有关。试证明 X 服从 $(0,1)$ 上的均匀分布。
>
> 先证对任意有理数 $x\in(0,1)$，我们有 $F(x)=x$。事实上，若 $x=\frac{m}{n}$，其中 $m<n$ 为正整数，则由题目条件可知 $F(0)=0$，$F(1)=1$，且
>
> $$F\left(\frac{1}{n}\right)-F(0)=F\left(\frac{2}{n}\right)-F\left(\frac{1}{n}\right)=\cdots=F(1)-F\left(\frac{n-1}{n}\right).$$
>
> 易知上式中各项之和为 1，故每项均等于 $\frac{1}{n}$. 从而对任意 $m<n$，有 $F\left(\frac{m}{n}\right)=\frac{m}{n}$. 再由连续函数的右连续性可知 $F(x)=x$ 对所有 $(0,1)$ 中的无理数也成立. 这就 证明了 $X\sim U(0,1)$.

> **例题4**：设连续型随机变量 X 的分布函数为
>
> $$F(x)=a+b\arctan x,\quad -\infty<x<\infty.$$
>
> 问 $\mathbb{E}X=0$ 是否成立? 不成立，因为$\int_{-\infty}^{\infty}|x|f(x)\mathrm{d}x=\infty$。

> **例题5**：设随机变量 X 的密度函数为 $f(x)=\frac{1}{a}x^2$，$0<x<3$，令 随机变量
>
> $$Y=\begin{cases}2, & X\leq 1\\ X, & 1<X<2\\ 1, & X>2\end{cases}$$
>
> 求概率 $\mathbb{P}(X\leq Y)$。
>
> **这里主要的错点是当$Y=1$时隐含有条件$X>2$，从而不可能有$X\leq Y$**。由全概率公式可解。

第三章 多维随机变量及其分布

这一章的所有公式都要看。

（1）对离散型随机变量联合分布的理解。即已知X,Y的边缘分布，和约束条件$h(X,Y)=0$，求联合分布。采取填表格的方法。

（2）泊松积分：$\int_0^{\infty}e^{-t^2}dt=\frac{\sqrt{\pi}}{2},\int_{-\infty}^{\infty}e^{-t^2}dt=\sqrt{\pi}$。

错题速览

> **例题1**：概统习题集第26页第21题。
>
> - 这题的启示：计算条件概率$P(X=x|Y=y)$一定要按照公式来计算！！

> **例题2**：概统习题集第27页第26题。

> **例题3**：概统习题集第27页第28题。（特定值条件期望）

> **例题4**：概统习题集第28页第31题。（第二次做仍然感觉有点问题）

> **例题5**：假设有 $n(n\geq 3)$ 个不同的盒子与 m 个相同的小球，每个小球独立地以概率 p_k 落入第 k 个盒子 $(k=1,2,\ \cdots,\ n)$. 分别以 X_1，X_2，$\cdots$，X_n 表示落入各个盒子 的球数. 试求
>
> (1) $(X_1,\ X_2,\ \cdots,\ X_n)$ 的联合分布;

(2) X_k 的边缘分布，其中 $k=1,2, \cdots, n$;

(3) (X_1, X_2) 的边缘分布;

(4) 在 $X_1=m_1$ 的条件下 $(X_2, \cdots, X_n)$ 的条件分布;

(5) $\mathbb{E}[X_2|X_1=k]$和$\mathrm{Var}(X_2|X_1=k)$;

(6) $\mathbb{E}[X_1+X_2]$和$\mathrm{Var}(X_1+\ldots X_k), k=2,\ldots,n-1$。

解答:

(1) $(X_1, X_2, \cdots, X_n)$ 服从多项分布 $M(m; p_1, p_2, \cdots, p_n)$:

$$\mathbb{P}(X_1=m_1,\cdots,X_n=m_n)=\frac{m!}{m_1!m_2!\cdots m_n!}p_1^{m_1}p_2^{m_2}\cdots p_n^{m_n},$$

其中 $m_1, m_2, \cdots, m_n$ 是任一使得 $m_1+m_2+\cdots+m_n=m$ 的非负整数列.

(2) $X_k \sim B(m, p_k)$.

(3) $\mathbb{P}(X_1=m_1, X_2=m_2)=\frac{m!}{m_1!m_2!(m-m_1-m_2)!}p_1^{m_1}p_2^{m_2}(1-p_1-p_2)^{m-m_1-m_2}$, 其中 m_1, m_2 是任一使得 $m_1+m_2 \leq m$ 的非负整数.

(4) 条件分布为

$$\begin{aligned}&\mathbb{P}(X_2=m_2,\cdots,X_n=m_n \mid X_1=m_1)\\=&\frac{(m-m_1)!}{m_2!\cdots m_n!}\left(\frac{p_2}{1-p_1}\right)^{m_2}\cdots\left(\frac{p_n}{1-p_1}\right)^{m_n},\end{aligned}$$

其中 $m_2, \cdots, m_n$ 是任一使得 $m_1+m_2+\cdots+m_n=m$ 的非负整数列.

(5) $\mathbb{E}[X_2 \mid X_1=k]=\frac{(m-k)p_2}{1-p_1}, \quad \mathrm{Var}(X_2 \mid X_1=k)=\frac{(m-k)p_2(1-p_1-p_2)}{(1-p_1)^2}$

(6) $\mathbb{E}[X_1+X_2]=m(p_1+p_2)$,

$\mathrm{Var}(X_1+\cdots+X_k)=m\sum_{i=1}^k p_i(1-p_i)-2m\sum_{i=1}^{k-1}\sum_{j=i+1}^k p_ip_j, k=1,\ldots,n.$

例题6：连续地郑一颗均匀的骰子，直到出现点数大于 2 为止，以 X 表示郑骰子的次数，以 Y 表示最后一次郑出的点数。

（1）求二维随机变量 (X, Y) 的联合分布以及边缘分布;

（2）问 X 和 Y 是否相互独立。

解答:

（1）$\mathbb{P}(X=i, Y=j)=\frac{1}{6}\times\left(\frac{1}{3}\right)^{i-1}, i=1,2, \cdots, j=3,4,5,6.$

$\mathbb{P}(X=i)=\frac{2}{3}\times\left(\frac{1}{3}\right)^{i-1}, i=1,2, \cdots, \quad \mathbb{P}(Y=j)=\frac{1}{4}, j=3,4,5,6$

（2）X, Y 独立。

第四章 数字特征

（1）如果X_i相互独立，则$E\prod X_i=\prod EX_i$。

（2）非负随机变量，$EX=\sum_{n=1}^{\infty}P(X\geq n)$，$EX=\int_0^{\infty}P(X\geq x)dx$。

$EX=E[\int_0^X dx]=E[\int_0^{\infty}I(X>x)dx]=\int_0^{\infty}EI(X>x)dx=\int_0^{\infty}P(X>x)dx$。

（3）$\int Pol(x)e^{bx}dx$，$b>0$时可以用分部积分法得到递推公式，$b<0$时利用伽马函数。

（4）利用$EX^2=E[X(X-1)]+EX$可以简化一些含阶乘和导数形式的概率密度函数计算。

（5）注意相关系数$\rho=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}$上面的是二次方，用$\sigma_{XY}=EXY-EXEY$，下面的是标准差，是$\sqrt{EX^2-(EX)^2}\sqrt{EY^2-(EY)^2}$。

（6）一般不相关不等于独立，但是对正态分布来说，不相关就等于独立。

（7）只要两个变量的协方差为0（不相关），就有$Var(X+Y)=VarX+VarY$。

（8）全期望公式

错题速览

例题1：设X为随机变量，则使得$E|X-a|$达到最小的常数$a=\text{中位数}[X]$。使得$E(X-a)^2$达到最小的常数$a=EX$。

例题2：将 n 个球依次放入 n 个盒子中，假设每个球放入每个盒子中是等可能的，试求放完后空盒子个数的期望，以及当 $n\to\infty$ 时空盒子的平均比例。

记$X_i=I[\text{第i个盒子无球}]$，则$X=\sum X_i$，$P(X_i=1)=\frac{(n-1)^n}{n^n}=(1-\frac{1}{n})^n$，则$EX=\sum EX_i=n(1-\frac{1}{n})^n$。

> **例题3**：某零食厂商设计了一种营销策略，即在产品中放入一套有趣的卡片。假设这套卡片由 $n=12$ 张不同的卡通人物头像组成，且在每袋零食中随机放入其中一张。某人想聚齐这套卡片，设他一共需要买 X_n 袋该零食。试求：
>
> （1） $\mathbb{E}[X_n]$；（2） $\lim_{n\to\infty}\frac{\mathbb{E}[X_n]}{n\ln n}$。

> **例题4**：设 $N(t)$ 是一个依赖于变量 t 的随机变量，对 $t>0, N(t)$ 服从分布为
>
> $$\mathbb{P}(N(t)=k)=e^{-\lambda t}\frac{(\lambda t)^k}{k!},\quad k=0,1,2,\ldots,$$
>
> 设 T 是一个均值为 a，方差为 $b>0$ 的非负随机变量。
>
> 求 $(1)\ \mathrm{Cov}(T,N(T));(2)\ \mathrm{Var}(N(T))$。

> **例题5**：设随机变量 X 与 Y 分别是均值为 $1/\lambda$ 和 $1/\mu$ 的独立的指数随机变量。
>
> (1) 证明在条件 $X>Y$ 下, 随机变量 $\min\{X,Y\}$ 和 $X-Y$ 是相互独立的.
>
> (2) 证明对任意正数 $c>0$，$\mathbb{E}[\min\{X,Y\}\mid X>Y+c]=\mathbb{E}[\min\{X,Y\}\mid X>Y]=\mathbb{E}[\min\{X,Y\}]=\frac{1}{\lambda+\mu}.$

第五章 参数估计

（1）判断是否是统计量：只和样本数据有关的是统计量，和未知参数有关的不是统计量。

（2）正态分布可加性：$X\sim N(\mu_1,s_1^2),Y\sim N(\mu_2,s_2^2)$，则$aX+bY\sim N(a\mu_1+b\mu_2,a^2s_1^2+b^2s_2^2)$。

（3）极大似然估计正态分布方差时，变量是σ^2而不是σ。

错题速览

> **例题1**：调查 50 个人对某件事情是 (1) 否 (0) 支持，假设每个人对该事情支持的可能性为 p，各人之间相互独立。则总体分布是什么? 若其中 10 个人的调查结果为 $x_1,\ldots,x_{10}$ (其中 x_i 只取 0 或 1)，则抽样分布是什么?
>
> 总体分布是$\mathcal{B}(1,p)$；抽样分布是$P(X_i=x_i)=p^{\sum x_i}(1-p)^{10-\sum x_i}$。

> **例题2**：设总体是参数为λ的指数分布，X_i是一个简单样本，求样本均值$\bar{X}$的分布。（证明$\sum_{i=1}^n 2\lambda X_i$服从$\chi^2_{2n}$分布）

> **例题3**：概统习题集50页第7题。