概率统计B

Probability and Statistics

张思容
zhangsirong@buaa.edu.cn

数学与系统科学学院，北京航空航天大学
School of Mathematics and System Sciences, BUAA

November 15, 2014

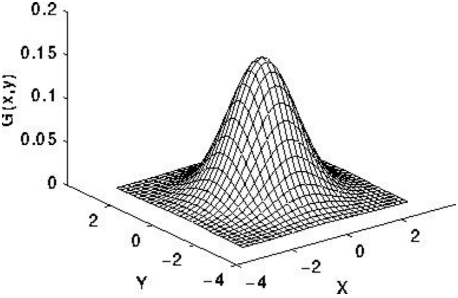
Review：回顾

RECALL:

- 随机变量与分布函数
- 常见离散分布：两点分布，二项分布，几何分布，泊松分布；
- 常见连续分布：均匀分布，正态分布，指数分布；
- 分布的特征与复合：

TODAY

- 随机向量：
- 常见二维分布：



二维正态分布(高斯分布)

随机向量的定义

Definition (随机向量)

随机向量$\vec{X}$是定义在样本空间Ω上的实值多元函数:即每一个样本点ω,$\vec{X}(\omega)$是个向量。 通常用$\vec{X}, \vec{Y}, \vec{Z}$或$\vec{\xi}, \vec{\eta}$等表示。

记n是向量的维数：一般考察二维向量，记为$\vec{X} = (X, Y)$.

EXAMPLE (身体指标)

$\vec{I} = (H, W)$。H是身高(m)，W是体重(kg)；假设H, W为正态分布；身体状况是怎么样的？身高与体重的关系？BMI指标$= W/H^2$。

EXAMPLE (多项分布)

掷骰子n次，结果可以为$(1-6)$,记其中为$X_i = n_i$为结果为i的次数；
则$\vec{X} = (X_1, X_2, \ldots, X_6)$是六维随机向量 ,其分布为
$P(X_1 = n_1, X_2 = n_2, \ldots, X_6 = n_6) = \frac{n!}{n_1!n_2!\ldots n_6!} p_1^{n_1} \ldots p_6^{n_6}$. 其中$\sum n_i = n$.
一般的称为多项分布。*(***边缘分布是二项分布)*

联合分布函数

Definition (联合累积分布)

给定二维随机向量(X,Y),定义 $F(x,y)=P(X\le x)\cap P(Y\le y)=P(X\le x, Y\le y)$。称为随机向量$(X,Y)$的分布函数，或随机变量$X,Y$的联合分布。

- ***分布函数可以决定任意事件(集合)的概率$P(\{(x,y)\in D\})$。(测度论)
- 分布函数的充分必要条件:
  1. $0\le F(x,y)\le 1. F(x,-\infty)=F(-\infty,y)=F(-\infty,-\infty)=0, F(+\infty,+\infty)=1.$
  2. $F(x,y)$对x或y单调非减
  3. $F(x,y)$对x或y右连续; ***
  4. 任意$x_1<x_2, y_1<y_2$，有$F(x_2,y_2)+F(x_1,y_1)-F(x_1,y_2)-F(x_2,y_1)\ge 0.$
- 分类：二维离散随机向量，二维连续随机向量，混合型随机向量；

二维离散分布

Definition (二维分布矩阵)

给定(X, Y)二维随机变量，取值至多可数；记$p_{i,j} = P(X = x_i, Y = y_j)$，称$p_{i,j}$为二维离散分布(矩阵)；$\sum_i \sum_j p_{i,j} = 1$.

EXAMPLE (人口调查)

设一个家庭至多三个小孩，$X = i, Y = j$对应男孩与女孩的个数。分布如下：

$i:j$	0	1	2	3	p_i
0	0.15	0.10	0.0875	0.0375	0.375
1	0.10	0.175	0.1125	0	0.3875
2	0.0875	0.1125	0	0	0.2
3	0.0375	0	0	0	0.0375
q_j	0.375	0.3875	0.2	0.0375	1

注意：令$p_i = \sum_j p_{i,j}, q_j = \sum_i p_{i,j}$为矩阵的行和与列和；
则$\sum p_i = \sum q_i = 1$ 称为边缘分布!

二维连续分布

Definition (二维连续随机向量)

给定(X,Y)二维随机变量，如果存在二元非负可积函数$f(x,y)$使得，任意区域D，有 $P((X,Y)\in D)=\iint_D f(x,y)dxdy$,称$(X,Y)$是连续随机向量。$f(x,y)$称为联合密度函数。 特别: $\iint_{R^2} f(x,y)dxdy=1.$

- 分布函数 $F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)dxdy.$
- 定理: 连续可微的联合分布函数有联合密度函数，所以是连续随机向量。 即 $\frac{\partial^2 F}{\partial x \partial y}=f(x,y).$
- 均匀分布: $f(x,y)=\frac{1}{Area(D)},(x,y)\in D.$

  约会问题: 两个人约定见面。假设在一小时内随机到达， 求第一个人等十分钟以上的概率。

  解答: 25/36

  $P(X+10<Y)=\int_D f(x,y)dxdy=\int_0^{50}dx\int_{x+10}^{60}(1/60)^2dy=25/72$

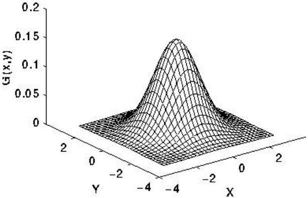
二维正态分布 bivariate normal distribution

EXAMPLE (二维正态分布)

给定(X, Y)二维随机变量，称为方程$N(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2, \rho)$的二维正态分布，

如果其概率密度函数是:

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



- 参数μ_1, μ_2是均值，σ_1, σ_2是差异，
- ρ是 相关系数。$|\rho| \leq 1$.

边缘分布

Definition (边缘分布函数)

给定二维随机向量及其分布函数$F(x,y)$, 可以得到X, Y是随机变量，其分布函数$F_X(x) = F(x,\infty), F_Y(y) = F(\infty, y)$. 称为边缘分布函数。

- 随机向量的投射得到坐标随机变量: 对应的分布函数即边缘分布。
- 离散情形: 边缘分布列即矩阵的行和或列和。
- 连续情形: 存在对应的边缘密度函数$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$, $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$.
- 联合分布决定边缘分布，反之不能。

分布函数的计算

EXAMPLE (P52 例2)

$f(x, y) = ae^{-2y}, 0 \le x \le 2, y > 0.$

- 确定常数a.
- 求分布函数$F(x, y)$.
- 求概率 $P(|Y| \le X)$.
- 求X, Y的边缘分布函数;
- 求X, Y的边缘密度函数;

边缘分布与联合分布

Proposition

(二维)正态分布的边缘分布还是正态分布。设 $(X, Y) \sim N(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2, \rho)$,则边缘分布$X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$。

注：配方求积分可得。

- 边缘分布不能得到联合分布($\rho =?$).
- 一般的联合分布与边缘分布的类型不一定一样。
  圆盘上的均匀分布的边缘分布不是均匀分布。

如何利用边缘分布得到联合分布？ 条件分布或独立性!

作业

北航教材:
P67 习题三. 2,4,6,12,13,15 补充： 计算习题4，6的边缘分布律。

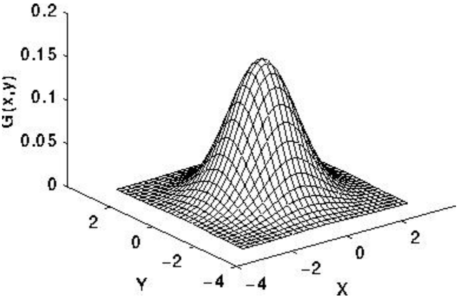
Review: 回顾

RECALL:

- 随机向量与联合分布函数
- 离散分布矩阵，连续分布密度；
- 常见随机向量分布：多项分布，均匀分布，正态分布；
- 边缘分布及边缘密度函数；

TODAY

- 条件分布及计算；
- 随机变量的独立性和相关性；



二维正态分布(高斯分布)

条件概率与随机变量

- 条件概率公式：$P(B|A) = \frac{P(AB)}{P(A)}$
- *** 关于事件A的随机变量条件概率公式：$P(X \in B|A) = \frac{P(X \in B \cap A)}{P(A)}$.
  (条件)随机变量: $P(X \in B|A) = \frac{\int_{A \cap B} f(x)dx}{P(A)}$. 特别密度函数$f_{X|A} = f(x)/P(X \in A)$.
- 关于随机变量Y的条件概率公式：$P(X = x|Y = y) = ?$
  $P(X = x|Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)}$.
  说明：希望上面对于任意x, y都成立！
  问题: 连续情形$P(X = x) = P(Y = y) = 0$.

条件分布：离散情形：

Definition (离散条件分布)

给定(X,Y)二维离散随机变量，记$p_{i,j}=P(X=x_i,Y=y_j)$，记$p_{.j}=\sum_i p_{i,j}$,$p_{i.}=\sum_j p_{i,j}$，则$P(X=x_i|Y=y_j)=p_{i,j}/p_{.j}$,称为条件$Y=y_j$下的条件分布律。类似有条件$X=x_i$下的条件分布律。

EXAMPLE (人口调查)

设一个家庭至多三个小孩，$X=i,Y=j$对应男孩与女孩的个数。分布如下：

$i:j$	0	1	2	3	$p_{i.}$
0	0.15	0.10	0.0875	0.0375	0.375
1	0.10	0.175	0.1125	0	0.3875
2	0.0875	0.1125	0	0	0.2
3	0.0375	0	0	0	0.0375
$p_{.j}$	0.375	0.3875	0.2	0.0375	1

有一个女孩的家庭的男孩分布律：$P(X=i|Y=1)=p_{i1}/p_{.1}$

条件分布：连续情形

设(X,Y)是连续的二维随机向量，密度函数为$f(x,y)$.

- 局部逼近：$P(x \le X \le x+\delta, y \le Y \le y+\delta) \approx f(x,y)\delta^2$
- 边缘分布: $P(y \le Y \le y+\delta) \approx f_Y(y)\delta$
- 条件分布：$P(x \le X \le x+\delta | y \le Y \le y+\delta) \approx \frac{f(x,y)}{f_Y(y)}\delta$

Definition (连续条件分布)

给定(X,Y)连续二维随机向量，密度函数$f(x,y)$.
则$f_{X|Y}(x) = f(x,y)/f_Y(y)$,称为条件$Y=y$下的条件概率密度。
对应有条件分布函数$F_{X|Y}(x) = \int_{-\infty}^{x} f(u,y)du/f_Y(y)$.

有乘法公式 $f(x,y) = f_{X|Y}(x)f_Y(y)$，可以推广到多个随机变量。

连续二维分布的例子

EXAMPLE (圆盘上均匀分布)

已知圆盘上均匀分布的概率密度为$f(x,y)=1/(\pi R^2), x^2+y2\leq R^2$.
则边缘密度$f_Y(y)=\frac{2}{\pi R^2}\sqrt{R^2-y^2}, |y|\leq R$.
条件密度函数$f_{X|Y}(x)=\frac{1}{2\sqrt{R^2-y^2}}, |x|\leq\sqrt{R^2-y^2}$.
结论：边缘分布非均匀分布但条件分布是均匀分布*(*小区间*)*。

Proposition (正态分布的边缘与条件分布)

*(*二维*)*正态分布的边缘分布还是正态分布。设
$(X,Y)\sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2,\rho)$,则边缘分布$X\sim N(\mu_1,\sigma_1^2)$,
$Y\sim N(\mu_2,\sigma_2^2)$。 特别条件分布还是正态分
布。$X|Y\sim N(\mu_1+\rho\sigma_1/\sigma_2(y-\mu_2),(\sigma_1\sqrt{1-\rho^2})^2)$.

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

独立性

Definition (独立性)

随机变量X, Y称为独立的，如果任意x, y,事件$\{X \leq x\}$与$\{Y \leq y\}$是独立的。即$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$. 可以推广到$n$个随机变量，称为独立序列。

X, Y相互独立等价于$F(x, y) = F_X(x)F_Y(y)$等价于$f(x, y) = f_X(x)f_Y(y)$等价于$p_{ij} = P_{i.}p_{.j}$. 一般的，随机向量的各个坐标分量独立可以得到简化的联合分布函数。

Theorem (独立的性质)

- 常数与任一随机变量独立；
- X, Y独立，则任意事件$X \in A$与$Y \in B$都独立；
- X, Y独立，任意函数$U = g(X), V = f(Y)$,则U, V是互相独立的；

二维正态分布的坐标随机变量相互独立的充要条件$\rho = 0$.

随机向量的期望

Definition (二维随机向量的期望)

设随机向量为$\vec{X} = (X, Y)$,则$\vec{X}$的期望是 (EX, EY)，记为$E\vec{X}$.类似有n维向量的期望。

Theorem (随机向量函数的期望)

设$\vec{X}$是随机向量，如果为连续随机向量，分布密度为$f(x_1, x_2, \ldots, x_n)$,则$Y = g(\vec{X})$的期望是
$EY = \int\int g(\vec{X}) f(x_1, \ldots, x_n) dx_1 \ldots dx_n$. 对于离散随机向量类似。

推论:

- +++线性:$E(\sum a_i X_i) = \sum a_i E(X_i)$.
- +++(独立)乘法：设X, Y独立，则$EXY = EXEY$. 可推广到n维。

复合函数期望的应用

Corollary (复合函数的方差)

- $VarX = DX = EX^2 - (EX)^2$;
- 线性变换: $Var(aX + b) = a^2 Var(X)$.
- 设X, Y独立,则$Var(X + Y) = VarX + VarY$.

EXAMPLE (配对问题)

n个信放在n个信封，求平均有几封正确放好。

解答： 设$X_i =$第i个信封放对。则$EX_i = 1/n$,
设Y为所有放对的信封数， $Y = X_1 + \cdots + X_n, EY = E(X_1 + \cdots + X_n) = 1$.即 平均一封信放对。

协方差

Definition (协方差)

设随机变量X, Y的期望存在,定义X, Y的协方差为$cov(X, Y) = E(X - EX)(Y - EY)$. 可记为$\sigma_{XY}$.
计算公式: $cov(X, Y) = E(XY) - EXEY$

- 对称性: $cov(X, Y) = cov(Y, X)$,特别$Cov(X, X) = Var(X)$
- (双)线性: $cov(aX, Y) = acov(X, Y)$
- 求和: $cov(\sum X_i, \sum Y_j) = \sum_i \sum_j cov(X_i, Y_j)$ 特别 $Var(\sum X_i) = Var(X_i) + \sum \sum_{i \neq j} cov(X_i, X_j)$

推广: n维随机向量期望$E(\vec{X}) = \vec{\mu}$,
协方差矩阵: $\Sigma = (\sigma_{X_i X_j})_{n \times n}$, $\Sigma = E(\vec{X} - \vec{\mu})^\tau (\vec{X} - \vec{\mu})$.
高维正态分布$f(\vec{X}) = Cexp^{-\frac{1}{2}(\vec{X} - \vec{\mu})^\tau \Sigma^{-1} (\vec{X} - \vec{\mu})}$

相关系数

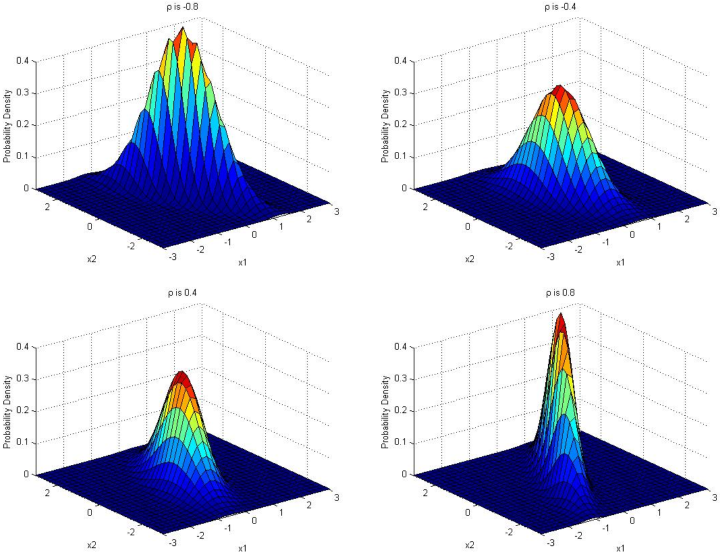
Definition (相关系数)

设随机变量X, Y的期望和方差存在,定义X, Y的相关系数为$\rho(X, Y) = \frac{cov(X,Y)}{\sqrt{VarXVarY}}$. 可记为$\rho_{XY}$.

计算公式: 引入标准化随机变量
$U_X = (X - \mu_X)/\sigma_X, U_Y = (Y - \mu_Y)/\sigma_Y$, 则 $\rho(X, Y) = E(U_X U_Y)$

- 相关系数: $-1 \le \rho(X, Y) \le 1$,即柯西许瓦茨不等式。
- 如果$\rho(X, Y) = 0$,称X, Y互不相关; 一般地如果$cov(X, Y) = 0$,称X, Y互不相关。
  独立必然不相关,反之不一定. P113 例5.
- $|\rho| = 1 \Longleftrightarrow P(Y = aX + b) = 1$. 注: $VarX = 0 \to X \equiv Const$.
  相关指线性相关: 不相关不一定没关系,更不一定独立。
  例子: $X \sim N(0, 1), Y = X^2$; $cov(X, Y) = 0$.
- 二维正态分布$N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$,$\rho$即相关系数,
  特别二维正态分布: 不相关等价于独立。

例子：二维正态分布



矩母函数***

Definition (矩母函数)

设随机变量X的期望存在,定义X的矩母函数 为$M_X(s) = E(exp^{sX})$. 可记为$M(s)$.

- 离散情形: $M(s) = \sum e^{sx_i} p(x_i)$.
- 连续情形: $M(s) = \int_{-\infty}^{\infty} e^{sx} f(x) dx$
- 对应有: Z变换: $M(z) = \sum z^i p(i)$.
  拉普拉斯变换 $L(f) = \int_0^{\infty} e^{-sx} f(x) dx$.
- 结论: 随机变量的矩即矩母函数在零点的导数。
  $EX^n = \frac{d^n M(s)}{ds^n}|_{s=0}$.
  注: 存在逆变换, 唯一决定随机变量的分布函数。

例子: 指数分布 $M(s) = \frac{\lambda}{\lambda - s}, s < \lambda$.
$EX = \frac{1}{\lambda}, DX = \frac{1}{\lambda^2}$.

作业

北航教材:
P68 习题三. 15， 17， 19， 20, 25, 29;
P90 习题五. 13, 18, 22, 23.

Review：回顾

RECALL:

- 随机向量及其分布
- 边缘分布与条件分布;
- 常见分布：均匀分布与二维正态分布;
- 独立性和相关性

TODAY

- 随机向量的函数;
- 常见二维随机向量函数的分布;

二维正态分布(高斯分布)

随机向量的函数

Remark (随机向量的函数)

随机向量$\vec{X}$是定义在样本空间Ω上的实值N元向量函数，$g: R^n \to R$的实值函数，记$Y = g(\vec{X})$,则Y是定义在Ω上的一个随机变量。一般考察二维向量，记为$Z = g(x, y)$.

例子(身体指标):H是身高(m)，W是体重(kg). 假设(H, W)为二维正态分布, 则 BMI指标$= W/H^2$ 是一个随机变量(分布是什么？)

- 一般复杂g得到的分布函数比较复杂；我们仅关心简单情形或分布函数的特征；
- 应用: 到达过程(随机变量求和); 失效模型(随机变量的极小值)；延迟模型(随机变量极大值);
- 计算: $F_Z(z) = P(g(x, y \leq z)) = \int\int_{g(x,y)\leq z} f(x, y)dxdy$
  类似二重积分的累次积分方法: $F_Z(z) = \int_{-\infty}^{z}\int_{g(x,y)=u} f(x, y)dl$ (后面是曲线积分). ***来源于二维变换$U = g_1(X, Y), V = g_2(X, Y)$,对应于二重积分的坐标变换公式。

离散随机向量的变换

- 一维情形: $Z = g(X)$, $p_Z(z) = \sum_{g(x)=z} p(x)$.
- 二维情形: $p_Z(z) = \sum_{g(x,y)=z} p(x,y)$. 沿曲线$g(x,y) =$常数，求和。(等高线!)

EXAMPLE (泊松分布求和:例3)

设$X \sim \mathcal{P}(\lambda_1)$, $Y \sim \mathcal{P}(\lambda_2)$, 则$Z = X + Y$服从$\mathcal{P}(\lambda_1 + \lambda_2)$.

注解: 两个到达过程的和即对应强度的相加的一个到达过程。

一般连续情形

- 一维情形:

Theorem (复合函数的密度函数)

设X的密度函数$f(x), Y=g(X)$;Y的取值范围为D, 对任意$y\in D$, 设其原像可以写成有限点(即$\sharp(g^{-1}(y))$有限).每一个可以写成一个逆函数$h_i(y)=x_i$; 则 Y的密度函数为 $f_Y(y)=\sum_{i=1}^{n}f(h_i(y))|h_i'(y)|$。

注记：一般地在每个单调区间内有一个逆函数。
参考作业解答:P90.13

- 二维情形：$F_Z(z)=\int\int_{g(x,y)\le z}f(x,y)dxdy$. 沿曲线$g(x,y)=$常数，求积分。
  特别地：边缘分布密度 $Z_1=X$, $f_{Z_1}(z)=\int_{-\infty}^{\infty}f(z,y)dy$.
  $Z_2=Y$, $f_{Z_2}(z)=\int_{-\infty}^{\infty}f(x,z)dx$.
  一般的有***$F_Z(z)=\int_{-\infty}^{z}\int_{g(x,y)=u}f(x,y)dl$

随机变量的和与差

随机变量的和: $Z = X + Y$, 设(X, Y)联合分布密度为$f(x, y)$,
则$P(Z \leq z) = \int_{X+Y\leq z} f(x, y)dxdy$

- 公式1: $f_Z(z) = \int_{-\infty}^{\infty} f(x, z - x)dx$;
- 公式2: $f_Z(z) = \int_{x+y=z} f(x, y)dy$ 沿直线积分。
- 特别X, Y独立有 $f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z - x)dx$
  通常称等式右边为卷积 $f_X(x) * f_Y(y)$.
- 例子: X, Y为独立的均匀分布，则$Z = X + Y$是三角分布。(参见信号处理).

EXAMPLE (约会问题: 随机变量的差)

设两个人约定八点见面，每人迟到的时间服从指数分布$\epsilon(\lambda)$. 设迟到时间相互独立，求$Z = X - Y$的分布函数。

答案: 分布密度函数$f_Z(z) = \frac{\lambda}{2}e^{-\lambda|z|}$，称为双边指数分布。

**随机变量的乘与除

设(X, Y)联合分布密度为$f(x, y)$,

- $Z = XY$, 则 $f_{XY} = \int_{-\infty}^{\infty} \frac{1}{|x|} f(x, z/x) dx$
- $Z = Y/X$, 则 $f_{Y/X} = \int_{-\infty}^{\infty} |x| f(x, xz) dx$
- 特别如果$Y > 0$,则$\log Y$是随机变量。常见有对数正态分布$\log Y = X \sim N(\mu, \sigma^2)$.
  以上乘除公式可以用加减法代替。

Remark (概念辨析)

一般的随机向量的有很多复合函数(从而导出很多随机变量)。 上面随机变量的加减乘除是：我们把复合函数得到的随机变量看成(坐标)边缘随机变量的运算的结果。 但知道边缘分布并不能知道复合函数的分布函数!!!
应用中我们考虑很多随机变量的关系(及运算)，则假定不同随机变量的独立性是有用的，甚至必须的。

极大极小随机变量

Definition (随机变量排序)

已知$X_1, X_2, \ldots, X_n$, 可以排序为
$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$,称$X_{(i)}, 1 \le i \le n$为次序统计量。

- 设X_i是独立同分布的；分布函数为$F(x)$;
- 最大(极大)值分布$F_{max}(x) = P(X_1 \le x, X_2 \le x, \ldots, X_n \le x) = P(X_1 \le x)P(X_2 \le x)\ldots P(X_n \le x) = F(x)^n$
- 最小(极小)值分布$F_{min}(x) = 1 - (1 - F(x))^n$.

EXAMPLE (灯泡现象)

教室的灯泡越多，则越容易坏。

解释： 假设每个灯泡服从指数分布$\mathcal{E}(\lambda)$. 共有n个灯泡， 第一个坏掉灯泡的概率分布即极小分布. $F_{min}(x) = 1 - (1 - F(x))^n = 1 - e^{-n\lambda x}$. 显然$n$越大，$F_{min}(x)$越大，寿命越短。

简单随机变量的和

$S = X_1 + X_2 + \cdots + X_n$. 假设随机变量互相独立,
则$F(S) = F(X_1)F(X_2)\cdots F(X_n)$.
离散情形:

- X_i是两点分布$B(1, p)$，则S是二项分布$B(n, p)$.
- X_i是几何分布, 则S是负二项分布$NB(n, p)$.
- X_i是泊松分布$\mathcal{P}(\lambda_i)$, 则S是泊松分布$\mathcal{P}(\sum_i \lambda_i)$.

连续情形:

- X_i是均匀分布，则S是？.
- X_i是指数分布, 则S是伽马分布$\Gamma(n, \lambda)$.
- X_i是正态分布, 则S是正态分布$N(\sum \mu_i, \sum \sigma_i^2)$.

可以推广到随机变量的线性组合。 特别正态分布的线性组合还是正态分布!
***可以推广到可数个随机变量的和。

例子1

EXAMPLE (债价格)

假设$S(n)$是1万元债在第n周的价格。一个流行的价格演化模型假定$S(n)/S(n-1)$是互相独立的对数正态分布$\log N(\mu, \sigma^2)$。假定$\mu = 0.0165, \sigma = 0.0730$, *(a)* 计算连续两周每周价格都上涨的概率。*(b)* 计算两周后价格上涨的概率。

解答:

- 设$Z \sim N(0,1), \log(S(n)/S(n-1)) = X \sim N(\mu, \sigma^2)$,则 任一周内上涨的概率$P(S(n)/S(n-1) > 1) = P(X > 0) = P((X-\mu)/\sigma > -\mu/\sigma)$ 即$P(Z > -0.0165/0.0730) = P(Z < 0.2260) = 0.5894$. 由独立性, 连续两周都上涨的概率$P = (0.5894^2) = 0.3474$
- 两周后上涨即$P(S(2)/S(0) > 1) = P(S(2)/s(1)s(1)/s(0) > 1) = P(X_1 + X_2 > 0)$. 注意$X_1, X_2 \sim X = N(\mu, \sigma^2)$, $X_1 + X_2 \sim N(2\mu, 2\sigma^2)$, 计算有$P(X_1 + X_2 > 0) = P(Z > -0.0330/(0.0730\sqrt{2})) = 0.6254$.

例子2

EXAMPLE (系统寿命)

假设一个系统由若干个独立的元件组成，包含并联，串联和备用元件等连接方式；则系统的寿命由元件的寿命和连接方式决定。设各个元件的寿命服从指数分布$\epsilon(\lambda)$.

基本连接方式:

- 串联: $S = \min(X, Y)$, $F_S = 1 - (1 - F(x))^2 = 1 - e^{-2\lambda s}$;
- 并联: $S = \max(X, Y)$, $F_S = F(x)^2 = (1 - e^{-\lambda s})^2$;
- 备用系统: $S = X + Y$, $f_S = \lambda e^{-\lambda s}\lambda s$. $S \sim \Gamma(2, \lambda)$.

一般可以一步步得到系统的寿命。

作业

北航教材:
P120 习题四. 2,15,16,18,19,26,28,31

QUIZ 小测验 二

1. 设随机变量X在$(-\pi/2, \pi/2)$上服从均匀分布,则$Y = \tan X$的概率密度 $f_Y(y) =$
2. 将红、白、黑三只球随机地逐个放入编号为1，2，3的三个盒内（每盒容纳球的个数不限），以X 表示有球盒子的最小号码，求：（1）随机变量X的分布律；（2）X的分布函数。
3. 某仪器上装有4只独立工作的同类元件。已知每只元件的寿命(以小时计) $X \sim N(5000, \sigma^2)$,当工作的元件不少于2只时,该仪器能正常工作。 则该仪器能正常工作5000小时以上的概率为_。
4. ***P63. 23题:设随机变量X的概率密度为$f(x) = \frac{a}{e^x + e^{-x}}$, (1)确定常数$a$; (2)求X的分布函数; (3)求$P(0 \le X \le \ln\sqrt{3})$.

QUIZ 小测验 二: 答案

1. $f_Y(y) = \frac{1}{\pi}\frac{1}{1+y^2}, -\infty < y < +\infty.$
2. （1）随机变量X的分布律;
   $P(X = 3) = 1/27; P(X = 2) = 7/27; P(X = 1) = 19/27;$
   （2）X的分布函数。
   $F(x) = 0, x < 1; 19/27, 1 \leq x < 2; 26/27, 2 \leq x < 3; 1, x \geq 3$
3. 一个元件正常工作概率$p = P(X > 5000) = 1/2;$
   元件工作个数Y服从$B(4, 0.5)$,
   仪器正常工作概率:
   $P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) = 1 - 1/16 - 4/16 = 11/16.$
4. $a = 2/\pi, F(x) = 2/\pi \arctan e^x, P = 1/6.$