Date.　　/

随机变量函数的分布

$g \backslash X$	离散	连续
可列	PMF	PMF
连续	PMF	PDF

两套工具：CDF 和 PMF/PDF.

选用表示方法：对 $g(X)$：

考察 $Y=g(X)$ 的分布. 若 $g(x)$ 连续 则可通过以下两步实现：

1° 得到 Y 的 CDF：$F_Y(y)=\mathbb{P}(Y\le y)$　　2° 求导得到 $f_Y(y)=\frac{dF_Y(y)}{dy}$.

例：设 $X\sim N(\mu,\sigma^2)$，则 $Y=(X-\mu)/\sigma\sim N(0,1)$.

$\mathbb{P}(Y\le y)=\mathbb{P}(X\le\mu+\sigma y)=\int_{-\infty}^{\mu+\sigma y}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx=\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)dt$

$=\int_{-\infty}^{y}\varphi(t)dt$. 为标准正态分布的形式.

设 $X\sim N(0,1)$，求 $Y=X^2$ 的概率密度.　$(y>0,\ Y\ge 0)$

$F_Y(y)=\mathbb{P}(Y\le y)=\mathbb{P}(-\sqrt{y}\le X\le\sqrt{y})=\Phi(\sqrt{y})-\Phi(-\sqrt{y})=2\Phi(\sqrt{y})-1$，是连续的.

$f_Y(y)=F_Y'(y)=2\varphi(\sqrt{y})\frac{1}{2\sqrt{y}}=\frac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}}$，即 $Y\sim\Gamma(\frac{1}{2},\frac{1}{2})$.（正态→Gamma）

· $X\sim\mathcal{E}(\lambda)$，$Y=\lambda X$. 证明 $Y\sim\mathcal{E}(1)$：

$F(y)=\mathbb{P}(Y\ge y)=\mathbb{P}(X\ge\frac{y}{\lambda})=\int_{\frac{y}{\lambda}}^{\infty}\lambda e^{-\lambda t}dt=\cancel{-}e^{-y}$，$f(y)=\frac{dF}{dy}=e^{-y}$.

· Gamma 分布 $\Gamma(\alpha,\lambda)$.　α 个 iid $\mathcal{E}(\lambda)$ 之和

α,λ 为正常数，若 X 的密度是 $f(x)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$，$x\ge 0$，

称 X 服从参数 (α,λ) 的 Gamma 分布，$X\sim\Gamma(\alpha,\lambda)$. $\Gamma(\alpha)=\int_0^{\infty}x^{\alpha-1}e^{-x}dx$ 为 Gamma 函数.

性质：$\Gamma(n)=(n-1)!$；$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$；$\Gamma(\frac{1}{2})=\sqrt{\pi}$.

α 称为形状参数 (shape parameter)，λ 称为尺度参数 (rate parameter).

考虑标准正态分布：$1=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx\overset{u=\frac{x^2}{2}}{=\!=}\Rightarrow\sqrt{2\pi}=2\int_0^{+\infty}\frac{1}{\sqrt{2u}}e^{-u}du=\sqrt{2}\int_0^{+\infty}u^{-\frac{1}{2}}e^{-u}du$.

$=\sqrt{2}\,\Gamma(\frac{1}{2})$.

$\alpha=1$ 时为指数分布，α 为正整数时称为 Erlang 分布.

线性变换（改变伸缩倍数和平移）

$Y=aX+b$，$f_Y(y)=\frac{1}{|a|}f_X\left(\frac{y-b}{a}\right)$. 从 PDF 到 PDF.

可以用该结论验证：若 X 是服从正态分布的随机变量，则 Y 也是正态的.

* 扩展：$Y=g(X)$，$g(x)$ 是严格单调函数.

那么事件 $\{x\le X\le x+\delta\}=\{g(x)\le Y\le g(x+\delta)\}$ 近似为 $g(x)+\delta\left|\frac{dg}{dx}(x)\right|$.

因此 $\delta f_X(x)=\delta f_Y(g(x))\cdot|g'(x)|$，即 $f_Y(y)=f_X(h(y))\,|h'(y)|$，$h(y)$ 是 $g(x)$ 的逆映射.

Date.   /

例：$X\sim\Gamma(\alpha,1)$，$f_X(x)=\frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-x}$，$x>0$，$Y=\frac{X}{\lambda}$，$f_Y(y)=\frac{\lambda^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-\lambda y}$，$Y\sim\Gamma(\alpha,\lambda)$.

计算技巧：$f_X(x)\left|\frac{dx}{dy}\right|=f_Y(y)$，$\left|\frac{dx}{dy}\right|=\left(\left|\frac{dy}{dx}\right|\right)^{-1}$.

更一般的定理：设X的概率密度为$f(x)$，$D\subset\mathbb{R}$，$Y=g(X)$，$\mathbb{P}(Y\in D)=1$. 若

1° 对$y\in D$，$\{Y=y\}=\bigcup_{i=1}^n\{X=h_i(y)\}$；

2° 每个$h_i(y)$是D到其值域D_i的可逆映射，在D内有连续的导数；

3° $D_1,D_2,\dots D_n$互不相交.

则Y的密度函数为 $f_Y(y)=\begin{cases}\sum_{i=1}^n f(h_i(y))\,|h_i'(y)|, & y\in D\\ 0, & y\in D^c\end{cases}$

例 $r>0$，$X\sim U(0,2\pi)$，求$Y=r\cos X$的概率密度.（两种方法）

方法一：记$D=(-r,r)$，$\mathbb{P}(Y\in D)=1$，$\{Y=y\}=\{X=\arccos\frac{y}{r},X=2\pi-\arccos\frac{y}{r}\}$，$h_1(x)=\arccos\frac{y}{r}$，$h_2(x)=2\pi-h_1(x)$

利用$F_X(x)=\frac{1}{2\pi}$：$f_Y(y)=\frac{1}{2\pi}\left(\left|\frac{d}{dy}h_1(y)\right|+\left|\frac{d}{dy}h_2(y)\right|\right)=\frac{1}{\pi\sqrt{r^2-y^2}}$，$-r<y<r$.

方法二：$\mathbb{P}(Y\le y)=\mathbb{P}(\cos X\le y/r)=\int_{x:\cos x\le y/r}\frac{1}{2\pi}dx=\frac{1}{2\pi}\left(2\pi-2\arccos\frac{y}{r}\right)$.

$(\arccos x)'=-\frac{1}{\sqrt{1-x^2}}$，$f(y)=\frac{1}{\pi\sqrt{r^2-y^2}}$，$-r<y<r$.

随机向量及其分布（特别关注多项分布、二元正态分布）

（定义）若$X_1,\dots X_n$都是概率空间$(\Omega,\mathcal{F},\mathbb{P})$上的随机变量，称$X=(X_1,\dots X_n)$是概率空间上的n维随机向量。

（定义）$\vec{X}$为n维随机向量，$\mathbb{R}^n$上的n元函数：$F(x_1,\dots x_n)=\mathbb{P}(X_1\le x_1,\dots X_n\le x_n)$

为$\vec{X}$的联合概率分布函数，简称分布函数或联合分布.

$F(x_1,\dots x_n)$性质：$\mathbb{R}^n$上的右连续函数，关于每个自变元x_i单调非降.

（定义）若$X_1,\dots X_n$都是离散型随机变量，称$\vec{X}=(X_1,\dots X_n)$为离散型随机向量.

若X所有的不同取值是$\vec{x}(j_1,\dots j_n)=(x_1(j_1),x_2(j_2),\dots x_n(j_n))$，$j_i\ge1$，

则称$p_{j_1,\dots,j_n}=\mathbb{P}(\vec{X}=\vec{x}(j_1,\dots j_n))$是$\vec{X}$的联合分布列.

性质：$p_{j_1,\dots,j_n}\ge0$，$\sum_{j_1,\dots,j_n}p_{j_1,\dots,j_n}=1$.（p是在单一取值下的可能概率）

表格形式：$\{p_i\}$是各行和，$\{q_j\}$是各列和.

p_{ij}	y_1	y_2	$\dots$	$\{p_i\}$
x_1	p_{11}	p_{12}	$\dots$	p_1
x_2	p_{21}	p_{22}		p_2
$\vdots$	$\vdots$			$\vdots$
$\{q_j\}$	q_1	q_2		$\Sigma=1$

$\{p_i\}$、$\{q_j\}$分别为X、Y的PMF.

(定义) 设离散型随机变量的联合分布列为 $p_{j_1,\dots,j_n}$，则称 $(j\ge 1)$

$P(X_1=x_1(j_1),\dots,X_k=x_k(j_k))=\sum_{j_{k+1},\dots,j_n} p_{j_1,j_2,\dots,j_n}$，为X的边缘分布列.

(定义) 边缘分布/边际分布 marginal distribution

对 $1\le k<n$，$S_k=\{i_1,\dots,i_k\}$ 为指标集，称 $X_{S_k}=(X_{i_1},\dots,X_{i_k})$ 的联合分布

$F_k(x_{i_1},x_{i_2},\dots,x_{i_k})=P(X_{i_1}\le x_{i_1},\dots,X_{i_k}\le x_{i_k})=P(X_j\le x_j, j\in S_k, X_t\le \infty, t\in S_k^c)$.

为 $F(x_1,\dots,x_n)$ 的k-维边缘分布. (一共有 2^n-2 个边缘分布)

离散型随机向量的独立性：设 $\vec{X}=(X_1,\dots,X_n)$ (离散随机向量) 有概率分布 $p_{j_1,\dots,j_n}=P(\vec{X}=\vec{x}(j_1,\dots,j_n))$，则 $X_1,\dots,X_n$ 相互独立的充要条件是：对任意 $(x_1(j_1),x_2(j_2),\dots,x_n(j_n))$，有

$P(X_1=x_1(j_1),\dots,X_n=x_n(j_n))=\prod_{i=1}^{n}P(X_i=x_i(j_i))$.

推论：(二维) 设离散型随机向量 (X,Y) 的所有不同取值是 (x_i,y_j)，$i,j\ge 1$，则X,Y相互独立的充要条件是 $\forall(x_i,y_j)$，有 $P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)$.

(证明) $\Leftarrow$：$P(X\le x,Y\le y)=\sum_{x_i\le x}\sum_{y_j\le y}P(X=x_i,Y=y_j)=\sum_{x_i\le x}\sum_{y_j\le y}P(X=x_i)P(Y=y_j)$

$=\sum_{i:x_i\le x}P(X=x_i)\sum_{j:y_j\le y}P(Y=y_j)=P(X\le x)P(Y\le y)$.

$\Rightarrow$：$A=\{x_i\}$ 和 $B=\{y_j\}$ 都是 Borel 集，$\{X\in A\}$ 与 $\{Y\in B\}$ 独立，$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)$

举例：多项分布. 设 $A_1,\dots,A_r$ 是试验S的完备事件组，对试验S进行n次独立重复试验时，用 X_i 表示 A_i 发生次数，则 $(X_1,\dots,X_n)$ 的概率分布是

$P(X_1=k_1,\dots,X_r=k_r)=\frac{n!}{k_1!\cdots k_r!}p_1^{k_1}\cdots p_r^{k_r}$，其中 $k_i\ge 0$，$p_i=P(A_i)$，$\sum_{i=1}^{r}k_i=n$.

对 $1\le m<n$，$(X_{j_1},\dots,X_{j_m})$ 的边缘分布为 $P(X_{j_1}=k_1,\dots,X_{j_m}=k_m)=\frac{n!}{k_1!\cdots k_m!\,k!}p_{j_1}^{k_1}\cdots p_{j_m}^{k_m}q^{k}$.

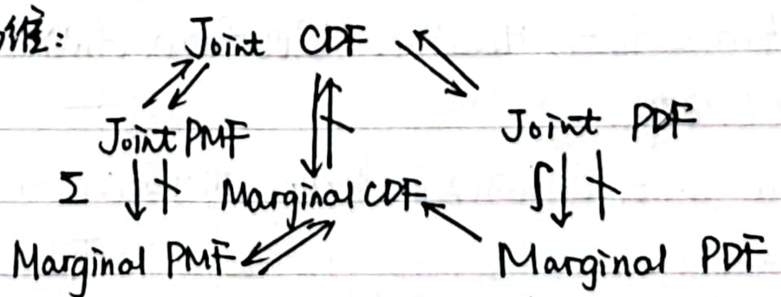
其中 $0\le k_i\le n$，$k=n-\sum_{i=1}^{m}k_i\ge 0$，$q=1-\sum_{i=1}^{m}p_{j_i}$.

特别地，$P(X_j=k)=\binom{n}{k}p_j^k(1-p_j)^{n-k}$，$k\ge 0$. (特殊情况下为二项分布)

*重要性质 若 $X_1,\dots,X_n$ 相互独立，则 $F(x_1,\dots,x_k)$ 与 $G(x_{k+1},\dots,x_n)$ 也相互独立.

一维：

高维：



边缘推联合(离散)

只能保证存在性，不能保证唯一性.　(连续)：既不保证唯一，也不保证存在.

(下一部分会更新此图)

Date. /

(定义) $\vec{X}=(X_1,\dots X_n)$ 为随机向量，若存在 $\mathbb{R}^n$ 上非负可积函数 $f(\vec{x})$，使得对 $\mathbb{R}^n$ 的任何子立方体，$D=\{(x_1,\dots x_n)\mid a_i<x_i<b_i, 1\le i\le n\}$ 有

$\mathbb{P}(X\in D)=\int_D f(x_1,x_2,\dots x_n)dx_1dx_2\dots dx_n=\int_D f(\vec{x})d\vec{x}$，则称 $\vec{X}$ 为连续型随机向量，$f(\vec{x})$ 是 X 的联合密度函数，简称联合密度/概率密度.

(定理) 对 $\forall$ n维 Borel 集 B 有 $\mathbb{P}(X\in B)=\int_B f(\vec{x})d\vec{x}$.

$\vec{X}$ 的联合概率密度不必唯一. $\mathbb{R}^n$ 是 Borel 集，$\int_{\mathbb{R}^n} f(x)dx=\mathbb{P}(X\in\mathbb{R}^n)=1$.

(定理)(会用即可) Fubini：设 D是 $\mathbb{R}^n$ 上的子区域，$\varphi(\vec{x})$ 是D上的非负函数或绝对可积函数(指取绝对值后积分有限)，则对区域D上的n重积分 $\int_D\varphi(\vec{x})d\vec{x}$ 可以进行累次积分计算，且积分次序可以交换.

二维边缘密度：联合密度 $f(x,y)$，$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy$，$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx$.

(定义) 称被积函数 $f_k(x_1,\dots x_k)=\int_{\mathbb{R}^{n-k}}f(x_1,\dots x_n)dx_{k+1}\dots dx_n$ 是 $(X_1,\dots X_k)$ 的联合密度，称为 X 的边缘概率密度函数，简称边缘密度.

注：求连续型随机变量概率密度一定要注意范围，建议用示性函数.

· 连续型随机向量的独立性

(PDF不表示事件，不能像离散型PMF一样直接拆开各项)

定理 设对每个 $i(1\le i\le n)$，随机变量 X_i 有概率密度 $f_i(x_i)$，则 $X_1,\dots X_n$ 相互独立的充要条件是 $\vec{X}=(x_1,\dots x_n)$ 有联合密度 $f_1(x_1)\dots f_n(x_n)$，$\vec{x}\in\mathbb{R}^n$.

证明 $\Leftarrow$：$\forall\vec{x}\in\mathbb{R}^n$，$\mathbb{P}(X_1\le x_1,\dots X_n\le x_n)=\int_{-\infty}^{x_1}\dots\int_{-\infty}^{x_n}\prod_{i=1}^{n}f_i(t_i)dt_1\dots dt_n$

$=\prod_{i=1}^{n}\int_{-\infty}^{x_i}f_i(x_i)dt_i=\prod\mathbb{P}(X_i\le x_i)$，故 $X_1,\dots X_n$ 相互独立.

$\Rightarrow$：对 $\mathbb{R}^n$ 的任意子立方体用 Fubini 定理. 得

$\mathbb{P}(\vec{X}\in D)=\mathbb{P}(a_1<X_1\le b_1)\dots\mathbb{P}(a_n<X_n\le b_n)=\int_{a_1}^{b_1}f_1(t_1)dt_1\dots\int_{a_n}^{b_n}f_n(t_n)dt_n$

$=\int_D f_1(t_1)\dots f_n(t_n)dt_1\dots dt_n$. 因此 $f_1(x_1)\dots f_n(x_n)$ 为联合密度.

*例：二元正态分布.

除了 μ_1、μ_2、σ_1、σ_2 外引入新参数 ρ，其中 σ_1、$\sigma_2>0$，$\rho\in(-1,1)$，

$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$，若 (X,Y) 有概率密度

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

引入矩阵 $\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$，$|\Sigma|$ 表示其行列式. $\vec{x}=(x,y)^T$，$\mu=(\mu_1,\mu_2)^T$.

证明独立的方法：析联合分布列或考虑条件分布.

前式简写为 $f(x)=\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\exp\left(-\frac{1}{2}[(x-\mu)^T\Sigma^{-1}(x-\mu)]\right)$.

例 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1,\sigma_2)$ 求 X、Y 的边缘密度.（$\mu_1=\mu_2=0,\sigma_1^2=\sigma_2^2=1$）

$f_X(x)=\int f(x,y)dy=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(\frac{y^2-2\rho xy+\rho x^2+x^2-\rho x^2}{2(1-\rho^2)}\right)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$

发现此时 $f_X(x)\sim N(0,1)$　　　注意：两个正态分布联合(X,Y)不一定构成二元正态分布.

定理 设 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1,\sigma_2;\rho)$，则 X、Y 独立的充要条件是 $\rho=0$.

联合密度 $f(x,y)\to$ 联合分布 $F(x,y)$：$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv$，$(x,y)$ 连续时 $f=\frac{\partial^2F}{\partial x\partial y}$.

联合分布 $\to$ 联合密度：若 $\frac{\partial F}{\partial x}$，$\frac{\partial F}{\partial y}$，$\frac{\partial^2F}{\partial x\partial y}$，$\frac{\partial^2F}{\partial y\partial x}$ 均连续，则

$f(x,y)=\frac{\partial^2F}{\partial x\partial y}=\frac{\partial^2F}{\partial y\partial x}$ 为 (X,Y) 的联合密度函数.

(定理) 说明：更一般地，设 (X,Y) 有连续的分布函数 $F(x,y)$，定义 $f(x,y)=\begin{cases}\frac{\partial^2F(x,y)}{\partial x\partial y}, & \text{混合偏导存在}\\ 0\end{cases}$

若 $\int_{\mathbb{R}^2}f(x,y)dxdy=1$ 则 $f(x,y)$ 是 (X,Y) 的联合密度.

若 X、Y 的分布函数连续，则其在联合分布 $F(x,y)$ 连续.

· 存在随机向量 (X,Y)：它有连续的联合分布，但不是连续型随机向量.

(反例) $X\sim U(0,1)$，$Y=X$，若有联合密度 $f(x,y)$，用 D 表示 $y=x$ 直线上的全部点，

$1=P(X=Y)=P((X,Y)\in D)=\int_D f(x,y)dxdy=0$，那么 (X,Y) 不是连续型随机向量.

条件密度和条件分布

· 离散型：joint PMF $p_{ij}=P(X=x_i,Y=y_j)$；marginal PMF $p_i=P(X=x_i)$，$q_j=P(Y=y_j)$.

(定义) 对每个固定的 j，称 $P(X=x_i\mid Y=y_j)=\frac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\frac{p_{ij}}{q_j}$ 为

给定条件 $Y=y_j$ 下，X 的条件分布列 (conditional PMF).

定理：X,Y 独立的充要条件：$\forall i,j\geq 1$，$P(X=x_i\mid Y=y_j)=P(X=x_i)$.

独立同分布序列 (iid)：随机变量 $\{X_i\}$ 相互独立并有相同的分布函数

例 (射击问题) 用 S_n 表示第 n 次击中目标时的射击次数，每次击中概率为 $p=1-q$.

$(X,Y)=(S_1,S_2)$ 的联合分布：$P(X=i,Y=j)=p^2q^{j-2}$，$j>i\geq1$.

边缘分布：$P(X=i)=pq^{i-1}$，$P(Y=j)=(j-1)p^2q^{j-2}$.

条件分布：$P(Y=j\mid X=i)=pq^{j-i-1}$，$P(X=i\mid Y=j)=\frac{1}{j-1}$，说明 $S_2=j$ 时，S_1 在 1 到 $j-1$

中的取值是等可能的. 若令 $X_1=S_1$，$X_2=S_2-S_1$，($X_i=S_i-S_{i-1}$)，发现

$P(X_1=i,X_2=j)=P(X_1=i)P(X_2=j)$. $X_1,\dots,X_n$ 是独立同分布的.

例(孵化问题) $N\sim P(\lambda)$ 为鸡蛋系数，X 为孵化成功，Y 为失败.

$X|N=n\sim B(n,p)$. $X+Y=N$，显然 $X \perp\!\!\!\perp Y$. 接下来证明之.

$P(X=i,Y=j)=P(X=i|N=i+j)P(N=i+j)=\frac{(i+j)!}{i!\,j!}p^iq^j\frac{\lambda^{i+j}}{(i+j)!}e^{-\lambda}$

$=\frac{(\lambda p)^i}{i!}\frac{(\lambda q)^j}{j!}e^{-\lambda}$ (Poisson分布核心项) $=\mathbb{P}(X=i)\mathbb{P}(Y=j)$. 它们其实相互独立.

技巧：识别核心项；归一化系数对应PMF/PDF；拆分快速判断独立性.

$X+Y=N$ 却相互独立是Poisson分布特性，可以用“密度”概念理解.

当发现联合分布列中 i,j 已经解耦(完全拆开)，可以直接判定为相互独立.

例(消费问题) 超市单日顾客数 N 是随机变量，单个顾客消费与 N 独立且服从 $P(\mu)$.

S 表示全天营业额，求 $\mathbb{P}(S=k|N)$.

X_j 表示第 j 个顾客的消费，$\{X_j\}$ 独立同分布. S_n 表示前 n 个顾客消费，$S_n\sim P(n\mu)$.

(要证：$S_n\sim P(n\mu)$：先证 $n=2$ 情况，按 $X+Y=k$ 分类讨论并求和，$S_2\sim P(2\mu)$.

之后用归纳法证明.)

$\mathbb{P}(S=k|N=m)=\mathbb{P}(S_m=k)=\frac{(m\mu)^k}{k!}e^{-m\mu}$，则 $\mathbb{P}(S=k|N)=\frac{(N\mu)^k}{k!}e^{-N\mu}$.

· 连续型：理解：对 $\forall y$(固定)，$f_{X|Y}(x|y)$ 是一个概率密度函数(PDF)

为了处理条件密度，考虑在 y 的周围施加一个微小扰动.

$\mathbb{P}(x\le X\le x+\Delta x|Y=y)\approx \mathbb{P}(x\le X\le x+\Delta x|y\le Y\le y+\Delta y)$.

$\approx\frac{f(x,y)\cdot\Delta x\cdot\Delta y}{f_Y(y)\,\Delta y}$ (直接用PDF计算) $\approx f_A(x)$. $F_A(x)=\int_{-\infty}^{x}f_A(t)\,dt$.

(先到CDF再求导得PDF) 若 $f_Y(y)$ 在 y 连续，$f_Y(y)>0$，$\frac{\partial F}{\partial y}$ 存在，则 $\lim_{\varepsilon\to 0}\mathbb{P}(X\le x|y-\varepsilon<Y\le y)$

$=\lim_{\varepsilon\to 0}\frac{F(x,y)-F(x,y-\varepsilon)}{F_Y(y)-F_Y(y-\varepsilon)}=\frac{\partial F/\partial y}{\partial F_Y/\partial y}=\frac{\frac{\partial}{\partial y}\int_{-\infty}^{y}\left(\int_{-\infty}^{x}f(s,t)\,ds\right)dt}{f_Y(y)}=\frac{\int_{-\infty}^{x}f(s,y)\,ds}{f_Y(y)}$.

(定义) 随机向量 (X,Y) 有联合密度 $f(x,y)$，Y 边缘密度 $f_Y(y)$，给定 $f_Y(y)>0$，则

$\mathbb{P}(X\le x|Y\le y)=\frac{\int_{-\infty}^{x}f(s,y)\,ds}{f_Y(y)}$ 为给定 $Y=y$ 时 X 的条件分布(函数)，记作 $F_{X|Y}(x|y)$.

称 $f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{f(x,y)}{\int_{-\infty}^{\infty}f(x,y)\,dx}$ 为 X 的条件概率密度.

· 条件密度与条件分布关系：若 y 使得 $f_Y(y)>0$，有

1° $F_{X|Y}(x|y)=\mathbb{P}(X\le x|Y=y)=\int_{-\infty}^{x}f_{X|Y}(s|y)\,ds$，$x\in\mathbb{R}$.

2° 若 $F(x,y)$ 关于 x 连续，且除去至多可列个点外有连续导数，则 X 条件密度 $f_{X|Y}(x|y)=\begin{cases}\frac{\partial F_{X|Y}(x|y)}{\partial x} & \text{偏导存在}\\ 0 & \text{不存在}\end{cases}$

(定理) X、Y 独立的充要条件：对 $y\in\{y|f_Y(y)>0\}$，$f_{X|Y}(x|y)=f_X(x)$，$x\in\mathbb{R}$.

随机向量条件分布+边缘分布→联合分布→另一变量的边缘分布

(定义) 设 $\vec{X}=(X_1,\dots,X_n)$ 和 $\vec{Y}=(Y_1,\dots,Y_m)$ 是随机向量，$f(\vec{x},\vec{y})$ 为其联合密度，Y的边缘密度

$f_Y(\vec{y})=\int_{\mathbb{R}^n} f(\vec{x},\vec{y})\,dx$. 给定y，$f_Y(y)>0$，则称

$P(X\le x\mid Y=y)=\frac{\int_{s\le x} f(s,y)\,ds}{f_Y(y)}$ 为X的条件分布函数(条件分布)，$F_{X|Y}(x|y)$.

$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$ 为X的条件概率密度(条件密度).

随机向量函数的分布（都与正态分布有关，统推里会讲）

- 卡方分布 Chi-Squared distribution (χ^2_v分布) 卡方检验 可以写作 χ^2分布.

  多个正态独立同分布平方和. $f(x)=\frac{1}{2^{v/2}\Gamma(v/2)}x^{\frac{v}{2}-1}e^{-\frac{x}{2}}$，$x\ge 0$.

  $X\sim\chi^2_v$ (参数为v) $X_i^2\sim\Gamma(\frac12,\frac12)$，于是 $X_n^2=\Gamma(\frac n2,\frac12)$. $EX=n$，$var(X)=2n$.

- Student's t_v分布 t_v检验 $f(x)=\frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\,\Gamma(v/2)}\left(1+\frac{x^2}{v}\right)^{-\frac{v+1}{2}}$.

  设有 $Z\sim N(0,1)$，$X\sim\chi_n^2$ 且X与Z独立，令 $T=\frac{Z}{\sqrt{X/n}}$，则有 $T\sim t_n$.

  对称性：若 $T\sim t_v$，则 $-T\sim t_v$. 比正态分布重尾. $n\ge 2$时 $ET=0$，$n\ge 3$时 $var(X)=\frac{n}{n-2}$.

  v很大时 t_v和标准正态分布很像；$v=1$时 性质特殊，也叫Cauchy分布.

- F分布 Fisher Snedecor distribution 方差分析、线性回归（两个卡方分布）

$f(x)=\frac{\Gamma(\frac{m+n}{2})m^{\frac m2}n^{\frac n2}x^{\frac m2-1}}{\Gamma(\frac m2)\Gamma(\frac n2)(mx+n)^{\frac{m+n}{2}}}$，$x\ge 0$，$X\sim F(m,n)$. $n\ge 3$时 $EX=\frac{n}{n-2}$.

(或 $X\sim\chi^2_m$，$Y\sim\chi^2_n$ 独立，$F=\frac{X/m}{Y/n}$)

独立同分布变量 $X_j\sim N(0,1)$，$j=1,\dots,m$ 和 $Y_j\sim N(0,1)$，$j=1,\dots,n$. $F=\frac{\frac1m\sum_i X_i^2}{\frac1n\sum_i Y_i^2}$，$F\sim F(m,n)$.

定理 n维随机向量X、Y有相同分布，$g:\mathbb{R}^n\to\mathbb{R}^m$ 可测，则 $g(\vec{X})$ 与 $g(\vec{Y})$ 有相同的分布.

P_A $P(x|y)\leftrightarrow$ Conditional $F(x|y)$ $\rightleftarrows$ $f(x|y)$ (连续时可推)

P

$$
\begin{tikzcd}
 & \text{Joint } F(x,y) \arrow[dr, shift left] & \\
P(x,y) \arrow[ur, leftrightarrow] \arrow[d, shift right, "\Sigma"'] & & f(x,y) \arrow[ul, dashed, shift left] \arrow[dl, "\int"] \arrow[d, "/"] \\
P_X(x),P_Y(y) \arrow[u, shift right, "f"'] \arrow[r, leftrightarrow] & \text{Marginal } F_X(x),F_Y(y) \arrow[r, leftrightarrow] & f_X(x),f_Y(y)
\end{tikzcd}
$$

独：

$$
\begin{tikzcd}
 & F(x,y)=F_X(x)F_Y(y) \arrow[dl, leftrightarrow] \arrow[dr, Rightarrow] & \\
P(x,y)=P_X(x)P_Y(y) \arrow[d, Leftrightarrow] & & f(x,y)=f_X(x)f_Y(y) \arrow[d, Leftrightarrow] \\
P(x|y)=P_X(x) & & f_{X|Y}(x|y)=f_X(x)
\end{tikzcd}
$$

CDF PDF/PMF

Date. /

对 X, Y (iid) $\sim N(0,1)$，考虑随机变量 $T=\frac{X}{|Y|}$.

$\mathbb{P}(T\le t) \Rightarrow \mathbb{P}(\frac{X}{|Y|}\le t)=\mathbb{P}(X\le t|Y|)=\int_{-\infty}^{\infty}\int_{-\infty}^{t|y|}\varphi(x)\varphi(y)\,dxdy$

$=\int_{-\infty}^{+\infty}\Phi(t|y|)\varphi(y)dy.$ $f_T(t)=2\int_0^{+\infty}\varphi(y)\varphi(ty)y\,dy\rightarrow$

$f_T(t)=\frac{2}{2\pi}\cdot\int_0^{+\infty}\exp(-\frac{y^2(1+t^2)}{2})y\,dy.$ 用 $u=-\frac{1}{2}y^2(1+t^2)$ 换元：

$f_T(t)=\frac{1}{\pi}\frac{1}{1+t^2}\int_0^{+\infty}e^{-u}du=\frac{1}{\pi(1+t^2)}$. 积分部分是 $\mathcal{E}(\lambda=1)$ 的CDF.

或者 $P(X\le t|Y|)=\int P(X\le t|y|)P(Y=y)\varphi(y)dy=\int\Phi(t|y|)\varphi(y)dy$.

Cauchy 分布：t_ν 分布在 $t_\nu=1$ 时的特例，$f(x)=\frac{1}{\pi(1+x^2)}$，$X\sim t_1$.

独立同分布变量 $Z_j\sim t_1$，令 $X=\frac{Z_1+\dots+Z_n}{n}$，则依然有 $X\sim t_1$.

例 设 (X,Y) 有联合分布列，求 $Z=X+Y$ 的分布列.

$P_Z(Z=z)=P(X+Y=z)=\sum_x P(X=x, Y=z-x)$.

特别地，当 X、Y 独立时，有分布列 $P_Z(z)=\sum_x\mathbb{P}(X=x)\mathbb{P}(Y=z-x)=\sum_x P_X(x)P_Y(z-x)$.

此时 Z 的分布列为 X 和 Y 的分布列的卷积 (convolution).

对连续型：(X,Y) 有联合密度 $f(x,y)$，则 $Z=X+Y$ 有密度函数

$f_Z(z)=\int_{-\infty}^{\infty}f(x,z-x)dx$. X、Y 独立时 有 $f_Z(z)=\int_{-\infty}^{\infty}f_X(x)f_Y(z-x)dx$.

证明：$\mathbb{P}(a<X+Y\le b)=\int_{\mathbb{R}^2}I_{a<x+y\le b}f(x,y)dxdy=\int_{\mathbb{R}}(\int_{a-x}^{b-x}f(x,y)dy)dx=\int_{\mathbb{R}}(\int_a^b f(x,z-x)dz)dx$

$=\int_a^b(\int_{\mathbb{R}}f(x,z-x)dx)dz$. 因此 $f_Z(z)=\int_{-\infty}^{\infty}f(x,z-x)dx=\int_{-\infty}^{+\infty}f(z-y,y)dy$.

或者用微元法：$\mathbb{P}(z<X+Y\le z+\Delta z)\approx f_Z(z)\Delta z\rightarrow\int_{-\infty}^{+\infty}\int_{z-x}^{z+\Delta z-x}f_{X,Y}(x,y)dydx=\int_{-\infty}^{+\infty}f_{X,Y}(x,z-x)dx$.

例 均匀分布的独立同分布推广

· X、Y 独立且 $\sim U(0,1)$，$Z=X+Y$. $f_X(x)=f_Y(y)=I_{0<x<1}$，$f_Z(z)=\int_0^1 I_{0<z-x<1}dx=\begin{cases}z, & z\in[0,1)\\ 2-z, & z\in[1,2]\\ 0, & \text{otherwise}\end{cases}$

X、Y 独立且 $\sim U(-1,1)$，$Z=X+Y$. $f_Z(z)=\int_{-1}^{1}I_{-1<z-x<1}\frac{dx}{4}=\begin{cases}\frac{1}{4}(2-|z|), & |z|\le 2\\ 0, & |z|>2\end{cases}$.

有 X_1、X_2、$X_3\sim U(-1,1)$，$f_{X_1+X_2+X_3}(x)=\frac{1}{16}(3-|x|)^2$，$1\le|x|\le 3$ 或 $\frac{1}{8}(3-x^2)$，$0\le|x|\le 1$. $|x|>3$ 无意义.

iid $X_1,\dots,X_n\sim U(-1,1)$：$f_{S_n}(x)=\frac{1}{2^n(n-1)!}\sum_{k=0}^{[\frac{n+x}{2}]}(-1)^k\binom{n}{k}(n+x-2k)^{n-1}$，$|x|\le n$.

· 对称法：$V=X-Y$ 有 $f_V(v)=\int_{-\infty}^{\infty}f(x,x-v)dx$. 特别地，当 X、Y 独立时

$f_V(v)=\int_{-\infty}^{+\infty}f_X(x)f_Y(x-v)dx=\int_{-\infty}^{+\infty}f_X(y+v)f_Y(y)dy$.

* 特别注意易错题：X、Y 独立同分布，且 $\sim N(0,1)$，$T=\frac{X}{Y}$ 的概率密度：

$f_T(t)=\int_{-\infty}^{\infty}f_X(ty)f_Y(y)dy\Rightarrow$ 做法错误. 卷积公式只对和差适用.

如果用微元理解：随 X、Y 变化，乘/除法不能直接提出 Δz.

用除法公式：$\int|y|f_X(t|y|)f_Y(y)dy$.

多个函数的联合密度　　　　　　　　　　(实n随机变量函数)

设 $\vec{X}=(X_1,\dots X_n)$ 有联合密度 $f(\vec{x})$，$Y_i=g_i(\vec{X})$，$\vec{Y}=(Y_1,\dots Y_n)$ 是n个X的函数.

定理(只考察二维) 设 (X,Y) 有联合密度 $f(x,y)$. $U=u(x,y)$，$V=v(x,y)$ 都是 (X,Y) 函数，平面上区域D使得 $P((U,V)\in D)=1$. 若存在D上函数 $x_i(u,v)$，$y_i(u,v)$，使得

1° 对 $(u,v)\in D$ 有 $\{U=u,V=v\}=\bigcup_{i=1}^{n}\{X=x_i,Y=y_i\}$；

2° Δ_i：$x_i=x_i(u,v)$，$y_i=y_i(u,v)$ 是D到其值域 D_i 的可逆映射，有连续的偏导数，且 Jacobi矩阵行列式绝对值 $\left|\frac{\partial(x_i,y_i)}{\partial(u,v)}\right|\neq 0$.　　3° $D_1,\dots D_n$ 互不相交.

则 (U,V) 有联合密度 $g(u,v)=\begin{cases}\sum_{i=1}^{n}f(x_i(u,v),y_i(u,v))\left|\frac{\partial(x_i,y_i)}{\partial(u,v)}\right| & (u,v)\in D\\ 0, & (u,v)\in D^c.\end{cases}$

·用事件换元法解决 $P(T=\frac{X}{Y}\le t)$：考虑 (x,y) 在二维平面内对应角度θ的均匀分布，$P=P(\theta\in[\alpha,\pi]\cup[\pi-\alpha,2\pi])$，$\alpha=\arctan t$.

或者 $P=1-\frac{2\alpha}{2\pi}=1-\frac{1}{\pi}\arctan t\Rightarrow f_T(t)=\frac{1}{\pi(1+t^2)}$.

例 设X、Y独立，$X,Y\sim N(0,1)$，考虑 $U=\frac{X}{Y}$ 和 $V=X^2+Y^2$.（增补变量法）

$f_{U,V}(u,v)=f_{X,Y}(x,y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right|=f_X(x)f_Y(y)\cdot\frac{1}{2(1+u^2)}=\frac{1}{4\pi}e^{-\frac{1}{2}v}\frac{1}{1+u^2}$.

但这样系数会差2：(u,v) 和 (x,y) 是一对2的关系，(x,y) 与 $(-x,-y)$ 结果相同.

$g(u,v)=\frac{1}{2\pi(1+u^2)}e^{-\frac{v}{2}}$，即U,V独立，U服从Cauchy分布，V服从指数分布.

总结：求 $\frac{X}{Y}$ 等随机变量分布的方法：

(General) CDF → $f(x,y)$　CDF → $f_{X|Y}\,f_Y$.

(Special) 微元法(特别关注卷积) 事件换元法. 增补变量法(见上)

·对于多个函数的联合密度：一般只考虑到二维.

注意点：$\left|\frac{\partial(x,y)}{\partial(u,v)}\right|$ 上下次序(可以先用逆矩阵 $\left|\frac{\partial(u,v)}{\partial(x,y)}\right|$ 取行列式倒数.

例：X、Y独立，$X\sim N(0,1)$，由极坐标变换 $\begin{cases}X=R\cos\theta\\Y=R\sin\theta\end{cases}$，求 (R,θ) 的联合密度.

$f(x,y)=\frac{1}{2\pi}\exp(-\frac{x^2+y^2}{2})$，定义 $D=\{(r,\theta)|r>0,\theta\in[0,2\pi)\}$. $P((R,\theta)\in D)=1$.

$\left|\frac{\partial(x,y)}{\partial(r,\theta)}\right|=r>0$，$g(r,\theta)=f(r\cos\theta,r\sin\theta)\cdot r=\frac{1}{2\pi}r\exp(-\frac{r^2}{2})=\frac{1}{2\pi}\exp(-\frac{r^2}{2})$.

r和θ独立，$g_R(r)=r\exp(-\frac{r^2}{2})$，为Rayleigh分布，$\theta\sim U(0,2\pi)$.

Date.   /

次序统计量 order statistics

(定义) 设 $X_1, X_2, \dots X_n$ 是概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的随机变量，对 $w\in\Omega$，将 $X_1(w), \dots X_n(w)$ 从小到大排列得到 $X_{(1)}(w) \le X_{(2)}(w) \dots \le X_{(n)}(w)$，称 $X_{(1)}, X_{(2)}, \dots X_{(n)}$ 为原 $\{X_i\}$ 的次序统计量.

以下设 $X_1, X_2, \dots X_n$ 为 iid，有共同的分布函数 $F(x)$ 和密度函数 $f(x)$.

性质 1° $(X_{(1)}, \dots X_{(n)})$ 有联合密度 $g(x_1, \dots x_n) = \begin{cases} n!\prod_{i=1}^n f(x_i), & x_1 < \dots < x_n \\ 0, & \text{otherwise} \end{cases}$

2° $X_{(k)}$ 有密度 $g_k(x_k) = n\binom{n-1}{k-1}[F(x_k)]^{k-1}[1-F(x_k)]^{n-k}f(x_k)$.

3° 对于 $-\infty \le a < x_1 \dots < x_k < b \le \infty$，有

$$\int_{a<x_1<\dots x_k<b} f(x_1)\dots f(x_k)\,dx_1\dots dx_k = \frac{1}{k!}(F(b)-F(a))^k.$$

3° 可以用归纳法证明. 由 Fubini 定理 $= \int_a^b \frac{1}{(k-1)!}(F(x_k)-F(a))^{k-1}f(x_k)dx_k$.

4° 对 $k_1<k_2$，$(X_{(k_1)}, X_{(k_2)})$ 有联合密度 $(x_{k_1}<x_{k_2})$

$$g(x_{k_1}, x_{k_2}) = \frac{n!}{(k_1-1)!(k_2-k_1-1)!(n-k_2)!}\times[F(x_{k_1})]^{k_1-1}[F(x_{k_2})-F(x_{k_1})]^{k_2-k_1-1}\times[1-F(x_{k_2})]^{n-k_2}f(x_{k_1})f(x_{k_2})$$

· Beta 分布 $Beta(\alpha, \beta)$，$\alpha, \beta>0$

$f(x) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$，$0<x<1$，其中 $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

常在 Bayes 统计方法中作为参数的先验分布. 特别地，它是二项分布的共轭分布.

$(X|p \sim B(n,p))$，p 的先验 $B(\alpha,\beta)$，观测 X 后得到后验 $Beta(\alpha+X, \beta+n-X)$.

· 例 灯泡问题. 若每个灯泡的寿命 $X_i \sim \mathcal{E}(\lambda)$，教室内原本装有 4 只灯泡，现在想换成 24 只，第一只灯泡损坏的时长 Z 如何变化？

由上述性质 2，4 只灯泡时 $g_Z(x_{(1)}) = \mathcal{E}(4\lambda)$，24 只时 $g_Z(x_{(1)}) = \mathcal{E}(24\lambda)$.

那么 24 只灯泡时更快出现灯泡损坏的概率更大.

例 均匀分布 $\{X_n\}$ iid，$X_i \sim U(0,1)$. 极差 $R = X_{(n)} - X_{(1)}$.

1°(随机向量函数) 求 R 的 pdf. 由公式 $f(X_{(1)}, X_{(n)}) = \frac{n!}{0!(n-2)!0!}x_1^0(x_n-x_1)^{n-2}(1-x_n)^0\cdot 1\cdot 1 = n(n-1)(x_n-x_1)^{n-2}$

$$\mathbb{P}(R\le x) = \int_0^{1-x}dx_1\int_{x_1}^{x_1+x} f(x_1,x_n)\,dx_n + \int_{1-x}^1 dx_1\int_{x_1}^1 f(x_1,x_n)\,dx_n = nx^{n-1}-(n-1)x^n.$$

R 的 pdf: $f_R(x) = n(n-1)x^{n-2}(1-x)$.

2° 条件密度：在 $X_{(n)} = y$ 条件下，求 $X_{(1)}$ 和 $(X_{(1)}, \dots X_{(n-1)})$ 条件密度. $\Rightarrow f_{X_{(n)}}(y) = ny^{n-1}$.

$$f_{X_{(1)}|X_{(n)}}(x_{(1)}|y) = \frac{f_{X_{(1)}X_{(n)}}(x_1, y)}{f_{X_n}(y)} = \frac{n(n-1)(y-x_1)^{n-2}}{ny^{n-1}} = \frac{(n-1)(y-x_1)^{n-2}}{y^{n-1}}.$$

$f(x_1, \dots x_n) = n!$ (联合密度) $f(x_1, \dots x_{n-1}|y) = \frac{n!}{ny^{n-1}} = \frac{(n-1)!}{y^{n-1}}$，$x_1\le \dots \le x_{n-1} \le y$.

随机变量的p分位数

设X是$(\Omega, \mathcal{F}, \mathbb{P})$上的r.v.，分布函数为$F(x)$. 　　　　该方式确定的p不唯一.

(直观定义) 对$p\in(0,1)$，若$\mathbb{P}(X\le x)\ge p$，$\mathbb{P}(X\ge x)\ge 1-p$，称$x$为X的p-分位数.

(正式定义) 对$p\in(0,1)$，定义$F^{-1}(p)=\inf\{x\mid F(x)\ge p\}$，称$F^{-1}(p)$为F或X的p分位数

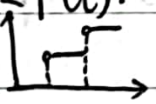
(p-quantile)，通常用ξ_p表示. $\xi_{1/2}$即为X或F的中位数. 定义等价于$F^{-1}(p)=\sup\{x\mid F(x)<p\}$.

常见应用：统计计算(要求唯一性)，统计推断(通常不要求唯一性)

性质 1° $F^{-1}(p)$单调非降

2° $F^{-1}(F(x))\le x$，$\forall x\in\mathbb{R}$　　3° $F(F^{-1}(p))\ge p$，$\forall p\in(0,1)$.

*4° $F^{-1}(p)\le t \Leftrightarrow p\le F(t)$.

5° $F^{-1}(p)$左连续. 　　6. 若F连续，则$F(F^{-1}(p))=p$，$\forall p\in(0,1)$.

(定理)：产生服从分布函数$F(x)$的随机变量：

设$X\sim U(0,1)$，$F(x)$是分布函数，则$Y=F^{-1}(X)\sim F$.

例 设$F(x)$为随机变量X的CDF，且$F(x)$是连续分布函数，则$F(X)\sim U(0,1)$.

设$Y=F(X)$，$\mathbb{P}(Y\le y)=\mathbb{P}(F(X)\le y)=\mathbb{P}(X\le F^{-1}(y))=F(F^{-1}(y))=y$，则$Y\sim U(0,1)$.

接下来考虑：易生成服从均匀分布的随机变量，如何生成服从下列分布函数的r.v：

1. $\mathcal{E}(\lambda)$. $X=-\lambda^{-1}\lg(1-U)$，$U\sim U(0,1)$.

2. Cauchy分布. $Z=X/Y$，其中X、Y相互独立且服从$N(0,1)$.

3° 令$\theta=2\pi U$，$R=\sqrt{-2\log V}$ ($U,V\sim U(0,1)$且相互独立) 则$X=R\cos\theta$和$Y=R\sin\theta$

相互独立，且都服从$N(0,1)$.

证明：(利用多个函数的联合密度) (参考前页极坐标变换例题) $R=\sqrt{-2\log U}\sim$ Rayleigh分布，$\theta=2\pi U\sim U(0,2\pi)$. R与θ独立. 由此前定理：X、Y同分布，$g:\mathbb{R}^n\to\mathbb{R}^m$可测，则$g(X)$与$g(Y)$同分布. 当时为了得到两个正态分布的极坐标变换构造了R、θ，那么$X=\sqrt{-2\log U}\cos(2\pi V)$，$Y=\sqrt{-2\log U}\sin(2\pi V)$，X Y独立且服从$N(0,1)$. (用于产生正态随机变量)

4° (离散型) $F_n(x)=\frac{1}{n}\sum_{i=1}^{n} I(x_i\le x)$，其中$x_1,\dots,x_n$是给定的观测值. (利用次序统计量)

$F_n(x)$是阶梯函数. 令$x_{(1)}\le\dots\le x_{(n)}$是$\{x_i\}$的次序观测值，因为

$F^{-1}(x)=x_{(1)}$ $(0<x\le\frac{1}{n})$，$x_{(2)}$ $(\frac{1}{n}<x\le\frac{2}{n})$，… $x_{(n)}$ $(\frac{n-1}{n}<x\le 1)$. 所以$F^{-1}(U)$的值域为$\{x_{(i)}\}$且$\mathbb{P}(F^{-1}(U)=x_{(k)})=\mathbb{P}(\frac{k-1}{n}<U\le\frac{k}{n})=\frac{1}{n}$，$k=1,\dots,n$.

期望与方差

目的：在事情发生前刻画整体情况；总结事件；便于决策。

(定义) 离散型：X 有概率分布 $\mathbb{P}(X=x_j)$，$j=1,2\dots$ 若 $\sum_{j=1}^{\infty}|x_j|\mathbb{P}(X=x_j)<+\infty$ (绝对收敛)，

称 $E(X)=\sum_{j=1}^{\infty}x_j\mathbb{P}(X=x_j)$ 为 X 的数学期望或均值。

随机变量总结为一个数：(加权)平均。有限个值的 r.v. 的期望总存在。

直观示例：A 为事件，I_A 是 A 的示性函数，则 I_A 服从两点分布，且 $EI_A=P(A)$。

· 二项分布 $X\sim B(n,p)$，$p_j=\binom{n}{j}p^j(1-p)^{n-j}$，$E(X)=np$。

(用归一化证明) $E(X)=\sum_{j=0}^{n}j\binom{n}{j}p^j(1-p)^{n-j}$，其中 $j\binom{n}{j}=n\binom{n-1}{j-1}$，

$E(X)=np\sum_{j=1}^{n}\binom{n-1}{j-1}p^{j-1}(1-p)^{n-j}=np(p+(1-p))^{n-1}=np$。

· 泊松分布 $X\sim P(\lambda)$ 同理 $E(X)=\lambda$。

· 几何分布 $X\sim G(p)$。 $E(X)=\frac{1}{1-p}$。

(定义) 连续型：设 X 有概率密度 $f(x)$，若 $\int_{-\infty}^{+\infty}|x|f(x)dx<+\infty$，

称 $E(X)=\int_{-\infty}^{+\infty}xf(x)dx$ 为 X 的数学期望或均值。

· 均匀分布 $X\sim U(a,b)$ $E(X)=\frac{a+b}{2}$。

· 指数分布 $X\sim \mathcal{E}(\lambda)$ $E(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_0^{\infty}\lambda xe^{-\lambda x}\overset{\text{(分部积分)}}{=}\frac{1}{\lambda}$。

例 任给一 Gamma 分布 ~~[illegible]~~ $f(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$，$x>0$。 $(X\sim\Gamma(\alpha,\beta))$。

$E(X)=\int_{-\infty}^{\infty}xf(x)dx=\int\frac{\beta^\alpha}{\Gamma(\alpha)}x^\alpha e^{-\beta x}dx=\frac{1}{\beta\Gamma(\alpha)}\int_0^{\infty}t^{\alpha+1-1}e^{-t}dt$，$t=\beta x$ 换元。

$\int_0^{\infty}t^{\alpha+1-1}e^{-t}dt=\Gamma(\alpha+1)\int_0^{\infty}f(t\sim\Gamma(\alpha+1,1))dt=\Gamma(\alpha+1)$，

$E(X)=\frac{1}{\beta\Gamma(\alpha)}\Gamma(\alpha+1)=\frac{\alpha}{\beta}$。

· 正态分布 $X\sim N(\mu,\sigma^2)$。先用 $x-\mu$ 换元。

由对称性、归一化，$E(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_{-\infty}^{\infty}\mu f(x)dx+\int_{-\infty}^{\infty}(x-\mu)f(x)dx$

$=\mu+\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)\underbrace{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx}_{\text{奇函数}}\to E(X)=\mu$。

·(直观利用"重心/质心")：$E(X)$ 存在，$f(x)$ 关于 $x=\mu$ 对称，即 $f(x+\mu)=f(\mu-x)$，则 $E(X)=\mu$。

一般随机变量的数学期望：$E(X)=\int_\Omega x\,d\mathbb{P}=\int_R x\,dF(x)$，$F$ 为分布函数。

(将离散的求和形式优化地写成积分形式) 对任意单调不减的右连续函数 $F(x)$，若其在 $[a,b]$ 上仅有可列个跳跃点且 $\sum_{j:x_j\in[a,b]}|f(x_j)||[F(x_j)-F(x_j-0)]|<\infty$ (绝对收敛)，则定义 $\int_a^b f(x)dF(x)=\sum_{j:x_j\in[a,b]}f(x_j)[F(x_j)-F(x_j-0)]$

若 $F(x)$ 可分解为 $F_1(x)+F_2(x)$，其中 存在 $f_0(x)$ 非负，使 $F_1(x)=\int_{-\infty}^{x}f_0(s)ds$，$F_2(x)$ 仅有可列个跳跃点，若 F_1、F_2 均绝对收敛，

则 EX 存在且 $EX=\int_{-\infty}^{+\infty}xf_0(x)dx+\sum_{j=1}^{\infty}x_j\mathbb{P}(X=x_j)=\int_{-\infty}^{\infty}x\,dF_1(x)+\int_{-\infty}^{\infty}x\,dF_2(x)=\int_{-\infty}^{+\infty}x\,dF(x)$。

随机变/向量函数的数学期望：

（定理）设 $X=(X_1, X_2, \dots X_n)$ 为随机向量（随机变量同理），$\vec{x}=(x_1, \dots x_n)\in\mathbb{R}^n$。若 X 有联合分布函数 $F(\vec{x})$，实函数 $g(\vec{x})$ 使得 $\int_{\mathbb{R}^n}|g(\vec{x})|dF(\vec{x})<\infty$，

则 $Y=g(\vec{x})$ 有期望 $E(Y)=E(g(\vec{x}))=\int_{\mathbb{R}^n}g(\vec{x})dF(\vec{x})$.

离散情形结论类似。对 $Y=g(X)$，$EY=gEX)$ 可以直接应用（g符合条件时）

$E(X+Y)=\int z f_Z(z)dz=\int(x+y)f_{X,Y}(x,y)dxdy$.

一般来说，$E(g(X))\neq g(E(X))$。以下为一些特例：

（定理）$E|X|<+\infty$，$E|Y|<+\infty$，$a,b,c\in\mathbb{R}$，则

1° $Ec=c$（常数）　　2° $|EX|\leq E|X|$.

*3° $E(aX+bY)=aEX+bEY$（Linearity，线性组合可直接提系数）

4° $X\leq Y$ a.s.，$\Rightarrow EX\leq EY$.（反过来不对）

*5° X 和 Y 独立，则 $E(XY)=(EX)(EY)$.（Fubini定理，可以直接拆乘法，要求独立）

（由此，二项分布相当于n个两点分布之和，Gamma分布可看作α个ε(β)之和）

例（邮票收藏问题）有一盒共N张邮票，从中有放回地每次抽取一张，要收集k张不同的邮票，那么抽取次数的期望？

想抽取到第i张不同的邮票，相当于概率为 $\frac{N-i+1}{N}$ 的几何分布，$\varepsilon_i=\frac{N}{N-i+1}$.

（如抽到第一张 $\varepsilon_1=1$）抽取次数的期望相当于 $\varepsilon_1\sim\varepsilon_k$ k个几何分布的期望之和（拆分成可计算的分布）$\sum_{i=1}^{k}\varepsilon_i=N\left(\sum_{m=1}^{N}\frac{1}{m}-\sum_{m=1}^{N-k}\frac{1}{m}\right)\approx N\ln\frac{N+1}{N-k+1}$.

定理 设 $EX=\mu$，$E(X^2)<\infty$，则 $E(X-\mu)^2\leq E(X-c)^2$，$\forall c\in\mathbb{R}$.

等价表达（优化角度）：$E(X)=\arg\min_{c\in\mathbb{R}} E(X-c)^2$.

证

$E(X-c)^2=E(X-\mu+\mu-c)^2=E(X-\mu)^2+2\underbrace{(\mu-c)E(X-\mu)}_{=0}+(\mu-c)^2=E(X-\mu)^2+(\mu-c)^2$.

中位数：X 有概率密度 $f(x)$，m 是其中位数，则 $m=\arg\min_{c\in\mathbb{R}}E|X-c|$.

中位数在社会统计资料中常用。优点：一定存在，不易受极值影响。

概统中期望比中位数更重要（性质优良；中位数可能不唯一）

例：$X\sim t_1$ 即 Cauchy 分布，求 $E(X)$.

考虑 $X,Y\sim N(0,1)$ 且独立，$\frac{X}{Y}\sim t_1$，$E(\frac{X}{Y})$ 是否存在？（接下页）

Date.

例 Cauchy 分布 ($X\sim t_1$) 的 EX：由对称性可直观判断 $EX=0$.

若 $X,Y\sim N(0,1)$ 且独立，则 $\frac{X}{Y}\sim t_1$. $f_X(x)=f_Y(y)=\frac{1}{\pi(1+x^2)}$.

$E\frac{X}{Y}\neq\frac{EX}{EY}$：乘/除法不能拆开. $\frac{EX}{EY}=\frac{0}{0}\Rightarrow$ 不存在.

但实际上 $E|X|=\int_{-\infty}^{+\infty}\frac{|x|}{\pi(1+x^2)}dx=\frac{2}{\pi}\int_0^{+\infty}\frac{x}{1+x^2}dx=\frac{2}{\pi}\cdot\frac{1}{2}\ln(1+x^2)\Big|_0^{+\infty}=+\infty$

Cauchy 有特殊性质，但其期望不存在，也称为 evil distribution.

例 $X\sim\Gamma(\alpha,1)$, $Y\sim\Gamma(\beta,1)$, $T=X+Y$, $W=\frac{X}{X+Y}$. X Y 相互独立. $\begin{cases}X=TW\\Y=T(1-W)\end{cases}$

1° 判断 T 和 W 是否相互独立，考虑联合 PDF 能否拆开.

$f_{T,W}(t,w)=f_{X,Y}(x,y)\left|\frac{\partial(x,y)}{\partial(t,w)}\right|=|t|f_X(x)f_Y(y)$.

$=|t|\frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-x}\frac{1}{\Gamma(\beta)}y^{\beta-1}e^{-y}=\frac{1}{\Gamma(\alpha)\Gamma(\beta)}e^{-t}(tw)^{\alpha-1}[t(1-w)]^{\beta-1}t$

$=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}w^{\alpha-1}(1-w)^{\beta-1}\cdot\frac{1}{\Gamma(\alpha+\beta)}t^{\alpha+\beta-1}e^{-t}$. 那么 $W\sim Beta(\alpha,\beta)$. $T\sim\Gamma(\alpha+\beta,1)$.

即 W 与 T 相互独立. 2° 此时 $EW=\frac{EX}{ET}=\frac{EX}{E(X+Y)}$. (一般情况不成立)

· 若 $E|X|=0$，则 $\mathbb{P}(X=0)=1$，即 $X=0$ a.s.

(证明) $P(|X|>\frac{1}{n})=P(n|X|>1)=EI_{(n|X|>1)}\leq E(n|X|I_{n|X|>1})\leq nE|X|=0$

由概率连续性，$P(|X|>0)=P(\bigcup_{n=1}^{\infty}\{|X|>\frac{1}{n}\})=\lim_{n\to\infty}P(|X|>\frac{1}{n})=0$. 最后 $P(X=0)=1$.

高阶矩 (定义) Moment 设 X 为 r.v. $m\in N^*$. 若 $E(|X|^m)<\infty$，称 $E(|X|^m)$ 为 X 的 m 阶原点矩，称 $E(X-EX)^m$ 为 X 的 m 阶中心矩.

$m>2$ 时以上原点/中心矩统称高阶矩. $m=3$ 偏度 skewness $m=4$ 峰度 kurtosis

(定义)：随机变量的方差 (Variance)

若随机变量 X 的期望 $\mu=EX$ 有限，称 $E(X-\mu)^2$ 为 X 的方差，记作 var(X) 或 σ_X^2.

当 $var(X)<\infty$ 时，称 X 的方差有限. 表示了 X 在期望附近的集中程度.

常用公式：$var(X)=E(X-EX)^2=E(X^2)-E^2(X)$. *

· 两点分布 $var(X)=p(1-p)=pq$，二项分布 $var(X)=npq$.

· 泊松分布 $E(X^2)=E(X(X-1))+E(X)=\sum_{k=0}^{\infty}k(k-1)\frac{\lambda^k}{k!}e^{-\lambda}+\lambda=\lambda^2\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda}+\lambda=\lambda^2+\lambda$,

则 $var(X)=\lambda$，即对泊松分布 $E(X)=var(X)=\lambda$.

· 几何分布 $X\sim G(p)$. $EX=\sum_{j=1}^{\infty}jpq^{j-1}=\sum_{j=0}^{\infty}jpq^{j-1}=p\sum_{j=0}^{\infty}(q^j)'$

$E(X^2)=E(X(X-1))+E(X)=\sum_{j=1}^{\infty}j(j-1)pq^{j-1}+\frac{1}{p}=pq(\sum_{j=0}^{\infty}q^j)''+\frac{1}{p}$

$=pq(\frac{1}{1-q})''+\frac{1}{p}=\frac{2q}{p^2}+\frac{1}{p}$，因此 $var(X)=\frac{q}{p^2}$.

· 均匀分布 $X\sim U(a,b)$. $E(X^2)=\int_a^b\frac{x^2}{b-a}dx=\frac{b^3-a^3}{3(b-a)}$ $(a\neq b)$

所以 $var(X)=\frac{b^3-a^3}{3(b-a)}-(\frac{a+b}{2})^2=\frac{1}{12}(b-a)^2$.

(转移到Gamma分布的PDF)

· 指数分布 $E(X^2)=\int_0^{\infty}x^2\lambda e^{-\lambda x}dx=\frac{1}{\lambda^2}\Gamma(3)=\frac{2}{\lambda^2}$，$var(X)=\frac{1}{\lambda^2}$.

· Gamma分布 $f(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}dx$. $E(X^2)=\int_0^{\infty}\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha+1}e^{-\beta x}dx=\frac{\Gamma(\alpha+2)}{\beta^2\Gamma(\alpha)}$.

(凑Gamma分布) $EX^2=\frac{\alpha(\alpha+1)}{\beta^2}$，$var(X)=\frac{\alpha(\alpha+1)}{\beta^2}-(\frac{\alpha}{\beta})^2=\frac{\alpha}{\beta^2}$.

· 正态分布 同时利用对称性、分部积分、归一化

$var(X)=\int_{-\infty}^{+\infty}(x-\mu)^2f(x)dx=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)^2\exp(-\frac{(x-\mu)^2}{2\sigma^2})dx=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t^2\exp(-\frac{t^2}{2})dt$

$=\sigma^2\int_{-\infty}^{\infty}\frac{t^2}{\sqrt{2\pi}}\exp(-\frac{t^2}{2})dt=\sigma^2$.（内部化为标准正态）

· 方差的性质（由于定义来自期望，性质也多与期望有关）设 $EX=\mu$，$var(X)<\infty$，则

1° $var(aX+b)=a^2var(X)$，（放缩，二阶期望）$a,b\in R$

2° $var(X)=E(X-\mu)^2\leq E(X-c)^2$，$\forall c\in R$

3° $var(X)=0\Leftrightarrow X=\mu$ a.s.

4° $X_1,\dots X_n$ 相互独立，则 $var(\sum_{k=1}^{n}X_k)=\sum_{k=1}^{n}var(X_k)$.

证明4°：$Var(X_1+X_2)=E((X_1+X_2)^2)-(EX_1+EX_2)^2$ ← X_1,X_2独立，$EX_1+EX_2=E(X_1+X_2)$

$=EX_1^2+2EX_1X_2+EX_2^2-(EX_1)^2-2EX_1EX_2-(EX_2)^2=Var(X_1)+Var(X_2)+2(E(X_1X_2)-EX_1EX_2)$

例 信封匹配问题：记 X_i 表示第i封匹配，共有n封信，X为总的匹配数.

$EX=\frac{1}{n}\sum_{i=1}^{n}X_i=\frac{1}{n}\cdot\sum_{i=1}^{n}EX_i=1$（互相独立）

$Var(X)=E(\sum_{i=1}^{n}X_i)^2-1=\sum EX_i+\sum_{i\neq j}E(X_iX_j)-1=1+\frac{n(n-1)}{n(n-1)}-1=1$.

即对于匹配问题，其期望、方差均为1.

(定义) 若随机变量X的期望 $\mu=EX$ 有限，记 $var(X)=E(X-\mu)^2$ 为X的方差，记作 σ_X^2，则称 $\sigma_X=\sqrt{var(X)}$ 为X的标准差 (Standard deviation)；

设 $var(X)<\infty$，令 $Y=\frac{(X-EX)}{\sqrt{Var(X)}}$. 则 $EY=0$，$var(Y)=1$. 此时称Y是X的标准化 (Standardization). 特别地，$X\sim N(\mu,\sigma^2)$ 时 $Y=\frac{(X-\mu)}{\sigma}\sim N(0,1)$.

计算公式：$var(X)=E(X^2)-E^2(X)$.