第三讲 球对称分布（II）

2025.3.3

在高维空间随机(均匀)取一个向量，它几乎与任何其它给定的或独立的随机向量正交。

球对称随机向量的计算机采样

球对称分布的表示（极分解）$\mathbf{x} \overset{d}{=} \mathbf{u}r$ 在理论证明或计算时非常有用。另一个重要的应用是用于蒙特卡洛采样/抽样（sampling），这里的采样/抽样指的是计算机随机数的生成。该表示。

蒙特卡洛（Monte Carlo,MC）需要产生给定分布的随机数或随机向量，我们默认$U(0,1)$随机数容易生成并直接利用。

Inversion sampling

Inversion sampling算法(适用于累积分布函数容易求逆的情形):
假设随机变量X的累积分布函数$F(x) = P(X < x)$连续，$\bar{F} = 1 - F$。
若$\xi \sim U(0,1)$,则$F^{-1}(\xi) \overset{d}{=} X$，$\bar{F}^{-1}(\xi) \overset{d}{=} X$.

例如，如果$R \sim p(r) = nr^{n-1}, 0 < r < 1, F(r) = r^n$。
为了生成随机数R， 可首先生成$\xi \sim U(0,1)$，令$R = \xi^{1/n}$即可。

$N(0,1)$非常重要。直接利用inversion sampling不能产生$N(0,1)$随机数, 这是因为我们没有累积分布函数Φ的精确显式表达，无法精确计算Φ^{-1}。但有趣的是，利用二元正态的极分解，我们可以很容易地同时产生两个$N(0,1)$随机数，原理如下：

已知 $\mathbf{x} \sim N_2(0, I_2) \stackrel{d}{=} \mathbf{u}R$， $\mathbf{u} \sim U(S^1), R \sim \chi_2$，只需产生$\mathbf{u}$和$R$:

- 产生$\theta \sim U(0,2\pi)$, 则 $\mathbf{u} = (\cos(\theta), \sin(\theta))^\top \sim U(S^1)$
- $R \sim \chi_2$的密度 $p_{\chi_2}(r) = r\exp(-r^2/2)$， $\bar{F}(r) = \exp(-r^2/2)$, 则 $\xi = \bar{F}(R) \sim U(0,1), R = \sqrt{-2\log(\xi)} \sim \chi_2$，其中$\xi \sim U(0,1)$，$\xi \perp\!\!\!\perp \theta$。

正态抽样：
Box-Muller
算法

Box-Muller方法 ($N_2(0, I_2)$抽样):

- 产生$\theta \sim U(0,2\pi)$，$\xi \sim U(0,1)$，独立;
- $R = \sqrt{-2\log(\xi)}$， $\mathbf{u} = (\cos(\theta), \sin(\theta))^\top$;
- $\mathbf{x} = R\mathbf{u} \sim N_2(0, I_2)$.

利用Box-Muller方法独立产生若干正态随机变量，合在一起即得到多元标准正态随机向量。所以正态随机向量非常容易产生。

$U(S^{n-1})$ 抽样

由第二讲定理1，若$\mathbf{z}\sim N_n(0,I_n)$，则$\mathbf{u}=\frac{\mathbf{z}}{\|\mathbf{z}\|}\sim U(S^{n-1})$，因此将$N_n(0,I_n)$随机向量单位化即得到球面均匀随机向量。

$U(S^{n-1})$抽样

- 产生$\mathbf{z}\sim N_n(0,I_n)$；
- 令 $\mathbf{u}=\frac{\mathbf{z}}{\|\mathbf{z}\|}$，则$\mathbf{u}\sim U(S^{n-1})$;

一般球对称随机向量的抽样

假设球对称随机向量$\mathbf{x}\sim f(\mathbf{x})=h(\|\mathbf{x}\|)$， 由第二讲定理1的推论，

$$\mathbf{x}\overset{d}{=}\mathbf{u}R,\ \mathbf{u}\perp\!\!\!\perp R,\ \mathbf{u}\sim U(S^{n-1}),R\sim p(r)=\frac{2\pi^{n/2}}{\Gamma(n/2)}h(r)r^{n-1}$$

球对称分布抽样：

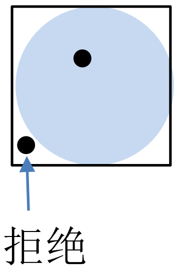
- 产生$\mathbf{z}\sim N_n(0,I_n)$；
- 令 $\mathbf{u}=\frac{\mathbf{z}}{\|\mathbf{z}\|}$，则$\mathbf{u}\sim U(S^{n-1})$;
- 产生一元$R\sim p(r)=\frac{2\pi^{n/2}}{\Gamma(n/2)}h(r)r^{n-1}$
- 令$\mathbf{x}=\mathbf{u}R$, 则$\mathbf{x}\sim f(\mathbf{x})=h(\|\mathbf{x}\|)$。

关键是产生随机变量R，可采用inversion sampling,；如果不可行，可采用拒绝抽样，importance sampling 等等 (参见附录1)

例如，$U(B^n)$采样: 生成$\xi\sim U(0,1)$，$\mathbf{z}\sim N_n(0,I_n)$，$\mathbf{x}=\frac{\mathbf{z}}{\|\mathbf{z}\|}\xi^{1/n}$

$U(B^n)$的拒绝抽样

拒绝采样或拒绝-接受采样是一种应用广泛的采样方式，尤其是均匀分布采样经常采用。拒绝采样的原理是若$x \sim U(C)$, $C_0 \subset C$, 则 $x|_{x\in C_0} \sim U(C_0)$.



若$u_1, \dots, u_n \ iid \sim U(-1,1)$, 则
$\mathbf{u} = (u_1, \dots, u_n)^\top \sim U([-1,1]^n)$

$U(B^n)$的拒绝抽样：

为了生成$\mathbf{x} \sim U(B^n)$，我们先在容易采样的区域$[-1,1]^n \supset B^n$中均匀采样$\mathbf{x} \sim U([-1,1]^n)$，如果得到的点落在$B^n$内，这就是我们需要的$\mathbf{x}$；若该点落在球外，则拒绝该点。

上述方法适用于n较小的情形，当n较大，比如$n = 10$时，单位球与超立方体$[-1,1]^{10}$的体积之比为0.0025，这意味着生成10000个$U([-1,1]^{10})$, 其中只有约25个落在单位球内，采样效率低下.

球对称分布的性质

我们已给出了球对称分布的极分解或随机表示

$$\mathbf{x} \overset{d}{=} \mathbf{u}r, \mathbf{u} \sim U(S^{n-1}), r \overset{d}{=} \|\mathbf{x}\|,$$

下面进一步利用旋转不变性考虑其分布性质，尤其是球面均匀分布及其边际分布，我们将大量采用同分布等号 " $\overset{d}{=}$ "

同分布等号

同分布等号 $\overset{d}{=}$ 可传递，但不是真正意义上的数学等号，同分布等号操作需要小心， 比如对于3个r.v. x, y, z

$$x \overset{d}{=} y \text{ 不一定蕴含 } x + z \overset{d}{=} y + z$$

> 引理 1：若随机向量 $\mathbf{x} \overset{d}{=} \mathbf{y}$， 则 $g(\mathbf{x}) \overset{d}{=} g(\mathbf{y})$，其中 g 是任一（可测）向量函数。特别地，$x_i \overset{d}{=} y_i$，$x_i x_j \overset{d}{=} y_i y_j$.

注：该引理的关键是等号两边的内部做同样的运算/函数，分布依旧相同。

证：$P(g(\mathbf{x}) \in C) = P\big(\mathbf{x} \in g^{-1}(C)\big) = P\big(\mathbf{y} \in g^{-1}(C)\big) = P(g(\mathbf{y}) \in C)$

问题：已知 $\mathbf{x} \overset{d}{=} \mathbf{y}$，何时 $\mathbf{x}+\mathbf{z} \overset{d}{=} \mathbf{y}+\mathbf{z}$?

由引理1，一个充分条件是 $\begin{pmatrix}\mathbf{x}\\ \mathbf{z}\end{pmatrix} \overset{d}{=} \begin{pmatrix}\mathbf{y}\\ \mathbf{z}\end{pmatrix}$

当$\mathbf{z}$与$\mathbf{x}$, $\mathbf{y}$都独立时上式成立, 特别地，当$\mathbf{z}$是常向量时成立。

Cramer-Wold定理说明两个随机向量同分布，当且仅当它们在任何单位方向的投影坐标同分布.

> 引理2 (Cramer-Wold定理) 假设随机向量$\mathbf{x}$，$\mathbf{y} \in R^n$，则
> $$\mathbf{x} \overset{d}{=} \mathbf{y} \Leftrightarrow \mathbf{t}^\top\mathbf{x} \overset{d}{=} \mathbf{t}^\top\mathbf{y}, \forall\, \mathbf{t} \in S^{n-1}。$$

证明:(⇒) 由引理1立得。

(⇐) 对任何$\forall\, \mathbf{t} \in R^n$, $\mathbf{t}/\|\mathbf{t}\| \in S^{n-1}$, $\mathbf{t}^\top\mathbf{x}/\|\mathbf{t}\| \overset{d}{=} \mathbf{t}^\top\mathbf{y}/\|\mathbf{t}\|$

$\Rightarrow \mathbf{t}^\top\mathbf{x} \overset{d}{=} \mathbf{t}^\top\mathbf{y} \Rightarrow$ 矩母函数 $E\left(\exp(\mathbf{t}^\top\mathbf{x})\right) = E\left(\exp(\mathbf{t}^\top\mathbf{y})\right) \Rightarrow \mathbf{x} \overset{d}{=} \mathbf{y}$

注：矩母（或特征）函数唯一决定$\mathbf{x}$的分布表达的含义实际上就是“所有一维投影坐标 $\mathbf{t}^\top\mathbf{x}$的分布决定$\mathbf{x}$整体的分布”，

球对称分布的一元边际

命题1：若$\mathbf{x}$服从球对称分布，则$\mathbf{x}$所有分量同分布(不独立)，且
$$E(\mathbf{x})=\mathbf{0}, var(\mathbf{x})=\sigma^2 I_n$$

证明：因为置换变换是正交变换，所以 $\mathbf{x}$各个分量同分布，方差相同；因为反射矩阵$H=diag(1,1,\dots,-1,1,\dots 1)$是正交变换，我们有$x_i \overset{d}{=} -x_i$，$x_i x_j \overset{d}{=} -x_i x_j$，从而$E(x_i)=0, cov(x_i,x_j)=0$.

球对称分布随机向量$\mathbf{x}$的各个分量x_i同分布，因为旋转不变性，它在任何方向 $\mathbf{t}\in S^{n-1}$上的投影坐标与x_i同分布

定理1. $\mathbf{x}$服从球对称分布 $\Leftrightarrow \mathbf{t}^\top\mathbf{x}\overset{d}{=}x_1, \forall\, \mathbf{t}\in S^{n-1}$。

证明：($\Rightarrow$)任取正交矩阵H,其第一行为$\mathbf{t}$, 则 $H\mathbf{x}\overset{d}{=}\mathbf{x}$，特别地两边的第一分量等分布，即$\mathbf{t}^\top\mathbf{x}\overset{d}{=}x_1$。

($\Leftarrow$) 对任何正交矩阵H，为了证明$H\mathbf{x}\overset{d}{=}\mathbf{x}$，由引理2，我们只需对任何$\mathbf{t}\in S^{n-1}$证明$\mathbf{t}^\top H\mathbf{x}\overset{d}{=}\mathbf{t}^\top\mathbf{x}$.因为$\mathbf{t}\in S^{n-1}$，$H^\top\mathbf{t}\in S^{n-1}$, 故 $\mathbf{t}^\top\mathbf{x}\overset{d}{=}x_1$, 且$(H^\top\mathbf{t})^\top\mathbf{x}\overset{d}{=}x_1$，从而$(H^\top\mathbf{t})^\top\mathbf{x}\overset{d}{=}\mathbf{t}^\top\mathbf{x}$

$U(S^{n-1})$的一维边际

定理2（第一讲P38推论1）. $\mathbf{u} = (u_1, \dots, u_n)^\top \sim U(S^{n-1})$，则$u_1$的概率密度函数为

$$f_n(u) = \frac{1}{B\left(\frac{n-1}{2}, \frac{1}{2}\right)} (1 - u_1^2)^{(n-3)/2},\ |u_1| < 1$$

（这是[−1,1]区间上的beta分布）

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

注：等价地，$\frac{u_1+1}{2} \sim beta\left(\frac{n-1}{2}, \frac{n-1}{2}\right)$，$\sqrt{n-1}\frac{u_1}{\sqrt{1-u_1^2}} \sim t_{n-1}$，

证明1：引入球坐标

$$\mathbf{u} = (\cos(\theta_1), \sin(\theta_1)\cos(\theta_2), \dots, \sin(\theta_1)\cdots\sin(\theta_{n-1}))^\top$$

由第二讲推论1，$\theta_k \sim p_k(\theta_k) \propto \sin^{n-k-1}(\theta_k), k = 1, \dots, n-1$，特别地

$$\theta_1 \sim p_1(\theta_1) \propto \sin^{n-2}(\theta_1)，0 < \theta_1 < \pi$$

因为$\int_0^\pi \sin^{n-k-1}(\theta_k)\, d\theta_k = B\left(\frac{n-k}{2}, \frac{1}{2}\right)$，所以$\theta_1$的概率密度为

$$p_1(\theta_1) = \frac{1}{B\left(\frac{n-1}{2}, \frac{1}{2}\right)} \sin^{n-2}(\theta_1)$$

由此可得$u_1 = \cos(\theta_1)$的概率密度

$$f_{n,1}(u) = \frac{1}{B\left(\frac{n-1}{2}, \frac{1}{2}\right)} (1 - u^2)^{(n-3)/2}$$

证明2：不妨假设$\mathbf{u}=\frac{\mathbf{x}}{\|\mathbf{x}\|}$, $\mathbf{x}\sim N_n(0,I_n)$, 所以

$\mathbf{x}\sim N_n(0,I_n)\Rightarrow \mathbf{x}/\|\mathbf{x}\|\sim U(S^{\,n-1})$.

$$u_1=\frac{x_1}{\sqrt{x_1^2+\cdots+x_n^2}},\quad x_1,\ldots,x_n\ iid \sim N(0,1)$$

则$u_1^2=\frac{x_1^2}{x_1^2+x_2^2+\cdots+x_n^2}\sim beta\left(\frac{1}{2},\ \frac{n-1}{2}\right)$,

$$\frac{\chi_a^2}{\chi_a^2+\chi_b^2}\overset{d}{=}beta\left(\frac{a}{2},\ \frac{b}{2}\right)$$

由此可求出u_1的概率密度。另外，显然

$$\sqrt{n-1}\frac{u_1}{\sqrt{1-u_1^2}}=\frac{\sqrt{n-1}x_1}{\sqrt{x_2^2+\cdots+x_n^2}}\sim t_{n-1}。$$

我们将采用第二种方法求$U(S^{\,n-1})$的多元边际分布（极坐标方法稍复杂）。

u_1的性质罗列

- $E(u_1) = 0, var(u_1) = \frac{1}{n}$

  $1 = E(u_1^2 + \cdots + u_n^2) = nE(u_1^2)$，所以$E(u_1^2) = \frac{1}{n}$.

- 一元边际u_1集中于0附近, 由切比雪夫不等式

  $$P(|u_1| < \varepsilon) > 1 - \frac{1}{n\varepsilon^2} \to 1, n \to \infty$$

  高维球面面积大部分在赤道带附近.

- $\sqrt{n}u_1 \to N(0,1), n \to \infty$

  $z = \sqrt{n}u_1$的密度

  $g_n(u) = \frac{1}{\sqrt{n}B\left(\frac{n-1}{2},\frac{1}{2}\right)}(1 - z^2/n)^{(n-3)/2} \to \frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2}{2})$

- $n = 2$时, $u_1 \sim f_{2,1}(u_1) = \frac{1}{\pi\sqrt{1-u_1^2}}$， $|u_1| < 1$ (arcsin law).
- $n = 3$时, $u_1 \sim f_{3,1}(u_1) = 1/2$, $|u_1| < 1$ (阿基米德);
- 球对称情形下(含正态)，相关系数可看作是u_1 （见后面）

$U(S^{n-1})$ 的随机投影

即使定理1中方向 $\mathbf{t}\in S^{n-1}$是随机的，投影坐标$\mathbf{t}^\top\mathbf{x}$仍与$x_i$同分布

推论1. 若$\mathbf{x}$服从球对称分布， 随机向量 $\mathbf{z}\in S^{n-1}$， $\mathbf{z}\perp\!\!\!\perp\mathbf{x}$， 则 $\mathbf{z}^\top\mathbf{x}\overset{d}{=}x_1$, 且 $\mathbf{z}^\top\mathbf{x}\perp\!\!\!\perp\mathbf{z}$ 。

证明：给定$\mathbf{z}$的条件下， $\mathbf{z}^\top\mathbf{x}$中$\mathbf{z}$是常数向量， 由定理1, $\mathbf{z}^\top\mathbf{x}\,|\,\mathbf{z}\overset{d}{=}x_1$， 该条件分布即$x_1$的分布不依赖于$\mathbf{z}$，所以 $\mathbf{z}^\top\mathbf{x}\perp\!\!\!\perp\mathbf{z}$ ， 且 $\mathbf{z}^\top\mathbf{x}\overset{d}{=}x_1$。

推论2. 若$\mathbf{u}\sim U(S^{n-1})$, $\mathbf{z}\perp\!\!\!\perp\mathbf{u}$, 则 $\mathbf{z}^\top\mathbf{u}\xrightarrow{\mathrm{P}}0$， $n\to\infty$

证明：$P(|\mathbf{z}^\top\mathbf{u}|>\varepsilon)=P(|u_1|>\varepsilon)<1/n\varepsilon^2\to 0$， $n\to\infty$。

注：高维空间中（ $n\to\infty$ ）随机/均匀取一个方向$\mathbf{u}$，它与任何独立向量$\mathbf{z}$几乎正交，与任何给定的方向$\mathbf{t}$几乎正交。

推论3：假设$\mathbf{u},\mathbf{v}$ iid$\sim U(S^{n-1})$, 则

$$\mathbf{u}^\top\mathbf{v}\overset{d}{=}u_1\overset{d}{=}v_1,\ \mathbf{u}^\top\mathbf{v}\perp\!\!\!\perp\mathbf{u},\ \mathbf{u}^\top\mathbf{v}\perp\!\!\!\perp\mathbf{v}$$

注：推论3说明“在球面/圆周上随机地选取两个点”等价于“给定一个点，随机地选取另外一个点”（参见贝特朗悖论解法1）。

$U(S^{n-1})$ 的唯一性

定理3. 球面均匀分布是球面上唯一的球对称/正交旋转不变的分布:

证明：已知球面均匀分布是球对称的。我们只需证明：
若随机向量$\mathbf{v}$ 在球面S^{n-1}上是球对称的，则必定$\mathbf{v}\sim U(S^{n-1})$。

取$\mathbf{u}\sim U(S^{n-1})$, $\mathbf{u}$与$\mathbf{v}$独立，下面证明$\mathbf{v}$与$\mathbf{u}$同分布。

$\mathbf{u}$，$\mathbf{v}\in S^{n-1}$都是球对称的，由推论1，$\mathbf{v}^\top\mathbf{u}\overset{d}{=}u_1$，且$\mathbf{v}^\top\mathbf{u}\overset{d}{=}v_1$，
所以$u_1\overset{d}{=}v_1$

（进而$u_i\overset{d}{=}v_i, i=1,\dots,n$,但这并不蕴含$\mathbf{u},\mathbf{v}$同分布）

对任何常数向量 $\mathbf{t}\in S^{n-1}$，因为$\mathbf{u}$，$\mathbf{v}$都是球对称的，

$$\mathbf{t}^\top\mathbf{u}\overset{d}{=}u_1,\quad \mathbf{t}^\top\mathbf{v}\overset{d}{=}v_1\quad\Rightarrow\quad \mathbf{t}^\top\mathbf{v}\overset{d}{=}\mathbf{t}^\top\mathbf{u}$$

由引理2（Cramer-Wold定理），$\mathbf{v}$与 $\mathbf{u}$同分布，从而$\mathbf{v}\sim U(S^{n-1})$。

证毕。

推论4. 若$\mathbf{x}$服从球对称分布，则$\mathbf{x}/\|\mathbf{x}\| \sim U(S^{n-1})$, 且与 $\|\mathbf{x}\|$独立。

证明：因为$\mathbf{x}$球对称，所以$\mathbf{u} \triangleq \mathbf{x}/\|\mathbf{x}\|$ 也是球对称的，且 $\mathbf{u} \in S^{n-1}$，由定理2，必定$\mathbf{u} \sim U(S^{n-1})$.

注：推论4的结论在第二讲定理1已经得到了。这里我们利用球面均匀分布唯一性（定理3）重新给出了一个证明。

Galton相关系数

$n \times 1$随机向量$\mathbf{x}, \mathbf{y}$的Galton相关系数

$$R = \mathbf{x}^\top \mathbf{y}/\|\mathbf{x}\|\|\mathbf{y}\| = \mathbf{u}^\top \mathbf{v}, \quad \mathbf{u} = \mathbf{x}/\|\mathbf{x}\|, \mathbf{v} = \mathbf{y}/\|\mathbf{y}\|$$

推论5. 假设$\mathbf{x}$是独立的球对称$n \times 1$随机向量，则$R = \mathbf{x}^\top \mathbf{y}/\|\mathbf{x}\|\|\mathbf{y}\|$具有概率密度

$$f_n(r) = \frac{1}{B\left(\frac{n-1}{2}, \frac{1}{2}\right)} (1 - r^2)^{(n-3)/2}, |r| \le 1$$

且$R^2 \sim beta\left(\frac{1}{2}, \frac{n-1}{2}\right)$，$\sqrt{n-1}\frac{R}{\sqrt{1-R^2}} \sim t_{n-1}$

证明：$\mathbf{u} \sim U(S^{n-1}), \mathbf{v} = \mathbf{y}/\|\mathbf{y}\| \in S^{n-1}$, 由推论2，$R = \mathbf{u}^\top \mathbf{v}$与 u_1同分布，应用定理2即得。

Pearson 相关系数

作业：假设独立的随机向量 $\mathbf{x},\mathbf{y}\in R^n$，假设 $\mathbf{y}$ 是球对称的，它们的Pearson相关系数 $r_{xy}=\frac{(\mathbf{x}-\mathbf{1}\bar{x})^{\top}(\mathbf{y}-\mathbf{1}\bar{y})}{\|\mathbf{x}-\mathbf{1}\bar{x}\|\|\mathbf{y}-\mathbf{1}\bar{y}\|}$,则

$$\sqrt{n-2}\frac{r_{xy}}{\sqrt{1-r_{xy}^2}}\sim t_{n-2}$$

等价地，$r_{xy}^2\sim beta\left(\frac{1}{2},\ \frac{n-2}{2}\right)$。

r_{xy}^2的密度：$f_n(r)=\frac{1}{B\left(\frac{n-2}{2},\frac{1}{2}\right)}(1-r^2)^{(n-4)/2},\ |r|<1$

特例：假设$(x_i,y_i), i=1,\dots,n$ iid$\sim N_2\left(\begin{pmatrix}\mu_x\\ \mu_y\end{pmatrix},\begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{pmatrix}\right)$

则$\mathbf{y}=(y_1,\dots,y_n)^{\top}\sim N_n(\mathbf{1}\mu_x,\sigma_x^2 I_n)$是球对称的，则当$\rho=0$时，$\mathbf{y}\perp\!\!\!\perp\mathbf{x}$，此时$\sqrt{n-2}\frac{r_{xy}}{\sqrt{1-r_{xy}^2}}\sim t_{n-2}$

注1：实际上(x_i,y_i)不必假设服从二元正态，只需要假设$y_i|x_i$服从一元正态，这正是线性模型中的假设。

注2：上述结果表明正态分布中的一些t、F分布理论结果在一般的球对称分布假设下也成立。

$U(S^{n-1})$的多元边际

定理4（第一讲定理1）. 假设 $\mathbf{u}=(u_1,\ldots,u_n)^\top \sim U(S^{n-1})$，则对任何$1\le k\le n-1, \mathbf{u}^{(k)}=(u_1,\ldots,u_k)^\top$的概率密度：

$$f_{n,k}(\mathbf{u}^{(k)})=\frac{\Gamma(n/2)}{\pi^{k/2}\Gamma((n-k)/2)}\left(1-\left\|\mathbf{u}^{(k)}\right\|^2\right)^{(n-k-2)/2}, \mathbf{u}^{(k)}\in B^k.$$

$k=1$时即定理2.

证明： $\mathbf{u}$球对称 $\Rightarrow \mathbf{u}^{(k)}=(u_1,\ldots,u_k)^\top$球对称，其密度只与 $r=\left\|\mathbf{u}^{(k)}\right\| = \sqrt{u_1^2+\cdots+u_k^2}$有关，我们只需求$r$的分布。

因为$\mathbf{u}\sim U(S^{n-1})$，不妨设$\mathbf{u}=\mathbf{x}/\|\mathbf{x}\|$， $\mathbf{x}\sim N_n(\mathbf{0},I_n)$，所以 $\left\|\mathbf{u}^{(k)}\right\|^2=\frac{x_1^2+\cdots+x_k^2}{x_1^2+\cdots+x_n^2}\sim Beta\left(\frac{k}{2},\frac{n-k}{2}\right) \Rightarrow \left\|\mathbf{u}^{(k)}\right\|$ 的密度

$$p(r)=\frac{2\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n-k}{2}\right)}r^{k-1}(1-r^2)^{\frac{n-k}{2}-1}$$

则由第二讲定理1（p12)或P14，$\mathbf{u}^{(k)}$的概率密度

$$f_{n,k}(\mathbf{u}^{(k)})=\frac{2\Gamma\left(\frac{k}{2}\right)}{2\pi^{k/2}\Gamma\left(\frac{n-k}{2}\right)\left\|\mathbf{u}^{(k)}\right\|^{k-1}}\times p\left(\left\|\mathbf{u}^{(k)}\right\|\right)$$

$$=\frac{\Gamma(n/2)}{\pi^{k/2}\Gamma((n-k)/2)}\left(1-\left\|\mathbf{u}^{(k)}\right\|^2\right)^{(n-k-2)/2}$$

注：另一种传统的证法没注意到$\mathbf{u}^{(k)}$的球对称性，利用Dirichelet分布求$\frac{x_1^2}{x_1^2+\cdots+x_n^2},\ldots,\frac{x_k^2}{x_1^2+\cdots+x_n^2}$的联合分布，比较复杂（参见附录2）

推广的阿基米德定理

推论6. (降2个维度仍然均匀) $U(S^{n-1})$的$k=n-2$元边际分布：

$$f\left(\mathbf{u}^{(n-2)}\right)=\frac{\Gamma(n/2)}{\pi^{(n-2)/2}}\ ,\quad \mathbf{u}^{(n-2)}=\mathbf{u}_{[1:(n-2)]}\in B^{n-2}$$

是R^{n-2}中单位球内的均匀分布$U(B^{n-2})$

该结果将单位球$B^{n-2}\subset R^{n-2}$和单位球面$S^{n-1}\subset R^n$联系在一起:

- $\mathbf{x}\sim U(B^n)\Rightarrow \mathbf{u}=\mathbf{x}/\|\mathbf{x}\|\sim U(S^{n-1})$
- $\mathbf{u}\sim U(S^{n-1})\ \Rightarrow\ \mathbf{u}_{[1:(n-2)]}\sim U(B^{n-2})$

两个问题可以互相转化

比如 $n\times 1$向量$\mathbf{x}\sim U(B^n)$ 的边际分布是否可能在低维球内均匀？
$\mathbf{x}$可以看作是$n+2$维欧氏空间中$\mathbf{u}\sim U\left(S^{(n+2)-1}\right)$的$n$维投影，
$\mathbf{x}=\ \mathbf{u}_{[1:n]}$，因为$\mathbf{u}$不可能有其它投影均匀，所以$\mathbf{x}$不可能在低维球内边际均匀。

$\mathbf{u} \sim U(S^{n-1})$，$\|\mathbf{u}\| = 1$.
$\mathbf{u}^{(n-1)} = (u_1, \dots, u_{n-1})^{\top}$的概率密度可以认为是$\mathbf{u}$的“概率密度”

$\mathbf{u} \sim U(S^{n-1})$的密度

推论7. 假设 $\mathbf{u} \sim U(S^{n-1})$， $\mathbf{u}^{(n-1)} = (u_1, \dots, u_{n-1})^{\top}$的概率密度

$$f\left(\mathbf{u}^{(n-1)}\right) = \frac{\Gamma(n/2)}{\pi^{n/2}} \frac{1}{\sqrt{1-u_1^2-\cdots-u_{n-1}^2}},\ \mathbf{u}^{(n-1)} \in B^{n-1}.$$

推论8. 假设 $\mathbf{u} \sim U(S^{n-1})$，则$\mathbf{v} = (u_1^2, \dots, u_{n-1}^2)^{\top}$的概率密度函数为

$$f(\mathbf{v}) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)^{n-1}\sqrt{\pi}} v_1^{-1/2} \cdots v_{n-1}^{-1/2} [1 - (v_1 + \cdots + v_{n-1})]^{-1/2},$$

$0 \le v_i \le 1,\ v_1 + \cdots + v_{n-1} \le 1$

记作 $\mathbf{v} \sim Dirichelet\left(\frac{1}{2}, \dots, \frac{1}{2}; \frac{1}{2}\right)$， 狄利克雷分布。

注：该分布描述的是

$$\mathbf{v} = \left(\frac{x_1^2}{x_1^2 + \cdots + x_n^2}, \dots, \frac{x_{n-1}^2}{x_1^2 + \cdots + x_n^2}\right)^{\top}$$

的联合分布，其中$x_1, \dots, x_n$ $iid \sim N(0,1)$. 参见附录。

附录1：拒绝抽样Rejection sampling（acceptance-rejection sampling）

均匀分布

定义. 假设集合$C \subset R^n$的体积$|C| < \infty$，如果对任何 $D \subset C$,

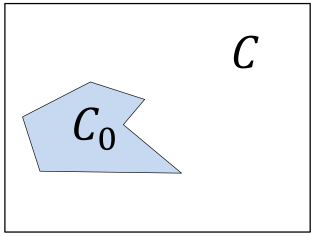
$$P(\mathbf{x} \in D) = |D|/\ |C|$$

则称 $\mathbf{x}$ 在C上均匀分布,记作$\mathbf{x} \sim U(C)$.

由均匀分布的定义，立得如下抽样基本原理

抽样基本原理1

假设 $\mathbf{x} \sim U(C)$，$C_0 \subset C$，则$\mathbf{x}|_{\mathbf{x} \in C_0} \sim U(C_0)$，即限制在子集$C_0$上也是均匀分布。



证明：任取$D \subset C_0$，

$$P\left(\mathbf{x} \in D|_{\mathbf{x} \in C_0}\right) = \frac{P(\mathbf{x} \in D)}{P(\mathbf{x} \in C_0)} = \frac{|D|/\ |C|}{|C_0|/\ |C|} = \frac{|D|}{|C_0|},$$

按照均匀分布的定义，上式表明$\mathbf{x}|_{\mathbf{x} \in C_0} \sim U(C_0)$。

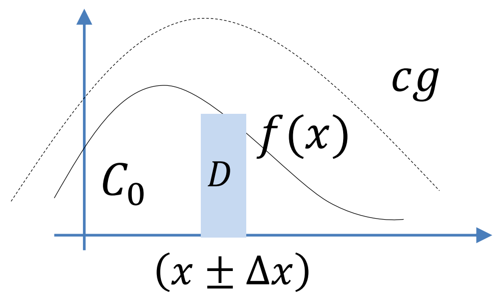
简单拒绝抽样

假设区域C_0不规则，难以从中生成均匀随机数。为了产生$\mathbf{x} \sim U(C_0)$，选取一个容易生成均匀随机数的规则区域C $\supset C_0$, 比如立方体， 产生$\mathbf{x} \sim U(C)$。若$\mathbf{x} \in C_0$，则$\mathbf{x}$即为所求；否则，舍弃/拒绝。

规则区域C包含C_0，C越小抽样效率越高。

抽样基本原理2

考虑从一元概率密度f中抽样，记$C_0 = sub(f) = \{(x,y): 0 \le y \le f(x)\}$为密度函数$f$的下方与横轴之间的区域（subgraph），若$(X,Y) \sim U(C_0)$，则$X \sim f$。



证明：注意C_0面积等于1

$$P(X \in (x \pm \Delta x))$$
$$= P((X,Y) \in D) = \frac{|D|}{|C_0|}$$
$$= f(x)\Delta x$$

拒绝抽样原理

为了从密度f抽样， 选取一个容易产生随机数的概率密度g, 假设存在$c \ge 1$, 使得 $cg \ge f$。 产生

$$(X,Y) \sim U(sub(cg)),$$

由基本原理2，若$Y \le f(X)$， 则X即为所求， 否则拒绝。

注意$(X,Y) \sim U\big(sub(cg)\big) \Leftrightarrow Y|X \sim U\big(0, cg(X)\big)\ X \sim g$，令$U \triangleq \frac{Y}{cg(X)}$ $\Rightarrow U|X \sim U(0,1)$，该分布与条件无关，故$U \perp\!\!\!\perp X$, 所以前述接受随机数的条件“$Y \le f(X)$”$\Leftrightarrow$“$U \le \frac{f(X)}{cg(X)}$”。基于这些观察，重述上述拒绝抽样过程如下（更容易，不必真的从$sub(cg)$抽样）：

拒绝抽样

为了从密度f抽样，选取一个容易产生随机数的概率密度g, 假设存在c ≥ 1, 使得 $cg \geq f$。

Rejection sampling：产生$X \sim g, U \sim U(0,1), U \perp\!\!\!\perp X$，若$U \leq \frac{f(X)}{cg(X)}$，则$X$即为所求，否则拒绝。

该方法首先产生一个$r.v.\ X \sim g$，计算$p = \frac{f(X)}{cg(X)}$。再独立地产生一个$U \sim U(0,1)$，若$U \ \leq p$, 则接受X.

所以拒绝抽样方法在产生X后采取了随机化方式（抛一个正面概率为p的硬币）来决定是否接受X。这看起来很神秘。
MCMC中的Metropolis-Hastings随机抽样方法原理与此类似。

Importance sampling

考虑计算积分$\int h(x)dx$,取某个概率密度函数f (比如正态密度), 改写

$$\int h(x)dx = \int \frac{h(x)}{f(x)} f(x)dx$$

Importance sampling: 采样$x_1, \dots, x_N$ iid $\sim f$, 计算$\sum \frac{h(x_i)}{f(x_i)} / N \approx \int h(x)dx$

$$\sum \frac{h(x_i)}{f(x_i)} / N \to E\left(\frac{h(x_i)}{f(x_i)}\right) = \int h(x)dx \ .$$

附录2：狄利克雷(Dirichlete)分布 (可忽略)

Dirichlet分布

定义1：若随机向量$\mathbf{x} = (x_1, \dots, x_n)^\top$的概率密度函数为

$$p(\mathbf{x}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_{n+1})}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{n+1})} \prod_{i=1}^{n} x_i^{\alpha_i - 1} \left(1 - \sum_{i=1}^{n} x_i\right)^{\alpha_{n+1} - 1},$$

$$0 < x_i < 1, 0 < x_1 + \cdots + x_n < 1,$$

我们称$\mathbf{x} \sim \mathrm{Dirichlet}(\alpha_1, \dots, \alpha_n; \ \alpha_{n+1})$.

分号表示地位不同

注1：beta分布是$n = 1$ 时的Dirichlet分布：

$$p(x) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1 - 1}(1 - x)^{\alpha_2 - 1}, \quad 0 < x < 1,$$

记作$x \sim \mathrm{Beta}(\alpha_1, \ \alpha_2)$或$x \sim \mathrm{Dirichlet}(\alpha_1; \alpha_2)$.

镜像对称：$x \sim \mathrm{Beta}(\alpha_1, \ \alpha_2) \Rightarrow 1 - x \sim \mathrm{Beta}(\alpha_2, \ \alpha_1)$。

注2：定义1 中若记$x_{n+1} = 1 - \sum_{i=1}^{n} x_i$，则密度具有完美的对称形式,因为$\sum_{i=1}^{n+1} x_i = 1$，$x_i > 0$，所以D分布是比率变量的分布。

Dirichelet随机向量可由独立Gamma或卡方随机变量生成：

命题A1. 假设$z_1,\dots,z_n,z_{n+1}$独立，$z_i\sim\text{Gamma}(\alpha_i),\alpha_i>0,i=1,\dots,n+1$. 令，

$$t=z_1+\cdots+z_n+z_{n+1},\quad x_i=z_i/t=z_i/(z_1+\cdots+z_{n+1}),\ i=1,\dots,n,$$

则$\mathbf{x}=(x_1,\dots,x_n)^\top$与$t$独立，$t\sim\text{Gamma}(\sum_{i=1}^{n+1}\alpha_i)$且$\mathbf{x}\sim\text{Dirichlet}(\alpha_1,\dots,\alpha_n;\ \alpha_{n+1})$.

证明：容易验证Jacobi $J=J(z_1,\dots,z_n,z_{n+1}\to x_1,\dots,x_n,t)=t^n$.

注意$z_i=x_it,1\le i\le n,\ z_{n+1}=t-t\sum_{i=1}^n x_i$, 所以$x_1,\dots,x_n,t$的联合密度

$$p(x_1,\dots,x_n,t)$$

$$=\left(\prod_{i=1}^n\frac{1}{\Gamma(\alpha_i)}(x_it)^{\alpha_i-1}e^{-x_it}\right)\frac{1}{\Gamma(\alpha_{n+1})}\left(t-t\sum_{i=1}^n x_i\right)^{\alpha_{n+1}-1}e^{-(t-t\sum_{i=1}^n x_i)}\times J$$

$$=\frac{\Gamma(\sum_{i=1}^{n+1}\alpha_i)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{n+1})}\prod_{i=1}^n x_i^{\alpha_i-1}\left(1-\sum_{i=1}^n x_i\right)^{\alpha_{n+1}-1}\frac{1}{\Gamma(\sum_{i=1}^{n+1}\alpha_i)}t^{\alpha_1+\cdots+\alpha_{n+1}-1}e^{-t}$$

所以$t=z_1+\cdots+z_{n+1}\sim\text{Gamma}(\sum_{i=1}^{n+1}\alpha_i)$， t与$x_1,\dots,x_n$独立，且 $\mathbf{x}=(x_1,\dots,x_n)^\top\sim\text{Dirichlet}(\alpha_1,\dots,\alpha_n;\ \alpha_{n+1})$.

因为 $\chi_k^2/2 \overset{d}{=} Gamma(k/2)$分布，所以有如下推论.

推论A1. 假设$z_1, \dots, z_n, z_{n+1}$独立，$z_i \sim \chi_{d_i}^2, i = 1, \dots, n+1$. 令

$$t = z_1 + \cdots + z_n + z_{n+1}，x_i \sim \frac{z_i}{t} = \frac{z_i}{z_1 + \cdots + z_{n+1}}, i = 1, \dots, n，$$

则$\mathbf{x} = (x_1, \dots, x_n)^\top$与$t$独立，且$\mathbf{x} \sim \text{Dirichlet}(d_1/2, \dots, d_n/2;\ d_{n+1}/2)$.

特别地，$n = 1$时，$\frac{\chi_n^2}{\chi_n^2 + \chi_m^2} \sim \text{Dirichelet}(n/2; m/2) = \text{Beta}(n/2, m/2)$
(两个卡方r.v.独立)

应用Dirichelet分布的Gamma表示（命题A1），容易得到：

推论A2. 假设$\mathbf{x} = (x_1, \dots, x_n)^\top \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_n;\ \alpha_{n+1})$, 则对任何
$1 \le k \le n$, $(x_1, \dots, x_k)^\top \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_k;\ \alpha_{k+1} + \cdots + \alpha_{n+1})$,
特别地，$x_i \sim \text{Dirichlet}(\alpha_i; \alpha - \alpha_i) = \text{Beta}(\alpha_i, \alpha - \alpha_i)$, 其中$\alpha = \sum_{j=1}^{n+1} \alpha_j$

我们可以利用命题A1 计算$U(S^{n-1})$的边际分布：

推论A3：假设 $x_1, \dots, x_n \; iid \sim N(0,1)$，则

$$\mathbf{v} = \left(\frac{x_1^2}{x_1^2+\cdots+x_n^2}, \dots, \frac{x_\mathrm{k}^2}{x_1^2+\cdots+x_n^2}\right)^\top \sim Dirichelet\left(\overbrace{\frac{1}{2}, \dots, \frac{1}{2}}^{k\text{个}}; \frac{n-k}{2}\right)$$

这实际上也是$\mathbf{v} = (u_1^2, \dots, u_k^2)^\top$的分布，这里$\mathbf{u} = (u_1, \dots, u_n)^\top \sim U(S^{n-1})$

基于推论A3，做变换$\mathbf{v} \to \mathbf{u}^{(k)} = \left(\frac{x_1}{\sqrt{x_1^2+\cdots+x_n^2}}, \dots, \frac{x_\mathrm{k}}{\sqrt{x_1^2+\cdots+x_n^2}}\right)^\top$,

即可解得$\mathbf{u} = (u_1, \dots, u_n)^\top \sim U(S^{n-1})$的$k$元边际。