hw2

2.1

假设随机向量 $(x, y) \in B^2$, 做极坐标变换

$$x = r\cos(\theta), y = r\sin(\theta), r > 0, 0 \le \theta < 2\pi,$$

假设 $\theta \sim U(0, 2\pi), r \sim \text{Beta}(1, \beta)(\beta > 0)$, 试求 (x, y) 的联合概率密度, 当 $\beta \to \infty$ 时, 该分布有什么特点?

Solution:

Jacobi 矩阵行列式为:

$$J = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix}^{-1} = \frac{1}{r}.$$

所以 (x, y) 的联合密度函数为

$$p(x, y) = \frac{1}{2\pi}\frac{1}{B(1, \beta)}(1 - r)^{\beta - 1}\frac{1}{r} = \frac{\beta(1 - r)^{\beta - 1}}{2\pi r}.$$

当 $\beta \to \infty$ 时，分布集中在 $(0, 0)$ 处 ■

2.2

假设三元随机向量 $(x, y, z)^\top$ 服从球对称分布, 其概率密度函数 $f(x, y, z) = h\left(\sqrt{x^2 + y^2 + z^2}\right)$, 令 $(x, y, z)^\top$ 的球坐标变换

$$x = r\cos(\theta), y = r\sin(\theta)\cos(\psi), z = r\sin(\theta)\sin(\psi), \quad r > 0, 0 \le \theta < \pi, 0 \le \psi < 2\pi.$$

(a) 证明 $\cos(\theta) \sim U(-1, 1), \psi \sim U(0, 2\pi)$ 。

(b) 如果已知 $r \sim U(0, 1)$, 试求 $(x, y, z)^\top$ 的概率密度。

Solution:

(a)Jacobi 矩阵行列式为:

$$J = \begin{vmatrix} \sin\psi\cos\theta & r\cos\psi\cos\theta & -r\sin\psi\sin\theta \\ \sin\psi\sin\theta & r\cos\psi\sin\theta & r\sin\psi\cos\theta \\ \cos\psi & -r\sin\psi & 0 \end{vmatrix} = r^2\sin\theta$$

所以 (r, θ, ψ) 的联合密度函数为

$$p(r, \theta, \psi) = r^2\sin\theta h(r) = (\frac{1}{2\pi})(\frac{1}{2}sin\theta)(4\pi r^2 h(r))$$

可知 r, θ, ψ 相互独立，密度函数为

$$p(r) = 4\pi r^2 h(r), p(\psi) = \frac{1}{2\pi}, p(\theta) = \frac{1}{2}\sin\theta$$

所以 $\cos(\theta) \sim U(-1, 1), \psi \sim U(0, 2\pi)$

(b) 由于 $r \sim U(0, 1)$,

$$p(r) = 1_{\{0 < r \le 1\}} = 4\pi r^2 h(r)$$

所以 $h(r) = \frac{1}{4\pi r^2}1_{\{0<r\le1\}}$

那么 $(x, y, z)^\top$ 的密度函数为

$$f(x, y, z) = \frac{1}{4\pi}\frac{1}{x^2 + y^2 + z^2}.$$

■

2.3

假设随机变量 $x \sim \mathrm{Beta}(1/2, n/2)$, 令 $r = \sqrt{(1-x)/x}$ 。假设 $\mathbf{u} \sim U\left(S^{n-1}\right)$ 且与 r 独立, 试验证 $\mathbf{x} = r\mathbf{u}$ 的密度函数为

$$p(\mathbf{x}) = \frac{\Gamma((n+1)/2)}{\pi^{(n+1)/2}} \frac{1}{\left(1+\|\mathbf{x}\|^2\right)^{(n+1)/2}}, \mathbf{x} \in R^n.$$

这是多元 Cauchy 分布。

Solution:

$$x = \frac{1}{1+r^2}$$

$$J = |\frac{\partial x}{\partial r}| = \frac{2r}{\left(1+r^2\right)^2}$$

由于 $r > 0$，r 的密度函数为

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

由于 $\mathbf{u} \sim U\left(S^{n-1}\right)$ 和第二讲定理 2，可知

$$p(\mathbf{x}) = \frac{\Gamma(n/2) p(\|\mathbf{x}\|)}{\pi^{n/2}\|\mathbf{x}\|^{n-1}} = \frac{\Gamma((n+1)/2)}{\pi^{(n+1)/2}} \frac{1}{\left(1+\|\mathbf{x}\|^2\right)^{(n+1)/2}}.$$

■

2.4

假设 $n \times 1$ 随机向量 $\mathbf{x}$ 服从球对称分布, 其概率密度函数仅与模长有关: $f(\mathbf{x}) = h(\|\mathbf{x}\|)$, 对于如下球坐标变换: $\mathbf{x} \rightarrow (r, \theta)$

$$x_1 = r\cos\left(\theta_1\right), x_2 = r\sin\left(\theta_1\right)\cos\left(\theta_2\right), \ldots, x_n = r\sin\left(\theta_1\right)\sin\left(\theta_2\right)\cdots\sin\left(\theta_{n-1}\right)$$

我们在第二讲定理 2 的证明过程中得到了 $\boldsymbol{\theta} = \left(\theta_1, \ldots, \theta_{n-1}\right)$ 的联合概率密度

$$q(\boldsymbol{\theta}) = \frac{\Gamma(n/2)}{2\pi^{n/2}} \sin^{n-2}\left(\theta_1\right)\cdots\sin\left(\theta_{n-2}\right), 0 \leq \theta_1, \ldots, \theta_{n-2} < \pi, 0 \leq \theta_{n-1} < 2\pi.$$

(a) 利用 $q(\boldsymbol{\theta})$ 分布, 求 θ_1 的边际概率密度。

(b) 记 $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\| = \left(u_1, \ldots, u_n\right)^{\top}$ 试求 $u_1 = \cos\left(\theta_1\right)$ 的边际概率密度。并说明只有 $n=3$ 时, $u_1 \sim U(-1,1)$.

(c) 对任何 $2 \leq i \leq n$, 求 u_i 的边际密度。

Solution:

(a) 由于 $q(\boldsymbol{\theta}) = \frac{\Gamma(n/2)}{2\pi^{n/2}} \sin^{n-2}\left(\theta_1\right)\cdots\sin\left(\theta_{n-2}\right)$

可知 $\theta_1, \ldots, \theta_{n-2}$ 相互独立

所以 θ_1 的边际密度函数为

$$p(\theta_1) = c\sin^{n-2}(\theta_1)$$

已知

$$\mathrm{B}(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \int_0^1 t^{p-1}(1-t)^{q-1}\,\mathrm{d}t.$$

令 $p = \frac{n-1}{2}$ 和 $q = \frac{1}{2}$，换元 $t = \frac{\cos^2\theta}{\sin^2\theta}$

$$\mathrm{B}(\frac{n-1}{2}, \frac{1}{2}) = 2\int_0^{\frac{\pi}{2}} \sin^{n-2}(\theta)\mathrm{d}\theta = \int_0^{\pi} \sin^{n-2}(\theta)\mathrm{d}\theta = \frac{\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})}{\Gamma(\frac{n}{2})}$$

所以

$$p(\theta_1) = \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})} \sin^{n-2}(\theta_1)$$

(b) 由 (a) 可得

$$p(u_1)=\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})}\sin^{n-2}(\arccos u_1)\frac{1}{\sin(\arccos u_1)}=\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})}\sin^{n-3}(\arccos u_1)=\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})}(1-u_1^2)^{\frac{n-3}{2}}$$

n=3 时，$p(u_1)=\frac{1}{2}$，$u_1\sim U(-1,1)$

(c) 因为球对称，所有边际分布相同

$$p(u_i)=\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})}(1-u_i^2)^{\frac{n-3}{2}}\quad 2\le i\le n$$

■

2.5

(a) 假设 $x/y \perp\!\!\!\perp y$, 证明 $E(x/y)=E(x)/E(y)$ 。

(b) 假设 $\mathbf{x}=(x_1,\ldots,x_n)^\top\sim \text{Dirichlete}\,(\alpha_1,\ldots,\alpha_n;\alpha_{n+1})$, 记 $a=\sum_{i=1}^{n+1}\alpha_i,\theta_i=\alpha_i/a,\boldsymbol{\theta}=(\theta_1,\ldots,\theta_n)^\top$. 试利用第三讲命题 5Dirichlete 分布的 gamma 表示证明

$$E(\mathbf{x})=\boldsymbol{\theta},\quad \text{var}(\mathbf{x})=\left[\text{diag}(\boldsymbol{\theta})-\boldsymbol{\theta}\boldsymbol{\theta}^\top\right]/(1+a).$$

Solution:

(a) 令 $u=x/y$，因为 $u \perp\!\!\!\perp y$ 和 $x=uy$.

$$E[x]=E[u]E[y]$$

所以 $E(x/y)=E(x)/E(y)$

(b) 由第三讲命题五, 我们知道存在独立随机变量 z_1,...,z_{n+1},$z_i\sim\Gamma(\alpha_i,1)$, 使得 $x_i=z_i/\sum_{j=1}^{n+1}z_j$, 且 $t=\sum_{i=1}^{n+1}z_i\sim\Gamma(a,1)$, $x_i=z_i/t$ 与 t 是独立的，由 (a) 可知

$$E(x_i)=E(z_i/t)=E(z_i)/E(t)=\alpha_i/\alpha=\theta_i\;E(x)=\theta$$

同理，知

$$\begin{aligned}\text{Var}[x]&=E[xx^\top]-E[x]E[x]^\top\\&=E\left[\frac{zz^\top}{t^2}\right]-\theta\theta^\top\\&=\left(\frac{E[zz^\top]}{E[t^2]}\right)-\theta\theta^\top\\&=\left(\frac{\text{diag}(\alpha)+\alpha\alpha^\top}{a^2+a}\right)-\theta\theta^\top\\&=\frac{\text{diag}\,\theta-\theta\theta^\top}{a+1}.\end{aligned}$$

另一种解法，直接按照定义积分：

$$\begin{aligned}\mathbb{E}[x_i]&=\int\int\frac{\Gamma(\alpha_1+\cdots+\alpha_{n+1})}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{n+1})}\left(\prod_{1\le j\le n,j\ne i}x_j^{\alpha_j-1}\right)x_i^{\alpha_i}\,\mathrm{d}\,x_1\cdots\mathrm{d}\,x_{n+1}\\&=\frac{\Gamma(\alpha_1+\cdots+\alpha_{n+1})}{\Gamma(\alpha_1+\cdots+\alpha_{n+1}+1)}\frac{\Gamma(\alpha_i+1)}{\Gamma(\alpha_i)}=\frac{\alpha_i}{a}=\theta_i,\quad 1\le i\le n.\end{aligned}$$

$$\begin{aligned}\mathbb{E}[x_i^2]&=\int\int\frac{\Gamma(\alpha_1+\cdots+\alpha_{n+1})}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{n+1})}\left(\prod_{1\le j\le n,j\ne i}x_j^{\alpha_j-1}\right)x_i^{\alpha_i+1}\,\mathrm{d}\,x_1\cdots\mathrm{d}\,x_{n+1}\\&=\frac{\Gamma(\alpha_1+\cdots+\alpha_{n+1})}{\Gamma(\alpha_1+\cdots+\alpha_{n+1}+2)}\frac{\Gamma(\alpha_i+2)}{\Gamma(\alpha_i)}=\frac{\alpha_i(\alpha_i+1)}{a(a+1)}.\quad 1\le i\le n.\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}[x_i x_k] &= \int\int \frac{\Gamma(\alpha_1+\cdots+\alpha_{n+1})}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{n+1})}\left(\prod_{1\le j\le n, j\ne i,k} x_j^{\alpha_j-1}\right) x_i^{\alpha_i} x_k^{\alpha_k}\,\mathrm{d}\,x_1\cdots\mathrm{d}\,x_{n+1} \\
&= \frac{\Gamma(\alpha_1+\cdots+\alpha_{n+1})}{\Gamma(\alpha_1+\cdots+\alpha_{n+1}+2)}\frac{\Gamma(\alpha_i+1)\Gamma(\alpha_k+1)}{\Gamma(\alpha_i)\Gamma(\alpha_k)} = \frac{\alpha_i\alpha_k}{a(a+1)}, \quad 1\le i,k\le n.
\end{aligned}
$$

因此有:

$$
var(x_i) = \mathbb{E}[x_i^2] - (\mathbb{E}[x_i])^2 = \frac{\alpha_i(\alpha_i+1)}{a(a+1)} - \frac{\alpha_i^2}{a^2} = \frac{\theta_i-\theta_i^2}{a+1}
$$

$$
cov(x_i,x_k) = \mathbb{E}[x_i x_k] - \mathbb{E}[x_i]\mathbb{E}[x_k] = \frac{\alpha_i\alpha_k}{a(a+1)} - \frac{\alpha_i\alpha_k}{a^2} = -\frac{\theta_i\theta_k}{a+1}
$$

即:

$$
E(\mathbf{x}) = \boldsymbol{\theta}, \quad var(\mathbf{x}) = [diag(\boldsymbol{\theta}) - \boldsymbol{\theta}\boldsymbol{\theta}^\top]/(1+a).
$$

■

2.6

假设 $\mathbf{u}=(u_1,\ldots,u_n)^\top \sim U(S^{n-1})$, 不妨假设 $\mathbf{u}=\mathbf{x}/\|\mathbf{x}\|$, 其中 $\mathbf{x}=(x_1,\ldots,x_n)^\top \sim N_n(\mathbf{0},I_n)$, 则 $u_1^2 = x_1^2/(x_1^2+x_2^2+\ldots+x_n^2)$, 试基于该表示（而不是直接应用第 1 讲定理 1 ）证明

(a) $u_1^2 \sim \text{Beta}\left(\frac{1}{2},\frac{n-1}{2}\right)$.

(b) u_1 的概率密度函数为

$$
p(u_1) = \frac{\Gamma(n/2)}{\sqrt{\pi}\Gamma((n-1)/2)} \times \left(1-u_1^2\right)^{(n-3)/2}, -1<u_1<1.
$$

Solution:

(a) $x\sim N(0,I_n)$, $x_1^2,\ldots,x_n^2$ iid $\sim \chi_1^2$

所以

$$
u_1^2 = \frac{x_1^2}{x_1^2+\cdots+x_n^2} \sim \text{Beta}(\frac{1}{2},\frac{n-1}{2}).
$$

(b) 同 2.4(b) ■

2.7

假设 $\mathbf{u}=(u_1,\ldots,u_n)^\top, \mathbf{v}=(v_1,\ldots,v_n)^\top$ iid $\sim U(S^{n-1})$, 记 $\mathbf{u},\mathbf{v}$ 之间的相似度（相关性）度量 $R=\mathbf{u}^\top\mathbf{v}$ 。

(a) 证明 R 与 $\mathbf{u}$ 独立, 也与 $\mathbf{v}$ 独立, 并且 $R \overset{d}{=} u_1$ 。

(b) 证明

$$
\sqrt{n-1}\frac{R}{\sqrt{1-R^2}} \sim t_{n-1}.
$$

(c) 假设 $\mathbf{x},\mathbf{y}$ iid $\sim N_n(\mathbf{0},\sigma^2 I_n)$, 令 Galton 样本相关系数 $R_G = \frac{\mathbf{x}^\top\mathbf{y}}{\|\mathbf{x}\|\cdot\|\mathbf{y}\|}$. 证明

$$
\sqrt{n-1}\frac{R_G}{\sqrt{1-R_G^2}} \sim t_{n-1}.
$$

(d) 条件同 (c), 令 Pearson 样本相关系数 $R_P = (\mathbf{x}-\mathbf{1}\bar{x})^\top(\mathbf{y}-\mathbf{1}\bar{y})/(\|\mathbf{x}-\mathbf{1}\bar{x}\|\cdot\|\mathbf{y}-\mathbf{1}\bar{y}\|)$. 证明

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

Solution:

(a) 参考第二讲定理 1

(b) 由（a）结论，有

$$\sqrt{n-1}\frac{R}{\sqrt{1-R^2}} \overset{d}{=} \sqrt{n-1}\frac{u_1}{\sqrt{u_2^2+\cdots+u_n^2}} \sim t_{n-1}.$$

(c) 令 $u = x/\|x\|, v = y/\|y\|$, 则 u, v iid $\sim U(S^{n-1}), \quad R_G = u^\top v$, 从而：

$$\sqrt{n-1}\frac{R_G}{\sqrt{1-R_G^2}} \overset{d}{=} \sqrt{n-1}\frac{R}{\sqrt{1-R^2}} \sim t_{n-1}.$$

(d) 令 H 为正交矩阵，其第一行为元素全为 $1/\sqrt{n}$, 则 $Hx = (\sqrt{n}\bar{x}, z^\top)^\top$,$Hy = (\sqrt{n}\bar{y}, w^\top)^\top$, 由于 H 是正交阵，有

$$z, w \quad \text{iid} \sim N_{n-1}(0, \sigma^2 I_{n-1})$$

注意到 $\|\mathbf{x}-\mathbf{1}\bar{x}\|^2 = \mathbf{x}^T\mathbf{x} - n\bar{x}^2 = \|\mathbf{z}\|^2, \|\mathbf{y}-\mathbf{1}\bar{y}\|^2 = \mathbf{y}^T\mathbf{y} - n\bar{y}^2 = \|\mathbf{w}\|^2$, 则:

$$R_P = (\mathbf{x}-\mathbf{1x})^\top(\mathbf{y}-\mathbf{1y})/(\|\mathbf{x}-\mathbf{1x}\| \cdot \|\mathbf{y}-\mathbf{1y}\|) = \frac{z^T w}{\|z\|\|w\|}$$

再结合 (c) 的结论，有

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

■