作业 1：常见概率分布

姓名（学号）

2020 年 02 月 12 日

说明

本次作业是写出**常用概率分布，要求不能少于 25 个分布。每个分布至少要给出分布的定义、概率函数或密度函数、数学期望和方差。**

每次作业，需提交 Rmarkdown 文档和所生成的 pdf 文档。

做作业之前，你可能需要做如下准备工作：

1. 安装 LaTex（建议安装完整版的 Texlive 或最新版的 MikTex）；
2. 安装 R 和 RStudio；
3. 正确配置 RStudio：
    - 打开 `Tools` => `global options`，然后
    - 点击 `sweave`，在 `weave rnw files using` 选择 **knitr**
    - 在 `Typeset LaTex into PDF using` 选择 **XeLaTex**
    - 在 `Code => Saving => Default text Encoding` 选择 `UTF-8`
4. 把中文 LaTex 模版文件 `template_article_zh.tex` 与作业文件放在同一个文件夹；
5. 安装需要的软件包（bookdown 等）。

更详细的说明可参看我的博客《如何用 R Markdown 写学术文档?》和《R Markdown 用法》。

本 Rmd 文档可作为作业模板和示例, 但提交的作业不要包含本说明。

1 离散型随机变量的分布

1.1 二项分布

在 n 次独立重复试验中，记 X 为事件 A 发生的次数，假设在每次试验中事件 A 发生的概率都是 p，则随机变量 X 的分布称为**二项分布**，记为 $X \sim \text{Binom}(n,p)$，其概率函数为

$$P(X=k|n,p)=\binom{n}{k}(1-p)^{n-k},(k=0,1,\ldots,n;0<p<1)$$

其中 $E(X)=np, D(X)=np(1-p)$。

2 连续型随机变量的分布

2.1 均匀分布

如果随机变量 X 的密度函数 (pdf) 为

$$X \sim f(x|a,b)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & \text{otherwise}\end{cases}$$

则 X 的分布称为 (a,b) 区间的均匀分布，记为 $X \sim \text{Unif}(a,b)$。

其数学期望和方差分别为

$$E(X)=\frac{a+b}{2}, D(X)=\frac{(b-a)^2}{12}.$$

3 常用多维分布

3.1 多项分布

假设进行 n 次独立重复试验，每次试验中有 k 个可能的结果，各个结果发生的概率分别为 $p_1, ..., p_k$ (其中 $p_i \ge 0, \sum_{i=1}^{k} p_i = 1$)。记 X_i 为第 i 个结果出现的次数，则称随机向量 $\boldsymbol{X} = (X_1, X_2, \ldots, X_k)^T$ 服从多项分布 (Multinomial distribution)，参数为 n 和 $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)^T$，记为 $\boldsymbol{X} \sim \mathrm{Multinom}(n, \boldsymbol{p})$，其联合概率函数 (pmf) 为

$$P(X_1 = x_1, \ldots, X_k = x_k) = \binom{n}{x_1 x_x \cdots x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}, \ (x_i \ge 0, \sum_{i=1}^{k} x_i = n).$$

多项分布 $\mathrm{Multinom}(n, \boldsymbol{p})$ 具有如下性质：

1. $\mathrm{E}(\boldsymbol{X}) = n\boldsymbol{p}$, 即 $\mathrm{E}(X_i) = np_i \, (i = 1, 2, \ldots, k)$
2. $\boldsymbol{X}$ 的协方差矩阵 Σ 是对称矩阵，其对角线元素为

   $$\sigma_i^2 = np_i(1 - p_i),$$

   非对角线元素为

   $$\mathrm{Cov}(X_i, X_j) = -np_i p_j \, (i \neq j).$$

3. $\boldsymbol{X}$ 的相关系数矩阵元素为

   $$\mathrm{Corr}(X_i, X_j) = \frac{-np_i p_j}{\sqrt{np_i(1 - p_i) \cdot np_j(1 - p_j)}} = -\sqrt{\frac{p_i p_j}{q_i q_j}}$$

4. 记 $\boldsymbol{X}_{(k-1)} = (X_1, X_2, \ldots, X_{k-1})^T$，则 $\boldsymbol{X}_{(k-1)} \sim \mathrm{Multinom}(n, \boldsymbol{p}_{(k-1)})$，其中

   $$\boldsymbol{p}_{(k-1)} = (p_1, \ldots, p_{k-2}, p_{k-1} + p_k)^T.$$

特别, $X_i \sim \mathrm{Bin}(n, p_i)$.