概率论与数理统计知识点整理

KujoStar-hmc

2022 年 1 月 5 日

前言

作者于 2022-2023 学年秋季学期上了唐宏岩老师的概率论与数理统计课程，一学期颇受折磨，但是唐老师教学水平高超、教学内容丰富，若是用心理解钻研，确实会学到不少东西。前几天刚刚考完期末考试，想着可以趁着得闲将一学期所学知识精简整理，希望可以对以后选修概率论与数理统计课程（不局限于唐老师）的同学有一点微薄的帮助。

除了本人笔记外，此文档也参考了 lambda 的听课笔记进行整理归纳，特此致谢。

一、常用的计算恒等式

指数函数的级数求和表示

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

与二项分布相关的常见积分

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2+a^2}{2}} \mathrm{d}x = \sqrt{2\pi} e^{-\frac{a^2}{2}}$$

$$\int_{-\infty}^{+\infty} e^{-\frac{(x+a)^2}{2\sigma^2}} \mathrm{d}x = \sqrt{2\pi}\sigma$$

与 Gamma 函数相关的计算式

$$\int_0^1 u^{a-1}(1-u)^{b-1} \mathrm{d}u = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} \mathrm{d}t$$

组合数相关恒等式

$$k\binom{n}{k} = n\binom{n-1}{k-1}$$

$$k^2\binom{n}{k}=n\binom{n-1}{k-1}+n(n-1)\binom{n-2}{k-2}$$
$$\sum_{k=0}^{n}\binom{n}{k}p^kq^{n-k}=(p+q)^n$$

另外，打好微积分和线性代数的基础、掌握这两门数学基础课中的一些思想对概统的学习非常重要。

二、事件的概率

1. 事件和与差的概率

$$P(A+B)=P(A)+P(B)-P(AB),P(A-B)=P(A)-P(AB)$$

对于事件和的概率推广到一般性，我们有

$$P\left(\sum_{i=1}^{n}A_i\right)=\sum_{r=1}^{n}(-1)^{r+1}\sum_{i_1<i_2<\cdots i_r}^{i_k\in\{1,\cdots,n\}}P\left(\prod_{k=1}^{r}A_{i_k}\right)$$

2. 条件概率

$$P(A|B)=\frac{P(AB)}{P(B)},P(AB)=P(B)P(A|B)$$
$$P(B|A)=\frac{P(B)P(A|B)}{P(A)}$$
$$P(A)=P(B)P(A|B)+P(\overline{B})P(A|\overline{B})$$

3.Bayes 公式

设 $B_i,i=1,2,\cdots,n$ 为全事件 Ω 的一个划分，则有求解条件概率的贝叶斯公式：

$$\forall i,P(B_i|A)=\frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n}P(B_j)P(A|B_j)}$$

三、随机变量及其分布

如无特殊说明，本节涉及的 $F(x)$ 均表示随机变量 X 的 cdf，$f(x)$ 表示随机变量 X 的 pmf/pdf，$E(X)$ 表示随机变量 X 的期望，$Var(X)$ 表示随机变量 X 的方差。

首先，我们有 $\sum_{i=1}^{\infty}f(x)=1$（离散）或 $\int_{\mathbb{R}}f(x)\mathrm{d}x=1$（连续）。这是由 pmf 和 pdf 的定义保证的。

1. 常见的离散分布

Bernouli 分布（两点分布）

若随机变量 X 满足

$$X = \begin{cases} 1, & p \\ 0, & 1-p \end{cases}, p \in (0,1)$$

则称 X 服从概率为 p 的 Bernouli 分布，记作

$$X \sim B(p)$$

我们有

$$E(X) = p, Var(X) = p(1-p)$$

二项分布

二项分布其实说明了 n 次独立的 Bernouli 实验的成功次数。记 X 为这一次数，我们有

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \cdots, n$$

则 X 服从二项分布，记作

$$X \sim B(n, p)$$

我们有

$$E(X) = np, Var(X) = np(1-p)$$

Poisson 分布

若随机变量 X 满足

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \lambda > 0, k \in \mathbb{N}$$

则称 X 服从参数为 λ 的 Poisson 分布，记作

$$X \sim P(\lambda)$$

我们有

$$E(X) = Var(X) = \lambda$$

Remark：二项分布可以用 Poisson 分布来近似。具体来说，若 $X \sim B(n,p)$，则可以近似认为 $X \sim P(\lambda)$，其中参数 $\lambda = np$。当 p 很小，n 很大，λ 不太大时，近似的效果很好，其误差上限为 $\min\{p, np^2\}$。

2. 常见的连续分布

均匀分布

若随机变量 X 的 pdf 为

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{otherwise} \end{cases}$$

则称 X 服从区间 $[a,b]$ 上的均匀分布，记作

$$X \sim U(a,b)$$

我们有

$$E(X) = \frac{a+b}{2}, Var(X) = \frac{1}{12}(a-b)^2$$

Remark：$X \sim U(0,1)$ 通常称为随机数。

正态分布

若随机变量 X 的 pdf 为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

则称 X 服从正态分布，记作

$$X \sim N(\mu, \sigma^2)$$

我们有

$$E(X) = \mu, Var(X) = \sigma^2$$

Remark1：$N(0,1)$ 称为标准正态分布。

Remark2：正态分布标准化：

$$X \sim N(\mu, \sigma^2) \Leftrightarrow Y = \frac{X-\mu}{\sigma^2} \sim N(0,1)$$

指数分布

若随机变量 X 的 pdf 为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \le 0 \end{cases}, \lambda > 0$$

则称 X 服从指数分布，记作

$$X \sim Exp(\lambda)$$

我们有

$$E(X) = \frac{1}{\lambda}, Var(X) = \frac{1}{\lambda^2}$$

Remark：此时随机变量 X 的 cdf 为

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

Beta 分布

若随机变量 X 的 pdf 为

$$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1},0\leq x\leq 1$$

其中

$$\Gamma(x)=\int_0^{+\infty}e^{-t}t^{x-1}\mathrm{d}t$$

则称 X 服从 Beta 分布，记作

$$X\sim \mathrm{Beta}(a,b)$$

我们有

$$E(X)=\frac{a}{a+b},Var(X)=\frac{ab}{(a+b)^2(a+b+1)}$$

Remark：Gamma 函数有如下性质：

$$\Gamma(x+1)=x\Gamma(x)$$

利用 Gamma 函数的定义以及分部积分法易证。

3. 随机变量的函数

设 X 为一个随机变量，$g(x)$ 为在 X 取值范围内有定义的函数，若随机变量 $Y=g(X)$，则称 Y 为随机变量 X 的函数。

pdf 公式推导

令 X 为连续型随机变量，其 pdf 为 $f(x)$，假设 $g(x)$ 为严格单调可微函数，则 $Y=g(X)$ 的 pdf 为

$$f_Y(y)=\begin{cases} f(h(y))h'(y), & \alpha\leq y\leq\beta \\ 0, & \text{otherwise}\end{cases}$$

其中

$$\alpha=\min\{g(-\infty),g(+\infty)\},\beta=\max\{g(-\infty),g(+\infty)\},h(x)=g^{-1}(x)$$

利用分布函数的定义易证。

典例

对数正态分布的概率密度函数。设随机变量 $Y>0$，$X=\ln Y\sim N(\mu,\sigma^2)$，则 Y 的 pdf 为

$$f(y)=\begin{cases} \frac{1}{\sqrt{2\pi}\sigma y}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}, & y>0 \\ 0, & y\leq 0\end{cases}$$

四、联合分布

本节研究的对象为随机向量：

$$(X_1, \cdots, X_n), \Omega \to \mathbb{R}^n$$

随机向量的每个分量都是同一个 Ω 上的随机变量。大部分时候主要研究二元分布，此时的随机向量记为 (X, Y)，联合 cdf 记为 $F(x, y)$，联合 pdf 记为 $f(x, y)$。

设 n 元分布的 pdf 为 $f(x_1, x_2, \cdots, x_n)$，由概率的定义，我们首先有

$$\sum_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) = 1$$

或者

$$\int_{\mathbb{R}^n} f(x_1, x_2, \cdots, x_n) \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n = 1$$

1. 常见的联合分布

多项分布

设某个随机试验有 n 个互斥的可能结果：$A_1, A_2, \cdots, A_n$，$\{A_i\}$ 是全事件 Ω 的一个划分，A_i 发生的概率是 p_i，A_i 出现的次数为 X_i。设 $\sum_{i=1}^n k_i = N$，N 为总实验次数，则我们有

$$P(X_1 = k_1, \cdots, X_n = k_n) = \binom{N}{k_1, \cdots, k_n} \prod_{i=1}^n p_i^{k_i}$$

其中

$$\binom{N}{k_1, \cdots, k_n} = \frac{N!}{\prod_{i=1}^n k_i!}$$

二元均匀分布

pdf 为

$$f(x, y) = \begin{cases} \frac{1}{(b-a)(d-c)}, & (x, y) \in (a, b) \times (c, d) \\ 0, & \text{otherwise} \end{cases}$$

Remark：二元分布又被称为矩形分布。

二元正态分布

$(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，其中 $|\rho| < 1$，则 (X, Y) 服从二元正态分布，其联合 pdf 为

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

2. 边际分布

边际分布的 cdf

我们考虑随机向量 $(X_1, X_2, \cdots, X_n)$，则我们定义 X_i 的边际 cdf:

$$F_i(x) := P(X_i \le x) = P(X_i \le x, \forall j \ne i, X_j \in \mathbb{R})$$

一般研究二元分布 (X, Y)，设联合 cdf 为 $F(x, y)$，则我们有 X 的边际 cdf 为

$$\begin{aligned} F_X(x) :=& P(X \le x) \\ =& P(X \le x, -\infty < y < +\infty) \\ =& \lim_{y \to +\infty} P(X \le x, -\infty < Y \le y) \\ =& \lim_{y \to +\infty} F(x, y) \end{aligned}$$

Y 的边际 cdf 可以类似得到。

边际分布的 pdf

边际分布描述了随即向量每个分量单独的分布情况。此处以连续型二元分布 (X, Y) 为研究的对象。设 (X, Y) 的联合 pdf 为 $f(x, y)$，由上面对边际 cdf 的讨论，我们有

$$F_X(x) = \lim_{y \to +\infty} F(x, y) = \int_{-\infty}^{x} \int_{\mathbb{R}} f(t, y) \mathrm{d}y \mathrm{d}t$$

我们熟知，随机变量 X 的 cdf$F(x)$ 与 pdf$f(x)$ 之间有如下关系：

$$F'(x) = f(x)$$

则对上面得到的式子对 x 求导，我们有

$$f_X(x) = \int_{\mathbb{R}} f(x, y) \mathrm{d}y$$

这就是 X 的边际 pdf。同理我们得到 Y 的边际 pdf 为

$$f_Y(y) = \int_{\mathbb{R}} f(x, y) \mathrm{d}x$$

典例

我们设随机向量 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则我们有如下结论：

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, x \in \mathbb{R}$$

也即

$$X \sim N(\mu_1, \sigma_1^2)$$

同理，我们可以得到

$$Y \sim N(\mu_2, \sigma_2^2)$$

结论的推导对二元正态分布的联合 pdf 直接单侧积分即可，读者可以检验自己的微积分水平（逃）。
Remark：联合分布可以确定边际分布，但是反之未必成立。

3. 条件分布

条件分布描述了在一些分量的值已经确定的情况下，随机向量中另外的分量的分布情况。绝大部分时候，我们仅仅以二元随机向量 (X,Y) 为研究对象。

离散型随机向量

我们设

$$P(X=a_i,Y=b_j)=p_{ij},\sum_{i,j}p_{ij}=1$$

则我们有

$$P(X=a_i|Y=b_j)=\frac{P(X=a_i,Y=b_j)}{P(Y=b_j)}=\frac{p_{ij}}{\sum_k p_{kj}}$$

Remark：

$$\sum_i P(X=a_i|Y=b_j)=1$$

连续型随机向量

我们设 (X,Y) 的 pdf 为 $f(x,y)$，则我们有

$$\begin{aligned}&P(X\le x|y\le Y\le y+\mathrm{d}y)\\=&\frac{P(X\le x,y\le Y\le y+\mathrm{d}y)}{P(y\le Y\le y+\mathrm{d}y)}\\=&\frac{\int_{-\infty}^{x}\left(\int_{y}^{y+\mathrm{d}y}f(s,t)\mathrm{d}t\right)\mathrm{d}s}{\int_{y}^{y+\mathrm{d}y}f_Y(s)\mathrm{d}s}\end{aligned}$$

对 x 求导可以得到 X 此时的条件概率密度为

$$f_X(x|y\le Y\le y+\mathrm{d}y)=\frac{\int_{y}^{y+\mathrm{d}y}f(x,t)\mathrm{d}t}{\int_{y}^{y+\mathrm{d}y}f_Y(s)\mathrm{d}s}$$

令 $\mathrm{d}y\to 0$，我们可以得到

$$f_X(x|Y=y)=\frac{f(x,y)}{f_Y(y)}$$

于是可以定义 X 的条件密度函数：

$$f_X(x|y):=\frac{f(x,y)}{f_Y(y)}$$

类似地，我们有

$$f_Y(y|x):=\frac{f(x,y)}{f_X(x)}$$

典例

设 $(X,Y) \sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则根据条件分布的定义，简单计算可得

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

也即当 $X=x$ 时，我们得到 Y 的分布：

$$Y \sim N\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),(1-\rho^2)\sigma_2^2\right)$$

4. 条件分布下的全概率公式

X 离散，Y 离散

$$P(X=x_i) = \sum_j P(X=x_i|Y=y_j)P(Y=y_j)$$

X 离散，Y 连续

$$P(X=x_i) = \int_{\mathbb{R}} P(X=x_i|Y=y)f_Y(y)\mathrm{d}y$$

X 连续，Y 离散

$$f_X(x) = \sum_i f_X(x|y_i)P(Y=y_i)$$

X 连续，Y 连续

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\mathrm{d}y = \int_{\mathbb{R}} f_X(x|y)f_Y(y)\mathrm{d}y$$

5. 条件分布下的 Bayes 公式

X 离散，Y 离散

$$P(Y=y_i|X=x_i) = \frac{P(X=x_i|Y=y_i)P(Y=y_i)}{P(X=x_i)} = \frac{P(X=x_i|Y=y_i)P(Y=y_i)}{\sum_j P(X=x_i|Y=y_j)P(Y=y_j)}$$

X 离散，Y 连续

$$P(Y=y|X=x_i) = \frac{P(X=x_i|Y=y)f_Y(y)}{P(X=x_i)} = \frac{P(X=x_i|Y=y)f_Y(y)}{\int_{\mathbb{R}} P(X=x_i|Y=y)f_Y(y)\mathrm{d}y}$$

X 连续，Y 离散

$$P(Y=y_i|X=x)=\frac{f_X(x|y_i)P(Y=y_i)}{f_X(x)}=\frac{f_X(x|y_i)P(Y=y_i)}{\sum_i f_X(x|y_i)P(Y=y_i)}$$

X 连续，Y 连续

$$f_Y(y|x)=\frac{f_X(x|y)f_Y(y)}{f_X(x)}=\frac{f_X(x|y)f_Y(y)}{\int_{\mathbb{R}} f_X(x|y)f_Y(y)\mathrm{d}y}$$

6. 随机向量分量之间的独立性

设随机向量 (X,Y) 的 cdf 为 $F(x,y)$，若 $F(x,y)=F_X(x)F_Y(y)$，则称 X,Y 独立。
Remark：X,Y 独立 $\Leftrightarrow f(x,y)=f_X(x)f_Y(y),\forall x,y\in\mathbb{R}$

五、多个随机变量的函数

设有 n 个随机变量 $X_i,i=1,2,\cdots,n$，函数 $g(x_1,x_2,\cdots,x_n)$ 为 n 元函数，设有随机变量 Y，满足 $Y=g(X_1,X_2,\cdots,X_n)$，则称 Y 为随机变量 $X_1,X_2,\cdots,X_n$ 的函数。本节主要研究 Y 的相关性质。

1. 随机变量的和的分布

相互独立的二项分布的和

设 $X_i\sim B(n_i,p)$，$i=1,2$，X_1,X_2 独立，设 $Y=X_1+X_2$，则有结论：

$$Y\sim B(n_1+n_2,p)$$

Remark：利用二项分布的概率表达式，对其中一个变量进行求和即可。也即利用

$$P(Y=k)=P(X_1+X_2=k)=\sum_{k_1=0}^{k}P(X_1=k_1)P(X_2=k-k_1)$$

相互独立的 Poisson 分布的和

设 $X_i\sim P(\lambda_i)$，$i=1,2$，X_1,X_2 独立，设 $Y=X_1+X_2$，则有结论：

$$Y\sim P(\lambda_1+\lambda_2)$$

二元正态分布分量的和

设 $(X_1,X_2)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则有

$$X_1+X_2\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)$$

2. 随机变量的商的分布

设 (X_1,X_2) 有联合 pdf$f(x_1,x_2)$，设 $Y=\dfrac{X_2}{X_1}$，其中 $X_1>0$，研究 Y 的分布。

根据定义，对 $\forall y$，我们有

$$\begin{aligned}
&P(Y\leq y)\\
=&P\left(\frac{X_2}{X_1}\leq y\right)\\
=&P(X_2\leq yX_1)\\
=&\iint_{x_2\leq yx_1}f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2\\
=&\int_0^{+\infty}\left(\int_{-\infty}^{yx_1}f(x_1,x_2)\mathrm{d}x_2\right)\mathrm{d}x_1\\
\xlongequal{x_2=tx_1}&\int_{-\infty}^{y}\left(\int_0^{+\infty}x_1f(x_1,x_1t)\mathrm{d}x_1\right)\mathrm{d}t
\end{aligned}$$

对 y 求导，得到 Y 的 pdf 为

$$l(y)=\int_0^{+\infty}x_1f(x_1,x_1y)\mathrm{d}x_1$$

3. 密度函数变换法

设随机向量 (X_1,X_2) 有联合 pdf$f(x_1,x_2)$，设 $(Y_1,Y_2)=(g_1(X_1,X_2),g_2(X_1,X_2))$，且该变换可逆可微，也即 $\exists h_1,h_2$，使得 $(X_1,X_2)=(h_1(Y_1,Y_2),h_2(Y_1,Y_2))$，则 (Y_1,Y_2) 的联合 pdf 为

$$f(h_1(y_1,y_2),h_2(y_1,y_2))\left|\det J\right|$$

其中矩阵 J 为变换的 Jacobi 矩阵，也即

$$J=\frac{\partial(h_1,h_2)}{\partial(y_1,y_2)}=\begin{bmatrix}\frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2}\\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2}\end{bmatrix}$$

典例

设 (X_1,X_2) 的 pdf 为 $f(x_1,x_2)$，利用密度函数变换法求 $Y=X_1+X_2$ 的 pdf。

我们令

$$Y=X_1+X_2,Z=X_1$$

则有逆变换为

$$X_1=Z,X_2=Y-Z$$

对应的 $\det J$ 为

$$\begin{vmatrix}0 & 1\\ 1 & -1\end{vmatrix}=-1$$

则 Y,Z 的联合 pdf 为

$$l(y,z)=f(z,y-z)|-1|=f(z,y-z)$$

Y 的 pdf 为

$$l_Y(y)=\int_{\mathbb{R}} l(y,z)\mathrm{d}z=\int_{\mathbb{R}} f(z,y-z)\mathrm{d}z$$

Remark：若 X_1,X_2 独立，则有

$$l_Y(y)=\int_{\mathbb{R}} f_{X_1}(z)f_{X_2}(y-z)\mathrm{d}z$$

这是卷积和积分变换的一种直观解释。

4. 次序统计量

设 $X_i(i=1,2,\cdots,n)$ 为来自同一总体的样本，总体 cdf 为 $F(x)$，总体 pdf 为 $f(x)$，$X_{(k)}$ 表示从小到大排列后次序为 k 的随机变量，例如 $X_{(1)}$ 表示 $\min\{X_1,X_2,\cdots,X_n\}$，$X_{(n)}$ 表示 $\max\{X_1,X_2,\cdots,X_n\}$。

单个次序统计量的分布

次序统计量 $X_{(k)}$ 的 pdf 为

$$f_k(x)=\frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k}f(x)$$

两个次序统计量的联合分布

任意两个次序统计量 $X_{(i)},X_{(j)}(i<j)$ 的联合 pdf 为

$$f_{i,j}(x,y)=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x)]^{i-1}[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j}f(x)f(y),x<y$$

以上结论均采用 pdf 的极限形式进行证明，可以参考作者的知乎专栏：前往奇妙世界