2022 春信息论 B 数学基础复习

复习内容：概率论基础　　课程助教：高源

1. 全概率公式
$$P(A)=P\left(B_1\right) P\left(A \mid B_1\right)+P\left(B_2\right) P\left(A \mid B_2\right)+\cdots$$
其中 $\{B_i\}$ 构成了完备事件群。

2. 贝叶斯公式
$$\begin{aligned} P(B_i \mid A) &= P\left(A B_i\right) / P(A) \\ &= P\left(B_i\right) P\left(A \mid B_i\right) / \sum_j P\left(B_j\right) P\left(A \mid B_j\right) \end{aligned}$$

3. 一维正态分布的概率密度函数
$$f(x)=(\sqrt{2 \pi} \sigma)^{-1} \mathrm{e}^{-(x-\mu)^2 / 2 \sigma^2},-\infty<x<\infty$$
记作 $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$。注意到，为了确定一个正态分布只需要两个参数，即均值和方差。除此以外，正态分布的概率密度函数验证中涉及的微积分技巧也建议掌握。首先作变量代换 $t=(x-\mu)/\sigma$，将问题转化为证明
$$I=\int_{-\infty}^{\infty} \mathrm{e}^{-t^2 / 2} \mathrm{~d} t=\sqrt{2 \pi}$$
进一步考虑（为了作极坐标代换）
$$I^2=\int_{-\infty}^{\infty} \mathrm{e}^{-t^2 / 2} \mathrm{~d} t \int_{-\infty}^{\infty} \mathrm{e}^{-u^2 / 2} \mathrm{~d} u=\int_{-\infty}^{\infty} \mathrm{e}^{-\left(t^2+u^2\right) / 2} \mathrm{~d} t \mathrm{~d} u$$
转换极坐标 $t=r \cos \theta, u=r \sin \theta$，则上述积分式可转化为
$$I^2=\int_0^{2 \pi} \mathrm{d} \theta \int_{-\infty}^{\infty} \mathrm{e}^{-r^2 / 2} r \mathrm{~d} r=2 \pi$$

4. 指数分布的概率密度函数
$$f(x)= \begin{cases}\lambda \mathrm{e}^{-\lambda x}, & \text{当} x>0 \\ 0, & \text{当} x \leqslant 0\end{cases}$$

5. 二维正态的概率密度函数

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

记作 $(X,Y)\sim\mathcal{N}\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho\right)$。其中 ρ 为相关系数，$\rho=0$ 时 X 和 Y 独立。

6. 多元正态分布的概率密度函数

$$f_{\mathbf{x}}\left(x_1,\ldots,x_k\right)=\frac{1}{\sqrt{(2\pi)^k|\Sigma|}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

7. 边缘分布。设 $X=\left(X_1,\cdots,X_n\right)$ 有概率密度函数 $f\left(x_1,\cdots,x_n\right)$，其中分量 X_1 的概率密度函数为（其余分量类似）

$$f_1\left(x_1\right)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f\left(x_1,x_2,\cdots,x_n\right)\mathrm{d}x_2\cdots\mathrm{d}x_n$$

8. 二维正态的边缘分布是正态分布。若 $\left(X_1,X_2\right)$ 有二维正态分布 $N\left(a,b,\sigma_1^2,\sigma_2^2,\rho\right)$，$X_1$，$X_2$ 的边缘分布分别是一维正态分布 $N\left(a,\sigma_1^2\right)$ 和 $N\left(b,\sigma_2^2\right)$.

9. 条件概率密度函数。设二维随机向量 $X=\left(X_1,X_2\right)$ 有概率密度函数 $f\left(x_1,x_2\right)$，有

$$f\left(x_1,x_2\right)=f_1\left(x_1\right)f_2\left(x_2\mid x_1\right)$$

10. 二维正态的条件概率分布是正态分布。设 $\left(X_1,X_2\right)$ 服从二维正态分布 $\mathcal{N}\left(a,b,\sigma_1^2,\sigma_2^2,\rho\right)$，有 $X_2\mid X_1\sim\mathcal{N}\left(b+\rho\sigma_2\sigma_1^{-1}\left(x_1-a\right),\sigma_2^2\left(1-\rho^2\right)\right)$。

11. 连续随机变量的独立性。若连续型随机向量 $\left(X_1,\cdots,X_n\right)$ 的概率密度函数 $f\left(x_1,\cdots,x_n\right)$ 可表为 n 个函数 $g_1,\cdots,g_n$ 之积，其中 g_i 只依赖于 x_i，即

$$f\left(x_1,\cdots,x_n\right)=g_1\left(x_1\right)\cdots g_n\left(x_n\right)$$

则 $X_1,\cdots,X_n$ 相互独立，且 X_i 的边缘密度函数 $f_i\left(x_i\right)$ 与 $g_i\left(x_i\right)$ 只相差一个常数因子。

12. 离散随机变量的独立性。设 $X_1,\cdots,X_n$ 都是离散型随机变量，若对任何常数 $a_1,\cdots,a_n$，都有

$$P\left(X_1=a_1,\cdots,X_n=a_n\right)=P\left(X_1=a_1\right)\cdots P\left(X_n=a_n\right)$$

则称 $X_1,\cdots,X_n$ 相互独立。

13. 离散随机变量的函数的概率分布求解。首先明确离散随机变量的函数仍为离散随机变量，依次确定其字母表和对应的概率值即可。

14. 连续随机变量的函数的概率密度求解。一般来讲有两种常用方法：

    - 根据定义求解概率分布函数，然后对其求导。例如考虑 $Y=X^2$，当 $y\leqslant 0$ 时 $P(Y\leqslant y)=0$，求导可得 $l(y)=0$。对于 $y>0$ 有

      $$P(Y\leqslant y)=P\left(X^2\leqslant y\right)=P(-\sqrt{y}\leqslant X\leqslant\sqrt{y})=\int_{-\sqrt{y}}^{\sqrt{y}}f(t)\mathrm{d}t$$

      对其求导可得

      $$l(y)=\frac{1}{2}y^{-1/2}[f(\sqrt{y})+f(-\sqrt{y})]$$

    - 构造一一对应的变量代换，基于多重积分变量代换公式求解得到变换后的概率密度函数，再进行边缘化得到所需的概率密度函数。以两个为例。设 (X_1,X_2) 的密度函数为 $f\left(x_1,x_2\right)$，为了求解 $Y_1=g_1\left(X_1,X_2\right)$ 的概率密度函数，构造 $Y_2=g_2\left(X_1,X_2\right)$，使得 (X_1,X_2) 到 (Y_1,Y_2) 存在一一对应变换，即要满足雅可比行列式不为 0。于是有 $X_1=h_1\left(Y_1,Y_2\right),X_2=h_2\left(Y_1,Y_2\right)$，和雅可比行列式

      $$J\left(y_1,y_2\right)=\left|\begin{array}{ll}\partial h_1/\partial y_1 & \partial h_1/\partial y_2\\ \partial h_2/\partial y_1 & \partial h_2/\partial y_2\end{array}\right|$$

      进而有

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

      这样我们就得到了随机向量的概率密度函数。为了我们的目标，求解我们的目标变量 Y_1 的边缘概率密度函数即可。

15. 随机变量和的概率密度函数公式（卷积公式）。设 (X_1,X_2) 的联合密度函数为 $f\left(x_1,x_2\right)$，则 $Y=X_1+X_2$ 的概率密度函数为

    $$l(y)=\int_{-\infty}^{\infty}f\left(x,y-x\right)\mathrm{d}x$$

16. 正态分布变量和仍为正态分布。$X_1,X_2\sim\mathcal{N}\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho\right)$，则

    $$Y=X_1+X_2\sim\mathcal{N}\left(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2\right)$$

17. 若干个随机变量之和的期望等于各变量的期望之和, 即

    $$E\left(X_1+X_2+\cdots+X_n\right)=E\left(X_1\right)+E\left(X_2\right)+\cdots+E\left(X_n\right)$$

注意，这里并不要求随机变量之间独立（和概率加法公式区别）。

18. 若干个独立随机变量之积的期望等于各变量的期望之积

$$E\left(X_1 X_2 \cdots X_n\right)=E\left(X_1\right) E\left(X_2\right) \cdots E\left(X_n\right)$$

19. 随机变量函数的期望。设随机变量 X 为离散型，有分布 $P\left(X=a_i\right)=p_i, i=1,2, \cdots$，或者为连续型，有概率密度函数 $f(x)$。则

$$E(g(X))=\sum_i g\left(a_i\right) p_i\left(\text{ 当} \sum_i\left|g\left(a_i\right)\right| p_i<\infty \text{ 时}\right)$$

或

$$E(g(X))=\int_{-\infty}^{\infty} g(x) f(x) \mathrm{d} x\left(\text{ 当} \int_{-\infty}^{\infty}|g(x)| f(x) \mathrm{d} x<\infty \text{ 时}\right)$$

这个结论的意义在于：为了计算 X 的某一函数 $g(X)$ 的期望，并不需要先算出 $g(X)$ 的密度函数，而可以就从 X 的分布出发，这大大方便了计算。因为在 g 较为复杂时 $g(X)$ 的密度很难求。

20. 条件期望。

$$E(Y \mid x)=\int_{-\infty}^{\infty} y f(y \mid x) \mathrm{d} y$$

21. 全期望公式。

$$E(Y)=E[E(Y \mid X)]$$

实际求解的时候，首先假定 X 为固定值 x，求解 $E(Y \mid x)$，再将 x 换成 X 即可。

22. 独立随机变量之和的方差等于各变量的方差之和

$$\operatorname{Var}\left(X_1+\cdots+X_n\right)=\operatorname{Var}\left(X_1\right)+\cdots+\operatorname{Var}\left(X_n\right)$$

23. 指数分布的均值为 λ^{-1}，方差为 λ^{-2}。

24. 协方差。称 $E\left[\left(X-m_1\right)\left(Y-m_2\right)\right]$ 为 X, Y 的协方差，并记为 $\operatorname{Cov}(X, Y)$。若 X, Y 独立, 则 $\operatorname{Cov}(X, Y)=0$。$[\operatorname{Cov}(X, Y)]^2 \leqslant \sigma_1^2 \sigma_2^2$. 等号当且仅当 X, Y 之间有严格线性关系时成立。

25. 相关系数。称 $\operatorname{Cov}(X, Y) /\left(\sigma_1 \sigma_2\right)$ 为 X, Y 的相关系数, 并记为 $\operatorname{Corr}(X, Y)$。注意，由独立可以得到相关系数为 0（不相关），但是相关系数为 0 并不一定独立。二维正态的参数 ρ 即为随机向量两个分量之间的相关系数。