第五次作业参考解答

1. (1) 随机向量 (X, Y) 的分布表如下：

$Y \backslash X$	0	1	2	3
0	$\frac{1}{22}$	$\frac{3}{22}$	$\frac{3}{44}$	$\frac{1}{220}$
1	$\frac{2}{11}$	$\frac{3}{11}$	$\frac{3}{55}$	0
2	$\frac{3}{22}$	$\frac{9}{110}$	0	0
3	$\frac{1}{55}$	0	0	0

(2) 用边际分布的视角：

$$\begin{aligned} P(X=1) &= \sum_{y=0}^{2} P(X=1, Y=y) \\ &= \frac{3}{22} + \frac{3}{11} + \frac{9}{110} \\ &= \frac{27}{55} \end{aligned}$$

也可以直接求：

$$P(X=1) = \frac{\binom{3}{1}\binom{9}{2}}{\binom{12}{3}} = \frac{27}{55}$$

2. 用联合累积分布函数的定义直接验证即可。

证明：

$$\begin{aligned} P(a < X \le b, c < Y \le d) &= P(X \le b, c < Y \le d) - P(X \le a, c < Y \le d) \\ &= [P(X \le b, Y \le d) - P(X \le b, Y \le c)] - [P(X \le a, Y \le d) - P(X \le a, Y \le c)] \\ &= F(b,d) - F(b,c) - F(a,d) + F(a,c) \end{aligned}$$

Tips: 本题可以用图形来辅助说明，但不建议只通过画图证明。

3. (1) (X, Y) 的概率密度函数

$$f(x,y) = \begin{cases} \frac{1}{\pi}, & x^2+y^2 \le 1 \\ 0, & x^2+y^2 > 1 \end{cases}$$

(2) 当 $-1 \le x \le 1$ 时，

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\mathrm{d}y = \frac{2\sqrt{1-x^2}}{\pi}$$

其他情况下 $f_X(x) = 0$.

$f_Y(y)$ 同理可求，

$$f_Y(y) = \begin{cases} \frac{2\sqrt{1-y^2}}{\pi}, & -1 \le y \le 1 \\ 0, & \text{其他情况} \end{cases}$$

(3) 考虑对应区域的积分：

$$P(R \le r) = \iint_{\sqrt{x^2+y^2} \le r} \frac{1}{\pi}\mathrm{d}x\mathrm{d}y = r^2, 0 < r < 1.$$

也可以考虑密度函数变换法，使用极坐标 (R,θ) 表示点的坐标。得到 (R,θ) 的概率密度函数

$$g(r,\theta)=\begin{cases}\dfrac{r}{\pi}, & 0\le r\le 1, 0\le\theta\le 2\pi\\ 0, & \text{其他情况}\end{cases}$$

进一步求出 $P(R\le r)=\int_0^r\int_0^{2\pi}\frac{r_0}{\pi}\mathrm{d}\theta\mathrm{d}r_0=r^2$

(4) R 的概率密度函数 $g(r)=2r$. 所求期望 $E(R)=\int_0^1 2r^2\mathrm{d}r=\frac{2}{3}$.

Tips: 注意计算期望不要对 $P(R\le r)$ 积分。

4. 按照课程讲义的定义和方式直接计算，关于多元正态分布更一般的定义和计算方法，见第三次习题课讲义。

二元正态分布的边际密度:

$$\begin{aligned}
f_X(x)&=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y\\
&=\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}[(\frac{x-\mu_1}{\sigma_1})^2+(\frac{y-\mu_2}{\sigma_2})^2-2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{y-\mu_2}{\sigma_2})]\right\}\mathrm{d}y\\
&=\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}[(\frac{x-\mu_1}{\sigma_1})^2+(\frac{y-\mu_2}{\sigma_2})^2-2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{y-\mu_2}{\sigma_2})]\right\}\mathrm{d}(\frac{y-\mu_2}{\sigma_2})\\
&=\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\exp\left\{Ay^2+By+C\right\}\mathrm{d}y\\
&=\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\exp\left\{A(y+\frac{B}{2A})^2-\frac{B^2}{4A}+C\right\}\mathrm{d}y\\
&=\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}e^{-\frac{B^2}{4A}+C}\int_{-\infty}^{\infty}\exp\left\{A(y+\frac{B}{2A})^2\right\}\mathrm{d}y\\
&=\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}e^{-\frac{B^2}{4A}+C}\int_{-\infty}^{\infty}e^{Ay^2}\mathrm{d}y\\
&=\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}e^{-\frac{B^2}{4A}+C}\sqrt{\frac{\pi}{-A}}
\end{aligned}$$

其中 $A=-\frac{1}{2(1-\rho^2)},B=\frac{\rho}{1-\rho^2}(\frac{x-\mu_1}{\sigma_1}),C=-\frac{1}{2(1-\rho^2)}(\frac{x-\mu_1}{\sigma_1})^2$. 因此

$$-\frac{B^2}{4A}+C=-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2$$

代入可得

$$\begin{aligned}
f_X(x)&=\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}e^{-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2}\sqrt{\pi 2(1-\rho^2)}\\
&=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}
\end{aligned}$$

5. 二元正态分布的条件密度：

$$
\begin{aligned}
f_{Y|X}(y|x) &= \frac{f(x,y)}{f_X(x)} \\
&= \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}[(\frac{x-\mu_1}{\sigma_1})^2+(\frac{y-\mu_2}{\sigma_2})^2-2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{y-\mu_2}{\sigma_2})]\right\}}{\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}} \\
&= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}[(\frac{x-\mu_1}{\sigma_1})^2+(\frac{y-\mu_2}{\sigma_2})^2-2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{y-\mu_2}{\sigma_2})]+\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}
\end{aligned}
$$

我们考察 exp 指数上关于 y 的二次表达式，y^2 的系数为

$$\frac{-1}{2(1-\rho^2)\sigma_2^2}$$

y 的系数为

$$-\frac{1}{2(1-\rho^2)}[-\frac{2\mu_2}{\sigma_2^2}-\frac{2\rho}{\sigma_2}(\frac{x-\mu_1}{\sigma_1})]$$

剩余的常数项为

$$-\frac{1}{2(1-\rho^2)}[(\frac{x-\mu_1}{\sigma_1})^2+\frac{\mu_2^2}{\sigma_2^2}+\frac{2\rho\mu_2}{\sigma_2}(\frac{x-\mu_1}{\sigma_1})]+\frac{(x-\mu_1)^2}{2\sigma_1^2}=-\frac{1}{2(1-\rho^2)}[(\frac{x-\mu_1}{\sigma_1})^2\rho^2+\frac{\mu_2^2}{\sigma_2^2}+\frac{2\rho\mu_2}{\sigma_2}(\frac{x-\mu_1}{\sigma_1})]$$

配方可得

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

Tips: 第四题和第五题应展示必要的推导和计算过程，不应当略过太多步骤直接展示结果。

6. (1) 所取的点在三角形区域中均匀分布

$$f(x,y)=\begin{cases}2, & x>0,y>0,\text{且}x+y<1\\ 0, & \text{其他情况}\end{cases}$$

Tips: 对于没有明确说累积分布函数还是概率密度函数的情况，一般可以认为二者皆可，以表示形式简单方便为宜。

(2) 当 $0\le y\le 1$ 时

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)\mathrm{d}x \\
&= \int_0^{1-y} 2\mathrm{d}x \\
&= 2(1-y)
\end{aligned}
$$

其他情况下 $f_Y(y)=0$.

(3) 当 $0\le y\le 1, 0\le x\le 1-y$ 时

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\
&= \frac{2}{2(1-y)} = \frac{1}{1-y}
\end{aligned}
$$

其他情况下 $f_{X|Y}(x|y)=0$.

7. (1)

$$\begin{aligned}
P(X_1=k|X_1+X_2=n) &= \frac{P(X_1=k,X_2=n-k)}{P(X_1+X_2=n)} \\
&= \frac{\frac{\lambda_1^k}{k!}e^{-\lambda_1}\frac{\lambda_2^{n-k}}{(n-k)!}e^{-\lambda_2}}{\frac{(\lambda_1+\lambda_2)^n}{n!}e^{-(\lambda_1+\lambda_2)}} \\
&= \frac{n!}{k!(n-k)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}
\end{aligned}$$

结果说明给定 $X_1+X_2=n$ 时, X_1 服从以 $p=\frac{\lambda_1}{\lambda_1+\lambda_2}$ 为参数的二项分布。

(2) 可以从 Poisson 分布的引入考虑，参数 λ 描述了一小段时间内发生一次小概率事件的概率。如果已知两类事件一共发生了 n 次，那么取其中的一个事件，是第一类事件的概率为 $\frac{\lambda_1}{\lambda_1+\lambda_2}$，进而可以用二项分布描述这个条件概率。

8. (1) 设甲的到达时间为下午 1 点之后 X 分钟，乙的到达时间为下午 1 点之后 Y 分钟，那么 (X,Y) 的概率密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{3600}, & 0\le x\le 60, 0\le y\le 60\\ 0, & \text{其他情况}\end{cases}$$

(2) 所求概率为

$$\begin{aligned}
P(|X-Y|\ge 10) &= P(X\le Y-10)+P(X\ge Y+10)\\
&= \int_{10}^{60}\int_0^{y-10}\frac{1}{3600}\mathrm{d}x\mathrm{d}y+\int_0^{50}\int_{y+10}^{60}\frac{1}{3600}\mathrm{d}x\mathrm{d}y\\
&= \int_{10}^{60}\frac{y-10}{3600}\mathrm{d}y+\int_0^{50}\frac{50-y}{3600}\mathrm{d}y\\
&= \frac{25}{36}
\end{aligned}$$

Tips: 应当明确写出随机变量的含义和取值范围。不能只给出概率密度函数而不加以解释。

9. (1)

$$\begin{aligned}
H_X(x) &= \lim_{y\to+\infty}H(x,y)\\
&= F(x)[\lim_{y\to+\infty}G(y)]\{1+\alpha[1-F(x)][1-\lim_{y\to+\infty}G(y)]\}\\
&= F(x)\\
H_Y(y) &= \lim_{x\to+\infty}H(x,y)\\
&= [\lim_{x\to+\infty}F(x)]G(y)\{1+\alpha[1-\lim_{x\to+\infty}F(x)][1-G(y)]\}\\
&= G(y)
\end{aligned}$$

(2) 利用题中所给的函数。区间 $[0,1]$ 上均匀分布的累积分布函数为

$$F(x)=\begin{cases}0, & x<0\\ x, & 0\le x\le 1\\ 1, & x>1\end{cases}$$

在 $H(X,Y)$ 的表达式中，将 $F(x),G(y)$ 都取为均匀分布对应的分布函数，再取 $\alpha\pm 1$ 即可。

10. 根据 Copula 函数的定义，有 $C(u,1)=u,\ C(1,v)=v$。取二元分布函数 $C(F(x),G(y))$ 即可满足。

Tips: 关于 Copula 函数更详细的解释见第三次习题课讲义。注意，一些手动构造 (例如 $F(x)+G(y)-1$) 并不满足要求. 需要依据 Copula 函数进行构造，即利用 $C(u,v)$ 的抽象形式。虽然直接利用第 9 题的函数，构造如 $H(x,y)=F(x)G(y)$ 也符合边际分布分别为 $F(x)$ 和 $G(y)$，但失去了一般性。

11.

- X 为离散型随机变量，Y 为离散型随机变量

$$P(X=x)=\sum_y P(X=x,Y=y)=\sum_y P(X=x|Y=y)P(Y=y)$$
$$P(Y=y|X=x)=\frac{P(X=x,Y=y)}{P(X=x)}=\frac{P(X=x|Y=y)P(Y=y)}{\sum_{y_i}P(X=x|Y=y_i)P(Y=y_i)}$$

- X 为离散型随机变量，Y 为连续型随机变量

$$P(X=x)=\int_{-\infty}^{\infty}P(X=x|Y=y)f_Y(y)\mathrm{d}y$$
$$f_{Y|X}(y|x)=\frac{f(x,y)}{P(X=x)}=\frac{P(X=x|Y=y)f_Y(y)}{\int_{-\infty}^{\infty}P(X=x|Y=t)f_Y(t)\mathrm{d}t}$$

- X 为连续型随机变量，Y 为离散型随机变量

$$f_X(x)=\sum_y f_{X|Y}(x|y)P(Y=y)$$
$$P(Y=y|X=x)=\frac{f(x,y)}{f_X(x)}=\frac{f_{X|Y}(x|y)P(Y=y)}{\sum_{y_i}f_{X|Y}(x|y_i)P(Y=y_i)}$$

- X 为连续型随机变量，Y 为连续型随机变量

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)\mathrm{d}y=\int_{-\infty}^{\infty}f_{X|Y}(x|y)f_Y(y)\mathrm{d}y$$
$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{f_{X|Y}(x|y)f_Y(y)}{\int_{-\infty}^{\infty}f_{X|Y}(x|t)f_Y(t)\mathrm{d}t}$$

Tips: 本题重点关注离散型和连续型随机变量的区别，包括什么时候使用概率 $P(\cdot)$，什么时候使用概率密度 $f(\cdot)$，以及对离散型随机变量使用求和 $\sum$，对连续型随机变量使用积分 $\int$. 另外，X 和 Y 分别为离散型和连续型的两种情况不是“同理”“一致”的关系。

12. (1) 根据概率密度函数的性质，有

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\mathrm{d}x\mathrm{d}y=1$$

作极坐标变换

$$
\begin{aligned}
1 &= \int_0^{2\pi}\int_0^1 \frac{c\rho}{1+\rho^2}\mathrm{d}\rho\mathrm{d}\theta \\
&= \int_0^{2\pi}\int_0^1 \frac{c}{2(1+t)}\mathrm{d}t\mathrm{d}\theta \\
&= \int_0^{2\pi} \frac{c}{2}\ln 2\mathrm{d}\theta \\
&= \pi c\ln 2
\end{aligned}
$$

因此 $c = \frac{1}{\pi\ln 2}$.

(2) X 的边际密度

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y \\
&= c\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{1+x^2+y^2}\mathrm{d}y \\
&= \frac{c}{1+x^2}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{1+(\frac{y}{\sqrt{1+x^2}})^2}\mathrm{d}y \\
&= \frac{c}{\sqrt{1+x^2}}\arctan\frac{y}{\sqrt{1+x^2}}\Bigg|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \\
&= \frac{2}{\pi\sqrt{1+x^2}\ln 2}\arctan\sqrt{\frac{1-x^2}{1+x^2}}, -1\le x\le 1
\end{aligned}
$$

同理，Y 的边际密度为

$$
f_Y(y) = \frac{2}{\pi\sqrt{1+y^2}\ln 2}\arctan\sqrt{\frac{1-y^2}{1+y^2}}, -1\le y\le 1
$$

容易验证 $f_X(x)f_Y(y)\neq f(x,y)$，因此 X,Y 不独立

13. 需要证明 $g(x,y)$ 在 $\mathbb{R}^2$ 上积分为 1，以及 $g(x,y)$ 非负。

证明：由于 X,Y 独立且都服从标准正态分布 $N(0,1)$, 其联合概率密度函数

$$
f(x,y) = f(x)f(y) = \frac{1}{2\pi}\exp(-\frac{1}{2}(x^2+y^2))
$$

根据对称性，有

$$
\iint_{x^2+y^2\le 1}\frac{xy}{100}\mathrm{d}x\mathrm{d}y = 0
$$

因此在整个 $\mathbb{R}^2$ 上，$\iint_{\mathbb{R}^2} g(x,y)\mathrm{d}x\mathrm{d}y = \iint_{\mathbb{R}^2} f(x,y)\mathrm{d}x\mathrm{d}y = 1$.

当 $x^2+y^2\le 1$ 时，$\frac{xy}{100}$ 的最小值为 $-\frac{1}{200}$. 但是在 $x^2+y^2\le 1$ 的区域中，$f(x,y)$ 的最小值为 $\frac{1}{2\pi}e^{-\frac{1}{2}}\approx 0.096 > 0.005$. 因此 $g(x,y)\ge 0$ 成立，所以 $g(x,y)$ 是一个二维概率密度函数。

(2) 在关于原点对称的区间 I 上，有 $\int_I \frac{xy}{100}\mathrm{d}x = \int_I \frac{xy}{100}\mathrm{d}y = 0$. 因此 U 的边际分布

$$
g_U(x) = \int_{-\infty}^{\infty} g(x,y)\mathrm{d}y = \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y = f_X(x)
$$

同理 $g_V(y) = f_Y(y)$. 即 U,V 都服从标准正态分布。但可以从 (U,V) 的联合概率密度函数 $g(x,y)$ 看出 (U,V) 的联合分布不是二元正态分布。

评价：本题最后一个设问是在问 $g(x,y)$ 是否服从二元正态分布。这是否应该是一个“显然”的事情？如果严格证明，可以先假设 $g(x,y)$ 是一个二元正态分布，那么由于在 $x^2+y^2>1$ 时 $g(x,y)=f(x,y)$，这个区域上的取值限制会完全固定二元正态分布函数中所有参数的取值，则 $g(x,y)$ 必然应当恒等于 $f(x,y)$，与 $x^2+y^2\leq 1$ 时的情况矛盾。或者也可以考虑固定 $y=\frac{\sqrt{2}}{2}$，那么 $g(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2})=f(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2})+\frac{1}{200}$，但是 $\lim_{x\to(\frac{\sqrt{2}}{2})^+} g(x,\frac{\sqrt{2}}{2})=f(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2})$. 这与正态分布函数的连续性矛盾。

14. 绘图略去。根据作业 4-10(3), 由 $x=-\ln(1-y)$ 得 $y=1-e^{-x}$，即 X 的累积分布函数为 $F(x)=1-e^{-x}$. 概率密度函数为 $f(x)=e^{-x}, x\geq 0$. 绘图结果应当可以展现二者图像的相似程度。如果在同一坐标尺度下绘制，注意要把 x 直方图中的数量转变为出现的频率。进行比较的指数分布参数 λ 取 1。

作业总结：

- 本次作业的内容主要涉及联合分布一节，包括二维情况下的概率密度，边际分布，条件分布等。
- 主要集中出现的问题包括：第 3 题第 6 题中混淆累积分布函数（cdf,$F(x)$）和概率密度函数（pdf,$f(x)$）。第 3 题第 (4) 问求期望的方法。第 4 题第 5 题的计算过程。第 10 题部分同学不能理解题意。第 11 题中缺失 X 和 Y 一个是离散型一个是随机型的情况。同时在第 11 题中，有一些错误的记号，包括对连续型随机变量 Y 求 $P(Y=y|X=x)$, 求和与积分使用不当，不加说明地使用 A,B 而不阐述与 X,Y 的关系。一些同学统一使用了 $f(\cdot)$ 表示，虽然讲义中在引入离散随机变量的 PMF 时也使用了 f，但这容易和概率密度函数混淆，建议使用 $P(\cdot)$. 第 12 题的第（2）问无法正确计算积分。第 13 题的第（1）问不证明 $g(x,y)$ 的非负性，或者认为这是显然的 (比如认为 $xy\geq 0$，不讨论 $f(x,y)$ 具体的取值范围而直接判定 $f(x,y)+\frac{xy}{100}\geq 0$)，少部分同学将 $f(x,y)$ 表达式写错。