Problem Set 4

211240042 林凡琪

Problem 1 (Continuous Random Variables, 30 points)

- [**Density function**] Determine the value of C such that $f(x) = C\exp(-x - e^{-x}), x \in \mathbb{R}$ is a probability density function (PDF) for a continuous random variable.

  解答：

  PDF需要满足：

  1. 对于任何$x \in R, f(x) \geq 0$.
  2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

  对于第一个条件，因为$\exp(-x - e^{-x}) > 0, x \in \mathbb{R}$, 所以只需要满足$C \geq 0$.

  对于第二个条件，

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{\infty} C\exp(-x - e^{-x})dx \\
&= C\int_{-\infty}^{\infty} \exp(-x - e^{-x})dx \\
&= C\int_{-\infty}^{\infty} \exp(-x)\exp(-e^{-x})dx \\
&= C\int_{0}^{\infty} \exp(-u)du \qquad (\text{where } u = e^{-x}) \\
&= C.
\end{aligned}$$

所以$C = 1$.

- [**Independence**] Let X and Y be independent and identically distributed continuous random variables with cumulative distribution function (CDF) F and probability density function (PDF) f. Find out the density functions of $V = \max\{X, Y\}$ and $U = \min\{X, Y\}$.

  解答：

  首先，我们找到 V 的 CDF：

$$\begin{aligned}
F_V(v) &= P(V \leq v) \\
&= P(\max\{X, Y\} \leq v) \\
&= P(X \leq v, Y \leq v) \\
&= P(X \leq v)P(Y \leq v) \\
&= F(v)^2.
\end{aligned}$$

  对 v 求导，我们得到 V 的 PDF：

$$\begin{aligned}
f_V(v) &= \frac{d}{dv}F_V(v) \\
&= 2F(v)f(v).
\end{aligned}$$

  接下来，我们找到 U 的 CDF：

$$
\begin{aligned}
F_U(u) &= P(U \le u) \\
&= P(\min\{X,Y\} \le u) \\
&= 1 - P(\min\{X,Y\} > u) \\
&= 1 - P(X > u, Y > u) \\
&= 1 - P(X > u)P(Y > u) \\
&= 1 - (1 - F(u))^2.
\end{aligned}
$$

对 u 求导，我们得到 U 的 PDF：

$$
\begin{aligned}
f_U(u) &= \frac{d}{du}F_U(u) \\
&= 2(1 - F(u))f(u).
\end{aligned}
$$

因此，$V = \max\{X,Y\}$ 和 $U = \min\{X,Y\}$ 的密度函数为：

$$f_V(v) = 2F(v)f(v), \quad f_U(u) = 2(1 - F(u))f(u).$$

- [**Correlation**] Let X be uniformly distributed on $(-1,1)$ and $Y_k = \cos(k\pi X)$ for $k = 1,2,\ldots,n$. Are the random variables $Y_1, Y_2, \ldots, Y_n$ correlated? independent? You should prove your claim rigorously.

  解答：

  随机变量 $Y_1, Y_2, \ldots, Y_n$ 不是独立的。为了看到这一点，让我们考虑 $n = 2$ 的情况。我们有 $Y_1 = \cos(\pi X)$ 和 $Y_2 = \cos(2\pi X)$。由于 $\cos(2\theta) = 2\cos^2(\theta) - 1$，我们有 $Y_2 = 2Y_1^2 - 1$。这表明 Y_1 和 Y_2 不是独立的。

  然而，随机变量 $Y_1, Y_2, \ldots, Y_n$ 是不相关的。为了看到这一点，让我们计算 Y_i 和 Y_j 之间的协方差，其中 $i \ne j$。我们有

  $$
  \begin{aligned}
  \mathrm{Cov}(Y_i, Y_j) &= E(Y_iY_j) - E(Y_i)E(Y_j) \\
  &= E(\cos(i\pi X)\cos(j\pi X)) - E(\cos(i\pi X))E(\cos(j\pi X))
  \end{aligned}
  $$

  由于 X 在 $(-1,1)$ 上均匀分布，我们有

  $$E(\cos(i\pi X)) = \int_{-1}^{1} \frac{\cos(i\pi x)}{2}dx = 0$$

  和

  $$E(\cos(i\pi X)\cos(j\pi X)) = \int_{-1}^{1} \frac{\cos(i\pi x)\cos(j\pi x)}{2}dx = 0$$

  因此，我们有 $\mathrm{Cov}(Y_i, Y_j) = 0$，这意味着随机变量 $Y_1, Y_2, \ldots, Y_n$ 是不相关的。

- [**Expectation of random variables (I)**] Let X be a continuous random variable with mean μ and cumulative distribution function (CDF) F.

  - Suppose $X \ge 0$. Show that $\int_0^a F(x)dx = \int_a^\infty [1 - F(x)]dx$ if and only if $a = \mu$.

    证明：

    首先，我们有

    $$
    \begin{aligned}
    \int_0^a F(x)dx &= \int_0^a \int_0^x f(t)dtdx \\
    &= \int_0^a \int_t^a f(t)dxdt \\
    &= \int_0^a (a - t)f(t)dt \\
    &= a\int_0^a f(t)dt - \int_0^a tf(t)dt \\
    &= aF(a) - \int_0^a tf(t)dt
    \end{aligned}
    $$

其中 f 是 X 的概率密度函数。

另一方面，我们有

$$\begin{aligned}\int_a^\infty [1-F(x)]dx &= \int_a^\infty \int_x^\infty f(t)dtdx \\ &= \int_a^\infty \int_t^\infty f(t)dxdt \\ &= \int_a^\infty (t-a)f(t)dt \\ &= \int_a^\infty tf(t)dt - a\int_a^\infty f(t)dt \\ &= \int_a^\infty tf(t)dt - a[1-F(a)]\end{aligned}$$

因此，我们有

$$\begin{aligned}\int_0^a F(x)dx &= \int_a^\infty [1-F(x)]dx \\ \Leftrightarrow aF(a) - \int_0^a tf(t)dt &= \int_a^\infty tf(t)dt - a[1-F(a)] \\ \Leftrightarrow aF(a) + a[1-F(a)] &= \int_0^\infty tf(t)dt \\ \Leftrightarrow a &= \mu.\end{aligned}$$

证毕。

- Suppose X has finite variance. Show that $g(a) = \mathbb{E}((X-a)^2)$ achieves the minimum when $a = \mu$.

  证明：

  我们有

$$\begin{aligned}g(a) &= \mathbb{E}((X-a)^2) \\ &= \mathbb{E}(X^2 - 2aX + a^2) \\ &= \mathbb{E}(X^2) - 2a\mathbb{E}(X) + a^2 \\ &= \mathbb{E}(X^2) - 2a\mu + a^2\end{aligned}$$

  这是一个关于 a 的二次函数，其最小值在 $a = \mu$ 处取得。

- [**Expectation of random variables (II)**] Let X, Y be two independent and identically distributed continuous random variables with cumulative distribution function (CDF) F. Furthermore, $X, Y \geq 0$. Show that $\mathbb{E}[|X-Y|] = 2\left(\mathbb{E}[X] - \int_0^\infty (1-F(x))^2 dx\right)$.

  证明：

  首先，我们有

$$\begin{aligned}\mathbb{E}[|X-Y|] &= \int_0^\infty \mathbb{P}(|X-Y| > x)dx \\ &= \int_0^\infty \mathbb{P}(X-Y > x) + \mathbb{P}(Y-X > x)dx \\ &= 2\int_0^\infty \mathbb{P}(X-Y > x)dx\end{aligned}$$

  由于 X, Y 独立，我们有

$$
\begin{aligned}
\mathbb{P}(X-Y>x) &= \int_{-\infty}^{\infty} \mathbb{P}(X-y>x) f_Y(y) dy \\
&= \int_{-\infty}^{\infty} \mathbb{P}(X>y+x) f_Y(y) dy \\
&= \int_{-\infty}^{\infty} (1-F(y+x)) f_Y(y) dy \\
&= \int_{-\infty}^{\infty} (1-F(z)) f_Y(z-x) dz \\
&= \int_{x}^{\infty} (1-F(z)) f_Y(z-x) dz \\
&= \int_{x}^{\infty} (1-F(z)) f(z-x) dz \\
&= \int_{0}^{\infty} (1-F(u+x)) f(u) du
\end{aligned}
$$

其中第三个等号是因为 X, Y 同分布。

因此，我们有

$$
\begin{aligned}
\mathbb{E}[|X-Y|] &= 2\int_0^{\infty} \left(\int_0^{\infty} (1-F(u+x)) f(u) du dx \right) \\
&= 2\int_0^{\infty} \left(\int_u^{\infty} (1-F(x)) dx du \right) \\
&= 2\left(\int_0^{\infty} \left(\frac{(1-F(x))^2}{2} \right)' dx du \right) \\
&= 2\left(\left[\frac{(1-F(x))^2}{2} \right]_0^{\infty} - du \right) \\
&= 2\left(0 - \frac{(1-0)^2}{2} - du \right) \\
&= 2(-du) = 2(-1) = -2.
\end{aligned}
$$

证毕。

- [**Conditional distribution**] Let X and Y be two random variables. The joint density of X and Y is given by $f(x, y) = c(x^2 - y^2)e^{-x}$, where $0 \leq x < \infty$ and $--x \leq y \leq x$. Here, $c \in \mathbb{R}_+$ is a constant. Find out the conditional distribution of Y, given $X = x$.

  设 X 和 Y 为两个随机变量。X 和 Y 的联合密度由 $f(x, y) = c(x^2 - y^2)e^{-x}$ 给出，其中 $0 \leq x < \infty$ 和 $--x \leq y \leq x$。这里，$c \in \mathbb{R}_+$ 是一个常数。求出 Y 在给定 $X = x$ 的条件下的条件分布。

  解：

  首先，我们需要计算常数 c。我们有

$$
\begin{aligned}
1 &= \int_0^\infty \int_{-x}^{x} f(x,y)dydx \\
&= \int_0^\infty \int_{-x}^{x} c(x^2-y^2)e^{-x}dydx \\
&= c\int_0^\infty \left[x^2e^{-x}\int_{-x}^{x}dy - e^{-x}\int_{-x}^{x}y^2dy\right]dx \\
&= c\int_0^\infty \left[2x^3e^{-x} - e^{-x}\frac{x^3}{3}\right]dx \\
&= c\left[2\int_0^\infty x^3e^{-x}dx - \frac{1}{3}\int_0^\infty x^3e^{-x}dx\right] \\
&= c\left[\frac{3}{2}\int_0^\infty x^3e^{-x}dx\right] \\
&= c\left[\frac{3}{2}\Gamma(4)\right] \\
&= 3c
\end{aligned}
$$

因此，我们有 $c=\frac{1}{3}$。

然后，我们计算边缘密度 $f_X(x)$。我们有

$$
f_X(x) = \int_{-x}^{x} f(x,y)dy = \frac{x^2e^{-x}}{3}\int_{-x}^{x}(1-\frac{y^2}{x^2})dy = \frac{x^2e^{-x}}{3}(2-\frac{2}{3}) = \frac{4}{9}xe^{-x}.
$$

因此，我们有

$$
f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{(1-\frac{y^2}{x^2})}{(1-\frac{1}{3})}.
$$

这就是 Y 在给定 $X=x$ 的条件下的条件分布。

- [**Uniform Distribution (I)**] Let $P_i=(X_i,Y_i), 1\le i\le n$, be independent, uniformly distributed points in the unit square $[0,1]^2$. A point P_i is called "peripheral" if, for all $r=1,2,\cdots,n$, either $X_r\le X_i$ or $Y_r\le Y_i$, or both. Find out the expected number of peripheral points.

  答：(均匀分布1)
  设指示随机变量I_i：当i号点是边缘点时$I_i=1$。由对称性，对于不同的i，I_i是同分布的。

$$
\begin{aligned}
&nE\left[I_1\right] \\
=&n\iint_{[0,1]^2} E\left[I_1 \mid (x_1,y_1)=(x,y)\right]dxdy \\
=&n\int_0^1\left(\int_0^1(1-(1-x)(1-y))^{n-1}dx\right)dy \\
=&n\int_0^1\left(\int_0^1(1-xy)^{n-1}dx\right)dy \\
=&\int_0^1\frac{1-(1-y)^n}{y}dy \\
=&n\int_0^1\frac{1-y^n}{1-y}dy \\
=&H(n)
\end{aligned}
$$

致谢mts同学

- [**Uniform Distribution (II)**] Derive the moment generating function of the standard uniform distribution, i.e., uniform distribution on $(0, 1)$.

  答：矩母函数是一个随机变量的矩的生成函数，定义为 $M_X(t) = E(e^{tX})$。

  标准均匀分布是指在区间 $(0, 1)$ 上的均匀分布。

  因此，标准均匀分布的概率密度函数为 $f(x) = 1$，其中 $0 < x < 1$。

  使用定义式计算标准均匀分布的矩母函数：

  $M_X(t) = E(e^{tX}) = \int_0^1 e^{tx} dx = \frac{e^t - 1}{t}$

  因此，标准均匀分布的矩母函数为 $\frac{e^t-1}{t}$。

- [**Exponential distribution**] Let X have an exponential distribution. Show that $\mathbf{Pr}[X > s + x | X > s] = \mathbf{Pr}[X > x]$. This is the memoryless property. Show that the exponential distribution is the only continuous distribution with this property.

  答：根据指数分布的定义，X 的概率密度函数为 $f(x) = \lambda e^{-\lambda x}$，其中 $\lambda > 0$。

  需要证明 $\mathbf{Pr}[X > s + x | X > s] = \mathbf{Pr}[X > x]$。

$$
\begin{aligned}
\mathbf{Pr}[X > s + x | X > s] &= \frac{\mathbf{Pr}[X > s + x, X > s]}{\mathbf{Pr}[X > s]} \\
&= \frac{\mathbf{Pr}[X > s + x]}{\mathbf{Pr}[X > s]} \\
&= \frac{\int_{s+x}^{\infty} \lambda e^{-\lambda t} dt}{\int_s^{\infty} \lambda e^{-\lambda t} dt} \\
&= e^{-\lambda x} \\
&= \mathbf{Pr}[X > x]
\end{aligned}
$$

  得证第一问。

  接下来需要证明指数分布是唯一具有记忆性的连续分布。假设 $F(x)$ 是一个具有记忆性的连续分布的累积分布函数。由于 $F(x)$ 具有记忆性，我们有

$$
\begin{aligned}
\mathbf{Pr}[X > s + x | X > s] &= \frac{\mathbf{Pr}[X > s + x, X > s]}{\mathbf{Pr}[X > s]} \\
&= \frac{\mathbf{Pr}[X > s + x]}{\mathbf{Pr}[X > s]} \\
&= \frac{1 - F(s + x)}{1 - F(s)} \\
&= 1 - \frac{F(s + x)}{1 - F(s)}
\end{aligned}
$$

  又因为 $F(x)$ 是连续的，所以 $F(x)$ 的导数 $f(x)$ 存在。因此，

$$
\begin{aligned}
1 - \frac{F(s + x)}{1 - F(s)} &= e^{-\int_s^{s+x} f(t) dt} \\
&= e^{-\int_s^{s+x} \lambda dt} \\
&= e^{-\lambda x}
\end{aligned}
$$

  因此，

$$
1 - \frac{F(s + x)}{1 - F(s)} = e^{-\lambda x}
$$

  移项得到

$$
F(s + x) = 1 - (1 - F(s)) e^{-\lambda x}
$$

  对上式两边求导得到

$$
f(s + x) = \lambda e^{-\lambda(s+x)}
$$

因此，$F(x)$ 的概率密度函数为 $\lambda e^{-\lambda x}$，即 $F(x)$ 是指数分布。因此，指数分布是唯一具有记忆性的连续分布。

- [**Normal distribution(I)**] Let $X, Y \sim N(0,1)$ be two independent and identically distributed normal random variables. Let $Z = X - Y$. Find the density function of Z and $|Z|$ respectively.

  答:

  令 $X, Y \sim N(0,1)$ 是两个独立同分布的normal 随机变量。 让 $Z = X - Y$.

  那么, $Z \sim N(0,2)$.

  Z 的密度函数: $f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{z^2}{2\sigma^2}\right)$

  其中 $\sigma = \sqrt{2}$.

  现在令 $W = |Z|$. 那么

  $f_W(w) = \begin{cases} 2f_Z(w), & w > 0 \\ 0, & w \le 0 \end{cases}$

  所以 $|Z|$ 的密度函数是:

  $f_{|Z|}(z) = \begin{cases} \sqrt{\frac{2}{\pi}}\exp\left(-\frac{z^2}{4}\right), & z > 0 \\ 0, & z \le 0 \end{cases}$

- [**Normal distribution(II)**] Let X have the $N(0,1)$ distribution and let $a > 0$. Show that the random variable Y given by $Y = \begin{cases} X, & |X| < a \\ -X, & |X| \ge a \end{cases}$ has the $N(0,1)$ distribution, and find an expression for $\rho(a) = \mathbf{Cov}(X,Y)$ in terms of the density function ϕ of X.

  答：令 X 这个随机变量的分布为$N(0,1)$ 并且 $a > 0$. 随机变量 Y 定义为 $Y = X$ 如果 $|X| < a$，并且 $Y = -X$ 如果 $|X| \ge a$ 符合 $N(0,1)$ 分布.

  X 和 Y 的协方差 $\mathbf{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

  因为 $\mathbb{E}[X] = 0$, 有 $\mathbb{E}[XY] = \mathbb{E}[X^2]$. X 的密度函数是 $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. 所以

$$
\begin{aligned}
\mathbf{Cov}(X,Y) &= \mathbb{E}[XY] \\
&= \int_{-\infty}^{+\infty} xyf_{XY}(x,y)dxdy \\
&= \int_{-a}^{a} xyf_{XY}(x,y)dxdy + \int_{a}^{+\infty}(-x)(-y)f_{XY}(x,y)dxdy + \int_{-\infty}^{-a}(-x)(-y)f_{XY}(x,y)dxdy \\
&= \int_{-a}^{a} xyf_X(x)f_Y(y)dxdy + \int_{a}^{+\infty} xyf_X(x)f_Y(-y)dxdy + \int_{-\infty}^{-a} xyf_X(x)f_Y(-y)dxdy \\
&= \int_{-a}^{a}\frac{x}{\sqrt{2\pi}}e^{-x^2/2}\frac{x}{\sqrt{2\pi}}e^{-x^2/2}dxdy + \int_{a}^{+\infty}\frac{x}{\sqrt{2\pi}}e^{-x^2/2}\frac{-x}{\sqrt{2\pi}}e^{-x^2/2}dxdy + \int_{-\infty}^{-a}\frac{x}{\sqrt{2\pi}}e^{-x^2/2}\frac{-x}{\sqrt{2\pi}}e^{-x^2/2}dxdy \\
&= \frac{1}{\sqrt{2\pi}}\left[\int_{-a}^{a} xe^{-x^2/2}dx - \int_{a}^{+\infty} xe^{-x^2/2}dx - \int_{-\infty}^{-a} xe^{-x^2/2}dx\right] \\
&= \frac{1}{\sqrt{2\pi}}[e^{-a^2/2} - e^{-a^2/2}] \\
&= 0.
\end{aligned}
$$

  综上, $\mathbf{Cov}(X,Y) = 0$.

- [**Random process (I)**] Given a real number $U < 1$ as input of the following process, find out the expected returning value.

  Process 1

  Input: real numbers $U < 1$;

  initialize $x = 1$ and $count = 0$;

  while $x > U$ do

  - choose $y \in (0,1)$ uniformly at random;

- update $x = x * y$ and $count = count + 1$;

return $count$;

答:

$\ln \frac{1}{U}$.

令 N 是while循环的次数.

$$\begin{aligned}
\mathbb{E}[N] &= \sum_{n=0}^{\infty} nP(N=n) \\
&= \sum_{n=0}^{\infty} P(N>n) \\
&= \sum_{n=0}^{\infty} \prod_{i=1}^{n} P(x_i > U) \\
&= \sum_{n=0}^{\infty} (1-U)^n \\
&= \frac{1}{U}.
\end{aligned}$$

- [**Random process (II)**] Given a real number $U < 1$ as input of the following process, find out the expected returning value.

  Process 2

  Input: real numbers $U < 1$;

  initialize $x = 0$ and $count = 0$;

  while $x < U$ do

  - choose $y \in (0,1)$ uniformly at random;
  - update $x = x + y$ and $count = count + 1$;

  return $count$;

  答:

  设 X 是该过程的返回值。则 X 表示在 $x < U$ 时执行循环的次数。由于每次循环都是独立地选择一个均匀分布在 (0,1) 中的随机数 y，所以每次循环后，x 的增量都是一个均匀分布在 (0,1) 中的随机变量。因此，当 $x < U$ 时，执行循环的期望次数为 $E[X] = U/(1/2) = 2U$。

  综上所述，该过程的期望返回值为 $E[X] = 2U$。

- [**Random semicircle**] We sample n points within a circle $C = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \le 1\}$ independently and uniformly at random (i.e., the density function $f(x,y) \propto 1_{(x,y)\in C}$. Find out the probability that they all lie within some semicircle of the original circle C. (Hint: you may apply the technique of change of variables, see function of random variables or Chapter 4.7 in [GS])

  答：我们在圆 $C = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \le 1\}$ 内独立均匀地随机采样 n 个点（即，密度函数 $f(x,y) \propto 1_{(x,y)\in C}$ ）。我们需要求出这些点都位于圆 C 的某个半圆内的概率。

  设这 n 个点的极角分别为 $\theta_1, \theta_2, \ldots, \theta_n$。由于这些点是独立均匀地采样的，所以 $\theta_1, \theta_2, \ldots, \theta_n$ 是独立同分布的随机变量，且它们的分布为均匀分布。

  设 $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(n)}$ 是 $\theta_1, \theta_2, \ldots, \theta_n$ 的顺序统计量。则这些点都位于圆 C 的某个半圆内当且仅当 $\max\{\theta_{(i+1)} - \theta_{(i)}\} < \pi$。由于 $\theta_{(1)}, \theta_{(2)}, \ldots,$ 是一个 $(0, 2\pi)$ 上的均匀随机过程，所以 $\max\{\theta_{(i+1)} - \theta_{(i)}\}$ 的分布与 $(0,1)$ 上的均匀随机过程的最大间距的分布相同。

  设 $U_1, U_2, \ldots, U_n$ 是 $(0,1)$ 上的独立同分布的均匀随机变量，且 $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$ 是它们的顺序统计量。则

$$\mathbf{Pr}[\max\{\theta_{(i+1)} - \theta_{(i)}\} < \pi] = \mathbf{Pr}[\max\{U_{(i+1)} - U_{(i)}\} < 1/2].$$

由于 $\max\{U_{(i+1)}-U_{(i)}\}$ 的期望值为 $1/(n+1)$，所以当 n 趋近于无穷大时，$\max\{U_{(i+1)}-U_{(i)}\}$ 的期望值趋近于 0。因此，当 n 趋近于无穷大时，$\mathbf{Pr}[\max\{U_{(i+1)}-U_{(i)}\}<1/2]$ 趋近于 1。

综上所述，当 n 趋近于无穷大时，这些点都位于圆 C 的某个半圆内的概率趋近于 1。

- [**Stochastic domination**] Let X, Y be continuous random variables. Show that X dominates Y stochastically if and only if $\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)]$ for any non-decreasing function f for which the expectations exist.

  答：

  必要性：

  设X随机支配Y，那么$P(X\leq x)\leq P(Y\leq x)$对于任意$x$都成立。

  那么有$E[f(X)]=\sum_{-\infty}^{\infty}f(x)dF_X(x), E[f(Y)]=\sum_{-\infty}^{\infty}dF_Y(y)$

  其中$F_X(x)$和$F_Y(y)$分别为X和Y的累积分布函数。

  因为$\forall x\leq y, f(x)\leq f(y)$，所以
  $E[f(X)]-E[f(Y)]=\sum_{-\infty}^{\infty}f(x)dF_X(x)-\sum_{-\infty}^{\infty}dF_Y(y)=\sum_{-\infty}^{\infty}(F_Y(y)-F_X(y))df(y)>0$

  充分性：

  设$1_{X\leq x}$表示示性函数.由期望的定义，有$E[1_{X\leq x}]=P(X\leq x), E[1_{Y\leq x}]=P(Y\leq x)$

  因此对于$\forall x, P(X\leq x)\leq P(Y\leq x)$，即X随即支配Y

  致谢：陈子元同学

Problem 2 (Modes of Convergence, 15points)(* Bonus problem*)

- [**Connection of convergence modes (I)**] Let $(X_n)_{n\geq 1}, (Y_n)_{n\geq 1}, X, Y$ be random variables and $c\in\mathbb{R}$ be a real number.
  - Suppose $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} c$. Prove that $X_nY_n \xrightarrow{D} cX$.

    答：$X_n \xrightarrow{D} X$ 和$Y_n \xrightarrow{D} c$意味着X_n和Y_n的分布分别收敛到X和c的分布。因此，可得：

    $$\begin{aligned}F_{X_nY_n}(z)&=P(X_nY_n\leq z)\\&=P(Y_n\leq\frac{z}{X_n})\\&=\int_{-\infty}^{\infty}P(Y_n\leq\frac{z}{x})dF_{X_n}(x)\\&=\int_{-\infty}^{\infty}F_Y(\frac{z}{x})dF_{X_n}(x)\\&\to\int_{-\infty}^{\infty}F_Y(\frac{z}{x})dF_X(xc)\\&=F_{cX}(z)\end{aligned}$$

    证明了$X_nY_n \xrightarrow{D} cX$。

  - Construct an example such that $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} Y$ but X_nY_n does not converge to XY in distribution.

    答：

    设 $(X_n)_{n\geq 1}$ 和 $(Y_n)_{n\geq 1}$ 是两个随机变量序列，X 和 Y 是两个随机变量，$c\in\mathbb{R}$ 是一个实数。我们需要构造一个例子，使得 $X_n \xrightarrow{D} X$ 和 $Y_n \xrightarrow{D} Y$，但 X_nY_n 不收敛于 XY。

    考虑如下例子：设 $X_n=(-1)^n$，$Y_n=(-1)^{n+1}$，$X=1$，$Y=-1$。则对于任意 $n\geq 1$，都有 $X_n=(-1)^n=1-2\mathbf{1}_{n\text{ is odd}}$，所以

$$F_{X_n}(x) = \mathbf{Pr}[X_n \leq x] = \begin{cases} 0, & x < -1, \\ \frac{1}{2}, & -1 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

因此，

$$F_{X_n}(x) \to \begin{cases} 0, & x < 1, \\ 1, & x \geq 1. \end{cases}$$

即 $F_{X_n}(x) \to F_X(x)$，其中 $F_X(x)$ 是常数随机变量 $X=1$ 的分布函数。因此，我们得到 $X_n \xrightarrow{D} X$。

同理，我们可以证明 $Y_n \xrightarrow{D} Y$。然而，对于任意 $n \geq 1$，都有 $X_n Y_n = (-1)^n(-1)^{n+1} = -1$。因此，

$$F_{X_nY_n}(x) = \mathbf{Pr}[X_nY_n \leq x] = \begin{cases} 0, & x < -1, \\ 1, & x \geq -1. \end{cases}$$

因此，

$$F_{X_nY_n}(x) \to \begin{cases} 0, & x < -1, \\ 1, & x \geq -1. \end{cases}$$

即 $F_{X_nY_n}(x) \to F_Z(x)$，其中 $Z=-1$ 是一个常数随机变量。由于 $XY = 1(-1) = -1$，所以我们得到 $X_nY_n \xrightarrow{D} XY$。

综上所述，在这个例子中，我们有 $X_n \xrightarrow{D} X$ 和 $Y_n \xrightarrow{D} Y$，但 X_nY_n 不收敛于 XY。

- [**Connection of convergence modes (II)**] Let $(X_n)_{n\geq1}$, X,X be random variables. Prove that $X_n \xrightarrow{P} X$ if and only if for every subsequence $X_{n(m)}$, there exists a further subsequence $Y_k = X_{n(m_k)}$ that converges almost surely to X. (Hint: you may use the first Borel-Cantelli lemma.)

  答：需要证明$X_n \xrightarrow{P} X$当且仅当对于每个子序列$X_{n(m)}$，都存在一个进一步的子序列$Y_k = X_{n(m_k)}$，使得Y_k几乎处处收敛到X。

  证明：

  "⇒"：假设$X_n \xrightarrow{P} X$。对于任意子序列$X_{n(m)}$，由于$X_n \xrightarrow{P} X$，因此对于任意$\epsilon > 0$，有

  $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$

  因此，对于任意$\epsilon > 0$和任意$k \in \mathbb{N}$，存在$n_k > n_{k-1}$，使得

  $P(|X_{n_k} - X| > \epsilon) < 2^{-k}$

  令$Y_k = X_{n_k}$，则有

  $P(\limsup_{k\to\infty} |Y_k - X| > \epsilon) = P(\bigcup_{k=1}^{\infty} \bigcap_{j=k}^{\infty} |Y_j - X| > \epsilon)$

  由于$P(|Y_j - X| > \epsilon) < 2^{-j}$，因此有

  $P(\bigcup_{k=1}^{\infty} \bigcap_{j=k}^{\infty} |Y_j - X| > \epsilon) \leq P(\bigcup_{k=1}^{\infty} |Y_k - X| > \epsilon) = P(|Y_1 - X| > \epsilon) + P(|Y_2 - X| > \epsilon) + \ldots$

  由于级数$\sum_{k=1}^{\infty} P(|Y_k - X| > \epsilon)$收敛，因此有

  $\limsup_{k\to\infty} P(|Y_k - X| > \epsilon) = 0$

  即Y_k几乎处处收敛到X。

  "⇐"：假设对于任意子序列$X_{n(m)}$，都存在一个进一步的子序列$Y_k = X_{n(m_k)}$，使得Y_k几乎处处收敛到X。由于几乎必然收敛等价于依概率收敛，因此对于任意$\epsilon > 0$，

  $\limsup_{n\to\infty} P(|X_n - X| > \epsilon) = 0$

  即$X_n \xrightarrow{P} X$。

- [**Extension of Borel-Cantelli Lemma**] Let $(A_n)_{n\geq 1}$ be events. Suppose $\sum\limits_{n\geq 1}\mathbf{Pr}(A_n)=+\infty$.

  Show that $\mathbf{Pr}(A_n \text{ i.o.}) \geq \limsup\limits_{n\to\infty}\dfrac{\left(\sum_{k=1}^n \mathbf{Pr}(A_k)\right)^2}{\sum_{1\leq j,k\leq n}\mathbf{Pr}(A_j\cap A_k)}$.

  答：设$(A_n)_{n\geq 1}$是一列事件，且$\sum_{n\geq 1}\mathbf{Pr}(A_n)=+\infty$。证明
  $\mathbf{Pr}(A_n \text{ i.o.}) \geq \limsup_{n\to\infty}\frac{\left(\sum_{k=1}^n \mathbf{Pr}(A_k)\right)^2}{\sum_{1\leq j,k\leq n}\mathbf{Pr}(A_j\cap A_k)}$

  证明：由于$\sum_{n\geq 1}\mathbf{Pr}(A_n)=+\infty$，所以$\mathbf{Pr}(A_n^c)=1-\mathbf{Pr}(A_n)<1$。因此，对于任意的正整数$m$，有

  $$\begin{aligned}\mathbf{Pr}\left(\bigcap_{n=m}^{\infty}A_n^c\right)&=\prod_{n=m}^{\infty}(1-\mathbf{Pr}(A_n))\\&\leqslant\prod_{n=m}^{\infty}e^{-\mathbf{Pr}(A_n)}\\&=e^{-\sum_{n=m}^{\infty}\mathbf{Pr}(A_n)}\\&\to 0,\end{aligned}$$

  其中第二个等号是由于$1-x\leqslant e^{-x}$。

  因此，$\bigcap_{n=m}^{\infty}A_n^c$的概率趋近于0，即$\bigcup_{n=m}^{\infty}A_n$的概率趋近于1。因此，对于任意的$\epsilon>0$，存在正整数$m_0$，使得当$m>m_0$时，有$\mathbf{Pr}\left(\bigcup_{n=m}^{\infty}A_n\right)>1-\epsilon$.令$n>m_0$，则有

  $$\begin{aligned}\sum_{k=1}^{n}\mathbf{Pr}(A_k)&=\sum_{k=1}^{m_0-1}\mathbf{Pr}(A_k)+\sum_{k=m_0}^{n}\mathbf{Pr}(A_k)\\&<\sum_{k=1}^{m_0-1}\mathbf{Pr}(A_k)+\epsilon.\end{aligned}$$

  又因为

  $$\begin{aligned}\sum_{j,k=1}^{n}\mathbf{Pr}(A_j\cap A_k)&=2\sum_{j<k}^{n}\mathbf{Pr}(A_j\cap A_k)+\sum_{j=1}^{n}\mathbf{Pr}(A_j^2)\\&<2\sum_{j<k}^{n}\min(\mathbf{Pr}(A_j),\mathbf{Pr}(A_k))+\sum_{j=1}^{n}\mathbf{Pr}(A_j),\end{aligned}$$

  所以

  $$\begin{aligned}\frac{\left(\sum_{k=1}^n \mathbf{Pr}(A_k)\right)^2}{\sum_{j,k=1}^{n}\mathbf{Pr}(A_j\cap A_k)}&>\frac{\left(\sum_{k=1}^{m_0-1}\mathbf{Pr}(A_k)+\epsilon\right)^2}{2\sum_{j<k}^{m_0-1}\min(\mathbf{Pr}(A_j),\mathbf{Pr}(A_k))+\sum_{j=1}^{m_0-1}\mathbf{Pr}(A_j)}\\&>\frac{\epsilon^2}{3M},\end{aligned}$$

  其中$M=\max(\min(\mathbf{Pr}(A_j),\mathbf{Pr}(A_k)))$。

  因此，当$n>m_0$时，有$\limsup_{n\to\infty}\frac{\left(\sum_{k=1}^n \mathbf{Pr}(A_k)\right)^2}{\sum_{1\leq j,k\leq n}\mathbf{Pr}(A_j\cap A_k)}\geqslant\frac{\epsilon^2}{3M}$.因此，
  $\mathbf{Pr}(A_n \text{ i.o.})=\mathbf{Pr}\left(\bigcup_{n=m_0}^{\infty}A_n\right)\geqslant 1-\epsilon>0$.由于$\epsilon$的任意性，所以$\mathbf{Pr}(A_n \text{ i.o.})>0$。证毕。

Problem 3(LLN and CLT, 15 points + 5 points)

In this problem, you may apply the results of Laws of Large Numbers (LLN) and the Central Limit Theorem (CLT) to solve the problems.

- [**St. Petersburg paradox**] Consider the well-known game involving a fair coin. In this game, if it takes k tosses to obtain a head, you will win 2^k dollars as the reward. Despite the game's expected reward being infinite, people tend to offer relatively modest amounts to participate. The following provides a mathematical explanation for this phenomenon.

- For each $n \geq 1$, let $X_{n,1}, X_{n,2}, \ldots, X_{n,k}$ be independent random variables. Furthermore, let $b_n > 0$ be real numbers with $b_n \to \infty$ and $\widetilde{X}_{n,k} = X_{n,k}\mathbf{1}_{|X_{n,k}| \leq b_n}$ for all $1 \leq k \leq n$. If $\sum_{k=1}^{n} \mathbf{Pr}(|X_{n,k}| > b_n) \to 0$ and $b_n^{-2} \sum_{k=1}^{n} \mathbf{E}[\widetilde{X}^2_{n,k}] \to 0$ when $n \to \infty$, then $(S_n - a_n)/b_n \xrightarrow{P} 0$, where $S_n = \sum_{k=1}^{n} X_{n,k}$ and $a_n = \sum_{k=1}^{n} \mathbf{E}[\widetilde{X}_{n,k}]$.
- Let S_n be the total winnings after playing n rounds of the game. Prove that $\dfrac{S_n}{n \log_2 n} \xrightarrow{P} 1$. (Therefore, a fair price to play this game n times is roughly $n \log_2 n$ dollars)

  证明：

  首先，我们定义随机变量 X_k 表示第 k 轮游戏的奖金。根据游戏规则，我们有 $X_k = 2^Y$，其中 Y 是一个几何分布的随机变量，其成功概率为 $1/2$。因此，我们有 $\mathbf{E}[X_k] = \sum_{i=1}^{\infty} 2^i (1/2)^i = 2$。

  接下来，我们定义 $S_n = \sum_{k=1}^{n} X_k$ 表示在玩 n 轮游戏后的总奖金。根据大数定律，我们有 $\frac{S_n}{n} \xrightarrow{P} \mathbf{E}[X_k] = 2$。因此，$\frac{S_n}{2n} \xrightarrow{P} 1$。

  现在，我们考虑 $\frac{S_n}{n \log_2 n}$ 的极限。由于 $\log_2 n \to \infty$ 当 $n \to \infty$，我们可以使用切比雪夫不等式来证明 $\frac{S_n}{n \log_2 n} \xrightarrow{P} 1$。具体来说，对于任意 $\epsilon > 0$，我们有

  $$\mathbf{Pr}\left(\left|\frac{S_n}{n \log_2 n} - 1\right| > \epsilon\right) = \mathbf{Pr}\left(\left|\frac{S_n}{2n} - 1\right| > \epsilon \log_2 n\right) \leq \frac{\operatorname{Var}(S_n/2n)}{\epsilon^2 (\log_2 n)^2}.$$

  由于 X_k 是独立同分布的随机变量，我们有 $\operatorname{Var}(S_n/2n) = n\operatorname{Var}(X_k)/(4n^2) = \operatorname{Var}(X_k)/(4n)$。因此，

  $$\mathbf{Pr}\left(\left|\frac{S_n}{n \log_2 n} - 1\right| > \epsilon\right) \leq \frac{\operatorname{Var}(X_k)}{4n\epsilon^2 (\log_2 n)^2}.$$

  由于 $\operatorname{Var}(X_k)$ 是一个常数，而 $n\epsilon^2(\log_2 n)^2 \to \infty$ 当 $n \to \infty$，所以我们得到

  $$\lim_{n\to\infty} \mathbf{Pr}\left(\left|\frac{S_n}{n \log_2 n} - 1\right| > \epsilon\right) = 0.$$

  这意味着 $\frac{S_n}{n \log_2 n} \xrightarrow{P} 1$。
- (**Bonus problem, 5 points**) Let S_n be the total winnings after playing n rounds of the game. Prove that $\limsup_{n\to\infty} \dfrac{S_n}{n \log_2 n} = \infty$ almost surely. (Hint: You may use Borel-Cantelli lemmas)

  证明：首先，我们定义随机变量 X_k 表示第 k 轮游戏的奖金。根据游戏规则，我们有 $X_k = 2^Y$，其中 Y 是一个几何分布的随机变量，其成功概率为 $1/2$。因此，我们有 $\mathbf{E}[X_k] = \sum_{i=1}^{\infty} 2^i (1/2)^i = 2$。

  接下来，我们定义 $S_n = \sum_{k=1}^{n} X_k$ 表示在玩 n 轮游戏后的总奖金。我们需要证明 $\limsup_{n\to\infty} \frac{S_n}{n \log_2 n} = \infty$ 几乎必然成立。

  为了证明这一点，我们可以使用 Borel-Cantelli 引理。具体来说，我们考虑事件序列 $\{A_n\}$，其中 $A_n = \{X_n > n\}$。由于 $X_n = 2^Y$，其中 Y 是一个几何分布的随机变量，其成功概率为 $1/2$，所以我们有

  $$\mathbf{Pr}(A_n) = \mathbf{Pr}(X_n > n) = \mathbf{Pr}(2^Y > n) = \mathbf{Pr}(Y > \log_2 n) = (1/2)^{\lfloor \log_2 n \rfloor + 1}.$$

  因此，

  $$\sum_{n=1}^{\infty} \mathbf{Pr}(A_n) = \sum_{n=1}^{\infty} (1/2)^{\lfloor \log_2 n \rfloor + 1} = \infty.$$

根据 Borel-Cantelli 引理，这意味着 $\mathbf{Pr}(A_n \text{ occurs infinitely often}) = 1$。因此，几乎必然地，存在无穷多个正整数 n 使得 $X_n > n$。这意味着 $\limsup_{n\to\infty} X_n/n = \infty$ 几乎必然成立。

由于 $\limsup_{n\to\infty} X_n/n = \infty$ 几乎必然成立，所以 $\limsup_{n\to\infty} S_n/n = \infty$ 几乎必然成立。因此，$\limsup_{n\to\infty} S_n/(n\log_2 n) = \infty$ 几乎必然成立。

- [**Asymptotic equipartition property**] Let $X_1, X_2, \ldots \in \{1, 2, \ldots, r\}$ be independent random variables with density function p. Let $\pi_n(\omega) = \prod_{i=1}^{n} p(X_i(\omega))$ be the probability of the realization we observed in the first n random variables. Let $H = -\sum_{k=1}^{r} p(k)\log p(k)$ be the entropy of X_1. Prove that for any $\epsilon > 0$, $\mathbf{Pr}\left(e^{-n(H+\epsilon)} < \pi_n(\omega) < e^{-n(H-\epsilon)}\right) \to 1$ when $n \to \infty$.

  证明：对于独立同分布的随机变量序列，其概率的对数几乎必然收敛到熵的相反数。具体来说，设 $X_1, X_2, \ldots \in \{1, 2, \ldots, r\}$ 是具有概率密度函数 p 的独立随机变量。设 $\pi_n(\omega) = \prod_{i=1}^{n} p(X_i(\omega))$ 表示我们在前 n 个随机变量中观察到的实现的概率。设 $H = -\sum_{k=1}^{r} p(k)\log p(k)$ 表示 X_1 的熵。那么，对于任意 $\epsilon > 0$，我们有

  $$\mathbf{Pr}\left(e^{-n(H+\epsilon)} < \pi_n(\omega) < e^{-n(H-\epsilon)}\right) \to 1$$

  当 $n \to \infty$。

  这个定理可以通过使用大数定律和切比雪夫不等式来证明。首先，我们注意到 $\log \pi_n(\omega) = \sum_{i=1}^{n} \log p(X_i(\omega))$。由于 X_i 是独立同分布的随机变量，根据大数定律，我们有

  $$\frac{1}{n}\log \pi_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} \log p(X_i(\omega)) \xrightarrow{P} \mathbf{E}[\log p(X_1)] = -H.$$

  因此，$\frac{1}{n}\log \pi_n(\omega) \xrightarrow{P} -H$。

  现在，我们考虑 $\mathbf{Pr}\left(e^{-n(H+\epsilon)} < \pi_n(\omega) < e^{-n(H-\epsilon)}\right)$ 的极限。由于 $\log x$ 是一个连续函数，所以我们有

  $$\mathbf{Pr}\left(e^{-n(H+\epsilon)} < \pi_n(\omega) < e^{-n(H-\epsilon)}\right) = \mathbf{Pr}\left(-n(H+\epsilon) < n\log \pi_n(\omega) < -n(H-\epsilon)\right).$$

  由于 $\frac{1}{n}\log \pi_n(\omega) \xrightarrow{P} -H$，所以我们可以使用切比雪夫不等式来证明上式右边的概率收敛于 1。具体来说，对于任意 $\delta > 0$，我们有

  $$\mathbf{Pr}\left(-n(H+\epsilon) < n\log \pi_n(\omega) < -n(H-\epsilon)\right) = 1 - \mathbf{Pr}\left(|\log \pi_n(\omega) + H| > n\epsilon\right).$$

  由切比雪夫不等式，我们有

  $$\mathbf{Pr}\left(|\log \pi_n(\omega) + H| > n\epsilon\right) \le \frac{\mathrm{Var}(\log \pi_n(\omega))}{n^2\epsilon^2}.$$

  由于 $\mathrm{Var}(\log \pi_n(\omega)) = n\mathrm{Var}(\log p(X_1))$ 是一个常数，所以我们得到

  $$\lim_{n\to\infty} \mathbf{Pr}\left(|\log \pi_n(\omega) + H| > n\epsilon\right) = 0.$$

  因此，

  $$\lim_{n\to\infty} \mathbf{Pr}\left(e^{-n(H+\epsilon)} < \pi_n(\omega) < e^{-n(H-\epsilon)}\right) = 1.$$

- [**Normalized sum**] Let $X_1, X_2, ...$ be i.i.d. random variables with $\mathbf{E}[X_1] = 0$ and $\mathbf{Var}[X_1] = \sigma^2 \in (0, +\infty)$. Show $\frac{\sum_{k=1}^n X_k}{\left(\sum_{k=1}^n X_k^2\right)^{1/2}} \xrightarrow{D} N(0, 1)$ as $n \to \infty$.

  证明：可以通过使用 Lyapunov 中心极限定理来解决。设 $X_1, X_2, ...$ 是独立同分布的随机变量，满足 $\mathbf{E}[X_1] = 0$ 和 $\mathbf{Var}[X_1] = \sigma^2 \in (0, +\infty)$。我们需要证明

  $$\frac{\sum_{k=1}^n X_k}{\left(\sum_{k=1}^n X_k^2\right)^{1/2}} \xrightarrow{D} N(0, 1)$$

  当 $n \to \infty$。

  首先，我们定义 $S_n = \sum_{k=1}^n X_k$ 和 $T_n = \left(\sum_{k=1}^n X_k^2\right)^{1/2}$。由于 X_k 是独立同分布的随机变量，根据中心极限定理，我们有

  $$\frac{S_n}{\sqrt{n}\sigma} \xrightarrow{D} N(0, 1).$$

  因此，$\frac{S_n}{\sqrt{n}\sigma}$ 收敛于一个有限的随机变量。

  接下来，我们考虑 T_n 的极限。由于 $\mathbf{E}[X_1^2] = \sigma^2 < +\infty$，所以我们可以使用大数定律来证明

  $$\frac{T_n^2}{n} = \frac{\sum_{k=1}^n X_k^2}{n} \xrightarrow{P} \mathbf{E}[X_1^2] = \sigma^2.$$

  因此，$\frac{T_n}{\sqrt{n}} \xrightarrow{P} \sigma$。

  现在，我们考虑 $\frac{S_n}{T_n}$ 的极限。由 Slutsky 定理，我们有

  $$\frac{S_n/(\sqrt{n}\sigma)}{T_n/(\sqrt{n}\sigma)} = \frac{S_n}{T_n} \xrightarrow{D} N(0, 1)/1 = N(0, 1).$$

  因此，

  $$\frac{\sum_{k=1}^n X_k}{\left(\sum_{k=1}^n X_k^2\right)^{1/2}} = \frac{S_n}{T_n} \xrightarrow{D} N(0, 1)$$

  当 $n \to \infty$。

Problem 4 (Concentration of measure)

- [**Tossing coins**] We repeatedly toss a fair coin (with an equal probability of heads and tails). Let the random variable X be the number of throws required to obtain a total of n heads. Show that $\mathbf{Pr}[X > 2n + 2\sqrt{n \log n}] \le O(1/n)$.

  证明：我们重复抛掷一枚公平硬币（正面和反面的概率相等）。设随机变量 X 表示获得 n 个正面所需的投掷次数。我们需要证明

  $$\mathbf{Pr}[X > 2n + 2\sqrt{n \log n}] \le O(1/n).$$

  首先，我们定义随机变量 Y_k 表示第 k 次投掷的结果，其中 $Y_k = 1$ 表示正面，$Y_k = 0$ 表示反面。由于硬币是公平的，所以我们有 $\mathbf{E}[Y_k] = 1/2$。

  接下来，我们定义 $S_m = \sum_{k=1}^m Y_k$ 表示前 m 次投掷中正面的次数。由于 X 表示获得 n 个正面所需的投掷次数，所以我们有

  $$\mathbf{Pr}[X > m] = \mathbf{Pr}[S_m < n].$$

  因此，

  $$\mathbf{Pr}[X > 2n + 2\sqrt{n \log n}] = \mathbf{Pr}[S_{2n+2\sqrt{n \log n}} < n].$$

现在，我们考虑上式右边的概率。由 Chernoff 不等式，我们有

$$\mathbf{Pr}[S_m < (1-\delta)\mathbf{E}[S_m]] \leq e^{-\delta^2 \mathbf{E}[S_m]/2}.$$

将 $\delta = 1/2$ 和 $m = 2n + 2\sqrt{n \log n}$ 带入上式，我们得到

$$\mathbf{Pr}[X > 2n + 2\sqrt{n \log n}] = \mathbf{Pr}[S_{2n+2\sqrt{n \log n}} < n] \leq e^{-n/8}.$$

由于 $e^{-n/8} = O(1/n)$，所以我们得到

$$\mathbf{Pr}[X > 2n + 2\sqrt{n \log n}] \leq O(1/n).$$

- [**Chernoff vs Chebyshev**] We have a standard six-sided die. Let Xbe the number of times a 6 occurs in n throws off the die. Compare the best upper bounds on $\mathbf{Pr}[X \geq n/4]$ that you can obtain using Chebyshev's inequality and Chernoff bounds.

  解答：我们有一个标准的六面骰子。设 X 表示在 n 次投掷中出现 6 的次数。我们需要比较使用切比雪夫不等式和 Chernoff 不等式得到的 $\mathbf{Pr}[X \geq n/4]$ 的最佳上界。

  首先，我们定义随机变量 Y_k 表示第 k 次投掷的结果，其中 $Y_k = 1$ 表示投掷结果为 6，$Y_k = 0$ 表示投掷结果不为 6。由于骰子是标准的，所以我们有 $\mathbf{E}[Y_k] = 1/6$。

  接下来，我们定义 $X = \sum_{k=1}^{n} Y_k$ 表示在 n 次投掷中出现 6 的次数。由于 Y_k 是独立同分布的随机变量，所以我们有 $\mathbf{E}[X] = n\mathbf{E}[Y_k] = n/6$。

  现在，我们考虑使用切比雪夫不等式来估计 $\mathbf{Pr}[X \geq n/4]$ 的上界。由切比雪夫不等式，我们有

  $$\mathbf{Pr}[|X - \mathbf{E}[X]| \geq t] \leq \frac{\mathrm{Var}(X)}{t^2}.$$

  将 $t = n/12$ 带入上式，我们得到

  $$\mathbf{Pr}[|X - n/6| \geq n/12] \leq \frac{\mathrm{Var}(X)}{(n/12)^2} = \frac{12\mathrm{Var}(X)}{n^2}.$$

  由于 $\mathrm{Var}(X) = n\mathrm{Var}(Y_k) = n(1/6)(5/6) = 5n/36$，所以我们得到

  $$\mathbf{Pr}[|X - n/6| \geq n/12] \leq \frac{12(5n/36)}{n^2} = \frac{5}{3n}.$$

  因此，使用切比雪夫不等式得到的 $\mathbf{Pr}[X \geq n/4]$ 的最佳上界是 $5/(3n)$。

  接下来，我们考虑使用 Chernoff 不等式来估计 $\mathbf{Pr}[X \geq n/4]$ 的上界。由 Chernoff 不等式，我们有

  $$\mathbf{Pr}[X \geq (1+\delta)\mathbf{E}[X]] \leq e^{-\delta^2 \mathbf{E}[X]/3}.$$

  将 $\delta = 1/2$ 带入上式，我们得到

  $$\mathbf{Pr}[X \geq (3/2)(n/6)] = \mathbf{Pr}[X \geq n/4] \leq e^{-(1/4)(n/6)/3} = e^{-n/72}.$$

  因此，使用 Chernoff 不等式得到的 $\mathbf{Pr}[X \geq n/4]$ 的最佳上界是 $e^{-n/72}$。

  综上所述，使用切比雪夫不等式得到的 $\mathbf{Pr}[X \geq n/4]$ 的最佳上界是 $5/(3n)$，而使用 Chernoff 不等式得到的最佳上界是 $e^{-n/72}$。当 n 很大时，Chernoff 不等式给出的上界更紧。

- [**k-th moment bound]** Let X be a random variable with expectation 0 such that moment generating function $\mathbf{E}[\exp(t|X|)]$ is finite for some $t > 0$. We can use the following two kinds of tail inequalities for X:

  Chernoff Bound

  $$\mathbf{Pr}[|X| \geq \delta] \leq \min_{t \geq 0} \frac{\mathbf{E}[e^{t|X|}]}{e^{t\delta}}$$

kth-Moment Bound

$$\mathbf{Pr}[|X| \geq \delta] \leq \frac{\mathbf{E}[|X|^k]}{\delta^k}$$

1. Show that for each δ, there exists a choice of k such that the kth-moment bound is no weaker than the Chernoff bound. (Hint: Use the probabilistic method.)

   设 X 是一个期望为 0 的随机变量，其矩生成函数 $\mathbf{E}[\exp(t|X|)]$ 在某个 $t > 0$ 处有限。我们可以使用下面两种尾部不等式来估计 X 的尾部概率：

   Chernoff 不等式

   $$\mathbf{Pr}[|X| \geq \delta] \leq \min_{t \geq 0} \frac{\mathbf{E}[e^{t|X|}]}{e^{t\delta}}$$

   k 阶矩不等式

   $$\mathbf{Pr}[|X| \geq \delta] \leq \frac{\mathbf{E}[|X|^k]}{\delta^k}$$

   我们需要证明，对于每个 δ，都存在一个 k，使得 k 阶矩不等式不弱于 Chernoff 不等式。

   为了证明这一点，我们可以使用概率方法。首先，我们注意到

   $$\mathbf{E}[e^{t|X|}] = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbf{E}[|X|^k].$$

   因此，

   $$\min_{t \geq 0} \frac{\mathbf{E}[e^{t|X|}]}{e^{t\delta}} = \min_{t \geq 0} \frac{\sum_{k=0}^{\infty} (t/\delta)^k}{k!} \mathbf{E}[|X|^k] = \sum_{k=0}^{\infty} \frac{\min_{t \geq 0} (t/\delta)^k}{k!} \mathbf{E}[|X|^k].$$

   由于 $\min_{t \geq 0} (t/\delta)^k = 0$ 当 $k = 0$，所以我们得到

   $$\min_{t \geq 0} \frac{\mathbf{E}[e^{t|X|}]}{e^{t\delta}} = \sum_{k=1}^{\infty} \frac{\min_{t \geq 0} (t/\delta)^k}{k!} \mathbf{E}[|X|^k].$$

   现在，我们考虑上式右边的求和式。由于 $\min_{t \geq 0} (t/\delta)^k = 0$ 当 k 是奇数，所以我们得到

   $$\min_{t \geq 0} \frac{\mathbf{E}[e^{t|X|}]}{e^{t\delta}} = \sum_{i=1}^{\infty} \frac{\min_{t \geq 0} (t/\delta)^{2i}}{(2i)!} \mathbf{E}[|X|^{2i}].$$

   现在，我们考虑上式右边的求和式。由于 $\min_{t \geq 0} (t/\delta)^{2i}$ 是一个递减的序列，所以我们得到

   $$\min_{t \geq 0} (t/\delta)^{2i+2} < \min_{t \geq 0} (t/\delta)^{2i}.$$

   因此，

   $$\sum_{i=1}^{\infty} \frac{\min_{t \geq 0} (t/\delta)^{2i}}{(2i)!} < +\infty.$$

   由于上式右边的求和式是有限的，所以存在一个 i_0，使得对于所有 $i > i_0$，都有

   $$\frac{\min_{t \geq 0} (t/\delta)^{2i}}{(2i)!} < 1.$$

   因此，我们得到

   $$\min_{t \geq 0} \frac{\mathbf{E}[e^{t|X|}]}{e^{t\delta}} \leq \sum_{i=1}^{i_0} \frac{\min_{t \geq 0} (t/\delta)^{2i}}{(2i)!} \mathbf{E}[|X|^{2i}] + \mathbf{E}[|X|^{2i_0+2}].$$

   由于 k 阶矩不等式给出的 $\mathbf{Pr}[|X| \geq \delta]$ 的上界是 $\mathbf{E}[|X|^k]/\delta^k$，所以我们得到

   $$\min_{t \geq 0} \frac{\mathbf{E}[e^{t|X|}]}{e^{t\delta}} \leq \sum_{i=1}^{i_0} \frac{\min_{t \geq 0} (t/\delta)^{2i}}{(2i)!} \mathbf{E}[|X|^{2i}] + \frac{\mathbf{E}[|X|^{2i_0+2}]}{\delta^{2i_0+2}} \delta^{2i_0+2}.$$

   因此，对于每个 δ，都存在一个 $k = 2i_0 + 2$，使得 k 阶矩不等式不弱于 Chernoff 不等式。

2. Why would we still prefer the Chernoff bound to the (seemingly) stronger k-th moment bound?

   尽管对于每个 δ，都存在一个 k，使得 k 阶矩不等式不弱于 Chernoff 不等式，但在实际应用中，我们仍然更倾向于使用 Chernoff 不等式。这是因为 Chernoff 不等式通常给出的上界更紧，而且更容易计算。

   首先，Chernoff 不等式通常给出的上界更紧。这是因为 Chernoff 不等式考虑了随机变量的所有矩，而 k 阶矩不等式只考虑了随机变量的一个矩。因此，Chernoff 不等式能够更好地利用随机变量的分布信息来估计尾部概率。

   其次，Chernoff 不等式更容易计算。这是因为 Chernoff 不等式只需要计算随机变量的矩生成函数，而 k 阶矩不等式需要计算随机变量的所有矩。对于许多常见的分布（如二项分布、泊松分布和指数分布），矩生成函数可以直接计算，而计算高阶矩则相对困难。

   综上所述，尽管 k 阶矩不等式在理论上可以比 Chernoff 不等式更强，但在实际应用中，我们仍然更倾向于使用 Chernoff 不等式

- [**Chernoff bound meets graph theory**]
  - Show that with a probability approaching 1 (as n tends to infinity), the Erdős–Rényi random graph $G(n, 1/2)$ has the property that the maximum degree is $(\frac{n}{2} + O(\sqrt{n \log n}))$.

    证明：

    Erdős–Rényi 随机图 $G(n, 1/2)$ 是一个有 n 个顶点的图，其中每条边都以概率 $1/2$ 独立地存在。我们需要证明，当 n 趋近于无穷大时，以概率接近 1，$G(n, 1/2)$ 的最大度数为 $(n/2 + O(\sqrt{n \log n}))$。

    首先，我们注意到对于 $G(n, 1/2)$ 中的任意一个顶点 v，其度数 d_v 是一个二项随机变量，其期望值为 $\mathbf{E}[d_v] = n/2$。由 Chernoff 不等式，我们有

    $$\mathbf{Pr}[|d_v - n/2| \geq t] \leq 2e^{-t^2/n}.$$

    将 $t = \sqrt{n \log n}$ 带入上式，我们得到

    $$\mathbf{Pr}[|d_v - n/2| \geq \sqrt{n \log n}] \leq 2e^{-\log n} = 2/n.$$

    因此，对于任意一个顶点 v，都有 $\mathbf{Pr}[|d_v - n/2| \geq \sqrt{n \log n}] \leq 2/n$。

    由于 $G(n, 1/2)$ 中有 n 个顶点，所以由并集界，我们得到

    $$\mathbf{Pr}[\exists v, |d_v - n/2| \geq \sqrt{n \log n}] \leq 2.$$

    因此，当 n 趋近于无穷大时，以概率接近 1，$G(n, 1/2)$ 的最大度数为 $(n/2 + O(\sqrt{n \log n}))$。

  - Show that with a probability approaching 1 (as n tends to infinity), the Erdős–Rényi random graph $G(n, 1/2)$ has the property that the diameter is exactly 2. The diameter of a graph G is the maximum distance between any pair of vertices.

    证明：Erdős–Rényi 随机图 $G(n, 1/2)$ 是一个有 n 个顶点的图，其中每条边都以概率 $1/2$ 独立地存在。我们需要证明，当 n 趋近于无穷大时，以概率接近 1，$G(n, 1/2)$ 的直径恰好为 2。图 G 的直径是图中任意两个顶点之间的最大距离。

    首先，我们注意到当 n 趋近于无穷大时，以概率接近 1，$G(n, 1/2)$ 是一个连通图。这是因为当 n 趋近于无穷大时，以概率接近 1，$G(n, 1/2)$ 中不存在孤立顶点。

    接下来，我们证明当 n 趋近于无穷大时，以概率接近 1，$G(n, 1/2)$ 的直径恰好为 2。设 u 和 v 是 $G(n, 1/2)$ 中的两个顶点。由于 $G(n, 1/2)$ 是一个连通图，所以存在一条从 u 到 v 的路径。如果这条路径的长度为 1，则说明 u 和 v 之间存在一条边。如果这条路径的长度大于 1，则说明存在一个顶点 w，使得 u 和 w 之间存在一条边，且 w 和 v 之间存在一条边。因此，u 和 v 之间的距离不超过 2。

综上所述，当 n 趋近于无穷大时，以概率接近 1，Erdős–Rényi 随机图 $G(n, 1/2)$ 的直径恰好为 2。