Homework 5

Name: 柯宇斌, ID: 2200013213

Problem 1 (10') . 设随机变量 $X_1, \dots, X_n \sim i.i.d.\pi(1)$, 定义 $S_n = X_1 + \cdots + X_n$. 求证:

$$P(S_n \geq 2n) \leq \frac{1}{n}$$

◀

Answer. 由于 $X_1, \dots, X_n$ 独立同分布，所以有

$$E(S_n) = \sum_{i=1}^{n} E(X_i) = n, D(S_n) = \sum_{i=1}^{n} D(X_i) = n$$

从而

$$P(S_n \geq 2n) = P(S_n - E(S_n) \geq n) \leq P(|S_n - E(S_n)| \geq n) \leq \frac{D(S_n)}{n^2} = \frac{1}{n}$$

◁

Problem 2 (10') . 设独立随机变量列 $\{X_n\}$ 满足: 对于正整数 k, $X_{2k-1} \sim B(3, 0.4)$,$X_{2k} \sim U(-1, 3)$ 请给出常数 c, 使得随机变量 $\frac{1}{2n}\sum_{i=1}^{2n} X_i$ 依概率收敛到 c, 并说明理由. ◀

Answer. 取 $c = 1.1$. 我们定义

$$S_n = \frac{1}{2n}\sum_{i=1}^{n} X_{2i-1}, T_n = \frac{1}{2n}\sum_{i=1}^{n} X_{2i}$$

由独立同分布, 我们有 $E(S_n) = \frac{3}{5}, D(S_n) = \frac{9}{50n}, E(T_n) = \frac{1}{2}, D(T_n) = \frac{1}{3n}$ 则 $\forall \epsilon > 0$, 我们有

$$P(|S_n - E(S_n)| \geq \frac{\epsilon}{2}) \leq \frac{18}{25n\epsilon^2}, P(|T_n - E(T_n)| \geq \frac{\epsilon}{2}) \leq \frac{4}{3n\epsilon^2}$$

而 $c = 1.1 = E(S_n) + E(T_n)$ 所以

$$P(|\frac{1}{2n}\sum_{i=1}^{2n} X_i - c| \geq \epsilon) \leq P(|S_n - E(S_n)| \geq \frac{\epsilon}{2})P(|T_n - E(T_n)| \geq \frac{\epsilon}{2}) \leq \frac{24}{25n^2\epsilon^4}$$

所以 $lim_{n\to+\infty}P(|\frac{1}{2n}\sum_{i=1}^{2n} X_i - c| \leq \epsilon) = 1$

所以依概率收敛到 $c = 1.1$. ◁

Problem 3 (10') . 考虑以下股票价格模型: 初始股票价格为 1, 且每经过一天后, 股票有 p 的概率上涨至原来的 a 倍, 有 $1-p$ 的概率下跌至原来的 b 倍 (即: 若现在股票价格为 r, 则经过一天后, 股票有 p 的概率上涨为 ar, 有 $1-p$ 的概率下跌为 br, 这里 $0 < b < 1 < a$), 假设不同时间股票的上涨和下跌相互独立, 经过 n 天后股票的价格为 X_n.

(1) 试求常数 c, 使得 $\frac{1}{n}\ln X_n$ 几乎必然收敛到 c.

(2) 若 $a = 1.1, b = 0.9, p = 0.5$, 给出两年半 (913 天) 后股票价格不低于初始价格的概率的近似值. ◀

Answer. (1) 由定义 $X_0 = r$. 记 T_i 为第 i 天是否上涨的示性函数，则 $T_i \sim B(1,p)$. 同时有

$$lnX_i = lnX_0 + (\sum_{j=1}^{i} T_j)(\ln a - \ln b) + i \ln b$$

记 $S_n = \frac{1}{n}\ln X_n = \frac{1}{n}(\ln X_0 + (\sum_{i=1}^{n} T_i)(\ln a - \ln b) + n\ln b)$, 有

$$E(S_n) = \frac{lnr}{n} + (p\ln a + (1-p)\ln b)$$

$$D(S_n) = \frac{p(1-p)}{n}(\ln a - \ln b)^2$$

考虑四阶矩 $T(X) = E((X - E(X))^4)$

若 X, Y 独立则

$$\begin{aligned}
G(X+Y) =& E((X+Y-E(X)-E(Y))^4) && (1)\\
=& E((X-E(X))^4) + E(4(X-E(X))^3(Y-E(Y))) + E(6(X-E(X))^3(Y-E(Y))) && (2)\\
&+ E(4(X-E(X))(Y-E(Y))^3) + E((Y-E(Y))^4) && (3)\\
=& G(X) + G(Y) + 6E((X-E(X))^2(Y-E(Y)^2)) && (4)\\
=& G(X) + G(Y) + 6D(X)D(Y) && (5)
\end{aligned}$$

所以有

$$G(\sum_{i=1}^{n} T_i) = \sum_{i=1}^{n} G(T_i) + \sum_{i=1}^{n-1} 6D(\sum_{j=1}^{i} T_j)D(T_{i+1})$$

$$G(\sum_{i=1}^{n} T_i) = np(1-p)(p^3 + (1-p)^3) + 6p^2(1-p)^2\frac{n(n-1)}{2}$$

所以

$$G(S_n) = \frac{(\ln a - \ln b)^4}{n^2}p(1-p)(\frac{(p^3 + (1-p)^3)}{n} + 6p(1-p)\frac{n-1}{2n}) \le \frac{A}{n^2}$$

其中 A 取 $c = (p\ln a + (1-p)\ln b)$, 则

$$P(|S_n - c| \ge \epsilon) \le P(|S_n - E(S_n)| \ge (\epsilon - \frac{\ln r}{n})) \le \frac{G(S_n)}{(\epsilon - \frac{\ln r}{n})^4} \le \frac{A}{n^2(\epsilon - \frac{\ln r}{n})^4}$$

所以 $\forall \epsilon > 0$, 取 N 满足 $\frac{\ln r}{n} \le \frac{\epsilon}{2}$, 则 $\forall n > N$ $\lim_{n\to+\infty}\sum_{j=n}^{+\infty} P(|S_n - c| \ge \epsilon) \le \lim_{n\to+\infty}\sum_{j=n}^{+\infty}\frac{4A}{n^2\epsilon^2} = 0$
所以 S_n 几乎必然收敛到 $c = (p\ln a + (1-p)\ln b)$.

(2) 记 $M_i = T_i\ln 1.1 + (1 - T_i)\ln 0.9$, 由定义,

$$E(M_i = \mu = \frac{1}{2}ln(0.99), D(M_i) = \sigma^2 = (\frac{\ln 0.9 - \ln 1.1}{2})^2$$

所以

$$P(\sum_{i=1}^{n} M_i \ge 0) = 1 - \Phi(\frac{-n\mu}{\sqrt{n}\sigma}) \approx 1 - \Phi(1.5133) = 0.066$$

◁

Problem 4 (10') . 设 $X_1, X_2, \ldots$ 是独立同分布的随机变量序列。

(1) 对 $x \in R$, 定义

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \le x\}}$$

设 X_1 的分布函数是 $F(x)$, 求证：对任意 $x \in R$, $F_n \to^P F(x)$

(2) 设 $f(x)$ 是定义在 $[0,1]$ 上的连续函数，$X_1 \sum U(0,1)$, 求证：

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to^P \int_0^1 f(x)dx$$

◀

Answer. (1) 注意到 $1_{\{X_i \le x\}}$ 的分布为 $\begin{cases} 1, F(x) \\ 0, 1-F(x) \end{cases}$, 所以 $E(1_{\{X_i \le x\}}) = F(x)$, 且独立同分布。所以由辛钦大数定律有,

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \le x\}} \to^P E(1_{\{X_i \le x\}}) = F(x)$$

(2) 因为 $f(x)$ 在 $[0,1]$ 上连续，那么它有界且可积，从而

$$E(f(X_i)) = \int_0^1 f(x)dx$$

有界。所以由辛钦大数定律有,

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to^P \int_0^1 f(x)dx$$

◁

Problem 5 (10') . 给定随机变量序列 $\{X_n\}$, 其中 $X_n \sim B(n, \lambda/n)$.

(1) 求 X_n 的特征函数 $\psi_{X_n}(t)$.

(2) 设 $Y \sim \pi(\lambda)$, 求证：X_n 依分布收敛到 Y. ◀

Answer. (1)

$$\begin{aligned}
\psi_{X_n}(t) =& E(e^{itX}) & (6)\\
=& \sum_{j=1}^{n} \binom{n}{j} (\frac{\lambda}{n})^j (1-\frac{\lambda}{n})^{n-j} e^{itj} & (7)\\
=& \sum_{j=1}^{n} \binom{n}{j} (\frac{\lambda}{n}e^{it})^j (1-\frac{\lambda}{n})^{n-j} & (8)\\
=& (\frac{\lambda}{n}e^{it} + 1 - \frac{\lambda}{n})^n & (9)\\
=& (\frac{\lambda}{n}(e^{it}-1)+1)^n & (10)
\end{aligned}$$

(2) 要证 X_n 依分布收敛到 Y, 只要证 $\{\psi_{X_n}(t)\}$ 收敛于 $\psi_Y(t)$. 而

$$\begin{aligned}\psi_Y(t) =& E(e^{itX}) && (11)\\ =& \sum_{k=0}^{+\infty} e^{-\lambda}\frac{\lambda^k}{k!}e^{itk} && (12)\\ =& \sum_{k=0}^{+\infty} e^{-\lambda}\frac{(\lambda e^{it})^k}{k!} && (13)\\ =& e^{-\lambda}e^{\lambda e^{it}} && (14)\\ =& e^{\lambda(e^{it}-1)} && (15)\end{aligned}$$

所以

$$\begin{aligned}\lim_{n\to+\infty}\psi_{X_n}(t) =& \lim_{n\to+\infty}(\frac{\lambda}{n}(e^{it}-1)+1)^n && (16)\\ =& \lim_{n\to+\infty}(\frac{\lambda}{n}(e^{it}-1)+1)^{\frac{1}{\frac{\lambda}{n}(e^{it}-1)}*n*\frac{\lambda}{n}(e^{it}-1)} && (17)\\ =& \lim_{n\to+\infty}(\frac{\lambda}{n}(e^{it}-1)+1)^{\frac{1}{\frac{\lambda}{n}(e^{it}-1)}*\lambda(e^{it}-1)} && (18)\\ =& e^{\lambda(e^{it}-1)} && (19)\\ =& \psi_Y(t) && (20)\end{aligned}$$

所以 $\{\psi_{X_n}(t)\}$ 收敛于 $\psi_Y(t)$ ◁

Problem 6 (10') . 有两个班级同时上一门课, Y 班有 25 人, L 班有 64 人. 已知两个班期中考试平均成绩均为为 78 分, 标准差为 14 分. 试问:L 班的平均成绩超过 80 分的概率大, 还是 Y 班的平均成绩超过 80 分的概率大? ◀

Answer. 我们认为同一个班的人的成绩独立同分布。

假设一个班有 n 人. 设每个人的成绩 X_i 满足

$$E(X_i)=\mu, D(X_i)=\sigma^2$$

则

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

那么,$\forall\epsilon>0$, 班级平均成绩超过 $\mu+\epsilon$ 的概率为

$$P(\sum_{i=1}^{n} X_i \geq n(\mu+\epsilon)) = 1-\Phi(\frac{n(\mu+\epsilon)-n\mu}{\sqrt{n}\sigma}) = 1-\Phi(\frac{\sqrt{n}\epsilon}{\sigma})$$

关于 n 单调递减，所以 Y 班概率大。 ◁

Problem 7 (10') . 燕园美发对每个顾客的服务时间服从均值为 1/3 小时的指数分布. 假设店里始终只有一个理发师在工作，理发师对每个顾客的服务相互独立，一个理发师不能同时服务两个顾客，且服务之间没有空闲时间，一个客户服务完之后总会有下一个客户需要服务.

(1) 求为 30 个顾客服务，总共需要的时间在 9 到 11 小时之间的概率.

(2) 求最大可能的 n，使得燕园美发有不低于 95% 的把握在 12 小时内服务完 n 个顾客 ◀

Answer. (1) 我们记 T_i 为对每个顾客的服务时间, 则 $\forall t, P(T_i = t) = \lambda e^{-\lambda t}$

$$E(T_i) = \frac{1}{\lambda}, D(T_i) = \frac{1}{\lambda^2}$$

所以

$$\lambda = 3, E(T_i) = \frac{1}{3}, D(T_i) = \frac{1}{9}$$

所以

$$S_n = \sum_{i=1}^{n} T_i \sim N(\frac{n}{3}, \frac{n}{9})$$

$$\frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^{n} T_i \sim N(\frac{1}{3}, \frac{1}{9n})$$

所以

$$P(\frac{3}{10} \leq \frac{S_{30}}{30} \leq \frac{11}{30}) \approx \Phi(\frac{\frac{11}{30} - \frac{1}{3}}{\frac{1}{3\sqrt{30}}}) - \Phi(\frac{\frac{3}{10} - \frac{1}{3}}{\frac{1}{3\sqrt{30}}}) \approx 0.415$$

(2)

$$P(S_n \leq 12) \approx \Phi(\frac{12 - \frac{n}{3}}{\frac{\sqrt{n}}{3}}) \geq 0.95 \Longrightarrow n \leq 27$$

所以 n 最大为 27。 ◁

Problem 8 (10') . 某大学选课系统需要服务 15000 名学生. 假设在选课开始时，每个学生访问选课网的概率是 0.6，且不同学生的访问行为相互独立. 当同时访问选课网的学生数量超出其所能承受的最大访问量时，选课系统就会崩溃. 如果要确保系统有至少 99.9% 的把握在选课开始时不崩溃，那么该系统至少需要承受多少学生的同时访问? ◀

Answer. 记 T_i 为学生 i 是否访问的示性函数, 则 T_i 独立同分布，且

$$E(T_i) = 0.6, D(T_i) = 0.24$$

记 $S_n = \sum_{i=1}^{n} T_i$, 则

$$S_n \sim N(0.6n, 0.24n)$$

而设最大承载量为 M 则有

$$P(S_n \leq M) \approx \Phi(\frac{M - 0.6n}{\sqrt{0.24n}}) \geq 0.999 \Longrightarrow M \geq 9185.8$$

所以至少承受 9186 名学生的访问。 ◁

Problem 9 (10') . 设随机变量列 X_n 相互独立，均服从参数为 $\frac{1}{2}$ 的两点分布，即

$$P(X_n=1)=P(X_n=0)=\frac{1}{2}, n=1,2,\ldots.$$

定义随机变量列

$$Y_n=\sum_{k=1}^{n}\frac{X_k}{2^k},$$

证明 Y_n 依分布收敛到 $[0,1]$ 上的均匀分布 $U(0,1)$ ◀

Answer. 设 $Z_n\sim U(0,1)$, 只需证特征函数依分布收敛。注意到

$$\psi_{Z_n}(t)=E(e^{itZ_n})=\int_0^1 e^{itz}dz=\frac{e^{it}-1}{it}$$

$$\begin{aligned}
\psi_{Y_n}(t)=&E(e^{itY_n}) &(21)\\
=&\prod_{j=1}^{n}E(e^{it\frac{X_j}{2^j}}) &(22)\\
=&\prod_{j=1}^{n}\frac{1}{2}(1+e^{it\frac{1}{2^j}}) &(23)\\
=&\frac{1}{2^n}\frac{1-e^{it}}{1-e^{it\frac{1}{2^n}}} &(24)
\end{aligned}$$

记 $M=\frac{1}{2^n}$, 只要证明

$$\lim_{M\to 0}M\frac{1-e^{it}}{1-e^{itM}}=\frac{e^{it}-1}{it}$$

这由洛必达法则即得。 ◁

Problem 10 (10') . 用概率论方法证明: 当 $n\to\infty$, $e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}\to\frac{1}{2}$

◀

Answer. 先证明泊松分布的再生性

$$\pi(\lambda_1)+\pi(\lambda_2)=\pi(\lambda_1+\lambda_2)$$

假设 $X\sim\pi(\lambda_1), Y\sim\pi(\lambda_2)$ $Z=X+Y$ 则 $k=0,1,2\ldots$, 有

$$\begin{aligned}
P(Z=k)=&\sum_{i=0}^{k}P(X=i)P(Y=k-i) &(25)\\
=&\sum_{i=0}^{k}e^{-\lambda_1}\frac{\lambda_1^i}{i!}e^{-\lambda_2}\frac{\lambda_2^{k-i}}{(k-i)!} &(26)\\
=&e^{-(\lambda_1+\lambda_2)}\sum_{i=0}^{k}\frac{\lambda_1^i}{i!}\frac{\lambda_2^{k-i}}{(k-i)!} &(27)\\
=&\frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{i=0}^{k}\binom{k}{i}\lambda_1^i\lambda_2^{k-i} &(28)\\
=&\frac{e^{-(\lambda_1+\lambda_2)}}{k!}(\lambda_1+\lambda_2)^k &(29)
\end{aligned}$$

所以 $Z \sim \pi(\lambda_1 + \lambda_2)$

下面回到原题假设 $X_i \sim \pi(1)$, 记 $S_n = \sum_{i=1}^{n} X_i \sim \pi(n)$ 而

$$E(X_i) = \mu = 1, D(X_i) = \sigma^2$$

所以 $S_n \sim N(n, n\sigma^2)$ 所以 $\lim_{n\to\infty} P(S_n \leq n) = \frac{1}{2}$ 所以 $\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} \to \frac{1}{2}$ $\triangleleft$