Introduction to Probability

Miscellaneous Problems and Homework

孙雪晖 516030910588

目录

1 Eggs and Boxes

将m个相同的鸡蛋放进n个盒子里，X_i表示第i个盒子里的鸡蛋数目，X_i可以为0。

对于给定任意$t_1, t_n (0 \leq t_1, t_n \leq m, t_1 + t_n \leq m)$，证明

$$P(X_1 \geq t_1 \cap X_n \geq t_n) \leq P(X_i \geq t_1) P(X_n \geq t_n)$$

证明. 由排列组合知识，x个相同的鸡蛋放进y个盒子里，盒子可以为空的总共放法共有：$\binom{x+y-1}{y-1}$ 种。

$$P(X_1 \geq t_1 \cap X_n \geq t_n) = \frac{\binom{m+n-t_1-t_n-1}{n-1}}{\binom{m+n-1}{n-1}} = \frac{\prod_{i=1}^{n-1}(m+n-t_1-t_n-i)}{\prod_{i=1}^{n-1}(m+n-i)}$$

$$P(X_i \geq t_1) P(X_n \geq t_n) = \frac{\binom{m+n-t_1-1}{n-1}}{\binom{m+n-1}{n-1}} \frac{\binom{m+n-t_n-1}{n-1}}{\binom{m+n-1}{n-1}} = \frac{\prod_{i=1}^{n-1}(m+n-t_1-i) \prod_{i=1}^{n-1}(m+n-t_n-i)}{(\prod_{i=1}^{n-1}(m+n-i))^2}$$

我们要证明$P(X_1 \geq t_1 \cap X_n \geq t_n) \leq P(X_i \geq t_1) P(X_n \geq t_n)$，即证

$$\frac{\prod_{i=1}^{n-1}(m+n-t_1-t_n-i)}{\prod_{i=1}^{n-1}(m+n-i)} \leq \frac{\prod_{i=1}^{n-1}(m+n-t_1-i) \prod_{i=1}^{n-1}(m+n-t_n-i)}{(\prod_{i=1}^{n-1}(m+n-i))^2}$$

即

$$\prod_{i=1}^{n-1}(m+n-t_1-t_n-i) \prod_{i=1}^{n-1}(m+n-i) \leq \prod_{i=1}^{n-1}(m+n-t_1-i) \prod_{i=1}^{n-1}(m+n-t_n-i)$$

$\because (m+n-t_1-t_n-i) + (m+n-i) = (m+n-t_1-i) + (m+n-t_n-i)$,

$|(m+n-t_1-t_n-i) - (m-n-i)| \geq |(m+n-t_1-i) - (m+n-t_n-i)|$

$\therefore (m+n-t_1-t_n-i)(m+n-i) \leq (m+n-t_1-i)(m+n-t_n-i)$

$\therefore \prod_{i=1}^{n-1}(m+n-t_1-t_n-i) \prod_{i=1}^{n-1}(m+n-i) \leq \prod_{i=1}^{n-1}(m+n-t_1-i) \prod_{i=1}^{n-1}(m+n-t_n-i)$, 得证

□

2 Prove Weak Law of Large Numbers by Chebyshev Inequality

Let $X_1, X_2, ..., X_n$ be an independent trials process with $\mu_i = E(X_i)$ and $\sigma_i^2 = Var(X_i)$. Suppose they are uniformly boundness, let $S_n = \sum_{i=1}^{n} X_i$, prove $\forall \epsilon > 0, P(\frac{S_n}{n} - E(\frac{S_n}{n}) < \epsilon) \to 1$ as $n \to \infty$.

证明.

$$Var(S_n) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} \sigma_i^2$$

令$M = max_{i=1}^{n} \sigma_i^2$, 则

$$Var(\frac{S_n}{n}) = \frac{\sum_{i=1}^{n} \sigma_i^2}{n^2} \leq \frac{nM}{n^2} = \frac{M}{n}$$

$$\because E(\frac{S_n}{n}) = \frac{\sum_{i=1}^{n} \mu_i}{n}$$

由切比雪夫不等式得，

$$P(\frac{S_n}{n} - E(\frac{S_n}{n}) \geq \epsilon) \leq \frac{Var(\frac{S_n}{n})}{\epsilon^2} = \frac{M}{n\epsilon^2}$$

$$\because \frac{M}{n\epsilon^2} \to 0(n \to \infty)$$

$$\therefore P(\frac{S_n}{n} - E(\frac{S_n}{n}) \geq \epsilon) \to 1(n \to \infty)$$

$$\therefore P(\frac{S_n}{n} - E(\frac{S_n}{n}) < \epsilon) \to 0(n \to \infty)$$

□

3 Hardy-Weinberg Law

哈迪-温伯格定律：理想状况下，种群内各等位基因的基因频率在遗传过程中稳定不变。
比如：假设$P(A) = p, P(a) = q = 1 - p$，下一代中：

$$P(A) = P(AA) + \frac{1}{2}P(Aa) = P(A)^2 + \frac{1}{2} \cdot 2P(A)P(a) = p^2 + pq = p^2 + p(1-p) = p$$

$$P(a) = P(aa) + \frac{1}{2}P(Aa) = P(a)^2 + \frac{1}{2} \cdot 2P(A)P(a) = q^2 + pq = q^2 + q(1-q) = q$$

故基因频率不变。当有多个等位基因的时候，该定律是否还满足？

依然满足。

证明. 考虑n个等位基因$A_1, A_2, ..., A_n$，设$P(A_i) = p_i$，若稳定遗传，下一代中：

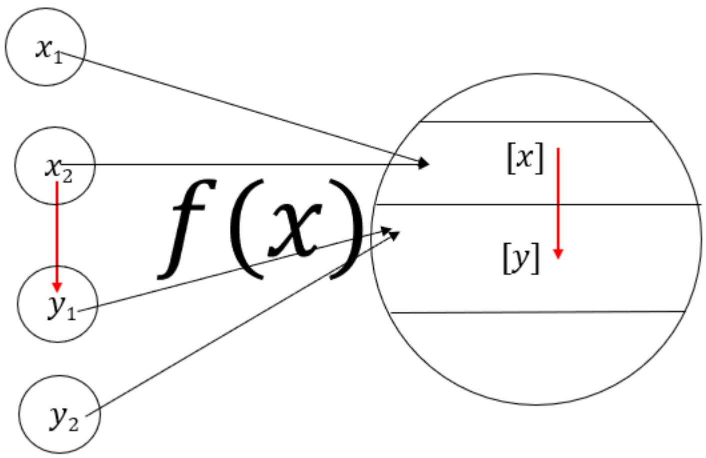
$$P(A_i) = P(A_iA_i) + \frac{1}{2}\sum_{j=i+1}^{n} P(A_iA_j) = p_i^2 + \frac{1}{2}\sum_{j\neq i} p_ip_j = p_i^2 + \frac{1}{2}p_i\sum_{j\neq i} p_j = p_i^2 + \frac{1}{2}\cdot p_i(1-p_i) = p_i$$

$\therefore$基因频率不变

□

4 Markov Chain

若$(x_0, x_1, ...)$是马尔可夫链，对于什么样的函数f，$(f(x_0), f(x_1), ...)$是马尔可夫链？



设$f : \Omega \mapsto \Omega^{\#}$，在函数$f$的值域空间$\Omega^{\#}$中，$f$将所有$x_i$映射到了一个等价类$[x_i]$。

定义$P(x, A) = \sum_{y\in A} P(x, y)$，则$\forall x, x', [x] = [x'] \Rightarrow P(x, [y]) = P(x', [y])$

定义$P^{\#}([x], [y])$为$(f(x_0), f(x_1), ...)$中等价类$[x]$到$[y]$的转移概率，则如果满足$P^{\#}([x], [y]) = P(x, [y])$，则在$(f(x_0), f(x_1), ...)$中，各等级类之间的转移依然满足无记忆性，因此$(f(x_0), f(x_1), ...)$是马尔可夫链。

$\therefore$满足$P^{\#}([x], [y]) = P(x, [y])$时，$(f(x_0), f(x_1), ...)$是马尔可夫链。

5 General Normal Distribution

$X \sim \mathcal{N}(0, 1), Y = \sigma X + \mu, \sigma > 0$，求随机变量$Y$的密度分布函数$f_Y(x)$。（标准正态分布→一般正态分布）

$$f_X(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

首先计算随机变量X与Y累计分布函数的关系：

$$\begin{aligned}F_Y(x) &= P(Y \le x)\\ &= P(\sigma X + \mu \le x)\\ &= P(X \le \frac{x-\mu}{\sigma})\\ &= F_X(\frac{x-\mu}{\sigma})\end{aligned}$$

求导得到概率密度函数的关系：

$$\begin{aligned}f_Y(x) &= \frac{dF_Y(x)}{dx}\\ &= \frac{1}{\sigma} f_X(\frac{x-\mu}{\sigma})\\ &= \frac{1}{\sigma}\frac{e^{-\frac{\frac{(x-\mu)^2}{\sigma^2}}{2}}}{\sqrt{2\pi}}\\ &= \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\end{aligned}$$

6 Sum of Normal Distributions

$X \sim \mathcal{N}(\mu_1, \sigma_1^2), Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$，$X, Y$相互独立，证明$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

证明.

$$\begin{aligned}f_{X+Y}(y) &= P(X + Y = y)\\ &= \sum_x P(X = x, Y = y - x)\\ &= \sum_x f_X(x) f_Y(y-x)\\ &= \int_{-\infty}^{\infty} f_X(x) f_Y(y-x) dx\\ &= \int_{-\infty}^{\infty} \frac{e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}}{\sqrt{2\pi}\sigma_1}\frac{e^{-\frac{(y-x-\mu_2)^2}{2\sigma_2^2}}}{\sqrt{2\pi}\sigma_2} dx\end{aligned}$$

经过上面的推导，$X + Y$概率密度函数被表示为X与Y的概率密度函数的卷积形式，接下来进行积分运算：

$$\begin{aligned}f_{X+Y}(y) &= \int_{-\infty}^{\infty} \frac{exp\Big[-\frac{(x-\mu_1)^2}{2\sigma_1^2}\Big]}{\sqrt{2\pi}\sigma_1}\frac{exp\Big[-\frac{(y-x-\mu_2)^2}{2\sigma_2^2}\Big]}{\sqrt{2\pi}\sigma_2} dx\\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{2\pi}\sigma_1\sigma_2} exp\Big[-\frac{\sigma_1^2(y-x-\mu_2)^2 + \sigma_2^2(x-\mu_1)^2}{2\sigma_1^2\sigma_2^2}\Big] dx\\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{2\pi}\sigma_1\sigma_2} exp\Big[-\frac{\sigma_1^2(y^2 + x^2 + \mu_2^2 - 2xy - 2x\mu_2 - 2y\mu_2) + \sigma_2^2(x^2 + \mu_1^2 - 2x\mu_1)}{2\sigma_1^2\sigma_2^2}\Big] dx\\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{2\pi}\sigma_1\sigma_2} exp\Big[-\frac{x^2(\sigma_1^2 + \sigma_2^2) - 2x(\sigma_1^2(y-\mu_2) + \sigma_2^2\mu_1) + \sigma_1^2(y^2 + \mu_2^2 - 2y\mu_2) + \sigma_2^2\mu_1^2}{2\sigma_1^2\sigma_2^2}\Big] dx\end{aligned}$$

换元，令$\sigma=\sqrt{\sigma_1^2+\sigma_2^2}$，则：

$$
\begin{aligned}
f_{X+Y}(y)&=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\frac{1}{\sqrt{2\pi}\frac{\sigma_1\sigma_2}{\sigma}}exp\Big[-\frac{x^2-2x\frac{\sigma_1^2(y-\mu_2)+\sigma_2^2\mu_1}{\sigma^2}+\frac{\sigma_1^2(y^2+\mu_2^2-2y\mu_2)+\sigma_2^2\mu_1^2}{\sigma^2}}{2(\frac{\sigma_1\sigma_2}{\sigma})^2}\Big]dx\\
&=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\frac{1}{\sqrt{2\pi}\frac{\sigma_1\sigma_2}{\sigma}}exp\Big[-\frac{(x-\frac{\sigma_1^2(y-\mu_2)+\sigma_2^2\mu_1}{\sigma^2})^2-(\frac{\sigma_1^2(y-\mu_2)+\sigma_2^2\mu_1}{\sigma^2})^2+\frac{\sigma_1^2(y-\mu_2)^2+\sigma_2^2\mu_1^2}{\sigma^2}}{2(\frac{\sigma_1\sigma_2}{\sigma})^2}\Big]dx\\
&=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}exp\Big[-\frac{\sigma^2(\sigma_1^2(y-\mu_2)^2+\sigma_2^2\mu_1^2)-(\sigma_1^2(y-\mu_2)+\sigma_2\mu_1)^2}{2\sigma^2(\sigma_1\sigma_2)^2}\Big]\\
&\cdot\frac{1}{\sqrt{2\pi}\frac{\sigma_1\sigma_2}{\sigma}}exp\Big[-\frac{(x-\frac{\sigma_1^2(y-\mu_2)+\sigma_2^2\mu_1}{\sigma^2})^2}{2(\frac{\sigma_1\sigma_2}{\sigma})^2}\Big]dx\\
&=\frac{1}{\sqrt{2\pi}\sigma}exp\Big[-\frac{(y-(\mu_1+\mu_2))^2}{2\sigma^2}\Big]\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\frac{\sigma_1\sigma_2}{\sigma}}exp\Big[-\frac{(x-\frac{\sigma_1^2(y-\mu_2)+\sigma_2^2\mu_1}{\sigma^2})^2}{2(\frac{\sigma_1\sigma_2}{\sigma})^2}\Big]dx
\end{aligned}
$$

显然积分项是一个$\mathcal{N}(\frac{\sigma_1^2(y-\mu_2)+\sigma_2^2\mu_1}{\sigma^2},(\frac{\sigma_1\sigma_2}{\sigma})^2)$的正态分布概率密度函数，由归一化条件，积分值为1。因此

$$
f_{X+Y}(y)=\frac{1}{\sqrt{2\pi}\sigma}exp\Big[-\frac{(y-(\mu_1+\mu_2))^2}{2\sigma^2}\Big]
$$

因此$X+Y\sim\mathcal{N}(\mu_1+\mu_2,\sigma^2)$，即$X+Y\sim\mathcal{N}(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$

□

7 Gamma Distribution and Poisson Distribution

$X \sim Gamma(\alpha, \beta), Y \sim Poisson(x\beta)$，证明$P(X \le x) = P(Y \ge \alpha)$.

证明.

$$P(Y \ge \alpha) = 1 - P(Y \le \alpha - 1) = 1 - e^{-x\beta} \sum_{k=0}^{\alpha-1} \frac{(x\beta)^k}{k!}$$

令$\Gamma(a, b)$表示Gamma函数从b到无穷的部分，即$\Gamma(a, b) = \int_b^\infty t^{a-1}e^{-t}dt$，由Gamma函数性质，对于正整数$n$：

$$\Gamma(n, b) = (n-1)! e^{-b} \sum_{k=0}^{n-1} \frac{b^k}{k!}$$

$$\Gamma(\alpha, x\beta) = (\alpha - 1)! e^{-x\beta} \sum_{k=0}^{\alpha-1} \frac{(x\beta)^k}{k!}$$

由于$\Gamma(\alpha) = (\alpha - 1)!$，所以

$$\begin{aligned}
P(Y \ge \alpha) &= 1 - \frac{\Gamma(\alpha, x\beta)}{\Gamma(\alpha)} \\
&= \frac{\Gamma(\alpha) - \Gamma(\alpha, x\beta)}{\Gamma(\alpha)} \\
&= \frac{1}{\Gamma(\alpha)} (\int_0^\infty t^{\alpha-1}e^{-t}dt - \int_{x\beta}^\infty t^{\alpha-1}e^{-t}dt) \\
&= \frac{1}{\Gamma(\alpha)} \int_0^{x\beta} t^{\alpha-1}e^{-t}dt \\
&= \frac{1}{\Gamma(\alpha)} \int_0^{x} (\beta t)^{\alpha-1}e^{-\beta t} \cdot \beta dt \\
&= \int_0^x \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1}e^{-\beta t}dt \\
&= \int_0^x f_X(t)dt \\
&= P(X \le x)
\end{aligned}$$

□

8 Gambler's Ruin

一个赌徒与赌场赌钱，一开始赌场有A元钱，赌徒有B元钱。每次抛一枚硬币，等概率会获得正面或反面。如果是正面，赌徒从赌场赢得一元钱，否则赌徒输给赌场一元钱。当赌徒或赌场某一方将所有钱输光时，游戏结束。求赌徒破产的概率。

设X_i为抛i次硬币后赌徒总共赢得的钱数，如果是输钱则X_i为负。$X_0 = 0$，第i次抛硬币后赢得的钱数只取决于第$i-1$次抛硬币后赢得的钱数，因此是一个马尔可夫链模型。

设p_i为赌徒有i元钱，赌场有$A+B-i$元，最终破产的概率。

$$\begin{cases} p_0 = 1 \\ p_i = \frac{1}{2}p_{i-1} + \frac{1}{2}p_{i+1}, 0 < i < A+B \\ p_{A+B} = 0 \end{cases}$$

由方程条件可知，$p_0, p_1, ..., p_{A+B}$构成等差数列，因此$p_i = 1 - \frac{i}{A+B}$
赌徒最终破产概率$p_B = 1 - \frac{B}{A+B} = \frac{A}{A+B}$

如果硬币正反面概率不相等？

设正面概率为p，与上面类似可以列出方程

$$\begin{cases} p_0 = 1 \\ p_i = pp_{i-1} + (1-p)p_{i+1}, 0 < i < A+B \\ p_{A+B} = 0 \end{cases}$$

逐项相减可得

$$p_i - p_{i-1} = \frac{p-1}{p}(p_{i-1} - p_{i-2})$$

因此方程的解为

$$p_i = 1 - \frac{1 - (\frac{1-p}{p})^i}{1 - (\frac{1-p}{p})^{A+B}}$$

$$p_B = \frac{1 - (\frac{1-p}{p})^B}{1 - (\frac{1-p}{p})^{A+B}}$$

经过程序模拟，一开始令赌徒的钱数B为100，赌场的钱数A从0开始不断增加，每次进行1000次重复试验，记录赌徒破产的次数，除以1000得到近似的破产概率。模拟结果发现，在硬币正面概率为0.47时，赌场最初只需要大约30元钱，就可以使赌徒破产的概率接近百分之百。