概率论复习-作业

#SP

1-5 多项分布

（a）由于所有事件出现的排列顺序可以是不定的，因此 $X_1, X_2, \cdots, X_r$ 的联合分布为：

$$P(X_1 = N_1, X_2 = N_2, \cdots, X_r = N_r; p_1, p_2, \cdots, p_r) = \frac{(\sum_i N_i)!}{\prod_i N_i!} \prod_i p_i^{N_i}$$

这被称为多项分布。
（b）在多项分布中，每个变量实际上都服从二项分布。因此，每个变量的方差都与二项分布的方差相同：

$$\mathrm{Var}(X_i) = \mathrm{Cov}(X_i, X_i) = np_i(1 - p_i)$$

而在 $i \neq j$ 时:

$$\mathrm{Cov}(X_i, X_j) = \mathrm{E}(X_i X_j) - \mathrm{E}(X_i)\mathrm{E}(X_j)$$

那么我们只需要求解 $\mathrm{E}(X_i X_j)$。为了求解这个期望，使用条件期望的结论，取条件于其中一个变量：

$$\mathrm{E}(X_i X_j) = \mathrm{E}[\mathrm{E}(X_i X_j | X_j = N_j)] = \mathrm{E}[X_j E(X_i | X_j = N_j)]$$

考虑这一取条件的意义：实际上，相当于我已经固定了 N_j 次独立重复实验，规定这些实验中只能出现第 j 个结果。因此，在取完条件后，我们可以认为剩余的所有随机变量 $X_1, X_2, \cdots, X_{j-1}, X_{j+1}, \cdots, X_r$ 服从新的多项分布，且这一多项分布的参数为 $\frac{p_1}{1-p_j}, \frac{p_2}{1-p_j}, \cdots, \frac{p_r}{1-p_j}$。那么

$$E[X_i | X_j = N_j] = \frac{p_i}{1 - p_j}(n - N_j)$$

从而继续计算

$$\begin{aligned}
&E(X_i X_j) \\
&\quad = E\left[\frac{p_i}{1 - p_j}(n - N_j)N_j\right] \\
&\quad = \frac{p_i}{1 - p_j} E[nN_j - N_j^2] \\
&\quad = \frac{p_i}{1 - p_j}(nE[N_j] - E[N_j^2]) \\
&\quad = \frac{p_i}{1 - p_j}(nE[N_j] - \mathrm{Var}(N_j) - E[N_j]^2) \\
&\quad = \frac{p_i}{1 - p_j}(n^2 p_j - n(1 - p_j)p_j - n^2 p_j^2) \\
&\quad = (n^2 - n)p_i p_j
\end{aligned}$$

因此：

$$\mathrm{Cov}(X_i, X_j) = (n^2 - n)p_i p_j - n^2 p_i p_j = -np_i p_j$$

（c）最后，我们再来计算未出现的结果数的均值和方差。选取指示变量 I_i，如果结果 i 不出现，则 $I_i = 1$，否则 $I_i = 0$。显然，在 n 次独立重复实验中，结果 i 不出现的概率是 $(1 - p_i)^n$，也就是说：

$$P(I_i = 1) = (1 - p_i)^n \Rightarrow E[I_i] = (1 - p_i)^n$$

那么，不出现的事件的数目之和的期望：

$$E=\sum_i E[I_i]=\sum_i(1-p_i)^n$$

由于变量 I_i 服从两点分布，因此其方差为：

$$\text{Var}[I_i]=(1-p_i)^n(1-(1-p_i)^n)$$

现在考虑 $\text{Cov}(I_i,I_j)$：

$$\text{Cov}(I_i,I_j)=E[I_iI_j]-E[I_i]E[I_j]$$

这里要想使得 $I_iI_j=1$，则必须 $I_i=1$ 且 $I_j=1$。也就是说，我们要求结果 i 和 j 都没有出现过。那么

$$E[I_iI_j]=P(I_iI_j=1)=(1-p_i-p_j)^n$$

因此：

$$\text{Cov}(I_i,I_j)=(1-p_i-p_j)^n-(1-p_i)^n(1-p_j)^n$$

不出现事件的数目的方差为：

$$\text{VAR}=\sum_i(1-p_i)^n(1-(1-p_i)^n)+\sum_{i\neq j}((1-p_i-p_j)^n-(1-p_i)^n(1-p_j)^n)$$

1-39 质点移动

使用 T_i 表征从第 $i-1$ 点走向第 i 点所需步数的期望值，那么，在 $i>1$ 的情况下，考虑质点从第 $i-1$ 个点走向第 i 个点的过程：质点可能从第 $i-1$ 点直接向右走一步，达到第 i 个点；也有可能从第 $i-1$ 个点向左走一步，到达第 $i-2$ 个点，那么质点就需要从第 $i-2$ 点线返回到第 $i-1$ 点，再返回到第 i 点，在这种情况下，质点需要行走的步数就是 $1+T_{i-1}+T_i$ 步。那么，根据全期望公式：

$$T_i=1+\frac{T_{i-1}+T_i}{2}$$

化简这个递推方程，得到：

$$T_i=2+T_{i-1}$$

利用边界条件 $T_0=0$，$T_1=1$，我们容易得到：

$$T_i=2i-1$$

那么，质点走到第 i 个节点所需的总步数就是：

$$E_i=\sum_i T_i=i^2$$

命题得证。

1-37 峰值

我们给出如下的论证：因为 $X_1,X_2,\cdots,X_n$ 是 iid 的，那么我们从中任选三个相邻的随机变量 X_{i-1},X_i,X_{i+1}，X_i 成为“峰值”的概率必然为 $1/3$。也就是说，现在共有 n 个随机变量 $X_1,X_2,\cdots,X_n$,设其中出现的“峰值”数目为 m 个。那么

$$E\left[\frac{n}{m}\right]=\frac{1}{3}$$

那么由大数定律我们知道，“峰值”出现的时间比例必然依概率 1 收敛到 $1/3$

1-34 失效率

利用条件概率的定义：

$$P(X_1 < X_2 | \min(X_1, X_2) = t) = \frac{P(X_1 < X_2, \min(X_1, X_2) = t)}{P(\min(X_1, X_2) = t)}$$

将分子和分母展开：

$$= \frac{P(X_1 = t, X_2 > X_1)}{P(X_1 = t, X_2 > X_1) + P(X_2 = t, X_1 > X_2)}$$

利用 X_1 和 X_2 之间的独立性：

$$= \frac{P(X_1 = t)P(X_2 > t)}{P(X_1 = t)P(X_2 > t) + P(X_2 = t)P(X_1 > t)}$$

利用失效率的定义，注意到关系：

$$P(X_1 = t) = P(X_1 > t)\lambda_1(t) \qquad P(X_2 = t) = P(X_2 > t)\lambda_2(t)$$

代入上式得到：

$$= \frac{P(X_1 > t)\lambda_1(t)P(X_2 > t)\lambda_2(t)}{P(X_1 > t)\lambda_1(t)P(X_2 > t) + P(X_2 > t)\lambda_2(t)P(X_1 > t)} = \frac{\lambda_1(t)}{\lambda_1(t) + \lambda_2(t)}$$

从而得证。

1-29 Gamma 分布

要证明 $\sum_{i=1}^n X_i$ 具有 Gamma 分布，只要证明它与 Gamma 分布有相同的矩母函数。指数分布的概率密度函数为

$$f(x) = \lambda \exp(-\lambda x)$$

那么，其矩母函数为：

$$M_X(t) = E[e^{tX}] = \int_0^{+\infty} \lambda \exp(-\lambda x) \exp(xt) dx = \int_0^{+\infty} \lambda \exp((t - \lambda)x) dx = \frac{\lambda}{\lambda - t}$$

由于 X_i 是独立同分布的，所以：

$$M_{X_1+X_2+\cdots X_n}(t) = E[\exp(t(X_1 + X_2 + \cdots X_n))] = E[te^{X_1}]E[te^{X_2}]\cdots E[te^{X_n}] = \left(\frac{\lambda}{\lambda - t}\right)^n$$

而 gamma 分布的矩母函数：

$$M_Y(t) = E[e^{tX}] = \int_0^{+\infty} \frac{\lambda \exp(-\lambda x)(\lambda x)^{n-1}}{(n-1)!} \exp(tx) dx = \left(\frac{\lambda}{\lambda - t}\right)^n$$

因此，原命题得证。

1-20 连续的填充问题

首先，如果区间的长度小于 1，那么我们无法选择一个单位区间。因此在 $x < 1$ 时，$M(x) = 0$。
在 $x > 1$ 时，我们取条件于第一次选择的单位区间的左端点。如果第一次选择的单位区间左端点为 y，那么我们在第二次选择时，相当于既要在单位区间的左侧，长度为 y 的区间内选择新的单位区间，又要在单位区间右侧，长度为 $x - y - 1$ 的区间内选择新的单位区间。由于 y 是在 $(0, x-1)$ 间均匀分布的，因此，由全期望公式：

$$
\begin{aligned}
&M(x)\\
&= E[M(x)|\text{the first select is (y,y+1)}]\\
&= \frac{1}{x-1}\int_0^{x-1}(M(y)+M(x-y-1))dy\\
&= \frac{1}{x-1}\left[\int_0^{x-1}M(y)dy+\int_{x-1}^{0}M(t)(-dt)\right]\\
&= \frac{2}{x-1}\int_0^{x-1}M(y)dy
\end{aligned}
$$

因此原命题得证。

1-17 次序统计量

对于 (a)，在 n 个随机变量中有 i 个小于 x 分两种情况：要么第 n 个随机变量小于 x，剩下 $n-1$ 个随机变量中还有 $i-1$ 个小于 x；要么第 n 个随机变量大于 x，剩下 $n-1$ 个随机变量中有 i 个小于 x。那么，根据以上论述得到：

$$F_{i,n}(x)=F(x)F_{i-1,n-1}(x)+\bar{F}(x)F_{i,n-1}(x)$$

对于 (b)，在 $n-1$ 个随机变量中有 i 个小于 x 也有两种情况：要么第 n 个随机变量位于 $X_1,X_2,\cdots,X_n$ 的前 i 个最小者中，那么所有的 n 个随机变量中总共有 $i+1$ 个小于 x；要么第 n 个随机变量不在 $X_1,X_2,\cdots,X_n$ 的前 i 个最小者中，那么所有的 n 个随机变量中总共只有 i 个小于 x。根据以上论述直接写出：

$$F_{i,n-1}(x)=\frac{i}{n}F_{i+1,n}(x)+\frac{n-i}{n}F_{i,n}(x)$$

1-22 条件方差

按照定义，X 的方差写为：

$$
\begin{aligned}
&\mathrm{Var}(X)\\
&= E[X^2]-E[X]^2\\
&= E[E[X^2|Y]]-E[E[X|Y]]^2
\end{aligned}
$$

将条件方差变形：

$$
\begin{aligned}
&\mathrm{Var}(X|Y)\\
&= E[(X-E[X|Y])^2|Y]\\
&= E[X^2-2XE[X|Y]+E[X|Y]^2|Y]\\
&= E[X^2|Y]-2E[XE[X|Y]|Y]+E[E[X|Y]^2|Y]\\
&= E[X^2|Y]-2E[X|Y]^2+E[X|Y]^2\\
&= E[X^2|Y]-E[X|Y]^2
\end{aligned}
$$

将条件方差的化简结果代入 $\mathrm{Var}(X)$，得到：

$$
\begin{aligned}
&\mathrm{Var}(X)\\
&= E[\mathrm{Var}(X|Y)+E[X|Y]^2]-E[E[X|Y]]^2\\
&= E[\mathrm{Var}(X|Y)]+E[E[X|Y]^2]-E[E[X|Y]]^2\\
&= E[\mathrm{Var}(X|Y)]+\mathrm{Var}(E[X|Y])
\end{aligned}
$$

原命题得证。
算例：用 X 表示矿工到达安全地的时间，Y 表示矿工选取的门，那么：

$$\mathrm{Var}(X)=\frac{1}{3}\mathrm{Var}(X|Y=1)+\frac{1}{3}\mathrm{Var}(X|Y=2)+\frac{1}{3}\mathrm{Var}(X|Y=3)+\mathrm{Var}(E[X|Y])$$

下面计算 $\mathrm{Var}(E[X|Y])$。在 $Y=1$ 时，$E[X|Y]=2$；在 $Y=2$ 时，$E[X|Y]=10+3=13$；在 $Y=3$ 时，$E[X|Y]=10+5=15$，那么 $\mathrm{Var}(E[X|Y])=98/3$。又考虑到 $\mathrm{Var}(X|Y=1)=0$（因为此时确定经过 2 小时可以走出矿井），$\mathrm{Var}(X|Y=2)=\mathrm{Var}(X|Y=3)=\mathrm{Var}(X)$（在走完一段时间确定的路程后，问题回到初始的问题），因此可以得到方程：

$$\frac{1}{3}\mathrm{Var}(X)=\frac{98}{3}$$

求矩母函数 $M(t)$ 的导数得到：在 $t=0$，$M'(t)=10, M''(t)=198$，那么，$\mathrm{Var}(X)=98$，验证了我们的结果。

1-11 整数变量的概率生成函数

(a)直接对生成函数求导，得到：

$$\frac{d^k}{dz^k}P(z)=k!P(X=k)+(k+1)!zP(X=k+1)+\frac{(k+2)!}{2!}z^2P(X=k+2)+\cdots$$

代入 $z=0$，得到：

$$\frac{d^k}{dz^k}P(z)|_{z=0}=k!P(X=k)$$

(b) 注意到

$$P(1)=\sum_{i=0}^{\infty}P(X=i)$$

而

$$P(-1)=\sum_{i=0}^{\infty}(-1)^iP(X=i)$$

那么

$$\frac{P(1)+P(-1)}{2}=P(X=0)+P(X=2)+P(X=4)+\cdots P(X=2k)$$

命题得证

(c) 若 X 服从二项分布，则 X 的概率生成函数为：

$$P(z)=\sum_{i=0}^{n}z^iC_n^ip^i(1-p)^{n-i}$$

在 $z=1$ 时，

$$P(z)=\sum_{i=0}^{n}C_n^ip^i(1-p)^{n-i}=(p+(1-p))^n=1^n=1$$

在 $z=-1$ 时，

$$P(z)=\sum_{i=0}^{n}C_n^i(-p)^i(1-p)^{n-i}=(1-p-p)^n$$

那么，利用 (b)中我们获得的结论：

$$P(X=2k,\forall k)=\frac{1+(1-2p)^n}{2}$$

(d) 若 X 服从泊松分布，则 X 的概率生成函数为：

$$P(z)=\sum_{i=0}^{\infty}\frac{z^i\lambda^i exp(-\lambda)}{i!}=\exp(z\lambda)\exp(-\lambda)$$

因此，$P(1)=1, P(-1)=\exp(-2\lambda)$，从而命题得证

(e)若 X 服从几何分布，那么概率生成函数为：

$$P(z)=\sum_{i=1}^{\infty}z^i(1-p)^{i-1}p=zp\sum_{i=1}^{\infty}[z(1-p)]^{i-1}$$

根据等比数列求和公式，$P(1)=1, P(-1)=-p/(2-p)$ ，利用上面的结论即得：

$$P(X=2k,\forall k)=\frac{1-p}{2-p}$$

(f) 若 X 服从负二项分布，那么概率生成函数为：

$$P(z)=\sum_{i=r}^{\infty}z^iC_{r-1}^{i-1}p^i(1-p)^{i-r}$$

显然 $P(1)=1$，而 $P(-1)=(1-)^r(\frac{p}{2-p})^r$ ，利用 (b)中结论即可证明。

1-6 记录值

(a)由于 $X_1,X_2,\cdots,X_n$ 是独立同分布的，因此在 X_i 加入这个随机变量序列时，它成为序列中最大的随机变量的概率为 $1/i$ 。因此，产生记录值的总个数为：

$$E[N_n]=\sum_{i=1}^{n}\frac{1}{i}$$

使用指示变量 I_i 来代表第 i 时刻是否产生了记录值，那么 I_i 应当服从二项分布。又由于 X_i 能否成为记录值只与 $\max(X_1,X_2,\cdots,X_{i-1})$ 有关，而与 $X_1,X_2,\cdots,X_{i-1}$ 等各个变量独立。因此：

$$\mathrm{Var}(N_n)=\sum_{i=1}^{n}\mathrm{Var}(I_n)=\sum_{i=1}^{n}\frac{1}{i}(1-\frac{1}{i})$$

(b) $P(T>n)$ 说明仅仅在 X_1 时产生了一个记录值，之后都没有产生记录值，那么 X_1 是 $X_1,X_2,\cdots,X_n$ 中的极大值，从而

$$P(T>n)=\frac{1}{n}$$

那么显然：

$$P(T<n+1)=1-\frac{1}{n}\quad\Rightarrow\quad P(T<+\infty)=1$$

由于

$$P(T=i)=\frac{1}{i}$$

所以

$$\mathrm{E}(T)=\sum_{i=2}^{\infty}P(T=i)i=\sum_{i=2}^{\infty}1\to\infty$$

(c)

不确定这个证法对不对

将 $X_1,X_2,\cdots,X_n$ 的分布记为 $G(x)$，则首个大于 y 的记录值出现的时刻服从指数分布。也就是说：

$$P(T_y=n)=(G(y))^n(1-G(y))$$

而 T_y 和 X_{T_y} 的联合分布：

$$P(T_y=n,X_{T_y}=x_0)=P(X_1<y,X_2<y,\cdots,X_{T_y}>y,X_{T_y}=x_0)=P(X_1<y)P(X_2<y)\cdots P(X_{T_y}>y,X_{T_y}=x_0)=(G(y))^n$$

而我们已经知道 $X_{T_y} > y$ （实际上，我们知道了 $T_y = n$ 的信息之后，就已经改变了 X_{T_y} 的分布，因此这里在求 X_{T_y} 的分布时，就要利用这一条件

$$P(X_{T_y} = x_0) = \frac{g(x_0)}{1 - G(y)}$$

那么注意到

$$P(T_y = n)P(X_{T_y} = x_0) = P(T_y = n, X_{T_y} = x_0)$$

这说明了二者独立。

答案的证法

$$P(X_{T_y} > x | T_y = n) = P(X_n > x | X_1 < y, X_2 < y, \cdots X_n > y)$$

由于 $X_1, X_2, \cdots, X_n$ 都是 iid 的随机变量，所以条件里的 $X_1 < y, X_2 < y, \cdots, X_{n-1} < y$ 可以去掉。那么：

$$P(X_n > x | T_y = n) = P(X_n > x | X_n > y) = \begin{cases} 1 & x < y \\ \bar{F}(x)/\bar{F}(y) & x > y \end{cases}$$

因此可以注意到 $X_n > x$ 的概率与 n 的取值是无关的。

1-35 倾斜密度

(a)按照定义

$$\mathrm{E}(h(x)) = \int_{-\infty}^{+\infty} h(x)f(x)dx$$

$$M(t)\mathrm{E}[\exp(-tX_t)h(X_t)] = M(t)\int_{-\infty}^{+\infty} \frac{1}{M(t)}\exp(tx_t)\exp(-tx_t)h(x_t)dx_t = \int_{-\infty}^{+\infty} h(x_t)dx_t$$

命题得证。
(b)按照定义

$$P(X > a) = \int_a^{+\infty} f(x)dx$$

$$P(X_t > a) = \int_a^{+\infty} \frac{1}{M(t)}\exp(tx_t)f(x_t)dx_t$$

那么

$$M(t)e^{-ta}P(X_t > a) = \int_a^{+\infty} \exp(t(x_t - a))f(x_t)dx_t > \int_a^{+\infty} f(x_t)dx_t$$

命题得证。
(c) 利用定义：

$$\mathrm{E}[X_t] = a \quad \Rightarrow \quad \int x\exp(t^\star x)f(x)dx = aM(t^\star)$$

要使得 $M(t)e^{-ta}$ 最小，那么

$$\frac{\mathrm{d}M(t)e^{-ta}}{\mathrm{d}t}|_{t=t^\star} = 0$$

解得

$$\int (x-a)\exp(t^\star(x-a))dx = 0$$

移向，方程两侧同时乘以 $\exp(t^\star a)$，得到：

$$\int x\exp(t^\star x)dx = a\int \exp(t^\star x)dx = aM(t^\star)$$

命题得证。