Lecture 6

Multinomial distributions

- as extension of binomial distribution

1. The experiment consists of Ω identical trials.
2. The outcome of each trial falls into one of k classes
3. Probability of falling into class i is P_i
4. The trials are independent
5. Let $X_1, \cdots, X_k$ be the number of trials in which the outcome falls into class $1, 2, \cdots, k$, then the vector $\begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$ is called a multinomial random vector.
$\longrightarrow$ the df is $k-1$

Note: $X_1 + \cdots + X_k = n$, x_i's are non-negative integers that sum to n.
i.e. space of possible outcomes: $\{(x_1, \cdots, x_k), x_i \geq 0, x_i \in N, \Sigma x_i = n\}$
Need to find $P(X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k)$

e.g. 1 2 3 4 5 6 7 8 9 10
G A O O B O O AB O O O
$P_1 \cdot P_4 \cdot P_4 \cdot P_3 \cdot P_4 \cdot P_4 \cdot P_3$ P_4 P_4 P_4

Idea:
For every argument of outcome, P(that particular outcome to occur) $= P_1^{x_1} P_2^{x_2} \cdots P_k^{x_k}$
$\Rightarrow$ desired probability $=$ [# of outcomes s.t. $X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k$] $\cdot$ $[P_1^{x_1} P_2^{x_2} \cdots P_k^{x_k}]$
$\hookrightarrow$ denoted as $\binom{n}{x_1 \cdots x_k} \Rightarrow \frac{n!}{x_1! x_2! \cdots x_k!}$

$\binom{n}{x_1 \cdots x_k} = \frac{n!}{x_0! x_1! \cdots x_k!}$ Note that if $k=2$ $\binom{n}{x_1 \; x_2} = \binom{n}{x_p}$

In summary, $P(X_1 = x_1, X_2 = x_2, \cdots X_k = x_k) = \frac{n}{x_1! x_2! \cdots x_s} P_1^{x_1} \cdots P_k^{x_k}$

Corollary:
1. For each i, $1 \leq i \leq k$, $E(X_i) = nP_i$, $Var(X_i) = nP_i(1-P)$
2. For any $i \neq j$, $Cov(X_i, X_j) = -nP_iP_j$ HW

2 possible outcomes
$1 \rightarrow P_i$ $0 \rightarrow (1-P_i)$
so $x_i \sim bin(n, P_i)$

talk a little bit look at independence.

Def: For events $A_1, \cdots A_n$, we call them indeptent if the indicator random variables $I\{A_1\}, I\{A_2\}, \cdots, I\{A_n\}$ are independent.

Thm: $A_1, \cdots A_n$ are independent iff $P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k})$ (*)
$\forall\ 1 \le i_1 < i_2 < \cdots < i_k \le n$

Proof: 1° $\Rightarrow$ If $A_1, \cdots, A_n$ are indp $\Rightarrow I(A_{i_1}), \cdots, I(A_{i_k})$ are indep for any $1 \le i_1 < \cdots < i_k \le n$

hence $P(I\{A_{i_1}\}=1, I\{A_{i_2}\}=1, \cdots, I\{A_{i_k}\}=1) = P(I\{A_{i_1}\}=1) P(I\{A_{i_2}\}=1) \cdots P(\quad)$ ✓

LHS of ✓ $= P(A_{i_1} A_{i_2} \cdots A_{i_k})$, RHS of ✓ $= P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k})$

So $P(A_{i_1} \cdots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k})$

2° "$\Leftarrow$" Need to show, if (*) holds, $I\{A_{i_1}\} \cdots I\{A_{i_k}\}$ are indep.

i.e. need: $P(I\{A_1\} = \binom{0}{1} \cdots I\{A_n\} = \binom{0}{1}) = P(I\{A_1\} = \binom{0}{1}) P(\quad) \cdots P(\cdots)$

Note: there are 2^n qualities to check!

e.g. $n=2$, 4 cases

① $P(I(A_1)=1, I(A_2)=1) = P(I(A_1)=1) P(I(A_2)=1)$

② $P(I(A_1)=1, I(A_2)=0) = P(I(A_1)=1) P(I(A_2)=0)$

③ $P(I(A_1)=0, I(A_2)=1) = P(I(A_1)=0) P(I(A_2)=1)$

④ $P(0, 0) = P(0) P(0)$

LHS of ① $= P(A_1 A_2)$, RHS $= P(A_1) P(A_2)$

$\because$ (*) for $n=2$ is $P(A_1 A_2) = P(A_1) P(A_2)$ $\therefore$ ① holds

LHS of ② $= P(A_1 \overline{A_2})$, RHS $= P(A_1) P(\overline{A_2})$

$\because A_1 = A_1 A_2 \cup A_1 \overline{A_2} \Rightarrow P(A_1) = P(A_1 A_2) + P(A_1 \overline{A_2})$

$\Rightarrow P(A_1 \overline{A_2}) = P(A_1) - P(A_1 A_2) \overset{\text{by (*)}}{=\!=} P(A_1) - P(A_1) P(A_2) = P(A_1)(1 - P(A_2))$
$= P(A_1) P(\overline{A_2})$
$\Rightarrow$ ② holds

(③ holds because of symmetry to ②)

LHS of ④ $= P(\overline{A_1} \overline{A_2})$, RHS of ④ $= P(\overline{A_1}) P(\overline{A_2})$

$P(\overline{A_1} \overline{A_2}) = 1 - P(A_1 \cup A_2) = 1 - [P(A_1) + P(A_2) - P(A_1 A_2)] = 1 - P(A_1) - P(A_2) + P(A_1) P(A_2)$
$= (1 - P(A_1))(1 - P(A_2))$
$= P(\overline{A_1}) P(\overline{A_2})$

Thus ①②③④ hold.

∎

HW: Show that if $n=k$, "$\Leftarrow$" holds then for $n=k+1$ "$\Leftarrow$" holds

Moment & Probability Generating functions

pgf / mgf — Def: sps that R is a r.v. ① $\Pi(z) = E(z^R)$ is called the prob. generating function of R.

② $M(z) = E(e^{zR})$ is called the moment generating function of R

Example 1: If R is integer-valued r.v. then $\Pi(z)=\sum_{j=0}^{\infty} P(R=j)z^j$

$$M(z)=\sum_{j=0}^{\infty} P(R=j)\, e^{zj}$$

Example 2: Sps $X \sim \text{Poisson}(\lambda)$

$$\Pi_X(z)=\sum P(X=j)z^j = e^{-\lambda}\sum \frac{\lambda^j}{j!} z^j = e^{-\lambda}e^{z\lambda} = e^{(z-1)\lambda}$$

HW: $M_X(z)=E(e^{zX})=\sum P(X=j)e^{zj}=e^{-\lambda}\sum \frac{\lambda^j}{j!}e^{zj}=e^{-\lambda}\sum\frac{(\lambda e^z)^j}{j!}$

$$=e^{-\lambda}e^{\lambda e^z}=e^{(e^z-1)\lambda}$$

Thm: If X, Y are independent, $\Pi_{X+Y}(z)=\Pi_X(z)\Pi_Y(z)$, $M_{X+Y}(z)=M_X(z)M_Y(z)$

Proof: $\Pi_{X+Y}(z)=E(z^{X+Y})=E(z^X z^Y)=E(z^X)E(z^Y)$ because $X \perp\!\!\!\perp Y$

$$=\Pi_X(z)\cdot\Pi_Y(z)$$

Example: How to calculate the distribution of $X+Y$?

$\to$ If X, Y are both integer-valued.

then $P(X+Y=t)=\sum P(X=j, Y=t-j)=\sum P(X=j)P(Y=t-j) \to$ convolution of distribution

Corollary: if $X_1,\cdots,X_n$ are indep. then $M_{X_1+\cdots+X_n}(z)=M_{X_1}(z)M_{X_2}(z)\cdots M_{X_n}(z)$

$\Leftrightarrow$ 若 X, Y dist 不同, 则 Π 和 M 都不同. Uniqueness of pgf: $\Pi_K(z)$

Thm: if $\Pi_X(z)$, $M_X(z)$ are finite on $[-\delta,\delta]$ for some $\delta>0$, either $\Pi_X(z)=\Pi_Y(z)$ on $[-\delta,\delta]$, or $M_X(z)=M_Y(z)$ on $[-\delta,\delta]$, then the distribution of X $\underline{\text{exactly}}$ distribution of Y.

Eg: Sps $X\sim \text{Pois}(\lambda)$, $Y\sim\text{Pois}(\mu)$ and $X \perp\!\!\!\perp Y$

Show $X+Y\sim\text{Poisson}(\lambda+\mu)$

Proof: $\Pi_{X+Y}(z)=\Pi_X(z)\Pi_Y(z)=e^{(z-1)\lambda}e^{(z-1)\mu}=e^{(z-1)(\lambda+\mu)}$

On the other hand, for a $\text{Pois}(\lambda+\mu)$ r.v. W.

$\Pi_W(z)=e^{(z-1)(\mu+\lambda)} \longrightarrow \Pi_{X+Y}(z)$

by the

get $X+Y\sim\text{Pois}(\lambda+\mu)$

[NOT REQUIRED] Thm: If $M_X(z)$ is finite on $(-\delta,\delta)$ for some $\delta>0$, then $E(X^k)=M_X^{(k)}(0)$ for all N^+, where $M^{(k)}(0)$ is the k-th derivative of M at 0.

Proof (only show case $k=1$, others similarly follow):

i.e. need to show $E(X)=M_X^{(1)}(0)$

Note that for any t, $\frac{M_X(t)-M_X(0)}{t}=\frac{E(e^{tX})-E(e^{t\cdot 0})}{t}=E\left(\frac{e^{tX}-e^{t\cdot 0}}{t}\right)$

note that as $t\to 0$, $\frac{e^{tX}-e^{0X}}{t}\to X \not\Rightarrow M'_X(0)=\frac{M_X(t)-M_X(0)}{t} \not\to E(X)$

"$\Rightarrow$"不一定成立

Recall that $X_k \to X \not\Rightarrow EX_k \to EX$

(Axiom 5 says $X_k \to X$ monotonically $\Rightarrow EX_k \to EX$

(DCT) Dominate Convergence Thm: $|X_k| \le Y$, $EY < \infty$, middle value thm

Note that if $|t| \le \sigma$, then $(**)$ $\dfrac{e^{tx} - e^{t\cdot 0}}{t} = xe^{t^*}$ for some $t^* \in (0,1)$

$\le |X|(e^{rx} + e^{-rx})$

want an integrable upperbound that has finite expectation

$e^x \ge x$ if $x \ge 1 \Rightarrow rx \le e^{rx}$ if $rx \ge 1 \Rightarrow 0 \le x < \frac{e^{rx}}{r}$ if $x \ge \frac{1}{r}$

if $x > r \cdots r \le \frac{1}{r} + e^{rx}/r$, hence $\forall x \in R, \forall r > 0$, $|X| \le \left(\frac{2}{r} + \frac{e^{rx} + e^{-rx}}{r}\right)$

$\Rightarrow (**) \le \left(\frac{2}{r} + \frac{e^{-rx} + e^{rx}}{r}\right)(e^{rx} + e^{-rx})$. let $r = \frac{\sigma}{4}$, let $Y = \dfrac{5(1 + e^{2rx} + e^{-2rx})}{r}$