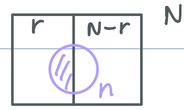
May 12

④ The Hypergeometric distribution

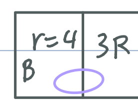
- Setup: population of size N, a subpopulation of size r. Take a sample of size n. Let X = number of items from the subpopulation included in the sample.



X is said to have a hypergeometric distribution with parameter: N, r, n.

Range of X?

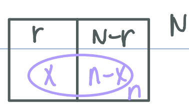
Ex: an urn contains 3 red balls & 4 blue. Take a sample of size $n=5$ from the urn. Let X = # of blue balls.



$N=7$
$n=5$

$X \in \{2, 3, 4\}$.

Back to the general case: $X \leq n$ and $X \leq r \Rightarrow X \leq \text{Min}(n, r)$.



$\underbrace{n - X \leq n}_{X \geq 0}$ and $n - X \leq N - r$

and $X \geq n + r - N$

$X \geq \text{Max}\{0, n+r-N\}$

$\text{Max}\{0, n+r-N\} \leq X \leq \text{Min}(n, r)$

p.f. of X: let k be an integer such that

$\text{Max}\{0, n+r-N\} \leq k \leq \text{Min}\{n, r\}$

$$P(X=k) = \frac{C_k^r \, C_{n-k}^{N-r}}{C_n^N}$$

- Ex: An urn contains 4 blue ball & 3 red ball. Take a sample of size $n=5$. $X = \#$ of blue balls.

$X \in \{2,3,4\}$. $P_X(2) = \frac{C_2^4 C_3^3}{C_5^7} = \frac{6}{\frac{7\times6}{2}} = \frac{2}{7}$

$P_X(3) = \frac{C_3^4 C_2^3}{\frac{7\times6}{2}} = \frac{4\times3}{\frac{7\times6}{2}} = \frac{4}{7}$

$P_X(4) = \frac{C_4^4 C_1^3}{\frac{7\times6}{2}} = \frac{1}{7}$

Prop: If X has a hypergeometric distribution with parameters N, r, n

then: ① $\mathbb{E}(X) = n\frac{r}{N}$

② $V(X) = n\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$

success

$N = 10^9$, $n=3$, $\frac{N-n}{N-1} \approx 1$, $r=100$

$P(3S) = \frac{10^2}{10^9} \times \frac{99}{10^9-1} \times \frac{98}{10^9-2}$ $\approx$ $P(3S) = \left(\frac{10^2}{10^9}\right)^3$

⑤ Prisson Distribution

$p: \mathbb{R} \to (0,1)$

$\{x, p(x) \neq 0\}$ is discrete and $\sum_x p(x) = 1$

Recall: $e^x = \sum_{k=0}^{+\infty} \frac{x^k}{k!}$ $\forall x \in \mathbb{R}$

- In particular, for $\lambda > 0$ fixed:

$e^\lambda = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!}$ ie. $1 = \sum_{k=0}^{+\infty} \underbrace{e^{-\lambda}\frac{\lambda^k}{k!}}_{P_X(k)}$

There exists a discrete r.v. X, such that $\text{Range}(X) = \{0,1,2,\dots\} = \mathbb{N}$

and $\forall k \in \mathbb{N}$ $P_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$.

Such r.v is said to have a Prisson distribution with parameter λ.

Notation: $X \sim \text{Prisson}(\lambda)$ $(a_k > 0)$

$$\sum_{k=0}^{+\infty} a_k = L > 0 \iff 1 = \sum_{k=0}^{+\infty} \frac{a_k}{L} \leftarrow p_X(h)$$

Ex: The number of defects in a certain fabric is believed to have a Poisson distribution with mean (parameter) $\lambda = 2$. What is the probability that the fabric contains more than 3 defects?

$X = \#$ of defects. $X \sim \text{Poisson}(\lambda = 2)$. $P(X > 3) = 1 - P(X \le 3)$

$$= 1 - \left(\sum_{k=0}^{3} P(X=k)\right) = 1 - \sum_{k=0}^{3} e^{-2} \frac{2^k}{k!} = 1 - e^{-2}(1+2+2) = 1 - 5e^{-2}$$

Exercise: let $X \sim \text{Poisson}(\lambda = 4)$. For which value(s) is $p_X(k)$ maximal (the highest)?

Prop: If $X \sim \text{Poisson}(\lambda)$

① $\mathbb{E}(X) = \lambda$

② $V(X) = \lambda$

Proof:

① $\mathbb{E}(X) = \sum_{k=0}^{+\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = 0 + e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \overset{(l=k-1)}{=}$

$$\lambda e^{-\lambda} \sum_{l=0}^{+\infty} \frac{\lambda^l}{l!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

② $V(X) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - (\mathbb{E}(X))^2 = \mathbb{E}(X(X-1)) + \lambda - \lambda^2$

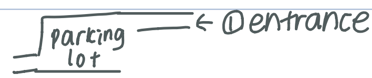
$$\mathbb{E}(X(X-1)) = \sum_{k=0}^{+\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} = 0 + 0 + \sum_{k=2}^{+\infty} \frac{k(k-1)}{k!} e^{-\lambda} \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \lambda^2 \sum_{k=2}^{+\infty} \frac{\lambda^{k-2}}{(k-2)!} \overset{(l=k-2)}{=} e^{-\lambda} \lambda^2 \sum_{l=0}^{+\infty} \frac{\lambda^l}{l!} = \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2$$

$$V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Exercise: Find $\mathbb{E}(X^3)$ and $\mathbb{E}(X^4)$

Exercise: Entrance ② →



X_i = # of cars which use ① in a given day.

$X_i \sim \text{Poisson}(\lambda_i)$. Moreover assume that a car will use either entrance independently of others (in practice the events $\{X_1 = k\}$ and $\{X_2 = \ell\}$ are independent). Find the distribution of $X_1 + X_2 = X$

$X \in \{0, 1, 2, \cdots\} = \mathbb{N}$.

$P(X=0) = P(X_1 = 0 \text{ and } X_2 = 0) = P(X_1 = 0)\, P(X_2 = 0) = e^{-\lambda_1} e^{-\lambda_2} = e^{-(\lambda_1 + \lambda_2)}$

$P(X=1)$

$\vdots$

$P(X=k)$

Moment generating functions (mgf)

Def: Let X be a discrete r.v. and P_X be its p.f. The moment generating function of X (mgf of X) is the function defined as:

$$m_X(t) = \mathbb{E}(e^{tX}) = \sum_x P_X(x)\, e^{tx}$$

Ex:

x	-1	0	1
$P_X(x)$	$1/2$	$1/3$	$1/6$

$m_X(t) = \sum_x P_X(x)\, e^{tx} = \frac{1}{2}e^{-t} + \frac{1}{3}e^{0t} + \frac{1}{6}e^{t}$

$m_X(t) = \frac{1}{3} + \frac{1}{2}e^{-t} + \frac{1}{6}e^{t} \quad \forall t \in \mathbb{R}$

Properties:

① $m_X(0) = \mathbb{E}(e^{0X}) = \mathbb{E}(1) = 1$

② If the range of X is a finite set then $m_X(t)$ is defined for every $t \in \mathbb{R}$. The domain of m_X is an interval

Thm: The mgf is a unique identifier of a distribution ie. No 2 diff distributions have the same mgf. $m_X \longleftrightarrow p_X$

Ex: Given $m_X(t) = \frac{1}{2}e^{-5} + \frac{1}{4} + \frac{1}{6}e^{2t} + \frac{1}{12}e^{4t}$ $t \in \mathbb{R}$.

m_X is the mgf of some discrete r.v. X. What is p_X? p.f of x

X	-5	0	2	4
$p_X(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$

Def: The n^{th} moment of a r.v. X (if defined) is denoted $\mu_n = \mathbb{E}(X^n)$

Connection between mgf and $\mathbb{E}(X^n)$?

$m_X(t) = \mathbb{E}(e^{tX}) = \mathbb{E}\left(\sum_{n=0}^{+\infty} \frac{(tX)^n}{n!}\right)$ under some conditions

$= \sum_{n=0}^{+\infty} \mathbb{E}\left(\frac{(tX)^n}{n!}\right) = \sum_{n=0}^{+\infty} \frac{t^n}{n!}\mathbb{E}(X^n) = \boxed{\sum_{n=0}^{+\infty} \frac{\mu_n}{n!} t^n}$ (where $\mu_n = \mathbb{E}(X^n)$) n^{th} moment of X.

(**)

Under some conditions: $\frac{d^n}{dt^n}(m_X(t))\Big|_{t=0} = \mu_n$. This follows from (**)

Thm: If the domain of m_X contains an interval of the form $(-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$ $(0 \to +\infty)$

Then $m_X(t) = \sum_{n=0}^{+\infty} \frac{\mu_n}{n!} t^n$ on some interval, around 0.

Consequently, $\mu_n = \mathbb{E}(X^n) = \frac{d^n}{dt^n}(m_X(t))\Big|_{t=0}$

Ex:

① Binomial Distribution: Since $X \sim Bin(n,p)$ has a finite range, then Domain of $m_X = \mathbb{R}$.

$m_X(t) = \mathbb{E}(e^{tX}) = \sum_{k=0}^{n} e^{kt} p_X(k)$

$$= \sum_{k=0}^{n} e^{kt} C_k^n p^k (1-p)^{n-k} = \sum_{k=0}^{n} C_k^n \underbrace{(pe^t)}_{a}^k \underbrace{(1-p)}_{b}^{n-k} = [pe^t + (1-p)]^n$$

→ If $X \sim Bin(n, p)$ then $m_X(t) = [pe^t + (1-p)]^n \quad \forall t \in \mathbb{R}$

$m_X(t) = (pe^t + (1-p))^n$

$\frac{d}{dt}(m_X(t)) = npe^t (pe^t + (1-p))^{n-1}$

→ $E(X) = \frac{d}{dt}(m_X(t)) \Big|_{t=0} = np(p + 1 - p)^{n-1} = np$

→ $\frac{d^2}{dt^2}(m_X(t)) = npe^t (pe^t + (1-p))^{n-1} + n(n-1)(pe^t)^2 (pe^t + (1-p))^{n-2}$

$\mathbb{E}(X^2) = \frac{d^2}{dt^2}(m_X(t)) \Big|_{t=0} = np + n(n-1)p^2 \ \ldots$