Discrete Random Variables (I)

Let

$X : S \longrightarrow \mathbb{R}$ $\rightsquigarrow$ random variable.

S = sample space of an experiment.

$\mathcal{X}$:= Values/image of X.

Often times when talking about $\rightsquigarrow$ we will only specify about X $\mathcal{X}$ (not worry about S)

Suppose

$F_X \rightsquigarrow$ cdf of X

p_X or $f_X \rightsquigarrow$ pdf of X.

} $\rightsquigarrow$ Can use these to calculate

$P(a \leq X \leq b)$

$\boxed{P(X \text{ takes values in the interval } [a,b])}$

X is said to be Discrete r.v $\rightsquigarrow$ if F_X is a step function

$\mathcal{X}$ will be a discrete subset of $\mathbb{R}$.

ie $\mathcal{X}$ is a

finite set $\Updownarrow$ in bijection with $\{1,2,3,\dots,N\}$ for some $N \in \mathbb{N}$

Countably infinite set $\Updownarrow$ in bijection with $\mathbb{N} = \{1,2,3,4,\dots\}$

Goal for today $\rightsquigarrow$ Study r.v's with finite $\mathcal{X}$.

ie calculate pmf $p_X(x)$, μ_X and σ_X^2

① Uniform Discrete Random Variable

Fix $N \in \mathbb{N}$.

$\boxed{\text{The r.v } X \text{ has "uniform discrete distribution" with parameter } N.}$ $\rightsquigarrow$ if $\mathcal{X} = \{1,2,3,\dots,N\}$ and $p_X(x) = P(X=x) = \frac{1}{N}$ for $x = 1,2,3,\dots,N$. , $p_X(x) = 0$ $x \notin \mathcal{X}$.

Note: ① $\sum_{x\in\mathcal{X}} p_X(x) = \underbrace{\frac{1}{N} + \frac{1}{N} + \cdots + \frac{1}{N}}_{N \text{ times}} = N \cdot \frac{1}{N} = 1$

and $p_X(x) = \frac{1}{N} > 0 \quad \forall x \in \mathcal{X}$

$\therefore$ p_X is a legitimate probability mass function!

② The distribution table for X is

x	1	2	3	4		N-1	N
$P(X=x)$	$\frac{1}{N}$	$\frac{1}{N}$	$\frac{1}{N}$	$\frac{1}{N}$		$\frac{1}{N}$	$\frac{1}{N}$

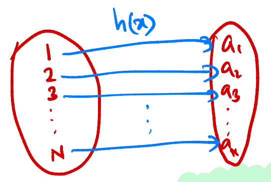
③ $E(X) = \sum_{x\in\mathcal{X}} x \cdot p_X(x)$, $\mathcal{X} = \{1,2,3,\dots,N\}$

$$\mu_X = E(X) = 1\cdot\frac{1}{N} + 2\cdot\frac{1}{N} + 3\cdot\frac{1}{N} + \cdots + N\cdot\frac{1}{N}$$
$$= \frac{1}{N}(1+2+3+\cdots+N)$$
$$= \frac{1}{N}\cdot\left(\frac{N\cdot(N+1)}{2}\right) = \frac{(N+1)}{2}$$ $\sim$ mid point of the values.

④ To calculate $V(X)$, $V(X) = \sum_{x\in\mathcal{X}} (x-\mu_X)^2 \cdot p_X(x) = \left(1 - \frac{(N+1)}{2}\right)^2 \cdot \frac{1}{N} + \cdots \left(N - \frac{(N+1)}{2}\right)^2 \cdot \frac{1}{N}$ $= E(X^2) - (E(X))^2$

$$E(X^2) = 1\cdot\frac{1}{N} + (2)^2\cdot\frac{1}{N} + 3^2\cdot\frac{1}{N} + \cdots + N^2\cdot\frac{1}{N}$$
$$= \frac{1}{N}\Big(\underbrace{1+2^2+3^2+\cdots+N^2}_{\text{sum of squares of first } N \text{ natural numbers}}\Big)$$
$$= \frac{1}{N}\left(\frac{N(N+1)(2N+1)}{6}\right) = \frac{(N+1)(2N+1)}{6}$$

$$\therefore V(X) = E(X^2) - E(X)^2 = \frac{(N+1)(2N+1)}{6} - \left(\frac{(N+1)}{2}\right)^2$$
$$= \frac{(N+1)(N-1)}{12} = \left(\frac{N^2-1}{12}\right).$$



⑤ Let $\{a_1, a_2, a_3, \dots, a_N\} \subseteq \mathbb{R}$

define $h : \mathcal{X} \longrightarrow \{a_1, a_2, \dots, a_N\}$, $i \longmapsto a_i$

$Y = h(X)$, $E(Y) = ?$

$E(h(X)) = \sum_{x\in\mathcal{X}} h(x)\cdot p_X(x)$.

then if X has uniform distribution $\rightsquigarrow$ $E(h(X)) = \frac{a_1 + a_2 + \cdots + a_N}{N}$

$V(h(X))$ = variance of the numbers $\{a_1, a_2, a_3, \dots, a_N\}$

$\frac{1}{n}\sum_{i=1}^{n}(a_i - \mu)^2$

Now fix $N, M, n \in \mathbb{N}$

Suppose a bag has N balls (identical except for color) M of which are red.

$E(x) = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n$

Experiment: Choose n balls from the bag

$\boxed{X = \# \text{ of red balls in the sample of } n \text{ balls.}}$

The calculation of $p_X(x) = P(X=x)$

depends on if

Sampling without replacement $\Updownarrow$ $\boxed{\text{Hypergeometric distribution}}$ with parameters: N, M, n

Sampling with replacement $\Updownarrow$ $\boxed{\text{Binomial Distribution}}$ with parameters: n, $p = \frac{N}{M} \cdot \frac{M}{N}$

Bag total # balls = N, M = red

choosing n. balls.

$\therefore$ total # ways to choose n balls from $N = \binom{N}{n}$

ways to choose favourable outcome $= \{X = x\}$

= want exactly x red balls in a sample of n balls.

② Hypergeometric Distribution

$\boxed{X \text{ is said to have hypergeometric distribution with parameters: } N, M, n}$ $\xrightarrow{\text{if}}$ $\mathcal{X} = \{0,1,2,3,\dots,n\}$

$$\boxed{p_X(x) = \frac{\binom{M}{x}\cdot\binom{N-M}{n-x}}{\binom{N}{n}}}$$ $= P(X=x)$

$= \binom{M}{x}\cdot\binom{N-M}{n-x}$

$P(X=x) = P\left(\begin{array}{l}\text{getting exactly } x \text{ successes in a sample of size } n \\ \text{sampled without replacement from a population} \\ \text{of size } N \text{ with } M \text{ successes}\end{array}\right)$

Note: x must satisfy:

$$\boxed{\max(0, n-N+M) \leq x \leq \min(n, M)}$$

sample size, population, # successes.

$N=10$
$n=15$
$P(X=11)=?$

$p_X(x) = \frac{\binom{M}{x}\cdot\binom{N-M}{n-x}}{\binom{N}{n}} \geq 0$

$\sum_{x\in\mathcal{X}} p_X(x) = 1$

Note: If N is large, $\frac{N}{M} \to p$

Can use the hypergeometric dist (N, M, n) to approximate Binomial $(N, p = \frac{M}{N})$

$\boxed{\text{If } X \text{ has hypergeometric distribution with parameters: } N, M, n}$ $\rightsquigarrow$ $E(X) = n \cdot \frac{M}{N}$

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

$N = 10000$

$n = 3$

$\frac{1}{10000} \cdot \frac{1}{9999} \cdot \frac{1}{9998}$

12

$\left(\frac{1}{10000}\right)^3$

if we set $p = \frac{M}{N}$

$$\boxed{E(X) = n\cdot p \quad \text{and} \quad V(X) = \left(\frac{N-n}{N-1}\right)^{\approx 1} \cdot n\cdot p\cdot(1-p)}$$

$N \to \infty$

③ Binomial Distribution

$\boxed{X \text{ has Binomial Distribution with parameters: } n, p}$ $\xrightarrow{\text{if}}$ $\mathcal{X} = \{0,1,2,\dots,n\}$

Binomial Experiment

for $x \in \mathcal{X}$

$$\boxed{p_X(x) = \binom{n}{x}\cdot p^x \cdot (1-p)^{(n-x)}}$$

Note:

$$1 = 1^n = (p + (1-p))^n = \sum_{x=0}^{n}\binom{n}{x} p^x \cdot (1-p)^{(n-x)} = \sum_{x\in\mathcal{X}} p_X(x)$$

$\therefore$ p_X is a pmf.

$\boxed{\text{If } X \text{ has Binomial Distribution with parameters: } n, p}$ $\rightsquigarrow$ $\boxed{E(X) = n\cdot p \quad V(X) = n\cdot p\cdot q}$ $(1-p)$

④ Bernoulli Trials

Experiment with exactly two outcomes: success (p), Failure ($1-p$)

Let $p \in (0,1)$ and let

$P(\text{Success}) = p$ $\boxed{P(\text{Failure}) = 1-p}$

$\boxed{X \text{ is a Bernoulli random variable with parameter: } p}$ $\rightsquigarrow$ $\mathcal{X} = \{0,1\}$ and

$$\boxed{p_X(x) = \begin{cases} p & \text{if } x=1 \\ (1-p) & \text{if } x=0 \end{cases}}$$

In this case we have

$\boxed{E(X) = p \quad \text{and} \quad V(X) = p\cdot(1-p)}$

Store. ⇝ receives a shipment of 400 refrigerators

Question for the store manager: Do you accept/reject the shipment.

Impractical strategy ⇝ Check if all 400 refrigerators are working.

Practical strategy: $N = 400$, $M = 30$ defectives inside the sample.

failures.

get a random sample $n = 10$

X = # of defectives of sample of size 10.

Hypergeometric distribution.

Acceptance sampling: