CHAP TWO

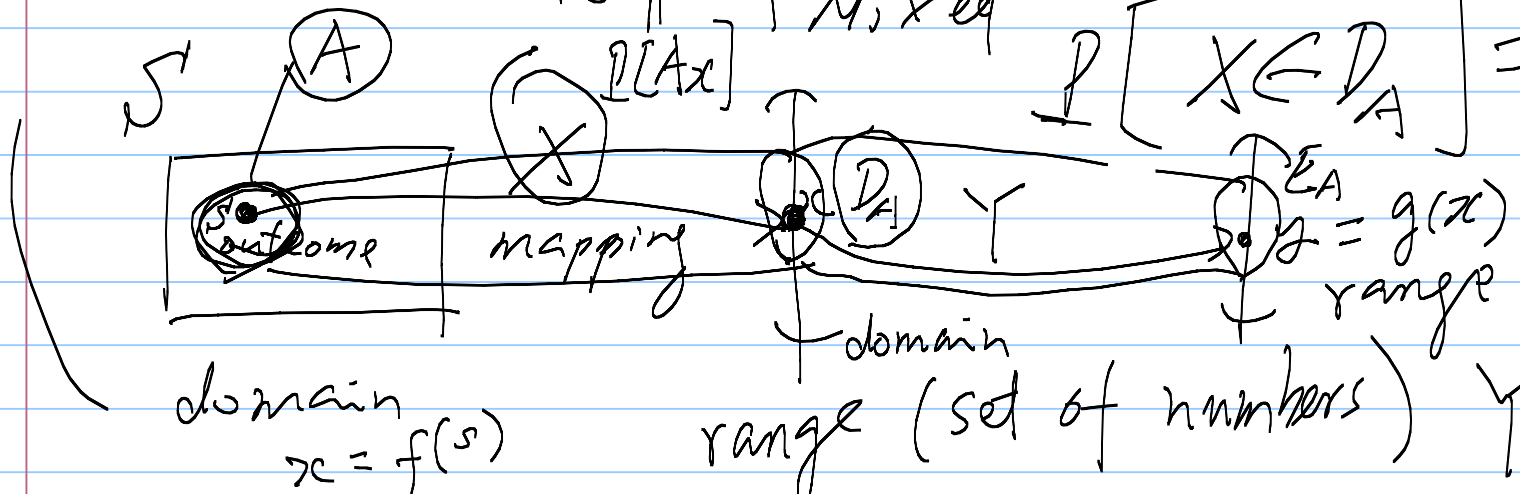
Experiment : Procedure and Observation

Physical Model $\Rightarrow$ Probability Model

Random Variable < Discrete ✓ / Continuous ✓ / Mixed

$P[A]$ $\quad P_X(x) = P[X=x] = P[A_x]$

$P[X \in D_A] = P[A] = P[Y \in E_A]$

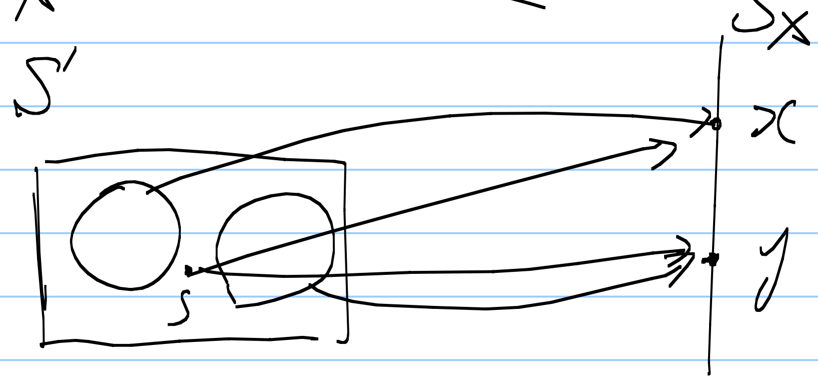


$Y \triangleq g(X)$

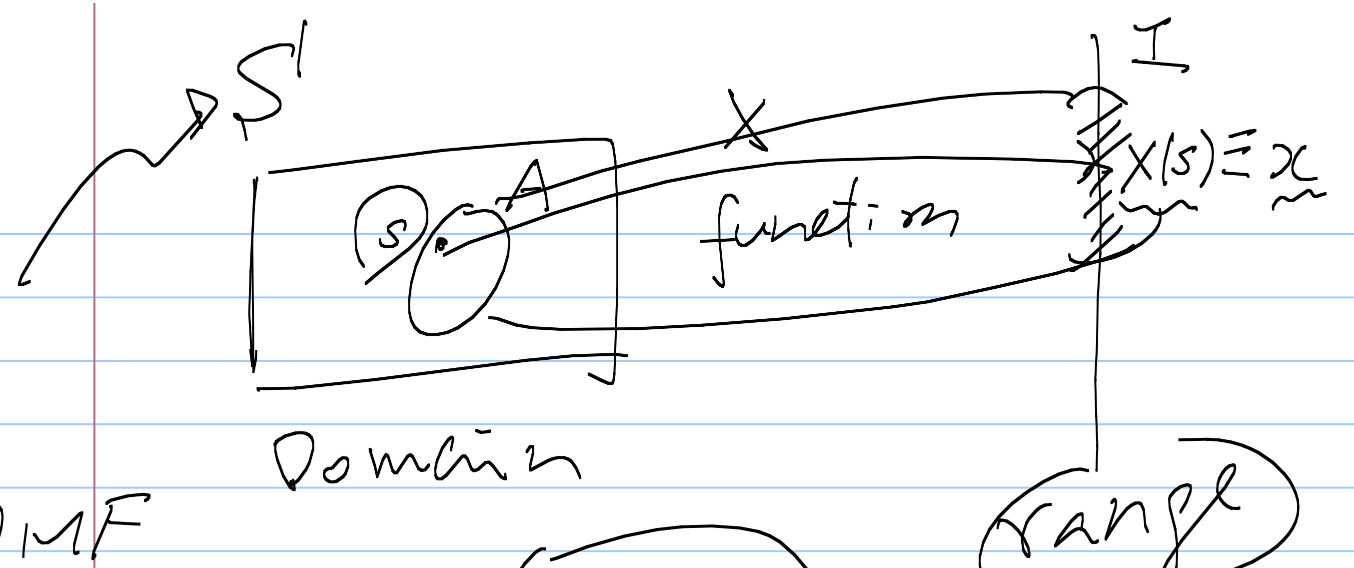
$$\boxed{P[X = x]} \equiv P[A] \text{ where}$$

$$A = \{s \mid X(s) = x\}$$

$$P_X(x) \triangleq P[X = x]$$



$\{X = x\} \cap \{X = y\} = \phi$ if $x \neq y$.



X(s)

PMF

$$P_X(x) \triangleq P[X=x] = P[\{s \mid X(s)=x\}]$$

$$\{X \in I\} = A$$

$$P[X \in I] = P[A].$$

A discrete random variable X is said to be completely statistically characterized by its $P_X(x)$.

$$\overbrace{\underbrace{X\ X\ \cdots\ X}_{x-1}\ r}^{x}$$

r or a

$k-1$ r

$x-(k-1)$ a

$$\underbrace{\binom{x-1}{k-1} p^{k-1} (1-p)^{(x-1)-(k-1)}}_{k-1 \text{ successes}} \cdot p$$

r a

aa ... ar

$\overbrace{a\,a\,r\,a}^{l-1} \cdots r$

l

$$\binom{l-1}{k-1} p \, p^{k-1} (1-p)^{l-k}$$

q

Recall: Consider Repeatable Experiment with <u>identical</u> and <u>independent</u> subexperiments.
Each subexperiment gives two possible outcomes, namely success, with probability p, and failure, with probability $1-p$.

Bernoulli (p) random variable represents the number of successes in one single trial

Geometric (p) random variable represents the number of trials until the first success (inclusive)

Binomial (n, p) random variable represents the number of successes in n trials.

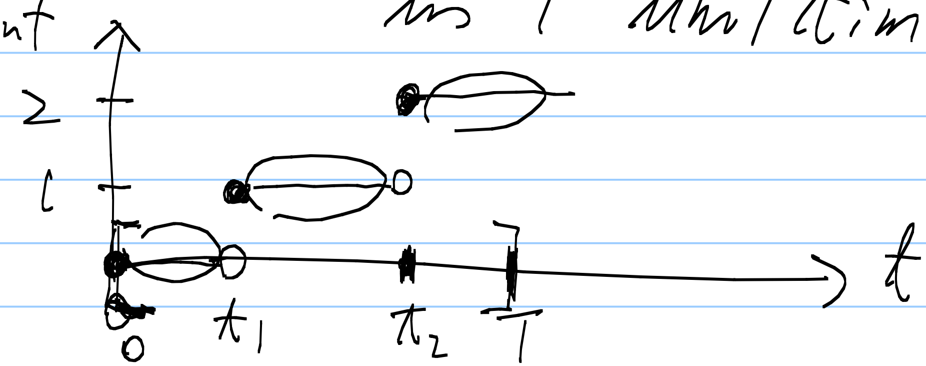
Note: Binomial $(1, p) =$ Bernoulli (p)

Pascal (k, p) random variable represents the number of trials until k successes (inclusive). Note: Pascal $(1, p)$ is equivalent to Geometric (p).

Poisson (α) Random Variable = number of arrivals in T unit times

Counter arrival

Inter-Arrivals occur independently



α denotes the average number of arrivals in T unit times

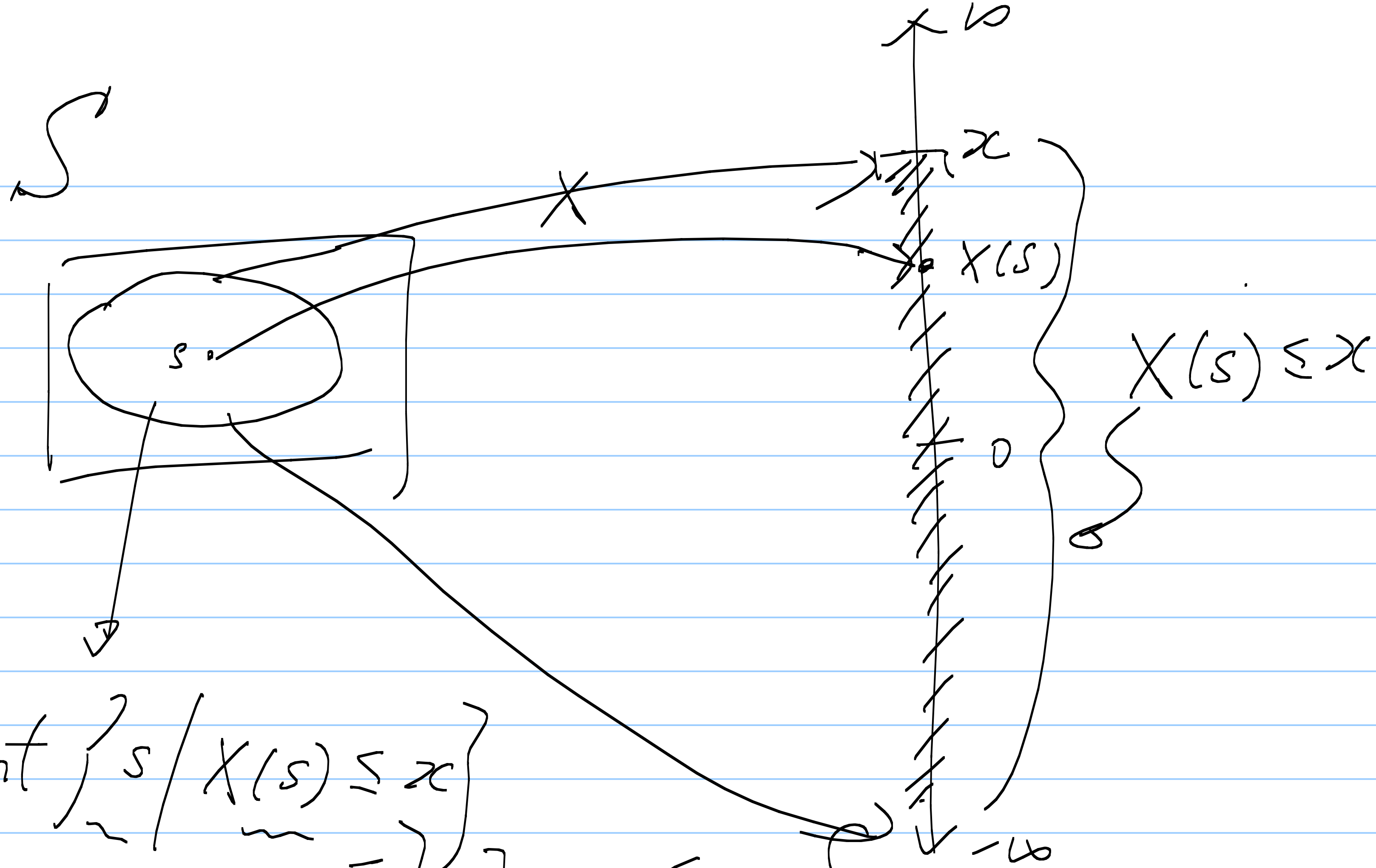
λ denotes the average number of arrivals in unit time

$\alpha = \lambda T$

$$P_K(k) = \begin{cases} \alpha^k e^{-\alpha}/k! & , k = 0, 1, \ldots \\ 0 & , \text{otherwise} \end{cases}$$

PMF describes a complete probability model of a discrete random variable.

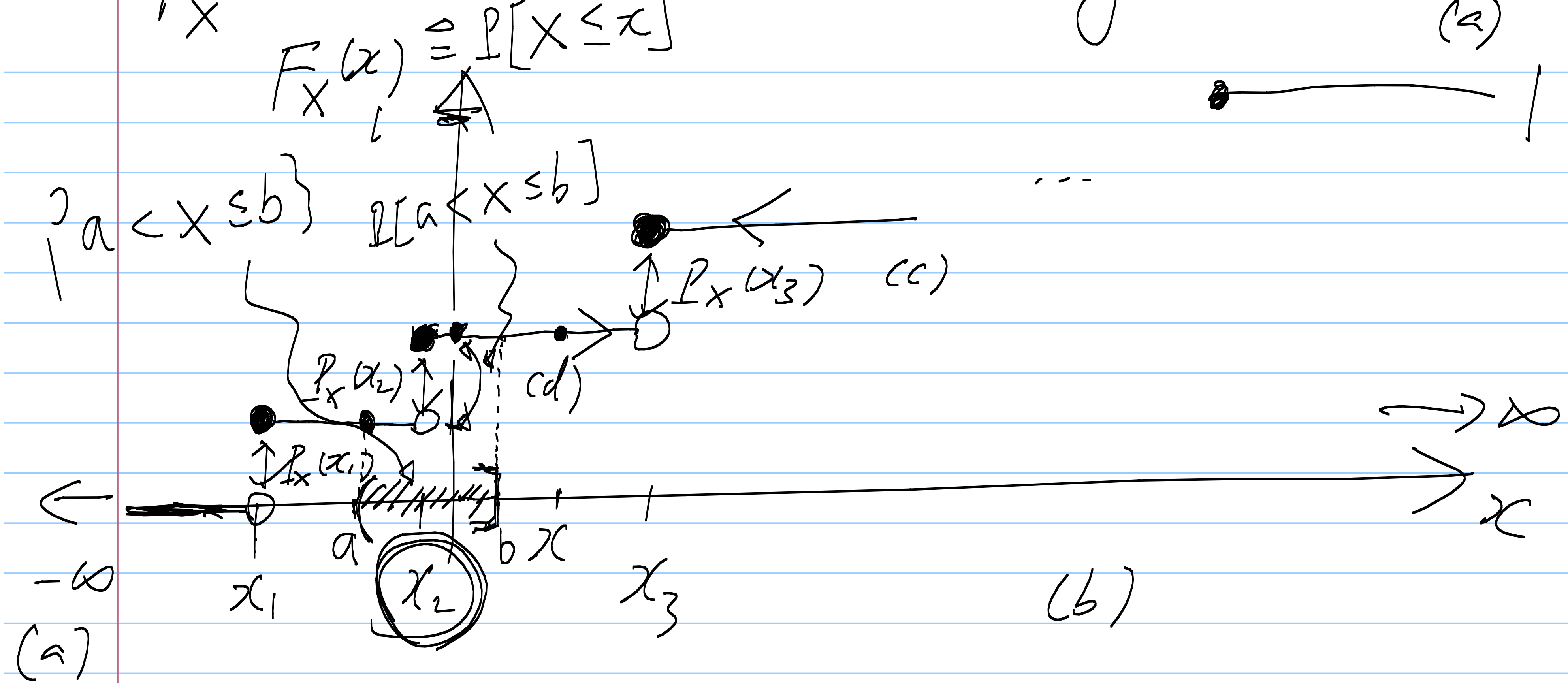
(Cumulative) Distribution (Function) (CDF) also decribes a complete probability model.



event $\{ s \mid X(s) \le x \}$

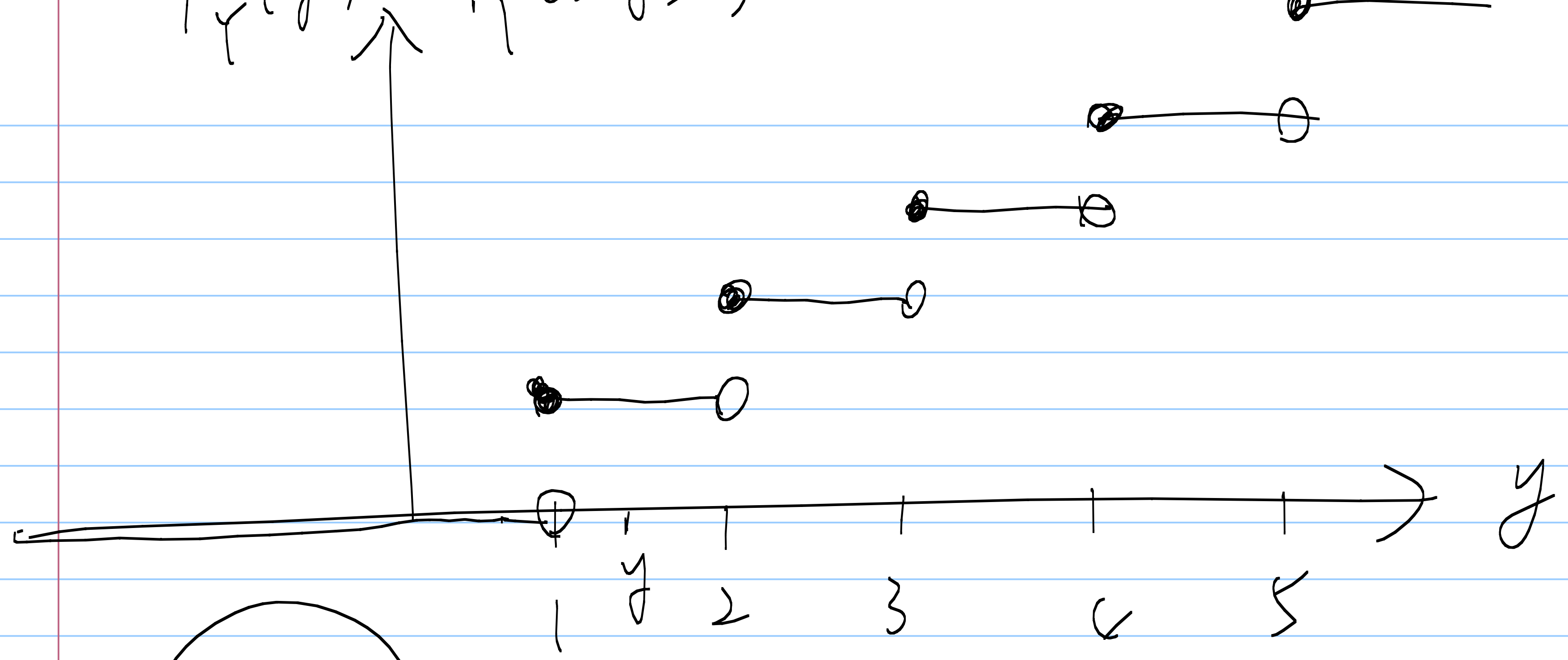
$F_X(x) \triangleq P[\ \{\ \}\ ] = P[X \le x] \neq P[X < x]$

$F_X(x)$ must be nondecreasing. $S_X = \{x_1, x_2, \dots\}$



(e) $F_X(x)$ is continuous from the right, but not necessarily from the left.

$$F_Y(y) = F_Y(\lfloor y \rfloor)$$



$\lfloor y \rfloor \equiv$ the largest integer that is not larger than y.

$\equiv$ integer part of y

$\lceil y \rceil \equiv$ the smallest integer that is not smaller than y

A statistic is a single number derived from (numerical)

- The <u>average</u> value of a set of experiment outcomes is a statistic of the outcomes.

Three averages of interest:

mean (or expected value, or expectation)

median

mode (unimodal, multimodal)

Consider an experiment that produces a rv X. Perform n independent trials of this experiment. Let x_i denote the ith sample value.

$$m_n \equiv \text{sample average} = \frac{1}{n}\underbrace{\sum_{i=1}^{n}}x_i$$

$$= \frac{1}{n}\underbrace{\sum_{x \in S_X}} x N_x \qquad \left(N_x \equiv \text{the number of sample value } x \text{ in } n \text{ trials} \right)$$

$$m_n = \sum_{x \in S_X} x \frac{N_x}{n}$$

We can interpret $\underbrace{P[X = x]} = \lim_{n \to \infty} \frac{N_x}{n}$

(relative frequency interpretation)

$$\lim_{n\to\infty} m_n = \lim_{n\to\infty} \sum_{x\in S_X} x \frac{N_x}{n}$$

$$= \sum_{x\in S_X} x \underbrace{\lim_{n\to\infty} \frac{N_x}{n}}_{P[X=x]} = \sum_{x\in S_X} x P[X=x] = E[X]$$

$$\sum_{x=0}^{\infty} q^x = \frac{1}{1-q} \quad \text{for } 0<q<1$$

$$\underbrace{\sum_{x=1}^{\infty} x q^x} = \sum_{x=0}^{\infty} x q^x = \sum_{x=0}^{\infty} q \frac{d}{dq} q^x$$

$$= q \frac{d}{dq} \underbrace{\sum_{x=0}^{\infty} q^x}_{\frac{1}{1-q}} = q \underbrace{\frac{d}{dq} (1-q)^{-1}}_{+(1-q)^{-2}}$$

$$= \frac{q}{(1-q)^2}$$

$$\Rightarrow \boxed{\frac{p}{\cancel{q}} \frac{\cancel{q}}{(1-q)^2}} = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

$$\lim_{n\to\infty} P_K(k) = \underbrace{\lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{n^k}}_{=1} \cdot \frac{\alpha^k}{k!} \cdot \lim_{n\to\infty} \underbrace{\left(1-\frac{\alpha}{n}\right)^{n-k}}_{= \frac{\left(1-\frac{\alpha}{n}\right)^n}{\left(1-\frac{\alpha}{n}\right)^k}}$$

$$= \frac{\alpha^k}{k!} e^{-\alpha}$$

$$= \frac{\lim_{n\to\infty} \left(1-\frac{\alpha}{n}\right)^n}{\lim_{n\to\infty} \left(1-\frac{\alpha}{n}\right)^k} \qquad \lim_{n\to\infty} \left(1-\frac{\alpha}{n}\right)^k = 1$$

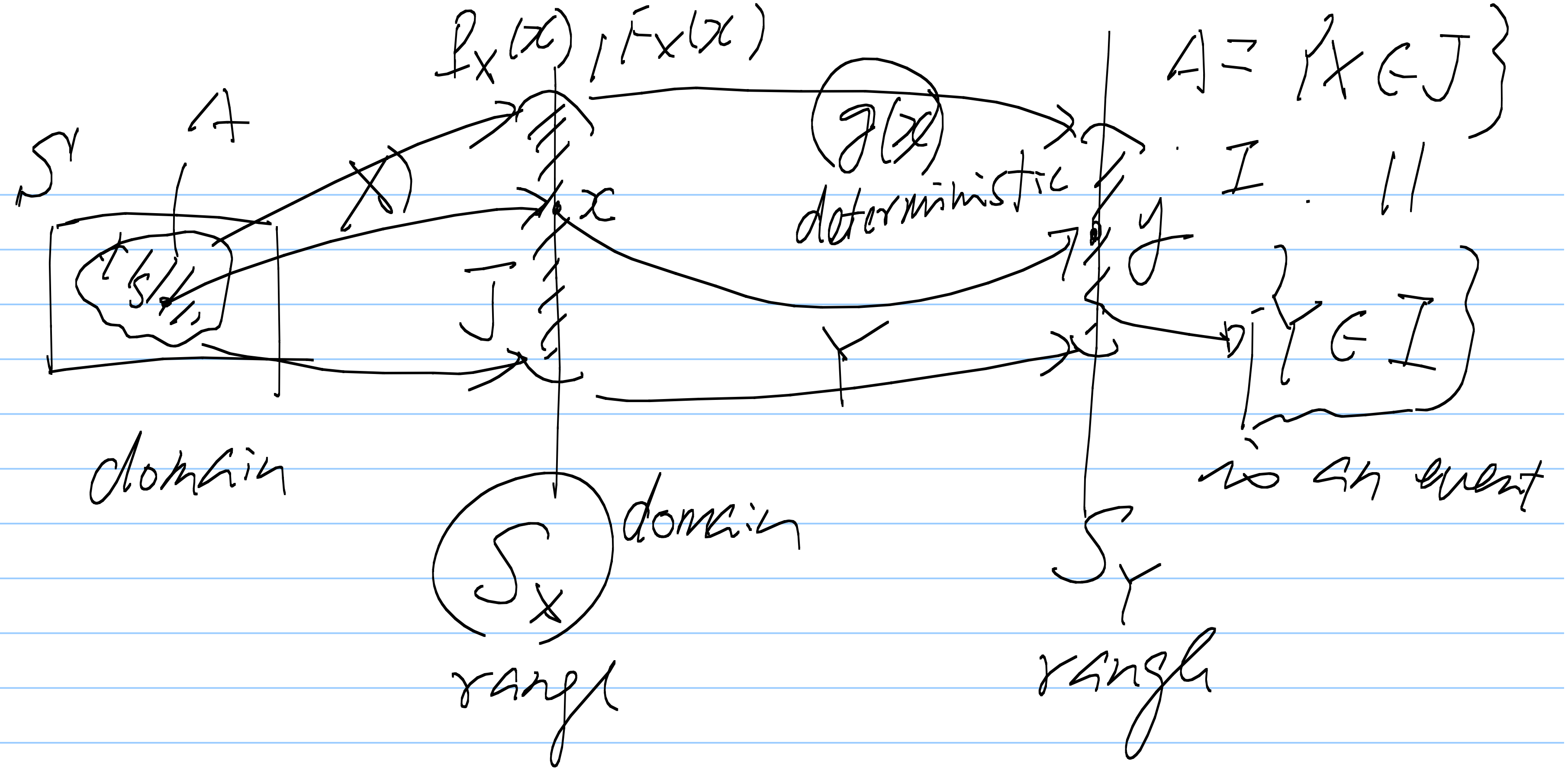
$$= \lim_{n\to\infty} \left(1-\frac{\alpha}{n}\right)^n = \lim_{n\to\infty} e^{n\ln\left(1-\frac{\alpha}{n}\right)}$$

$$\frac{0}{0}$$

$$= e^{\lim_{n\to\infty} n \ln\left(1-\frac{\alpha}{n}\right)} = e^{\lim_{n\to\infty} \frac{\ln(1-\alpha/n)}{1/n}}$$

L'Hopital law

$$\overset{\text{L'Hopital law}}{=} e^{\lim_{n\to\infty} \frac{\frac{\alpha/n^2}{1-\alpha/n}}{-n^{-2}}} = e^{\lim_{n\to\infty} \frac{-\alpha}{1-\alpha/n}} = e^{-\alpha}$$

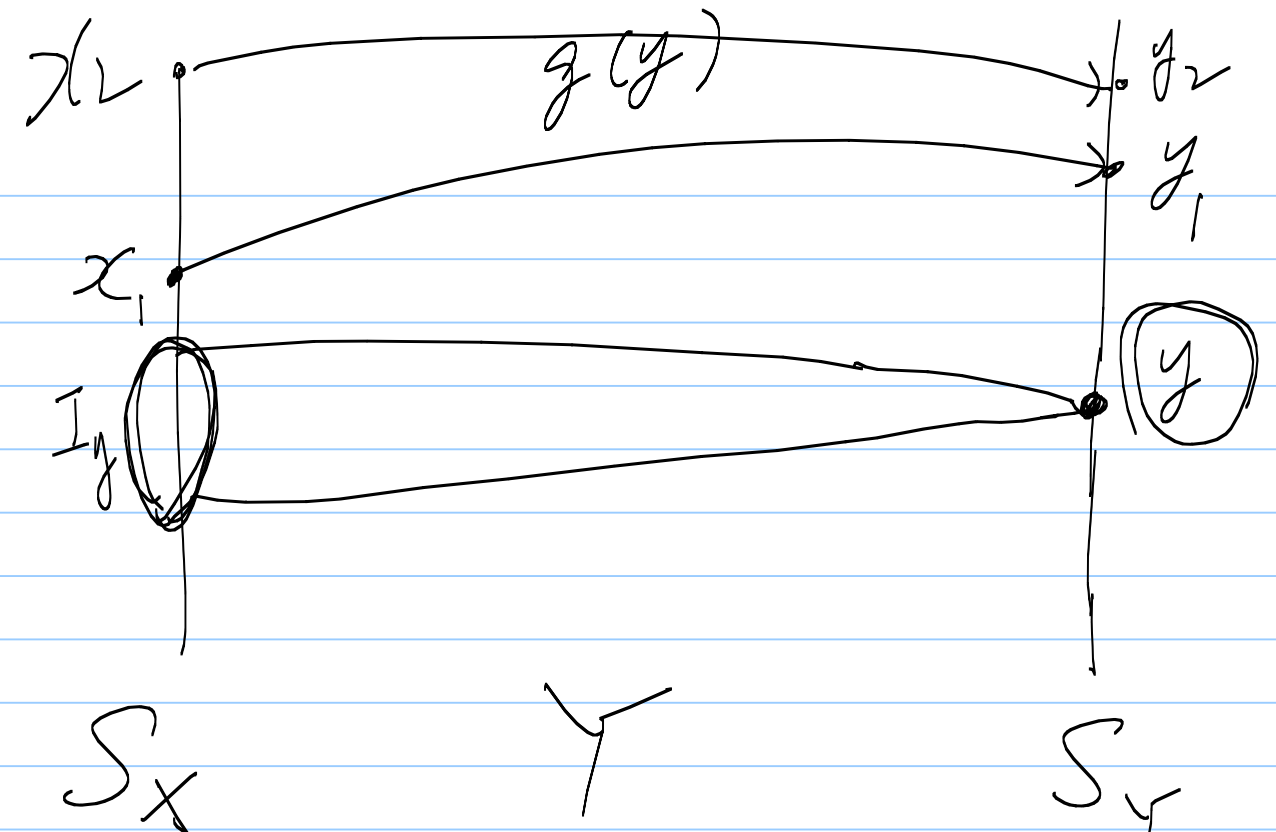


X is completely statistically characterized if $f_X(x)$ or $F_X(x)$ is known.

$A = \{s \mid X(s) \in J\} = \{s \mid Y(s) \in I\}$

Y is called the derived random variable.

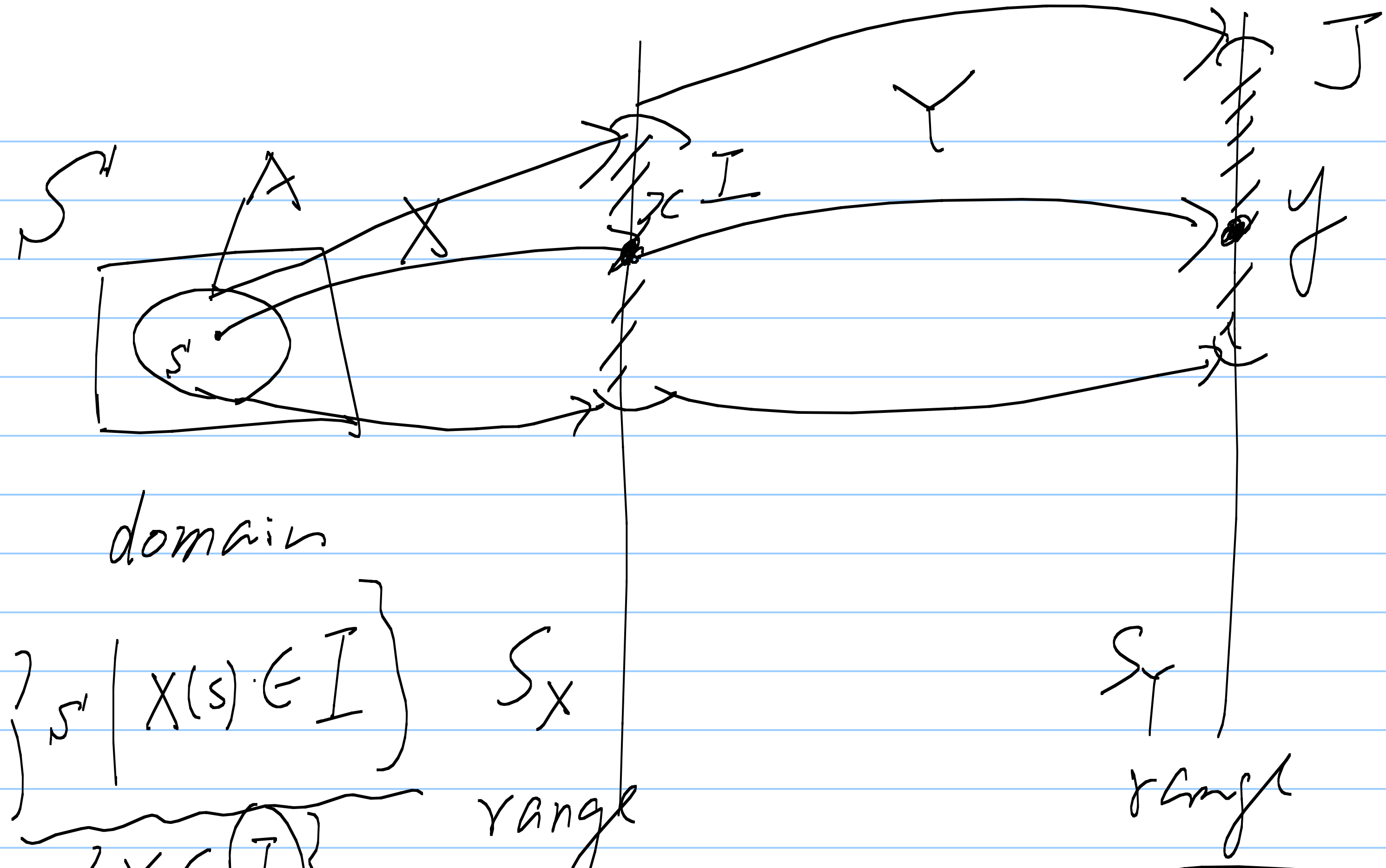
Denote $\boxed{Y = g(X)}$ ✓



① When $g(x_1) \neq g(x_2)$ for $x_1 \neq x_2$ (i.e., $g(x)$ is one-to-one), $P_Y(y) = P_X(x)$ if $y = g(x)$.

② When $y = g(x)$ for all $x \in I_y \triangleq \{x \mid g(x) = y\}$,

$$P_Y(y) = \sum_{x \in I_y} P_X(x)$$



$$A = \{ s \mid X(s) \in I \}$$

$$= \{ X \in I \}$$

$$= \{ Y(X(s)) \in J \} = \{ Y \in J \}$$

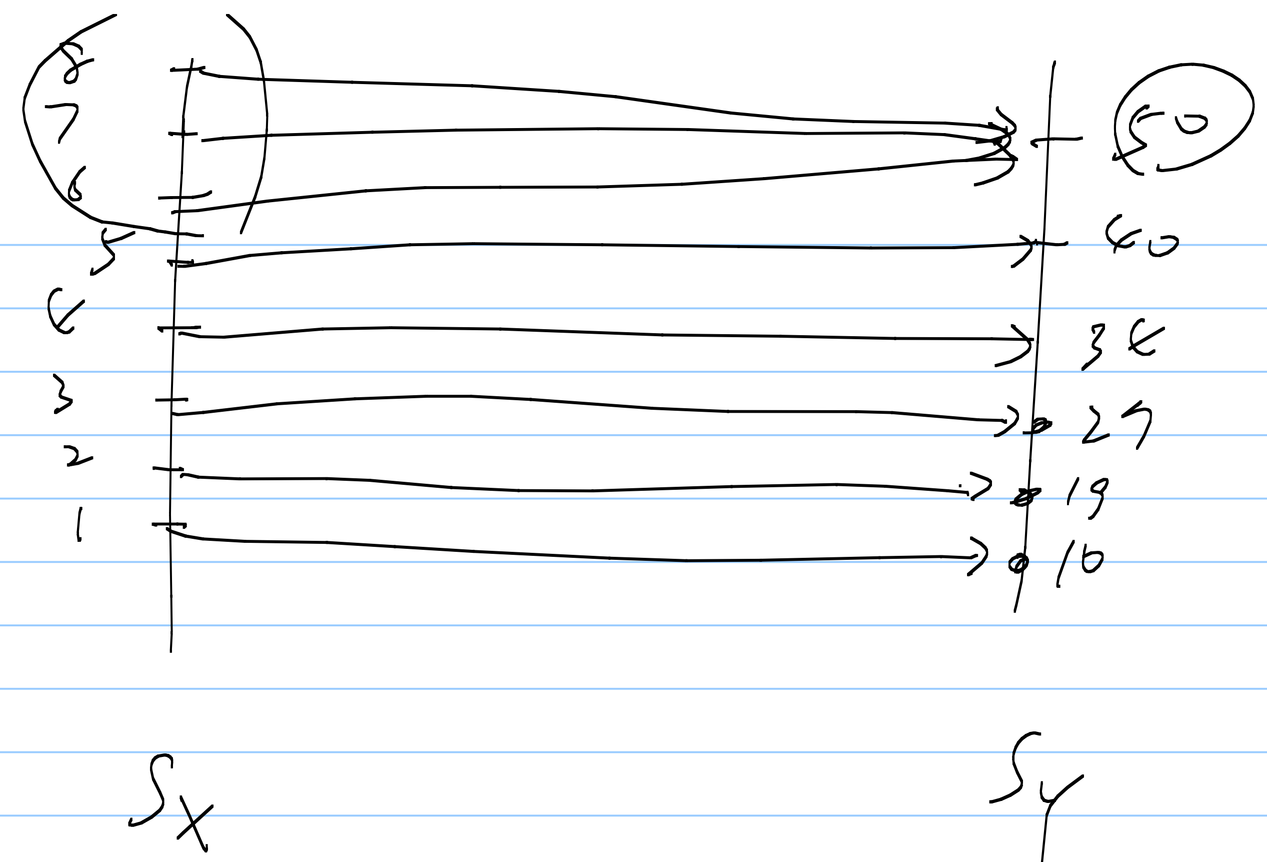
domain

$$Y = g(X)$$

When $J = \{y\}$, $A = \{x \mid g(x) = y\} = \{s \mid g(x(s)) = y\}$ and

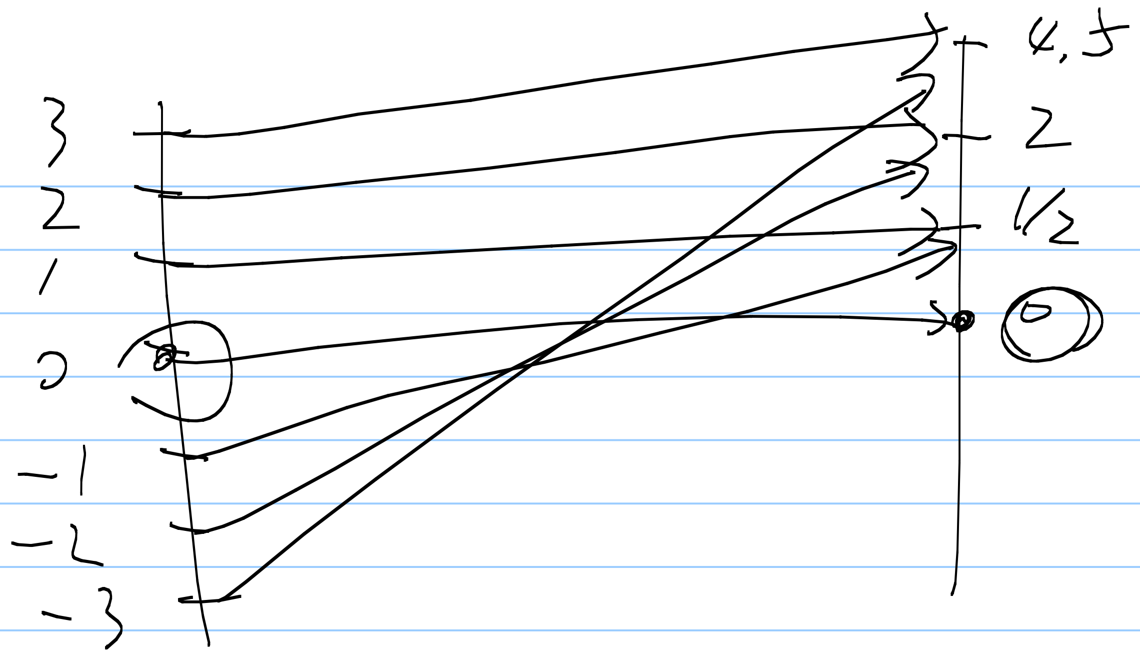
$$P[Y \in J] = P[Y = y] = P[A] = \sum_{x : g(x) = y} P_X(x)$$

$$P[Y = y] = P_Y(y)$$



many-to-one mapping

Given $P_X(x)$ and $y = g(x)$, this section deals with the approach of finding $P_Y(y)$.

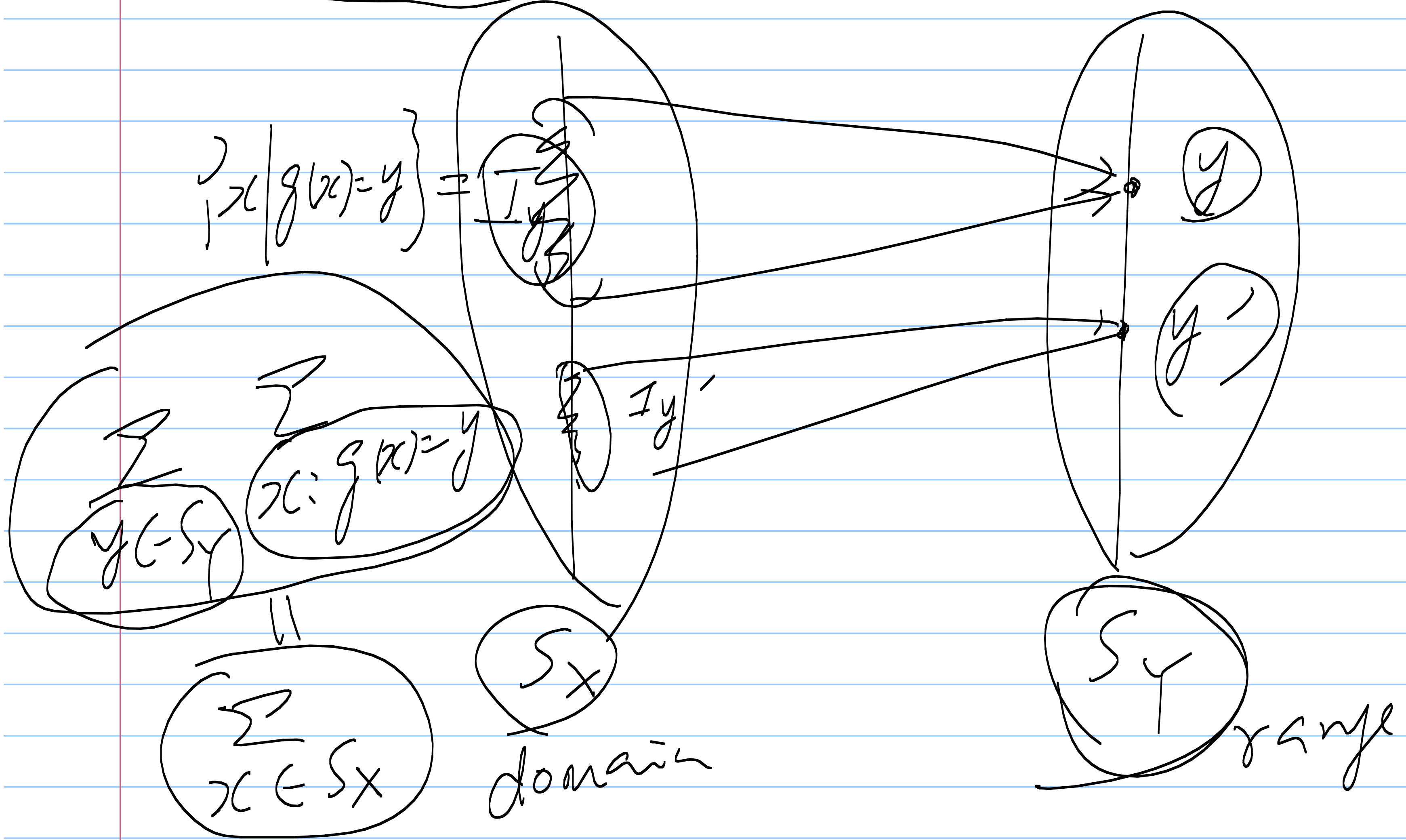


S_X S_Y

$\{Y=0\} = \{X=0\}$, $\{Y=\frac{1}{2}\} = \{X \in \{-1, 1\}\}$

$\{Y=2\} = \{X \in \{-2, 2\}\}$, $\{Y=4.5\} = \{X \in \{-3, 3\}\}$

Law of Unconscious Statistician



- An operator $\overrightarrow{\mathbb{T}}[x]$ is called linear (affine) iff for any a_n's, b, and any N, and for any arguments x_n's

$$\underbrace{\overrightarrow{\mathbb{T}}\left[\sum_{n=1}^{N} a_n x_n\right] + b}_{} = \sum_{n=1}^{N} a_n \overrightarrow{\mathbb{T}}[x_n] + b$$

(Principle of Superposition).

- For any two linear operators $\overrightarrow{\mathbb{T}}[x]$ and

$S[Y]$, $\qquad \underbrace{\mathbb{E}[S[Y]]} = S[\underbrace{\mathbb{E}[Y]}]$ ②

∴ Expectation is linear. $\qquad Y \to \boxed{S} \to \boxed{\mathbb{E}} \to ①$

$$\begin{array}{c} \| \\ \end{array}$$

∵ pf for Theorem 2.12: $\qquad Y \to \boxed{\mathbb{E}} \to \boxed{S} \to ②$

Let $Y = g(X) = aX + b$. Then, from theorem 2.10,

$$E[Y] = \sum_{x \in S_X} g(x) P_X(x) = \underbrace{\sum_{x \in S_X} (ax + b) P_X(x)}$$

$$= a \underbrace{\sum_{x \in S_X} x P_X(x)}_{\mu_X} + b \underbrace{\sum_{x \in S_X} P_X(x)}_{=1} \quad \text{Q.E.D.}$$

If $W = g(R)$ where $g(r)$ is nonlinear, then

$$E[W] \neq g(E[R])$$

If $W = g(R)$ where $g(r)$ is linear, then

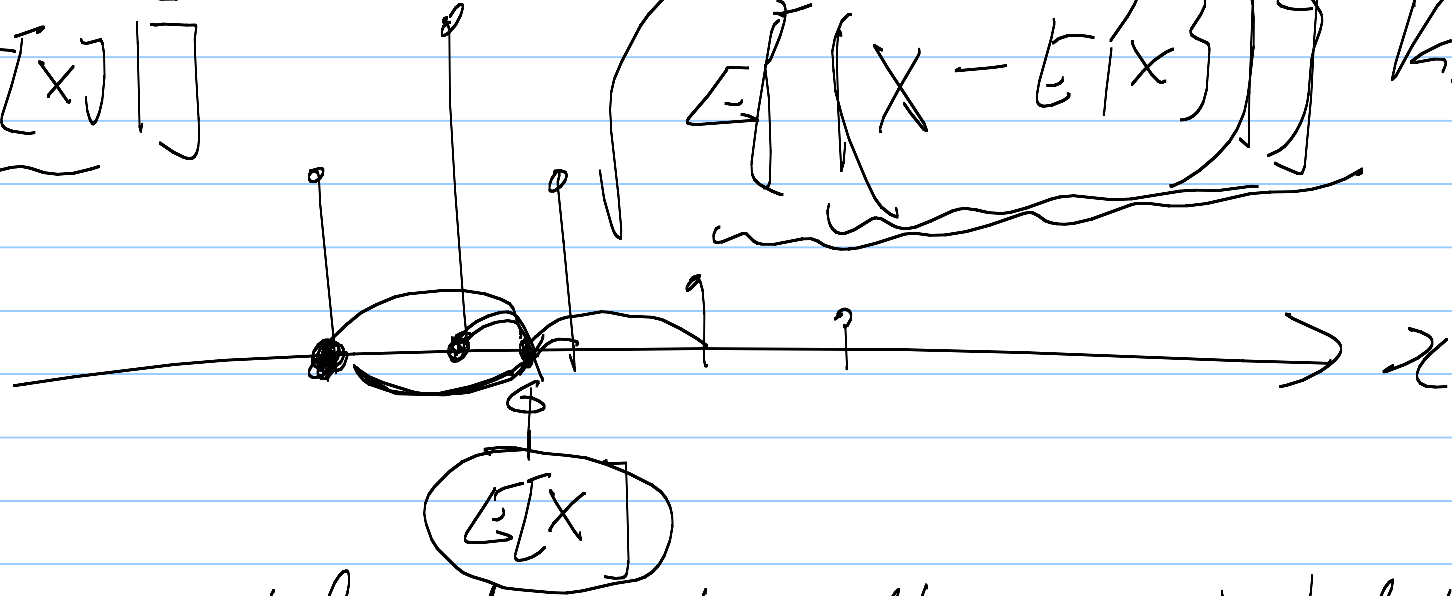
$$E[W] = g(E[R]).$$

$P_X(x)$

$E[X - E[X]] = 0$

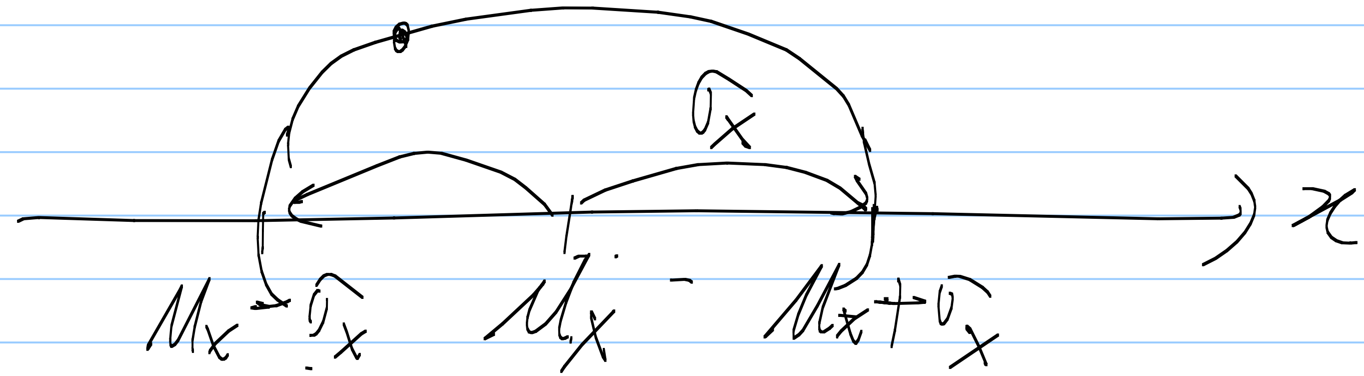
$E[\,|X - E[X]|\,]$

$\sqrt{E[(X - E[X])^2]}$ Variance



$P_X(x)$ completely describe the probability model of X, and so does $F_X(x)$.

For some X,



$$P[X \in (\mu_X - \sigma_X, \mu_X + \sigma_X)]$$

$Var[X]$ or σ_X describes the dispersion of X relative to μ_X.

$$\mathrm{Var}[X] \triangleq E\left[\underbrace{(X-\mu_X)^2}\right]$$

$$= \sum_{x \in S_X} \underbrace{(x-\mu_x)^2}_{(x^2 - 2\mu_x x + \mu_x^2)} P_X(x)$$

law of unconscious statistician

$E[X^2]$ is commonly called the <u>mean square value</u> of X.

All the moments $\{E[X^n] \mid n \in \{1, 2, \dots\}\}$ completely describe the probability model of X.

$$Var[X] \triangleq E[(X - \mu_X)^2] \geq 0$$

$$= E[X^2] - \mu_X^2 \geq 0$$

$$\Longrightarrow \boxed{E[X^2] \geq \mu_X^2 \text{ for any random variable } X.}$$

Recall: $P[A|B]$ denotes the conditional probability of event A given that event B occurs, with $P[B] > 0$. That is, we define

$$P[A|B] = \frac{P[A \cap B]}{P[B]}.$$

If we let $A = \{X = x\}$ for a discrete random variable, then

$$P[A|B] = P[X = x | B] \equiv P_{X|B}(x)$$

$$P_{X|B}(x) \triangleq P[X=x \mid B] = \frac{P[X=x, B]}{P[B]}$$

$P_{X|B}(u)$, $P_{X|B}(\cdot)$

dummy argument

If $x \in B$,

$$P_{X|B}(x) = \frac{P[X=x, B]}{P[B]} = \frac{P[X=x, X \in B]}{P[B]}$$

$$= \frac{P[X=x]}{P[B]} = \frac{P_X(x)}{P[B]}$$

If $x \notin B$,

$$P_{X|B}(x) = \frac{0}{P[B]} = 0$$

Q.E.D.

Note that conditional expectation $E[X|B]$ satisfies the law of unconscious statistician, i.e.,

$$E[g(X)|B] = \sum_{x \in B} g(x)\, p_{X|B}(x),$$

and is linear, i.e., for any constants a, b, $E[aX+b|B] = a\,E[X|B]+b$.

$\{X = x\}$ is an event representing

$\{ s \in S \mid \underbrace{X(s) = x}_{\text{condition}} \}$

argument value

$\{X = x\} = \{ s \mid X(s) = x \}$

S_A is countable

discrete

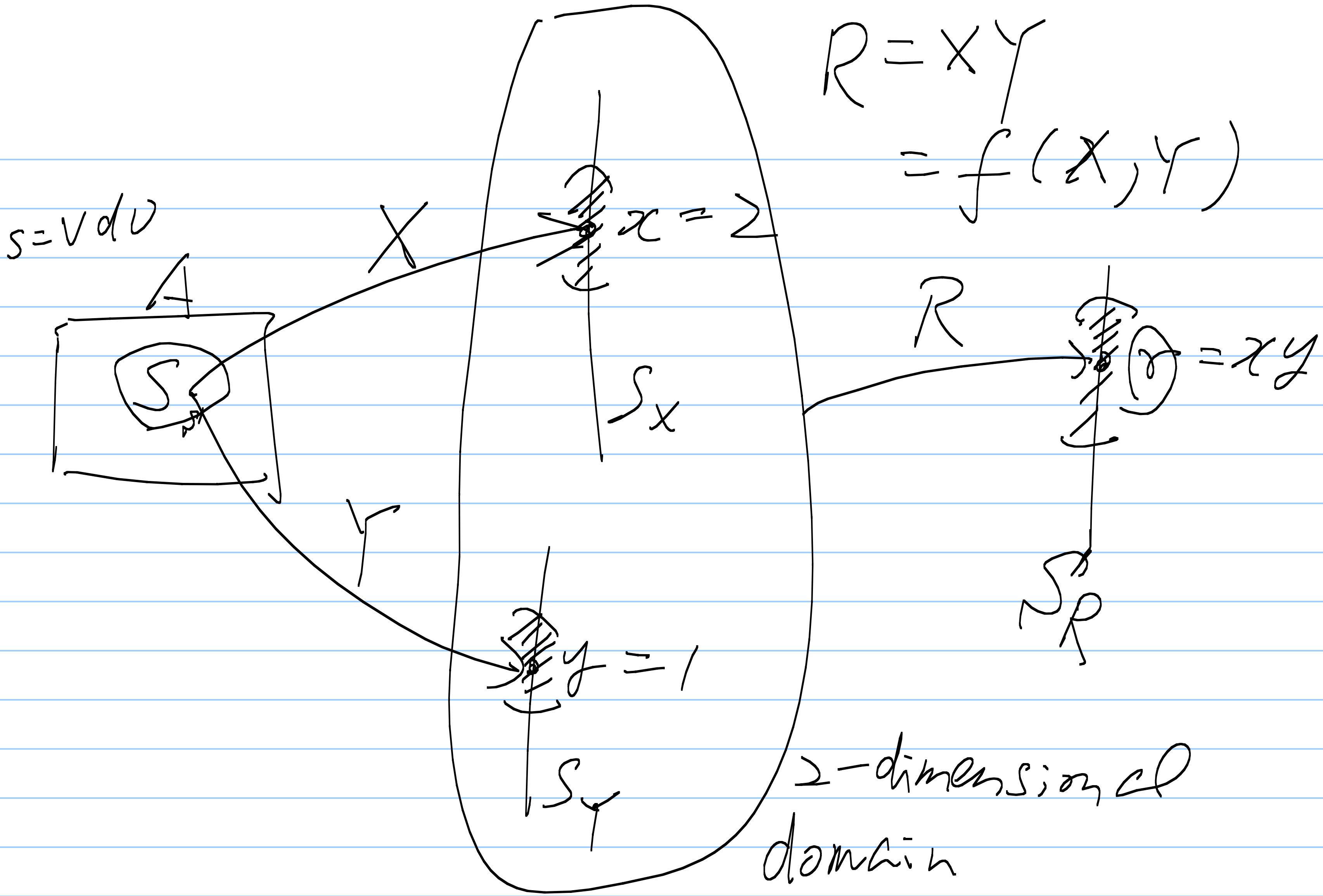
$A \equiv$ # of students asleep (discrete)

$C \equiv$ # of phone calls (S_C is countable)

continuous

$M \equiv$ # of minutes (S_M is NOT countable)

Grade (letter)	$\longrightarrow$	Grade (number)
A		4
B		3
C		2
D		1
E		0


s = v dv
A
S
X
Y
x = 2
y = 1
S_X
S_Y
2-dimensional domain
R
r = xy
S_R
R = XY
= f(X, Y)

$$\underline{P_X(x)} \triangleq P[X = x]$$

$$= P\big[\underbrace{\{ s \mid X(s) = x \}}\big]$$

an event of outcomes satisfying the condition $X(s) = x$

Define event $A = \{X = x\}$

$$\underline{P}[A] = \underline{P}[X = x] = \underline{P}_X(x)$$

Axiom 1 : For any event A,

$$\underline{P}[A] \geq 0 \iff \underline{P}_X(x) \geq 0$$

for any x

Axiom 2 : $\underline{P}[S] = 1$.

$$S = S_X \Rightarrow \underline{P}[S] = \underline{P}[X \in S_X] = \boxed{\sum_{x \in S_X} \underline{P}_X(x)}$$

$=$

$$\overset{2}{=} \mathbb{P}\Big[\bigcup_{x\in S_X} \underbrace{\{X=x\}}\Big] \underset{\text{Axiom 3}}{=} \sum_{x\in S_X} P_X(x) \underset{\text{Axiom 2}}{=} 1$$

(c) For any $B \subset S_X$, the probability that X is in the set B is

$$P[X \in B] = P\Big[\bigcup_{x \in B} \underbrace{\{X = x\}}\Big]$$

$$\text{Axiom 3} \longrightarrow = \sum_{x \in B} \underbrace{P[X = x]}_{P_X(x)}$$

$$= \sum_{x \in B} P_X(x)$$

#

A statistic is a single number (nonrandom or random) that is derived from a probability model.

* repeatable independent trials
* sample space for single trial $= \{0, 1\}$ with $P[0] = 1-p$ and $P[1] = p$.

$$\overbrace{\underbrace{0\,0 \cdots 0}_{x-1}\ 1}^{10} \qquad P[X = x] = (1-p)^{x-1}\, p$$

$1/p$

Since $E[\ ]$ is linear,

$$E[(X-\mu_x)^2] = E[(X^2 - 2\mu_x X + \mu_x^2)]$$

$$= E[X^2] - 2\underbrace{\mu_x E[X]}_{\mu_x^2} + \underbrace{E[\mu_x^2]}_{= \mu_x^2 \ \phi \text{ constant}}$$

$$= \underline{E[X^2] - \mu_x^2}$$

$$P[A] = \sum_{i=1}^{m} P[A \cap B_i]$$

$$= \sum_{i=1}^{m} P[A | B_i] P[B_i]$$

$Y = x$