Ch 4 : Distribution Functions and Discrete Random Variables

便箋標題 2009/10/2

experiment, outcome, event

⇒ some function of the outcome

random variables

- rolling two fair dice

X : the sum   X : {2, 3, 4, ... 12}

$p(X=2) = p(\{1,1\}) = \frac{1}{36}$

$p(X=3) = p(\{1,2\}, \{2,1\}) = \frac{2}{36}$

$\vdots$

Def:

S: the sample space of an experiment

Random variable X: a function that assigns a real value to each outcome in S

$X: S \to \mathbb{R}$

For any set of real numbers A, the prob. that X will assume a value that is contained in A is equal to

the prob. that the outcome of the experiment is contained in $X^{-1}(A)$

$P(X \in A) = P(X^{-1}(A))$ where

$X^{-1}(A)$ is the event consisting of all points $s \in S$

such that $X(s) \in A$

$X(\{1,1\}) = 2$ $X(\{1,2\}) = 3$ $X(\{2,1\}) = 3$

If $A = \{3, 4\}$

$X^{-1}(A) = \{(1,2), (2,1), (1,3), (2,2), (3,1)\}$

$p(X \in A) = p(X^{-1}(A)) = \frac{2}{36} + \frac{3}{36} = \frac{5}{36}$

$\underbrace{X^{-1}(A)}_{\text{event}}$

Ex 4.3

In the U.S., the # of twin births is $\sim \frac{1}{90}$

Let X be the number of births in a certain hospital until the first twins are born.

X is a random variable

T: twin births N: single births

X : a real-valued function defined on the sample space

$$S = \{T, NT, NNT, NNNT, \cdots\}$$ by

$$X(\underbrace{NNN\cdots N}_{i-1}T) = i$$

The set of all possible values of X is $\{1, 2, 3 \cdots\}$ and

$$p(X=i) = p(\{\underbrace{NNN\cdots N}_{i-1}T\}) = \left(\frac{89}{90}\right)^{i-1} \cdot \frac{1}{90}$$

X, Y be two r.v.s over S

$X: S \to \mathbb{R}$ $\quad Y: S \to \mathbb{R}$ are real-valued functions

having the same domain.

$X+Y$; $X-Y$; $aX+bY$

X^2, $\sin X$, $\cos X^2$, e^X, X^3-2X are r.v.s

Ex 4.5

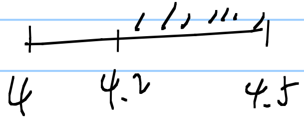
a flat metal disk : diameter is a random number

between 4 and 4.5

What is the prob. that the area of such a flat disk chosen at random is at least 4.41π ?

D : diameter

$$p\left(\left(\frac{D}{2}\right)^2\pi > 4.41\pi\right) = p(D^2 > 17.64) = p(D > 4.2)$$



$$= \frac{4.5-4.2}{4.5-4} = \frac{3}{5}$$

$$\therefore p\left(\frac{D^2}{4}\pi > 4.41\pi\right) = \frac{3}{5}$$

Ex 4.6

A random number is selected from $(0, \frac{\pi}{2})$

What is the prob. that its sine is greater than its cosine?

X: the selected number

$$p(\sin X > \cos X) = p(\tan X > 1) = p(X > \frac{\pi}{4})$$

$\downarrow$

45°

$$= \frac{\frac{\pi}{2} - \frac{\pi}{4}}{\frac{\pi}{2} - 0} = \frac{1}{2}$$

4.2 Distribution functions

r.v. $X : S \to \mathbb{R}$

$P(X = a)$ $P(X < a)$ $P(X \le a)$ $P(X > b)$

-- $P(b < X < a)$

$P(b \le X < a)$

function

F defined on $(-\infty, +\infty)$

$F(t) = P(X \le t) \Rightarrow$ accumulates all prob. of

the values of X up to and including t

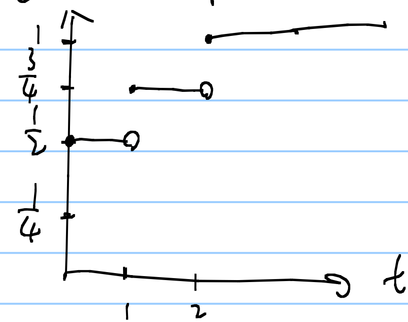
$\Rightarrow$ cumulative distribution function (CDF) of X

CDF $F(t) = p(X \le t)$

$p(X=0) = \frac{1}{2}$

$p(X=1) = \frac{1}{4}$

$p(X=2) = \frac{1}{4}$



$p(X \le \frac{1}{2})$ $\rightarrow$ $t = \frac{1}{2}$

$F(\frac{1}{2})$

Properties of CDFs

1. F : nondecreasing $\quad F(t) \le F(u)$ if $t < u$

2. $\lim_{t \to \infty} F(t) = 1 \qquad$ 3. $\lim_{t \to -\infty} F(t) = 0$

4. F is <u>right continuous</u>

$X \le a \quad F(a)$

$X > a \quad 1 - F(a)$

$X < a \quad F(a^-)$

$X \ge a \quad 1 - P(X < a) = 1 - F(a^-)$

$X = a \quad F(a) - F(a^-)$

$a < X \le b \equiv (X \le b) - (X \le a)$

$= F(b) - F(a)$

$a < X < b \equiv (X < b) - (X \le a)$

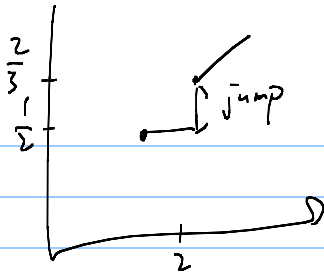
$= F(b^-) - F(a)$

$a \le X \le b = F(b) - F(a^-)$

$a \le X < b = F(b^-) - F(a^-)$

Ex 4.7 $F(x) = \begin{cases} 0 & x<0 \\ \frac{x}{4} & 0 \le x < 1 \\ \frac{1}{2} & 1 \le x < 2 \\ \frac{1}{12}x + \frac{1}{2} & 2 \le x < 3 \\ 1 & x \ge 3 \end{cases}$

X: compound: discrete + continuous



(a) $p(X<2) = \frac{1}{2}$

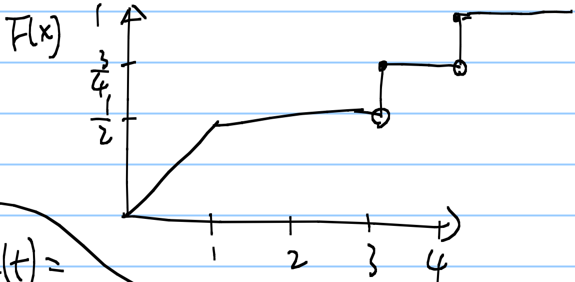
(b) $p(X=2) = F(2) - F(2^-) = \frac{1}{12}\cdot 2 + \frac{1}{2} - \frac{1}{2} = \frac{1}{6}$

(c) $p(1 \le X < 3) = F(3^-) - F(1^-) = \frac{1}{12}\cdot 3 + \frac{1}{2} - \frac{1}{4} = \frac{1}{2}$

(d) $p(X > \frac{3}{2}) = 1 - p(X \le \frac{3}{2}) = 1 - \frac{1}{2} = \frac{1}{2}$

(e) $p(X = \frac{5}{2}) = F(\frac{5}{2}) - F(\frac{5}{2}^-) = 0$

coin
- H → [0, 1] random selection
- T → 3, 4 random selection



$F(t) = p(X \le t)$

(f) $p(2 < X \le 7)$

$= F(7) - F(2)$

$= 1 - \left(\frac{2}{12} + \frac{1}{2}\right) = \frac{1}{3}$

Ex 4.10

$$F(t) = \begin{cases} 0 & t<0 \\ \frac{1}{2}t^2 & 0 \le t < 1 \\ k(4t-t^2) & 1 \le t < 2 \\ 1 & t \ge 2 \end{cases}$$

$X < 2$

$P(X<2) = 1$ $\quad F(2^-) = 1 \Rightarrow k(4t-t^2) = 1$

(a) $(8-4)k = 1 \Rightarrow k = \frac{1}{4}$

(b) A : event between 0.5 and 1.5

$P(0.5 \le X \le 1.5) = F(1.5) - F(0.5^-) = \frac{15}{16} - \frac{1}{8} = \frac{13}{16}$

B : $X > 1$ $\quad P(X>1) = 1 - F(1) = 1 - \frac{3}{4} = \frac{1}{4}$

(c) A B independent?

$AB: \quad 1 < X \leq 1.5$

$$p(AB) = p(1 < X \leq 1.5) = F(1.5) - F(1) = \frac{15}{16} - \frac{3}{4} = \frac{3}{16}$$

$p(A)p(B) \neq p(AB) \Rightarrow A$ & B dependent.

Remark 4.1

F: right-continuous, nondecreasing on $(-\infty, \infty)$

$$\lim_{t\to\infty} F(t) = 1 \qquad \lim_{t\to-\infty} F(t) = 0$$

$\Rightarrow$ find a sample space S, and a r.v. over S

4.3 Discrete r.v.s

$F(t) = P(X \le t)$

- Number of tails in flipping a coin twice

  finite set $\{0, 1, 2\}$

- Number of flips until the first heads

  $\{1, 2, 3, \dots\} \Rightarrow$ countable set

discrete r.v.s

- Amount of next year's rainfall $\{x : x \ge 0\} \Rightarrow$ uncountable set

Def: probability mass function p of X whose set of possible values is $\{x_1, x_2, \cdots\}$ is a function from $\mathbb{R}$ to $\mathbb{R}$ that satisfies

(a) $p(x) = 0$ if $x \notin \{x_1, x_2, \cdots\}$

(b) $p(x_i) = P(X = x_i)$

(c) $\sum_{i=1}^{\infty} p(x_i) = 1$

Ex 4.12

Can
$$p(x) = \begin{cases} c\left(\frac{2}{3}\right)^x & x = 1, 2, 3 \cdots \\ 0 & \text{elsewhere} \end{cases}$$
be a probability mass function?

$$\sum_{x=1}^{\infty} c\left(\frac{2}{3}\right)^x = 1$$

$$c\sum_{x=1}^{\infty}\left(\frac{2}{3}\right)^x = 1$$

$$c\left(\frac{\frac{2}{3}}{1-\frac{2}{3}}\right) = 1 \qquad c = \frac{1}{2}$$

Ex 4.13

X : # of births in a hospital until the first girl is born

$\Rightarrow$ probability mass function and distribution function of X ?

$$p(i) = P(X=i) = \underbrace{\left(\frac{1}{2}\right)^{i-1}}_{\text{boy}} \cdot \underbrace{\frac{1}{2}}_{\text{girl}} = \left(\frac{1}{2}\right)^i$$

$F(t) = ?$ $\quad n-1 \le t < n \quad$ where $\quad n \ge 2$

$$F(t) = p(X \le t) = \left(\tfrac{1}{2}\right)^1 + \left(\tfrac{1}{2}\right)^2 + \cdots \left(\tfrac{1}{2}\right)^{n-1} = \sum_{i=1}^{n-1} \left(\tfrac{1}{2}\right)^i$$

$$= 1 - \left(\tfrac{1}{2}\right)^{n-1}$$

$$F(t) = \begin{cases} 0, & t < 1 \\ 1 - \left(\tfrac{1}{2}\right)^{n-1} & n-1 \le t < n \end{cases}$$

4.4 Expectations of discrete r.v.s

$$E[X] = \sum_{x \in A} x\, p(x)$$

mean, mathematical expectation, expectation of X

$E[X]$, EX, μ_X, μ expected value

Ex 4.1) lottery : pick six different integers between 1 and 49

if all six numbers matched → 1,200,000

if exactly five numbers matched → 800

exactly four numbers matched → 35

⇒ the expected value of the amount a player wins in one game ?

$$p(X=1{,}200{,}000) = \frac{1}{\binom{49}{6}}$$

$$p(X=800) = \frac{\binom{6}{5}\binom{43}{1}}{\binom{49}{6}}$$

$$p(X=35) = \frac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}}$$

$$p(X=1{,}200{,}000) \times 1{,}200{,}000 + p(X=800) \times 800 + p(X=35) \times 35$$

$$= 0.13$$

If the cost per game is 50 cents (0.5)

On the average, lose 39 cents per game

$\Rightarrow$ 10,000 games $\Rightarrow$ loss ≈ 3900

Ex 4.20

The tanks of a country's army are numbered 1 to N.

loses n random tanks to the enemy

If $X_1, X_2, \dots X_n$ are the numbers of the captured tanks,

what is $E(\max X_i)$?

use $E(\max X_i)$ to find an estimate of N

Sol: Let $Y = \max X_i$

$$p(Y=k) = \frac{\binom{k-1}{n-1}}{\binom{N}{n}}$$

$$E[Y] = \sum_{k=n}^{N} k\, p(Y=k)$$

$$= \sum_{k=n}^{N} k \frac{\binom{k-1}{n-1}}{\binom{N}{n}}$$

$$= \frac{1}{\binom{N}{n}} \sum_{k=n}^{N} k \frac{(k-1)!}{(n-1)!\,(k-n)!}$$

$$= \frac{n}{\binom{N}{n}} \sum_{k=n}^{N} \frac{k!}{n!\,(k-n)!}$$

$$= \frac{n}{\binom{N}{n}} \underbrace{\sum_{k=n}^{N} \binom{k}{n}}$$

$\sum_{k=n}^{N}\binom{k}{n}$ is the coefficient of x^n in the polynomial $\sum_{k=n}^{N}(1+x)^k$

$$\sum_{k=n}^{N}(1+x)^k = (1+x)^n\sum_{k=0}^{N-n}(1+x)^k = (1+x)^n\left[\frac{(1+x)^{N-n+1}-1}{(1+x)-1}\right]$$

$$= \frac{1}{x}\left[(1+x)^{N+1}-(1+x)^n\right]$$

$\Rightarrow$ x^n

coefficient of $\Rightarrow \binom{N+1}{n+1}$

$$E[Y] = n\frac{\binom{N+1}{n+1}}{\binom{N}{n}} \qquad N = \frac{n+1}{n}E[Y]-1$$

For example, 12 tanks captured maximum of the numbers is 117

$$N \approx \frac{13}{12} \times 117 - 1 \approx 126$$

Thm 4.1

X : constant c

$$E[X] = c$$

Thm 4.2

$$\underline{E[g(X)] = \sum_{x \in A} g(x)\, p(x)}$$

X : discrete r.v.

A : set of possible values

$p(x)$: prob. mass function

$$E[\underbrace{g(X)}_{Y}] = \sum_{\text{all } y} y \, P_Y(g(X) = y)$$

previous fair dice rolling

X = sum

$Y = g(X) = 2X$

$$p(X=2) = p(Y = 2 \cdot 2) = p(\{1,1\})$$

$$p(X=3) = p(Y = 2 \cdot 3) = p(\{1,2\}, \{2,1\})$$

$$E[Y = g(X) = 2X] = \sum_{\text{all } y} y \, P_Y(g(X) = y)$$

$$= \sum_{\text{all } x} \underbrace{g(x)}_{=y} \cdot p(x)$$

Corollary

$$E[\alpha_1 g_1(x) + \alpha_2 g_2(x) + \cdots \alpha_n g_n(x)]$$

$$= \alpha_1 E[g_1(x)] + \alpha_2 E[g_2(x)] + \cdots \alpha_n E[g_n(x)]$$

$E[X]$ is linear.

$$E[(\alpha X + \beta)] = \alpha E[X] + \beta$$

$$E[X^2] = \sum_{X \in A} X^2 p(x)$$

$$E[X^2-2X+4] = \sum_{x\in A} (x^2-2x+4)\,p(x)$$

$$E[e^X] = \sum_{x\in A} e^x\, p(x)$$

$$E[X\cos X] = \sum_{x\in A} (x\cos x)\,p(x)$$

Ex 4.23

$$p(x) = \begin{cases} \frac{x}{15} & x=1,2,3,4,5 \\ 0 & \text{otherwise} \end{cases}$$

$$E[X(6-X)] = 5\cdot\frac{1}{15} + 8\cdot\frac{2}{15} + 9\cdot\frac{3}{15} + 8\cdot\frac{4}{15} + 5\cdot\frac{5}{15} = 7$$

Ex 4.24

A box contains 10 disks of radii 1, 2, ... and 10.

What is the expected value of the area of a disk randomly selected ?

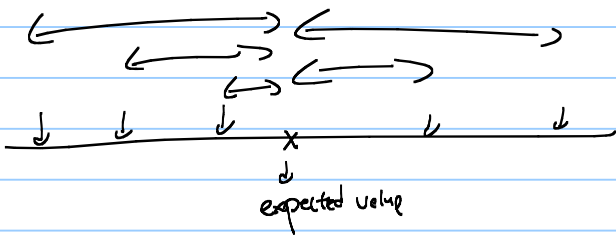
R : radius

$$E[\pi R^2] = \pi E(R^2)$$

$$= \pi \sum_{i=1}^{10} i^2 p(i)$$

$$= \pi \sum_{i=1}^{10} \frac{1}{10}(i^2) = 38.5\pi$$

4.5 Variances and moments of discrete r.v.s



$Var(X)$ σ_X : standard deviation

variance

$$Var[X] = E[(X-\mu)^2] = \sum_{\substack{\text{all } x \\ \in A}} (x-\mu)^2 p(x)$$

$$\sigma_X = \sqrt{Var[X]} = \sqrt{E[(X-\mu)^2]}$$

Ex 4.26

Karen plays two games : Keno and Bolita

Bolita : buys a ticket for \$1

draw a ball at random from a box of 100 balls

if matched ⇒ wins \$75 , otherwise loses

Keno: bet 1$ on a single number that has a 25% chance to win

if win ⇒ return her dollar plus 2 dollars more

otherwise ⇒ they keep the dollar.

$E(B) : \underline{\underline{74}} \cdot 0.01 + (-1)\, 0.99 = -0.25$

$E(K) : 2 \cdot 0.25 + (-1) \cdot 0.75 = -0.25$

$Var(B) : E[(B-\mu)^2] = (74+0.25)^2 \cdot 0.01 + (-1+0.25)^2\, 0.99$

$= 55.69$

$Var(K) : E[(K-\mu)^2] = (2+0.25)^2 (0.25) +$

$(-1+0.25)^2 \cdot 0.75 = 1.6875$

Thm 4.3 $\quad Var[X] = E[X^2] - (E[X])^2$

$$Var[X] = E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2]$$
$$= E[X^2] - E[2\mu X] + E[\mu^2]$$
$$= E[X^2] - 2\mu E[X] + \mu^2$$
$$= E[X^2] - 2E[X]E[X] + (E[X])^2$$
$$= E[X^2] - (E[X])^2 \leftarrow$$

$$Var[X] \geq 0 \Rightarrow E[X^2] \geq (E[X])^2$$

Ex 4.27 the variance of X, the outcome of rolling a fair dice?

$$E(X) = \sum_{x=1}^{6} x\,p(x) = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2}$$

$$E(X^2) = \sum_{x=1}^{6} x^2 p(x) = \frac{91}{6}$$

2nd moment

of X

$$\mathrm{Var}[X] = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

#

Thm 4.4

$\mathrm{Var}[X] = 0$ iff X is a constant with prob. 1

Thm 4.5

$$\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}[X]$$

$$\sigma_{aX+b} = |a|\,\sigma_X$$

$$\mathrm{Var}[aX+b] = E[(aX+b)^2] - (E[aX+b])^2$$
$$= E[a^2X^2+2abX+b^2] - [(aE[X]+b)^2]$$
$$= a^2E[X^2] + 2abE[X] + b^2 - [a^2(E[X])^2 + 2abE[X] + b^2]$$
$$= a^2(E[X^2] - (E[X])^2)$$
$$= a^2\,\mathrm{Var}[X]$$

$$\sigma_{aX+b} = \sqrt{\mathrm{Var}[aX+b]} = |a|\sqrt{\mathrm{Var}[X]} = |a|\,\sigma_X$$

Ex 4.28

$E[X] = 2$, $E[X(X-4)] = 5$

Find the variance and standard deviation of $-4X+12$

$$E[X(X-4)] = E[X^2-4X] = E[X^2] - 4\underset{\substack{\Downarrow \\ 2}}{E[X]} = 5$$

$$\therefore E[X^2] = 8+5 = 13$$

$$\underset{\substack{\shortparallel \\ Var[X]}}{Variance_X} = \underset{\substack{\shortparallel \\ 13}}{E[X^2]} - (\underset{\substack{\shortparallel \\ 2}}{E[X]})^2 = 9$$

$$\sigma_X = \sqrt{Var[X]} = 3$$

$$Var[-4X+12] = (-4)^2 \cdot 9 = 144$$

$$\sigma_{-4X+12} = |-4| \times 3 = 12$$

- Concentrated

$$P(|\underset{=}{Y}-w| \le t) \le P(|\underset{=}{X}-w| \le t)$$

X is more concentrated about w that is Y

Thm 4.6 $E[X] = E[Y] = \mu$

X is more concentrated about μ then is Y

$$Var(X) \le Var(Y)$$

Moment

$E[g(X)]$	
$E[X^n]$	n-th moment of X
$E[\lvert X\rvert^r]$	r-th absolute moment of X
$E[X-c]$	1-st moment of X about c
$E[(X-c)^n]$	n-th moment of X about c
$E[(X-\mu)^n]$	n-th central moment of X

$E[X]$ — first moment (expectation)

$E[X^2]$ — second moment

$$Var[X] = E[X^2] - (E[X])^2$$

4.6 Standardized random variables

$$X^* = \frac{X-\mu}{\sigma}$$

$$E[X^*] = \overset{0}{?} \qquad Var[X^*] = \overset{1}{?}$$

$$E[X^*] = E\left[\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right] = \frac{1}{\sigma}\underbrace{E[X]}_{=\mu} - \frac{\mu}{\sigma} = 0$$

$$Var[X^*] = Var\left[\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right] = \frac{1}{\sigma^2}Var[X] = \frac{1}{\sigma^2}\cdot\sigma^2 = 1$$

A student's grade in a prob. test is 72

history test is 85

history: mean 82, standard deviation 7

prob. " 68 " " 4

standardized history

$$\frac{85-82}{7} = 0.43$$

Standardized prob.

$$\frac{72-68}{4} = 1$$

Ex: Norton owns two appliance stores

store 1: average 13 TVs per week s.d. = 5

store 2: average 7 TVs s.d. = 4

Two applicants for a new position

one → store 1
each for 1 week
the other → store 2

in store 1 → 10 sets
⇒ who should be hired?
2 → 6 sets

$$\frac{10-13}{5} = -0.6 \qquad \frac{6-7}{4} = -0.25$$