5장 연속확률분포 : 확률모형.

4장 이산확률분포 : 확률모형

1. 이산균등분포
2. 이항분포
3. 기하분포
4. 음이항분포
5. 푸아송분포
6. 초기하분포
7. 다항분포

§4.1. 이산균등분포

Example

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

$\omega \mapsto X(\omega)$ = 앞면의 개수

(1) 동전을 한 번 던지는 확률실험에서 표본공간은

$$\Omega = \{ \text{앞면}(H), \text{뒷면}(T) \}$$

이다. 앞면의 수를 확률변수 X라 하면, X의 치역공간은 $R_X = \{0, 1\}$이고 확률질량함수는 다음과 같다.

2개의 값.

$$f_X(x) = \begin{cases} \frac{1}{2}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

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

$\omega \mapsto X(\omega)$ = 주사위 눈의 갯수

(2) 주사위를 한 번 던지는 확률실험에서 표본공간은

$$\Omega = \{ ⚀, ⚁, ⚂, ⚃, ⚄, ⚅ \}$$

이다. 주사위 윗변에 나온 눈의 수를 확률변수 X라 하면, X의 치역공간은 $R_X = \{1, 2, 3, 4, 5, 6\}$이고 확률질량함수는 다음과 같다.

6개의 값.

$$f_X(x) = \begin{cases} \frac{1}{6}, & x = 1, 2, 3, 4, 5, 6 \\ 0, & \text{otherwise} \end{cases}$$

Definition

n개의 값

이산확률변수 X의 치역공간이 $R_X = \{x_1, x_2, \cdots, x_n\}$이고 확률질량함수가

$$f_X(x) = \begin{cases} \frac{1}{n}, & x = x_1, x_2, \cdots, x_n \\ 0, & \text{otherwise} \end{cases}$$

모수(parameter)가 n인

이면, X는 **이산균등분포**(discrete uniform distribution)를 따른다고 하고, 기호로 $X \sim DU(n)$으로 나타낸다.

follow

n개의 값.

Theorem

확률변수 X의 치역공간이 $R_X=\{x_1,x_2,\cdots,x_n\}$이고,

$X\sim DU(n)$이면, 다음이 성립한다.

(1) $E[X]=\dfrac{1}{n}\sum_{i=1}^{n}x_i=\mu_X$

(2) $Var[X]=\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\mu_X)^2$

(3) $m_X(t)=\dfrac{1}{n}\sum_{i=1}^{n}e^{tx_i}$

$=$ $E[e^{tX}]$ 적률생성함수

(proof)

$f_X(x_i)=\frac{1}{n}$

(1) $E[X]=\sum_{i=1}^{n}x_i f_X(x_i)=\sum_{i=1}^{n}x_i\dfrac{1}{n}=\dfrac{1}{n}\sum_{i=1}^{n}x_i$

$x\in R_X$, $f_X(x_i)=\frac{1}{n}$

(2) $Var[X]=\sum_{i=1}^{n}(x_i-E[X])^2 f_X(x_i)=\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\mu_X)^2$

$x\in R$, $f_X(x_i)=\frac{1}{n}$

(3) $m_X(t)=E[e^{tX}]=\sum_{i=1}^{n}e^{tx_i}f_X(x_i)=\dfrac{1}{n}\sum_{i=1}^{n}e^{tx_i}$

$x\in R_X$, $f_X(x_i)=\frac{1}{n}$

대등정리 이용 X.

Example

(1) 동전을 한 번 던지는 확률실험에서 앞면의 수를 X라 할 때, X의 평균과 분산을 구하여라.

$\swarrow$ 2개.

$R_X = \{0, 1\}$

(2) 주사위를 한 번 던지는 확률실험에서 주사위 윗면에 나온 눈의 수를 X라 할 때, X의 평균과 분산을 구하여라.

대등정리 이용 O (*sol*)

$R_X = \{1, 2, 3, 4, 5, 6\}$ $E[X] = \frac{6+1}{2} = 3.5.$ $Var[X] = \frac{6^2 - 1}{12} = \frac{35}{12}$

자연수 1~6

(1) $R_X = \{0, 1\}$이고, $X \sim DU(2)$이므로,

$$E[X] = \frac{1}{2}(0+1) = 0.5,$$

$$Var[X] = \frac{1}{2}\{(0-0.5)^2 + (1-0.5)^2\} = 0.25$$

(2) $R_X = \{1, 2, 3, 4, 5, 6\}$이고, $X \sim DU(6)$이므로,

$$\cancel{E[X] = \frac{1}{6}(1+2+3+4+5+6) = 3.5},$$

$$\cancel{Var[X] = \frac{1}{6}\{(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2\}}$$

$$\cancel{= \frac{1}{6}\{6.25 + 2.25 + 0.25 + 0.25 + 2.25 + 6.25\}} = \frac{35}{12}$$

Corollary

n개의 값

확률변수 X의 치역공간이 $R_X = \{1, 2, \cdots, n\}$이고, X가

자연수 1~n

이산균등분포를 따르면,

$$E[X] = \frac{n+1}{2}, \quad Var[X] = \frac{n^2-1}{12}$$

이다.

$(proof)$

(1) $E[X] = \frac{1}{n}\sum_{x=1}^{n} x = \frac{1}{\cancel{n}} \frac{\cancel{n}(n+1)}{2} = \frac{n+1}{2}$

(2) $E[X^2] = \frac{1}{n}\sum_{x=1}^{n} x^2 = \frac{1}{\cancel{n}} \frac{\cancel{n}(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$

$$Var[X] = E[X^2] - (E[X])^2 = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}$$

$= \frac{2(n+1)(2n+1) - 3(n+1)^2}{12} = \frac{4n^2 + \cancel{6n} + 2 - (3n^2 + \cancel{6n} + 3)}{12}$

$n^2 - 1$

Example 1

1에서 10까지 숫자를 적은 동일한 모양의 카드가 들어 있는 주머니에서 임의로 하나 꺼낸 카드의 번호를 X라 할 때, 다음을 구하여라.

(1) X의 확률질량함수

(2) X의 평균과 분산

(3) $Y=2X-1$의 평균과 분산을 구하여라.

(sol)

10개의 값.

(1) $R_X=\{1,2,\cdots,10\}$이고

1~10

$$f_X(x)=\begin{cases}\frac{1}{10}, & x=1,2,\cdots,10\\ 0, & \text{otherwise}\end{cases}$$

이므로, $X\sim DU(10)$ 이다.

(2) $E[X]=\frac{10+1}{2}=5.5$, $Var[X]=\frac{10^2-1}{12}=\frac{33}{4}$

(3) $E[Y]=E[2X-1]=2E[X]-1=2(5.5)-1=10$

상수배

$Var[Y]=Var[2X-1]=2^2\,Var[X]=2^2\cdot\frac{33}{4}=33$

1. 이산균등분포
2. 이항분포
3. 기하분포
4. 음이항분포
5. 푸아송분포
6. 초기하분포
7. 다항분포

§4.3. 이항분포

■ 베르누이 분포

Definition

확률실험의 결과가 '성공'과 '실패'와 같이 상반되는 두 가지로 나오는 경우, 즉 표본공간이

$$\Omega=\{성공(s), 실패(f)\}$$

이면, 이와 같은 실험을 **베르누이 시행**(Bernoulli trial)이라 한다. 베르누이 시행에서 성공률이 $p\,(0\le p\le 1)$이면, 실패율은 $q=1-p$이다.

만약 확률변수 X를 성공이면 1, 실패이면 0으로 정의하면, X의 치역공간은 $R_X=\{0,1\}$이고 확률질량함수는

$$f_X(x)=\begin{cases} p, & x=1 \\ q, & x=0 \\ 0, & \text{otherwise} \end{cases}$$

이다.

이와 같은 확률질량함수를 갖는 확률변수 X의 확률분포를 모수 p인 **베르누이 분포**(Bernoulli distribution)라 하고, 기호로 $X\sim B(1,p)$로 나타낸다.

$X: \Omega \longrightarrow \mathbb{R}$

$\omega \longmapsto X(\omega)=\begin{cases} 1, & \omega=성공 \\ 0, & \omega=실패 \end{cases}$

binomial

시행횟수

Theorem

$X \sim B(1,p)$이면, 다음이 성립한다.

(1) $E[X]=p$

(2) $Var[X]=p(1-p)=pq$ (단, $q=1-p$)

(3) $m_X(t)=pe^t+q$

$(proof)$ $R_X=\{0,1\}$

실패율

(1) $E[X]=\sum_{x=0,1} x\cdot f_X(x)=0(1-p)+1\cdot p=p$

(2) $E[X^2]=\sum_{x=0,1} x^2\cdot f_X(x)=0^2(1-p)+1^2\cdot p=p$

$Var[X]=E[X^2]-(E[X])^2=p-p^2=p(\underbrace{1-p}_{=q})=pq$

(3) $m_X(t)=E[e^{tX}]=\sum_{x=0,1} e^{tx}\cdot f_X(x)=\underbrace{e^{t\cdot 0}\cdot(1-p)}_{=1-p=q}+\underbrace{e^{t\cdot 1}\cdot p}_{=e^t\cdot p}=pe^t+q$

Example 1

앞면이 나올 가능성이 $\frac{1}{3}$인 찌그러진 동전을 던지는 게임에서, 앞면이 나오면 성공한다고 하자. 이때 앞면이 나오는 사건에 대한 확률질량함수와 평균 그리고 분산을 구하여라.

뒷면 — 실패.

(*sol*) X(앞면)$=1$, X(뒷면)$=0$이면, X의 확률질량함수는

$$f_X(x)=\begin{cases} \frac{1}{3} & , \quad x=1 \\ \frac{2}{3} & , \quad x=0 \\ 0 & , \quad \text{otherwise} \end{cases}$$

이다. $X \sim B(1, \frac{1}{3})$

$$(\therefore)\ E[X]=\frac{1}{3},\quad Var[X]=\frac{1}{3}\cdot\frac{2}{3}=\frac{2}{9}$$

Example 2

[예제 1]에서의 동전을 세 번 던져서 앞면이 나온 횟수를 X라 할 때, X의 확률질량함수를 구하여라.

(*sol*)

확률실험의 표본공간이 $\Omega=\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$이므로,

확률변수 X의 치역공간은 $R_X=\{0,1,2,3\}$이다.

첫 번째 시행에서 확률변수 X_1을 $X_1(H)=1$, $X_1(T)=0$로 정의하고,

두 번째 시행에서 확률변수 X_2을 $X_2(H)=1$, $X_2(T)=0$로 정의하고,

세 번째 시행에서 확률변수 X_3을 $X_3(H)=1$, $X_3(T)=0$로 정의하면

각 시행은 독립적이므로, $X=X_1+X_2+X_3$이다.

$$P(X=0)=P(X_1=0, X_2=0, X_3=0)=P(X_1=0)P(X_2=0)P(X_3=0)=(\frac{2}{3})^3$$

$$\begin{aligned}P(X=1)&=P(X_1=1, X_2=0, X_3=0 \text{ or } X_1=0, X_2=1, X_3=0 \text{ or } X_1=0, X_2=0, X_3=1)\\&=P(X_1=1, X_2=0, X_3=0)+P(X_1=0, X_2=1, X_3=0)+P(X_1=0, X_2=0, X_3=1)\\&=3(\frac{1}{3})^1(\frac{2}{3})^2\end{aligned}$$

$$\begin{aligned}P(X=2)&=P(X_1=1, X_2=1, X_3=0 \text{ or } X_1=1, X_2=0, X_3=1 \text{ or } X_1=0, X_2=1, X_3=1)\\&=P(X_1=1, X_2=0, X_3=0)+P(X_1=1, X_2=0, X_3=1)+P(X_1=0, X_2=1, X_3=1)\\&=3(\frac{2}{3})^2(\frac{2}{3})^1\end{aligned}$$

$$P(X=3)=P(X_1=1, X_2=1, X_3=1)=P(X_1=1)P(X_2=1)P(X_3=1)=(\frac{1}{3})^3$$

$$f_X(x)=\begin{cases}\binom{3}{x}(\frac{1}{3})^x(\frac{2}{3})^{3-x}, & x=0,1,2,3\\ 0, & \text{otherwise}\end{cases}$$

▣ 이항 분포

성공
실패

n번 시행에서

성공횟수 = X

순서상관없이!

Definition

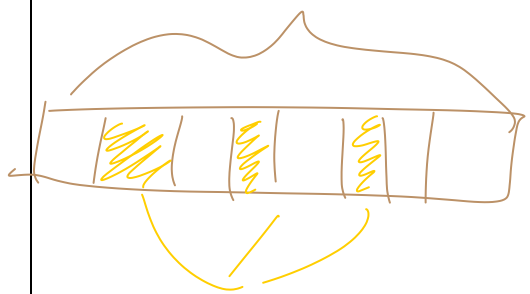
성공할 확률이 p인 베르누이 시행을 독립적으로 n번 반복하는 확률실험에서 성공 횟수를 X라 하면, X의 치역공간은

$$R_X = \{0, 1, 2, \cdots, n\}$$

n+1개의 값.

이고, X의 확률질량함수는

n번



$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, \cdots, n \\ 0, & \text{otherwise} \end{cases}$$

이다.

자리.

이와 같은 확률질량함수를 갖는 확률변수 X의 확률분포를 모수가 n과 p인 이항분포(binomial distribution)라 하고, 기호로 $X \sim B(n, p)$로 나타낸다.

실패 = n−x번

binomial

시행횟수

성공률.

Theorem

$X \sim B(n,p)$이면, 다음이 성립한다.

(1) $E[X] = np$

(2) $Var[X] = npq$ (단, $q = 1-p$)

(3) $m_X(t) = (pe^t + q)^n$

$(proof)$

$R_X = \{0, 1, 2, \cdots, n\}$

※ 이항정리 $(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$

(3) $m_X(t) = E[e^{tX}] = \sum_{x=0}^{n} e^{tx} f_X(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} = (pe^t + (1-p))^n = (pe^t + q)^n$$

a, b, q

t에 관해 미분

$m_X'(t) = n(pe^t + q)^{n-1} pe^t$

$m_X''(t) = n(n-1)(pe^t + q)^{n-2}(pe^t)^2 + n(pe^t + q)^{n-1} pe^t$

(1) $E[X] = m_X'(0) = n(pe^0 + q)^{n-1} pe^0 = np$

(2) $E[X^2] = m_X''(0) = n(n-1)p^2 + np$

$Var[X] = E[X^2] - (E[X])^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$ $= npq$

$\cancel{n^2p^2} - np^2 + np - \cancel{n^2p^2}$

Example 2

$n=3.$

앞면이 나올 가능성이 $\frac{1}{3}$인 찌그러진 동전을 세 번 던져서 앞면이 나온 횟수를 X라 할 때, X의 확률질량함수와 $E[X]$와 $Var[X]$를 구하여라.

$R_X=\{0,1,2,3\}$

(sol)

$n=3$이고 $p=\frac{1}{3}$이므로, X의 확률질량함수는

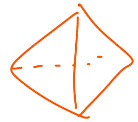
$$f_X(x)=\begin{cases}\binom{3}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{3-x}, & x=0,1,2,3\\ 0, & \text{otherwise}\end{cases}$$

이고, X의 평균과 분산은 다음과 같다.

$$\sum_{x\in R_X} x f_X(x) = \mu_X=E[X]=np=3\cdot\frac{1}{3}=1$$

$$\sigma_X^{\ 2}=Var[X]=npq=3\cdot\frac{1}{3}\cdot\frac{2}{3}=\frac{2}{3}$$

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



$R_X=\{0.1.2.3.4.5\}$

Example 3

$n=5$

1부터 4의 숫자가 적힌 사면체를 5번 던질 때, 숫자 1이 나온 횟수를 X라 한다.

= 성공.

숫자 2.3.4 = 실패.

(1) X의 확률질량함수를 구하여라.
(2) 앞면이 한 번 나올 확률을 구하여라.
(3) 앞면이 많아야 한 번 나올 확률을 구하여라.
(4) 앞면이 적어도 두 번 이상 나올 확률을 구하여라.
(5) X의 평균과 분산을 구하여라.

(*sol*)

(1) $n=5$이고 $p=\dfrac{1}{4}$이므로, X의 확률질량함수는 다음과 같다.

$$f_X(x)=\begin{cases}\dbinom{5}{x}\left(\dfrac{1}{4}\right)^x\left(\dfrac{3}{4}\right)^{5-x}, & x=0,1,2,3,4,5\\ 0, & \text{otherwise}\end{cases}$$

(2) $f_X(1)=\dbinom{5}{1}\left(\dfrac{1}{4}\right)^1\left(\dfrac{3}{4}\right)^4=\dfrac{405}{1024}$

(3) $P(X\le 1)=f_X(0)+f_X(1)=\dbinom{5}{0}\left(\dfrac{1}{4}\right)^0\left(\dfrac{3}{4}\right)^5+\dbinom{5}{1}\left(\dfrac{1}{4}\right)^1\left(\dfrac{3}{4}\right)^4$

$$=\frac{243}{1024}+\frac{405}{1024}=\frac{648}{1024}=0.6328125$$

(4) $P(X\ge 2)=1-P(X\le 1)=1-\dfrac{648}{1024}=\dfrac{376}{1024}$

(5) $\mu_X=np=5\cdot\dfrac{1}{4}=\dfrac{5}{4}$, $\sigma_X{}^2=npq=5\cdot\dfrac{1}{4}\cdot\dfrac{3}{4}=\dfrac{15}{16}$

$X\sim B(n,p)$

Example 4

$n=10.$

정답을 맞출 확률. $=\frac{1}{5}=0.2$

〃 못맞출확률 $=\frac{4}{5}$

5지 선다형으로 주어진 10문제에서 임의로 답안을 선정한다. 다음 이항누적분포표를 이용하여 다음을 구하여라.

(1) 정답을 선택한 문항수가 2개일 확률

(2) 적어도 5문제 이상 정답을 선택할 확률

누적이항분포표. 493쪽.

시행횟수 성공횟수

n	x	p = 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	0	0.5987	0.3487	0.1969	0.1074	0.0562	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.464	0.0233	0.0107
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230

(sol)

(1) $P(X=2)=P(X\le 2)-P(X\le 1)=0.6778-0.3758=0.3020$

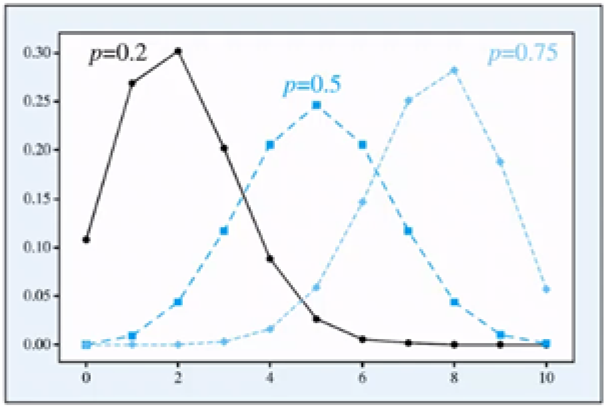
(2) $P(X\ge 5)=1-P(X\le 4)=1-0.9672=0.0328$

Remark

(1) $p<0.5$이면, 이항분포는 왼쪽으로 치우치고 오른쪽으로 긴꼬리 모양을 가진다.

(2) $p>0.5$이면, 이항분포는 오른쪽으로 치우치고 왼쪽으로 긴꼬리 모양을 가진다.

(3) $p=0.5$이면, 이항분포는 n에 관계없이 $m_X=\dfrac{n}{2}$을 중심으로 좌우대칭이고, 이 경우 **대칭이항분포**(symmetric binomial distribution)라 한다.



p에 따른 이항분포의 비교

Theorem

(1) $X_1, X_2, \cdots, X_n$이 독립이고 모두 같은 베르누이 분포 $B(1,p)$를 따르는 확률변수이면, $X=X_1+X_2+\cdots+X_n$는 $B(n,p)$를 따른다. 이때,

$$\mu_X = E[X_1+\cdots+X_n] = \sum_{i=1}^{n} E[X_i] = np$$

$$\sigma_X{}^2 = Var[X_1+\cdots+X_n] = \sum_{i=1}^{n} Var[X_i] = np(1-p)$$

이다.

(2) 두 확률변수 X와 Y가 독립이고 각각

$$X \sim B(m,p),\quad Y \sim B(n,p)$$

인 이항분포를 이룬다면, $Z=X+Y$일 때

$$P(Z=k) = \binom{m+n}{k} p^k (1-p)^{(m+n)-k}$$

이다. 즉, $Z \sim B(m+n,p)$이다.

(proof)

(2) $P(Z=k) = \sum_{x \in R_X} P(X=x, Y=k-x) = \sum_{x \in R_X} P(X=x)P(Y=k-x)$

$$= \sum_{x \in R_X} \binom{m}{x} p^x (1-p)^{m-x} \binom{n}{k-x} p^{k-x} (1-p)^{n-(k-x)}$$

$$= \sum_{x \in R_X} \binom{m}{x}\binom{n}{k-x} p^k (1-p)^{(m+n)-k} = \binom{m+n}{k} p^k (1-p)^{(m+n)-k}$$

Example 5

임의로 선정된 남학생과 여학생이 5지 선다형으로 주어진 두 종류의 문제에서 임의로 답안을 선정한다. 남학생은 A형 문제지로 10문제를 풀고 여학생은 B형 문제지로 5문제를 푼다. 남학생이 정답을 선정한 문제 수를 X, 여학생이 정답을 선정한 문제 수를 Y라 할 때, 다음을 구하여라.

(1) 15문제에서 두 학생이 정답을 선택한 평균 문제 수

(2) 15문제에서 두 학생이 적어도 5문제 이상 정답을 선택할 확률

(sol)

(1) $X \sim B(10, 0.2)$, $Y \sim B(5, 0.2)$이고, X와 Y는 독립이다.

$Z = X + Y \sim B(15, 0.2)$이므로, 평균은 $\mu_Z = 15\left(\frac{1}{5}\right) = 3$

(2) $P(Z \geq 5) = 1 - P(Z \leq 4) = 1 - 0.8358 = 0.1642$

n	x	p = 0.20
15	0	0.0352
	1	0.1671
	2	0.3980
	3	0.6482
	4	0.8358
	5	0.9386

Theorem

$X \sim B(n,p)$일 때, 확률변수 $Y = \frac{X}{n}$는 성공의 비율을 나타내며, Y를 **표본비율**(sample proportion)이라 하고,

$$E[Y] = p, \quad Var[Y] = \frac{pq}{n}$$

이다.

$(proof)$

$$E[Y] = E[\frac{X}{n}] = \frac{1}{n} E[X] = \frac{1}{n} np = p$$

$$Var[Y] = Var[\frac{X}{n}] = \frac{1}{n^2} Var[X] = \frac{1}{n^2} npq = \frac{pq}{n}$$