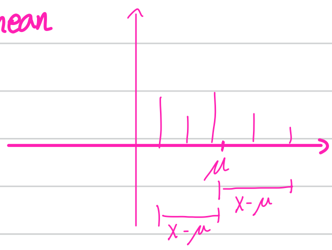
강의 06

Variance

Variance : A quantity of the spread of a PMF

Let $\mu = E[X] =$ mean



distance from the mean, μ

$= X - \mu$

The average of the distances

$\rightarrow E[X-\mu] \xrightarrow{\text{Linearity}} E[X] - \mu = \mu - \mu = 0$

Definition of variance

$$V(X) = E[(X-\mu)^2]$$

$$\rightarrow \sum_x (x-\mu)^2 P_X(x)$$

<Properties>

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

$$= E[X^2] - 2\mu E[X] + \mu^2$$

$$= E[X^2] - 2\mu\mu + \mu^2$$

$$= E[X^2] - E[X]$$

Let $Y = X + b$

$$V(Y) = E[(X+b-\underbrace{\mu-b}_{E[Y]})^2] = E[(X-\mu)^2]$$

$$= V(X)$$

$$V(aY) = E[(aX - a\mu)^2]$$
$$= E[a^2(X-\mu)^2]$$
$$= a^2 E[(X-\mu)^2]$$
$$= a^2 V(X)$$

Variance of Bernolli

$$X = \begin{cases} 1 & w.p.\ p \\ 0 & w.p.\ 1-p \end{cases} \qquad E[X] = p$$

$$V(X) = \underbrace{\sum_{x} (x-\mu)^2 P_X(x)}_{E[(x-\mu)^2]} = \underbrace{(1-\mu)^2 P_X(1)}_{x=1} + \underbrace{(0-\mu)^2 P_X(0)}_{x=0}$$

$E[X] = p = \mu$

$P_X(1) = p$

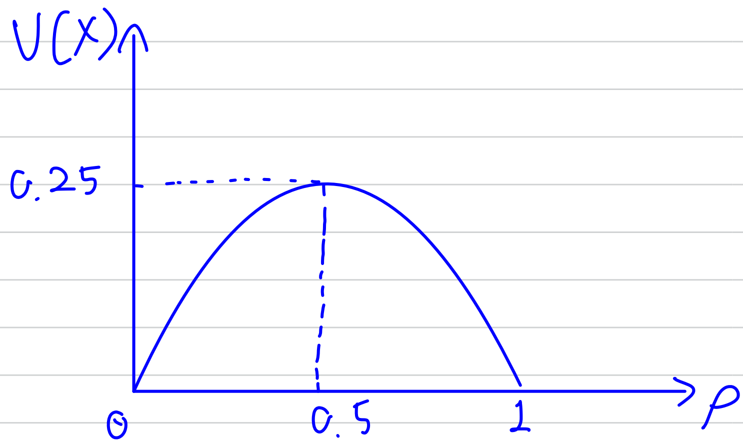
$P_X(0) = 1-p$

$$(1-p)^2 p + (0-p)^2 (1-p)$$

$$= p(1-p)(1-p+p)$$

$$= p(1-p)$$

p에 따른 variance



V of uniform

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

$$= \frac{1}{n+1}\left(0^2 + 1^2 \cdots n^2\right) - \left(\frac{n}{2}\right)^2$$

$$= \frac{1}{n+1} \times \frac{1}{6} \times n(n+1)(2n+1) - \left(\frac{n}{2}\right)^2$$

$$= \frac{1}{12} n(n+2)$$

Conditional PMF

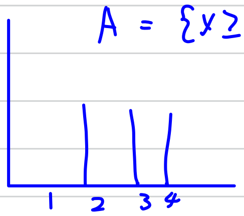
$P_{X|A}(x) = P(X = x \mid A)$

-> Probability X where A occured.

$E[X|A] = \sum x P_{X|A}(x)$

ex) $P_X(x)$

$P_{X|A}(x)$ $A = \{x \geq 2\}$



$E[X] = 2.5$

$V(X) = \frac{5}{4}$

$E[X|A] = 3$

$V(X|A) = \frac{(4-3)^2 + (3-3)^2 + (2-3)^2}{3}$

정의를 따라간다.

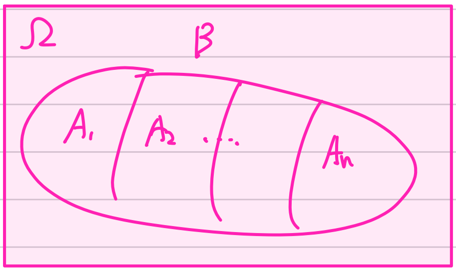
$= \frac{2}{3}$

☆ Conditional variance

다를바없다.

-> 똑같은 계산을 해야함

Total expectation theorem



$$P(B) = P(A_1)P(B|A_1) + \cdots P(A_n)P(B|A_n)$$

$B = \{X = x\}$ 라고 하면

$$p_X(x) = P(A_1)p_{X|A_1}(x) \cdots P(A_n)p_{X|A_n}(x)$$

Σ를 씌우면

$$\sum_x x\left(p_X(x) = P(A_1)p_{X|A_1}(x) \cdots P(A_n)p_{X|A_n}(x)\right)$$

$$\rightsquigarrow \underbrace{\sum_x x\, p_X(x)}_{E[X]} = \underbrace{\sum_x x P(A_1) p_{X|A_1}(x)}_{\sum_x x\, p_{X|A_1}(x) = E[X|A_1]} \cdots \sum_x x P(A_n) p_{X|A_n}(x)$$

$$\boxed{\rightsquigarrow E[X] = P(A_1)E[X|A_1] \cdots P(A_n)E[X|A_n]}$$

Geometric r.v.

$$P_X(x) = (1-p)^{k-1} p$$

첫 번째 동전 던지기에서 T가 나온 경우 다음과 같다

이때 첫 번째 T가 다음에 발생할 event에 영향을 주지 아니한데,

이것을 **Memorylessness** 라고 한다.

Conditioning Geometric $\quad P(X-1=3 \mid X>1) = ?$

첫 번째에 T가 나옴

$$P(X-1=3 \mid X>1) = P(T_2 T_3 H_4 \mid T_1)$$

4번째에 H가 나옴

T_1과 $T_2 T_3 H_4$ 은 독립적이기에 "unconditioning과 동일"하다 하심

$$P(T_2 T_3 H_4)$$

$$= (1-p)^2 p = P_X(3)$$

$$= P_{X-1|X>1}(3)$$

$$\Rightarrow P_{X-1|X>1}(k) = P_X(k)$$

Generalization

$$\Rightarrow P_{X-n|X>n}(k) = P_X(k)$$

Mean of geometric

$$E[X] = \sum_{x=1}^{\infty} x\, p_X(x) = \sum_{x=1}^{\infty} x(1-p)^{x-1} p$$

→ 복잡

→ 다른 방법으로 접근

$$E[X] = E[1+X-1] \xrightarrow{\text{Linearity}} 1+E[X-1]$$

$$\xrightarrow[\text{정리}]{\text{Total expectation}} 1+pE[X-1 \mid X=1] + (1-p)E[X-1 \mid X>1]$$

↳ X=1일 확률
= 첫번째에서 H가 나올 확률
= p

↳ X>1에서 H가 나올 확률
=(X=1) 에서 T가 나올 확률

$$\xrightarrow[\text{정리}]{\text{Conditioning geometric}} 1+p\cdot 0 + (1-p)E[X]$$

첫번째에 H가 나왔을 때
두번째 이상에서 H가 나오는건
Geometric에서 고려X

$$\rightsquigarrow E[X] = 1+(1-p)E[X]$$

$$\rightsquigarrow E[X] = \frac{1}{p}$$

Multiple r.v. and joint PMFs

$X : P_X$
$Y : P_Y$ → marginal PMFs

Joint PMF : $P_{X,Y}(x,y) = P(X=x \text{ and } Y=y)$

y \ x	1	2	3	4
4				
3				2/20
2		1/20	1/20	1/20
1		1/20		

$P_{X,Y}(2,2) = \frac{1}{20}$

$\sum_x \sum_y P_{X,Y}(x,y) = 1$

$P_X(x) = \sum_y P_{X,Y}(x,y)$

Mean of binomial

$$P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$X_i = 1$: i번째 시도에서 성공

$X_i = 0$: otherwise

$X =$ 1번 성공 + 2번성공 ⋯ n번 성공

$= X_1 + X_2 \cdots X_n$

$$E[X] = \overbrace{E[X_1] + E[X_2] \cdots E[X_n]}^{n}$$

$1 \cdot p + 0 \cdot (1-p) = p$

$1 \cdot p + 0 \cdot (1-p)^2 = p$

$\Rightarrow np$