ash

1 Properties of Expectation

1.1 Linearity of Expectation (Additivity)

For any random variables X and Y (regardless of independence):

$$
\begin{aligned}
E(X+Y) &= \sum_x \sum_y (x+y)P(X=x,Y=y) \\
&= \sum_x \sum_y xP(X=x,Y=y) + \sum_x \sum_y yP(X=x,Y=y) \\
&= \sum_x x \underbrace{\left(\sum_y P(X=x,Y=y)\right)}_{=P(X=x)\text{ (Marginal)}} + \sum_y y \underbrace{\left(\sum_x P(X=x,Y=y)\right)}_{=P(Y=y)} \\
&= \sum_x xP(X=x) + \sum_y yP(Y=y) \\
&= E(X) + E(Y)
\end{aligned}
$$

Note: This can be generalized to $E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i)$.

1.2 Expectation of Product (Independence)

If X and Y are **independent**, then $P(X=x,Y=y) = P(X=x)P(Y=y)$. Thus:

$$
\begin{aligned}
E(XY) &= \sum_x \sum_y xyP(X=x,Y=y) \\
&= \sum_x \sum_y xyP(X=x)P(Y=y) \quad (\because \text{Independence}) \\
&= \left(\sum_x xP(X=x)\right)\left(\sum_y yP(Y=y)\right) \\
&= E(X)E(Y)
\end{aligned}
$$

2 Covariance and Correlation

2.1 Covariance

Using the linearity of expectation derived above:

$$\begin{aligned} Cov(X,Y) &= E[(X-E(X))(Y-E(Y))] \\ &= E[(X-\mu_X)(Y-\mu_Y)] \\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X\mu_Y \\ &= E(XY) - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y \\ &= E(XY) - \mu_X\mu_Y \end{aligned}$$

If X, Y are independent, we know $E(XY) = E(X)E(Y) = \mu_X\mu_Y$. Substituting this into the covariance formula:

$$Cov(X,Y) = \mu_X\mu_Y - \mu_X\mu_Y = 0$$

if $Cov(X,Y) = 0$, we say that X and Y are uncorrelated(선형관계가 없다)
(Note: The converse($Cov = 0 \to$ independent) is generally not true.) [536,537]

2.2 Correlation

$$corr(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

3 sum of distributions

3.1 expectation

using Linearity of Expectation

$$E\sum(X_1 + \cdots X_n) = \sum E(X_1 + \cdots X_n)$$

3.2 variance

$$E\left(\sum_{i=1}^{n} X_i\right)^2 = \sum_{i=1}^{n} E(X_i^2) + \sum_{i\neq j} E(X_iX_j) \quad (1)$$

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + \sum_{i\neq j} Cov(X_i, X_j) \quad (2)$$

Proof. First, considering the algebraic expansion of a square:

$$\left(\sum_{i=1}^{n} X_i\right)^2 = \sum_{i=1}^{n} X_i^2 + \sum_{i\neq j} X_i X_j$$

Using the definition of variance $Var(Z) = E[(Z - E[Z])^2]$:

$$\begin{aligned} Var\left(\sum_{i=1}^{n} X_i\right) &= E\left[\left(\sum_{i=1}^{n} X_i - E\left[\sum_{i=1}^{n} X_i\right]\right)^2\right] \\ &= E\left[\left(\sum_{i=1}^{n}(X_i - E[X_i])\right)^2\right] \\ &= \sum_{i=1}^{n} Var(X_i) + \underbrace{\sum_{i\neq j} E\left[(X_i - E[X_i])(X_j - E[X_j])\right]}_{=\sum_{i\neq j} Cov(X_i, X_j)} \end{aligned}$$

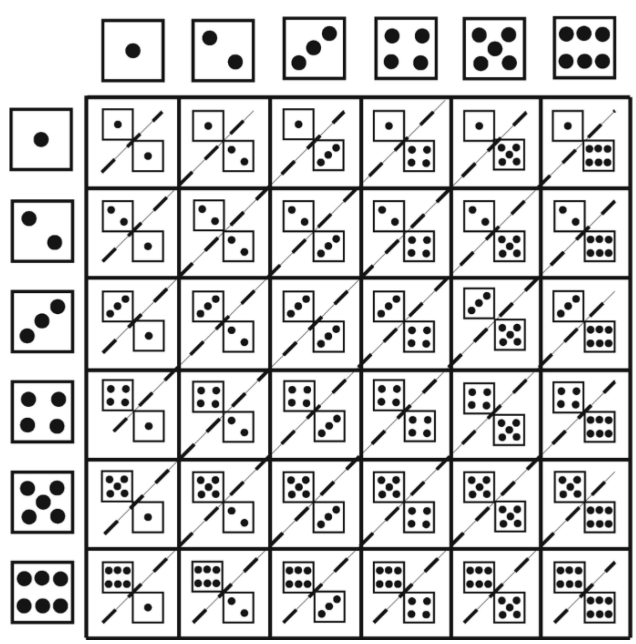
If $X_1, \ldots, X_n$ are **pairwise independent**, then $Cov(X_i, X_j) = 0$ for all $i \neq j$. Therefore:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

□

4 Convolution [540, 541]

- convolution of dice : $f(2), f(3), \cdots, f(12)$



Convolution은 두 다항식을 곱하는 것과 동일한 원리이다.
두 다항식 $f(x), g(x)$를 다음과 같다고 하자.

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = \sum_{i=0}^{n} a_i x^i$$

$$g(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n = \sum_{j=0}^{n} b_j x^j$$

두 다항식의 곱 $h(x) = f(x) \cdot g(x)$은 다음과 같으며, 각 항의 계수는 convolution형태로 나타난다.

$$h(x) = \sum_{k=0}^{2n} \left(\sum_{i+j=k} a_i b_j \right) x^k$$

이때, $i + j = k$를 만족하는 모든 i, j에 대한 계수의 곱($a_i b_j$)을 연산한 후 더한 값이 x^k의 새로운 계수가 된다.

Convolution of probability measures도 이와 유사한 구조를 가진다. continuous case에서도 $x + y = z$위의 density를 적분하는 것과 동일하다. 만약 a_i와 b_j를 각각 확률 $P(X = i)$, $P(Y = j)$로 생각한다면, 위 연산은 두 독립 확률변수의 합 $Z = X + Y$의 분포를 구하는 과정과 정확히 일치한다.

- **Marginal Density**
  If f is the joint density of (X, Y), then the two marginal densities are given by:

  $$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

  $$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

  joint CDF is given by:

  $$F_{X,Y}(x, y) = P(X \le x, Y \le y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v)\, du\, dv$$

  If X and Y are **independent**, then:

  $$f(x, y) = f_X(x) f_Y(y)$$

4.1 Convolution of probability measures [540, 541]

$Z = X + Y$의 확률밀도함수(PDF)를 유도하기 위해, 먼저 Z의 누적분포함수(CDF)를 구한다.

Let X and Y be independent random variables with probability density functions f and g, respectively.

$$\begin{aligned} F_Z(z) = P(X+Y \le z) &= \iint_{x+y\le z} f(x)g(y)\,dx\,dy \\ &= \int_{-\infty}^{\infty} g(y)\left(\int_{-\infty}^{z-y} f(x)\,dx\right)dy \end{aligned}$$

To find the PDF $f_Z(z)$, we differentiate $F_Z(z)$ with respect to z:

$$\begin{aligned} f*g(z) = f_Z(z) = \frac{d}{dz}F_Z(z) &= \int_{-\infty}^{\infty} g(y)\left(\frac{d}{dz}\int_{-\infty}^{z-y} f(x)\,dx\right)dy \\ &= \int_{-\infty}^{\infty} g(y)f(z-y)\,dy \qquad (1) \end{aligned}$$

□

4.2 Convolution of Discrete Case

Assume that you have a pair (X, Y) of independent discrete random variables X and Y, with PMFs $f(x)$ and $g(y)$ respectively. Their joint probability mass function is given by:

$$p(x,y) = P(X=x, Y=y) = P(X=x)P(Y=y) = f(x)g(y)$$

Using the indicator function I_A:

$$I_A(x,y) = \begin{cases} 1 & \text{if } (x,y)\in A \\ 0 & \text{if } (x,y)\notin A \end{cases}$$

$$P((X,Y)\in A) = \sum_{(x,y)\in A} p(x,y) = \sum_x\sum_y p(x,y)\cdot I_A(x,y)$$

We want to know the distribution of $Z = X+Y$.
Let $A = \{(x,y) \mid x+y=z\}$. Then:

$$\begin{aligned} P(Z=z) = P((X,Y)\in A) &= \sum_y p(z-y,y) \\ &= \sum_y P(X=z-y)P(Y=y) \\ &= \sum_y f(z-y)g(y) \end{aligned}$$

4.3 Convolution on Finite Groups

Assume that the independent random variables X and Y take values on a finite group G with group operation $\circ$. Let X and Y be distributed according to τ and η respectively.

$$p(x,y) = P(X=x, Y=y) = P(X=x)P(Y=y) = \eta(x)\tau(y)$$

Let $Z = Y \circ X$ and $A = \{(x, y) \in G \times G \mid y \circ x = z\}$. Since $y \circ x = z \iff y = z \circ x^{-1}$, we have:

$$\begin{aligned} P(Z = z) = P((X, Y) \in A) &= \sum_{(x,y)\in A} p(x, y) \\ &= \sum_{x\in G} p(x, z \circ x^{-1}) \\ &= \sum_{x\in G} P(Y = z^{-1} \circ x)P(X = x) \\ &= \sum_{x\in G} \eta(z \circ x^{-1})\tau(x) \end{aligned}$$

5 Gamma distribution

- Gamma function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$$
$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$$

- Gamma distribution

$$f(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad x > 0$$

5.1 Gamma Distribution : Expectation and Variance

- $E(X)$

$$\begin{aligned}
E(X) &= \int_0^\infty x f(x;\alpha,\beta)\,dx \\
&= \int_0^\infty x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\,dx \\
&= \int_0^\infty \frac{\beta^\alpha x^\alpha e^{-\beta x}}{\Gamma(\alpha)}\,dx \\
&= \int_0^\infty \alpha \frac{\beta^\alpha x^\alpha e^{-\beta x}}{\Gamma(\alpha+1)}\,dx \quad (\because \Gamma(\alpha+1) = \alpha\Gamma(\alpha)) \\
&= \frac{\alpha}{\beta} \underbrace{\int_0^\infty \frac{\beta^{\alpha+1} x^\alpha e^{-\beta x}}{\Gamma(\alpha+1)}\,dx}_{\text{PDF of Gamma distribution}(\alpha+1,\beta)} \\
&= \frac{\alpha}{\beta} \times 1 \\
&= \frac{\alpha}{\beta}
\end{aligned}$$

- $E(X^2)$

$$\begin{aligned}
E(X^2) &= \int_0^\infty x^2 \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\,dx \\
&= \int_0^\infty \frac{1}{\beta} \frac{\beta^{\alpha+1} x^{\alpha+1} e^{-\beta x}}{\Gamma(\alpha)}\,dx \\
&= \int_0^\infty \frac{\alpha(\alpha+1)}{\beta} \frac{\beta^{\alpha+1} x^{\alpha+1} e^{-\beta x}}{\Gamma(\alpha+2)}\,dx \quad (\because \Gamma(\alpha+2) = \alpha(\alpha+1)\Gamma(\alpha)) \\
&= \frac{\alpha(\alpha+1)}{\beta^2} \underbrace{\int_0^\infty \frac{\beta^{\alpha+2} x^{\alpha+1} e^{-\beta x}}{\Gamma(\alpha+2)}\,dx}_{\text{PDF of Gamma distribution}(\alpha+2,\beta)} \\
&= \frac{\alpha(\alpha+1)}{\beta^2} \times 1 \\
&= \frac{\alpha^2}{\beta^2} + \frac{\alpha}{\beta^2}
\end{aligned}$$

- $Var(X) = E(X^2) - (EX)^2$

$$\begin{aligned}
Var(X) =& \frac{\alpha^2}{\beta^2} + \frac{\alpha}{\beta^2} - \left(\frac{\alpha}{\beta}\right)^2 \\
=& \frac{\alpha}{\beta^2}
\end{aligned}$$

6 Weak Law of large number

6.1 Markov inequality and Chebyshev inequality

- Markov's Inequality

$$P(X \geq a) \leq \frac{E[X]}{a} \quad \text{(for } X \geq 0, a > 0\text{)}$$

Proof. Let indicator function $I_{\{X \geq a\}}$ be defined as:

$$I = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{if } X < a \end{cases}$$

Since $X \geq 0$ and $a > 0$, the following inequality holds:

$$\begin{aligned} I_{\{X \geq a\}} &\leq \frac{X}{a} \quad \text{(since } X \geq 0\text{)} \\ E\left[I_{\{X \geq a\}}\right] &\leq E\left(\frac{1}{a}X\right) \\ P(X \geq a) &\leq \frac{E[X]}{a} \end{aligned}$$

□

- Chebyshev inequality

$$P(\mid X - \mu \mid \leq k) \geq \frac{\sigma^2}{k^2}$$

Proof. by Markov indequality:

$$P(\mid X - \mu \mid \leq k) = P((X-\mu)^2 \leq k^2) \geq \frac{1}{k^2}E(X-\mu)^2 = \frac{1}{k^2}Var(X)$$

□

6.2 weak LLN

Let $X_1, X_2 \cdots$ be iid random variables with a finite first moment, $EX_i = \mu$ then:

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu$$

as $n \to \infty$

Proof. Denote $\mu = E[X], \sigma^2 = \mathrm{Var}(X)$, and let $S_n = X_1 + \cdots + X_n$. Then:

$$E[S_n] = E[X_1] + \cdots + E[X_n] = n\mu \quad \text{and} \quad \mathrm{Var}(S_n) = n\sigma^2$$

Therefore, by Chebyshev's inequality:

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) &= P\left(|S_n - n\mu| > n\epsilon\right) \\ &\leq \frac{n\sigma^2}{(n\epsilon)^2} \\ &= \frac{\sigma^2}{n\epsilon^2} \to 0 \quad \text{as } n \to \infty \end{aligned}$$

□

[534]