

다차원 확률변수의 확률분포

이산형 이차원 확률 변수

정의

두 확률변수 X, Y 의 순서쌍인 (X, Y) 가 가질 수 있는 순서쌍들의 집합이

$$\{(X_j, Y_k) | j = 1, 2, \dots, k = 1, 2, \dots\}$$

일 때, 각 순서쌍에 그 순서쌍을 가질 확률을 대응시키는 함수

$$f(x_j, y_k) = P(X = x_j, Y = y_k)$$

로 정의된 함수 f 를 이산형 bivariate random vector (X, Y) 의 확률 밀도함수라고 한다.

이산형 pdf의 정의

$$(a) f(x, y) \geq 0 \quad \forall x, y : -\infty < x < \infty, -\infty < y < \infty$$

$$f(x, y) = 0 \quad \forall (x, y) \neq (x_j, y_k) \quad (j = 1, 2, \dots, k = 1, 2, \dots)$$

$$(b) \sum_x \sum_y f(x, y) = \sum_{x=1}^{\infty} \sum_{y=1}^{\infty} f(x_j, y_k) = 1$$

$$(c) \sum_{x:a \leq x \leq b} \sum_{y:c \leq y \leq d} f(x, y) = P(a \leq X \leq b, c \leq Y \leq d)$$

연속형 이차원 확률 변수

정의

두 개의 확률변수 X, Y 가 모두 실수 구간의 값을 가질 수 있고 그에 관한 확률이 적분으로 주어질 때, 두 확률변수의 순서쌍인 (X, Y) 를 연속형의 bivariate random vector라고 하며 (X, Y) 에 관한 확률을 정해주는 함수, 즉

$$\int_{x:a \leq x \leq b} \int_{y:c \leq y \leq d} f(x, y) dy dx = P(a \leq X \leq b, c \leq Y \leq d) \quad (a < b, c < d)$$

인 함수 f 를 확률밀도함수라고 한다.

연속형 pdf의 정의

- (a) $f(x, y) \geq 0 \forall x, y : -\infty < x < \infty, -\infty < y < \infty$
 (b) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dy dx = 1$
 (c) $\int_a^b \int_c^d f(x, y) dy dx = P(a \leq X \leq b, c \leq Y \leq d) \ (a < b, c < d)$

Marginal PD , Joint PD

두 확률변수 X, Y 가 주어지면, 이들의 순서쌍(X, Y)의 분포를 Joint PD라고 하고, X 의 분포와 Y 의 분포는 각 확률변수의 Marginal PD라고 한다.

Joint PD로 부터 Marginal PD를 얻는 방법은 다음과 같다.

$$P(a \leq X \leq b) = P(a \leq X \leq b, -\infty < Y < +\infty)$$

$$= \begin{cases} \sum_{x:a \leq x \leq b} \sum_y f(x, y) & (if, (X, Y) \text{ are discrete}) \\ \int_a^b \int_{-\infty}^{+\infty} f(x, y) dy dx & (if, (X, Y) \text{ are continuous}) \end{cases}$$

따라서,

$$f_1(x) = \begin{cases} \sum_y f(x, y) \\ \int_{-\infty}^{+\infty} f(x, y) dy \end{cases}$$

로 정의된 함수 f_1 에 대하여

$$P(a \leq X \leq b) = \begin{cases} \sum_{x:a \leq x \leq b} f_1(x) & (if, (X, Y) \text{ are discrete}) \\ \int_a^b f_1(x) dx & (if, (X, Y) \text{ are continuous}) \end{cases}$$

y 에 대한 함수도 같은 방식으로 $f_2(y)$ 를 통해 정의할 수 있다.

X, Y 의 Joint CDF

$$F(x, y) = P(X \leq x, Y \leq y), -\infty < x < +\infty, -\infty < y < +\infty$$

(이산형)

X, Y 가 가질 수 있는 값들이 각각 $x_1 < x_2 < \dots < x_m < \dots, y_1 < y_2 < \dots < y_n < \dots$ 일 때,

$$F(x_m, y_n) = P(X \leq x_m, Y \leq y_n) = \sum_{j=1}^m \sum_{k=1}^n f(x_j, y_k)$$

$$f(x_j, y_k) = P(X = x_j, Y = y_k) \\ = \{F(x_j, y_k) - F(x_{j-1}, y_k)\} - \{F(x_j, y_{k-1}) - F(x_{j-1}, y_{k-1})\}$$

(연속형)

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(t, u) du dt$$

f 가 연속인 점 (x, y) 에서 $f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$

Marginal CDF도 Joint CDF를 통해 얻을 수 있다.

$$F_1(x) = \lim_{y \rightarrow \infty+} F(x, y)$$

$$F_2(y) = \lim_{x \rightarrow \infty+} F(x, y)$$

Joint CDF의 성질

오른쪽 연속성

$$\lim_{h \rightarrow 0+} F(x+h, y) = F(x, y)$$

$$\lim_{k \rightarrow 0+} F(x, y+k) = F(x, y)$$

결합확률분포의 특성치

Joint PD에 대한 기댓값

$$E[g(X, Y)] = \begin{cases} \sum_x \sum_y g(x, y) f(x, y) \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dy dx \end{cases}$$

가 실수로 정의되면 그 값을 $g(X, Y)$ 의 기댓값이라 하고, 기호로는 $E[g(X, Y)]$ 로 나타낸다.

기댓값의 성질

(1)

$$E[c_1 g_1(X, Y) + c_2 g_2(X, Y)] = c_1 E[g_1(X, Y)] + c_2 E[g_2(X, Y)]$$

(2)

$$g_1(X, Y) \leq g_2(X, Y)$$

이면,

$$E[g_1(X, Y)] \leq E[g_2(X, Y)]$$

기댓값의 정의를 통해 하나의 확률변수에만 의존하는 함수의 기댓값은 Marginal PD에 대한 기댓값임을 알 수 있다.

$$E[g(X)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x) f(x, y) = \int_{-\infty}^{+\infty} g(x) \left\{ \int_{-\infty}^{+\infty} f(x, y) dy \right\} dx$$

이때, X의 Marginal PDF는 $f_1(x) = \int_{-\infty}^{+\infty} f(x, y) dy$ 이므로, g(X)의 기댓값은 Marginal PD에 대한 기댓값으로 주어진다.

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) \int_{-\infty}^{+\infty} f(x, y) dy dx = \int_{-\infty}^{+\infty} g(x) f_1(x) dx$$

Covariance와 Correlation coefficient

(Covariance)

확률변수 X의 평균과 표준편차를 각각 μ_1 과 σ_1 , Y의 평균과 표준편차를 각각 μ_2 과 σ_2 라고 할 때, $(X - \mu_1)(Y - \mu_2)$ 의 기댓값을 X와 Y의 Covariance라고 하고, $Cov(X, Y)$, $\sigma_{X,Y}$, $\sigma_{1,2}$ 로 나타낸다.

$$Cov(X, Y) = E[(X - \mu_1)(Y - \mu_2)]$$

(Correlation coefficient) ρ

두 확률변수 X, Y의 표준편차 σ_1 , σ_2 가 0이 아닐 때, X와 Y의 Covariance를 두 표준편차의 곱으로 나눈 수를 X와 Y의 Correlation coefficient라고 한다.

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_1 \sigma_2}$$

의미는 다음과 같다.

Covariance는 두 변수가 각각의 기준인 평균으로부터 변화하는 방향과 변화하는 양에 대한 정보를 주는 특성치이다.

한 변수가 그 평균보다 커질 때 다른 변수도 평균보다 커지는 확률이 많다면 $(X - \mu_1)(Y - \mu_2)$ 의 값이 양수로서 큰 값이 될 것을 기대할 수 있다.

즉, 공분산의 값이 큰 양수일 것이고, 서로가 반대 방향으로 변화하는 확률이 클 경우에는 공분산의 값이 음수로 주어질 것이다.

이와 달리 각각의 변화가 서로에게 영향을 주지 않는다면, 0에 가까운 값이 될 것이다.

Correlation coefficient는 Covariance가 두 변수의 변화 관계의 방향과 크기를 표현해주지만, 수치의 크기가 측정단위에 의존하는 문제가 있어 측정 단위에 기인한 규모의 차이인지 변화량의 크기에 기인한

것인지를 구분하기 어려운 문제가 있다.

따라서, 각 변수의 측정 단위로 인한 규모 변화를 방지하고자, 표준편차로 나누어 줌으로써 표준화된 Covariance를 의미한다.

- 참고사항

회귀계수는 두 변수 간에 원인과 결과의 직선적 관계를 설명하며, 상관계수는 두 변수 간 직선적 상관관계의 강도만을 나타낸다. 따라서 상관계수를 가지고 원인과 결과의 관계를 설명할 수 없으며 해서도 안된다. 예를 들면 벼 시비실험에서 분얼수(새끼 친 수)와 간장 간에는 높은 정상관이 있으나, 이것을 가지고 분얼수가 많아지는 것은 간장이 길어진 효과 때문이라든지, 간장이 길어진 것은 분얼수가 많아진 효과 때문이라고 해석해서는 안 된다. 왜냐하면 분얼수와 간장이 함께 증가한 것은 시비량을 증가시켰기 때문이다. 이와 같이 두 변수 사이에 제3의 요인이 관여하여 생긴 상관관계를 무의미상관(nonsense correlation)이라고 한다.

상관계수가 0일 때 이것은 두 변수 간에 단지 직선적 관계가 없다는 뜻이며, 두 변수 사이에 관계가 전혀 없다는 의미는 아니다. 두 변수 간에 직선적 관계가 없어도 곡선적 관계는 얼마든지 있을 수 있다.

회귀분석이 유의해도 그것이 원인과 결과의 관계가 아닌 경우도 있다. 예컨대 연도별 농가인구와 농업소득과의 관계는 비록 통계적으로 유의하더라도 실제로는 별다른 의미가 없는 것이다.

(*출처: [상관분석\(correlation analysis\)](#))

Covariance의 성질과 계산 공식

(1) $Cov(X, Y) = Cov(Y, X), Cov(X, X) = Var(X)$

(2) $Cov(aX + b, cY + d) = acCov(X, Y), (a, b, c, d \text{ are constants})$

(3) $Cov(X, Y) = E(XY) - E(X)E(Y)$

(2)에 대한 증명

$$V = aX + b, W = cY + d$$

라고 하면

$$V - E(V) = a(X - \mu_1), W - E(W) = c(Y - \mu_2)$$

$$\begin{aligned} Cov(V, W) &= E[\{V - E(V)\}\{W - E(W)\}] \\ &= E[a(X - \mu_1)c(Y - \mu_2)] \\ &= acE[(X - \mu_1)(Y - \mu_2)] \\ &= acCov(X, Y) \end{aligned}$$

Correlation coefficient(ρ)의 성질

(1)

$$Var\left(\frac{Y - \mu_2}{\sigma_2} - \rho \frac{X - \mu_1}{\sigma_1}\right) = 1 - \rho^2$$

$$(2) -1 \leq \rho \leq 1$$

$$(3) \begin{cases} \rho = 1 \Leftrightarrow P\left(\frac{Y - \mu_2}{\sigma_2} = \frac{X - \mu_1}{\sigma_1}\right) = 1 \\ \rho = -1 \Leftrightarrow P\left(\frac{Y - \mu_2}{\sigma_2} = -\frac{X - \mu_1}{\sigma_1}\right) = 1 \end{cases}$$

(1)은 $var(Z) = E(Z^2) = 1$ 임을 통해 증명이 되며, (1)을 통해 (2)가 성립됨을 알 수 있고, (3)은 $Var(X) = 0$ 인 경우로써, $P(X = \mu) = 1$ 을 통해 증명이 가능하다.

위의 성질을 통해 Correlation coefficient의 절댓값이 커질수록 (X,Y)의 분포는 직선 $\frac{y - \mu_2}{\sigma_2} = \rho \frac{x - \mu_1}{\sigma_1}$ 주위에 가깝게 분포되어 나타날 것이다. (비례관계)

이러한 뜻에서 상관계수는 두 변수 사이의 직선 관계를 나타내는 특성치로 해석할 수 있다.

Joint moment

Joint moment

$E(|X^r Y^s|) < +\infty$ 일 때,

$$E(X^r Y^s) = \begin{cases} \sum_x \sum_y x^r y^s f(x, y) \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^r y^s f(x, y) dy dx \end{cases}$$

를 (X,Y)의 (r,s)차의 (r,s)번째 joint moment라고 한다.

기호 : $m_{r,s}(X, Y)$ 또는 $m_{r,s}$

직선 관계의 특성을 나타내는 Covariance, Correlation coefficient처럼 Joint PD에 대한 특성을 나타낸다.

하나의 PD에서와 같이 Joint moment generating function은 유사한 방식으로 전개된다.

$$e^{t_1 X} e^{t_2 Y} = \sum_{r=0}^{\infty} \frac{(t_1 X)^r}{r!} \sum_{s=0}^{\infty} \frac{(t_2 Y)^s}{s!}$$

$$E(e^{t_1 X + t_2 Y}) = \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{E(X^r Y^s)}{r! s!} t_1^r t_2^s$$

Joint MGF는 Joint Moment를 생성하며, Joint PD를 결정하는 성질을 갖는다.

Joint MGF 정의

0을 포함하는 열린구간들의 t_1, t_2 값에 대하여 $E(e^{t_1 X + t_2 Y})$ 가 실수일 때, 함수

$$M(t_1, t_2) = E(e^{t_1 X + t_2 Y}), -h_1 < t_1 < h_1, -h_2 < t_2 < h_2 (\exists h_1 > 0, h_2 > 0)$$

를 Joint MGF라고 한다.

Joint MGF 성질

(1) 이차원 확률변수 (X, Y) 의 Joint MGF가 존재하면, 즉

$$M(t_1, t_2) = E(e^{t_1 X + t_2 Y}) < +\infty \\ -h_1 < t_1 < h_1, -h_2 < t_2 < h_2 (\exists h_1 > 0, h_2 > 0)$$

이면, (X, Y) 의 모든 Joint moment가 존재하고,

$$E(X^r Y^s) = \left[\frac{\partial^{r+s}}{\partial t_1^r \partial t_2^s} M(t_1, t_2) \right]_{t_1=0, t_2=0} \\ E(e^{t_1 X + t_2 Y}) = M(t_1, t_2) = \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{E(X^r Y^s)}{r! s!} t_1^r t_2^s \\ -h_1 < t_1 < h_1, -h_2 < t_2 < h_2 (\exists h_1 > 0, h_2 > 0)$$

(2) (분포 결정성)

두 이차원 확률변수 (X_1, x_2) 와 (Y_1, Y_2) 의 Joint MGF가 0을 포함하는 열린구간들에서 일치 즉,

$$M_{X_1, X_2}(t_1, t_2) = M_{Y_1, Y_2}(t_1, t_2) \forall t_i : -h_i < t_i < h_i (\exists h_i > 0) (i = 1, 2)$$

이면, (X_1, X_2) 와 (Y_1, Y_2) 의 PD가 일치한다. 즉, Joint PDF와 Joint CDF가 일치한다.

Joint CGF 정의

Joint MGF가 $M(t_1, t_2)$ 가 존재하면,

$$C(t_1, t_2) = \log M(t_1, t_2) = \log E(e^{t_1 X + t_2 Y}), -h_i < t_i < h_i (\exists h_i > 0) (i = 1, 2)$$

를 Joint CGF(cumulant generating function)라고 한다.

$$C(t_1, t_2) = \sum_{r=0, r+s \geq 1}^{\infty} \sum_{s=0}^{\infty} \frac{C^{(r,s)}(0, 0)}{r! s!} t_1^r t_2^s$$

$$C^{(r,s)}(0,0) = \left[\frac{\partial^{r+s}}{\partial t_1^r \partial t_2^s} \right]_{t_1=0, t_2=0}$$

$C^{(r,s)}(0,0)$ 을 (X,Y) 의 $(r+s)$ 차의 (r,s) 번째 joint cumulant,
기호: $c_{r,s}(X,Y)$ 또는 $c_{r,s}$

joint cumulant와 joint moment의 관계를 통해

$$c_{1,0} = m_{1,0} = E(X)$$

$$c_{0,1} = m_{0,1} = E(Y)$$

$$c_{2,0} = m_{2,0} - (m_{1,0})^2 = Var(X)$$

$$c_{1,1} = m_{1,1} - m_{1,0}m_{0,1} = Cov(X,Y)$$

$$c_{0,2} = m_{0,2} - (m_{0,1})^2 = E(Y^2) - [E(Y)]^2 = Var(Y)$$

Marginal MGF, CGF

$$M_X(s) = M_{X,Y}(s,0), C_X(s) = C_{X,Y}(s,0)$$

$$M_Y(t) = M_{X,Y}(0,t), C_Y(t) = C_{X,Y}(0,r)$$

조건부분포와 조건부기댓값

Conditional PDF의 정의

확률변수 X,Y 가 모두 이산형일 때, $X=x$ 인 조건에서 Y 에 관한 CD가

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

로 정의되기 위해서는 $P(X = x) > 0$ 이어야 한다. 즉 이산형의 경우에는 변수가 가질 수 있는 값이 주어진 조건에서의 CD는 언제나 정의된다.

이차원 확률변수 (X,Y) 가 이산형으로서 (X,Y) 의 Joint probability가 $f_{1,2}(x,y)$ 이고, X 의 Marginal pmf가 $f_1(x)$ 일 때, $X = x$ 인 condition에서 Y 의 가능한 값에 conditional probability를 대응시키는 함수

$$f_{1,2} = \frac{f_{1,2}(x,y)}{f_1(x)} = \frac{P(X = x, Y = y)}{P(X = x)} = P(Y = y|X = x)$$

를 $X = x$ 인 조건에서 Y 의 conditional pdf라고 한다.

이산형인 경우의 conditional pdf

$$f_{2|1}(y|x) = \frac{f_{1,2}(x, y)}{f_1(x)} = P(Y = y|X = x) \quad (x : f_1(x) > 0)$$

$$(1) f_{2|1}(y|x) \geq 0 \quad \forall y : -\infty < y < +\infty$$

$$f_{2|1}(y|x) = 0 \quad \forall y : y \neq y_k (k = 1, 2, \dots)$$

$$(2) \sum_y f_{2|1}(y|x) = 1$$

$(X = x)$ event로 구성된 sample space에서 y 의 분포로 해석할 수 있다.

연속형인 경우의 conditional pdf

$$f_{2|1}(y|x) = \frac{f_{1,2}(x, y)}{f_1(x)} \quad (x : f_1(x) > 0)$$

$$P(c \leq Y \leq d|X = x) = \lim_{h \rightarrow 0+} P(c \leq Y \leq d|x \leq X \leq x + h) = \int_c^d f_{2|1}(y|x) dy$$

$$(1) f_{2|1}(y|x) \geq 0 \quad \forall y : -\infty < y < +\infty$$

$$(2) \int_{-\infty}^{+\infty} f_{2|1}(y|x) dy = 1$$

$$(3) P(c \leq Y \leq d|X = x) = \int_c^d f_{2|1}(y|x) dy$$

$P(c \leq Y \leq d, X = x) = 0, P(X = x) = 0$ 이므로, 양수 h 에 대한 범위 극한을 통해 정의한다.

conditional pdf의 성질

$$P(a \leq X \leq b, c \leq Y \leq d) = \begin{cases} \sum_{a \leq x \leq b} P(c \leq Y \leq d|X = x) f_1(x) & (if \text{ discrete}) \\ \int_a^b P(c \leq Y \leq d|X = x) f_1(x) & (if \text{ continuous}) \end{cases}$$

실직선 위의 구간에 분포되는 것을 $X = x$ 인 조건에서 Y 의 conditional pdf는 간략히 조건부분포라고 한다.

Conditional Mean

$X = x$ 인 조건에서 Y 의 CDF가 $f_{2|1}(y|x)$ 일 때,

$$\mu_{2|1}(x) = E(Y|X = x) = \begin{cases} \sum y f_{2|1}(y|x) & (if \text{ discrete}) \\ \int_{-\infty}^{+\infty} y f_{2|1}(y|x) dy & (if \text{ continuous}) \end{cases}$$

우항이 실수로 정의될 때, 좌항을 $X = x$ 일때, Y 의 conditional mean이라 한다.

Conditional Expectation

$$E(g(X, Y)|X = x) = \begin{cases} \sum_y g(x, y) f_{2|1}(y|x) & (if \text{ discrete}) \\ \int_{-\infty}^{+\infty} g(x, y) f_{2|1}(y|x) dy & (if \text{ continuous}) \end{cases}$$

conditional expected value of Y , given $X = x$

Conditional Variance

$$\begin{aligned} Var(Y|X = x) &= E[(Y - \mu_{2|1}(x))^2|X = x] \\ &= \begin{cases} \sum (y - \mu_{2|1}(x))^2 f_{2|1}(y|x) \\ \int_{-\infty}^{+\infty} (y - \mu_{2|1}(x))^2 f_{2|1}(y|x) dy \end{cases} \end{aligned}$$

Conditional Expectation의 성질 및 유도식

- (1) $E[c_1 g_1(Y) + c_2 g_2(Y)|X = x] = c_1 E[g_1(Y)|X = x] + c_2 E[g_2(Y)|X = x]$
- (2) $E[c(X)g(Y)|X = x] = c(X)E[g(Y)|X = x]$ * X 에 대한 확률변수이나 기댓값은 y 에 대해 적분을 수행하므로 자명
- (3) $g_1(Y) \leq g_2(Y)$ 이면, $E[g_1(Y)|X = x] \leq E[g_2(Y)|X = x]$

$$Var(Y|X = x) = E(Y^2|X = x) - \{E(Y|X = x)\}^2$$

conditional mean $\mu_{2|1}(x) = E(Y|X = x) = E(Y|X)$ 는 확률변수 X 의 값 x 에 따라 그 값이 정해지는 함수이다. 따라서, $\mu_{2|1}(x)$ 는 X 의 분포에 따라 $\mu_{2|1}(x)$ 값들을 가지는 새로운 확률변수이다. 이러한 확률변수를 X 가 주어진 조건에서 Y 의 Conditional mean이라 한다.

마찬가지로 conditional variance는 $\sigma_{2|1}^2(x) = Var(Y|X = x) = Var(Y|X)$ 로, X 의 분포에 따라 결정되는 확률변수이다.

확률변수로서의 Conditional Expectation $E[g(Y)|X]$

$X=x$ 인 조건에서 $g(Y)$ 의 conditional Expectation을

$$h(x) = E[g(Y)|X = x] = E[g(Y)|X] = h(X)$$

라고 할 때, X 의 분포에 따라 $h(x)$ 값들을 가지는 확률변수를 X 가 주어진 조건에서 $g(Y)$ 의 Conditional Expectation이라 한다.

Conditional Expectation의 성질

$$(1) E[E(Y|X)] = E(Y), E(Y) = E[E(Y|X)]$$

$$(2) Cov(Y - E(Y|X), v(X)) = 0, \forall v(X)$$

Proof

$$\begin{aligned} E[E(Y|X)] &= \int_{-\infty}^{+\infty} E(Y|X) f_1(x) dx \quad \because E(Y|X) = g(X) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f_{2|1}(y|x) dy f_1(x) dx \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x, y) dy dx \\ &= E(Y) \quad \because y = g(y), \text{ vice versa} \end{aligned}$$

$$\begin{aligned} \text{let, } Z &= Y - E(Y|X) \\ E(Z|X) &= E(Y|X) - E\{E(Y|X)|X\} = E(Y|X) - E(Y|X) = 0 \\ \because E(Y|X) &= g(X), E[g(X)|X] = E[g(X)] = E[Y|X] \\ E(Z) &= E[E(Z|X)] = E(0) = 0 \end{aligned}$$

$$\begin{aligned} Cov(Z, v(X)) &= E(Zv(X)) - E(Z)E(v(X)) \quad \because E(Z) = 0 \\ &= E\{E(Zv(X)|X)\} \\ &= E\{v(X)E(Z|X)\} \quad \because E(Z|X) = 0 \dots (1) \\ &= 0 = Cov(Y - E(Y|X), v(X)) \end{aligned}$$

회귀함수

Conditional Expectation $E(Y|X)$ 는 두 확률변수 Y 와 X 사이의 관계를 설명하는 데에 특별한 의미를 가진다. 특히나, $E(Y|X)$ 는 Y 를 예측하는 가장 좋은 X 의 함수를 의미하는 회귀함수라고 한다.

least square predictor

확률변수 X 의 함수 $u(X)$ 로서 $E[(Y - u(X))^2]$ 를 최소로 하는 함수는 $E(Y|X)$ 이다. 즉,

$$E[(Y - E(Y|X))^2] \leq E[(Y - u(X))^2], \forall u(X)$$

이고, mean squared prediction error인 $E[(Y - u(X))^2]$ 의 최소값은

$$\begin{aligned} E[(Y - E(Y|X))^2] &= E\{E[(Y - E(Y|X))^2|X]\} = E[Var(Y|X)] \\ &\because E[(Y - \mu_{2|1}(x))^2|X = x] = E[Y|X = x] \end{aligned}$$

• 참고사항

회귀계수는 두 변수 간에 원인과 결과의 직선적 관계를 설명하며, 상관계수는 두 변수 간 직선적 상관관계의 강도만을 나타낸다. 따라서 상관계수를 가지고 원인과 결과의 관계를 설명할 수 없으며 해서도 안된다. 예를 들면 벼 시비실험에서 분얼수(새끼 친 수)와 간장 간에는 높은 정상관이 있으나, 이것을 가지고 분얼수가 많아지는 것은 간장이 길어진 효과 때문이라든지, 간장이 길어진 것은 분얼수가 많아진 효과 때문이라고 해석해서는 안 된다. 왜냐하면 분얼수와 간장이 함께 증가한 것은 시비량을 증가시켰기 때문이다. 이와 같이 두 변수 사이에 제3의 요인이 관여하여 생긴 상관관계를 무의미상관(nonsense correlation)이라고 한다.

상관계수가 0일 때 이것은 두 변수 간에 단지 직선적 관계가 없다는 뜻이며, 두 변수 사이에 관계가 전혀 없다는 의미는 아니다. 두 변수 간에 직선적 관계가 없어도 곡선적 관계는 얼마든지 있을 수 있다.

회귀분석이 유의해도 그것이 원인과 결과의 관계가 아닌 경우도 있다. 예컨대 연도별 농가인구와 농업소득과의 관계는 비록 통계적으로 유의하더라도 실제로는 별다른 의미가 없는 것이다.

(*출처: [상관분석\(correlation analysis\)](#))

분산의 분해

$$Var(Y) = E[Var(Y|X)] + Var[E(Y|X)]$$

$$\because Y - \mu = (Y - E(Y|X)) \oplus (E(Y|X) - \mu)$$

$$E[(Y - \mu)^2] = E[(Y - E(Y|X))^2] - 2Cov(Y - E(Y|X), E(Y|X)) + E[(E(Y|X) - \mu)^2] = E[Var(Y|X)] + Var(E(Y|X))$$

$$Cov(Y - E(Y|X), E(Y|X)) = 0$$

확률변수의 독립성

정의

확률변수 X 가 어떠한 범위의 값을 갖든 확률변수 Y 에 관한 사건의 가능성에 아무런 영향을 주지 않는 경우. 즉 a, b, c, d 값에 상관없이

$$P(c \leq Y \leq d | a \leq X \leq b) = P(c \leq Y \leq d)$$

$$P(a \leq X \leq b, c \leq Y \leq d) = P(a \leq X \leq b)P(c \leq Y \leq d) \forall a, b, c, d$$

가 성립하는 경우로서, X에 관한 어떠한 사건도 Y에 관한 사건과 서로 독립이다. 두 확률변수 x와 Y가 *mutually independent*라고 한다.

Probabiltiy Function에서의 독립성

$$(1) \text{Joint CDF } cdf_{1,2}(x, y) = cdf_1(x)cdf_2(y)$$

$$(2) \text{Joint PDF } pdf_{1,2}(x, y) = pdf_1(x)pdf_2(y)$$

$$(3) \text{Joint mgf } mgf_{1,2}(x, y) = mgf_1(x)mgf_2(y)$$

$$(4) \text{Probabiltiy Measure } P(X \in A, Y \in B) = P(X \in A)P(Y \in B), \forall A, B^*$$

*Probabiltiy Measure $X \in A$ 은 $\{s : X(s) \in A\}$, 즉 역사상을 의미한다.

Joint PDF를 통해 독립성을 밝힐때, 반드시 Marginal PDF를 구할필요는 없다.

즉,

$$pdf_{1,2}(x, y) = g_1(x)g_2(y) \forall x, y \exists g_1, g_2$$

를 만족하는 $g_1(x), g_2(y)$ 가 존재하는 것을 증명하기만 하면 된다.

$$g_1(x)g_2(y) = \frac{g_1(x)}{\int_{-\infty}^{+\infty} g_1(x)dx} \frac{g_2(y)}{\int_{-\infty}^{+\infty} g_2(y)dy} = pdf_1(x)pdf_2(y)$$

$$\text{이므로. } (\because 1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} pdf_{1,2}(x, y)dydx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g_1(x)g_2(y)dydx = \int_{-\infty}^{+\infty} g_1(x)dx \int_{-\infty}^{+\infty} g_2(y)dy)$$

(e.g)

$$f_{1,2}(x, y) = 2e^{-x-2y}I_{(x \geq 0, y \geq 0)} = 2e^{-x}I_{(x \geq 0)}e^{-2y}I_{(y \geq 0)} = g_1(x)g_2(y)$$

$\therefore X$ and B are mutually independent

같은 방식으로 위의 (1)과 (3)의 Marginal cdf, Marginal mgf를 구하지 않아도 독립성을 판단할 수 있다.

독립인 확률변수들의 함수

확률변수 X, Y 가 서로 독립이면 각각의 함수인 $g_1(X), g_2(Y)$ 도 서로 독립이다.

Proof

g 의 inverse를 $g^{-1}(A) = \{x : g(x) \in A\}$ 로 나타내면

$$(g_1(X) \in A, g_2(Y) \in B) = (X \in g_1^{-1}(A), Y \in g_2^{-1}(B))$$

$$(g_1(X) \in X) = (X \in g_1^{-1}(A)), (g_2(Y) \in B) = (Y \in g_2^{-1}(B))$$

이때, X, Y가 독립이면,

$$P(X \in g_1^{-1}(A), Y \in g_2^{-1}(B)) = P(X \in g_1^{-1}(A))P(Y \in g_2^{-1}(B))$$

이므로,

$$\therefore P(g_1(X) \in A, g_2(Y) \in B) = P(g_1(X))P(g_2(Y))$$

Expectation Variance의 독립성

X, Y가 서로 독립이면,

$$E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$$

$$Cov(X, Y) = 0 \quad *$$

*역은 성립하지 않는다. ($E[XY] = E[X] = E[Y] = 0$ 인 경우가 반례)

$$\begin{aligned} Var(X + Y) &= Var(X) + Var(Y) + 2Cov(X, Y) \\ &\quad \text{if, } X, Y \text{ are independent} \\ Var(X + Y) &= Var(X) + Var(Y) \end{aligned}$$

다차원 확률변수의 분포

정의

Random Variable X_1, X_2, \dots, X_k 를 각 Entry로 하는 vector $(X_1, X_2, \dots, X_k)^t$ 를 k-차원 Random Variable 혹은 *k variate random vector*라고 한다.

Multivariate Random Variable Joint PDF (discrete)

$$\begin{aligned} (a) & f(x_1, \dots, x_k) \geq 0 \quad \forall x_i : -\infty \leq x_i \leq +\infty \quad (i = 1, 2, \dots, k) \\ & f(x_1, \dots, x_k) = 0 \quad \forall x_i \notin \{x_{i1}, x_{i2}, \dots\} \quad (i = 1, 2, \dots, k) \\ (b) & \sum_{x_1} \dots \sum_{x_k} f(x_1, \dots, x_k) = 1 \\ (c) & \sum_{x_1: a_1 \leq x_1 \leq b_1} \dots \sum_{x_k: a_k \leq x_k \leq b_k} f(x_1, \dots, x_k) = P(a_1 \leq X_1 \leq b_1, \dots, a_k \leq X_k \leq b_k) \end{aligned}$$

Multivariate Random Variable Joint PDF (continuous)

$$(a) f(x_1, \dots, x_k) \geq 0 \quad \forall x_i : -\infty < x_i < +\infty$$

$$(b) \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_k) dx_k \dots dx_1 = 1$$

$$(c) \int_{-a_1}^{+b_1} \dots \int_{-a_k}^{+b_k} f(x_1, \dots, x_k) dx_k \dots dx_1 = P(a_1 \leq X_1 \leq b_1, \dots, a_k \leq X_k \leq b_k)$$

Marginal PDF

Multivariate Random Variable $(X_1, X_2, \dots, X_k)^t$ 의 Joint PDF가 $f(x_1, \dots, x_k)$ 일때, X_1 과 $(X_1, X_2)^t$ 의 Marginal PDF $f_1(x)$ 와 $f_{1,2}(x, y)$ 는 각각 다음과 같이 주어진다.

$$(a) f_1(x) = \begin{cases} \sum_{x_2} \dots \sum_{x_k} f(x, x_2, \dots, x_k) \\ \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x, x_2, \dots, x_k) dx_k \dots dx_2 \end{cases}$$

$$(b) f_{1,2}(x, y) = \begin{cases} \sum_{x_3} \dots \sum_{x_k} f(x, y, x_3, \dots, x_k) \\ \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x, y, x_3, \dots, x_k) dx_k \dots dx_3 \end{cases}$$

Function of Random Vector's Expectation의 정의

random vector $(X_1, \dots, X_k)^t$ 의 Joint PDF가 $f(x_1, \dots, x_k)$ 일 때, 실수 값 함수 $g(x_1, \dots, x_k)$ 에 대해

$$\begin{cases} \sum_{x_1} \dots \sum_{x_k} g(x_1, \dots, x_k) f(x_1, \dots, x_k) \\ \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(x_1, \dots, x_k) f(x_1, \dots, x_k) dx_k \dots dx_1 \end{cases}$$

가 실수로 정의되면 그 값을 $g(X_1, \dots, X_k)$ 의 기댓값($E[g(X_1, \dots, X_k)]$)이라 한다.

Expectation의 성질

(선형성)

$$E[c_1 g_1(X_1, \dots, X_k) + c_2 g_2(X_1, \dots, X_k)] = c_1 E[g_1(X_1, \dots, X_k)] + c_2 E[g_2(X_1, \dots, X_k)]$$

Mean & Variance Matrix의 정의

분포의 특성을 나타내기 위해 각 변수의 Mean과 Variance를 이용하며, 변수 사이의 특성은 Covariance을 통해 나타낸다.

random vector $X = (X_1, \dots, X_k)^t$ 의 성분들인 X_1, \dots, X_k 의 Mean, Variance, Covariance인

$$\mu_i = E(X_i), \sigma_{i,j} = Cov(X_i, X_j) \quad (i, j = 1, 2, \dots, k)$$

를 대응하는 원소로 갖는 벡터와 행렬을 각각 X 의 mean vector, variance-covariance matrix라고 한다. 간단히 X 의 mean, variance matrix라고도 표현하며, 다음과 같이 정의한다.

$$E(X) = (\mu_1, \dots, \mu_k)^t = (E(X_1), \dots, E(X_k))^t$$

$$Var(X) = \begin{pmatrix} \sigma_{1,1} \dots \sigma_{x_1,k} \\ \dots \dots \dots \\ \sigma_{k,1} \dots \sigma_{x_k,x_k} \end{pmatrix} = (Cov(X_i, X_j))_{1 \leq i, j \leq k}$$

Mean & Variance Matrix의 성질

확률변수의 행렬 $V = (V_{i,j}), W = (W_{i,j})$ 에 대해 다음이 성립

- (a) $E(CWD) = CE(W)D$ (C,D는 모든 원소가 상수인 행렬)
- (b) $E(V + W) = E(V) + E(W)$

Covariance Matrix의 정의

$X = (X_1, \dots, X_k)^t$ 의 평균을 $\mu = (\mu_1, \dots, \mu_k)^t$ 라고 하면

$$(X - \mu)(X - \mu)^t = ((X_i - \mu_i)(X_j - \mu_j))$$

이므로, X의 variance matrix를 다음과 같이 나타낼 수 있다.

$$Var(X) = (Cov(X_i, X_j)) = (E[(X_i - \mu_i)(X_j - \mu_j)]) = E[(X - \mu)(X - \mu)^t]$$

또한, variance matrix를 일반화하여 $X = (X_1, \dots, X_k)^t$ 와 $Y = (Y_1, \dots, Y_l)^t$ 의 covariance matrix를

$$Cov(X, Y) = (Cov(X_i, Y_j))_{1 \leq i \leq k, 1 \leq j \leq l} \quad (k \times l)$$

라고 정의하며, $Y = (Y_1, \dots, Y_l)^t$ 의 평균을 $\eta = (\eta_1, \dots, \eta_l)^t$ 라고 하여 covariance matrix를

$$Cov(X, Y) = (Cov(X_i, Y_j)) = (E[(X_i - \mu_i)(Y_j - \eta_j)]) = E[(X - \mu)(Y - \eta)^t]$$

와 같이 나타낼 수 있다.

Mean Vector와 Covariance Matrix의 성질

A, C 는 상수의 행렬, b, d 는 상수의 벡터 일때,

- (a) $E(AX + b) = AE(X) + b$
- (b) $Var(AX + b) = AVar(X)A^t$
- (c) $Cov(AX + b, CY + d) = ACov(X, Y)C^t$
- (d) $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z),$
 $Cov(X, Z + W) = Cov(X, Z) + Cov(X, W)$
- (e) $Cov(Y, X) = (Cov(X, Y))^t, Var(X) = Cov(X, X)$
- (f) $Var(X + Y) = Var(X) + Var(Y) + Cov(X, Y) + Cov(Y, X)$

(Proof)

$$(b) Var(AX + b) = E[(AX - A\mu)(AX - A\mu)^t] = AVar(X)A^t$$

$$(c) Cov(AX + b, CX + d) = E[(AX - A\mu_x)(CY - C\mu_y)^t] = ACov(X, Y)C^t$$

Variance Matrix의 성질

$X = (X_1, \dots, X_k)^t$ 의 variance matrix $Var(X)$ 는 음이 아닌 nonnegative definite의 symmetric matrix이다. 즉

$$Var(X) = (Var(X))^t, \quad a^t Var(X) a \geq 0, \quad \forall a \in R^k$$

$$\because a^t Var(X) a = Var(a^t X) \geq 0, \quad \forall a \in R^k$$

*Definite Matrix

선형대수에서, 스칼라가 0이 아닌 모든 실수 열벡터에 대해, 즉 임의의 a^t 에 대해 $a^t M(X) a$ 의 부호가 결정 정적인 Matix M을 의미한다.

Multivariate Random Variable's Joint Moment, Joint MGF

$E(|X_1^{r_1} \dots X_k^{r_k}|) < +\infty$ 일 때

$$E(X_1^{r_1} \dots X_k^{r_k}) = \begin{cases} \sum_{x_1} \dots \sum_{x_k} x_1^{r_1} \dots x_k^{r_k} f(x_1, \dots, x_k) \\ \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} x_1^{r_1} \dots x_k^{r_k} f(x_1, \dots, x_k) dx_k \dots dx_1 \end{cases}$$

를 $X = (X_1, \dots, X_k)^t$ 의 (r_1, \dots, r_k) 차 joint moment($m_{r_1, \dots, r_k}(X_1, \dots, X_k)$ or m_{r_1, \dots, r_k})라고 한다.

0을 포함하는 열린구간들의 t_1, \dots, t_k 값에 대해 $E(e^{e_1 X_1 + e_2 X_2 + \dots + e_k X_k})$ 가 실수일 때, 함수

$$M(t_1, \dots, t_k) = E(e^{t_1 X_1 + \dots + t_k X_k}), \quad -h_i < t_i < h_i, \quad (\exists h_i > 0)(i = 1, \dots, k)$$

를 $X = (X_1, \dots, X_k)^t$ 의 Joint MGF라고 한다.

Joint MGF의 성질

(joint moment 생성)

Multivariate Random Variable $X = (X_1, \dots, X_k)^t$ 의 Joint MGF가 존재하면, 즉

$$M(t_1, \dots, t_k) = E(e^{t_1 X_1 + \dots + t_k X_k}) < +\infty, \quad -h_i < t_i < h_i, \quad (\exists h_i > 0)(i = 1, \dots, k)$$

이면, $X = (X_1, \dots, X_k)^t$ 의 모든 joint moment가 존재하고

$$E(X^{r_1} \dots X_k^{r_k}) = \left[\frac{\partial^{r_1 + \dots + r_k}}{\partial t_1^{r_1} \dots \partial t_k^{r_k}} \right]_{t_1 = \dots = t_k = 0}$$

$$M(t_1, \dots, t_k) = \sum_{r_1=0}^{\infty} \dots \sum_{r_k=0}^{\infty} \frac{E(X_1^{r_1} \dots X_k^{r_k})}{r_1! \dots r_k!} t_1^{r_1} \dots t_k^{r_k}$$

$$-h_i < t_i < h_i (\exists h_i > 0) (i = 1, \dots, k)$$

(분포 결정성)

$X = (X_1, \dots, X_k)^t, Y = (Y_1, \dots, Y_k)$ 의 MGF $M_{X_1, \dots, X_k}(t_1, \dots, t_k), M_{Y_1, \dots, Y_k}(t_1, \dots, t_k)$ 가 존재하고 0을 포함한 열린구간에서 일치하면, 즉

$$M_{X_1, \dots, X_k}(t_1, \dots, t_k) = M_{Y_1, \dots, Y_k}(t_1, \dots, t_k)$$

이면, X와 Y의 확률분포가 일치, 즉 $X = (X_1, \dots, X_k)^t$ 와 $Y = (Y_1, \dots, Y_k)^t$ 의 Joint PDF와 Joint CDF가 일치한다.

Joint CGF

2차원과 동일하게 정의된다.

$$C(t_1, \dots, t_k) = \log M(t_1, \dots, t_k) = \log E(e^{t_1 X_1 + \dots + t_k X_k})$$

를 $X = (X_1, \dots, X_k)^t$ 의 Joint CGF라고 한다. 마찬가지로 미분과 역급수전개를 통해 2차항까지 전개하면 mean과 variance matrix를 쉽게 구할 수 있다.

$$\left[\frac{\partial}{\partial t_i} C(t_1, \dots, t_k) \right]_{t_1 = \dots = t_k = 0} = E[X_i]$$

$$\left[\frac{\partial^2}{\partial t_i \partial t_j} C(t_1, \dots, t_k) \right]_{t_1 = \dots = t_k = 0} = \text{Cov}(X_i, X_j)$$

$$C(t) = \log M(t) = \log E(e^{t_1 X_1 + \dots + t_k X_k})$$

C(t)의 gradient vector를 $\dot{C}(t)$ Hessian matrix를 $\ddot{C}(t)$ 라고 하면

$$C(t) = \dot{C}(0)t + \frac{1}{2}t^2\ddot{C}(0) + \dots, E(X) = \dot{C}(0), Var(X) = \ddot{C}(0)$$

$$*\dot{C}(0) = \begin{pmatrix} \frac{\partial}{\partial t_1} C(0) \\ \dots \\ \frac{\partial}{\partial t_k} C(0) \end{pmatrix} = \begin{pmatrix} E(X_1) \\ \dots \\ E(X_k) \end{pmatrix}$$

Multivariate Random Variable에 대한 Conditional PDF

Multivariate Random Variance X, Y 에 대해, X 와 Y 의 Joint PDF가 $f_{X,Y}(x, y)$ 이고 X 의 marginal PDF가 $f_X(x)$ 일 때, $X = x$ 인 조건에서 Y 의 Conditional PDF는

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} (x : f_X(x) > 0)$$

Conditional Expectation

$X = x$ 인 조건에서 Y 의 Conditional PDF가 $f_{Y|X}(y|x)$ 일 때, $X = x$ 인 조건에서 실수 값 함수 $g(X, Y)$ 의 Conditional expectation을 다음과 같이 정의한다.

$$E(g(X, Y)|X = x) = \begin{cases} \sum_{y_1} \dots \sum_{y_l} g(x, y_1, \dots, y_l) f_{Y|X}(y_1, \dots, y_l|x) \\ \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(x, y_1, \dots, y_k) f_{Y|X}(y_1, \dots, y_l|x) dy_l \dots dy_1 \end{cases}$$

Conditional Expectation의 성질

- (a) $E[E(Y|X)] = E(Y)$
- (b) $Cov(Y - E(Y|X), v(X)) = 0, \forall v(X)$

Variance Matrix 분해

$$Var(Y) = E[Var(Y|X)] + Var[E(Y|X)]$$

(Proof)

$$Cov(Y - E(Y|X), E(Y|X)) = 0 \because E(Y|X) = v(X)$$

이므로,

$$Var(Y) = Var[(Y - E(Y|X)) \oplus E(Y|X)] = Var(Y - E(Y|X)) + Var(E(Y|X)) \because Cov = 0$$

$$let W = Y - E(Y|X)$$

$$E(WW^t|X) = E[(Y - E(Y|X))(Y - E(Y|X))^t|X] = Var(Y|X)$$

이므로,

$$Var(W) = E(WW^t) = E[E(WW^t|X)] = E[Var(Y|X)]$$

$$\therefore Var(Y) = Var(W) + Var[E(Y|X)] = E[Var(Y|X)] + Var[E(Y|X)]$$

Multivariate Random Variable의 Least Squares Predictor

Multivariate Random Variable X 의 벡터 값 함수 $u(X) = (u_1(X), \dots, u_k(X))^t$ 로서

$$E[||Y - u(X)||^2] = E[(Y_1 - u_1(X))^2 + \dots + (Y_k - u_k(X))^2]$$

을 최소로 하는 벡터 값 함수 $u(X)$ 는 $E(Y|X) = (E(Y_1|X), \dots, E(Y_k|X))^t$ 이다.

$$E[||Y - E(Y|X)||^2] \leq E[||Y - u(X)||^2], \forall u(X)$$

Multivariate Random Variable의 Independence

X_1, \dots, X_n 이 서로 독립일 필요 충분조건

(a) PDF

$$pdf_{1,\dots,n}(x_1, \dots, x_n) = pdf_1(x_1) \dots pdf_n(x_n), \forall x_i (i = 1, \dots, n)$$

(b) MGF

$$mgf_{1,\dots,n}(t_1, \dots, t_n) = mgf_1(t_1) \dots mgf_n(t_n),$$

$$\forall t_i : ||t_i|| < h_i (\exists h_i > 0) (i = 1, \dots, n)$$

두 확률변수의 독립성을 판단할때와 마찬가지로 marginal PDF나 marginal MGF를 반드시 구할 필요는 없다. 즉,

$$pdf_{1,\dots,n}(x_1, \dots, x_n) = g_1(x_1) \dots g_n(x_n), \forall (i = 1, \dots, n)$$

$$mgf_{1,\dots,n}(t_1, \dots, t_n) = m_1(t_1) \dots m_n(t_n),$$

$$\forall t_i : ||t_i|| < h_i (\exists h_i > 0) (i = 1, \dots, n)$$

또한, 여러 개의 서로 독립인 Multivariate Random Variable X_1, \dots, X_n 에 대해서도 다음과 같은 성질을 만족한다.

(a) X_1, \dots, X_n 인 각각의 함수인 $g_1(X_1), \dots, g_n(X_n)$ 도 서로 독립이다.

(b) g_1, \dots, g_n 에 대하여

$$E[g_1(X_1) \dots g_n(X_n)] = E[g_1(X_1)] \dots E[g_n(X_n)]$$

(c)

$$Cov(X_i, X_j) = 0 (i \neq j) (i, j = 1, \dots, n)$$

(d)

$$Var(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n)$$

$$\begin{aligned} \because Var(X_1 + \dots + X_n) &= Cov\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n Cov(X_i, X_j) \quad \because Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z) \end{aligned}$$

(e)

$$mgf_{X_1 + \dots + X_n}(t) = mgf_{X_1}(t) \dots mgf_{X_n}(t)$$