- **Theorem** (MGF for linear transformation). For constants a and b,

$Y=a+bX$

$M_Y(t) = E_Y(e^{tY})$

$$M_{a+bX}(t) = e^{at} M_X(bt).$$

can be used to identify the dist. of $a+bX$ from the dist. of X

Proof. $M_{a+bX}(t) = E_X[e^{t(a+bX)}] = e^{at} E_X[e^{(bt)X}] = e^{at} M_X(bt)$.

- **Theorem** (MGF for SUM of independent r.v.'s). If $X_1, \ldots, X_n$ are independent each with mgfs $M_{X_1}(t), \ldots, M_{X_n}(t)$, respectively, then the mgf of $S=X_1+\cdots+X_n$ is

Recall. LNp.8-36

t_1, t_n

$\because$ independent

$S: \Omega \to \mathbb{R}^1$
$M_S: \mathbb{R}^1 \to \mathbb{R}^1$

$$M_S(t) = M_{X_1}(t) \times \cdots \times M_{X_n}(t). \quad \text{--- } (*)$$

same t

Proof. $M_S(t) = E_S(e^{tS}) = E_{X_1,\ldots,X_n}[e^{t(X_1+\cdots+X_n)}]$ $(\triangle)$

$\because Y_1, \ldots, Y_n$ are indep.

$$= E_{X_1,\ldots,X_n}(\underbrace{e^{tX_1}}_{Y_1} \times \cdots \times \underbrace{e^{tX_n}}_{Y_n})$$

$$= E_{X_1}(e^{tX_1}) \times \cdots \times E_{X_n}(e^{tX_n}) = M_{X_1}(t) \times \cdots \times M_{X_n}(t).$$

Note. geometric is a special case of negative binomial with $n=1$.

➢ Example. If $X_1, \ldots, X_n$ are i.i.d. $\sim$ geometric(p), then

can use convolution approach (LNp.7-29) to prove it

$$S=X_1+\cdots+X_n \sim \text{negative binomial}(n, p).$$

By uniqueness Thm & mgf in LNp.8-35

Proof. $M_S(t) = M_{X_1}(t) \times \cdots \times M_{X_n}(t)$

$$= \frac{pe^t}{1-(1-p)e^t} \times \cdots \times \frac{pe^t}{1-(1-p)e^t} = \left[\frac{pe^t}{1-(1-p)e^t}\right]^n.$$

cf. convolution approach (LNp.7-32, 7-34)

➢ Example. If $X_1, \ldots, X_n$ are independent and

$$X_i \sim \text{normal}(\mu_i, \sigma_i^2), \text{ for } i=1, \ldots, n.$$

Let $S = a_0 + \underbrace{a_1 X_1}_{Y_1} + \cdots + \underbrace{a_n X_n}_{Y_n}$, then $\rightarrow$ $\begin{cases} E(S) = \mu_* & \text{(LNp.8-4)} \\ Var(S) = \sigma_*^2 & \text{(LNp.8-13)} \end{cases}$

$Y_1 + \cdots + Y_n = Z$

$M_S(t) = e^{a_0 t} \cdot M_Z(t) = e^{a_0 t} \cdot \left[\prod_{i=1}^n M_{Y_i}(t)\right]$ & $Y_i \sim \text{normal}(a_i\mu_i, a_i^2\sigma_i^2)$

$$S \sim \text{normal}\Big(\underbrace{a_0 + a_1\mu_1 + \cdots + a_n\mu_n}_{=\mu_*}, \underbrace{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2}_{=\sigma_*^2}\Big).$$

By uniqueness Thm & mgf in LNp.8-35

Proof. $M_S(t) = e^{a_0 t} \times \prod_{i=1}^n e^{\mu_i(a_i t) + (\sigma_i^2/2)(a_i t)^2}$

Recall. joint distribution

$$= e^{(a_0 + \overbrace{a_1\mu_1 + \cdots + a_n\mu_n}^{\mu_*})t + \left[(\overbrace{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2}^{\sigma_*^2})/2\right]t^2}.$$

- **Definition** (Joint Moment Generating Function). For random variables $X_1, \ldots, X_n$, their joint mgf is defined as

multi-dimensional Laplace transform

$(X_1, \ldots, X_n): \Omega \to \mathbb{R}^n$
$M_{X_1,\ldots,X_n}: \mathbb{R}^n \to \mathbb{R}^1$

$$M_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = E_{X_1,\ldots,X_n}\left(e^{t_1 X_1 + \cdots + t_n X_n}\right)$$

different t_i's

cf. $(\triangle)$ in LNp.8-38

provided that the expectation exists.

➢ Example. If $X_1, \ldots, X_m \sim$ multinomial($n, m, p_1, \ldots, p_m$), the joint pmf is:

$$\binom{n}{x_1, \ldots, x_m} p_1^{x_1} \cdots p_m^{x_m}$$

$0 \le x_i \le n$, $i=1, \ldots, m$.
$x_1 + \cdots + x_m = n$

$$M_{X_1,\ldots,X_m}(t_1,\ldots,t_m)$$

$$= \sum_{\substack{0\le x_i\le n,\ i=1,\ldots,m\\ x_1+\cdots+x_m=n}} \underbrace{e^{t_1x_1+\cdots+t_mx_m}}_{\substack{\| \\ e^{t_1x_1}\times e^{t_2x_2}\times\cdots\times e^{t_mx_m}}} \binom{n}{x_1,\ldots,x_m} p_1^{x_1}\cdots p_m^{x_m}$$

$$= \sum_{\substack{0\le x_i\le n,\ i=1,\ldots,m\\ x_1+\cdots+x_m=n}} \binom{n}{x_1,\ldots,x_m} (p_1e^{t_1})^{x_1}\cdots(p_me^{t_m})^{x_m}$$

By the multinomial expansion in LNp.7-16

$$= (p_1e^{t_1}+\cdots+p_me^{t_m})^n.$$

- Some Properties of Joint mgf

from definition

can be any X_i

➢ $M_{X_1}(t) = M_{X_1,X_2,\ldots,X_n}(t,0,\ldots,0)$.

cf. joint cdf → marginal cdf; joint pdf/pmf → marginal pdf/pmf

cf. relationship between joint mgf & marginal mgf.

➢ uniqueness theorem

$E(\prod_{i=1}^n Z_i)=\prod_{i=1}^n E(Z_i)$, $Z_i = e^{t_iX_i}$

Note. same property holds for pdf/pmf, cdf, mgf.

➢ $X_1, \ldots, X_n$ are independent if and only if

When setting $t_1=\cdots=t_n=t$, it becomes the mgf of $X_1+\cdots+X_n$

$$M_{X_1,\ldots,X_n}(t_1,\ldots,t_n) = M_{X_1}(t_1)\times\cdots\times M_{X_n}(t_n).$$

joint mgf = $\prod_{i=1}^n$ marginal mgf

different t_i's

cf. (*) in LNp.8-38

➢ $$\frac{\partial^{k_1+\cdots+k_n}}{\partial t_1^{k_1}\cdots\partial t_n^{k_n}} M_{X_1,\ldots,X_n}(0,\ldots,0) = E_{X_1,\ldots,X_n}(X_1^{k_1}\times\cdots\times X_n^{k_n}).$$

similar to the Thm & proof in LNp.8-36~37

❖ **Reading**: textbook, Sec 7.7

- moment
- central moment

their relationship

moment generating function

- Def. (Laplace transform)
  $M_X(t)=E(e^{tX})$
  $M_{X_1,\ldots,X_n}(t_1,\ldots,t_n)=E(e^{t_1X_1+\cdots+t_nX_n})$
- uniqueness theorem: mgf, pdf/pmf, cdf
- moment & mgf
  $M_X^{(k)}(0)=\mu_k$
  $\frac{\partial^{k_1+\cdots+k_n}}{\partial t_1^{k_1}\cdots\partial t_n^{k_n}}M_{X_1,\ldots,X_n}(0,\ldots,0)=E(X_1^{k_1}\cdots X_n^{k_n})$
- mgf & linear transformation
  $M_{a+bX}(t)=e^{at}M_X(bt)$
- mgf & sum of indep. r.v.'s
  $M_{X_1+\cdots+X_n}(t)=M_{X_1}(t)\times\cdots\times M_{X_n}(t)$
- joint mgf & independent
  $M_{X_1,\ldots,X_n}(t_1,\ldots,t_n)=M_{X_1}(t_1)\times\cdots\times M_{X_n}(t_n)$
- joint mgf & marginal mgf
  $M_{X_1,\ldots,X_n}(t,0,\ldots,0)=M_{X_1}(t)$

multinomial

mgfs of some dist.
- discrete r.v.
- Poisson
- binomial
- geometric
- negative binomial
- uniform
- exponential
- gamma
- beta
- normal

expectation

conditional expectation (expectation + conditional dist.)

$E_{Y|X}[h(Y)|x] = \begin{cases}\sum h(y)p(y|x)\\ \int h(y)f(y|x)dy\end{cases}$

加權平均: 平均 $h(y)$; 權重 $p(y|x)/f(y|x)$ (cf.)

Intuition: "center" of conditional dist.

understand their differences
- compare
  - $E_{Y|X}(Y|X=x)$
  - $E_{Y|X}(Y|X)$ (cf. compare its dist. with the marginal dist. of Y)
  - $E_Y(Y)$
- compare
  - $Var_{Y|X}(Y|X=x)$
  - $Var_{Y|X}(Y|X)$
  - $Var_Y(Y)$

Definition. $Y=h(X)$

$E(Y)=\begin{cases}\sum h(x)p(x)\\ \int h(x)f(x)dx\end{cases}$

加權平均: 平均 $h(x)$; 權重: $p(x)/f(x)$

- mean
- variance
- covariance
- correlation
- skewness
- kurtosis

- $\sigma_{XY}=cov(X,Y)$
  $=E[(X-\mu_X)(Y-\mu_Y)]$
  $=E(XY)-E(X)E(Y)$
- $cov(X,X)=var(X)$
- $cov>0, <0, =0$
- measure degree of association
- $cov(\sum X_i, \sum Y_j) = \sum_i\sum_j cov(X_i, Y_j)$
- $var(\sum X_i)=\sum Var(X_i) + 2\sum\sum_{i<j} cov(X_i,X_j)$

$X_1,\ldots,X_n$ indep. $Var(\sum X_i)=\sum Var(X_i)$

- expectation & Σ
  $E(\sum X_i)=\sum E(X_i)$
- expectation & Π
  X, Y: indep.
  $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$
- h convex $\Rightarrow E[h(Y)]>h(E(Y))$; h concave $\Rightarrow E[h(Y)]<h(E(Y))$

cf.

$X\le Y \Rightarrow E(X)\le E(Y)$
$a\le X\le b \Rightarrow a\le E(X)\le b$

$E(XY)=E(X)E(Y)$
indep $\overset{\times}{\rightleftarrows}$ uncorrelated

$X_1,\ldots,X_n$ have same mean μ, var σ^2 & indep.
- $\bar{X}$: $E(\bar{X})=\mu$, $Var(\bar{X})=\sigma^2/n$
- S^2: $E(S^2)=\sigma^2$, $Var(S^2)\to 0$ as $n\to\infty$

- $\rho_{XY}=cor(X,Y)$
  $=\frac{cov(X,Y)}{\sigma_X\sigma_Y}$
  $=E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right]$
- unit free
- measure strength of linear association
- $cor(a_0+a_1X, b_0+b_1Y) = sign(a_1b_1)cor(X,Y)$
- $-1\le\rho_{XY}\le 1$
- $\rho_{XY}=\begin{cases}1\Rightarrow Y=aX+b, a>0\\ -1\Rightarrow Y=aX+b, a<0\end{cases}$

- $E_{Y|X}[h(Y)|X=x]$: a function of x, free of Y
- $E_{Y|X}[h(X)|X=x]=h(x)$
- X, Y: indep.
  $E_{Y|X}[h(Y)|X=x]=E_Y[h(Y)]$
  $\Rightarrow$ a constant function of x

Variance decomposition
- $Var_Y(Y)=E_X[Var_{Y|X}(Y|X)] + Var_X[E_{Y|X}(Y|X)]$
- $Var_Y(Y)=E_X[Var_{Y|X}(Y|X)]$
  $\Leftrightarrow E_{Y|X}(Y|X)=E_Y(Y)$
- $Var_Y(Y)=Var_X[E_{Y|X}(Y|X)]$
  $\Leftrightarrow Y=E_{Y|X}(Y|X)$

law of total expectation (cf.)
- $E_{X,Y}[h(Y)]$: $E_Y[h(Y)]$, $E_XE_{Y|X}[h(Y)|X]$
  $\Rightarrow E_Y(Y)=E_XE_{Y|X}(Y|X)$
- $E_{X,Y}[h(X,Y)]$: $E_YE_{X|Y}[h(X,Y)|Y]$, $E_XE_{Y|X}[h(X,Y)|X]$

∵ multiplication law

Prediction
- criterion: MSE
- best constant predictor
  $\mu_Y \to MSE=\sigma_Y^2$
- best linear predictor
  $\mu_Y-\rho_{XY}\frac{\sigma_Y}{\sigma_X}(x-\mu_X) \to MSE=\sigma_Y^2(1-\rho_{XY}^2)$
- best predictor
  $E_{Y|X}(Y|X=x)\to MSE=E_X[Var_{Y|X}(Y|X)]$

- better predictor
- smaller MSE
- need more information about the dist. of X,Y