Lecture Notes 4: Multinomial Distribution

Professor: Zhihua Zhang

3 Multivariate Random Variables

3.1 Bivariate Distribution

Given a pair of discrete random variable X and Y, define the joint mass distribution by $f_{X,Y}(X=x,Y=y)=\mathbb{P}(X=x,Y=y)=\mathbb{P}(X=x \text{ and } Y=y)$.

Definition 3.1. *In the continuous case, we call a function $f(x,y)$ a probability density function, if*

1. *$f(x,y)\geq 0$ for all x,y.*
2. *$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$.*
3. *for any set $A\subset \mathbb{R}\times\mathbb{R}$, $\mathbb{P}((X,Y)\in A)=\iint\limits_A f(x,y)dxdy$.*

The cumulative distribution function of joint (X,Y) is given by $F_{X,Y}(x,y)=\mathbb{P}(X\leq x, Y\leq y)$.

Example 3.1. *Let (X,Y) have density $f(x,y)=\begin{cases} cx^2y & x^2\leq y\leq 1\\ 0 & otherwise\end{cases}$, then*

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = c\int_{-1}^{1}\int_{x^2}^{1} x^2ydydx = c\int_{-1}^{1}\frac{1}{2}x^2(1-x^2)dx=\frac{4}{21}c=1$$

so $c=\frac{21}{4}$. And $P(X\geq Y)=\frac{21}{4}\int_0^1\int_{x^2}^{x} x^2ydydx$.

Definition 3.2. *If random variable X and Y have joint probability density function $f_{X,Y}(x,y)$, then the marginal distribution function is given by $f_X(x)=P(X=x)=\sum_y P(X=x,Y=y)=\sum_y f_{X,Y}(x,y)$ and $f_Y(y)=P(Y=y)=\sum_x P(X=x,Y=y)=\sum_x f_{X,Y}(x,y)$. For continous case, $f_X(x)=\int f_{X,Y}(x,y)dy$ and $f_Y(y)=\int f_{X,Y}(x,y)dx$.*

Definition 3.3. *Random variables X and Y are independent, if for every A and B, $\mathbb{P}(X\in A, Y\in B)=\mathbb{P}(X\in A)\mathbb{P}(Y\in B)$.*

Theorem 3.1. *Random variables X and Y have joint probability density function $f_{X,Y}$, then X and Y are independent if and only if $f_{X,Y}(x,y)=f_X(x)f_Y(y)$ for all x and y.*

Definition 3.4. *If $f_Y(y)>0$, then the conditional density function given Y is $f_{X|Y}(x|y)=\mathbb{P}(X=x|Y=y)=\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}=\frac{f_{X,Y}(x,y)}{f_Y(y)}$. And $\mathbb{P}(X\in A|Y=y)=\int\limits_A f_{X|Y}(x|y)dx$.*

Definition 3.5. *Let $X = (X_1, X_2, ..., X_n)$ where X_i is a random variable. We call X a random vector, its probability density function is $f_{X_1,...,X_n}(x_1, x_2, ..., x_n)$, and the marginal is $f(x_i) = \sum_{x_1,...,x_{i-1},x_{i+1},...,x_n} f(x_1, ...x_n)$ for discrete case. For continuous case, we will use integral instead.$X_1, X_2, ..., X_n$ are independent if for every A_i, $\mathbb{P}(X_1 \in A_1, ..., X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i)$. Which means that $f(x_1, ..., x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$.*

Definition 3.6. *If $X_1, \ldots, X_n$ are independent and each has the same marginal distribution with CDF F, we say that $X_1, \ldots, X_n$ are i.i.d.(independent and identically distributed), $X_1, \ldots, X_n \overset{i.i.d.}{\sim} F$.*

Definition 3.7. *Let $f(x_1, x_2, ..., x_n)$ be the joint density function of $X_1, X_2, ..., X_n$, $\pi_1, \pi_2, ..., \pi_n$ is a permutation of $\{1, 2, ..., n\}$. If $f(x_1, x_2, ..., x_n) = f(x_{\pi_1}, x_{\pi_2}, ..., x_{\pi_n})$, then $X_1, ..., X_n$ are exchangeable.*

Theorem 3.2. *(de Finetti) Let $X_i \subset X$ for all $i \in \{1, 2, ..\}$. Suppose that for any n, $x_1, x_2, ..., x_n$ are exchangeable. Then we have*

$$f(x_1, x_2, ...x_n) = \int \Pi_{i=1}^{n} f(x_i|\theta) f(\theta) d\theta$$

for some parameter θ with prior distribution $f(\theta)$.

Theorem 3.3. *If $\theta \sim f(\theta)$ and $X_1, X_2, ..., X_n$ are conditionally iid given θ, then marginally $X_1, X_2, ...X_n$ are exchangeable.*

Proof.

$$\begin{aligned} f(x_{\pi_1}, x_{\pi_2}, ..., x_{\pi_n}) &= \int f(x_{\pi_1}, x_{\pi_2}, ..., x_{\pi_n}|\theta) f(\theta) d\theta \\ &= \int \prod_{i=1}^{n} f(x_{\pi_i}|\theta) f(\theta) d\theta = \int \prod_{i=1}^{n} f(x_i|\theta) f(\theta) d\theta = f(x_1, x_2, ...x_n) \end{aligned}$$

□

3.2 Expectations and Moments

Definition 3.8. *The mean of a random variable X is $E(x) = \int x dF(x) = \begin{cases} \sum_x x f(x) & x \text{ is discrete} \\ \int x f(x) dx & x \text{ is continous} \end{cases}$.*

Note that $E(x)$ exists if $\int |x| dF(x) < \infty$.

Example 3.2. *$Ga(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$, $x > 0$, then $E(x) = \int_0^\infty x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\beta x} dx = \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+1-1} e^{-\beta x} \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} dx = \frac{\alpha}{\beta}$. While, from another point of view, $\int x^{\alpha-1} e^{-\beta x} dx = \frac{\Gamma(\alpha)}{\beta^\alpha}$, and $\int x x^{\alpha-1} e^{-\beta x} dx = \frac{\Gamma(\alpha)\alpha}{\beta^{\alpha+1}}$, so $\int x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = \frac{\alpha}{\beta}$.*

Example 3.3. *$f(x) = \frac{\pi}{1+x^2}$, then $\int_{-\infty}^{+\infty} \pi \frac{|x|}{1+x^2} dx = 2\pi \int_0^{+\infty} \frac{x}{1+x^2} dx = \pi log(1+x)|_0^{+\infty} = \infty$, so $E(x)$ doesn't exist.*

Definition 3.9. *Let $Y = g(X)$, then $E(Y) = E(g(X)) = \int g(x)dF_X(x)$. If $Z = g(X,Y)$, then $E(Z) = E(g(X,Y)) = \iint g(x,y)dF_{X,Y}(x,y)$.*

Definition 3.10. *The mean of k^{th} moment of X is $E(X^k) = \int x^k dF(x)$, assuming $E(X^k)$exists.*

Theorem 3.4. *If the k^{th} moment exists, then the j^{th} moment for $j < k$ exists.*

Proof. $E(|x|) = \int_{-\infty}^{+\infty} |x|^j dF(x) = \int\limits_{|x|\leq 1} |x|^j dF(x) + \int\limits_{|x|>1} |x|^j dF(x) \leq \int\limits_{|x|\leq 1} dF(x) + \int\limits_{|x|>1} |x|^k dF(x) \leq \int\limits_{R} dF(x) + \int\limits_{R} |x|^k dF(x) = 1 + E(|x|^k) < \infty$. □

Theorem 3.5. *If $X_1, \ldots, X_n$ are random variables, and $a_1, \ldots, a_n$ are constants, then $E(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n E(X_i)$.*

Theorem 3.6. *If $X_1, \ldots, X_n$ are independent random variables, then $E(\prod_{i=1}^n X_i) = \prod_{i=1}^n E(X_i)$.*

3.3 Variance and Convariance

The variance is $\sigma^2 = E((x-\mu)^2) = Var(X) = \int (x-\mu)^2 dF(x)$, the standard deriation(std) is $std(X) = \sqrt{Var(X)}$.

1. $Var(X) = E(X^2) - \mu^2$.
2. If a and b are constants, then $Var(aX + b) = a^2 Var(X)$.
3. If $X_1, \ldots, X_n$ are independent and $a_1, \ldots, a_n$ are constants, then $Var(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a^2 Var(X_i)$.

Sample mean is $\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$, and the sample variance is $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x}_n)^2$.

Theorem 3.7. *If $X_1, \ldots, X_n$ are independent random variables, and $\mu = E(X_i)$, $\sigma^2 = Var(X_i)$. Then $E(\overline{X}_n) = \mu$, $Var(\overline{X}_n) = \frac{\sigma^2}{n}$ and $E(S_n^2) = \sigma^2$.*

Example 3.4. *If $X \sim Binomial(n,q)$, then $E(X) = \sum_{x=0}^n x\binom{n}{x} q^x (1-q)^{n-x} = \sum_{x=1}^n x\binom{n}{x} q^x (1-q)^{n-x} = \sum_{x=1}^n n\binom{n-1}{x-1} q^x (1-q)^{n-x} = \sum_{y=0}^{n-1} n\binom{n-1}{y} q^{y+1}(1-q)^{n-y-1} = nq$, $Var(X) = \sum_{x=0}^n x^2\binom{n}{x} q^x(1-q)^{n-x} - n^2q^2 = nq(1-q)$.*

Definition 3.11. *X and Y are random variables with μ_X, μ_Y, σ_X and σ_Y, then*

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

the correlation

$$\rho = \rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

1. *$Cov(X,Y) = E(XY) - E(X)E(Y)$.*
2. *$-1 \leq \rho(X,Y) \leq 1$.*
3. *If $Y = aX + b$, a, b are constants, then $\rho(X,Y) = 1$ if $a > 0$ and $\rho(X,Y) = -1$ if $a < 0$.*

4. *If X and Y are independent, then $Cov(X,Y) = \rho = 0$. Note that the converse is not true in general.*

5.

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$
$$Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$$

generally,

$$Var(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 Var(X_i) + 2\sum_{i<j} a_i a_j Cov(X_i, X_j)$$

Definition 3.12. *Let random vector $\mathbf{X} = (X_1, \ldots, X_n)^T$. Then mean of $\mathbf{X}$ is*

$$\mu = (\mu_1, \ldots, \mu_n)^T = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n])^T.$$

The covariance matrix $\mathbf{\Sigma}$ is

$$\mathbf{\Sigma} = Var(\mathbf{X}) = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_n) \\ Cov(X_1, X_2) & Var(X_2) & \cdots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \cdots & Var(X_n, X_n) \end{pmatrix}.$$

Theorem 3.8. *If $\mathbf{a}$ is a vector and $\mathbf{X}$ is a random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, then*

$$\mathbb{E}[\mathbf{a}^T\mathbf{X}] = \mathbf{a}^T\boldsymbol{\mu} \quad and \quad Var(\mathbf{a}^T\mathbf{X}) = \mathbf{a}^T\mathbf{\Sigma}\mathbf{a}$$

. If $\mathbf{A}$ is a matrix then

$$\mathbb{E}[\mathbf{A}\mathbf{X}] = \mathbf{A}\boldsymbol{\mu} \quad and \quad Var(\mathbf{A}\mathbf{X}) = \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T$$

.

3.4 Conditional Expectation

$\mathbb{E}(X)$ is a number, $\mathbb{E}(X|Y=y)$ is a function of y, and $\mathbb{E}(X|Y)$ is a random variable whose value is $\mathbb{E}(X|Y=y)$.

$$\mathbb{E}(X|Y=y) = \begin{cases} \sum_x x f_{X|Y}(x|y) & \text{x is discrete} \\ \int x f_{X|Y}(x|y)dy & \text{x is continous} \end{cases}$$

$$\mathbb{E}(g(X,Y)|Y=y) = \begin{cases} \sum_x g(x,y) f_{X|Y}(x|y) & \text{x is discrete} \\ \int g(x,y) f_{X|Y}(x|y)dy & \text{x is continous} \end{cases}$$

Example 3.5. *Suppose we draw $Y \sim Unif(0,1)$. After we observe $Y = y$, we draw $[X|Y = y] \sim Unif(y, 1)$.*

$$f_{X|Y}(x|y) = \frac{1}{1-y}, (y < x < 1)$$

$$\mathbb{E}(X|Y = y) = \int_y^1 \frac{x}{1-y} dx = \frac{1+y}{2}$$

$$\mathbb{E}(X|Y) = \frac{1+Y}{2}$$

Theorem 3.9 (The rule of iterated expectation)**.** *For X and Y, assuming the expectations exist, we have $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$ and $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$. Generally,*

$$\mathbb{E}(\mathbb{E}(g(X,Y)|X)) = \mathbb{E}(g(X,Y)) = \int g(x,y)dF(x,y)$$

Proof.

$$\mathbb{E}(\mathbb{E}(Y|X)) = \int \mathbb{E}(Y|X = x)f_X(x)dx = \iint y f_{Y|X}(y|x)f_X(x)dxdy = \iint y f_{X,Y}(x,y)dxdy = \mathbb{E}(Y)$$

□

Definition 3.13. $Var(Y|X = x) = \int (y - \hat{\mu}(x))^2 f(y|x)dy$*, where* $\hat{\mu}(x) = \mathbb{E}(Y|X = x)$*.*

Theorem 3.10. $Var(Y) = \mathbb{E}(Var(Y|X)) + Var(\mathbb{E}(Y|X))$*, so* $Var(Y) \geq Var(\mathbb{E}(Y|X))$*.*

Example 3.6. *Draw a document at random from the web, then draw n words at random from the document. Let X be the number of those words who have a certain string. If Q denotes the proportion of words in that document with the string, then Q is also a random variable because it varies from document to document.*

Given $Q = q$, we have that $X \sim Binomial(n, q)$, suppose $Q \sim Uniform(0, 1)$. Then

$$\mathbb{E}(X|Q = q) = nq, Var(X|Q = q) = nq(1-q)$$

$$Var(X) = \mathbb{E}Var(X|Q) + Var\mathbb{E}(X|Q) = n\mathbb{E}(Q(1-Q)) + nVar(Q)$$

3.5 Transformation

Random variable X has pdf f_X and cmf F_X. Let $Y = g(X)$ be a function of X. In the discrete case, the pmf of Y is $f_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(g(X) = y) = \mathbb{P}(x \in g^{-1}(y))$.

Example 3.7. *Suppose* $\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \frac{1}{4}$ *and* $\mathbb{P}(X = 0) = \frac{1}{2}$*. Let* $Y = X^2$*. So* $\mathbb{P}(Y = 0) = \frac{1}{2}$, $\mathbb{P}(Y = 1) = \frac{1}{2}$*.*

In the continuous case, the steps to find density of transformation variable is given by:

1. For each y, find set $A_y = \{x : g(x) \leq y\}$.
2. Find CDF, $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(x) \leq y) = \mathbb{P}(\{x : g(x) \leq y\}) = \int_{A_y} f_X(x)dx$.

3. $f_Y(y) = F'_Y(y)$.

Example 3.8. *$f_X(x) = e^{-x}$ for $x > 0$, and $Y = g(X) = \log X$. Then $F_X(x) = \int_0^x f_X(u)du = 1 - e^{-x}$. $A_Y = \{x : x \le e^y\}$. $F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(\log x \le y) = \mathbb{P}(x \le e^y) = F_X(e^y) = 1 - e^{-e^y}$. $f_Y(y) = (1 - e^{-e^y})' = e^y e^{-e^y}$.*

Example 3.9. *$X \sim Uniform(-1, 3)$, $Y = X^2$. $f_X(x) = \begin{cases} \frac{1}{4} & x \in (-1,3) \\ 0 & o.w. \end{cases}$. Now let us think about the distribution density of Y. Y can take value in $(0, 9)$.*

1. *$0 < Y < 1$. $A_y = \{X : X^2 \le y\} = [-\sqrt{y}, \sqrt{y}]$. $F_Y(y) = \int_{A_y} f_X(x)dx = \frac{1}{2}\sqrt{y}$.*
2. *$1 \le Y < 9$. $A_y = [-1, -\sqrt{y}]$. $F_Y(y) = \int_{A_y} \frac{1}{4}dx = \frac{1}{4}(1 + \sqrt{y})$.*

So, $f_Y(y) = \begin{cases} \frac{1}{4\sqrt{y}} & 0 < y < 1 \\ \frac{1}{8\sqrt{y}} & 1 \le y < 9 \end{cases}$

If random variable $Z = g(X, Y)$, then the way to find density of Z is given by:

1. For each z, find $A_z = \{(x, y) : g(x, y) \le z\}$.
2. Find CDF $F_Z(z) = \mathbb{P}(Z \le z) = \iint\limits_{A_z} f_{X,Y}(x, y)dxdy$.
3. $f_Z(z) = F'_Z(z)$.

Example 3.10. *Let $X_1, X_2 \overset{iid}{\sim} Uniform(0, 1)$, $Y = X_1 + X_2$. $f_{X_1,X_2}(x_1, x_2) = \begin{cases} 1 & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & o.w. \end{cases}$.*

$F_Y(y) = \mathbb{P}(\{(x_1, x_2) : (x_1 + x_2) \le y\}) = \iint\limits_{A_y} f(x_1, x_2)dx_1dx_2 = \begin{cases} \frac{1}{2}y^2 & 0 < y < 1 \\ 1 - \frac{(1-y)^2}{2} & 1 \le y \le 2 \\ 1 & y > 2 \\ 0 & y \le 0 \end{cases}$. So,

$f_Y(y) = \begin{cases} y & 0 \le y \le 1 \\ 1 - y & 1 < y \le 2 \\ 0 & o.w. \end{cases}$

Theorem 3.11. *Let X have CDF $F_X(x)$ and $Y = g(X)$, and let $\mathcal{X} = \{x : f_X(x) > 0\}$, $\mathcal{Y} = \{y : y = g(x)$ for some $x \in X\}$*

1. *if g is a strictly incresing function on $\mathcal{X}$, $F_Y(g) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.*
2. *if g is a strictly decreasing function on $\mathcal{X}$ and X is a continuous random variable. $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$*

Theorem 3.12. *Let X have continuous pdf $f_X(x)$, $Y = g(X)$, and g is strictly monotone function, then $f_Y(y) = f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)|$*

Proof. According to two case in theorem 3.4.

1. g is a strictly increasing function on $\mathcal{X}$, then $f_Y(y) = \frac{dF_Y(y)}{dy} = f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy}$

2. g is a strictly decreasing function on $\mathcal{X}$, then $f_Y(y) = \frac{dF_Y y}{dy} = -f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy}$.

So, we can combine them to $f_Y(y) = f_X(g^{-1}(y))|\frac{dg^{-1}(y)}{dy}|$. □

Theorem 3.13. *(Probability integral transformation) Let X has a continuous cdf $F_X(x)$, $Y = F_X(x)$. Then Y has uniform distribution on $(0,1)$, i.e. $\mathbb{P}(Y \leq y) = y$ where $0 \leq y \leq 1$.*

Proof. $\mathbb{P}(Y \leq y) = \mathbb{P}(F_X(x) \leq y) = \mathbb{P}(F_X^{-1}(F_X(x)) \leq F_X^{-1}(y)) = \mathbb{P}(x \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y$. □

3.6 Moments Generating Function(MGF)

Definition 3.14. *In probability theory and statistics, the moment-generating function of a random variable X is*

$$M_X(t) = \mathbb{E}[e^{tX}] = \int e^{tx} f_X(x)dx$$

One property about moment-generating function is that we can get $\mathbb{E}[X^k]$ from $M_X^{(k)}(0)$, as we can see $M_X^{(k)}(t) = \int x^k e^{tx} f_X(x)dx$, where we assume we can put the derivation inside. So $M_X^{(k)}(0) = \mathbb{E}[X^k]$.

Definition 3.15. *Laplace transformation $\mathcal{L}(t) = \int e^{-tx}dF(x)$.*

1. *If $Y = aX + b$, then $M_Y(t) = e^{bt}M_X(at)$*

2. *If $X_1, \ldots, X_n$ are independent and $Y = \sum\limits_i X_i$, then $M_Y(t) = \prod\limits_i M_{X_i}(t)$.*

3. *Let X and Y be random variables, if $M_X(t) = M_Y(t)$ for all t in an open integral around 0, then denote $X \stackrel{d}{=} Y$.*

4. *$\mathcal{L}(\mu) = \int e^{-\mu x}d\mu(x)$, $x, \mu \geq 0$ and $\mu(x)$ is non-decreasing and the integral converges for $\mu \in (0, +\infty)$. Then $\mathcal{L}'(\mu) = \int_0^\infty -e^{-\mu x}xd\mu(x)$, and $\mathcal{L}^{(k)}(\mu) = \int_0^\infty (-1)^k e^{-\mu x}x^k d\mu(x)$. So $(-1)^k\mathcal{L}^{(k)}(\mu) = \int_0^\infty e^{-\mu x}x^k d\mu(x) \geq 0$.*

Definition 3.16. *A function $f : (0, \infty) \to \mathbb{R}$ is completely monotone function if and only if f is of class C^∞(infinitely derivable), and $(-1)^n f^{(n)}(\lambda) \geq 0$ for all $n \in N \cup \{0\}$, and $\lambda > 0$.*

Theorem 3.14. *(Bernstein) Let $g : (0, \infty) \to \mathbb{R}$ be a completely monotone function. Then it is the Laplace transform of an unique measure μ on $[0, \infty]$, i.e. for all $\lambda > 0$,*

$$g(\lambda) = \mathcal{L}(\mu; \lambda) = \int_{[0,\infty)} e^{-\lambda t}\mu(dt)$$

Conversely, whenever $\mathcal{L}(\mu; \lambda) < \infty$ for every $\lambda > 0$, $\lambda \mapsto \mathcal{L}(\mu; \lambda)$ is a completely monotone function. Furthermore, $\mu(x)$ is a probability distribution iff $g(0) = 1$.

Proof. Assume $g(0+) = 1$ and $g(+\infty) = 0$. By Taylor's formula

$$\begin{aligned} f(\lambda) &= \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(\lambda - a)^k + \int_a^\lambda \frac{f^{(n)}(s)}{(n-1)!}(\lambda - s)^{n-1} ds \\ &= \sum_{k=0}^{n-1} \frac{(-1)^k f^{(k)}(a)}{k!}(a - \lambda)^k + \int_\lambda^a \frac{(-1)^n f^{(n)}(s)}{(n-1)!}(s - \lambda)^{n-1} ds \end{aligned} \tag{1}$$

where $a > 0$ and $n \in \mathbb{N}$. Let $a \to \infty$, then

$$\begin{aligned} \lim_{a\to\infty} \int_\lambda^a \frac{(-1)^n f^{(n)}(s)}{(n-1)!}(s - \lambda)^{n-1} ds &= \int_\lambda^\infty \frac{(-1)^n f^{(n)}(s)}{(n-1)!}(s - \lambda)^{n-1} ds \\ &\le f(\lambda). \end{aligned}$$

So the sum in (1) converges for every $n \in \mathbb{N}$ as $a \to \infty$. Let

$$\rho_n(\lambda) = \lim_{a\to\infty} \frac{(-1)^n f^{(n)}(a)}{n!}(a - k)^n$$

. This limit doesn't depend on $\lambda > 0$. Indeed, for $k > 0$,

$$\begin{aligned} \rho_n(k) &= \lim_{a\to\infty} \frac{(-1)^n f^{(n)}(a)}{n!}(a - k)^n \\ &= \lim_{a\to\infty} \frac{(-1)^n f^{(n)}(a)}{n!}(a - \lambda)^n \frac{(a - k)^n}{(a - \lambda)^n} \\ &= \rho_n(\lambda). \end{aligned}$$

So we can get

$$f(\lambda) = \sum_{k=0}^{n-1} \rho_k(\lambda) + \int_\lambda^\infty \frac{(-1)^n f^{(n)}(s)}{(n-1)!}(s - \lambda)^{n-1} ds$$

Let $\lambda \to \infty$, since $f(+\infty) = 0$, so $\rho_k(\lambda) = 0$. Then we can get

$$f(\lambda) = \int_\lambda^\infty \frac{(-1)^n f^{(n)}(s)}{(n-1)!}(s - \lambda)^{n-1} ds \tag{2}$$

. And since $f(0+) = 1$, we can get:

$$1 = \lim_{\lambda\to 0+} f(\lambda) = \int_0^\infty \frac{(-1)^n f^{(n)}(s)}{(n-1)!} s^{n-1} ds$$

And (2) can also be written as:

$$f(\lambda) = \int_0^\infty (1 - \frac{\lambda}{s})_+^{n-1} \frac{(-1)^n f^{(n)}(s)}{(n-1)!} s^{n-1} ds.$$

Let $t = \frac{n}{s}$, then

$$f(\lambda) = \int_0^\infty (1 - \frac{\lambda t}{n})_+^{n-1} \frac{(-1)^n}{n!} f^{(n)}(\frac{n}{t})(\frac{n}{t})^{n+1} dt$$

. Since $\lim_{n\to\infty}(1-\frac{\lambda t}{n})_+^{n-1} = e^{-\lambda t}$. So

$$f(\lambda) = \int_0^\infty e^{-\lambda t}\frac{(-1)^n}{n!}f^{(n)}(\frac{n}{t})(\frac{n}{t})^{n+1}dt.$$

For the converse, let $f(\lambda) = \mathcal{L}(\mu;\lambda) = \int_0^\infty e^{-\lambda t}\mu(dt)$. So

$$(-1)^n f^{(n)}(\lambda) = \int_0^\infty t^n e^{-\lambda t}\mu(dt) \geq 0$$

□

Corollary Let $g(t)$ be a function that is symmetric about the origin, integrable, convex and twice differentible on $(0,\infty)$ and $g(0^+) = 1$, $g(+\infty) = 0$ then

$$g(t) = \int_0^\infty (1-\frac{t}{s})_+ s g''(s)ds$$

Theorem 3.15. *A density function $f(x)$(symmetric about 0) can be represented as a Gaussian scale mixture iff $f(\sqrt{x})$ is completely monotone on $(0,\infty)$.*

Proof.

Let $g(x) = f(\sqrt{x})$.

$$\begin{aligned} & f(\sqrt{x}) \textit{ is completely monotone,} \\ \Longleftrightarrow & g(x) \textit{ is completely monotone.} \end{aligned}$$

By Bernstein :

$$\begin{aligned} \Longleftrightarrow & g(x) = \int_0^\infty e^{-xt}\mu(\mathrm{d}t) \\ \Longleftrightarrow & f(\sqrt{x}) = \int_0^\infty e^{-xt}\mu(\mathrm{d}t) \\ \Longleftrightarrow & f(x) = \int_0^\infty e^{-x^2 t}\mu(\mathrm{d}t) = C\int_0^\infty N(x \mid 0, \frac{1}{2t})\mu(\mathrm{d}t), \quad \textit{and} \int_0^\infty \mu(dt) = 1 \\ \Longleftrightarrow & f(x) \textit{ can be represented as a Gaussian scale mixture.} \end{aligned}$$

□

Theorem 3.16. *If $f(x) > 0$, then $e^{-uf(x)}$ is completely monotone for every $u > 0$ iff $f'(x)$ is completely monotone.*

Proof. If $e^{-uf(x)}$ is completely monotone for every $u > 0$:

$$e^{-\mu f(x)} = \sum_{j=0}^{\infty}\frac{(-1)^j\mu^j}{j!}[f(x)]^j$$

and all of its formal derivatives converge uniformly, so we can calculate $\frac{d^n}{dx^n}e^{-\mu f(x)}$ by termwise differentiation. Since $e^{-\mu f}$ is completely monotone, we have:

$$0 \leq (-1)^n\frac{d^n}{dx^n}e^{-\mu f(x)} = \sum_{j=1}^{\infty}\frac{\mu^j}{j!}(-1)^{n+j}\frac{d^n}{dx^n}[f(x)]^j$$

As $\mu > 0$, dividing μ, there is:

$$0 \leq (-1)^{n+1} \frac{d^n}{dx^n} f(x) + \sum_{j=2}^{\infty} \frac{\mu^{j-1}}{j!} (-1)^{n+j} \frac{d^n}{dx^n} [f(x)]^j$$

Then let $\mu \to 0$:

$$0 \leq (-1)^{n-1} \frac{d^{n-1}}{dx^{n-1}} f'(x)$$

Eventually, $f'(x)$ is completely monotone.

If $f'(x)$ is completely monotone:

$$(-1)^{n-1} \frac{d^n}{dx^n} f(x) \geq 0$$

Let $g(\lambda) = e^{-\lambda},\ \lambda = f(x)$:

$$h(x) = e^{-f(x)} = g(\lambda) \circ f(x)$$

And there is a formula for the n-th derivative of the composition $h = g \circ f$:

$$h^{(n)}(\lambda) = \sum_{(m, i_1, \ldots, i_l)} \frac{n!}{i_1! \ldots i_l!} g^{(m)}(f(\lambda)) \prod_{j=1}^{l} (\frac{f^{(j)}(\lambda)}{j!})^{i_j},$$

where $\sum_{j=1}^{l} j \cdot i_j = n$ and $\sum_{j=1}^{l} i_j = m$.

We can see that $n = m + \sum_{j=1}^{l} (j-1) \cdot i_j$.

We have $(-1)^m g^{(m)}(f(x)) \geq 0$ and $(-1)^{j-1} f^{(j)} \lambda \geq 0$.

So $(-1)^n h^{(n)}(x) \geq 0$ which means $e^{-f(x)}$ is completely monotone.

And $e^{-\mu f(x)}$ is completely monotone.

$\square$