Lecture 14: Multivariate mgf's and chf's

Multivariate mgf and chf

For an n-dimensional random vector X, its mgf is defined as

$$M_X(t) = E(e^{t'X}), \qquad t \in \mathscr{R}^n$$

and its chf is defined as

$$\phi_X(t) = E(e^{it'X}), \qquad t \in \mathscr{R}^n$$

Simple properties of mgf's and chf's

- $M_X(t)$ is either finite or ∞, and $M_X(0) = 1$.
- The chf $\phi_X(t)$ is a continuous function of $t \in \mathscr{R}^n$ and $\phi_X(0) = 1$.
- For any fixed $k \times n$ matrix A and vector $b \in \mathscr{R}^k$, the mgf and chf of the random vector $AX + b$ are, respectively,

  $$M_{AX+b}(t) = e^{b't} M_X(A't) \quad \text{and} \quad \phi_{AX+b}(t) = e^{ib't}\phi_X(A't), \quad t \in \mathscr{R}^k$$

  In particular, if Y is the first m components of X, $m < n$, then the mgf and chf of Y are, respectively,

  $$M_Y(s) = M_X((s,0)) \quad \text{and} \quad \phi_Y(s) = \phi_X((s,0)), \quad s \in \mathscr{R}^m,\ 0 \in \mathscr{R}^{n-m}$$

Calculation of moments

If $X = (X_1, ..., X_n)$ and $M_X(t)$ is finite in a neighborhood of 0, then $E(X_1^{r_1} \cdots X_n^{r_n})$ is finite for any nonnegative integers $r_1, ..., r_n$, and

$$E(X_1^{r_1} \cdots X_n^{r_n}) = \left. \frac{\partial^{r_1+\cdots+r_n} M_X(t)}{\partial t_1^{r_1} \cdots \partial t_n^{r_n}} \right|_{t=0}$$

In particular,

$$\left. \frac{\partial M_X(t)}{\partial t} \right|_{t=0} = E(X) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}$$

and

$$\left. \frac{\partial^2 M_X(t)}{\partial t \partial t'} \right|_{t=0} = E(XX') = \begin{pmatrix} E(X_1^2) & E(X_1X_2) & \cdots & E(X_1X_n) \\ E(X_2X_1) & E(X_2^2) & \cdots & E(X_2X_n) \\ \cdots & \cdots & \cdots & \cdots \\ E(X_nX_1) & E(X_nX_2) & \cdots & E(X_n^2) \end{pmatrix}$$

When the mgf is not finite in any neighborhood of 0, then we can use the chf to calculate moments.

Suppose that $E|X_1^{r_1} \cdots X_k^{r_n}| < \infty$ for some nonnegative integers $r_1, ..., r_n$ with $r = r_1 + \cdots + r_n$.

Since

$$\left| \frac{\partial^r e^{\imath t' X}}{\partial t_1^{r_1} \cdots \partial t_k^{r_n}} \right| = \left| \imath^r X_1^{r_1} \cdots X_k^{r_n} e^{\imath t' X} \right| \le |X_1^{r_1} \cdots X_k^{r_n}|$$

which is integrable, we can switch integration and differentiation to get

$$\frac{\partial^r \phi_X(t)}{\partial t_1^{r_1} \cdots \partial t_k^{r_n}} = \imath^r E\left(X_1^{r_1} \cdots X_k^{r_n} e^{\imath t' X} \right)$$

and

$$\left. \frac{\partial^r \phi_X(t)}{\partial t_1^{r_1} \cdots \partial t_k^{r_n}} \right|_{t=0} = \imath^r E(X_1^{r_1} \cdots X_k^{r_n}).$$

In particular,

$$\left. \frac{\partial \phi_X(t)}{\partial t} \right|_{t=0} = \imath EX, \qquad \left. \frac{\partial^2 \phi_X(t)}{\partial t \partial t^\tau} \right|_{t=0} = -E(XX')$$

Theorem M1.

If the mgf $M_X(t)$ of an n-dimensional random vector X is finite in a neighborhood of 0, then the chf of X is $\phi_X(t) = M_X(\imath t)$, $t \in \mathscr{R}^n$.

Proof (we need a proof because $M_X(t)$ may not be finite for all t)

We first prove this for a univariate X.

If $M_X(t)$ is finite in a neighborhood of 0, $(-\delta,\delta)$, $0<\delta<1$, then $M_X(t)$ is differentiable of all order in $(-\delta,\delta)$, $E|X|^r<\infty$ for all $r=1,2...$, and

$$M_X(t)=\sum_{k=0}^{\infty}\frac{E(X^k)t^k}{k!}, \qquad t\in(-\delta,\delta)$$

Using the inequality

$$\left|e^{isx}\left[e^{itx}-\sum_{k=0}^{n}\frac{(itx)^k}{k!}\right]\right|\le\frac{|tx|^{n+1}}{(n+1)!} \qquad t,s\in\mathscr{R}$$

we obtain that

$$\left|\phi_X(s+t)-\sum_{k=0}^{n}\frac{i^k t^k}{k!}E(X^k e^{isX})\right|\le\frac{|t|^{n+1}E|X|^{n+1}}{(n+1)!}, \qquad |t|<\delta,\ s\in\mathscr{R}$$

which implies that

$$\phi_X(s+t)=\sum_{k=0}^{\infty}\frac{\phi_X^{(k)}(s)t^k}{k!} \qquad |t|<\delta,\ s\in\mathscr{R}$$

where $g^{(k)}(s)=d^k g(s)/ds^k$.

Setting $s = 0$, we obtain that

$$\phi_X(t) = \sum_{k=0}^{\infty} \frac{\phi_X^{(k)}(0)t^k}{k!} = \sum_{k=0}^{\infty} \frac{i^k E(X^k)t^k}{k!} = \sum_{k=0}^{\infty} \frac{E(X^k)(it)^k}{k!} = M_X(it) \quad |t| < \delta$$

For $|t| < \delta$ and $|s| < \delta$,

$$\begin{aligned} \phi_X(s+t) &= \sum_{k=0}^{\infty} \frac{\phi_X^{(k)}(s)t^k}{k!} = \sum_{k=0}^{\infty} \frac{t^k}{k!} \sum_{j=k}^{\infty} \frac{i^j E(X^j)s^{j-k}}{(j-k)!} \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{j} i^j E(X^j) \frac{t^k s^{j-k}}{k!(j-k)!} = \sum_{j=0}^{\infty} \frac{i^j E(X^j)(s+t)^j}{j!} \end{aligned}$$

This proves that

$$\phi_X(s) = M_X(is), \qquad |s| < 2\delta$$

Continuing this process, we can show that $\phi_X(s) = M_X(is)$ for $|s| < 3\delta$, $|s| < 4\delta$,..., and hence, $\phi_X(t) = M_X(it)$ for $t \in \mathscr{R}$.

Consider now an n-dimensional X.

From the proof result for the univariate case, we have

$$\phi_X(t) = E(e^{it'X}) = \phi_{t'X}(1) = M_{t'X}(i) = M_X(it), \qquad t \in \mathscr{R}^n.$$

Theorem M2 (uniqueness)

(i) Random vectors $X \sim Y$ iff the chf's $\phi_X(t) = \phi_Y(t)$ for $t \in \mathscr{R}^n$.

(ii) If the mgf's of random vectors X and Y satisfy $M_X(t) = M_Y(t) < \infty$ for t in a neighborhood of 0, then $X \sim Y$.

Proof.

It is clear that the distribution of X determines its mgf and chf.

Part (i) follows from the following multivariate inversion formula: for any $A = (a_1, b_1) \times \cdots \times (a_n, b_n)$ such that the distribution of X is continuous at all points in the boundary of A,

$$P\big(X \in A\big) = \lim_{c \to \infty} \int_{-c}^{c} \cdots \int_{-c}^{c} \frac{\phi_X(t_1, ..., t_n)}{i^n (2\pi)^n} \prod_{i=1}^{n} \frac{e^{-it_i a_i} - e^{-it_i b_i}}{t_i} dt_i.$$

To establish (ii), we use the relationship between mgf and chf shown in Theorem M1.

If the mgf's $M_X(t) = M_Y(t) < \infty$ for t in a neighborhood of 0, then

$$\phi_X(t) = M_X(it) = M_Y(it) = \phi_Y(t), \qquad t \in \mathscr{R}^n$$

Then by (i), X and Y have the same distribution.

Similar to the univariate case, convergence of chf's and cdf's are equivalent and convergence of mgf's implies convergence of cdf's. We introduce the following result without giving proofs.

Theorem M3.

Suppose that $X_1, X_2, \dots$ is a sequence of k-dimensional random vectors with mgf's $M_{X_n}(t)$ and chf's $\phi_{X_n}(t)$, $t \in \mathscr{R}^k$.

(i) If $\lim_{n\to\infty} M_{X_n}(t) = M_X(t) < \infty$ for all t in a neighborhood of 0, where $M_X(t)$ is the mgf of a k-dimensional random vector X, then $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for all x where $F_X(x)$ is continuous.

(ii) A necessary and sufficient condition for the convergence of $F_{X_n}(x)$'s in (i) is that $\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$ for all $t \in \mathscr{R}^k$, where $\phi_X(t)$ is the chf of a k-dimensional random vector X.

(iii) In (ii), if $\lim_{n\to\infty} \phi_{X_n}(t) = g(t)$ for all $t \in \mathscr{R}^k$ and $g(t)$ is a continuous function on $\mathscr{R}^k$, then $g(t)$ must be a chf of some k-dimensional random vector X and the result in (ii) holds.

Note that Theorem M3(ii) gives a necessary and sufficient condition, while the converse of Theorem M3(i) is not true in general.

The next result concerns the relationship between independence and chf's and mgf's.

Theorem M4.

Let X and Y be two random vectors with dimensions n and m, respectively.

(i) X and Y are independent iff their joint and marginal chf's satisfy $\phi_{(X,Y)}(t) = \phi_X(t_1)\phi_Y(t_2)$ for all $t = (t_1, t_2)$, $t_1 \in \mathscr{R}^n$, $t_2 \in \mathscr{R}^m$.

(ii) If the mgf of (X, Y) is finite in a neighborhood of $0 \in \mathscr{R}^{n+m}$, then X and Y are independent iff their joint and marginal mgf's satisfy $M_{(X,Y)}(t) = M_X(t_1)M_Y(t_2)$ for all $t = (t_1, t_2)$ in the neighborhood of 0.

Proof.

Because of Theorem M1, (ii) follows from (i).

The only if part of (i) follows from Theorem 4.2.12A.

It remains to show the if part of (i).

Let $F(x,y)$ be the cdf of (X, Y), $F_X(x)$ be the cdf of X, and $F_Y(y)$ be the cdf of Y, $x \in \mathscr{R}^n$, $y \in \mathscr{R}^m$.

Define

$$\tilde{F}(x,y) = F_X(x)F_Y(y), \qquad x \in \mathscr{R}^n,\ y \in \mathscr{R}^m$$

It can be easily shown that $\tilde{F}$ is a cdf on $\mathscr{R}^{n+m}$.

There exist random vectors $\tilde{X}$ (n-dimensional) and $\tilde{Y}$ (m-dimensional) such that the cdf of $(\tilde{X}, \tilde{Y})$ is $\tilde{F}(x,y)$.

From the form of $\tilde{F}$, we know that $\tilde{X}$ and $\tilde{Y}$ are independent, $\tilde{X} \sim F_X$ with chf ϕ_X, and $\tilde{Y} \sim F_Y$ with chf ϕ_Y.

Hence the chf of $\tilde{F}$ must be $\phi_X(t_1)\phi_Y(t_2)$, $t_1 \in \mathscr{R}^n$, $t_2 \in \mathscr{R}^m$.

If $\phi_{(X,Y)}(t) = \phi_X(t_1)\phi_Y(t_2)$ for all $t = (t_1, t_2)$, $t_1 \in \mathscr{R}^n$, $t_2 \in \mathscr{R}^m$, then, by uniqueness, $F(x,y) = \tilde{F}(x,y) = F_X(x)F_Y(y)$ for all $x \in \mathscr{R}^n$, $y \in \mathscr{R}^m$.

This proves that X and Y are independent.

As an example, we now consider the mgf's in a family of multivariate distributions that is an extension of the univariate normal distribution family.

n-dimensional multivariate normal distribution

Let $\mu \in \mathscr{R}^n$, Σ be a positive definite $n \times n$ matrix, and $|\Sigma|$ be the determinant of Σ.

The following pdf on $\mathscr{R}^n$,

$$f(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{2}\right), \qquad x \in \mathscr{R}^n,$$

is called the n-dimensional normal pdf and the corresponding distribution is called the n-dimensional normal distribution and denoted by $N(\mu,\Sigma)$.

Using transformation, we can show that $f(x)$ is indeed a pdf on $\mathscr{R}^n$.

If $X \sim N(0,\Sigma)$, then, for any $t \in \mathscr{R}^n$,

$$\begin{aligned} M_X(t) &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \int_{\mathscr{R}^n} e^{t'x} \exp\left(-\frac{x'\Sigma^{-1}x}{2}\right) dx \\ &= \frac{e^{t'\Sigma t/2}}{(2\pi)^{n/2}|\Sigma|^{1/2}} \int_{\mathscr{R}^n} \exp\left(-\frac{(x-\Sigma t)'\Sigma^{-1}(x-\Sigma t)}{2}\right) dx \\ &= e^{t'\Sigma t/2} \end{aligned}$$

If $X \sim N(\mu,\Sigma)$, an application of the property of mgf and the previous result leads to

$$M_X(t) = e^{\mu' t + t'\Sigma t/2} \qquad t \in \mathscr{R}^n$$

From the property of mgf, we can differentiate $M_X(t)$ to get

$$E(X) = \left.\frac{\partial M_X(t)}{\partial t}\right|_{t=0} = (\mu + \Sigma t)\, e^{\mu' t + t'\Sigma t/2}\Big|_{t=0} = \mu$$

$$\begin{aligned} E(XX') &= \left.\frac{\partial^2 M_X(t)}{\partial t \partial t'}\right|_{t=0} = \frac{\partial}{\partial t}(\mu + \Sigma t)\, e^{\mu' t + t'\Sigma t/2}\Big|_{t=0} \\ &= \left[\Sigma e^{\mu' t + t'\Sigma t/2} + (\mu + \Sigma t)(\mu + \Sigma t)' e^{\mu' t + t'\Sigma t/2}\right]\Big|_{t=0} \\ &= \Sigma + \mu\mu' \end{aligned}$$

Therefore,

$$\mathrm{Var}(X) = E(XX') - E(X)E(X') = \Sigma$$

- Like the univariate case, μ and Σ in $N(\mu, \Sigma)$ are respectively the mean vector and covariance matrix of the distribution.
- Another important consequence from the previous result is that the components of $X \sim N(\mu, \Sigma)$ are independent iff Σ is a diagonal matrix, i.e., components of X are uncorrelated.

We now study properties of the multivariate normal distribution.

The bivariate normal distribution is a special case of the multivariate normal distribution with $n = 2$.

Bivariate normal distributions

First, we want to show that the definition for the general multivariate normal distribution with $n = 2$ is consistent with the early definition of the bivariate normal distribution.

Let $\mu_1 \in \mathscr{R}$, $\mu_2 \in \mathscr{R}$, $\sigma_1 > 0$, $\sigma_2 > 0$, and $-1 < \rho < 1$ be constants, and

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \qquad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Then

$$|\Sigma| = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2(1-\rho^2)$$

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

$$(x-\mu)'\Sigma^{-1}(x-\mu) = \frac{1}{|\Sigma|} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}' \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}$$

$$= \frac{1}{|\Sigma|}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}'\begin{pmatrix} \sigma_2^2(x_1-\mu_1)-\rho\sigma_1\sigma_2(x_2-\mu_2) \\ -\rho\sigma_1\sigma_2(x_1-\mu_1)+\sigma_1^2(x_2-\mu_2) \end{pmatrix}$$
$$= \frac{\sigma_2^2(x_1-\mu_1)^2-2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2)+\sigma_1^2(x_2-\mu_2)^2}{\sigma_1^2\sigma_2^2(1-\rho^2)}$$
$$= \frac{(x_1-\mu_1)^2}{\sigma_1^2(1-\rho^2)}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2(1-\rho^2)}+\frac{(x_2-\mu_2)^2}{\sigma_2^2(1-\rho^2)}$$

Thus, when $n=2$,

$$\begin{aligned} f(x) &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left(-\frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{2}\right) \\ &= \frac{\exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma_1^2(1-\rho^2)}+\frac{\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2(1-\rho^2)}-\frac{(x_2-\mu_2)^2}{2\sigma_2^2(1-\rho^2)}\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \end{aligned}$$

which is the same as the bivariate normal pdf defined earlier.

Transformation and marginals

If X is an n-dimensional random vector $\sim N(\mu,\Sigma)$ and $Y = AX+b$, where A is a fixed $k\times n$ matrix of rank $k\le n$ and b is a fixed k-dimensional vector, then $Y\sim N(A\mu+b, A\Sigma A')$.

This can be proved by using multivariate transformation, or the mgf.

We showed that the mgf of $X \sim N(\mu, \Sigma)$ is

$$M_X(t) = e^{\mu' t + t'\Sigma t/2} \qquad t \in \mathscr{R}^n$$

From the properties of mgf, $Y = AX + b$ has mgf

$$M_Y(t) = e^{b't} M_X(A't) = e^{b't} e^{\mu'(A't) + (A't)'\Sigma(A't)/2} = e^{(A\mu+b)'t + t'(A\Sigma A')t/2},$$

which is exactly the mgf for $N(A\mu + b, A\Sigma A')$, noting that $(A\Sigma A')^{-1}$ exists since A has rank $k \le n$ and Σ has rank n.

The result seems still hold even if $A\Sigma A'$ is singular, but we have not defined the normal distribution with a singular covariance matrix.

If we take $A = (I_k \; 0)$, where I_k is the identity matrix of order k and 0 is the $k \times (n-k)$ matrix of all 0's, then AX is exactly the vector containing the first k components of X and, therefore, we have shown that

- If Y is multivariate normal, then any sub-vector of Y is also normally distributed.

Is the converse true? That is, if both X and Y are normal, should the joint pdf of (X, Y) be always normal?

A counter-example

Consider

$$f(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}[1+g(x,y)], \qquad (x,y) \in \mathscr{R}^2$$

where

$$g(x,y) = \begin{cases} xy & -1 < x < 1,\ -1 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Then $f(x,y) \geq 0$. (Is it a pdf?)

$$\begin{aligned} \int_{-\infty}^{\infty} f(x,y)dy &= \frac{1}{2\pi}\left[\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}dy + \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} g(x,y)dy\right] \\ &= \frac{1}{2\pi} e^{-x^2/2}\left[\int_{-\infty}^{\infty} e^{-y^2/2}dy + xI(|x|<1)\int_{-1}^{1} ye^{-y^2/2}dy\right] \\ &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \end{aligned}$$

i.e., the marginal of X is $N(0,1)$ (which also shows that $f(x,y)$ is a pdf). Similarly, the pdf of Y is $N(0,1)$.

However, this $f(x,y)$ is certainly not a bivariate normal pdf.

Other joint distributions may not have the property that the marginal distributions are of the same type as the joint distribution.

Uniform distributions

A general n-dimensional uniform distribution on a set $A \subset \mathscr{R}^n$ has pdf

$$f(x) = \begin{cases} 1/C & x \in A \\ 0 & x \notin A \end{cases} \qquad C = \int_A dx$$

For $n = 1$ and $A =$ an interval, C is the length of the interval.
For $n = 2$, C is the area of the set A.
Consider for example $n = 2$ and A is the rectangle

$$A = [a,b] \times [c,d] = \{(x,y) \in \mathscr{R}^2 : \ a \le x \le b, \ c \le y \le d\}$$

where a, b, c and d are constants.
In this case, the two marginal distributions are uniform distributions on intervals $[a,b]$ and $[c,d]$, since, when $x \in [a,b]$,

$$\int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{C}\int_c^d dy = \frac{1}{(b-a)(d-c)}\int_c^d dy = \frac{1}{b-a}$$

and $\int_{-\infty}^{\infty} f(x,y)dy = 0$ when $x \notin [a,b]$.

However, if A is a disk

$$A = \{(x,y) \in \mathscr{R}^2 : \; x^2 + y^2 \leq 1\}$$

then $C = \pi$ and, when $x \in [-1,1]$,

$$\int_{-\infty}^{\infty} f(x,y)dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2\sqrt{1-x^2}}{\pi}$$

and $\int_{-\infty}^{\infty} f(x,y)dy = 0$ when $x \notin [-1,1]$.
This shows that the marginal distributions are not uniform.