

## Lecture Notes 5: Jacobian and Wedge

Professor: Zhihua Zhang

## 5.1 More About Mixture Distribution

**Theorem 5.1.** Let  $\mathbf{X}$  be an  $m \times 1$  random vector having a density function  $f_{\mathbf{X}}(\mathbf{x})$ , which is positive on a set  $\mathcal{X} \subset \mathbb{R}^m$ . Suppose the transform  $\mathbf{y} = \mathbf{y}(\mathbf{x}) = (y_1(\mathbf{x}), y_2(\mathbf{x}), \dots, y_m(\mathbf{x}))^T$  is 1-1 for some  $\mathcal{Y}$ , where  $\mathcal{Y}$  denotes the image of  $\mathcal{X}$  under  $\mathbf{y}$ , s.t. the inverse transformation  $\mathbf{x} = \mathbf{x}(\mathbf{y})$  exists for  $\mathbf{y} \in \mathcal{Y}$ . Assuming that the partial derivatives  $\frac{\partial x_i}{\partial y_j}$ 's, ( $i, j = 1, 2, \dots, m$ ) exist and continuous on  $\mathcal{Y}$ , it is well known that the density function of random vector  $\mathbf{Y} = \mathbf{y}(\mathbf{X})$  is

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}(\mathbf{y})) |J(\mathbf{x} \rightarrow \mathbf{y})|, \mathbf{y} \in \mathcal{Y} \quad (1)$$

where  $J(\mathbf{x} \rightarrow \mathbf{y})$  is the determinant of the Jacobian of the transformation, i.e.

$$J(\mathbf{x} \rightarrow \mathbf{y}) = \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_m} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_m}{\partial y_1} & \frac{\partial x_m}{\partial y_2} & \cdots & \frac{\partial x_m}{\partial y_m} \end{pmatrix} \quad (2)$$

Then we define the exterior product, or wedge product, which is a useful tool to calculate the determinant.

**Definition 5.1** (Wedge Product). The exterior product or wedge product of  $dx$  and  $dy$ , denoted as  $dx \wedge dy$ , has the property that  $dx \wedge dy = -dy \wedge dx$ .

Therefore, it's easy to prove that when  $x = y$ ,  $dx \wedge dx = 0$ .

**Theorem 5.2.** If  $d\mathbf{y} = (dy_1, dy_2, \dots, dy_m)^T$  is an  $m \times 1$  vector of differentials and if  $d\mathbf{x} = (dx_1, dx_2, \dots, dx_m)^T = \mathbf{B}d\mathbf{y}$ , where  $\mathbf{B}$  is an  $m \times m$  nonsingular matrix, then

$$\bigwedge_{i=1}^m dx_i = \det(\mathbf{B}) \bigwedge_{i=1}^m dy_i \quad (3)$$

*Proof.* We'll prove it by induction.

$m = 2$ :

$$\begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} dy_1 \\ dy_2 \end{pmatrix} = \begin{pmatrix} B_{11}dy_1 + B_{12}dy_2 \\ B_{21}dy_1 + B_{22}dy_2 \end{pmatrix} \quad (4)$$

Since  $dy_i \wedge dy_i = 0$  and  $dy_2 \wedge dy_1 = -dy_1 \wedge dy_2$ , we have

$$dx_1 \wedge dx_2 = (B_{11}dy_1 + B_{12}dy_2) \wedge (B_{21}dy_1 + B_{22}dy_2) \quad (5)$$

$$= (B_{11}B_{22} - B_{12}B_{21})dy_1 \wedge dy_2 = \det(\mathbf{B})dy_1 \wedge dy_2 \quad (6)$$

Suppose eq. 3 holds for  $m-1$ . Now consider the case of  $m$ . Let  $\mathbf{B} = \begin{pmatrix} \mathbf{A}_{(m-1) \times (m-1)} & \mathbf{b} \\ \mathbf{a}^T & B_{mm} \end{pmatrix}$  and  $\mathbf{Q} = \begin{pmatrix} \mathbf{I}_{(m-1) \times (m-1)} & -\mathbf{b}B_{mm}^{-1} \\ 0 & 1 \end{pmatrix}$ .

$$\begin{aligned} d\mathbf{x} &= \begin{pmatrix} \mathbf{A}_{(m-1) \times (m-1)} & \mathbf{b} \\ \mathbf{a}^T & B_{mm} \end{pmatrix} d\mathbf{y} \\ \Rightarrow \mathbf{Q}d\mathbf{x} &= \mathbf{Q}\mathbf{B}d\mathbf{y} \\ \Rightarrow \begin{pmatrix} d\mathbf{x}_{1:(m-1)} - B_{mm}^{-1}\mathbf{b}dx_m \\ dx_m \end{pmatrix} &= \begin{pmatrix} (\mathbf{A} - B_{mm}^{-1}\mathbf{b}\mathbf{a}^T)d\mathbf{y}_{1:(m-1)} \\ \mathbf{a}^T d\mathbf{y}_{1:(m-1)} + B_{mm}dy_m \end{pmatrix} \end{aligned}$$

where  $d\mathbf{x}_{1:(m-1)} = (dx_1, \dots, dx_{m-1})$  and  $d\mathbf{y}_{1:(m-1)}$  is defined similarly. Calculating the wedge for both sides, we have

$$\begin{aligned} \bigwedge_{i=1}^m dx_i &= \bigwedge ((\mathbf{A} - B_{mm}^{-1}\mathbf{b}\mathbf{a}^T)d\mathbf{y}_{1:(m-1)}) \bigwedge (B_{mm}dy_m) \\ &= B_{mm} \det(\mathbf{A} - B_{mm}^{-1}\mathbf{b}\mathbf{a}^T) \bigwedge_{i=1}^m dy_i \\ &= \det \begin{pmatrix} \mathbf{A} - B_{mm}^{-1}\mathbf{b}\mathbf{a}^T & \mathbf{0} \\ \mathbf{a}^T & B_{mm} \end{pmatrix} \bigwedge_{i=1}^m dy_i \\ &= \det(\mathbf{Q}\mathbf{B}) \bigwedge_{i=1}^m dy_i = \det(\mathbf{Q}) \det(\mathbf{B}) \bigwedge_{i=1}^m dy_i = \det(\mathbf{B}) \bigwedge_{i=1}^m dy_i \end{aligned}$$

□

**Example 5.1.** Convert rectangular coordinates  $x_1, x_2, \dots, x_m$  to polar coordinates  $r, \theta_1, \theta_2, \dots, \theta_{m-1}$ , where

$$x_1 = r \sin \theta_1 \sin \theta_2 \cdots \sin \theta_{m-2} \sin \theta_{m-1} \quad (7)$$

$$x_2 = r \sin \theta_1 \sin \theta_2 \cdots \sin \theta_{m-2} \cos \theta_{m-1} \quad (8)$$

$$\vdots \quad (9)$$

$$x_{m-1} = r \sin \theta_1 \cos \theta_2 \quad (10)$$

$$x_m = r \cos \theta_1 \quad (11)$$

$$(r > 0, 0 < \theta_i \leq \pi (i = 1, 2, \dots, m-2), 0 < \theta_{m-1} \leq 2\pi). \quad (12)$$

Then

$$J(\mathbf{x} \rightarrow r, \theta_1, \dots, \theta_{m-1}) = r^{m-1} \sin^{m-2} \theta_1 \sin^{m-3} \theta_2 \cdots \sin \theta_{m-2}. \quad (13)$$

*Proof.*

$$\begin{aligned}
x_1^2 &= r^2 \sin^2 \theta_1 \sin^2 \theta_2 \cdots \sin^2 \theta_{m-2} \sin^2 \theta_{m-1} \\
x_1^2 + x_2^2 &= r^2 \sin^2 \theta_1 \sin^2 \theta_2 \cdots \sin^2 \theta_{m-2} \\
&\vdots \\
x_1^2 + \cdots + x_{m-1}^2 &= r^2 \sin^2 \theta_1 \\
x_1^2 + \cdots + x_{m-1}^2 + x_m^2 &= r^2 \\
\Rightarrow 2x_1 dx_1 &= 2r^2 \sin^2 \theta_1 \cdots \sin^2 \theta_{m-2} \sin \theta_{m-1} \cos \theta_{m-1} d\theta_{m-1} + \text{terms of } dr, d\theta_1, \dots, d\theta_{m-2} \\
2x_1 dx_1 + 2x_2 dx_2 &= 2r^2 \sin^2 \theta_1 \cdots \sin^2 \theta_{m-3} \sin \theta_{m-2} \cos \theta_{m-2} d\theta_{m-2} + \text{terms of } dr, d\theta_1, \dots, d\theta_{m-3} \\
&\vdots \\
2x_1 dx_1 + \cdots + 2x_m dx_m &= 2r dr
\end{aligned}$$

Take wedge product for LHS and RHS simultaneously. we have

$$2^m \prod_{i=1}^m x_i \bigwedge_{i=1}^m dx_i = 2^m r^{2m-1} \sin^{2m-3} \theta_1 \cdots \sin \theta_{m-1} \prod_{i=1}^{m-1} \cos \theta_i dr \bigwedge_{i=1}^{m-1} d\theta_i \quad (14)$$

where

$$\prod_{i=1}^m x_i = r^m \sin^{m-1} \theta_1 \cdots \sin \theta_{m-1} \prod_{i=1}^{m-1} \cos \theta_i. \quad (15)$$

Then we can prove eq. 13 easily.  $\square$

**Definition 5.2.** For any matrix  $\mathbf{X} = (x_{ij})_{n \times m}$ ,  $d\mathbf{X} \stackrel{\text{def}}{=} (dx_{ij})$ ,  $d(\mathbf{XY}) = \mathbf{X}d\mathbf{Y} + d\mathbf{X}\mathbf{Y}$ , where the symbol  $(d\mathbf{X}) = \bigwedge_{i=1}^n \bigwedge_{j=1}^m dx_{ij}$ .

If  $\mathbf{X}$  is a symmetric  $m \times m$  matrix, the symbol  $(d\mathbf{X}) = \bigwedge_{1 \leq i \leq j \leq m} dx_{ij}$ .

**Theorem 5.3.** Let  $\mathbf{X}$  and  $\mathbf{Y}$  be two  $n \times m$  matrix, and  $\mathbf{X} = \mathbf{BYC}$  where  $B_{n \times n}$  and  $C_{m \times m}$  are nonsingular. Then we have

$$(d\mathbf{X}) = (\det \mathbf{B})^m (\det \mathbf{C})^n (d\mathbf{Y}), \text{ i.e. } J(\mathbf{X} \rightarrow \mathbf{Y}) = (\det \mathbf{B})^m (\det \mathbf{C})^n \quad (16)$$

*Proof.*

$$\text{vec}(\mathbf{X}) = \text{vec}(\mathbf{BYC}) \quad (17)$$

$$= (\mathbf{C}^T \otimes \mathbf{B}) \text{vec}(\mathbf{Y}), \quad (18)$$

where  $\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B})$  is the Kronecker product of  $\mathbf{A}$  and  $\mathbf{B}$  and  $\text{vec}(\mathbf{A})$  is the result of concatenating the columns of  $\mathbf{A}$ .

$$(d\mathbf{X}) = \det(\mathbf{C}^T \otimes \mathbf{B}) (d\mathbf{Y}) \quad (19)$$

$$= (\det \mathbf{C})^n (\det \mathbf{B})^m (d\mathbf{Y}) \quad (20)$$

$$(21)$$

$\square$

Consider the symmetric case, we have the following.

**Theorem 5.4.** *If  $\mathbf{X} = \mathbf{B}\mathbf{Y}\mathbf{B}^T$  where  $\mathbf{X}$  and  $\mathbf{Y}$  are  $m \times m$  symmetric matrices and  $\mathbf{B}$  is a nonsingular matrix, then*

$$(d\mathbf{X}) = (\det \mathbf{B})^{m+1}(d\mathbf{Y}), J(\mathbf{X} \rightarrow \mathbf{Y}) = (\det \mathbf{B})^{m+1} \quad (22)$$

*Proof.*

$$d\mathbf{X} = \mathbf{B}d\mathbf{Y}\mathbf{B}^T \Rightarrow (d\mathbf{X}) = (\mathbf{B}d\mathbf{Y}\mathbf{B}^T) = \rho(\mathbf{B})(d\mathbf{Y}) \quad (23)$$

where  $\rho(\mathbf{B})$  is a polynomial of elements of  $\mathbf{B}$ .

**Lemma 5.1.** *if  $\rho$  is a polynomial and has the property that  $\rho(\mathbf{X}_1\mathbf{X}_2) = \rho(\mathbf{X}_1)\rho(\mathbf{X}_2)$ , then  $\rho(\mathbf{X}) = (\det \mathbf{X})^k$  for some  $k$ .*

It's easy to show  $\rho(\mathbf{B}_1\mathbf{B}_2) = \rho(\mathbf{B}_1)\rho(\mathbf{B}_2)$ :

$$(\mathbf{B}_1\mathbf{B}_2d\mathbf{Y}\mathbf{B}_2^T\mathbf{B}_1^T) = (\mathbf{B}_1(\mathbf{B}_2d\mathbf{Y}\mathbf{B}_2^T)\mathbf{B}_1^T) = \rho(\mathbf{B}_1)(\mathbf{B}_2d\mathbf{Y}\mathbf{B}_2^T) = \rho(\mathbf{B}_1)\rho(\mathbf{B}_2)(d\mathbf{Y}) \quad (24)$$

Because of this property, we have  $\rho(\mathbf{B}) = (\det \mathbf{B})^k$  for some  $k$ . To determine  $k$ , we consider a simple case where  $\mathbf{B} = \text{diag}(b, 1, \dots, 1)$  and let  $\mathbf{Y} = (y_{ij})$ . Then

$$\mathbf{B}\mathbf{Y}\mathbf{B}^T = \begin{pmatrix} b^2y_{11} & by_{12} & \cdots & by_{1m} \\ by_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ by_{m1} & y_{m2} & \cdots & y_{mm} \end{pmatrix}.$$

So considering the upper triangular part, we have

$$(d\mathbf{X}) = (\mathbf{B}d\mathbf{Y}\mathbf{B}^T) = b^{m+1}(d\mathbf{Y}) = (\det \mathbf{B})^{m+1}(d\mathbf{Y}).$$

Therefore,  $k = m + 1$ . □

**Theorem 5.5.** *If  $\mathbf{X}_{m \times m} = \mathbf{Y}^{-1}$ , then*

$$(d\mathbf{X}) = (\det \mathbf{Y})^{-2m}(d\mathbf{Y}).$$

*Further, if  $\mathbf{Y}$  is symmetric, then*

$$(d\mathbf{X}) = (\det \mathbf{Y})^{-(m+1)}(d\mathbf{Y}).$$

*Proof.* Since

$$\mathbf{X} = \mathbf{Y}^{-1}\mathbf{Y}\mathbf{Y}^{-1},$$

If  $\mathbf{Y}$  is asymmetric, by Theorem. 5.3, we have

$$(d\mathbf{X}) = (\det \mathbf{Y}^{-1})^{2m}(d\mathbf{Y}) = (\det \mathbf{Y})^{-2m}(d\mathbf{Y}).$$

If  $\mathbf{Y}$  is symmetric, by Theorem. 5.4, we have

$$(d\mathbf{X}) = (\det \mathbf{Y}^{-1})^{m+1}(d\mathbf{Y}) = (\det \mathbf{Y})^{-(m+1)}(d\mathbf{Y}).$$

□

**Theorem 5.6.** If  $\mathbf{A}$  is an  $m \times m$  symmetric positive definite matrix, by Cholesky decomposition,  $\mathbf{A} = \mathbf{T}^T \mathbf{T}$ , where  $\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1m} \\ & t_{22} & \cdots & t_{2m} \\ & * & \ddots & \vdots \\ & & & t_{mm} \end{pmatrix}$  is an upper triangular matrix with positive diagonal elements. Then we have

$$J(\mathbf{A} \rightarrow \mathbf{T}) = 2^m \prod_{i=1}^m t_{ii}^{m+1-i}. \quad (25)$$

*Proof.* Let  $\mathbf{A} = (a_{ij})_{m \times m}$ .

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix} = \begin{pmatrix} t_{11} & & & \\ t_{12} & t_{22} & & * \\ \vdots & \vdots & \ddots & \\ t_{1m} & t_{2m} & \cdots & t_{mm} \end{pmatrix} \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1m} \\ & t_{22} & \cdots & t_{2m} \\ & * & \ddots & \vdots \\ & & & t_{mm} \end{pmatrix}$$

So

$$\begin{aligned} a_{11} = t_{11}^2 & \implies da_{11} = 2t_{11}dt_{11} \\ a_{12} = t_{11}t_{12} & \implies da_{12} = dt_{11}t_{12} + t_{11}dt_{12} \\ \vdots & \vdots \\ a_{ii} = t_{1i}^2 + \cdots + t_{ii}^2 & \implies da_{ii} = \text{terms of } dt_{1i}, \dots, dt_{(i-1),i} + 2t_{ii}dt_{ii} \\ (i < j) \ a_{ij} = t_{1i}t_{1j} + \cdots + t_{ii}t_{ij} & \implies da_{ij} = \text{terms of } dt_{1j}, \dots, dt_{(i-1),j}, dt_{1i}, \dots, dt_{(i-1),i} + dt_{ii}t_{ij} + t_{ii}dt_{ij} \\ \vdots & \vdots \\ a_{mm} = t_{1m}^2 + \cdots + t_{mm}^2 & \implies da_{mm} = \text{terms of } dt_{1m}, \dots, dt_{(m-1),m} + 2t_{mm}dt_{mm} \end{aligned}$$

When taking wedge of both sides, we have

$$(d\mathbf{A}) = 2^m \prod_{i=1}^m t_{ii}^{m+1-i} (d\mathbf{T})$$

□

## Homework

1. Compute the Laplace transforms of Gamma, Negative Binomial, Poisson distributions.
2. Consider that

$$\begin{aligned} w_1 &= w\alpha, & w_2 &= w(1-\alpha), \\ u_1 &= u - \beta\sigma\sqrt{\frac{w_2}{w_1}}, & u_2 &= u + \beta\sigma\sqrt{\frac{w_1}{w_2}}, \\ \sigma_1^2 &= r(1-\beta^2)\sigma^2w/w_1, & \sigma_2^2 &= (1-r)(1-\beta^2)\sigma^2w/w_2, \end{aligned}$$

where  $\alpha, \beta, r \in (0, 1)$ . Compute the Jacobian from  $(w_1, w_2, u_1, u_2, \sigma_1^2, \sigma_2^2)$  to  $(w, u, \sigma^2, \alpha, \beta, r)$

## 5.2 Random variables and their properties

**Definition 5.3.** Let random vector  $\mathbf{X} = (X_1, \dots, X_m)^T$ . Then mean of  $\mathbf{X}$  is

$$\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)^T = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_m])^T.$$

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

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

**Lemma 5.2.** If  $\mathbf{a}$  is a vector and  $\mathbf{X}$  is a random vector with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ , then

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

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

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

### 5.2.1 Examples

**Definition 5.4** (The Multinomial Distribution). A discrete random vector  $\mathbf{X} = (X_1, \dots, X_k)$  has Multinomial distribution of dimension  $k$  with parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  and  $n$  ( $0 \leq \theta_i \leq 1, \sum_{i=1}^k \theta_i \leq 1, n = 1, 2, \dots$ ). If it's p.d.f. is

$$\text{Mul}(\mathbf{x}|\boldsymbol{\theta}, n) = \frac{n!}{\prod_{i=1}^k x_i! (n - \sum_{i=1}^k x_i)!} \prod_{i=1}^k \theta_i^{x_i} (1 - \sum_{t=1}^k \theta_t)^{n - \sum_{j=1}^k x_j}. \quad (26)$$

The mean vector and covariance matrix

$$\mathbb{E}[X_i] = n\theta_i \quad (27)$$

$$\text{Var}(X_i) = n\theta_i(1 - \theta_i) \quad (28)$$

$$\text{Cov}(X_i, X_j) = -n\theta_i\theta_j. \quad (29)$$

**Theorem 5.7.** The marginal distribution of  $\mathbf{X}^{(m)} = (X_1, \dots, X_m)^T, (m < k)$  is the multinomial distribution

$$M_m(\mathbf{x}^{(m)} | (\theta_1, \dots, \theta_m), n).$$

The conditional distribution of  $\mathbf{X}^{(m)}$  given the remaining  $X_i$ 's is also Multinomial

$$f(x^{(m)} | x_{m+1}, \dots, x_k) \sim M_{m-1}(\mathbf{x}^{(m)} | (\theta'_1, \dots, \theta'_m), n - s),$$

where  $\theta'_i = \frac{\theta_i}{\sum_{j=1}^m \theta_j}, (1 \leq i \leq m)$  and  $s = \sum_{i=m+1}^k x_i$ .

Its corresponding prior is

**Definition 5.5** (Dirichlet Distribution). A continuous random vector  $\mathbf{X} = (X_1, \dots, X_k)$  has a Dirichlet distribution of dimension  $k$  with parameter  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{k+1})$ , ( $\alpha_i > 0, i = 1, \dots, k+1$ ) if its p.d.f. is

$$Dir(\mathbf{x}|\boldsymbol{\alpha}) = \frac{\Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\prod_{i=1}^{k+1} \Gamma(\alpha_i)} x_1^{\alpha_1-1} \dots x_k^{\alpha_k-1} (1 - \sum_{i=1}^k x_i)^{\alpha_{k+1}-1} \quad (30)$$

The mean vector and covariance matrix

$$\mathbb{E}[X_i] = \alpha_i / \sum_{j=1}^{k+1} \alpha_j \quad (31)$$

$$Var(X_i) = \frac{\mathbb{E}[X_i](1 - \mathbb{E}[X_i])}{1 + \sum_{j=1}^{k+1} \alpha_j} \quad (32)$$

$$Cov(X_i, X_j) = \frac{\mathbb{E}[X_i]\mathbb{E}[X_j]}{1 + \sum_{t=1}^{k+1} \alpha_t}. \quad (33)$$

**Theorem 5.8.** The marginal distribution of  $\mathbf{X}^{(m)} = (X_1, \dots, X_m)^T$ , ( $m < k$ ) is the Dirichlet distribution

$$Dir(\mathbf{x}^{(m)} | (\alpha_1, \dots, \alpha_m), \sum_{i=m+1}^{k+1} \alpha_i).$$

The conditional distribution of  $\mathbf{X}^{(m)}$  given the  $X_{m+1}, \dots, X_k$  of  $Y_i = \frac{X_i}{1 - \sum_{j=m+1}^k X_j}$  is also Dirichlet

$$Dir(\mathbf{y}^{(m)} | \alpha_1, \dots, \alpha_m, \alpha_{k+1}).$$

**Theorem 5.9.** A random vector  $\mathbf{X} = (X_1, \dots, X_k) \sim Dir(\mathbf{x}|\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{k+1}))$ . If  $\mathbf{Z} = (Z_1, \dots, Z_t)$  satisfies

$$Z_1 = X_1 + \dots + X_{i_1} \quad (34)$$

$$Z_2 = X_{i_1+1} + \dots + X_{i_2} \quad (35)$$

$$\vdots \quad (36)$$

$$Z_t = X_{i_{t-1}+1} + \dots + X_k, \quad (37)$$

then

$$\mathbf{Z} \sim Dir(\mathbf{z}|\boldsymbol{\beta}),$$

where  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{t+1})$  and

$$\beta_1 = \alpha_1 + \dots + \alpha_{i_1}$$

$$\beta_2 = \alpha_{i_1+1} + \dots + \alpha_{i_2}$$

$$\vdots$$

$$\beta_t = \alpha_{i_{t-1}+1} + \dots + \alpha_k$$

$$\beta_{t+1} = \alpha_{k+1}.$$

### Homework

1. Show the conditional distribution of multinomial distribution in Theorem 5.7.
- 2.

$$\begin{aligned}\mathbb{P}(\mathbf{X}|\boldsymbol{\theta}, n) &\sim \textit{Multinomial Distribution}, \\ \mathbb{P}(\boldsymbol{\theta}|\boldsymbol{\alpha}) &\sim \textit{Dirichlet Distribution}.\end{aligned}$$

Compute  $\mathbb{P}(\boldsymbol{\theta}|\mathbf{X})$ .