

# Brief Article

The Author

October 14, 2016

## **I 2.1**

Trivial.

## **2 2.2**

Trivial.

## **3 2.3**

The proof of (2.262) is trivial.

Proof of (2.263) (binomial theorem): The case of  $N = 1$  is trivial. The induction step is also easy, as long as you have patience.

The proof of (2.264) is trivial.

## **4 2.4**

By rearranging the result of differentiation, we have

$$\frac{1}{1-\mu} E(x) = \frac{1}{1-\mu} N. \quad (\text{I})$$

Then everything is trivial.

The proof of variation: skip...

## **5 2.5**

Too hard for me... see the manual... too many techniques in calculus...

## **6 2.6**

(2.267): make use of  $\Gamma(x+1) = x\Gamma(x)$ . (2.268): first calculate  $E(\mu^2)$ , then we are done.

(2.269): trivial...

Hint: as long as you prove the normalization, everything else becomes easy.

## **7 2.7**

Trivial... why 2 stars...

## **8 2.8**

(2.270): trivial. (2.271): see the manual...

Hint: sometimes, symbol manipulation is powerful.

## **9 2.9**

See the manual... too hard... but I should understand the upper and lower limits of integrand better. Here,  $\mu_1, \dots, \mu_{M-2}$  are constants. The upper and lower limits of  $\mu_{M-1}$  is a function of these  $M-2$  values.

## **10 2.10**

(2.273): trivial. (2.274): trivial. (2.275): trivial. There's a similar shortcut of covariance like variance.

## I1 2.I1

Since  $\mu_j^{\alpha_j-1}$  can be expressed as  $\exp((\alpha_j - 1) \ln \mu_j)$ , we have

$$\frac{\partial}{\partial \alpha_j} \prod_{k=1}^M \mu_k^{\alpha_k-1} = \frac{\partial}{\partial \alpha_j} \prod_{k=1}^M \exp((\alpha_k - 1) \ln \mu_k) \quad (2)$$

$$= \frac{\partial}{\partial \alpha_j} \exp((\alpha_j - 1) \ln \mu_j) \prod_{k=1, k \neq j}^M \mu_k^{\alpha_k-1} \quad (3)$$

$$= \ln \mu_j \exp((\alpha_j - 1) \ln \mu_j) \prod_{k=1, k \neq j}^M \mu_k^{\alpha_k-1} \quad (4)$$

$$= \ln \mu_j \prod_{k=1}^M \mu_k^{\alpha_k-1} \quad (5)$$

For the rest, see the manual...

Hint: if you see fraction  $a/b$ , try to fit them in a form containing  $\ln(b/a)$ ; if you see something  $-a$ , you can fit it in a form containing  $\ln(a^{-1})$ . If you see addition in something related to log, you should think about multiplication, and vice versa; if you see subtraction, you should think about division!

## I2 2.I2

Normalized: trivial. Mean:  $(a + b)/2$ ; trivial. Variance:  $(b - a)^2/12$ ; trivial.

## I3 2.I3

See the manual and refer to matrix cookbook... I think the trace trick is very useful...

I think the most important thing is... See the sum of entry-wise products of elements about 2 matrices as a trace... Consider quadratic form  $\mathbf{x}^T A \mathbf{x}$ , we can consider it as the sum of entry-wise products of elements about  $A$  and  $\mathbf{x}\mathbf{x}^T$ .

Proof of cyclicity of trace:  $\text{Tr}(ABC) = \text{Tr}(A(BC)) = \text{Tr}((BC)A) = \text{Tr}(BCA) = \text{Tr}(CAB)$ . (cyclicity for 2 is trivial)

## I4 2.I4

Skip... I really don't know variation calculus...

## I5 2.I5

$$H(\mathbf{x}) = - \int \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \ln \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\} \quad (6)$$

$$= \int \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \left( \frac{D}{2} \ln(2\pi) + \frac{1}{2} \ln |\boldsymbol{\Sigma}| + \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right) \quad (7)$$

$$= \frac{D}{2} \ln(2\pi) + \frac{1}{2} \ln |\boldsymbol{\Sigma}| + \frac{1}{2} \int \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{Tr}(\boldsymbol{\Sigma}^{-1} [(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T]) \quad (8)$$

$$= \frac{D}{2} \ln(2\pi) + \frac{1}{2} \ln |\boldsymbol{\Sigma}| + \frac{1}{2} \text{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}) \quad (9)$$

$$= \frac{D}{2} \ln(2\pi) + \frac{1}{2} \ln |\boldsymbol{\Sigma}| + \frac{D}{2} \quad (10)$$

The conversion from quadratic form to a trace representation is the most important thing. See the hint in 2.I4. Also, we exchange the order of integration, and the multiplication of two matrices in the trace operator, which is correct.

## I6 2.I6

$$p(x) = \int p_1(x - x_2) p_2(x_2) dx_2 \quad (11)$$

$$= \int N(x - x_2 | \mu_1, \tau_1^{-1}) N(x_2 | \mu_2, \tau_2^{-1}) dx_2 \quad (12)$$

$$= \int N(x | \mu_1 + x_2, \tau_1^{-1}) N(x_2 | \mu_2, \tau_2^{-1}) dx_2 \quad (13)$$

Using equation (2.II5), we have  $p(x)$  has a variance of  $1/\tau_1 + 1/\tau_2$ . So  $H(x) = \ln(2\pi(\tau_1 + \tau_2)/(\tau_1\tau_2))$ .

## I7 2.I7

Trivial... And we have to note that the determinant of the covariance matrix just works as part of the normalization constant. What really matters in the thing in the exponent.

## 18 2.18

Given the definition of eigenvalue, we have

$$\Sigma \mathbf{u}_i = \lambda_i \mathbf{u}_i. \quad (14)$$

Next we have some transformations...

$$\Sigma \mathbf{u}_i^* = \lambda_i^* \mathbf{u}_i^* \quad \text{taking complex conjugate} \quad (15)$$

$$(\mathbf{u}^*)_i^T \Sigma \mathbf{u}_i = \lambda_i (\mathbf{u}_i^*)_i^T \mathbf{u}_i \quad (\text{left multiply } (\mathbf{u}^*)_i^T) \quad (16)$$

$$\mathbf{u}_i^T \Sigma \mathbf{u}_i^* = \lambda_i^* \mathbf{u}_i^T \mathbf{u}_i^* \quad \text{left multiply (15)} \quad (17)$$

$$\mathbf{u}_i^T \Sigma \mathbf{u}_i^* = \lambda_i \mathbf{u}_i^T \mathbf{u}_i^* \quad \text{transpose of 16} \quad (18)$$

$$\lambda_i^* \mathbf{u}_i^T \mathbf{u}_i^* = \lambda_i \mathbf{u}_i^T \mathbf{u}_i^* \quad \text{right sides of above two equations} \quad (19)$$

$$\lambda_i^* = \lambda_i \quad \text{norm of a eigenvector is non-zero} \quad (20)$$

For the orthogonality, see the manual. The last question is trivial. Since we can always construct an orthonormal basis.

## 19 2.19

Trivial...

## 20 2.20

Trivial...

## 21 2.21

Trivial...

## 22 2.22

Let  $AA^{-1} = I$ . Because  $A$  is symmetric, we have  $A^T A^{-1} = I$ . Thus,  $A^{-1} = (A^T)^{-1} = (A^{-1})^T$ .

## **23 2.60**

Just see the manual...

## **24 2.61**

See the manual... Here, it considers the points that are very far from every sample points we observed, so that the radius of the  $V$  can be practically seen as the distance from the point to the origin... What's more, some variable transformation in integration is used.