CS470 Introduction to Artificial Intelligence
Fall 2017 Prof. Keeeung Kim
Homework 4. Basic Probability 20140174 Hyungrok Kim

Problem 1.

a) Derive the mean and variance of Bernoulli distribution Bern(x; p).

$$p(x) = p^x(1-p)^{1-x}$$

$$mean = \mu = E[X] = \sum_x x * p(x) = 0 * p(0) + 1 * p(1) = p$$

$$variance = Var(X) = E[(X-\mu)] = \sum_x (x-\mu)^2 * p(x)$$

$$= \sum_x (x^2p(x) - 2\mu xp(x) + \mu^2 p(x))$$

$$= \sum_x \left(x^2p(x)\right) - 2\mu^2 + \mu^2 = \sum_x \left(x^2p(x)\right) - \mu^2 = E[X^2] - (E[X])^2$$

$$= \sum_x x^2 * p(x) - p^2 = p - p^2 = p(1-p)$$

$$mean = p,\ \ variance = p(1-p)$$

b) Derive the mean and variance of Bernoulli distribution B(x; n, p).

The binomial distribution B(n, p) is the discrete probability distribution of the number of successes in a sequence of n independent trials with success probability p for each trial. Thus, Bernoulli distribution is a special case of binomial distribution with n = 1. Because the trials are independent, we can say that random variable X in binomial distribution is sum of n random variables of Bernoulli distribution.

$$Binomial\ random\ variable\ X = X_1 + X_2 + \cdots + X_n$$

$$X_i\ \ are\ independent\ Bernoulli\ distribution. (i = 1, 2, \dots, n)$$

$$E[X_i] = p,\ \ Var(X_i) = p(1-p)$$

$$E[X] = E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n] = np$$

$$Var(X) = Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n)$$

$$= np(1-p)$$

$$mean = np,\ \ variance = np(1-p)$$

Problem 2.

a) For D-dimensional vector x, derive the mean and variance of multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{x};\ \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

- Deriving the mean

$$E[X] = \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\} x dx$$

assume $z = x - \mu$

$$E[X] = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} z^T\Sigma^{-1}z\right\}(z+\mu)dz$$

$$= \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \left(\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} z^T\Sigma^{-1}z\right\} z dz + \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} z^T\Sigma^{-1}z\right\} \mu\, dz\right)$$

we can derive that

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} z^T\Sigma^{-1}z\right\} z dz = \int_{0}^{\infty} \exp\left\{-\frac{1}{2} z^T\Sigma^{-1}z\right\} z dz - \int_{0}^{-\infty} \exp\left\{-\frac{1}{2} z^T\Sigma^{-1}z\right\} z dz$$

assume $z_1 = -z$

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} z^T\Sigma^{-1}z\right\} z dz = \int_{0}^{\infty} \exp\left\{-\frac{1}{2} z^T\Sigma^{-1}z\right\} z dz - \int_{0}^{\infty} \exp\left\{-\frac{1}{2} {z_1}^T\Sigma^{-1}z_1\right\} z_1 dz_1 = 0$$

from above,

$$E[X] = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} z^T\Sigma^{-1}z\right\} \mu\, dz = \mu \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\} dx$$

because $\frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$ is the probability density function,

$$\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\} dx = 1$$

so,

$$E[X] = \mu \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\} dx = \mu$$

$$mean = E[X] = \mu$$

- Deriving the variance

$$first,\ \ from\ \ (x-\mu)^T\Sigma^{-1}(x-\mu)$$

$$let\ A = \Sigma^{-1}$$

$$\mathrm{A} = \frac{1}{2}(\mathrm{A}+\mathrm{A}^T) + \frac{1}{2}(\mathrm{A}-\mathrm{A}^T)$$

$$let\ \ B = \frac{1}{2}(\mathrm{A}+\mathrm{A}^T), \qquad C = \frac{1}{2}(\mathrm{A}-\mathrm{A}^T)$$

B is symmetric,

C is antisymmetric which means $\mathrm{c}_{ij} = -\mathrm{c}_{ji}$

$$(x-\mu)^T\Sigma^{-1}(x-\mu) = (x-\mu)^T\mathrm{A}(x-\mu) = (x-\mu)^T(\mathrm{B}+\mathrm{C})(x-\mu)$$

$$= (x-\mu)^T\mathrm{B}(x-\mu) + (x-\mu)^T\mathrm{C}(x-\mu)$$

$$(x-\mu)^T\mathrm{C}(x-\mu) = \sum_{i=j+1}\sum_{j=1}\left(\mathrm{c}_{ij}+\mathrm{c}_{ji}\right)(x-\mu)_i(x-\mu)_j = 0$$

because the antisymmetric component vanishes from the equation,

we can take Σ^{-1} $to\ be\ symmetric\ without\ loss\ of\ generality.$

$$if\ D = \Sigma^{-1}\ is\ symmetric,$$

$$D\mathrm{D}^{-1} = I,$$

$$(\mathrm{D}^{-1})^T\mathrm{D}^T = (\mathrm{D}^{-1})^T D = I$$

$$\mathrm{D}^{-1} = (\mathrm{D}^{-1})^T$$

$$so\ the\ inverse\ of\ \Sigma^{-1}, which\ is\ \Sigma\ is\ also\ symmetric.$$

Σ is a real, symmetric matrix, its eigenvalues will be real,

and its eigenvectors can be chosen to be an orthonormal set.

$$\Sigma\mathrm{u}_i = \lambda_i\mathrm{u}_i$$

$$\mathrm{u}_i\mathrm{u}_j = \begin{cases} 1, & if\ i=j \\ 0, & otherwise \end{cases}$$

with eigendecomposition $\Sigma = [\mathrm{u}_1 \quad \cdots \quad \mathrm{u}_D]\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_D \end{bmatrix}\begin{bmatrix} {\mathrm{u}_1}^T \\ \vdots \\ {\mathrm{u}_D}^T \end{bmatrix}$

$$= [\lambda_1\mathrm{u}_1 \quad \cdots \quad \lambda_D\mathrm{u}_D]\begin{bmatrix} {\mathrm{u}_1}^T \\ \vdots \\ {\mathrm{u}_D}^T \end{bmatrix} = \sum_{i=1}^{D}\lambda_i\mathrm{u}_i{\mathrm{u}_i}^T$$

$$P = [\mathrm{u}_1 \quad \cdots \quad \mathrm{u}_D]\ is\ an\ orthonormal\ matrix, P^{-1} = P^T$$

$$\Sigma^{-1} = \begin{bmatrix} \mathrm{u}_1{}^T \\ \vdots \\ \mathrm{u}_D{}^T \end{bmatrix}^{-1} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_D \end{bmatrix}^{-1} [\mathrm{u}_1 \quad \cdots \quad \mathrm{u}_D]^{-1} = [\mathrm{u}_1 \quad \cdots \quad \mathrm{u}_D] \begin{bmatrix} \frac{1}{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\lambda_D} \end{bmatrix} \begin{bmatrix} \mathrm{u}_1{}^T \\ \vdots \\ \mathrm{u}_D{}^T \end{bmatrix}$$

$$= \sum_{i=1}^{D} \frac{1}{\lambda_i} \mathrm{u}_i \mathrm{u}_i{}^T$$

$$\Sigma^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i} \mathrm{u}_i \mathrm{u}_i{}^T$$

$$(x-\mu)^T \Sigma^{-1} (x-\mu) = (x-\mu)^T \left(\sum_{i=1}^{D} \frac{1}{\lambda_i} \mathrm{u}_i \mathrm{u}_i{}^T \right) (x-\mu)$$

$$let \quad y_i = \mathrm{u}_i{}^T (x-\mu),$$

$$(x-\mu)^T \Sigma^{-1} (x-\mu) = \sum_{i=1}^{D} \frac{y_i{}^2}{\lambda_i}$$

$$from \quad y_i = \mathrm{u}_i{}^T (x-\mu)$$

$$let \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_D \end{bmatrix}, \qquad U = \begin{bmatrix} \mathrm{u}_1{}^T \\ \vdots \\ \mathrm{u}_D{}^T \end{bmatrix}$$

$$y = U(x-\mu)$$

$$(x-\mu) = U^{-1} y = U^T y = [\mathrm{u}_1 \quad \cdots \quad \mathrm{u}_D] \begin{bmatrix} y_1 \\ \vdots \\ y_D \end{bmatrix} = \sum_{i=1}^{D} y_i \, \mathrm{u}_i$$

now, let's derive the covariance matrix

$$cov[X] = E[(X - E[X])(X - E[X])^T] = E[(X-\mu)(X-\mu)^T]$$
$$= E[XX^T - X\mu^T - \mu X^T + \mu\mu^T] = E[XX^T] - 2\mu\mu^T + \mu\mu^T$$
$$= E[XX^T] - \mu\mu^T$$

$$E[XX^T] = \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right\} x x^T dx$$

assume $z = x - \mu$

$E[XX^T] = \frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}z^T\Sigma^{-1}z\right\}(z+\mu)(z+\mu)^T dz =$

$\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}z^T\Sigma^{-1}z\right\}(zz^T + z\mu^T + \mu z^T + \mu\mu^T)dz$

involving $z\mu^T$ and μz^T will vanish by symmetry like in the process I've shown when deriving the mean.

$$E[XX^T] = \frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}z^T\Sigma^{-1}z\right\}(zz^T + \mu\mu^T)dz$$

$$= \left(\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}z^T\Sigma^{-1}z\right\}zz^T dz\right) + \mu\mu^T$$

$$from \quad z = (x-\mu) = \sum_{i=1}^{D} y_i \mathrm{u}_i \quad and \; (x-\mu)^T\Sigma^{-1}(x-\mu) = \sum_{i=1}^{D}\frac{{y_i}^2}{\lambda_i}$$

$$\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}z^T\Sigma^{-1}z\right\}zz^T dz$$

$$= \frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}\sum_{i=1}^{D}\sum_{j=1}^{D}\mathrm{u}_i{\mathrm{u}_j}^T\int_{-\infty}^{\infty} \exp\left\{-\sum_{k=1}^{D}\frac{{y_k}^2}{2\lambda_k}\right\}y_i y_j dy = \sum_{i=1}^{D}\lambda_i\mathrm{u}_i{\mathrm{u}_i}^T$$

$$= \Sigma$$

now we have

$$cov[X] = E[XX^T] - \mu\mu^T = (\Sigma + \mu\mu^T) - \mu\mu^T = \Sigma$$

because it is the natural generalization to higher dimensions of the 1 dimensional variance, we can also call this the variance.

$$var[X] = cov[X] = \Sigma$$

$$mean = E[X] = \mu$$

$$variance = var[X] = \Sigma$$

b)

given

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \quad and\ D\ is\ invertible$$

$$then, assume\ that\ D's\ shape\ is\ (q \times q)\ and\ A's\ shape\ is\ (p \times p).$$

$$Let \quad L = \begin{pmatrix} I_p & 0 \\ -D^{-1}C & I_q \end{pmatrix} \quad where\ I_l\ represent\ identity\ matrix\ of\ shape\ (l \times l)$$

$$ML = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I_p & 0 \\ -D^{-1}C & I_q \end{pmatrix} = \begin{pmatrix} A - BD^{-1}C & B \\ 0 & D \end{pmatrix}$$

$$also, \ \begin{pmatrix} A - BD^{-1}C & B \\ 0 & D \end{pmatrix} can\ be\ shown\ to\ be$$

$$\begin{pmatrix} A - BD^{-1}C & B \\ 0 & D \end{pmatrix} = \begin{pmatrix} I_p & BD^{-1} \\ 0 & I_q \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix}$$

$$L^{-1} = \begin{pmatrix} I_p & 0 \\ D^{-1}C & I_q \end{pmatrix}$$

$$M = MLL^{-1} = \begin{pmatrix} I_p & BD^{-1} \\ 0 & I_q \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I_p & 0 \\ D^{-1}C & I_q \end{pmatrix}$$

$$finally, we\ can\ derive \quad M^{-1} = \begin{pmatrix} I_p & 0 \\ -D^{-1}C & I_q \end{pmatrix} \begin{pmatrix} \left(A - BD^{-1}C\right)^{-1} & 0 \\ 0 & D^{-1} \end{pmatrix} \begin{pmatrix} I_p & -BD^{-1} \\ 0 & I_q \end{pmatrix}$$

$$= \begin{pmatrix} \left(A - BD^{-1}C\right)^{-1} & 0 \\ -D^{-1}C\left(A - BD^{-1}C\right)^{-1} & D^{-1} \end{pmatrix} \begin{pmatrix} I_p & -BD^{-1} \\ 0 & I_q \end{pmatrix}$$

$$M^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{pmatrix}$$

c)

given

$$x = \begin{pmatrix} x_a \\ x_b \end{pmatrix}, \qquad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

$$Let \ \Sigma^{-1} = \Lambda = \begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix}$$

as I said in (b) Σ and $\Sigma^{-1} = \Lambda$ is symmetric.

so, $\Sigma_{aa}, \Sigma_{bb}, \Lambda_{aa}, \Lambda_{bb}$ are symmetric, and ${\Sigma_{ab}}^T = \Sigma_{ba},\ {\Lambda_{ab}}^T = \Lambda_{ba}$

Let's begin with joint probability $p(x) = p(x_a, x_b)$

First,

$$-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) = -\frac{1}{2}\begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix}^T \begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix} \begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix}$$

$$= -\frac{1}{2}(x_a-\mu_a)^T\Lambda_{aa}(x_a-\mu_a) - \frac{1}{2}(x_a-\mu_a)^T\Lambda_{ab}(x_b-\mu_b)$$

$$-\frac{1}{2}(x_b-\mu_b)^T\Lambda_{ba}(x_a-\mu_a) - \frac{1}{2}(x_b-\mu_b)^T\Lambda_{bb}(x_b-\mu_b)$$

and 'complete the square', exponent in a general Gaussian distribution can be written,

$$-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) = -\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu + constant$$

from this, because this quadratic equation is in the exponent, if we assume x_b is given as a constant, and express the equation from top of the page to this form we can find out the covariance and the mean of $p(x_a|x_b)$.

If we get the term from second order in x_a,

$$-\frac{1}{2}{x_a}^T\Lambda_{aa}x_a \quad matches \quad -\frac{1}{2}x^T\Sigma^{-1}x$$

$$\Sigma_{a|b} = {\Lambda_{aa}}^{-1}$$

now let's consider the term linear in x_a,

$$\frac{1}{2}{x_a}^T\Lambda_{aa}\mu_a + \frac{1}{2}{\mu_a}^T\Lambda_{aa}x_a - \frac{1}{2}{x_a}^T\Lambda_{ab}(x_b-\mu_b) - \frac{1}{2}(x_b-\mu_b)^T\Lambda_{ba}x_a$$

because Λ_{aa} is symmetric and ${\Lambda_{ab}}^T = \Lambda_{ba}$

the equation above

$$= {x_a}^T\{\Lambda_{aa}\mu_a - \Lambda_{ab}(x_b-\mu_b)\}$$

from the general quadratic equation,

$${\Sigma_{a|b}}^{-1}\mu_{a|b} = \Lambda_{aa}\mu_a - \Lambda_{ab}(x_b-\mu_b)$$

$$\mu_{a|b} = {\Lambda_{aa}}^{-1}\Lambda_{aa}\mu_a - {\Lambda_{aa}}^{-1}\Lambda_{ab}(x_b-\mu_b) = \mu_a - {\Lambda_{aa}}^{-1}\Lambda_{ab}(x_b-\mu_b)$$

from

$$\begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}^{-1} = \begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix}, and$$

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A-BD^{-1}C)^{-1} & -(A-BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A-BD^{-1}C)^{-1} & D^{-1}+D^{-1}C(A-BD^{-1}C)^{-1}BD^{-1} \end{pmatrix}$$

$$\Lambda_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}{}^{-1}\Sigma_{ba})^{-1}$$

$$\Lambda_{ab} = -(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}{}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}{}^{-1}$$

now we can get,

$$\Sigma_{a|b} = \Lambda_{aa}{}^{-1} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}{}^{-1}\Sigma_{ba}$$

$$\mu_{a|b} = \mu_a + \left(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}{}^{-1}\Sigma_{ba}\right)\left(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}{}^{-1}\Sigma_{ba}\right)^{-1}\Sigma_{ab}\Sigma_{bb}{}^{-1}(x_b - \mu_b)$$

$$= \mu_a + \Sigma_{ab}\Sigma_{bb}{}^{-1}(x_b - \mu_b)$$

so,

$$mean = \mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}{}^{-1}(x_b - \mu_b)$$

$$variance = \Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}{}^{-1}\Sigma_{ba}$$

d)

$$marginal\, probability \quad p(x_a) = \int p(x_a, x_b)dx_b$$

recall,

$$-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) = -\frac{1}{2}\begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix}^T \begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix}\begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix}$$

$$= -\frac{1}{2}(x_a-\mu_a)^T\Lambda_{aa}(x_a-\mu_a) - \frac{1}{2}(x_a-\mu_a)^T\Lambda_{ab}(x_b-\mu_b)$$

$$-\frac{1}{2}(x_b-\mu_b)^T\Lambda_{ba}(x_a-\mu_a) - \frac{1}{2}(x_b-\mu_b)^T\Lambda_{bb}(x_b-\mu_b)$$

because we need to integrate out x_b, consider only the terms involving x_b.

$$-\frac{1}{2}x_b{}^T\Lambda_{bb}x_b + x_b{}^T(\Lambda_{bb}\mu_b - \Lambda_{ba}(x_a-\mu_a))$$

like we did while (c), consider matching terms only with those involving x_b from quadratic equation in exponent.

to make it easier,

$$Let\ m = \Lambda_{bb}\mu_b - \Lambda_{ba}(x_a - \mu_a)$$

and we can organize the equation like below,

$$-\frac{1}{2}x_b{}^T\Lambda_{bb}x_b + x_b{}^Tm = -\frac{1}{2}\left(x_b - \Lambda_{bb}{}^{-1}m\right)^T\Lambda_{bb}\left(x_b - \Lambda_{bb}{}^{-1}m\right) + \frac{1}{2}m^T\Lambda_{bb}{}^{-1}m$$

we can see that first term is dependent to x_b, second term is not involving x_b, let's consider the first term,

$$-\frac{1}{2}\left(x_b - \Lambda_{bb}{}^{-1}m\right)^T \Lambda_{bb}\left(x_b - \Lambda_{bb}{}^{-1}m\right)$$

then the integration of exponential of this quadratic form over x_b looks like this.

$$\int -\frac{1}{2}\left(x_b - \Lambda_{bb}{}^{-1}m\right)^T \Lambda_{bb}\left(x_b - \Lambda_{bb}{}^{-1}m\right)dx_b$$

This is integral over an un-normalized Gaussian. Result will be the reciprocal of the normalization coefficient. The normalization coefficient is independent of the mean and dependent on the determinant of the covariance matrix. Therefore, we can integrate this out and the only term remaining from the form at the bottom of the last page will be,

$$\begin{aligned}\frac{1}{2}m^T\Lambda_{bb}{}^{-1}m &= \frac{1}{2}\left(\Lambda_{bb}\mu_b - \Lambda_{ba}(x_a-\mu_a)\right)^T \Lambda_{bb}{}^{-1}\left(\Lambda_{bb}\mu_b - \Lambda_{ba}(x_a-\mu_a)\right) \\ &= -\frac{1}{2}x_a{}^T\left(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}{}^{-1}\Lambda_{ba}\right)x_a + x_a{}^T\left(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}{}^{-1}\Lambda_{ba}\right)^{-1}\mu_a \\ &+ const\end{aligned}$$

again, just like how we derived mean and variance of the conditional distribution, let's match this form with the general quadratic form.

$$\Sigma_a = \left(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}{}^{-1}\Lambda_{ba}\right)^{-1}$$

$$\mu_a = \Sigma_a\left(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}{}^{-1}\Lambda_{ba}\right)\mu_a = \mu_a$$

from (b) and

$$\begin{pmatrix}\Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb}\end{pmatrix}^{-1} = \begin{pmatrix}\Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb}\end{pmatrix}$$

$$\left(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}{}^{-1}\Lambda_{ba}\right)^{-1} = \Sigma_{aa}$$

so, the mean and the variance of the marginal distribution $p(x_a)$ is like below.

$$\text{mean} = E[x_a] = \mu_a$$

$$variance = cov[x_a] = \Sigma_{aa}$$

Problem 3. Using the Fig. 1 and the Gaussian distribution $\boldsymbol{\mathcal{N}(x;\ \mu, \sigma^2)}$

a) Let $\boldsymbol{\mu = 5}$ and $\boldsymbol{\sigma = 3}$. Compute the probability that $\boldsymbol{x \in [2.3, 8.3]}$

by normalizing to $Z = \frac{X-\mu}{\sigma}$ we can derive the standard normal distribution

$$\mathcal{N}(Z; 0, 1)$$

$$P(2.3 \le x \le 8.3) = P\left(\frac{2.3-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \le \frac{8.3-\mu}{\sigma}\right) = P\left(\frac{2.3-5}{3} \le z \le \frac{8.3-5}{3}\right)$$

$$= P(-0.9 \le z \le 1.1) = P(0 \le z \le 1.1) + P(-0.9 \le z \le 0)$$

$$= P(0 \le z \le 1.1) + P(0 \le z \le 0.9) = 0.36433 + 0.31594 = 0.68027$$

$$(because\ the\ standard\ normal\ distribution\ is\ symmetric\ respect\ to\ zero)$$

so the answer is,

$$P(2.3 \le x \le 8.3) = 0.68027$$

b) Let $\boldsymbol{\mu = 10}$ and $\boldsymbol{\sigma = 5}$. Compute y$\boldsymbol{\in R_+}$ $\boldsymbol{which\ satisfies\ the\ following}$:

$$\boldsymbol{Pr(x \in [10-y, 10+y]) = 0.95}$$

$$P(10-y \le x \le 10+y) = 0.95$$

$$P(10-y \le x \le 10+y) = P\left(\frac{10-y-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \le \frac{10+y-\mu}{\sigma}\right)$$

$$= P\left(\frac{10-y-10}{5} \le z \le \frac{10+y-10}{5}\right) = P\left(-\frac{y}{5} \le z \le \frac{y}{5}\right)$$

$$= P\left(0 \le z \le \frac{y}{5}\right) + P\left(-\frac{y}{5} \le z \le 0\right)$$

$$= P\left(0 \le z \le \frac{y}{5}\right) + P\left(0 \le z \le \frac{y}{5}\right) = 2P\left(0 \le z \le \frac{y}{5}\right) = 0.95$$

$$P\left(0 \le z \le \frac{y}{5}\right) = 0.475$$

from the table,

$$P(0 \le z \le 1.96) = 0.475$$

$$\frac{y}{5} = 1.96$$

so the answer is,

$$y = 9.8$$