RANDOM DISTRIBUTIONS

Lecture Notes

Ahmet Ademoglu, *PhD*
Bogazici University
Institute of Biomedical Engineering

UNIVARIATE DISTRIBUTIONS

Gaussian Distribution

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

Show that the area under $\mathcal{N}(x|\mu,\sigma^2) = 1$.

Making a change of variable $x_1 = x_2 = x - \mu$, we define

$$I = \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}x_1^2\right\} dx_1 = \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}x_2^2\right\} dx_2$$

$$I^2 = \iint_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}(x_1^2 + x_2^2)\right\} dx_1 dx_2$$

In polar coordinates,

$$I^2 = \int_0^{\infty}\int_0^{2\pi} \exp\left\{-\frac{1}{2\sigma^2}r^2\right\} rdrd\theta = \int_0^{\infty}\int_0^{2\pi} \exp\left\{-\left(\frac{r}{(2\sigma^2)^{1/2}}\right)^2\right\} d\left(\frac{r}{(2\sigma^2)^{1/2}}\right)^2 \sigma^2 d\theta$$

$I^2 = \int_0^{\infty}\int_0^{2\pi} \exp\{-v\}\, dv\sigma^2 d\theta = 2\pi\sigma^2$ and $I = (2\pi\sigma^2)^{1/2}$.

Show that the mean and variance of Gaussian distribution is μ and σ^2, respectively.

Differentiating both sides of

$$\int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx = 1 \text{ w.r.t. } \sigma^2,$$

we get

$$\int_{-\infty}^{\infty} (-1/2) \frac{1}{(2\pi)^{1/2}(\sigma^2)^{3/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx$$

$$+ \int_{-\infty}^{\infty} \frac{1}{2(\sigma^2)^2}(x-\mu)^2 \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx = 0 \text{ which yields}$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx$$

For the mean,

$$E\{x\} = \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} x \cdot dx$$

Changing variable $z = x - \mu$

$$E\{x\} = E\{z+\mu\} = \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}z^2\right\} (z+\mu) dz = \mu.$$

The first integral is 0 because of the oddness of its integrand, and the second one is μ.

$$E\{(x-\mu)^2\} = \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} (x-\mu)^2 \cdot dx.$$

Bernoulli Distribution (*Flipping a coin*)

$$Bern(x|\mu) = \mu^x(1-\mu)^{1-x}$$

Show that the Bernouilli distribution has a mean and variance, μ and $\mu(1-\mu)$, respectively.

$E\{x\} = P(x=0)\cdot 0 + P(x=1)\cdot 1 = 1\cdot(1-\mu)\cdot 0 + \mu\cdot 1\cdot 1 = \mu$
$E\{x^2\} = P(x=0)\cdot 0^2 + P(x=1)\cdot 1^2 = \mu$
$Var\{x\} = E\{x^2\} - E\{x\}^2 = \mu - \mu^2 = \mu(1-\mu)$

Binomial Distribution (*Flipping Coin N times*)

$$Bin(m|N,\mu) = \begin{pmatrix} \text{N} \\ \text{m} \end{pmatrix} \mu^m(1-\mu)^{N-m}$$

Show that the Binomial distribution has a mean and variance, $N\mu$ and $N\mu(1-\mu)$, respectively.

$$E\{m\} = \sum_{m=0}^{N} P(m)m = \sum_{m=1}^{N} \begin{pmatrix} \text{N} \\ \text{m} \end{pmatrix} m\mu^m(1-\mu)^{N-m} = \sum_{p=0}^{N-1} \begin{pmatrix} \text{N} \\ \text{p+1} \end{pmatrix} (p+1)\mu^{p+1}(1-\mu)^{N-p-1}$$

$$E\{m\} = N\mu \sum_{p=0}^{N-1} \frac{(N-1)!}{p!(N-1-p)!}\mu^p(1-\mu)^{N-1-p} = N\mu\cdot 1 = N\mu.$$

$$E\{m^2\} = N\mu \sum_{p=0}^{N-1} (p+1)\frac{(N-1)!}{(p)!(N-p-1)!}\mu^p(1-\mu)^{N-p-1} = N\mu[(N-1)\mu+1] = N^2\mu^2 - N\mu^2 + N\mu$$

Poisson Distribution

$$Poisson(k|\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}$$

Show that the Poisson distribution has a mean and variance both equal to λ.

$$E\{k\} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda}k = \sum_{k=1}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda}k = \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{(k+1)!}e^{-\lambda}(k+1) = \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda} = \lambda \cdot 1$$

$$E\{k^2\} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda}k^2 = \sum_{k=1}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda}k^2 = \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{(k+1)!}e^{-\lambda}(k+1)^2$$

$$= \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda}(k+1) = \lambda(\lambda+1)$$

$$Var\{k\} = E\{k\}^2 - E\{k\}^2 = \lambda(\lambda+1) - \lambda^2 = \lambda$$

Beta Distribution (μ over $[0,1]$)

$$Beta(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1} \text{ with } \Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$$

Show that the Beta distribution has a mean $= \frac{a}{a+b}$ and variance $= \frac{ab}{(a+b)^2(a+b+1)}$.

$$E\{\mu\} = \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}\mu d\mu = \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a+1-1}(1-\mu)^{b-1}d\mu$$

$$\Gamma(z+1) = \int_0^\infty t^z e^{-t}dt = -t^z e^{-t}\Big|_0^\infty + z\int_0^\infty t^{z-1}e^{-t}dt = z\Gamma(z)$$

$$E\{\mu\} = \int_0^1 \frac{a}{a+b}\frac{\Gamma(a+1+b)}{\Gamma(a+1)\Gamma(b)}\mu^{a+1-1}(1-\mu)^{b-1}d\mu = \frac{a}{a+b}\cdot 1 = \frac{a}{a+b}$$

$$E\{\mu^2\} = \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}\mu^2 d\mu = \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a+2-1}(1-\mu)^{b-1}d\mu$$

$$= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}\mu^2 d\mu = \int_0^1 \frac{a(a+1)}{(a+b)(a+b+1)}\frac{\Gamma(a+2+b)}{\Gamma(a+2)\Gamma(b)}\mu^{a+2-1}(1-\mu)^{b-1}d\mu$$

$$= \frac{a(a+1)}{(a+b)(a+b+1)}$$

$$E\{\mu^2\} - E\{\mu\}^2 = \frac{a(a+1)}{(a+b)(a+b+1)} - \left(\frac{a}{a+b}\right)^2 = \frac{ab}{(a+b)^2(a+b+1)}$$

Gamma Distribution

$Gam(\lambda|a,b) = \frac{1}{\Gamma(a)} b^a \lambda^{a-1} e^{-b\lambda}$

Show that the Gamma distribution has a mean $= \frac{a}{b}$ and variance $= \frac{a}{b^2}$.

$E\{\lambda\} = \int\limits_0^\infty \frac{1}{\Gamma(a)} b^a \lambda^{a-1} e^{-b\lambda} \lambda d\lambda = \frac{a}{b} \int\limits_0^\infty \frac{1}{\Gamma(a+1)} b^{a+1} \lambda^{a+1-1} e^{-b\lambda} d\lambda = \frac{a}{b} \cdot 1$

$E\{\lambda^2\} = \int\limits_0^\infty \frac{1}{\Gamma(a)} b^a \lambda^{a-1} e^{-b\lambda} \lambda^2 d\lambda = \frac{a(a+1)}{b^2} \int\limits_0^\infty \frac{1}{\Gamma(a)} b^{a+2} \lambda^{a+2-1} e^{-b\lambda} d\lambda = \frac{a(a+1)}{b^2} \cdot 1$

$E\{\lambda^2\} - E\{\lambda\}^2 = \frac{a(a+1)}{b^2} - \left(\frac{a}{b}\right)^2 = \frac{a}{b^2}.$

Inverse Gamma Distribution

$$IGam(\lambda|a,b) = \frac{b^a}{\Gamma(a)}\lambda^{-a-1}e^{-\frac{b}{\lambda}}$$

Show that Inverse Gamma distribution has a mean $= \frac{b}{a-1}$ and variance $= \frac{b^2}{(a-1)^2(a-2)}$.

$$E\{\lambda\} = \int_0^\infty \frac{b^a}{\Gamma(a)}\lambda^{-a-1}e^{-\frac{b}{\lambda}}\lambda d\lambda = \int_0^\infty \frac{b^a}{\Gamma(a)}\lambda^{-a+1-1}e^{-\frac{b}{\lambda}}d\lambda$$

$$= \frac{b}{a-1}\int_0^\infty \frac{b^{a-1}}{\Gamma(a-1)}\lambda^{-a+1-1}e^{-\frac{b}{\lambda}}d\lambda = \frac{b}{a-1}\cdot 1 = \frac{b}{a-1}$$

$$E\{\lambda^2\} = \int_0^\infty \frac{b^a}{\Gamma(a)}\lambda^{-a-1}e^{-\frac{b}{\lambda}}\lambda^2 d\lambda$$

$$= \int_0^\infty \frac{b^a}{\Gamma(a)}\lambda^{-a+2-1}e^{-\frac{b}{\lambda}}d\lambda$$

$$= \frac{b^2}{(a-1)(a-2)}\int_0^\infty \frac{b^{a-2}}{\Gamma(a-2)}\lambda^{-a+2-1}e^{-\frac{b}{\lambda}}d\lambda = \frac{b}{a-1}\cdot 1 = \frac{b}{a-1} = \frac{b^2}{(a-1)(a-2)}\cdot 1$$

$$= \frac{b^2}{(a-1)(a-2)} - \left(\frac{b}{a-1}\right)^2 = \frac{b^2}{(a-1)^2(a-2)}$$

Student's t Distribution : (Infinite mixture of Gaussians)

$$\int_0^\infty \mathcal{N}(x|\mu,\tau^{-1}) Gam(\tau|a,b)d\tau = \int_0^\infty \frac{b^a e^{-b\tau}\tau^{a-1}}{\Gamma(a)} \left(\frac{\tau}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\tau}{2}(x-\mu)^2} d\tau$$

$$= \frac{b^a}{\Gamma(a)(2\pi)^{\frac{1}{2}}} \int_0^\infty e^{-(b+\frac{(x-\mu)^2}{2})\tau} \tau^{(a+\frac{1}{2}-1)} d\tau$$

$$= \frac{b^a}{\Gamma(a)(2\pi)^{\frac{1}{2}}(b+\frac{(x-\mu)^2}{2})^{(a+\frac{1}{2})}} \int_0^\infty e^{-t} t^{a+\frac{1}{2}-1} dt =$$

$$\frac{b^a \left[b+\frac{(x-\mu)^2}{2}\right]^{-(a+\frac{1}{2})} \Gamma(a+\frac{1}{2})}{\Gamma(a)(2\pi)^{\frac{1}{2}}}$$

Defining $\lambda = \frac{a}{b}$, $\nu = 2a$

$$St(x|\mu,\lambda,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(\frac{\lambda}{\pi\nu}\right)^{1/2} \left[1 + \frac{\lambda(x-\mu)^2}{\nu}\right]^{-\frac{\nu+1}{2}},$$

with ν *degrees of freedom*,

Show that the Student's t distribution has a mean μ and variance $\frac{1}{\lambda}\frac{\nu}{\nu-2}$.

$$E\{x\} = \int\limits_{-\infty}^{\infty} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\left(\frac{\lambda}{\pi\nu}\right)^{1/2}\left[1+\frac{\lambda(x-\mu)^2}{\nu}\right]^{-\frac{\nu+1}{2}} xdx \qquad \text{Replace } z = x-\mu,$$

$$E\{x\} = \int\limits_{-\infty}^{\infty} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\left(\frac{\lambda}{\pi\nu}\right)^{1/2}\left[1+\frac{\lambda}{\nu}z^2\right]^{-\frac{\nu+1}{2}} (z+\mu)dz = \int\limits_{-\infty}^{\infty} p(z|0,\lambda,\nu)(z+\mu)dz$$

$$\int\limits_{-\infty}^{\infty} p(z|0,\lambda,\nu)zdz = \int\limits_{-\infty}^{0} p(z|0,\lambda,\nu)zdz + \int\limits_{0}^{\infty} p(z|0,\lambda,\nu)zdz, \qquad \text{Replace } z = -t$$

$$= \int\limits_{\infty}^{0} p(-t|0,\lambda,\nu)tdt + \int\limits_{0}^{\infty} p(z|0,\lambda,\nu)zdz = -\int\limits_{0}^{\infty} p(-t|0,\lambda,\nu)tdt + \int\limits_{0}^{\infty} p(z|0,\lambda,\nu)zdz,$$

Integrals are finite for $\nu > 1$ and since p is symmetric,

$$= -\int\limits_{0}^{\infty} p(t|0,\lambda,\nu)tdt + \int\limits_{0}^{\infty} p(z|0,\lambda,\nu)zdz = 0, \longrightarrow E\{x\} = \mu$$

Self Study Question: Prove the following identities.

$$\beta(a,b) = \int\limits_{0}^{\infty} t^{a-1}(1+t)^{-(a+b)}dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

Identities for beta function $\beta(a,b) = \int_0^\infty t^{a-1}(1+t)^{-(a+b)}dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Replacing $z = x - \mu$, $E\{x^2\} = \int_{-\infty}^{\infty} p(x|\mu,\nu)x^2dx = \int_{-\infty}^{\infty} p(z|0,\nu)(z+\mu)^2dz$

$$\int_{-\infty}^{\infty} p(z|\nu)z^2dz = \int_{-\infty}^{0} p(z|\nu)z^2dz + \int_0^\infty p(z|0,\lambda,\nu), z^2dz$$

Replace $z = -t$ in the first term and use symmetry

$$= -\int_\infty^0 p(-t|0,\lambda,\nu)t^2dt + \int_0^\infty p(z|0,\lambda,\nu)z^2dz = 2\int_0^\infty p(z|0,\lambda,\nu)z^2dz$$

Replace $t = \frac{\lambda}{\nu}z^2$ so that $(\frac{\nu}{\lambda}t)^{\frac{1}{2}} = z$ and $(\frac{\nu}{\lambda})^{\frac{1}{2}}\frac{1}{2}t^{-\frac{1}{2}}dt = dz$

$$= 2\int_0^\infty \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\left(\frac{\lambda}{\pi\nu}\right)^{1/2}[1+t]^{-\frac{\nu+1}{2}}(\frac{\nu}{\lambda})t(\frac{\nu}{\lambda})^{\frac{1}{2}}\frac{1}{2}t^{-\frac{1}{2}}dt$$

$$= \frac{\nu}{\lambda}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\Gamma(\frac{1}{2})}\int_0^\infty t^{\frac{3}{2}-1}[1+t]^{-\frac{3}{2}-(\frac{\nu}{2}-1)}dt \;, = \nu\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\Gamma(\frac{1}{2})}\frac{\Gamma(\frac{1}{2}+1)\Gamma(\frac{\nu}{2}-1)}{\Gamma(\frac{\nu+1}{2})}$$

$$= \frac{\nu}{\lambda}\frac{\Gamma(\frac{\nu+1}{2})}{(\frac{\nu}{2}-1)\Gamma(\frac{\nu}{2}-1)\Gamma(\frac{1}{2})}\frac{\frac{1}{2}\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2}-1)}{\Gamma(\frac{\nu+1}{2})} = \frac{1}{\lambda}\frac{\nu}{\nu-2} \rightarrow E\{x^2\} = \frac{1}{\lambda}\frac{\nu}{\nu-2} + \mu^2$$

$$Var\{x\} = E\{x^2\} - E\{x\}^2 = \frac{1}{\lambda}\frac{\nu}{\nu-2} + \mu^2 - \mu^2 = \frac{1}{\lambda}\frac{\nu}{\nu-2}$$

Multivariate Distributions

Multinomial Distribution : (Throwing a dice: $\mathbf{x} = [0\,0\,1\,0\,0\,0]^T$)

$$Mult(m_1, m_2, \ldots, m_K|\mu, N) = \begin{pmatrix} N \\ m_1 m_2 \ldots m_K \end{pmatrix} \prod_{i=1}^{K} \mu_k^{m_k}$$

$$\sum_{i=1}^{K} \mu_i = 1 \text{ and } \sum_{i=1}^{K} m_i = N.$$

A binary random vector $x = [x_1\ x_2\ x_3]$ is observed $N = 10$ times as it was given below; What is the probability of occurence of $x_1 = 1$ to be $m_1 = 4$, $x_2 = 1$, $m_2 = 3$ and $x_3 = 1$, $m_3 = 3$.

0	1	0
0	0	1
0	1	0
1	0	0
1	0	0
0	0	1
1	0	0
0	1	0
1	0	0
0	0	1

$$\begin{pmatrix} N \\ m_1 \end{pmatrix} \begin{pmatrix} N - m_1 \\ m_2 \end{pmatrix} \begin{pmatrix} N - m_1 - m_2 \\ m_3 \end{pmatrix} = \begin{pmatrix} N \\ m_1 m_2 m_3 \end{pmatrix}$$

For each occurence the probability is $\mu_1^{m_1} \mu_2^{m_2} \mu_3^{m_3}$

The total probability $= \begin{pmatrix} N \\ m_1 m_2 m_3 \end{pmatrix} \mu_1^{m_1} \mu_2^{m_2} \mu_3^{m_3}$

$$= \frac{10!}{4!3!3!} \mu_1^4 \mu_2^3 \mu_3^3$$

Show that the Multinomial distribution has a mean $m_k = N\mu_k$ and variance $N\mu_k(1-\mu_k)$.

Choose $k=1$, the proof will be similar for any m_k.

$$E\{m_1\} = \sum_{m_1=0}^{N} \sum_{m_2=0}^{N-m_1} \sum_{m_3=0}^{N-m_1-m_2} \cdots \begin{pmatrix} \mathrm{N} \\ m_1 m_2 \ldots m_K \end{pmatrix} \prod_{k=1}^{K} \mu_k^{m_k} m_1$$

$$= \sum_{m_1=0}^{N} \frac{N}{m_1!} m_1 \mu_1^{m_1} \sum_{m_2=0}^{N-m_1} \sum_{m_3=0}^{N-m_1-m_2} \cdots \frac{(N-1)!}{m_2! \ldots m_k! (N-m_1-\cdots-m_k)!} \prod_{k=2}^{K} \mu_k^{m_k}$$

$$= \sum_{m_1=1}^{N} \frac{N}{m_1!} m_1 \mu_1^{m_1} \sum_{m_2=0}^{N-m_1} \sum_{m_3=0}^{N-m_1-m_2} \cdots \frac{(N-1)!}{m_2! \ldots m_k! (N-m_1-\cdots-m_k)!} \prod_{k=2}^{K} \mu_k^{m_k}$$

$$= \sum_{m_1=0}^{N-1} \frac{N}{(m_1+1)!} (m_1+1) \mu_1^{m_1+1} \sum_{m_2=0}^{N-1-m_1} \sum_{m_3=0}^{N-1-m_1-m_2} \cdots \frac{(N-1)!}{m_2! \ldots m_k! (N-1-m_1-\cdots-m_k)!} \prod_{k=2}^{K} \mu_k^{m_k}$$

$$= N\mu_1 \sum_{m_1=0}^{N-1} \mu_1^{m_1} \sum_{m_2=0}^{N-1-m_1} \sum_{m_3=0}^{N-1-m_1-m_2} \cdots \frac{(N-1)!}{m_1! m_2! \ldots m_k! (N-1-m_1-\cdots-m_k)!} \prod_{k=2}^{K} \mu_k^{m_k}$$

$$= N\mu_1 \sum_{m_1=0}^{N-1} \sum_{m_2=0}^{N-1-m_1} \sum_{m_3=0}^{N-1-m_1-m_2} \cdots \begin{pmatrix} N-1 \\ m_1 m_2 \ldots m_K \end{pmatrix} \prod_{k=1}^{K} \mu_k^{m_k}$$

$$= N\mu_1 \cdot (\mu_1 + \mu_2 + \cdots + \mu_K)^{N-1} = N\mu_1$$

$$E\{m_1^2\} = \sum_{m_1=0}^{N} \sum_{m_2=0}^{N-m_1} \sum_{m_3=0}^{N-m_1-m_2} \cdots \left(\begin{array}{c} \text{N} \\ m_1 m_2 \ldots m_K \end{array} \right) \prod_{k=2}^{K} \mu_k^{m_k} m_1^2$$

$$= \sum_{m_1=1}^{N} \frac{N}{m_1!} m_1^2 \mu_1^{m_1} \sum_{m_2=0}^{N-m_1} \sum_{m_3=0}^{N-m_1-m_2} \cdots \frac{(N-1)!}{m_2! \ldots m_k! (N-m_1-\cdots-m_k)!} \prod_{k=2}^{K} \mu_k^{m_k}$$

$$= \sum_{m_1=0}^{N-1} \frac{N!}{(m_1+1)!} (m_1+1)^2 \mu_1^{m_1+1} \sum_{m_2=0}^{N-1-m_1} \sum_{m_3=0}^{N-1-m_1-m_2} \cdots \frac{(N-1)!}{m_2! \ldots m_k! (N-1-m_1-\cdots-m_k)!} \prod_{k=1}^{K} \mu_k^{m_k}$$

$$= N\mu_1 \sum_{m_1=0}^{N-1} \frac{(N-1)!}{m_1!} (m_1+1) \mu_1^{m_1} \sum_{m_2=0}^{N-1-m_1} \sum_{m_3=0}^{N-1-m_1-m_2} \cdots \prod_{k=2}^{K} \mu_k^{m_k}$$

$$= N\mu_1 \left[\sum_{m_1=0}^{N-1} \sum_{m_2=0}^{N-1-m_1} \sum_{m_3=0}^{N-1-m_1-m_2} \cdots \left(\begin{array}{c} \text{N-1} \\ m_1 m_2 \ldots m_K \end{array} \right) \prod_{k=1}^{K} \mu_k^{m_k} (m_1+1) \right]$$

$$E\{m_1^2\} = N\mu_1[(N-1)\mu_1 + (\mu_1 + \mu_2 + \cdots + \mu_K)^{N-1}] = N\mu_1[(N-1)\mu_1 + 1]$$

$$E\{m_1^2\} - E\{m_1\}^2 = (N\mu_1)^2 - N\mu_1^2 + N\mu_1 - (N\mu_1)^2 = N\mu_1(1-\mu_1)$$

Dirichlet distribution

Dirichlet is a multivariate distribution over K random variables $0 \leq \mu_k \leq 1$, where $k = 1, \ldots, K$, subject to the constraints

$$0 \leq \mu_k \leq 1 \text{ and } \sum_{k=1}^{K} \mu_k = 1.$$

$\mu = [\mu_1 \, \mu_2 \ldots \mu_K]^T$ and $\alpha = [\alpha_1 \, \alpha_2 \ldots \alpha_K]^T$

$$Dir(\mu|\alpha) = C(\alpha) \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}$$

with $C(\alpha) = \frac{\Gamma(\hat{\alpha})}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_K)}$ and $\hat{\alpha} = \sum_{k=1}^{K} \alpha_k$

Show that the Dirichlet distribution has a mean $\frac{\alpha_k}{\hat{\alpha}}$ and variance $\frac{\alpha_k(\hat{\alpha}-\alpha_k)}{\hat{\alpha}^2(\hat{\alpha}+1)}$

$$E\{\mu_1\} = \int_0^1 C(\alpha) \prod_{k=1}^{K} \mu_k^{\alpha_k-1} \mu_1 d\mu = \int_0^1 C(\alpha)\mu_1^{1+\alpha_1-1} \dots \mu_K^{\alpha_K-1} d\mu_1 \dots d\mu_k$$

$$= \int_0^1 \frac{\Gamma(\alpha_1+\dots+\alpha_K)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_K)} \mu_1^{1+\alpha_1-1} \dots \mu_K^{\alpha_K-1} d\mu_1 \dots d\mu_k$$

$$= \frac{\alpha_1}{\alpha_1+\dots+\alpha_K} \int_0^1 \frac{\Gamma(1+\alpha_1+\dots+\alpha_K)}{\Gamma(1+\alpha_1)\dots\Gamma(\alpha_K)} \mu_1^{1+\alpha_1-1} \dots \mu_K^{\alpha_K-1} d\mu_1 \dots d\mu_k = \frac{\alpha_1}{\hat{\alpha}}$$

$$E\{\mu_1^2\} = \int_0^1 C(\alpha) \prod_{k=1}^{K} \mu_k^{\alpha_k-1} \mu_1^2 d\mu = \int_0^1 C(\alpha)\mu_1^{2+\alpha_1-1} \dots \mu_K^{\alpha_K-1} d\mu_1 \dots d\mu_k$$

$$= \int_0^1 \frac{\Gamma(\alpha_1+\dots+\alpha_K)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_K)} \mu_1^{2+\alpha_1-1} \dots \mu_K^{\alpha_K-1} d\mu_1 \dots d\mu_k$$

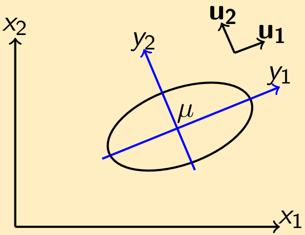
$$= \frac{\alpha_1(\alpha_1+1)}{\hat{\alpha}(1+\hat{\alpha})} \int_0^1 \frac{\Gamma(2+\alpha_1+\dots+\alpha_K)}{\Gamma(2+\alpha_1)\dots\Gamma(\alpha_K)} \mu_1^{2+\alpha_1-1} \dots \mu_K^{\alpha_K-1} d\mu_1 \dots d\mu_k = \frac{\alpha_1(\alpha_1+1)}{\hat{\alpha}(1+\hat{\alpha})}$$

$$Var\{\mu_1\} = E\{\mu_1^2\} - E\{\mu_1\}^2 = \frac{\alpha_1(\alpha_1+1)}{\hat{\alpha}(1+\hat{\alpha})} - \frac{\alpha_1^2}{\hat{\alpha}^2} = \frac{\alpha_k(\hat{\alpha}-\alpha_k)}{\hat{\alpha}^2(\hat{\alpha}+1)}$$

Multivariate Gaussian Distribution

$$\mathcal{N}(\mathbf{x}|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)}$$

Geometry of Multivariate Gaussian



$$\Delta^2 = (\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)$$

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

$$y_i = \mathbf{u}_i^T(\mathbf{x}-\mu)$$

$$\Delta^2 = \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}$$

Show that $E\{\mathbf{x}\} = \mu$ for a Multivariate Gaussian.

$$E\{\mathbf{x}\} = \int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)}\mathbf{x}d\mathbf{x} = \int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{z}^T\Sigma^{-1}\mathbf{z}}(\mathbf{z}+\mu)d\mathbf{z}$$

$$= \int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{z}^T\Sigma^{-1}\mathbf{z}}\mathbf{z}d\mathbf{z} + \mu\int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{z}^T\Sigma^{-1}\mathbf{z}}d\mathbf{z} = 0 + \mu\cdot 1 = \mu$$

Show that $Var\{\mathbf{x}\} = \Sigma$ for a Multivariate Gaussian.

$\Sigma = \sum_{i=1}^{D} \lambda_i \mathbf{u}_i \mathbf{u}_i^T, \quad \mathbf{x} - \mu = \mathbf{U}\mathbf{y} = \sum_{i=1}^{D} \mathbf{u}_i y_i \quad \frac{\partial x_i}{\partial y_j} = \mathbf{J}_{ij} = \mathbf{U}_{ji}, \quad |\mathbf{J}| = |\mathbf{U}^T|$ since $\mathbf{U}^T = \mathbf{U}^{-1}$

$\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $|\mathbf{U}^T\mathbf{U}| = |\mathbf{U}^T||\mathbf{U}| = 1, \ |\mathbf{J}| = 1 \ \ dx_i = |\frac{\partial x_i}{\partial y_j}| dy_j$ or $d\mathbf{x} = |\mathbf{J}| d\mathbf{y} = d\mathbf{y}$

Also $p(\mathbf{y}) = |\mathbf{J}| p(\mathbf{x})$

$$E\{(\mathbf{x}-\mu)(\mathbf{x}-\mu)^T\} = \int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)} (\mathbf{x}-\mu)(\mathbf{x}-\mu)^T d\mathbf{x}$$

$$= \frac{1}{(2\pi)^{D/2}\prod_{p=1}^{D}\lambda_p^{1/2}} \int \sum_{i,j} \mathbf{u}_i y_i y_j \mathbf{u}_j^T e^{-\frac{1}{2}\sum_{k=1}^{D}\frac{y_k^2}{\lambda_k}} d\mathbf{y}$$

$$= \frac{1}{(2\pi)^{D/2}\prod_{p=1}^{D}\lambda_p^{1/2}} \left\{ \left[\mathbf{u}_1\mathbf{u}_1^T \int y_1^2 e^{\frac{y_1^2}{2\lambda_1}} dy_1 \int e^{\frac{y_2^2}{2\lambda_2}} dy_2 \dots \right] + \right.$$

$$\left. \left[\mathbf{u}_2\mathbf{u}_2^T \int e^{\frac{y_1^1}{2\lambda_1}} dy_1 \int y_2^2 e^{\frac{y_2^2}{2\lambda_2}} dy_2 \dots \right] + \dots + \left[\mathbf{u}_D\mathbf{u}_D^T \int e^{\frac{y_1^1}{2\lambda_1}} dy_1 \dots \int y_D^2 e^{\frac{y_D^2}{2\lambda_D}} dy_D \right] \right\}$$

$$= \mathbf{u}_1\mathbf{u}_1^T\lambda_1 + \mathbf{u}_2\mathbf{u}_2^T\lambda_2 + \dots = \sum_{i=1}^{D} \lambda_i \mathbf{u}_i \mathbf{u}_i^T = \Sigma$$

Show that $\frac{1}{(2\pi\lambda)^{1/2}} \int y^2 e^{-\frac{y^2}{2\lambda}} dy = \lambda$

Wishart Distribution with ν degrees of freedom

$$\mathcal{W}(\Lambda|\mathbf{W},\nu) = B|\Lambda|^{(\nu-D-1)/2} e^{-\frac{1}{2}Tr(\mathbf{W}^{-1}\Lambda)}$$

$$B = |\mathbf{W}|^{-\nu/2}\left(2^{\nu D/2}\pi^{D(D-1)/4}\prod_{i=1}^{D}\Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1}$$

Normal-Wishart Distribution

$$p(\mu,\Lambda|\mu_0,\beta,\mathbf{W},\nu) = \mathcal{N}(\mu|\mu_0,(\beta\Lambda)^{-1})\mathcal{W}(\Lambda|\mathbf{W},\nu)$$

$\Lambda = \Sigma^{-1}$: Precision matrix

Multivariate Student's t Distribution

$$St(\mathbf{x}|\mu,\Lambda,\nu) = \frac{\Gamma(\frac{D+\nu}{2})}{\Gamma(\frac{\nu}{2})}\frac{|\Lambda|^{\frac{1}{2}}}{(\pi\nu)^{\frac{D}{2}}}\left[1+\frac{\Delta^2}{\nu}\right]^{-\frac{D+\nu}{2}}$$

$$\Delta^2 = (\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)$$

Conditional Gaussian Distributions

Assume $\mathbf{x}$ is a D dimensional vector with Gaussian distribution $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mu, \Sigma)$ and that we partition x into two disjoint subsets $\mathbf{x} = [\mathbf{x}_a \; \mathbf{x}_b]^T$ with mean and covariance matrix

$$\mu = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix} = \Lambda^{-1}$$

where $\Lambda = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}$ is the precision matrix.

Determine the conditional density $p(\mathbf{x}_a|\mathbf{x}_b) = p(\mathbf{x}_a, \mathbf{x}_b)p(\mathbf{x}_b)$ in terms of μ and Λ.

$-\frac{1}{2}(\mathbf{x}-\mu)^T\Lambda(\mathbf{x}-\mu) = -\frac{1}{2}(\mathbf{x}_a-\mu_a)^T\Lambda_{aa}(\mathbf{x}_a-\mu_a) - \frac{1}{2}(\mathbf{x}_a-\mu_a)^T\Lambda_{ab}(\mathbf{x}_b-\mu_b)$
$-\frac{1}{2}(\mathbf{x}_b-\mu_b)^T\Lambda_{ba}(\mathbf{x}_a-\mu_a) - \frac{1}{2}(\mathbf{x}_b-\mu_b)^T\Lambda_{bb}(\mathbf{x}_b-\mu_b)$
$= -\frac{1}{2}\mathbf{x}_a^T\Lambda_{aa}\mathbf{x}_a + \mathbf{x}_a^T\Lambda_{aa}\mu_a \; -\mathbf{x}_a^T\Lambda_{ab}(\mathbf{x}_b-\mu_b) + const$

If we assume that $p(\mathbf{x}_a|\mathbf{x}_b) = \mathcal{N}(\mathbf{x}_a|\mu_{a|b}, \Sigma_{a|b})$ then

$-\frac{1}{2}(\mathbf{x}_a-\mu_{a|b})^T\Sigma_{a|b}^{-1}(\mathbf{x}_a-\mu_{a|b}) = -\frac{1}{2}\mathbf{x}_a^T\Sigma_{a|b}^{-1}\mathbf{x}_a + \mathbf{x}_a^T\Sigma_{a|b}^{-1}\mu_{a|b} + const$

equating the quadratic and linear terms of $\mathbf{x}_a$:

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

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

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

Using *Schur's Complement*

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{M} & -\mathbf{MBD}^{-1} \\ -\mathbf{D}^{-1}\mathbf{CM} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{CMBD}^{-1} \end{pmatrix}$$

where $\mathbf{M} = (\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1}$

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

$$\mu_{a|b} = \Lambda_{aa}^{-1}\Sigma_{a|b}^{-1}\mu_a + \Lambda_{aa}^{-1}\Lambda_{ab}(\mathbf{x}_b - \mu_b) = \mu_a - \Sigma_{ab}\Sigma_{bb}^{-1}(\mathbf{x}_b - \mu_b)$$

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

Gaussian conditional distribution $p(\mathbf{x}_a|\mathbf{x}_b) = \mathcal{N}(\mathbf{x}_a|\mathbf{A}\mathbf{x}_b + \mathbf{b}, \mathbf{L}^{-1})$,

The Marginal Distribution of Multivariate Gaussian Distributions

$p(\mathbf{x}_a) = \int p(\mathbf{x}_a, \mathbf{x}_b) d\mathbf{x}_b$: obtained by integrating terms with $\mathbf{x}_b$.

Determine the marginal distribution $p(\mathbf{x}_a)$.

Exponential part of $p(\mathbf{x}_a, \mathbf{x}_b)$;

$$-\frac{1}{2}(\mathbf{x}_a - \mu_a)^T \Lambda_{aa}(\mathbf{x}_a - \mu_a) - \frac{1}{2}(\mathbf{x}_a - \mu_a)^T \Lambda_{ab}(\mathbf{x}_b - \mu_b) - \frac{1}{2}(\mathbf{x}_b - \mu_b)^T \Lambda_{ba}(\mathbf{x}_a - \mu_a)$$
$$-\frac{1}{2}(\mathbf{x}_b - \mu_b)^T \Lambda_{bb}(\mathbf{x}_b - \mu_b)$$

Collect quadratic and linear terms in $\mathbf{x}_b$;

$$-\frac{1}{2}\mathbf{x}_b^T \Lambda_{bb}\mathbf{x}_b - \mathbf{x}_b^T \Lambda_{ba}(\mathbf{x}_a - \mu_a) + \mathbf{x}_b^T \Lambda_{bb}\mu_b = -\frac{1}{2}\mathbf{x}_b^T \Lambda_{bb}\mathbf{x}_b + \mathbf{x}_b^T \underbrace{(\Lambda_{bb}\mu_b - \Lambda_{ba}(\mathbf{x}_a - \mu_a))}_{\mathbf{m}}$$

Add and subtract $(\frac{1}{2}\mathbf{m}^T \Lambda_{bb}^{-1})\Lambda_{bb}(\Lambda_{bb}^{-1}\mathbf{m})$ to make it a full Gaussian;

$$-\frac{1}{2}(\mathbf{x}_b - \Lambda_{bb}^{-1}\mathbf{m})^T \Lambda_{bb}(\mathbf{x}_b - \Lambda_{bb}^{-1}\mathbf{m}) + \frac{1}{2}\mathbf{m}^T \Lambda_{bb}^{-1}\mathbf{m}$$

Integrate the above full Gaussian term which is a function of $\mathbf{x}_b$;

$$\int \exp\{-\frac{1}{2}(\mathbf{x}_b - \Lambda_{bb}^{-1}\mathbf{m})^T \Lambda_{bb}(\mathbf{x}_b - \Lambda_{bb}^{-1}\mathbf{m})\} d\mathbf{x}_b = (2\pi)^{D_b/2}|\Lambda_{bb}|^{-1/2}$$

where D_b is the dimension of $\mathbf{x}_b$

$$-\frac{1}{2}(\mathbf{x}_a-\mu_a)^T\Lambda_{aa}(\mathbf{x}_a-\mu_a)-\frac{1}{2}(\mathbf{x}_a-\mu_a)^T\Lambda_{ab}(\mathbf{x}_b-\mu_b)-\frac{1}{2}(\mathbf{x}_b-\mu_b)^T\Lambda_{ba}(\mathbf{x}_a-\mu_a)$$
$$-\frac{1}{2}(\mathbf{x}_b-\mu_b)^T\Lambda_{bb}(\mathbf{x}_b-\mu_b)$$

Collect the remaining quadratic and linear terms of $\mathbf{x}_a$ from above and those from $\frac{1}{2}\mathbf{m}^T\Lambda_{bb}^{-1}\mathbf{m}=(\Lambda_{bb}\mu_b-\Lambda_{ba}(\mathbf{x}_a-\mu_a))^T\Lambda_{bb}^{-1}(\Lambda_{bb}\mu_b-\Lambda_{ba}(\mathbf{x}_a-\mu_a)$

$$-\frac{1}{2}(\mathbf{x}_a-\mu_a)^T\Lambda_{aa}(\mathbf{x}_a-\mu_a)+(\mathbf{x}_a-\mu_a)^T\Lambda_{ab}\mu_b$$
$$+\frac{1}{2}(\Lambda_{bb}\mu_b-\Lambda_{ba}(\mathbf{x}_a-\mu_a))^T\Lambda_{bb}^{-1}(\Lambda_{bb}\mu_b-\Lambda_{ba}(\mathbf{x}_a-\mu_a))$$

Quadratic Terms : $-\frac{1}{2}\mathbf{x}_a^T\Lambda_{aa}\mathbf{x}_a+\frac{1}{2}\mathbf{x}_a^T\Lambda_{ba}^T\Lambda_{bb}^{-1}\Lambda_{ba}\mathbf{x}_a$

$\mathbf{x}_a^T(.)\mathbf{x}_a : \mathbf{x}_a^T(\Lambda_{aa}-\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})\mathbf{x}_a$

Linear Terms : $+\mathbf{x}_a^T(\Lambda_{aa}\mu_a+\Lambda_{ab}\mu_b)+\frac{2}{2}\left[-\mathbf{x}_a^T\Lambda_{ba}^T\Lambda_{bb}^{-1}\left[\Lambda_{bb}\mu_b+\Lambda_{ba}\mu_a\right]\right]$

$$=\mathbf{x}_a^T\left[\Lambda_{aa}\mu_a+\Lambda_{ab}\mu_b-\Lambda_{ba}^T\Lambda_{bb}^{-1}\Lambda_{bb}\mu_b-\Lambda_{ba}^T\Lambda_{bb}^{-1}\Lambda_{ba}\mu_a\right]$$
$$=\mathbf{x}_a^T\left[\Lambda_{aa}-\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\right]\mu_a$$

$$\mathbf{x}_a^T(.) : \mathbf{x}_a^T\left(\Lambda_{aa}-\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\right)\mu_a\longrightarrow\left(\Lambda_{aa}-\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\right)\stackrel{?}{=}\Lambda_a$$
$(.)=\Sigma_a^{-1}\mu_a=[\Lambda_{aa}-\Lambda_{ba}\Lambda_{bb}^{-1}\Lambda_{ab}]\mu_a$ or $\Sigma_a=[\Lambda_{aa}-\Lambda_{ba}\Lambda_{bb}^{-1}\Lambda_{ab}]^{-1}$

Self Study Question:

Show that $\left(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\right) = \Lambda_a$.

Hint:

Remember that the integration yields $(2\pi)^{D_b/2}|\Lambda_{bb}|^{-1/2}$ and the normalization constant for $p(\mathbf{x}_a, \mathbf{x}_b) = (2\pi)^{-\frac{D_a+D_b}{2}} \begin{vmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{vmatrix}^{\frac{1}{2}}$.

Also remember the matrix identity for the determinant of block matrices *i.e.*

$$\begin{vmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{B} & \mathbf{D} \end{vmatrix} = \left| \begin{bmatrix} \mathbf{I} & \mathbf{C} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{A} - \mathbf{C}\mathbf{D}^{-1}\mathbf{B} & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{B} & \mathbf{I} \end{bmatrix} \right| = |\mathbf{D}| \cdot |\mathbf{A} - \mathbf{C}\mathbf{D}^{-1}\mathbf{B}| = |\mathbf{D}| \cdot |\mathbf{F}_{\mathbf{A}}|$$

.

Bayes' theorem for Gaussian variables

Given a Gaussian marginal distribution $p(x) = \mathcal{N}(\mathbf{x}|\mu, \Lambda^{-1})$ and a Gaussian conditional distribution $p(y|x) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1})$, the joint distribution of $\mathbf{x}$ and $\mathbf{y}$ is $p(\mathbf{z})$ where $\mathbf{z} = (\mathbf{x}\,\mathbf{y})^T$.

Determine $p(\mathbf{z}|\mu_{\mathbf{z}}, \mathbf{R}_{\mathbf{z}})$ where $\mathbf{z} = (\mathbf{x}\,\mathbf{y})^T$.

$\ln p(z) = \ln p(x) + \ln p(y|x) = -\frac{1}{2}(\mathbf{x} - \mu)^T \Lambda(\mathbf{x} - \mu) + -\frac{1}{2}(\mathbf{y} - \mathbf{A}\mathbf{x} - \mathbf{b})^T \mathbf{L}(\mathbf{y} - \mathbf{A}\mathbf{x} - \mathbf{b})$

Quadratic terms of $\mathbf{z}$: $-\frac{1}{2}\mathbf{z}^T \mathbf{R}_{\mathbf{z}} \mathbf{z}$

Quadratic terms of $\mathbf{x}$: $-\frac{1}{2}(\Lambda + \mathbf{A}^T \mathbf{L} \mathbf{A})$ Quadratic terms of $\mathbf{y}$: $-\frac{1}{2}\mathbf{L}$

Cross terms of $\mathbf{x}$ and $\mathbf{y}$: $\frac{1}{2}\mathbf{A}^T \mathbf{L}$ Cross terms of $\mathbf{y}$ and $\mathbf{x}$: $\frac{1}{2}\mathbf{L}\mathbf{A}$

$$-\frac{1}{2}\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}^T \begin{bmatrix} \mathbf{\Lambda} + \mathbf{A}^T \mathbf{L}\mathbf{A} & -\mathbf{A}^T \mathbf{L} \\ -\mathbf{L}\mathbf{A} & \mathbf{L} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = -\frac{1}{2}\mathbf{z}^T \mathbf{R}_{\mathbf{z}} \mathbf{z}$$

Using Schur's inversion formula, $\mathbf{R}_{\mathbf{z}}{}^{-1} = Cov\{\mathbf{z}\} = \begin{bmatrix} \mathbf{\Lambda}^{-1} & \Lambda^{-1}\mathbf{A}^T \\ \mathbf{A}\Lambda^{-1} & \mathbf{L}^{-1} + \mathbf{A}\Lambda^{-1}\mathbf{A}^T \end{bmatrix}$

Linear terms of $\mathbf{x}$: $\mathbf{x}^T(\Lambda\mu - \mathbf{A}^T \mathbf{L}\mathbf{b})$ Linear terms of $\mathbf{y}$: $\mathbf{L}\mathbf{b}$

Linear terms of $\mathbf{z}$: $\mathbf{z}^T \mathbf{R}_{\mathbf{z}} \mu_{\mathbf{z}} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}^T \begin{bmatrix} (\Lambda\mu - \mathbf{A}^T \mathbf{L}\mathbf{b}) \\ \mathbf{L}\mathbf{b} \end{bmatrix} \rightarrow$

$$\mu_{\mathbf{z}} = \mathbf{R}_{\mathbf{z}}{}^{-1} \begin{bmatrix} (\Lambda\mu - \mathbf{A}^T \mathbf{L}\mathbf{b}) \\ \mathbf{L}\mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}^{-1} & \Lambda^{-1}\mathbf{A}^T \\ \mathbf{A}\Lambda^{-1} & \mathbf{L}^{-1} + \mathbf{A}\Lambda^{-1}\mathbf{A}^T \end{bmatrix} \begin{bmatrix} (\Lambda\mu - \mathbf{A}^T \mathbf{L}\mathbf{b}) \\ \mathbf{L}\mathbf{b} \end{bmatrix}$$

$$\mu_{\mathbf{z}} = \begin{bmatrix} \mu \\ \mathbf{A}\mu + \mathbf{b} \end{bmatrix} = \begin{bmatrix} E\{\mathbf{x}\} \\ E\{\mathbf{y}\} \end{bmatrix}$$

Determine the conditional expectation and covariance of $p(\mathbf{x}|\mathbf{y})$

Remembering from the conditional density moments that

$$\Sigma_{a|b}^{-1}\mu_{a|b} = \Lambda_{aa}\mu_a - \Lambda_{ab}(\mathbf{x}_b - \mu_b) \text{ and } \Sigma_{a|b} = \Lambda_{aa}^{-1}$$

$$Cov\{\mathbf{x}|\mathbf{y}\} = \Sigma_{\mathbf{x}|\mathbf{y}} = (\Lambda + \mathbf{A}^T\mathbf{L}\mathbf{A})^{-1}$$

$$(\Lambda + \mathbf{A}^T\mathbf{L}\mathbf{A})\mu_{\mathbf{x}|\mathbf{y}} = (\Lambda + \mathbf{A}^T\mathbf{L}\mathbf{A})\mu + \mathbf{A}^T\mathbf{L}(\mathbf{y} - \mathbf{A}\mu - \mathbf{b}) = \mathbf{A}^T\mathbf{L}(\mathbf{y} - \mathbf{b}) + \Lambda\mu$$

$$\mu_{x|y} = E\{\mathbf{x}|\mathbf{y}\} = (\Lambda + \mathbf{A}^T\mathbf{L}\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{L}(\mathbf{y} - \mathbf{b}) + \Lambda\mu)$$