PROBABILITY & MULTIVARIATE RANDOM PROCESSES

Lecture Notes

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

Discrete Random Variables

A *random variable* is a number whose value depends upon the outcome of a random experiment.

Mathematically, a random variable X is a real-valued function on the space of outcomes which maps a probabilistic event to a real number:

$$X : \Omega \longrightarrow \mathcal{R} \tag{1}$$

A discrete random variable X has finitely or countably many values x_i, $i = 1, 2, \ldots$ and $p(x_i) = P(X = x_i)$ with $i = 1, 2, \ldots$ is called the probability mass function (pmf) of X.

Flipping a coin experiment

A real valued variable X is defined as

- If the flip is heads then $\omega_1 = \{H\} \longrightarrow X = 1$
- If the flip is tails then $\omega_2 = \{T\} \longrightarrow X = -1$

An urn contains 20 balls numbered $1, \ldots, 20$. Select 5 balls at random, without replacement. Let X be the largest number among selected balls. Determine its *pmf* and the probability that at least one of the selected numbers is 15 or more.

$$p(X=i) = \binom{\text{i-1}}{4} \Big/ \binom{20}{5}$$

$$P(\text{at least one of the selected numbers is} \geq 15\ )$$
$$= \sum_{i=15}^{20} P(X=i) = 1 - \binom{14}{5} \Big/ \binom{20}{5}$$

An urn contains 11 balls, 3 white, 3 red, and 5 blue balls. Take out 3 balls at random, without replacement. You win \$1 for each red ball you select and lose a \$1 for each white ball you select. Determine the *pmf* of X, the amount you win.

$$P(X=-3) = P(X=3) = \binom{3}{3} \Big/ \binom{11}{3}$$

$$P(X=-2) = P(X=2) = \left[\binom{3}{2} \cdot \binom{5}{1}\right] \Big/ \binom{11}{3}$$

$$P(X=-1) = P(X=1) = \left[\binom{3}{1} \cdot \binom{3}{2} + \binom{3}{1} \cdot \binom{5}{2}\right] \Big/ \binom{11}{3}$$

$$P(X=0) = \left[\binom{5}{1} \cdot \binom{3}{1} \cdot \binom{3}{1} + \binom{5}{3}\right] \Big/ \binom{11}{3}$$

Assume that X is a discrete random variable with possible values $x_i, i = 1, 2 \ldots$.

Expected value, also called expectation, average, or mean, of X is

$$E\{X\} = \sum_i x_i P(X = x_i) = \sum_i x_i p(x_i)$$

For any function, $g : \mathcal{R} \longrightarrow \mathcal{R}$,

$$E\{g(X)\} = \sum_i g(x_i) P(X = x_i)$$

Variance, σ^2 of X is

$$\sigma^2 = E\left\{[X - E\{X\}]^2\right\} = E\{X^2\} - E\{X\}^2$$

Rolling a dice N times

$E\{X\} = \frac{1 \cdot n_1 + 2 \cdot n_2 + 3 \cdot n_3 + 4 n_4 + 5 \cdot n_5 + 6 \cdot n_6}{N} = 1 \cdot \frac{n_1}{N} + 2 \cdot \frac{n_2}{N} + \cdots + 6 \cdot \frac{n_6}{N}$

For very large N, $\frac{n_1}{N} = \frac{n_2}{N} \cdots = \frac{n_6}{N} \approx \frac{1}{6}$

$E\{X\} = \frac{1}{6} \cdot (1 + 2 + \cdots + 6) = \frac{3}{2}$

Continuous Random Variables

A random variable X is continuous if there exists a nonnegative function f so that, for every interval B,

$$P(X \in B) = \int_B f(x)dx$$

The function $f = f_X$ is called the density of X.

The function $F = F_X$ given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)ds$$

is called the distribution function of X.

On an open interval where f is continuous,

$$F'(x) = f(x).$$

Expected Value

By analogy with discrete random variables, we define,

$$E\{X\} = \int_{-\infty}^{\infty} xf(x)dx,$$

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x)f(x)dx$$

and variance is computed by the same formula:

$$Var(X) = E\{X^2\} - (E\{X\})^2.$$

$$f(x) = \begin{cases} cx & 0 < x < 4, \\ 0 & otherwise. \end{cases}$$

(a) Determine c. (b) Compute $P(1 \leq X \leq 2)$. (c) Determine $E\{X\}$ and $Var\{X\}$.

(a) $\int_{-\infty}^{\infty} f(x)dx = 1 = \int_{0}^{4} cxdx = c\frac{x^2}{2}\Big|_0^4 = c\frac{16}{2} \longrightarrow c = \frac{1}{8}$ (b)

$P(1 \leq X \leq 2) = \int_{1}^{2} \frac{1}{8}xdx = \frac{x^2}{16}\Big|_1^2 = \frac{4-1}{16} = \frac{3}{16}$

(c) $E\{X\} = \int_{0}^{4} \frac{1}{8}x^2dx = \frac{x^3}{24}\Big|_0^4 = \frac{64}{24} = \frac{8}{3}$, $E\{X^2\} = \int_{0}^{4} \frac{1}{8}x^3dx = \frac{x^4}{32}\Big|_0^4 = \frac{256}{32} = 8$, $Var\{X\} = 8 - \frac{64}{9} = \frac{8}{9}$

Assume that X has density $f(x) = \begin{cases} 3x^2 & \text{if } x \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$

Compute the density f_Y of $Y = 1 - X^4$.

$x \in [0,1]$ and $y \in [0,1]$.

$F_Y(y) = P(Y \leq y) = P(1 - X^4 \leq y) = P(1 - y \leq X^4) = P((1-y)^{\frac{1}{4}} \leq X)$

$= 1 - P(X \leq (1-y)^{\frac{1}{4}}) = 1 - \int_0^{(1-y)^{\frac{1}{4}}} 3x^2 dx = 1 - x^3 \Big|_0^{(1-y)^{\frac{1}{4}}} = 1 - (1-y)^{\frac{3}{4}}$

$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{3}{4}(1-y)^{-\frac{1}{4}}$ for $y \in [0,1]$ and $f_y(Y) = 0$, otherwise.

Assume that X is uniform on $[0,1]$. What is the probability that the binary expansion of X starts with 0.010?

Smallest $X = 0.010 = \frac{1}{4}$ and largest $X = 0.011 = \frac{3}{8}$,

$P(\frac{1}{4} \leq X < \frac{3}{8}) = \frac{3}{8} - \frac{1}{4} = \frac{1}{8}$

Let X be a continuous random variable with probability density function on $0 \leq x \leq 1$, $f(x) = 3x^2$ What is the pdf of $Y = X^2$.

$P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) = \int_0^{\sqrt{y}} 3t^2 dt = y^{\frac{3}{2}} = F_Y(y)$, $0 \leq y \leq 1$

$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{3}{2} y^{\frac{1}{2}}$ for $0 \leq y \leq 1$.

A uniform random number X divides $[0,1]$ into two segments. Let R be the ratio of the smaller versus the larger segment. Compute the density of R.

$$R = \begin{cases} \frac{x}{1-x} & \textit{if } x \in [0, \frac{1}{2}], \\ \frac{1-x}{x} & \textit{if } x \in [\frac{1}{2}, 1], \end{cases}$$

$$F_R(r) = P(R \le r) = P(X \le \tfrac{1}{2}, \tfrac{X}{1-X} \le r) + P(X > \tfrac{1}{2}, \tfrac{1-X}{X} \le r)$$

$$= P(X \le \tfrac{1}{2}, X \le \tfrac{r}{r+1}) + P(X > \tfrac{1}{2}, X \ge \tfrac{1}{r+1}) = P(X \le \tfrac{r}{r+1}) + P(X \ge \tfrac{1}{r+1})$$

$$F_R(r) = \frac{r}{r+1} + 1 - \frac{1}{r+1} = \frac{2r}{r+1}$$

$$f(r) = \frac{d}{dr} F_R(r) = \frac{2}{(r+1)^2}$$

Assume that a lightbulb lasts on average 100 hours. Assuming *exponential* distribution, compute the probability that it lasts more than 200 hours and the probability that it lasts less than 50 hours.

Exponential density $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0, \\ 0 & x < 0 \end{cases}$

$$\lambda = \frac{1}{100}$$

$$P(X > 200) = 1 - P(X \le 200) = 1 - \int_0^{200} \frac{1}{100} e^{-\frac{x}{100}}\, dx = 1 + e^{-\frac{x}{100}} \Big|_0^{200} = 1 + e^{-2} - 1$$

$$P(X \le 50) = \int_0^{50} \frac{1}{100} e^{-\frac{x}{100}}\, dx = -e^{-\frac{x}{100}} \Big|_0^{50} = 1 - e^{\frac{1}{2}}$$

Theorem

If X is a continuous random variable, then the pdf of $y = g(X)$ is

$$f_Y(y) = \sum_{i=1}^{k} \frac{f_x(x^i)}{|g'(x^i)|}$$

where $x^1, \ldots, x^k$ are the roots of the equation $y = g(x)$.

Suppose X has a Gaussian distribution with a mean of 0 and variance of 1 and $Y = X^2 + 4$. Find the pdf of Y.

$y = g(x) = x^2 + 4$, $g^{-1}(y) = \pm\sqrt{y-4}$, $g'(x) = 2x$

$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-4)} \cdot \frac{1}{2\sqrt{y-4}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-4)} \cdot \frac{1}{2\sqrt{y-4}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-4)} \cdot \frac{1}{\sqrt{y-4}}$

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-4)} \cdot \frac{1}{\sqrt{y-4}} & 4 \leq y \\ 0 & otherwise \end{cases}$$

Let X be a normal distributed, $\mathcal{N}(\mu, \sigma^2)$ random variable and let $Y = \alpha X + \beta$, with $\alpha > 0$. How is Y distributed?

$$f_X(x) = \mathcal{N}(\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$F_Y(y) = P(Y \leq y) = P(\alpha X + \beta \leq y) = P(X \leq \frac{y-\beta}{\alpha}) = \int_{-\infty}^{\frac{y-\beta}{\alpha}} f_X(x)dx$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \int_{-\infty}^{\frac{y-\beta}{\alpha}} f_X(x)dx = f_Y(\frac{y-\beta}{\alpha})\frac{1}{\alpha}$$

$$f_Y(y) = \frac{1}{(2\pi\sigma^2\alpha^2)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2\alpha^2}(y-\alpha\mu-\beta)^2} = f_Y(y) = \mathcal{N}(\alpha\mu + \beta, \sigma^2\alpha^2)$$

Standard Normal Distribution

If $X \sim \mathcal{N}(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma}$ is *standard normal*

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

and its distribution function, also called as *error function* is

$$\Phi(z) = F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}x^2} dx$$

Jointly Distributed Random Variables

Flipping two coins experiment :

A real valued variable $x(i)$ with $i = 1, 2$ is defined as

- If the flip is heads then $\omega_1 = \{H\} \longrightarrow x(i) = 1$
- If the flip is tails then $\omega_2 = \{T\} \longrightarrow x(i) = -1$

Joint distribution function

$$F_{x(1),x(2)}(\alpha_1, \alpha_2) = Pr\{x(1) \leq \alpha_1, x(2) \leq \alpha_2\}$$

Joint density function

$$f_{x(1),x(2)} = \frac{\partial^2}{\partial\alpha_1\partial\alpha_2} F_{x(1),x(2)}(\alpha_1, \alpha_2)$$

Joint Moments

Correlation

$$r_{xy} = E\{xy^*\}$$

Covariance

$$r_{xy} = Cov(x, y) = E\Big\{[x - m_x][y - m_y]^*\Big\} = E\Big\{xy^*\Big\} - m_x m_y^*$$

Correlation Coefficient

$$\rho_{xy} = \frac{E\Big\{xy^*\Big\} - m_x m_y^*}{\sigma_x \sigma_y}$$

$$|\rho_{xy}| \leq 1$$

Joint Distributions and Independence

Discrete Case

Assume that you have a pair (X, Y) of discrete random variables X and Y. Their joint probability mass function is given by $p(x, y) = P(X = x, Y = y)$ so that

$$P((X, Y) \in A) = \sum_{(x,y)\in A} p(x, y).$$

The marginal probability mass functions are the *pmf*'s of X and Y, given by

$$P(X = x) = \sum_{y} P(X = x, Y = y) = \sum_{y} p(x, y)$$

$$P(Y = y) = \sum_{x} P(X = x, Y = y) = \sum_{x} p(x, y)$$

An urn has 2 red, 5 white, and 3 green balls. Select 3 balls at random and let X be the number of red balls and Y the number of white balls. Determine (a) joint *pmf* of (X, Y), (b) marginal *pmf's*, (c) $P(X \geq Y)$, and (d) $P(X = 2|X \geq Y)$.

$x = \{0, 1, 2\}$ and $Y = \{0, 1, 2, 3\}$

$$P(X = x, Y = y) = \frac{\binom{2}{x} \cdot \binom{5}{y} \cdot \binom{3}{3-(x+y)}}{\binom{10}{3}}$$

$y \backslash x$	0	1	2	$P(Y = y)$
0	1/120	$2 \cdot 3/120$	3/120	10/120
1	$5 \cdot 3/120$	$2 \cdot 5 \cdot 3/120$	5/120	50/120
2	$10 \cdot 3/120$	$10 \cdot 2/120$	0	50/120
3	10/120	0	0	10/120
$P(X = x)$	56/120	56/120	8/120	1

$P(X \geq Y) = \frac{(1+2\cdot5\cdot3+2\cdot3+5+3)}{120} = \frac{3}{8}$

$P(X = 2|X \geq Y) = \frac{P(X=2, X \geq Y)}{P(X \geq Y)} = \frac{\frac{3+5}{120}}{\frac{3}{8}} = \frac{8}{45}$.

Independence

Two random variables X and Y are independent if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

for all intervals A and B. In the discrete case, X and Y are independent exactly when

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

for all possible values x and y of X and Y, that is, the joint *pmf* is the product of the marginal *pmf*'s.

In the previous example, are X and Y independent?

$P(X = 1, Y = 3) = 0$ and $P(X = 1)P(Y = 3) = \frac{56}{120} \cdot \frac{10}{120}$

$$P(X = 1, Y = 3) \neq P(X = 1)P(Y = 3).$$

Therefore, X and Y are *not* independent.

Continuous Case

We say that (X, Y) is a jointly continuous pair of random variables if there exists a joint density $f(x, y) \geq 0$ so that

$$P((X, Y) \in S) = \iint_S f(x, y) dxdy,$$

where S is some nice (say, open or closed) subset of $\mathcal{R}^2$.

$$f(x, y) = \begin{cases} cx^2y & \text{if } x^2 \leq y \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

Determine *(a)* the constant c, *(b)* $P(X \geq Y)$, *(c)* $P(X = Y)$, and *(d)* $P(X = 2Y)$.

(a) $\int_{-1}^{1} \int_{x^2}^{1} cx^2 ydxdy = 1$, $c = \frac{21}{4}$,

(b) $P(X \geq Y) = \int_{x=0}^{1} \int_{y=x^2}^{x} \frac{21}{4} x^2 ydxdy = \frac{3}{20} \leftrightarrow$ Area between $y = x$ and $y = x^2$.

(c) $= 0$.

(d) $= 0$.

Marginal Densities

If f is the joint density of (X, Y), then the two marginal densities, which are the densities of X and Y, are computed by integrating out the other variable:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

Two jointly continuous random variables X and Y are independent exactly when the joint density is the product of the marginal ones:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

for all x and y.

Previous example, continued. Compute marginal densities and determine whether X and Y are independent.

$$f_X(x) = \int_{x^2}^{1} \frac{21}{4} x^2 y dy = \frac{21}{8} x^2 (1 - x^4) \text{ for } x \in [-1, 1] \text{ and 0 elsewhere.}$$

$$f_X(x) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4} x^2 y dx = \frac{7}{2} y^{\frac{5}{2}} \text{ for } y \in [0, 1] \text{ and 0 elsewhere.}$$

$f_{X,Y}(x, y) \neq f_X(x) f_Y(y)$, so X and Y are not independent.

Mr. and Mrs. Smith agree to meet "between 5 and 6 *p.m.*" Assume that they both arrive there at a random time between 5 and 6 and that their arrivals are independent. (a) Find the density for the time one of them will have to wait for the other. (b) Mrs. Smith later tells you she had to wait; given this information, compute the probability that Mr. Smith arrived before 5 : 30.

Let X and Y be the time when when Mr. and Mrs. Smith arrives, respectively;

(a) Let $T = |X - Y|$, which has possible values in $[0, 1]$ hour.
$P(T \leq t) = P(|X - Y| \leq t) = P(-t \leq X - Y \leq t)$
$= P(X - t \leq Y \leq X + t) = 1 - (1 - t)^2 = 2t - t^2$,

(b) $P(X \leq 0.5 | X > Y) = \frac{P(X \leq 0.5, X > Y)}{P(X > Y)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$.

Assume that you are waiting for two phone calls, from Alice and from Bob. The waiting time T_1 for Alices call has expectation 10 minutes and the waiting time T_2 for Bob's call has expectation 40 minutes. Assume T_1 and T_2 are independent exponential random variables. What is the probability that Alice's call will come first?

Assuming a time unit of 10 minutes

$f_{T_1} = e^{-t_1}$, $f_{T_1} = \frac{1}{4}e^{-t_2/4}$

$f_{T1,T2} = f_{T_1}(t_1) \cdot f_{T_2}(t_2) = \frac{1}{4}e^{-t_1-t_2/4}$

$$P(Alice's\ call\ first) = P(T_2 > T_1) = \int_0^\infty \int_{t_1}^\infty \frac{1}{4}e^{-t_1-t_2/4}dt_1 dt_2 = \frac{4}{5}$$

Conditional distributions

The conditional *pmf* of X given $Y = y$ is, in the discrete case, given simply by

$$p_X(x|Y = y) = P(X = x|Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)}.$$

For a jointly continuous pair of random variables X and Y , the conditional density of X given $Y = y$ is

$$f_X(x|Y = y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

where $f_{X,Y}(x, y)$ is the joint density of (X, Y).

Suppose (X, Y) has joint density $f(x,y) = \begin{cases} \frac{21}{4}x^2y & x^2 \le y \le 1, \\ 0 & otherwise \end{cases}$

Compute $f_X(x|Y=y)$.

$$f_X(x|Y=y) = \frac{f_{X,Y}(x,y)}{f_Y(y)},\ f_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4}x^2y dx = \frac{7}{2}y^{\frac{5}{2}}$$

$$f_X(x|Y=y) = \frac{\frac{21}{4}x^2y}{\frac{7}{2}y^{\frac{5}{2}}} = \frac{3}{2}x^2y^{-\frac{3}{2}}$$

Transformation of Random Variables

Theorem
Let $f(\mathbf{x})$ be the value of the probability density of the multivariate continuous random variable $\mathbf{X}$. If the vector-valued function $\mathbf{y} = U(\mathbf{x})$ is differentiable and invertible for all values within the domain of $\mathbf{x}$, then for corresponding values of $\mathbf{y}$, the probability density of $\mathbf{Y} = U(\mathbf{Y})$ is given by

$$f(\mathbf{y}) = f_x(U^{-1}(\mathbf{y})) \cdot \left|\frac{\partial}{\partial \mathbf{y}} U^{-1}(\mathbf{y})\right|$$

Consider a bivariate random variable $\mathbf{X}$ with states $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and probability density function $f(\mathbf{x}) = \frac{1}{2\pi} \exp(-\frac{1}{2}\mathbf{x}^T\mathbf{x})$. If $\mathbf{A}$ is a transformation matrix defined as $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, what is the probability density function of $\mathbf{y} = \mathbf{A}\mathbf{x}$?

$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\mathbf{y}$, $f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}) = \frac{1}{2\pi}\exp(-\frac{1}{2}\mathbf{y}^T\mathbf{A}^{-T}\mathbf{A}\mathbf{y})$

$\frac{\partial}{\partial \mathbf{y}}\mathbf{A}^{-1}\mathbf{y} = \mathbf{A}^{-T}$, $\left|\frac{\partial}{\partial \mathbf{y}}\mathbf{A}^{-1}\mathbf{y}\right| = \left|\mathbf{A}^{-\mathbf{T}}\right| = \frac{1}{|\mathbf{A}|} = \frac{1}{ad-bc}$

$f(\mathbf{y}) = \frac{1}{2\pi}\exp(-\frac{1}{2}\mathbf{y}^T\mathbf{A}^{-T}\mathbf{A}\mathbf{y}) \cdot \frac{1}{ad-bc}$

Let two resistors, having independent resistances, X_1 and X_2, uniformly distributed between 9 and 11 ohms, be placed in parallel. Find the probability density function of resistance Y_1 of the parallel combination.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad y_1 = \frac{x_1 x_2}{x_1 + x_2},\ y_2 = x_2 \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{x_1 x_2}{x_1+x_2} \\ x_2 \end{bmatrix}$$

$$\mathbf{y} = U(\mathbf{x}),\ U^{-1}(\mathbf{y}) = \begin{bmatrix} \frac{y_1 y_2}{y_2 - y_1} \\ y_2 \end{bmatrix},\ \frac{\partial}{\partial \mathbf{y}} U^{-1}(\mathbf{y}) = \begin{bmatrix} \frac{y_2^2}{(y_2-y_1)^2} & \frac{-y_1^2}{(y_2-y_1)^2} \\ 0 & 1 \end{bmatrix}$$

$$\left| \frac{\partial}{\partial \mathbf{y}} U^{-1}(\mathbf{y}) \right| = \frac{y_2^2}{(y_2-y_1)^2}$$

$$9 \le x_2 \le 11 \longrightarrow 9 \le y_2 \le 11, \qquad 9 \le x_1 \le 11 \longrightarrow \frac{9 \cdot 9}{9+9} = \frac{9}{2} \le y_1 \le \frac{11 \cdot 11}{11+11} = \frac{11}{2}$$

$$f_Y(y) = \begin{cases} \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{y_2^2}{(y_2-y_1)^2} & \frac{9}{2} \le y_1 \le \frac{11}{2},\ 9 \le y_2 \le 11 \\ 0 & \textit{otherwise} \end{cases}$$

Determine the pdf of $Y = X_1 + X_2$ where X_1 and X_2 are independent random variables.

$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $\quad y_1 = x_1 + x_2,\ y_2 = x_2 \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{Ax}$

$\mathbf{y} = U(\mathbf{x}), \quad U^{-1}(\mathbf{y}) = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \left| \frac{\partial}{\partial \mathbf{y}} U^{-1}(\mathbf{y}) \right| = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \mathbf{A}^{-1}$

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{A}^{-1}\mathbf{y}) \left| \frac{\partial}{\partial \mathbf{y}} U^{-1}(\mathbf{y}) \right| = f_{\mathbf{X}}(y1 - y2, y2) \cdot 1 = f_{X_1}(y2 - y1) f_{X_2}(y2)$$

$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(y_1, y_2) = f_{X_1}(y1 - y2) f_{X_2}(y2)$

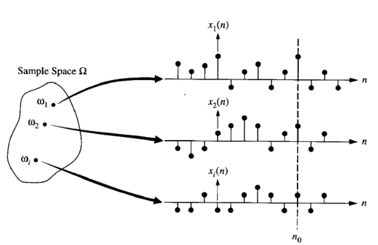
$$f_{Y_1}(y_1) = f_Y(y) = \int f_{\mathbf{X}}(y1 - y2, y2) dy_2 = \int f_{X_1}(y_2 - y_1) \cdot f_{X_2}(y_2) dy_2$$

Random Processes

$$x[n] = A\cos[\omega_0 n]$$

$$\ldots, x[-2], x[-1], x[0], x[1], x[2], \ldots$$

$$f_{x[n]}(\alpha) = Pr\{x[n] \leq \alpha\},\ f_{x[n]}(\alpha) = \frac{d}{d\alpha}F_{x[n]}(\alpha)$$



$$m_x[n] = E\left\{x[n]\right\}$$

$$\sigma_x^2(n) = E\left\{|x[n] - m_x(n)|^2\right\}$$

Crosscovariance between $x[n]$ and $y[n]$:

$$c_{xy(k,l)} = E\left\{x[k]y[l]^*\right\} - E\{x(k)\}E\{y[l]^*\}$$

Crosscorrelation between $x[n]$ and $y[n]$:

$$r_{xy(k,l)} = E\left\{x[k]y[l]^*\right\}$$

The Harmonic Process

$$x[n] = A\sin[n\omega_0 + \phi], \quad f_\Phi(\phi) = \begin{cases} \frac{1}{2\pi} & |\phi| \le \pi \\ 0 & otherwise \end{cases}$$

$m_x[n] = 0$ and $r_x(k,l) = \frac{|A|^2}{2}\cos((k-l)\omega)$

The autocorrelation matrix of a random vector $\mathbf{x}$ is

$$E\{\mathbf{x}\mathbf{x}^H\}.$$

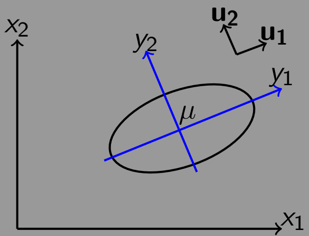
and its autocovariance matrix is

$$E\{(\mathbf{x} - E\{\mathbf{x}\})(\mathbf{x} - E\{\mathbf{x}\})^H\} = E\{\mathbf{x}\mathbf{x}^H\} - E\{\mathbf{x}\}E\{\mathbf{x}\}^H.$$

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.

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

$\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $|\mathbf{U}^T\mathbf{U}| = |\mathbf{U}^T||\mathbf{U}| = |\mathbf{U}^T||\mathbf{U}^T| = 1$, $|\mathbf{J}| = 1$

$dx_i = |\frac{\partial x_i}{\partial y_j}| dy_j$ or $d\mathbf{x} = |\mathbf{J}| d\mathbf{y} = d\mathbf{y}$

and remembering that for $\mathbf{y} = g(\mathbf{x})$ $f(\mathbf{y}) = f_x(g^{-1}(\mathbf{y})) \cdot \left| \frac{\partial}{\partial \mathbf{y}} g^{-1}(\mathbf{y}) \right|$

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

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

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

$$\left[\mathbf{u}_2\mathbf{u}_2^T \int e^{-\frac{1}{2}\frac{y_1^1}{\lambda_1}} dy_1 \int y_2^2 e^{\frac{y_2^2}{\lambda_2}} dy_2 \ldots \right] + \cdots +$$

$$\left[\mathbf{u}_D\mathbf{u}_D^T \int e^{-\frac{1}{2}\frac{y_1^1}{\lambda_1}} dy_1 \ldots \int y_D^2 e^{-\frac{1}{2}\frac{y_D^2}{\lambda_D}} dy_D \right] = \mathbf{u}_1\mathbf{u}_1^T\lambda_1 + \mathbf{u}_2\mathbf{u}_2^T\lambda_2 + \cdots = \sum_{i=1}^{D} \lambda_i \mathbf{u}_i \mathbf{u}_i^T = \Sigma$$

Maximum Likelihood Estimation

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\Big\{ -\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu) \Big\} d\mathbf{x}$$

log-likelihood of a multivariate Gaussian distribution
$\log p(\mathbf{x}|\mu, \Sigma) = -\frac{D}{2}\log(2\pi) - \frac{1}{2}|\Sigma| - \frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)$

$$\frac{\partial \log p(\mathbf{x})}{\partial \mathbf{x}} = -\frac{1}{2}\frac{\partial}{\partial \mathbf{x}}[-\mu^T \Sigma^{-1}\mathbf{x} + \mathbf{x}^T \Sigma^{-1}\mathbf{x} - \mathbf{x}^T \Sigma^{-1}\mu] = 0 \rightarrow \mathbf{x} = \mu$$

Maximum Likelihood Estimation

Given i.i.d. data $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)^T$ the log-likelihood function is given by

$$\ln p(\mathbf{X}|\mu, \Sigma) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n - \mu)^T \Sigma^{-1}(\mathbf{x_n} - \mu)$$

$$\frac{\partial}{\partial \mu}\ln p(\mathbf{X}|\mu, \Sigma) = \sum_{n=1}^{N}\Sigma^{-1}(\mathbf{x} - \mu) = 0 \rightarrow \mu_{ml} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$$

$$\frac{\partial}{\partial \Sigma}\ln p(\mathbf{X}|\mu, \Sigma) = 0 = -\frac{N}{2}\mathbf{\Sigma}^{-T} + \frac{1}{2}\sum_{n=1}^{N}\mathbf{\Sigma}^{-T}(\mathbf{x}_n - \mu)x_n - \mu)^T\mathbf{\Sigma}^{-T}$$

$$\Sigma = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \mu)(\mathbf{x}_n - \mu)^T$$

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*

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

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

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

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

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

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

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

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

$(.)=\Sigma_a^{-1}\mu_a=[\Lambda_{aa}-\Lambda_{ba}\Lambda_{bb}^{-1}\Lambda_{ab}]\mu_a$ or $\Sigma_a=[\Lambda_{aa}-\Lambda_{ba}\Lambda_{bb}^{-1}\Lambda_{ab}]^{-1}$

Bayes' theorem for Gaussian variables

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

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

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

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

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

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

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

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

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

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

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