10.5 Properties of Gaussian PDF

To help us develop some general MMSE theory for the Gaussian Data/Gaussian Prior case, we need to have some solid results for joint and conditional Gaussian PDFs.

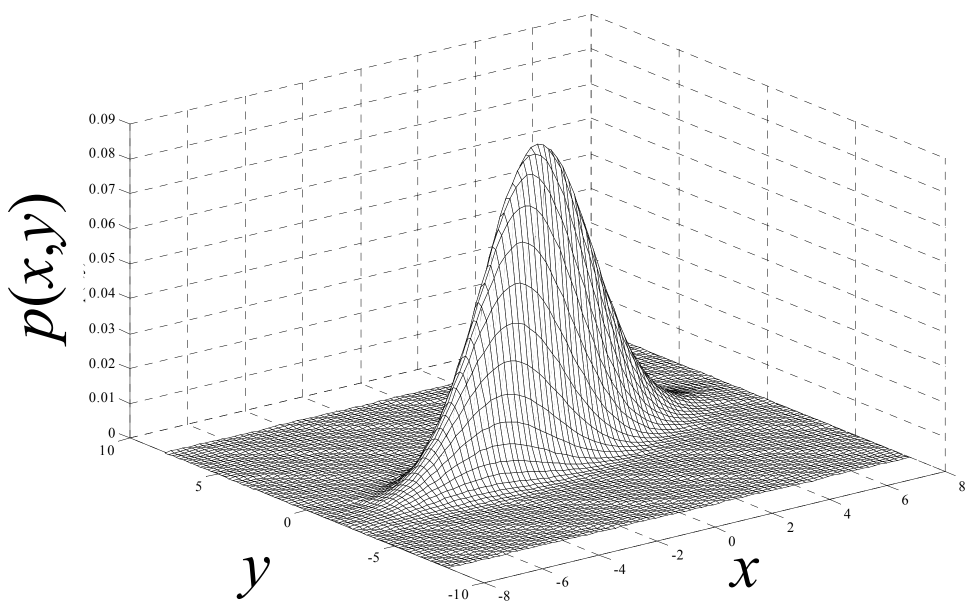
We'll consider the bivariate case but the ideas carry over to the general N-dimensional case.

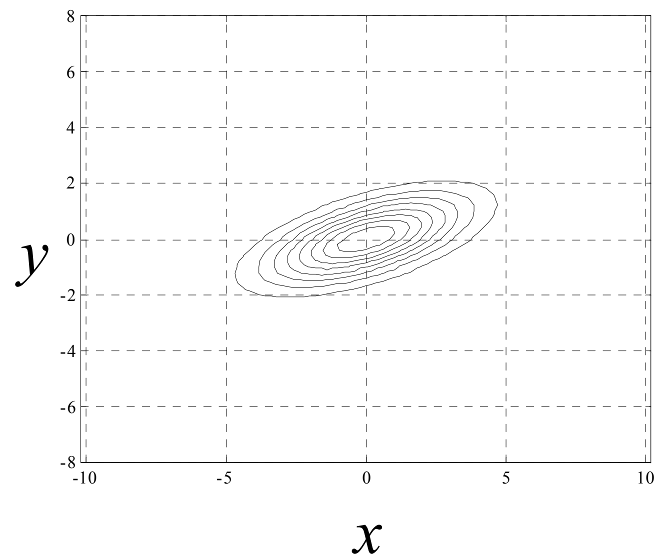
Bivariate Gaussian Joint PDF for 2 RV's X and Y

$$p(x,y) = \frac{1}{2\pi\,|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2} \underbrace{\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}^T \mathbf{C}^{-1} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}}_{\text{quadratic form}} \right)$$

$$E\left\{ \begin{bmatrix} X \\ Y \end{bmatrix} \right\} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} \mathrm{var}(X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(Y,X) & \mathrm{var}(Y) \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{YX} & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$





Marginal PDFs of Bivariate Gaussian

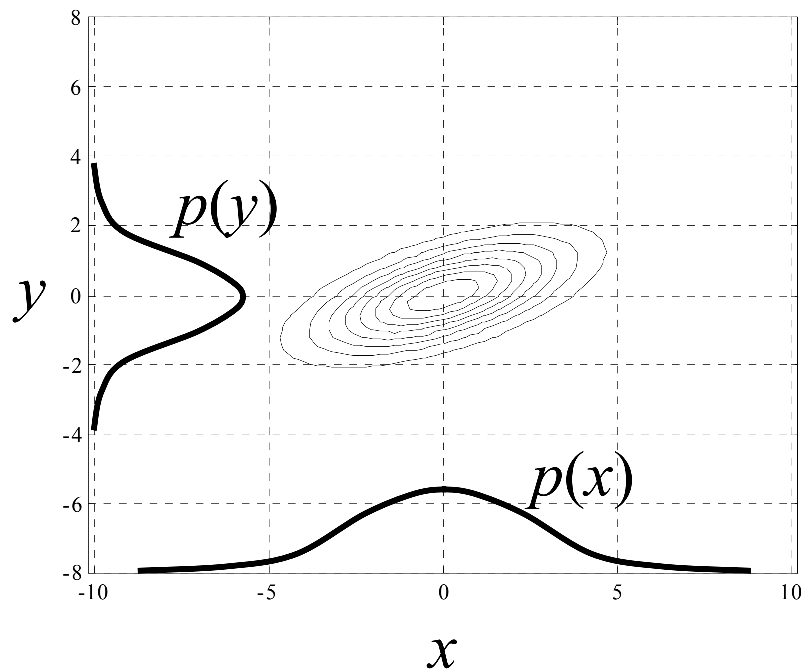
What are the marginal (or individual) PDFs?

We know that we can get them by integrating:

$$p(x) = \int_{-\infty}^{\infty} p(x,y)\, dy \qquad\qquad p(y) = \int_{-\infty}^{\infty} p(x,y)\, dx$$

After performing these integrals you get that:

$$X \sim N(\mu_X, \operatorname{var}\{X\}) \qquad\qquad Y \sim N(\mu_Y, \operatorname{var}\{Y\})$$



<u>Comment on "Jointly" Gaussian</u>

See Reading Notes on "Counter Example" posted on BB

We have used the term "Jointly" Gaussian…

Q: EXACTLY what does that mean?

A: That the RVs have a joint PDF that is Gaussian

$$p(x,y) = \frac{1}{2\pi |\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}^T \mathbf{C}^{-1} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix} \right)$$

Example for 2 RVs

We've shown that jointly Gaussian RVs also have Gaussian marginal PDFs

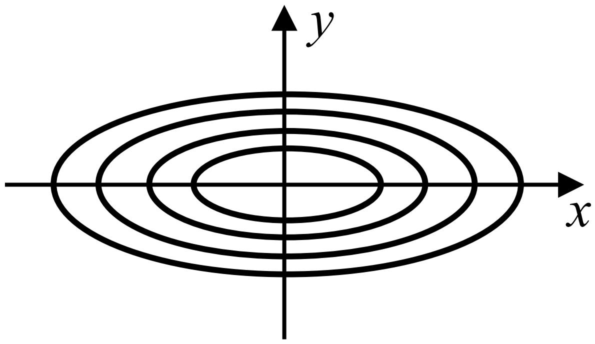
Q: Does having Gaussian Marginals imply Jointly Gaussian?

In other words… if X is Gaussian and Y is Gaussian is it always true that X and Y are <u>jointly</u> Gaussian???

A: No!!!!!

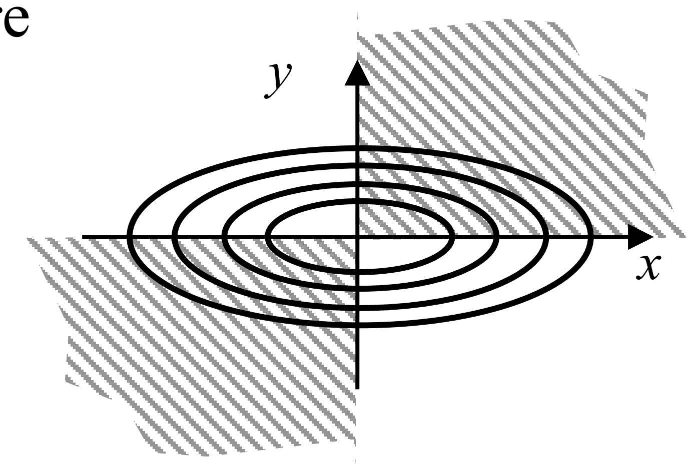
We'll construct a counterexample: start with a zero-mean, uncorrelated 2-D joint Gaussian PDF and modify it so it is no longer 2-D Gaussian but still has Gaussian marginals.

$$p_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y}\exp\left\{\frac{-1}{2}\left(\frac{x^2}{\sigma_X^2}+\frac{y^2}{\sigma_Y^2}\right)\right\}$$



But if we modify it by:

- Setting it to 0 in the shaded regions
- Doubling its value elsewhere



We get a 2-D PDF that is not a joint Gaussian but the marginals are the same as the original!!!!

Conditional PDFs of Bivariate Gaussian

What are the conditional PDFs?

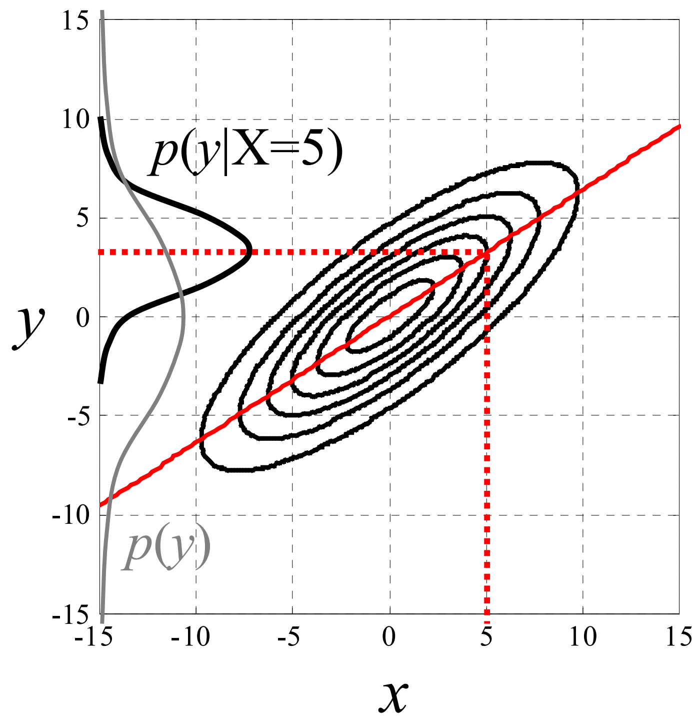
If you know that X has taken value $X = x_o$, how is Y distributed?

Slice @ x_o

$$\boxed{p(y \mid x_0) = \frac{p(x \mid x_0)}{p(x_0)} = \frac{p(x_0, y)}{\int_{-\infty}^{\infty} p(x_0, y)\, dy}}$$

Normalizer

$$C = \begin{bmatrix} 25 & 0.8\sqrt{25 \times 16} \\ 0.8\sqrt{25 \times 16} & 16 \end{bmatrix}$$



Slope of Line

$\mathrm{cov}\{X,Y\}/\mathrm{var}\{X\} = \rho\sigma_Y/\sigma_X$

Note: Conditioning on correlated RV

- **shifts mean**
- **reduces variance**

<u>Theorem 10.1</u>: Conditional PDF of Bivariate Gaussian

Let X and Y be random variables distributed <u>jointly</u> Gaussian with mean vector $[E\{X\} \;\; E\{Y\}]^T$ and covariance matrix

$$\mathbf{C} = \begin{bmatrix} \mathrm{var}(X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(Y,X) & \mathrm{var}(Y) \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{YX} & \sigma_Y^2 \end{bmatrix}$$

Then $p(y|x)$ is also Gaussian with mean and variance given by:

$$\boxed{\begin{aligned} E\{Y \mid X = x_o\} &= E\{Y\} + \frac{\sigma_{XY}}{\sigma_X^2}\left(x_o - E\{X\}\right) \\ &= E\{Y\} + \frac{\rho\sigma_Y}{\sigma_X}\left(x_o - E\{X\}\right) \end{aligned}}$$

Slope of Line

$$\boxed{\begin{aligned} \mathrm{var}\{Y \mid X = x_o\} &= \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} \\ &= \sigma_Y^2 - \rho^2\sigma_Y^2 = \left(1 - \rho^2\right)\sigma_Y^2 \end{aligned}}$$

<u>Amount</u> of Reduction

Reduction <u>Factor</u>

Impact on MMSE

We know the MMSE of RV Y after observing the RV $X = x_o$:

$$\boxed{\hat{Y} = E\left\{ Y \mid X = x_o \right\}}$$

So… using the ideas we have just seen:

if the data and the parameter are <u>jointly</u> Gaussian, then

$$\boxed{\hat{Y}_{MMSE} = E\{Y \mid X = x_o\} = E\{Y\} + \frac{\sigma_{XY}}{\sigma_X^2}\left(x_o - E\{X\}\right)}$$

It is the <u>correlation</u> between the RVs X and Y that allow us to perform Bayesian estimation.

Theorem 10.2: Conditional PDF of Multivariate Gaussian

Let $\mathbf{X}$ ($k\times1$) and $\mathbf{Y}$ ($l\times1$) be random vectors distributed jointly Gaussian with mean vector $[E\{\mathbf{X}\}^T \;\; E\{\mathbf{Y}\}^T]^T$ and covariance matrix

$$\mathbf{C} = \begin{bmatrix} \mathbf{C_{XX}} & \mathbf{C_{XY}} \\ \mathbf{C_{YX}} & \mathbf{C_{YY}} \end{bmatrix} = \begin{bmatrix} (k\times k) & (k\times l) \\ (l\times k) & (l\times l) \end{bmatrix}$$

Then $p(\mathbf{y}|\mathbf{x})$ is also Gaussian with mean vector and covariance matrix given by:

$$\boxed{E\{\mathbf{Y} \mid \mathbf{X} = \mathbf{x}_o\} = E\{\mathbf{Y}\} + \mathbf{C_{YX}}\mathbf{C_{XX}^{-1}}\left(\mathbf{x}_o - E\{\mathbf{X}\}\right)}$$

$$\boxed{\mathbf{C}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}_o} = \mathbf{C_{YY}} - \mathbf{C_{YX}}\mathbf{C_{XX}^{-1}}\mathbf{C_{XY}}}$$

$$E\{Y \mid X = x_o\} = E\{Y\} + \frac{\sigma_{XY}}{\sigma_X^2}\left(x_o - E\{X\}\right) \qquad \mathrm{var}\{Y \mid X = x_o\} = \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2}$$

Compare to Bivariate Results

For the Gaussian case… the cond. covariance does not depend on the conditioning x-value!!!

10.6 Bayesian Linear Model

Now we have all the machinery we need to find the MMSE for the "Bayesian Linear Model"

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

$N\times 1$

$N\times p$ known

$p\times 1$ $\sim N(\mu_\theta, \mathbf{C}_\theta)$

$N\times 1$ $\sim N(\mathbf{0}, \mathbf{C}_\mathbf{w})$

Clearly, **x** is Gaussian and θ is Gaussian…
But are they jointly Gaussian???

If yes… then we can use Theorem 10.2 to get the MMSE for θ!!!

Answer = Yes!!

Bayesian Linear Model is Jointly Gaussian

θ and $\mathbf{w}$ are each Gaussian and are independent

Thus their joint PDF is a product of Gaussians…

…which has the form of a jointly Gaussian PDF

Can now use: a linear transform of jointly Gaussian is jointly Gaussian

$$\begin{bmatrix} \mathbf{x} \\ \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \mathbf{H} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{w} \end{bmatrix}$$

Jointly Gaussian

Thus, Thm. 10.2 applies! Posterior PDF is…

- Joint Gaussian
- Completely described by its mean and variance

Conditional PDF for Bayesian Linear Model

To apply Theorem 10.2, notationally let $\mathbf{X} = \mathbf{x}$ and $\mathbf{Y} = \theta$.

First we need
$$E\{\mathbf{X}\} = \mathbf{H}\, E\{\theta\} + E\{\mathbf{w}\} = \mathbf{H}\mu_\theta$$

$$E\{\mathbf{Y}\} = E\{\theta\} = \mu_\theta$$

And also $\boxed{\mathbf{C}_{\mathbf{YY}} = \mathbf{C}_{\boldsymbol{\theta}}}$

$$\begin{aligned}
\mathbf{C}_{\mathbf{XX}} &= E\left\{(\mathbf{x} - E\{\mathbf{x}\})(\mathbf{x} - E\{\mathbf{x}\})^T\right\} \\
&= E\left\{[\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}}) + \mathbf{w}][\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}}) + \mathbf{w}]^T\right\} \\
&= \mathbf{H}\underbrace{E\left\{(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})^T\right\}}_{\mathbf{C}_{\boldsymbol{\theta}}}\mathbf{H}^T + E\left\{\mathbf{w}\mathbf{w}^T\right\}
\end{aligned}$$

Cross Terms are Zero because θ and $\mathbf{w}$ *are independent*

$$\boxed{\mathbf{C}_{\mathbf{XX}} = \mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T + E\left\{\mathbf{w}\mathbf{w}^T\right\}}$$

Similarly… $\mathbf{C}_{\mathbf{YX}} = \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}} = E\left\{(\boldsymbol{\theta}-\boldsymbol{\mu}_{\boldsymbol{\theta}})(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})^T\right\}$

Use $E\{\boldsymbol{\theta}\mathbf{w}\} = \mathbf{0}$
$E\{\boldsymbol{\mu}_{\boldsymbol{\theta}}\mathbf{w}\} = \mathbf{0}$

$$= E\left\{(\boldsymbol{\theta}-\boldsymbol{\mu}_{\boldsymbol{\theta}})(\mathbf{H}\boldsymbol{\theta}+\mathbf{w}-\mathbf{H}\boldsymbol{\mu}_{\boldsymbol{\theta}})^T\right\}$$

$$= E\left\{(\boldsymbol{\theta}-\boldsymbol{\mu}_{\boldsymbol{\theta}})(\boldsymbol{\theta}-\boldsymbol{\mu}_{\boldsymbol{\theta}})^T\mathbf{H}^T\right\}$$

$$\Rightarrow \boxed{\mathbf{C}_{\boldsymbol{\theta}\mathbf{x}} = \mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T}$$

Then Theorem 10.2 gives the conditional PDF's mean and cov
(and we know the conditional mean is the MMSE estimate)

Posterior Mean:

$$\boxed{\begin{aligned}\hat{\boldsymbol{\theta}}_{MMSE} &= E\{\boldsymbol{\theta}\,|\,\mathbf{x}\}\\ &= \boldsymbol{\mu}_{\boldsymbol{\theta}} + \underbrace{\overbrace{\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T}^{\text{Cross Correlation } \mathbf{C}_{\theta x}}\overbrace{\left(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T+\mathbf{C}_{\mathbf{w}}\right)^{-1}}^{\text{Relative Quality}}}_{\text{Update Transformation: Maps unpredictable part}}\underbrace{(\mathbf{x}-\mathbf{H}\boldsymbol{\mu}_{\boldsymbol{\theta}})}_{\text{Data Prediction Error}}\end{aligned}}$$

a priori estimate

Posterior Covariance:

$$\boxed{\mathbf{C}_{\boldsymbol{\theta}|\mathbf{x}} = \mathbf{C}_{\boldsymbol{\theta}} - \underbrace{\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T+\mathbf{C}_{\mathbf{w}}\right)^{-1}\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}}_{\text{Reduction Due to Data}}}$$

a priori covariance

Bayesian MMSE Estimator

Ex. 10.2: DC in AWGN w/ Gaussian Prior

Data Model: $x[n] = A + w[n]$ A & $w[n]$ are independent

$A \sim N(\mu_A, \sigma_A^2)$ $w[n] \sim N(0, \sigma^2)$

Write in linear model form:

$$\mathbf{x} = \mathbf{1}A + \mathbf{w} \quad \text{with } \mathbf{H} = \mathbf{1} = [\,1 \;\; 1 \;\; \dots \;\; 1]^T$$

Now General Result gives the MMSE estimate as:

$$\hat{A}_{MMSE} = E\{A \,|\, \mathbf{x}\} = \mu_A + \sigma_A^2 \mathbf{1}^T (\sigma_A^2 \mathbf{1}\mathbf{1}^T + \sigma^2 \mathbf{I})^{-1} (\mathbf{x} - \mathbf{1}\mu_A)$$

$$= \mu_A + \frac{\sigma_A^2}{\sigma^2} \mathbf{1}^T \underbrace{\left(\mathbf{I} + \frac{\sigma_A^2}{\sigma^2} \mathbf{1}\mathbf{1}^T\right)^{-1}}_{} (\mathbf{x} - \mathbf{1}\mu_A)$$

Can simplify using
"The Matrix Inversion Lemma"

Aside: Matrix Inversion Lemma

$$\left(\underbrace{\mathbf{A}}_{n\times n}+\underbrace{\mathbf{B}}_{n\times m}\underbrace{\mathbf{C}}_{m\times m}\underbrace{\mathbf{D}}_{m\times n}\right)^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}\mathbf{A}^{-1}\mathbf{B}+\mathbf{C}^{-1}\right)^{-1}\mathbf{D}\mathbf{A}^{-1}$$

Special Case ($m = 1$):

$$\left(\underbrace{\mathbf{A}}_{n\times n}+\underbrace{\mathbf{u}}_{n\times 1}\mathbf{u}^{T}\right)^{-1}=\mathbf{A}^{-1}-\frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{u}^{T}\mathbf{A}^{-1}}{1+\mathbf{u}^{T}\mathbf{A}^{-1}\mathbf{u}}$$

Continuing the Example… Apply the Matrix Inversion Lemma:

$$\hat{A}_{MMSE} = \mu_A + \frac{\sigma_A^2}{\sigma^2}\mathbf{1}^T\left(\mathbf{I} + \frac{\sigma_A^2}{\sigma^2}\mathbf{1}\mathbf{1}^T\right)^{-1}(\mathbf{x} - \mathbf{1}\mu_A)$$

Use Matrix Inv Lemma

$$= \mu_A + \frac{\sigma_A^2}{\sigma^2}\mathbf{1}^T\left(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^T}{N + \sigma^2/\sigma_A^2}\right)(\mathbf{x} - \mathbf{1}\mu_A)$$

Pass through $\mathbf{1}^T$ & use $\mathbf{1}^T\mathbf{1} = N$

$$= \mu_A + \frac{\sigma_A^2}{\sigma^2}\left(\mathbf{1}^T - \frac{N}{N + \sigma^2/\sigma_A^2}\mathbf{1}^T\right)(\mathbf{x} - \mathbf{1}\mu_A)$$

Factor Out $\mathbf{1}^T$ & use $\mathbf{1}^T\mathbf{1} = N$

$$= \mu_A + \frac{\sigma_A^2}{\sigma^2}\left(1 - \frac{N}{N + \sigma^2/\sigma_A^2}\right)(N\bar{x} - N\mu_A)$$

Algebraic Manipulation

$$\boxed{\hat{A}_{MMSE} = \mu_A + \underbrace{\left(\frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N}\right)}_{\text{“Gain” Factor}}\underbrace{(\bar{x} - \mu_A)}_{\text{Error Between Data-Only Est. \& Prior-Only Est.}}}$$

***a priori* estimate**

- When data is bad ($\sigma^2/N >> \sigma_A^2$), gain is small, data has little use

$$\boxed{\hat{A}_{MMSE} \approx \mu_A}$$

- When data is good ($\sigma^2/N >> \sigma_A^2$), gain is large, data has large use

$$\boxed{\hat{A}_{MMSE} \approx \bar{x}}$$

Using similar manipulations gives:

$$\mathrm{var}(A\,|\,\mathrm{x}) \quad = \quad \frac{\left(\dfrac{\sigma^2}{N}\right)\sigma_A^2}{\sigma_A^2 + \dfrac{\sigma^2}{N}} \quad = \quad \frac{1}{\dfrac{1}{\sigma_A^2} + \dfrac{1}{\sigma^2 / N}}$$

Like || resistors… small one wins!
⇒ var (A|**x**) is ≈ the smaller of:
- data estimate variance
- prior variance

Or… looking at it another way:

$$\frac{1}{\mathrm{var}(A\,|\,\mathrm{x})} \quad = \quad \frac{1}{\sigma_A^2} + \frac{1}{\sigma^2 / N}$$

… additive "information"!

10.7 Nuisance Parameters

One difficulty in classical methods is that nuisance parameters must explicitly dealt with.

In Bayesian methods they are simply "Integrated Away"!!!!

Recall Emitter Location: $[x \;\; y \;\; z \;\; f_0]$

Nuisance Parameter

In Bayesian Approach…

From $p(x, y, z, f_0 \,|\, \mathbf{x})$ can get $p(x, y, z \,|\, \mathbf{x})$:

$$p(x, y, z \,|\, \mathbf{x}) = \int p(x, y, z, f_0 \,|\, \mathrm{x}) \, df_0$$

Then… find conditional mean for the MMSE estimate!