Recall $f_{XY}(u,v) = f_X(u) f_{Y|X}(v|u)$ so

$$E[(Y-g(X))^2] = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} (v-g(u))^2 f_{Y|X}(v|u)\,dv \right) f_X(u)\,du.$$

for each u fixed, we have ~~com~~ ~~integral~~ $\cancel{\int_{-\infty}^{\infty} (y-\delta)^2 f_Y(y)\,dy}$ inside parentheses, an expression

$$\begin{aligned} E[(Y-\delta)^2] &= E[((Y-EY)+(EY-\delta))^2] \\ &= E[(Y-EY)^2 + 2(Y-EY)(EY-\delta) + (EY-\delta)^2] \\ &= var(Y) + (EY-\delta)^2. \end{aligned}$$

This is minimized by $\delta = EY$.

So for each u, integral in parentheses minimized by using $g(u) = g^*(u)$

where $g^*(u) = E[Y|X=u] = \int_{-\infty}^{\infty} v f_{Y|X}(v|u)\,dv.$

We write $E[Y|X]$ for $g^*(X)$, so minimum MSE is

$$E[(Y-E[Y|X])^2] = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} (v-g^*(u))^2 f_{Y|X}(v|u)\,dv \right) f_X(u)\,du$$

using $var(Y) = E[Y^2] - E[Y]^2$:

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} (v^2 - (g^*(u))^2) f_{Y|X}(v|u)\,dv \right) f_X(u)\,du$$

$$= E[Y^2] - E[(E[Y|X])^2].$$

to summarize for MMSE use conditional mean, $E[Y|X]$.

Suppose we don't allow arbitrary g^* since $E[Y|X]$ might be difficult to find and compute. ~~instead~~ (Integration could be hard)

Restrict to g that is linear, i.e. Linear estimator of Y given X:

$L(X) = aX + b$, we just need to find a and b rather than a whole function g^*.

MSE of $aX+b$ is

$$E[(Y-(aX+b))^2], \text{ which can be regrouped as}$$

$$E[((Y-aX)-b)^2]$$

So we see that for a given value of a, the constant b should be the minimum MSE constant estimator of $Y-aX$, i.e. the expected value

$$b = E[Y-aX] = \mu_Y - a\mu_X.$$

Therefore the LMMSE has the form

$$aX + \mu_Y - a\mu_X \quad \text{or equivalently} \quad \mu_Y + a(X-\mu_X).$$

Corresponding MSE is

$$E[(Y-\mu_Y - a(X-\mu_X))^2]$$

$$= var(Y-aX) = cov(Y-aX, Y-aX)$$

$$= var(Y) - 2a\, Cov(Y,X) + a^2 var(X).$$

To find best a, take derivative and set equal to zero.

$$-2 cov(Y,X) + 2a\, var(X) = 0$$

$$a^* = \frac{Cov(Y,X)}{var(X)}.$$

$$L^*(X) = \mu_Y + \left(\frac{\mathrm{cov}(Y,X)}{\mathrm{var}(X)}\right)(X-\mu_X)$$

$$= \mu_Y + \sigma_Y \rho_{XY}\left(\frac{X-\mu_X}{\sigma_X}\right)$$

The performance of this estimator is, by substituting a^* into $\mathrm{var}(Y) - 2a\,\mathrm{cov}(X,Y) + a^2\mathrm{var}(X)$:

$$\sigma_Y^2 - \frac{(\mathrm{cov}(X,Y))^2}{\mathrm{var}(X)} = \sigma_Y^2(1-\rho_{XY}^2).$$

Note that 4 out of 5 constants in LMMSE formula are means and standard deviations.

If X, Y are standardized (mean-zero, unit variance) then reduces to

$L^*(X) = \rho_{XY} X$ and it achieves MSE $= 1-\rho_{XY}^2$.

Correlation coefficient important in assessing ability to do linear estimation.

Ex Let $X = Y + N$ (signal + noise model) where $Y \sim \mathrm{Exp}(\lambda)$ and $N \sim \mathcal{N}(0, \sigma_N^2)$, $Y \perp\!\!\!\perp N$.

$E[Y] = \frac{1}{\lambda}$, $\mathrm{var}[Y] = \frac{1}{\lambda^2}$.

Find $L^*(X)$ and its performance.

Since $Y \perp\!\!\!\perp N$, $\mathrm{cov}(Y,N) = 0$, so

$$\mathrm{cov}(Y,X) = \mathrm{cov}(Y, Y+N) = \mathrm{cov}(Y,Y) + \mathrm{cov}(Y,N) = \mathrm{var}[Y] = \frac{1}{\lambda^2}$$

$$\mathrm{var}(X) = \mathrm{var}(Y+N) = \mathrm{var}(Y) + \mathrm{var}(N) = \frac{1}{\lambda^2} + \sigma_N^2$$

So

$$L^*(X) = \frac{1}{\lambda} + \frac{1/\lambda^2}{1/\lambda^2 + \sigma_N^2}\left(X - \frac{1}{\lambda}\right)$$

$$= \frac{1}{\lambda} + \frac{1}{1+\lambda^2\sigma_N^2}\left(X - \frac{1}{\lambda}\right) = \frac{X + \lambda\sigma_N^2}{1+\lambda^2\sigma_N^2}.$$

MSE is $\sigma_Y^2 - \dfrac{\sigma_Y^4}{\sigma_Y^2+\sigma_N^2} = \dfrac{\sigma_Y^2\sigma_N^2}{\sigma_Y^2+\sigma_N^2} = \dfrac{\sigma_N^2}{1+\lambda^2\sigma_N^2}.$