GAUSSIANITY OF OPTIMAL NOISE IN AN LQ MODEL

1. Simplified Problem

Suppose we wish to maximize the entropy $H(X)$ of the random variable X subject to the requirement that $EX^2 \leq \sigma^2$. Writing the problem out in terms of the density of X, we get

$$\max - \int \log\big(p(x)\big)\, p(x)\, dx \tag{1}$$

subject to

$$\int x^2 p(x)\, dx \leq \sigma^2 \tag{2}$$

$$\int p(x)\, dx = 1 \tag{3}$$

$$p(x) \geq 0\,,\ \text{all } x\,. \tag{4}$$

The objective function here is concave in p, because $-p\log(p)$ is concave in p for each x. The constraints are all linear in p, so their intersection defines a concave constraint set. We can therefore form a Lagrangian and be sure that a solution to the FOC's will solve the problem. If the problem involved derivatives of p we might have to be concerned with transversality, but here we can just optimize pointwise. Doing so and assuming at first that the $p \geq 0$ constraint never binds, leads to the FOC

$$-1 - \log p = \lambda_0 + \lambda_1 x^2\,. \tag{5}$$

Exponentiating, we see that the result is

$$p(x) = \gamma_0 \exp(-\lambda_0 - \lambda_1 x^2)\,. \tag{6}$$

A pdf of this form is Gaussian with mean zero. Pinning down its variance at σ^2 then fully characterizes the pdf.

This result is easy, but not by itself of much use. Entropy itself depends on the underlying base measure in an arbitrary way. It is *changes* in entropy, or mutual information between random variables, that has an interpretation in terms of costly information flow.

2. Static Control Problem

Now we consider a more interesting problem. Suppose we wish to minimize

$$E\big[(Y-X)^2\big] \tag{7}$$

subject to

$$-H(Y,X)+H(Y)+H(X)<\kappa \tag{8}$$
$$X \text{ has pdf } p(x)\,. \tag{9}$$

Here H is entropy and κ is a bound on the mutual information between Y and X.

Our object of choice is the joint pdf $q(x,y)$ of X and Y. Reformulating the problem in terms of q, it becomes

$$\min \int_{-\infty}^{\infty} q(x,y)(x-y)^2\,dx\,dy \tag{10}$$

subject to

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} &\log\big(q(x,y)\big)q(x,y)\,dx\,dy \\ &- \int_{-\infty}^{\infty}\log\big(p(x)\big)p(x)\,ds \\ &\quad - \int_{-\infty}^{\infty}\log\left(\int_{-\infty}^{\infty} q(x^*,y)\,dx^*\right)\left(\int_{-\infty}^{\infty} q(x,y)\,dx\right)dy \le \kappa\end{aligned} \tag{11}$$

$$\int_{-\infty}^{\infty} q(x,y)\,dy = p(x)\,,\ \text{all } x \tag{12}$$

$$q(x,y)\ge 0\,,\ \text{all } x,y. \tag{13}$$

Since there are no derivatives of q involved here, we can try to solve the problem point by point. That is, we just set up a Lagrangian and differentiate with respect to $q(x,y)$, separately for each x,y pair. This gives us the first order conditions

$$-(x-y)^2=\lambda_0\left(1+\log q(x,y)-1-\log\left(\left(\int q(x,y)\,dy\right)\right)+\lambda_1(x)\,,\right. \tag{14}$$

where again we begin by assuming the $q(x,y)\ge 0$ constraint does not bind. Note that if we define

$$h(x\mid y)=\frac{q(x,y)}{\int q(x,y)\,dx}\,, \tag{15}$$

the FOC (14) can be rewritten as

$$h(x\mid y)=\theta_0(x)e^{-\theta_1(x-y)^2}\,. \tag{16}$$

This of course has to be a pdf for every y, meaning that its integral with respect to x is one for every y. It is clear that we can meet this requirement by simply setting $\theta_0(x)$ to be a constant, choosing the constant so that it makes the right-hand side a Gaussian pdf. Note that for this to work, we need to use the fact that the range of x is the entire $(-\infty,\infty)$ interval, so our results will not extend to cases where $p(x)$ has bounded support.

A sketch of the argument, which you may follow easily only if you have studied Fourier analysis and linear operators, goes as follows. (16) makes

$$\int h(x \mid y dx = c\theta_0 * \phi(y) \equiv 1\ , \tag{17}$$

where ϕ is a Gaussian pdf with variance $\theta_1/2$ and c is a normalizing constant. But the only function that, convoluted with a Gaussian pdf, yields a constant is the constant function. To see that, note that in the frequency domain (17) becomes $\tilde{h} = c\tilde{\theta}_0\tilde{\phi}$, and the Fourier tranform of a constant is a Dirac delta function. Thus the only possible form for $\tilde{\theta}_0$ is itself a Dirac delta.

With this result in hand, we can now conclude that

$$q(x,y) = h(x \mid y)g(y) = \sqrt{\theta_1}\phi(\theta_1(x-y)^2)g(y)\ , \tag{18}$$

where $g(y)$ is the marginal pdf of y.

So our result is that it is optimal to choose the joint distribution in such a way that the conditional distribution of the given signal, x, given the observation, y, is Gaussian. The variance of that distribution will determine the amount of mutual information.

This result is not universal, however, because we have not checked whether it is possible to choose $g(y)$ so as to satisfy (12)'s specification of the marginal for x. It is always possible to do so if p is a Gaussian pdf, but there are forms for p that make it impossible.

An easy way to see why there are pdf's for which the result cannot hold is to note that our formula (18) for the joint pdf implies that

$$p(x) = \int \sqrt{\theta_0}\phi(\theta_1(x-y)^2)q(y)\,dy\ . \tag{19}$$

It is then apparent that

$$p'(x) = \int 2\theta_1(x-y)\sqrt{\theta_0}\phi(\theta_1(x-y)^2)q(y)dy\ . \tag{20}$$

The differentiation under the integral sign is justified because we know $|x-y|\,\phi(\theta_1(x-y)^2)$ is bounded and we know q is integrable. In fact, the same sort of argument shows that derivatives of p of every order exist. But many pdf's are not differentiable everywhere — for example the double-exponential or a uniform distribution on a finite interval.

An example of a pdf that is at the borderline of the range of p's for which our result holds is the pdf formed as the convolution of a $U(0,1)$ pdf with a normal, which is of course also the pdf of the sum of a uniform with an independent normal.