The Gaussian and entropy maximization

Peter Haggstrom
mathsatbondibeach@gmail.com
https://gotohaggstrom.com

February 6, 2022

1 Background

In the theory of entropy in the continuous case, one can ask what function y maximises the integral:

$$H(y) = -\int_{-\infty}^{\infty} y \ln y \; dx \text{ where } y \geq 0 \tag{1}$$

subject to these constraints:

$$\int_{-\infty}^{\infty} y \, dx = 1 \;, \; \int_{-\infty}^{\infty} x^2 \, y \, dx = \sigma^2 \tag{2}$$

where $y \to 0$ as $x \to \pm\infty$.

If we view y as a probability density then the first constraint in (2) tells as that the total probability sums to 1 while the second tells us that the variance is σ^2.

Applying the theory of the calculus of variations (see below for the details) one finds that $y = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}$ which is, of course, the normal distribution density function - the paradigm Gaussian.

It was Claude Shannon in his paper, "A Mathematical Theory of Communication" [1], who introduced the concept of entropy in the context of information. In so doing he proposed a definition of entropy in both the discrete and continuous cases which had to satisfy certain basic conditions. In the discrete case the entropy H is a function of the known probabilities p_i (where $1 \leq i \leq n$) and as Shannon put it: "Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome?" ([1] page 10). Shannon required H to satisfy the following conditions:

1. H should be continuous in the p_i.
2. If all the p_i are equal, $p_i = \frac{1}{n}$, then H should be a monotonic increasing function of n . With equally likely events there is more choice, or uncertainty, when there are more possible events.
3. If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H.

Shannon shows that the only H satisfying the above condition is $H = -k \sum_i^n p_i \ln p_i$ where k is a positive constant and in what follows we will assume $k = 1$. For a given n, H is a maximum and equal to $\ln n$ (note here that although the focus of Shannon's paper was on binary communication channels where one would use logarithms to base 2, his proof of his entropy formula is independent of any specific base) when all the p_i are equal (i.e. $\frac{1}{n}$). This is also intuitively the most uncertain situation.

The concept of entropy arises in statistical mechanics where, in the discrete case, the entropy H in Boltzmann's "H Theorem" is given by:

$$H = -\sum_i p_i \ln p_i \tag{3}$$

A detailed treatment of Boltzmann's H Theorem can be found, for instance, in Chapter VI of [2]. Although this book was first published in 1938 it contains relatively detailed derivations of important relations and reflects input from people such as Robert Oppenheimer, to whom the book was dedicated. In that construction p_i can be viewed as the probability of of a system being in cell i of its phase space.

The Gaussian arises from certain partial differential equations such as in Einstein's derivation of Brownian motion [4] which involved a heat equation. Maxwell's velocity distrbution law which also involves a Gaussian was derived from considerations of independence of velocity components and radial symmetry [5]. Of course, Laplace's proof of the central limit theorem also led to a Gaussian [9]. We know from Fourier theory that the Fourier transform of a Gaussian is a Gaussian and this is a telling property since if something is concentrated in the time domain, say, when we take the Fourier transform it will be spread out in the frequency domain, say. This is actually at the heart of the Heisenberg Uncertainty Principle in the sense that you cannot simultaneously localise a function and its Fourier transform. If we take the Fourier transform of something highly localised such as the Dirac delta "function" we get something that is smeared out in the transform space. Thus we go from localisation to non-localisation and vice versa. Here is an example of a Fourier transform pair of a rectangle function on the left (non-localised) and its Fourier transform on the right (localised).

Figure 1:

The fact that the Fourier transform of a Gaussian is a Gaussian suggests that because we do not get any additional non-localisation or localisation something fundamental is going on. If one thinks of a Fourier transform as mapping functions from one space to another, the fact that the Fourier transform of a Gaussian is also a Gaussian suggests that it is a fixed point in that space. In other words no additional information is gained by the Fourier transform acting on the Gaussian. This gives yet another perspective on the concept that the Gaussian maximises entropy or uncertainty. Note that in Figure 1 we go from a relatively uncertain context since for any x the value of $H(x)$ is constant to relative certainty in the sense that something relatively significant is going on at $x = 0$ where a lot of information seems to be localised.

Stein and Sharkarchi show ([6], pages 158-9) that the Heisenberg Uncertainty Principle can be expressed in the following form (leaving aside the physical constant). If ψ is a function in the Schwartz space $\mathcal{S}(\mathbb{R})$ (see [7] for more details on Schwartz space) which satisfies the normalising condition $\int_{-\infty}^{\infty} |\psi(x)|^2 \, dx = 1$, then:

$$\boxed{\left(\int_{-\infty}^{\infty} x^2 |\psi(x)|^2 \, dx \right) \left(\int_{-\infty}^{\infty} \xi^2 |\hat{\psi}(\xi)|^2 \, d\xi \right) \geq \frac{1}{16\pi^2}} \tag{4}$$

where $\hat{\psi}(\xi)$ is the Fourier transform of $\psi(x)$. Equality holds if and only if $\psi(x) = Ae^{-Bx^2}$ where $B > 0$ and $A^2 = \sqrt{\frac{2B}{\pi}}$. Note that in the equality case the form of $\psi(x)$ is a Gaussian.

In physics the Uncertainty Principle is expessed as:

The product of the uncertainty in position and the uncertainty in momentum is necessarily greater than a quantity of order $\hbar$.

The probability that a particle is located in (a, b) is $\int_a^b |\psi(x)|^2 \, dx$ and the probability that the momentum ξ of a particle belongs to the interval (a, b) is $\int_a^b |\hat{\psi}(\xi)|^2 \, d\xi$. Thus (4) expresses the idea that there is a lower bound for the product of the uncertainty of position and the uncertainty of momentum (as represented by $\int_{-\infty}^{\infty} x^2 |\psi(x)|^2 \, dx$ and $\int_{-\infty}^{\infty} \xi^2 |\hat{\psi}(\xi)|^2 \, d\xi$) . If we are more certain about position we are less certain about momentum and vice versa.

Interestingly, Heisenberg's original derivation involved Gaussian functions in the context of probability amplitudes and these had the form:

$$e^{\left[-\frac{(q-q^*)^2}{2q_1^2}-\frac{ip^*}{\hbar}(q-q^*)\right]} \tag{5}$$

He then concludes that:

$$p_1\, q_1 = \hbar \tag{6}$$

Thus in the Gaussian case you get the product equal to the irreducible uncertainty limit. The precise process by which Heisenberg arrived at this result can be found in the translated version of his paper [10]. Masanao Ozawa and many others have pointed out that Kennard then considered other types of functions and arrived at the inequality that is now known as Heisenberg's Uncertainty Principle - see [11].

To prove (4) we need integration by parts and some fundamental results from Fourier theory. We start with the normalisation condition which essentially says that the probability amplitudes of the wave function sum to 1. In what follows because it is assumed $\psi(x)$ is in Schwartz space it decays rapidly so that $x|\psi(x)|^2 \to 0$ as $x \to \pm\infty$.

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty} |\psi(x)|^2\, dx \\
&= \underbrace{\Big[x|\psi(x)|^2\Big]_{-\infty}^{\infty}}_{=0} - \int_{-\infty}^{\infty} x\frac{d}{dx}\Big(|\psi(x)|^2\Big)\, dx \\
&= - \int_{-\infty}^{\infty} x\frac{d}{dx}\Big(|\psi(x)|^2\Big)\, dx \\
&= - \int_{-\infty}^{\infty} x\frac{d}{dx}\Big(\psi(x)\,\overline{\psi(x)}\Big)\, dx \\
&= - \int_{-\infty}^{\infty} \Big(x\overline{\psi(x)}\,\psi'(x) + x\overline{\psi'(x)}\,\psi(x)\Big)\, dx \\
\therefore |1| &= \Big| - \int_{-\infty}^{\infty} \Big(x\overline{\psi(x)}\,\psi'(x) + x\overline{\psi'(x)}\,\psi(x)\Big)\, dx\, \Big| \\
1 &\leq \int_{-\infty}^{\infty} \Big|x\,\overline{\psi(x)}\,\psi'(x)\Big|\, dx + \int_{-\infty}^{\infty} \Big|x\,\overline{\psi'(x)}\,\psi(x)\Big|\, dx \\
&= \int_{-\infty}^{\infty} \Big|x\Big|\,\Big|\overline{\psi(x)}\Big|\,\Big|\psi'(x)\Big|\, dx + \int_{-\infty}^{\infty} \Big|x\Big|\,\Big|\overline{\psi'(x)}\Big|\,\Big|\psi(x)\Big|\, dx \\
&=2\int_{-\infty}^{\infty} \Big|x\Big|\,\Big|\psi(x)\Big|\,\Big|\psi'(x)\Big|\, dx \\
&\leq 2\left(\int_{-\infty}^{\infty} x^2|\psi(x)|^2\, dx\right)^{\frac{1}{2}}\left(\int_{-\infty}^{\infty} |\psi'(x)|^2\, dx\right)^{\frac{1}{2}}
\end{aligned} \tag{7}$$

In the last line of (7) we used the Cauchy-Schwarz inequality:

$$\int f(x)g(x)\,dx \leq \Big(\int f^2(x)\,dx\Big)^{\frac{1}{2}}\Big(\int g^2(x)\,dx\Big)^{\frac{1}{2}} \tag{8}$$

At this stage we need two results from Fourier theory. First we need to know what the Fourier transform of derivative is and, second, we need the Plancherel formula which connects the norms of the function and its Fourier transform. The Fourier transform of $f'(x)$ is:

$$\hat{f}'(x) = 2\pi i\xi\hat{f}(x) \tag{9}$$

Plancherel's formula states that if $f \in \mathcal{S}(\mathbb{R})$ (which is assumed to be the case) then:

$$\|\hat{f}\| = \|f\| \tag{10}$$

where $\|f\| = \left(\int_{-\infty}^{\infty}|f(x)|^2\,dx\right)^{\frac{1}{2}}$.
Using (9) and (10) we see that:

$$\begin{aligned}\int_{-\infty}^{\infty}|\psi'(x)|^2\,dx =& \int_{-\infty}^{\infty}|2\pi\,i\,\xi\,\hat{\psi}(\xi)|^2\,d\xi \\ =&4\pi^2\int_{-\infty}^{\infty}\xi^2|\hat{\psi}(\xi)|^2\,d\xi\end{aligned} \tag{11}$$

Thus using the last line of (7) and (11) we have:

$$\begin{aligned}\left(\int_{-\infty}^{\infty}x^2|\psi(x)|^2\,dx\right)\left(\int_{-\infty}^{\infty}|\psi'(x)|^2\,dx\right) \geq& \frac{1}{4} \\ \left(\int_{-\infty}^{\infty}x^2|\psi(x)|^2\,dx\right)\left(4\pi^2\int_{-\infty}^{\infty}\xi^2|\hat{\psi}(\xi)|^2\,d\xi\right) \geq& \frac{1}{4} \\ \left(\int_{-\infty}^{\infty}x^2|\psi(x)|^2\,dx\right)\left(\int_{-\infty}^{\infty}\xi^2|\hat{\psi}(\xi)|^2\,d\xi\right) \geq& \frac{1}{16\pi^2}\end{aligned} \tag{12}$$

What is left is working out the constants stated above ie A in $\psi(x) = Ae^{-Bx^2}$. In the last line of (7) of we have equality we must have equality in the step where we have used the Cauchy-Schwarz inequality ie:

$$\int_{-\infty}^{\infty} \left|x\right| \left|\psi(x)\right| \left|\psi'(x)\right| dx = \left(\int_{-\infty}^{\infty} x^2 |\psi(x)|^2 \, dx\right)^{\frac{1}{2}} \left(\int_{-\infty}^{\infty} |\psi'(x)|^2 \, dx\right)^{\frac{1}{2}} \tag{13}$$

Therefore:

$$\psi'(x) = \beta x \, \psi(x) \tag{14}$$

for some constant β. Solving (14) gives us:

$$\psi(x) = Ae^{\frac{\beta x^2}{2}} \tag{15}$$

Because we want $\psi(x)$ to be a Schwarz function it has to decay fast so we need $\beta = -2B < 0$ for some constant $B > 0$. The normalisation condition:

$$\int_{-\infty}^{\infty} |\psi(x)|^2 \, dx = 1 \tag{16}$$

implies that:

$$\begin{aligned} \int_{-\infty}^{\infty} |\psi(x)|^2 \, dx =& A^2 \int_{-\infty}^{\infty} e^{-2Bx^2} \, dx \\ =& \frac{A^2}{\sqrt{2B}} \int_{-\infty}^{\infty} e^{-x^2} \, dx \\ =& \frac{A^2}{\sqrt{2B}} \sqrt{\pi} \\ =& 1 \end{aligned} \tag{17}$$

Thus $A^2 = \sqrt{\frac{2B}{\pi}}$. Note that we have used the standard result that $\int_{-\infty}^{\infty} e^{-x^2} \, dx = \sqrt{\pi}$.

Let us assume that there is a function $\phi(x)$ different from $\psi(x)$ ie $\phi(x) \neq \psi(x)$ which satisfies the following:

$$\Bigg(\underbrace{\int_{-\infty}^{\infty} x^2 |\phi(x)|^2 \, dx}_{=\Phi}\Bigg) \Bigg(\underbrace{\int_{-\infty}^{\infty} \xi^2 |\hat{\phi}(\xi)|^2 \, d\xi}_{=\hat{\Phi}}\Bigg) = \frac{1}{16\pi^2} \tag{18}$$

ie this different function also equals the bound like a Gaussian.

$$\underbrace{\int_{-\infty}^{\infty} x^2 |\phi(x)|^2 \, dx}_{=\Phi} > \underbrace{\int_{-\infty}^{\infty} x^2 |\psi(x)|^2 \, dx}_{=\Psi} \tag{19}$$

ie this different function has a greater uncertainty than the Gaussian.

$$\underbrace{\int_{-\infty}^{\infty} \xi^2 |\hat{\phi}(\xi)|^2 \, d\xi}_{=\hat{\Phi}} > \underbrace{\int_{-\infty}^{\infty} \xi^2 |\hat{\psi}(\xi)|^2 \, d\xi}_{=\hat{\Psi}} \tag{20}$$

ie this different function's Fourier transform is more uncertain than the Gaussian.

If multiply (19) by $\hat{\Psi}$ we have (noting that it is non-negative):

$$\Phi \, \hat{\Psi} > \Psi \, \hat{\Psi} \tag{21}$$

Similarly if we multiply (20) by Φ (again non-negative) we have:

$$\Phi \, \hat{\Phi} > \Phi \, \hat{\Psi} \tag{22}$$

But ex hypothesi $\Phi \, \hat{\Psi} = \Phi \, \hat{\Phi} = \frac{1}{16\pi^2}$ so (21) and (22) imply that $\Phi \, \hat{\Psi} > \Phi \, \hat{\Psi}$, a contradiction hence the Gaussian is the only form of function which equals the bound in (4).

1.1 Using the calculus of variations to derive the Gaussian from entropy considerations

As noted above we need to find the function y which maximises the entropy functional $H(y) = -\int_{-\infty}^{\infty} y \ln y \; dx$ where $y \geq 0$ subject to the constraints $\int_{-\infty}^{\infty} y \, dx = 1$ and $\int_{-\infty}^{\infty} x^2 \, y dx = \sigma^2$ where $y \to 0$ as $x \to \pm\infty$.

The theoretical background to this type of problem can be found in Section 12 of [3] and many other textbooks.

Our starting point is this equation (see [3], pages 4250):

$$\frac{\partial}{\partial y}(F + \lambda G) - \frac{d}{dx}\frac{\partial}{\partial y'}(F + \lambda G) = 0 \tag{23}$$

where $F = -y \ln y$ and $G = (1 + x^2) \, y$. G comes from combining the integral constraints ie $\int_{-\infty}^{\infty} G(x, y, y') \, dx = \int_{-\infty}^{\infty} (y + x^2 \, y) \, dx = 1 + \sigma^2$.

Since $F + \lambda G = -y \ln y + \lambda(1 + x^2) \, y$, (23) yields the following:

$$\ln y = \lambda(1 + x^2) - 1 \tag{24}$$

From (24) it follows that:

$$y = e^{\lambda(1+x^2)-1} \tag{25}$$

Now because $y \to 0$ as $x \to \pm\infty$, it must be the case that $\lambda < 0$ ie $\lambda = -\gamma^2$. Thus we have that:

$$y = e^{-(1+\gamma^2)} \,.\, e^{-\gamma^2 x^2} \tag{26}$$

From the normalisation condition $\int_{-\infty}^{\infty} y\,dx = 1$ we get the following (upon making the substitution $u = \gamma x$):

$$\int_{-\infty}^{\infty} y\,dx = \int_{-\infty}^{\infty} e^{-(1+\gamma^2)} \,.\, e^{-\gamma^2 x^2}\,dx = \frac{e^{-(1+\gamma^2)}}{\gamma} \int_{-\infty}^{\infty} e^{-u^2}\,du = \frac{e^{-(1+\gamma^2)}}{\gamma}\sqrt{\pi} = 1 \tag{27}$$

Thus $e^{-(1+\gamma^2)} = \frac{\gamma}{\sqrt{\pi}}$ and so using (26):

$$y = \frac{\gamma}{\sqrt{\pi}}\, e^{-\gamma^2 x^2} \tag{28}$$

To find γ we use the second constraint $\int_{-\infty}^{\infty} x^2\,y\,dx = \sigma^2$ as follows:

$$\int_{-\infty}^{\infty} x^2\,y\,dx \quad = \quad \int_{-\infty}^{\infty} x^2\,\frac{\gamma}{\sqrt{\pi}}\,e^{-\gamma^2 x^2}\,dx \quad = \quad \frac{\gamma}{\sqrt{\pi}} \int_{-\infty}^{\infty} x\,(x\,e^{-\gamma^2 x^2})\,dx \tag{29}$$

Now (29) can be evaluated by integration by parts with $dv = x\,e^{-\gamma^2 x^2}\,dx$, $v = \frac{-e^{-\gamma^2 x^2}}{2\gamma^2}$, $u = x$ and $du = dx$. Performing the integration we get:

$$\begin{aligned}
\frac{\gamma}{\sqrt{\pi}} \int_{-\infty}^{\infty} x\,(x\,e^{-\gamma^2 x^2})\,dx =& \frac{\gamma}{\sqrt{\pi}} \left\{ \underbrace{\frac{-xe^{-\gamma^2 x^2}}{2\gamma^2}\Big|_{-\infty}^{\infty}}_{=0} + \frac{1}{2\gamma^2} \int_{-\infty}^{\infty} e^{-\gamma^2 x^2}\,dx \right\} \\
=& \frac{\gamma}{\sqrt{\pi}}\,\frac{1}{2\gamma^2} \int_{-\infty}^{\infty} e^{-\gamma^2 x^2}\,dx \\
=& \frac{1}{2\gamma\,\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\gamma^2 x^2}\,dx \\
=& \frac{1}{2\gamma\,\sqrt{\pi}}\,\frac{\sqrt{\pi}}{\gamma} \\
=& \frac{1}{2\gamma^2}
\end{aligned} \tag{30}$$

Hence $\frac{1}{2\gamma^2} = \sigma^2$ and so $\gamma = \frac{1}{\sqrt{2\sigma^2}}$ and substituting into (28) gives:

$$y = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{x^2}{2\sigma^2}} \tag{31}$$

2 Conclusions

The entropy maximisation property of the Gaussian is one of many reasons why the Gaussian is so ubiquitous. Its role in the Heisenberg Uncertainty Principle is yet another reason for that ubiquity.

3 References

[1] C E Shannon, A Mathematical Theory of Communication', The Bell System Technical Journal, Vol. XXVII, No.3, July 1948.

[2] Richard C Tolman, The Principles of Statistical Mechanics, Dover, 1979.

[3] I M Gelfand and S V Fomin, Calculus of Variations , Dover Publications, 1963

[4] Peter Haggstrom, How Einstein did Brownian motion in 1905, https://gotohaggstrom.com/How%20Einstein%20did%20Brownian%20motion%20in%201905.pdf

[5] Peter Haggstrom, How Maxwell derived his velocity distribution law, https://gotohaggstrom.com/How%20Maxwell%20derived%20his%20velocity%20distribution%20law_24April2019.pdf

[6] Elias M Stein & Rami Shakarchi, Fourier Analysis: An Introduction, Princeton University Press, 2003.

[7] Peter Haggstrom, Basic Fourier Integrals, http://www.gotohaggstrom.com/Basic%20Fourier%20integrals.pdf

[8] Sarvesh Kumar Yadav, Rajib Saha, Investigating non-Gaussianity in Cosmic Microwave Background Temperature Maps using Spherical Harmonic Phases, https://arxiv.org/abs/2001.10960

[9] Peter Haggstrom, The central limit theorem - How Laplace actually proved it , https://www.gotohaggstrom.com/The%20central%20limit%20theorem%20-%20how%

20Laplace%20actually%20proved%20it.pdf

[10] Werner Heisenberg, The Physical Content of Quantum Kinematics and Mechanics, https://www.informationphilosopher.com/solutions/scientists/heisenberg/Heisenberg_Uncertai

[11] Masanao Ozawa, Heisenberg's original derivation of the uncertainty principle and its universally valid reformulations, https://arxiv.org/abs/1507.02010v3.pdf. This link does not want to work in the pdf so you will need to copy it and search.

4 History

Created 06 February 2022