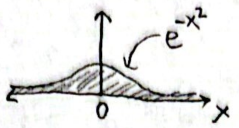
Some Useful Integrals for Info Theory (Part 1)

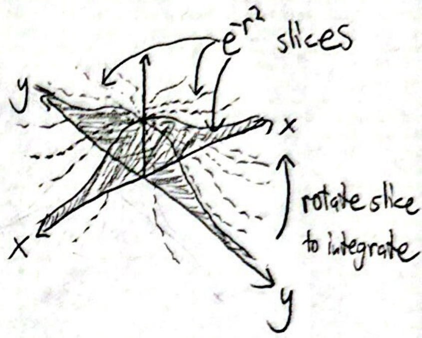
In this lecture, I will cover single-dimension Integrals that are good to have in your pocket when we study the entropy of Gaussian distributions.

Integral #1: $\boxed{\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}}$



The classic approach here is to think about a radius, $r^2 = x^2 + y^2$, so we are thinking about circles in a 2D plane. In general, quadratic terms play better in 2D.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\,dx\,dy \xrightarrow{\text{Translate to rotational coordinates}} \int_0^{2\pi}\int_0^{\infty} e^{-r^2}\,\underbrace{r\,dr\,d\theta}_{\text{rotational differential}}$$



Evaluate integral:

$$\int_0^{2\pi}\int_0^{\infty} e^{-r^2} r\,dr\,d\theta = \int_0^{2\pi}\left[-\tfrac{1}{2}e^{-r^2}\right]_0^{\infty} d\theta = \int_0^{2\pi} \tfrac{1}{2}\,d\theta = \pi$$

Let's look at this in rectangular coordinates again: $\pi = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\,dx\,dy = \int_{-\infty}^{\infty} e^{-x^2}\,dx \int_{-\infty}^{\infty} e^{-y^2}\,dy$

No coupled terms! We can split this integral.

These two are the same! $A = \int_{-\infty}^{\infty} e^{-x^2}\,dx = \int_{-\infty}^{\infty} e^{-y^2}\,dy$

Therefore, $\pi = A^2$, and $A = \sqrt{\pi} = \int_{-\infty}^{\infty} e^{-x^2}\,dx$ ✓

Integral #2: $\boxed{\int_{-\infty}^{\infty} x^2 e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}}$

This one is pretty evil. It's an integration by parts, but the parts are very unintuitive. The expression is begging for a u-substitution strategy, which doesn't help.

Split the integral as follows: $u = x \quad dv = xe^{-x^2} \quad du = dx \quad v = -\frac{1}{2}e^{-x^2}$

$$\int_{-\infty}^{\infty} x^2 e^{-x^2}\,dx = \underbrace{\int_{-\infty}^{\infty} u\,dv = [uv]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} v\,du}_{\text{The by-parts method}} = \underbrace{\left[-\frac{x}{2}e^{-x^2}\right]_{-\infty}^{\infty}}_{\text{This is } 0} - \underbrace{\int_{-\infty}^{\infty} -\frac{1}{2}e^{-x^2}\,dx}_{\text{Use Integral \#1}} = \frac{\sqrt{\pi}}{2}$$ ✓

$\boxed{\text{Integral \#3}}$: $\boxed{\int_{-\infty}^{\infty} e^{-kx^2}\,dx = \sqrt{\frac{\pi}{k}}}$ A small but Important variation of <u>Integral #1</u>.

Using the rotational coordinates: $\int_0^{2\pi}\int_0^{\infty} e^{-kr^2}\, r\,dr\,d\theta = \int_0^{2\pi}\left[-\frac{1}{2k}e^{-kr^2}\right]_0^{\infty} d\theta = \frac{1}{2k}\int_0^{2\pi} d\theta = \frac{\pi}{k}$

Use same argument as in #1, $\int_{-\infty}^{\infty} e^{-kx^2}\,dx = \sqrt{\frac{\pi}{k}}$ ✓

$\boxed{\text{Integral \#4}}$: $\boxed{\int_{-\infty}^{\infty} x^2 e^{-kx^2}\,dx = \frac{1}{2k}\sqrt{\frac{\pi}{k}}}$ Important variation of <u>Integral #2</u>.

By-parts: $u = x \quad dv = xe^{-kx^2}, \quad du = dx \quad v = -\frac{1}{2k}e^{-kx^2}$

$$\underbrace{\left[-\frac{x}{2k}e^{-kx^2}\right]_{-\infty}^{\infty}}_{0,\text{ again}} + \frac{1}{2k}\underbrace{\int_{-\infty}^{\infty} e^{-kx^2}\,dx}_{\text{Use Integral \#3}} = \frac{1}{2k}\sqrt{\frac{\pi}{k}}$$ ✓

The Gaussian maximizes entropy (mostly)

The concept of "Maximum Entropy" is not only a math concept, but also carries with it crucial epistimological insight. I recommend looking up "Maximum Entropy Principle" and ET Jayne's paper on it. Therefore, it's in our interest to see what the math does here.

By "entropy", I will be referring to "differential entropy" in these notes, defined as follows:

$$H(X) = \int_{-\infty}^{\infty} -p(x)\ln p(x)\,dx, \quad \text{where } p(x) \text{ is a probability } \underline{\text{density}} \text{ function over } x.$$

(continuous)

Technically "differential entropy" and "entropy" are not strongly equivalent, but for the most part it is innocuous to pretend they are the same.

Theorem: The probability distribution that maximizes entropy given a <u>fixed variance σ^2</u> is the Gaussian.

Proving this will require some prerequisites in Lagrange multipliers (very necessary!) and Calculus of Variations (less necessary, I can briefly explain). In general, optimizing over the space of continuous functions is an idea that takes getting used to.

$$\max_{p(x)} \int_{-\infty}^{\infty} -p(x)\ln p(x)\,dx \longrightarrow \text{Objective, entropy}$$

$$\text{such that} \int_{-\infty}^{\infty} p(x)\,dx = 1 \longrightarrow \text{probabilities must add to } 1. \text{ Assign this constraint to the multiplier } \lambda_1.$$

$$\int_{-\infty}^{\infty} (\underset{\substack{\uparrow \\ \text{mean}}}{x-\mu})^2 p(x) = \underset{\substack{\uparrow \\ \text{variance}}}{\sigma^2} \Rightarrow \text{This is our fixed variance. Assign this constraint to the multiplier } \lambda_2.$$

Augment objective with Lagrange multipliers: $\int_{-\infty}^{\infty} -p(x)\ln p(x)\,dx - \lambda_1 \int_{-\infty}^{\infty} p(x)\,dx + \lambda_1 - \lambda_2 \int_{-\infty}^{\infty} (x-\mu)^2 p(x)\,dx + \lambda_2$

$$= \int_{-\infty}^{\infty} -p(x)\ln p(x) - \lambda_1 p(x) - \lambda_2 (x-\mu)^2 p(x)\,dx + \lambda_1 + \lambda_2$$

The maximum point occurs when the derivative with respect to $p(x)$ inside the integral is 0.

$$-\ln p(x) - 1 - \lambda_1 - \lambda_2 (x-\mu)^2 = 0 \implies p(x) = e^{-1-\lambda_1-\lambda_2(x-\mu)^2}$$

A few notes:

1. Entropy is a concave function, so we know the stationary point (0-derivative) is a maximum, and not a minimum or saddle point.
2. Setting this one derivative to 0 is actually setting an infinite number of derivatives to 0 in disguise. To illustrate, see what happens when the integral becomes a discrete sum.

$$\mathcal{L}(p(x)) = \int_{-\infty}^{\infty} f(p(x))\,dx \implies \mathcal{L}(p(x_0), p(x_1), \ldots, p(x_N)) = \sum_{i=0}^{N} f(p(x_i))$$

(just use f for any placeholder function)

$$\left.\begin{aligned} \frac{\partial \mathcal{L}}{p(x_0)} &= \frac{df}{dp} = 0 \\ \frac{\partial \mathcal{L}}{p(x_1)} &= \frac{df}{dp} = 0 \\ &\vdots \\ \frac{\partial \mathcal{L}}{p(x_N)} &= \frac{df}{dp} = 0 \end{aligned}\right\}$$

need this whole gradient to be 0, as you would for a multivariable function. When your sum "acts like" an integral, you get to simplify to one constraint.

$$\frac{df}{dp} = 0$$

This is good intuition for continuous integrals, but it's not rigorous. The real justification is the "fundamental lemma of the calculus of variations."

Back to our derivative constraint: $p(x) = e^{-1-\lambda_1-\lambda_2(x-\mu)^2}$

All that's left to do is to solve for λ_1, λ_2 with our constraints.

$$\int_{-\infty}^{\infty} p(x)\,dx = 1$$

$$\int_{-\infty}^{\infty} e^{-\lambda_2(x-\mu)^2}\,dx = e^{1+\lambda_1}$$

Using the "Useful Integral" notes,*

$$\boxed{\sqrt{\frac{\pi}{\lambda_2}} = e^{1+\lambda_1}}$$

* due to translational symmetry, the mean μ is insignificant.

$$\int_{-\infty}^{\infty} (x-\mu)^2 p(x)\,dx = \sigma^2$$

$$\int_{-\infty}^{\infty} (x-\mu)^2 e^{-\lambda_2(x-\mu)^2}\,dx = \sigma^2 e^{1+\lambda_1}$$

Another useful integral to know,

$$\boxed{\frac{1}{2\lambda_2}\sqrt{\frac{\pi}{\lambda_2}} = \sigma^2 e^{1+\lambda_1}}$$

Substitute $e^{1+\lambda_1}$ from one to another: $\frac{1}{2\lambda_2} = \sigma^2 \longrightarrow \boxed{\lambda_2 = \frac{1}{2\sigma^2}}$

Get λ_1: $\sqrt{2\pi\sigma^2} = e^{1+\lambda_1} \longrightarrow \ln\sqrt{2\pi\sigma^2} = 1+\lambda_1 \longrightarrow \boxed{\lambda_1 = \ln\sqrt{2\pi\sigma^2} - 1}$

Put them in $p(x)$,

$$p(x) = e^{-1-(\ln\sqrt{2\pi\sigma^2}-1)-\frac{(x-\mu)^2}{\sigma^2}} = \boxed{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{\sigma^2}}} \checkmark$$

This is a Gaussian.