ORF 526 · Homework 1 · due Sep 7

Entropy, and why the Gaussian maximizes it

In Lecture 1 we defined the Gaussian by its Fourier transform and claimed, without proof, a second characterization: *among all laws with a given mean and covariance, the Gaussian maximizes entropy.* This assignment proves it. Along the way you will build the notion of entropy from scratch, which we will need again in Lecture 3 — where maximum entropy turns out to be the reason Gaussians appear at all.

recalled from lecture For $\Sigma \succ 0$, Fourier inversion applied to Definition 1.4 gives $\mathcal{N}(\mu, \Sigma)$ the density

$$q(x) \;=\; \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\Big(-\tfrac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\Big),$$

which in one dimension is $(2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/2\sigma^2}$. This is the only fact from Lecture 1 used below.

1. Surprise

Before entropy, surprise. Suppose an event of probability p occurs. How surprised should you be? Whatever the answer $s(p)$ is, two things ought to hold: certainty is unsurprising, so $s(1) = 0$; and if two *independent* things happen, the surprises should add, so $s(pq) = s(p) + s(q)$.

Problem 1. core Show that the only continuous $s : (0, 1] \to [0, \infty)$ with $s(1) = 0$ and $s(pq) = s(p) + s(q)$ is $s(p) = -c \log p$ for some $c \geq 0$. *Hint: substitute $p = e^{-u}$ and see what the equation becomes.*

So surprise is $-\log p$, and the base of the logarithm is only a choice of units. We take $c = 1$ and use natural logs throughout.

2. Entropy is expected surprise

If X takes value x with probability $p(x)$, the surprise you actually experience is the random quantity $-\log p(X)$. Its expectation is the **entropy**

$$H(p) \;=\; \mathbb{E}\big[-\log p(X)\big] \;=\; -\sum_x p(x) \log p(x).$$

Entropy is the surprise you should expect *before* looking: how uncertain the distribution is, in units of surprise.

Problem 2. core Compute H for a coin with bias q. Sketch it as a function of $q \in [0, 1]$, and find where it is largest. Compute H for the uniform distribution on n points.

Your answer to the second part suggests a general principle: uncertainty is greatest when no outcome is favoured. Let us prove it.

Problem 3. Let p be any distribution on n points and u the uniform one. Show $H(p) \leq H(u) = \log n$, with equality only for $p = u$. *Hint:* log *is concave, so Jensen's inequality applies to* $\mathbb{E}[\log(1/p(X))]$.

This is the first instance of the pattern we are after: *fix a constraint, and entropy is maximized by the least committal distribution satisfying it.* On a finite set with no constraint at all, that is the uniform law. The Gaussian will be the answer to the same question with a different constraint.

3. The continuous case, and a warning

For a random variable with density p on $\mathbb{R}^n$, the same formula defines the **differential entropy**

$$h(p) \;=\; -\int p(x) \log p(x)\, dx.$$

It is not the discrete entropy in disguise, and it misbehaves in two ways worth meeting now.

Problem 4. core Compute h for: the uniform law on $[0, a]$; the exponential law with rate λ; and $N(0, \sigma^2)$. *You should find* $h(N(0, \sigma^2)) = \frac{1}{2}\log(2\pi e\sigma^2)$. Then exhibit a distribution with *negative* differential entropy, and explain why no discrete entropy can be negative.

Problem 5. Show $h(aX) = h(X) + \log|a|$ for $a \neq 0$. Conclude that differential entropy is not invariant under a change of units — so "the entropy of a length" is not a well-defined number.

The cure is to stop measuring entropy in absolute terms and measure it *relative* to a reference.

4. Relative entropy

For densities p and q, the **relative entropy** (or Kullback–Leibler divergence) is

$$D(p\,\|\,q) \;=\; \int p(x)\log\frac{p(x)}{q(x)}\,dx \;=\; \mathbb{E}_p\left[\log\frac{p(X)}{q(X)}\right].$$

Read it as the extra surprise you suffer by expecting q when the truth is p — which predicts it should never be negative.

Problem 6. CORE Prove **Gibbs' inequality**: $D(p\,\|\,q) \geq 0$, with equality if and only if $p = q$. *Hint: apply Jensen to $-\log$, exactly as in Problem 3.* Then check that $D(p\,\|\,q)$ is unchanged if you rescale X — unlike h.

5. The theorem

Now fix a mean μ and a covariance $\Sigma \succ 0$, and let q be the Gaussian density recalled on page 1. Let p be *any* density with that same mean and covariance. We want $h(p) \leq h(q)$.

Start from $D(p\,\|\,q) \geq 0$ and expand:

$$0 \;\leq\; D(p\,\|\,q) \;=\; -h(p) - \int p(x)\log q(x)\,dx.$$

Everything now hinges on the second term, and on one structural fact about the Gaussian.

Problem 7. Write out $\log q(x)$ explicitly and observe that it is a *quadratic polynomial* in x. Deduce that $\int p\log q$ depends on p **only through its mean and covariance** — and therefore that

$$\int p(x)\log q(x)\,dx \;=\; \int q(x)\log q(x)\,dx \;=\; -h(q).$$

Problem 8. Combine the last two displays to conclude $h(p) \leq h(q)$, with equality only when $p = q$. Then compute $h(q)$ itself and check that it equals $\frac{1}{2}\log\big((2\pi e)^n \det\Sigma\big)$.

That is the theorem. Note how little the Gaussian had to do: the argument turned entirely on $\log q$ being quadratic, so that integrating it against p could only see p's first two moments. *A quadratic is exactly what a covariance constraint can measure.*

6. The same statement in Fourier

We defined the Gaussian by its characteristic function, and have now characterized it a second way. The two are closer than they look. Write the cumulant expansion

$$\log\varphi_p(t) \;=\; i\langle t,\mu\rangle - \tfrac{1}{2}t^\top\Sigma t \;+\; (\text{higher order}),$$

whose coefficients are the cumulants of p. Fixing the mean and covariance fixes the first two coefficients, and the Gaussian is the law that has no others. So the theorem you just proved reads:

maximum entropy at fixed covariance means setting every cumulant beyond the second to zero.

Problem 9. Verify directly from Definition 1.4 that $\log\varphi$ of $\mathcal{N}(\mu,\Sigma)$ is exactly quadratic, so that every cumulant beyond the second vanishes.

CORE problems are short and self-contained. The first quiz, on Wednesday Sep 9, draws on the core problems of this set and of Homework 2. AI assistance is permitted and encouraged throughout.