Math Notes

Kieran Harvie

Contents

0.1 Mean and Variance, and the arbitrariness thereof

For some time I have wondered about the arbitrariness around the mean and variance. For example why the arithmetic mean instead of the geometric or root-mean-squared? And why square root the variance to give the standard deviation?

Well the strictness of Markov's and Chebyshev's might provide a reason. Both rely of the conditional expected value, so just to reiterate:

$$E[X|X \geq a] \geq a$$

Since everything X can be is greater then a it's expected value must be greater than a. Notice the strictness of the inequality, this will be used to make the following inequalities much stricter.

Markov

$$\begin{aligned}\mu =&E[X]\\ =&P(X \leq a)E[X|X \leq a] + P(X \geq a)E[X|X \geq a]\\ \geq&0 \cdot E[X|X \leq a] + P(X \geq a)a\\ \frac{\mu}{a} \geq&P(X \geq a)\end{aligned}$$

Chebyshev

$$\begin{aligned}&E[(X-a)^2]\\ =&P(|X-a| \leq b)E[(X-a)^2||X-a| \leq b] + P(|X-a| > b)E[(X-a)^2||X-a| > b]\end{aligned}$$

A General Relation

Assume:

$$f(S') \geq 0, \quad g(S) \geq 0$$

Then through:

$$E[f(X)] = P(X \in S)E[f(X)|X \in S] + P(X \in S')E[f(X)|X \in S']$$

We have:

$$1 - \frac{E[g(X)]}{E[g(X)|X \in S']} \le P[X \in S] \le \frac{E[f(X)]}{E[f(X)|X \in S]}$$

With dual equality if:

$$f = 1_S, \quad g = 1_{S'}$$

Covariance

Lets try to find the lest squares regression between X and Y such that:

$$E[X] = E[Y] = 0, E[X^2] = E[Y^2] = 1$$

Since the expected values are both zero the line is through the origin

$$\begin{aligned}\sum_n (mx_n + c - y_n)^2 =& nE[(mX + c - Y)^2] \\ =& n\Big(E[m^2X^2] + E[c^2] + E[Y^2] + E[2cmX] + E[-2mXY] + E[-2cY]\Big) \\ =& n(m^2 + c^2 + 1 - 2mE[XY])\end{aligned}$$

Trying to minimize this value by our selection of trivially gets:

$$c = 0, \quad m = E[XY]$$

Just expanding the definitions gives:

$$COV[X, Y] = E[(X - E[X])(Y - E[Y])] = E[XY] = m$$

Hence the covariance can 'naturally' be interpreted and the first order function between the valuables.

$$E[f(X)] \approx E[f_0 + f_1X + f_2X^2/2] = f_0 + \mu f_1 + \sigma^2 f_2/2$$

$$E[f(X)] \approx E[f(\mu) + (X - \mu)f'(\mu) + (X - \mu)^2/2f''(\mu)] = f(\mu) + \frac{f''(\mu)}{2}\sigma^2$$

0.2 Interest Identities

Let P be the principle invested at a rate of r. Consider four different investment scenarios:

- Not invested: $P_0 = P$.
- Fully Invested at the beginning, one instalment at the end:

$$P_1 = (1 + r)P$$

- Continuously invested, continuous installments:

$$P_2 = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{P}{n}\left(1 + \frac{r}{n}\right)^k = \frac{\exp(r) - 1}{r}P$$

- Fully Invested at the beginning, continuous instalments:

$$P_3 = \lim_{n\to\infty} P\left(1 + \frac{r}{n}\right)^n = \exp(x)P$$

Interestingly the relative size of P_2 and P_1 depend on r. P_1 starts on P_2 but switches as r increases.

P_3 is always the best, the proof for P_1 and P_0 are obvious. $P_3 > P_2$ follows from:

$$0 < \int_0^r t\exp(t)\,dt = \big[(t-1)\exp(t)\big]_0^r = (r-1)\exp(r) + 1$$

The following interesting identities hold:

$$P_3 = rP_2 + P_0$$

$$P_3 - P = r(P_2 - P) + (P_1 - P)$$

The breaks first neatly breaks P_3 into a nice linear sum. The second does similar for the profit of the investment, total yield minus principle.

0.3 Discontinuities in a Non-decreasing Function

Let f be a non-decreasing function.

Define the jump function J as:

$$J(x) = \inf\{f(t)|t > x\} - \sup\{f(t)|t < x\}$$

This function is well defined since the sets are appropriately bound by $f(x)$. And it is clear that $J(d) \neq 0$ iff d is a discontinuity and that J is non-negative.

Let $U = (x_0, x_1)$. For $d_n \in U$ with $n < m \Rightarrow d_n < d_m$ we have:

$$f(x_1) - f(x_0) \geq \sum_k J(d_k)$$

Proof:

$$\begin{aligned}
&\sum_k J(d_k) \\
=&\inf\{f(t)|t > d_n\} - \sup\{f(t)|t < d_0\} + \sum_k \left[\inf\{f(t)|t > d_{k-1}\} - \sup\{f(t)|t < d_k\}\right] \\
\leq& f(d_n) - f(d_0) + \sum_k \left[f(d_{k-1}) - f(d_k)\right] \\
\leq& f(d_n) - f(d_0) \\
\leq& f(x_n) - f(x_0)
\end{aligned}$$

Let $S_n = \{d \in U|J(d) > \frac{1}{n}(f(x_1) - f(x_0))\}$ From the previous inequality there are at most n elements in S_n. Hence:

$$\{d \in U|J(d) > 0\} = \bigcup_n \left\{d \in U|J(d) > \frac{1}{n}(f(x_1) - f(x_0))\right\}$$

Is countable, hence the number of discontinuities of f on U is countable.

By corollary the discontinuities of a non-decreasing function on $\mathbb{R}$ are countable:

Let f be a non-decreasing function on $\mathbb{R}$. Let $X_n = (n-1, n+1)$, clearly $\mathbb{R} = \bigcup_n X_n$[1]. Assume f has an uncountable number of discontinuities then at least one X_n contains uncountable discontinuities. Otherwise there would be a countable set of countable sets of discontinuities, making them countable. But f being non-decreasing function and having an uncountable number of discontinuities in X_n is a contradiction.

[1]Having them overlap simplify the proof by avoiding literal edge-cases.

0.4 Integrals and Symmetry

0.4.1 Domain Symmetry

Let U be a subset of $\mathbb{R}^n$ and let $\phi : U \to U$ be a function such that:

$$\begin{aligned} \phi(U) &= U \\ |\det \phi'(\mathbf{u})| &= 1 \end{aligned}$$

Basically ϕ is a linear permutation[2] on U, this is a symmetry in the most direct sense. We obtain the following:

$$\begin{aligned} \int_U f(\mathbf{v})\, d\mathbf{v} &= \int_{\phi(U)} f(\mathbf{v})\, d\mathbf{v} \\ &= \int_U f(\phi(\mathbf{u}))|\det \phi'(\mathbf{u})|\, d\mathbf{u} \\ &= \int_U f(\phi(\mathbf{u}))\, d\mathbf{u} \end{aligned}$$

In particular we get:

$$0 = \int_U \big(f(\mathbf{u}) - f(\phi(\mathbf{u}))\big)\, d\mathbf{u}$$

This integral is important since a function can be split into a vanishing and non-vanishing part:

$$\begin{aligned} f(\mathbf{u}) =& \frac{1}{2}\big(f(\mathbf{u}) + f(\phi(\mathbf{u})) + \frac{1}{2}\big(f(\mathbf{u}) - f(\phi(\mathbf{u}))\big) \\ \int_U f(\mathbf{u})\, d\mathbf{u} =& \frac{1}{2}\int_U \big(f(\mathbf{u}) + f(\phi(\mathbf{u}))\, d\mathbf{u} + \frac{1}{2}\int_U \big(f(\mathbf{u}) - f(\phi(\mathbf{u}))\big)\, d\mathbf{u} \\ =& \frac{1}{2}\int_U \big(f(\mathbf{u}) + f(\phi(\mathbf{u}))\, d\mathbf{u} \end{aligned}$$

[2]Note that ϕ being a permutation requires that if the magnitude of the determinate is constant it must be unity, this can be seen by setting f to a constant.

For example, consider the classic odd function on an integral centered at 0.

$$\phi(x) = -x$$
$$U = [-1, 1]$$

We get the familiar:

$$\int_{-1}^{1} f(x)\, dx = \frac{1}{2}\int_{-1}^{1} \big(f(x) + f(-x)\big)\, dx$$

The utility of this relation can be seen by applying it to the basis of a class of function. Let $V = \langle 1, x, x^2 \rangle$, this is a basis for all parabolas. Notice that x base element vanishes, simplify the evaluation of integrals.

For a 2-D example recall that for two dimensional change of variables:

$$(x, y) = \phi(u, v)$$

We have:

$$|\det \phi'(\mathbf{v})| = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

The rotation symmetry for a regular triangle is:

$$(x, y) = \frac{1}{2}(-u - \sqrt{3}v, \sqrt{3}u - v)$$

Obviously the symmetries act like a group and with functions being a vector space we can use group representations.

0.4.2 Function Symmetry

Let f and ϕ be functions such that:

$$f(t) = \phi'(t) f(\phi(t))$$

Then for arbitrary x_0 and x_1 we have:

$$\begin{aligned}
\int_{x_0}^{x_1} f(t)\,dt &= \int_{\phi(x_0)}^{\phi(x_1)} \phi'(t) f(\phi(t))\,dt \\
&= \int_{\phi(x_0)}^{\phi(x_1)} f(t)\,dt \\
&= \int_{x_1}^{\phi(x_1)} f(t)\,dt + \int_{\phi(x_0)}^{x_1} f(t)\,dt \\
\int_{x_0}^{\phi(x_0)} f(t)\,dt &= \int_{x_1}^{\phi(x_1)} f(t)\,dt
\end{aligned}$$

Hence the integral value is independent of x_n, in particular if ϕ has a fixed point then the integral is zero.

0.5 Lagrange Multiplier

Recall that local extrema x of the function $f : \mathbb{R}^n \to \mathbb{R}$ subject to contrasts g_i satisfy:

$$\nabla f(x) = \sum_i \lambda_i \nabla g_i(x)$$

The core observation is that if ∇f has a component outside the span of ∇g_i then you can move in that direction while keeping g_i's constant, contradicting the point being an extrema.

But the actual constants λ_i have a useful interpretation as the rate the value of f at the extrema changes as the constant g_i changes. To see this pick a particular g_j and construct a d such that:

$$d \cdot \nabla g_i = D\delta_{i,j}$$

You can achieve this by iteratively removing components in some matter like the following:

$$d_0 = \nabla g_0,\ d_{n+1} = d_n - d_n \cdot \nabla g_n$$

Now scale d down such that functions around the extrema can be approximated through targets[3]. We have:

$$\begin{aligned}
g_i(x+d) =& g_i(x) + d \cdot \nabla g_i(x) \\
=& g_i(x) + D\delta_{i,j} \\
f(x+d) =& f(x) + d \cdot \nabla f(x) \\
=& f(x) + \sum_i \lambda_i d \cdot \nabla g_i(x) \\
=& f(x) + D\lambda_j \\
\nabla f(x+d) =& \nabla(f(x) + D\lambda) \\
=& \nabla f(x) \\
=& \sum_i \lambda_i \nabla g_i(x) \\
=& \sum_i \lambda_i \nabla \big(g_i(x+d) - D\delta_{i,j}\big) \\
=& \sum_i \lambda_i \nabla g_i(x+d)
\end{aligned}$$

[3] Those so inclined are free to chase $\epsilon - \delta$'s

From the these equation we can see that $x + d$ satisfy the requirement to be an extrema. We can also see that a change of D in g_i created a change of $\lambda_j D$ in the value at the extrema, hence giving a rate of change of λ_j.

To-Do: Add and example of minimizing height when the two contrasts are parabolic and linear. (You will need to use logs to get the change for the parabola to be a change in it's width and not height.)

0.6 Isosets

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ let the isosets[4] be sets S in the domain of f such that $f(S)$ is constant.

Given some isoset S and point $x \in \mathbb{R}^n$ of f we want some kind of function that returns some type of measure $d \in \mathbb{R}$ of the distance of x to S. This function will be used in fragment shader rendering, which is why the mission statement is so vague. We only need some general measure since it's better to efficiently get that measure and tweak coefficients then to get something perfectly accurate.

A particular application is the $n = 2$ case where we are looking to find the distance for the contour line.

0.6.1 Naïve Solution

The first idea is to compare the distance of $f(x)$ to $f(S)$ to d:

$$d = |f(x) - f(S)|$$

The problem with this solution is that the measure changes as a function of $|\nabla f(x)|$. That is to say that the faster f changes at x the closer x has to be to S to get the same value of d.

This method might work for some shader effects but not others. For example it won't work to draw a constant width contour as the width would be inversely proportional to $|\nabla f(x)|$

0.6.2 Better Solution

If the underestimation is proportional to $|\nabla f(x)|$ the obvious solution is to divide by $|\nabla f(x)|$:

$$d = \frac{|f(x) - f(S)|}{|\nabla f(x)|}$$

This is the solution currently used but deserves more analysis.

[4]Not the proper name, but I can't recall the proper name right now.

For starters consider the case that x is near the point $s \in S$. Then $\nabla f(s)$ points away from S, that is that if a tangent to S exists at s then $\nabla f(s)$ is at orthogonal to the tangent, by definition of S being a set such that f is constant. The combination of orthogonality and closeness lets us recover the original expression by use of the tangent surface:

$$\begin{aligned}
f(x) =& f(s) + (x-s) \cdot \nabla f(s) \\
|f(x) - f(s)| =& |(x-s) \cdot \nabla f(s)| \\
=& |x-s||\nabla f(s)| \\
\frac{|f(x) - f(s)|}{|\nabla f(s)|} =& |x-s|
\end{aligned}$$

Now consider the case where x is not close to S and the use case of drawing a constant width contour line. Under what conditions do we avoid a false positive? (That is the function thinks x is closer than it is.) Well we need some constraints on the rate at which $\nabla f(x)$ can grow. To see this consider a point far away from S but whose rate of change is very slow between most of x and s, so that $|f(x) - f(s)|$ is small, but suddenly increases at x, such that $|\nabla f(x)|$ is large. This causes their ratio to be small despite $|x - s|$ being large, false saying x should be colored as part of the contour line. If we reverse the set up, rate of change is large at first then slow, we will get a false negative. (That the point is further than we think it is)

To see how a constraint would be useful, consider the following one:

$$|\nabla f(x+d)| \leq k|d||\nabla f(x)|$$

$$
\begin{aligned}
|f(x) - f(s)| &= \left| \int_0^1 \nabla f(x + (s-x)t) \cdot (s-x)\, dt \right| \\
&= \left| (s-x) \cdot \int_0^1 \nabla f(x + (s-x)t)\, dt \right| \\
&\leq |(s-x)| \left| \int_0^1 \nabla f(x + (s-x)t)\, dt \right| \\
&\quad \text{Cauchy-Schwartz} \\
&\leq |(s-x)| \int_0^1 |\nabla f(x + (s-x)t)| \; dt \\
&\quad \text{ML Bound} \\
&\leq |(s-x)| \int_0^1 k|(s-x)t||\nabla f(x)|\, dt \\
\frac{|f(x) - f(s)|}{|\nabla f(x)|} &\leq \frac{k}{2}|s-x|^2
\end{aligned}
$$

This avoids a false negative as x must be at least $\sqrt{\frac{2d}{k}}$ away.

To-do: we need a inequities like $d \geq p(|x-s|)$ to get a bound on false positives.

The condition:

$$|\nabla f(x+d)| \leq k|d| + |\nabla f(x)|$$

Gives:

$$d \leq |x-s| \left(1 + \frac{|x-s|}{|\nabla f(x)|} \right)$$

0.7 Padé Approximant

We wish to approximate a function f by creating a rational function that agrees with f's first N derivatives at zero.

Let T_N be the Nth degree Maclaurin series of f. Consider the steps of finding the polynomial greatest common division by the extended Euclid algorithm of T_N with x^{N+1} where we prematurely stop:

$$\begin{aligned}
x^{N+1} =& 1 \cdot x^{N+1} + & 0 \cdot T_N(x) \\
T_N(x) =& 0 \cdot x^{N+1} + & 1 \cdot T_N(x) \\
r_1(x) =& 1 \cdot x^{N+1} + & -q_1(x) \cdot T_N(x) \\
& \vdots & \\
P(x) =& K(x) x^{N+1} + & Q(x) T_N(x)
\end{aligned}$$

By inspection we get the useful relation:

$$P(x)/Q(x) \equiv T_N(x) \mod x^{N+1}$$

Hence satisfying the original objective. Note that we can chose decrease the degree of P by simply continuing the algorithm.

0.7.1 The Reverse

Say I want to do the reverse, that I have $f = g/h$ where I have the power series for g and h but want to effectively find f We have:

$$\begin{aligned}
x^{N+1} =& 0 \cdot f(x) + & 1 \cdot x^{N+1} \\
g(x) =& h(x) \cdot f(x) + & 0 \cdot x^{N+1}
\end{aligned}$$

You can remove some higher term of f into a residue function K on x^{N+1}, since we don't really care about it.

0.7.2 Differential

What if g and h are related buy a differential equation? We can use it like how we used the quotient-remainder equation. The derivative is linear after all.

0.7.3 Chinese Remainder Theorem

Since I have modulo relations can I combine them with the Chinese Remainder Theorem? The moduli will have to be pairwise coprime $(x - a_i)^n$ stand out.

0.8 p-adic numbers

0.8.1 Valuation

A function v of a field is called a valuation if:

$$\begin{aligned} v(x) =& \infty \quad \text{iff } x = 0 \\ v(xy) =& v(x) + v(y) \\ v(x+y) \geq& \min(v(x), v(y)) \quad \text{with equality if } v(x) \neq v(y) \end{aligned}$$

(Note that the codomain is only required to be an abelian totally ordered group extended with ∞, But I will treat is as the natural numbers.)

There are three immediate corollaries of the definitions.
1:By induction on the inequality we have:

$$v\left(\sum_k x_k\right) \geq \min\left(\bigcup_k \{v(x_k)\}\right)$$

2: By setting $x = y = 1$ in the second equality we have $v(1) = 2v(1)$ and hence $v(1) = 0$.
3: By setting $xy = 1$ in the second equality we have $v(1/x) = -v(x)$.

0.8.2 p-adic numbers

Interestingly, the amount of times a prime p divides a rational number x is a valuation. Let $x = p^n \frac{a}{b}$ where a and b are coprime, then $v_p(x) = n$ is a valuation. (Assuming you set $v_p(0) = \infty$). This is easy to prove, if you need help remember that for a prime p we have: $p|ab \Rightarrow p|a$ or $p|b$.

The reason this is interesting is because $|x|_p = p^{-v_p(x)}$ is a metric on $\mathbb{Q}$. Meaning we can make a new field $\mathbb{Q}_p$ by taking all the limits in $\mathbb{Q}$ as we would nomarlly do to make $\mathbb{R}$ with $|\cdot|$. And through something called Ostrowski's theorem becomes a lot more motivated and less arbitrary way to complete $\mathbb{Q}$.

But the valuation alone also provides two cool proofs that simplify previous proofs.

0.8.3 Irrationality of $\sqrt{2}$

Consider the equation:

$$x^2 = 2$$

Valuating both sides gives:

$$2v_2(x) = 1$$

But $v_2(x)$ is an integer for all rational x hence the LHS is always even but the right is odd. Hence there is no x satisfying the equation and $\sqrt{2}$ is irrational.

0.8.4 Valuation of the Harmonic Numbers

The valuation of harmonic numbers is given by $v_2(H_n) = -\lfloor \log_2(n) \rfloor$.

Lemma: Let k be the power of the largest power of 2 less than n, i.e. $k = \lfloor \log_2(n) \rfloor$. Let S be the set $[1, n]$ excluding 2^k, then from the maximality of k we have:

$$\max(v_2(S)) \leq k - 1$$

Since if we assume there is an $s \in S$ such that $v_2(s) > k-1$ with the codomain of v_2 being integers means $v_2(s) \geq k$. This means there exists an integers a and b coprime to each other and 2 such that $s = 2^k \frac{a}{b}$. s being a positive integer means $b = 1$. $a \neq 1$ since it would make $s = 2^k$, which was excluded from S. But $a \geq 2$ would means $2^{k+1} \in S$, contradicting the maximality of k.

Now H_n can be written as the following sum:

$$H_n = \sum_{s \in S} \frac{1}{s} + 2^{-k}$$

Where the valuation of first term is bound by:

$$\begin{aligned} v_2\left(\sum_{s \in S} \frac{1}{s}\right) &\geq \min\{v_2(1/s) \mid s \in S\} \\ &= \min\{-v_2(s) \mid s \in S\} \\ &= -\max\{v_2(s) \mid s \in S\} \\ &= -k + 1 \end{aligned}$$

Hence the valuations are not equal since:

$$v_2(2^{-k}) = -k < -k + 1 \le v_2\left(\sum_{s\in S} \frac{1}{s}\right)$$

Hence

$$v_2(H_n) = \min\left\{v_2\left(\sum_{s\in S} \frac{1}{s}\right), v_2(2^{-k})\right\} = -k$$

As required.

Note that for $n \ge 2$ we have $k \ge 1$ and hence $v_2(H_n) \le -1$ meaning H_n isn't an integer for $n \ne 1$

0.8.5 Valuation of the Harmonic Numbers v2

I think the lemma's of the previous proof can be separated and cleaned up.

Lemma: Let S be a subset of the v's domain with an element $s_0 \in S$ such that:

$$v(S\backslash\{s_0\}) > v(s_0)$$

Then:

$$v\left(\sum_{s\in S} s\right) = v(s_0)$$

Proof: Plugging the inequality into the valuation inequality axiom gives:

$$v\left(\sum_{s\in S\backslash\{s_0\}} s\right) \ge v(S\backslash\{s_0\}) > v(s_0)$$

Hence the term of the far left and far right are not equal meaning:

$$\begin{aligned} v\left(\sum_{s\in S} s\right) &= v\left(s_0 + \sum_{s\in S\backslash\{s_0\}} s\right) \\ &= \min\left\{v(s_0), v\left(\sum_{s\in S\backslash\{s_0\}} s\right)\right\} \\ &= v(s_0) \quad \square \end{aligned}$$

Lemma: If $n \in \mathbb{N}$ then $v_p(n) \leq \log_p(n)$ with equality iff n is a power of p.

Proof: Equality in the case of n being a power is trivial. Now consider n not a power of 2 meaning $n = a\, p^{v_p(n)}$ where $a > 1$, hence:

$$p^{v_p(n)} < n = p^{\log_p(n)}$$

$\log_p$ is strictly monotonic, hence:

$$v_p(n) < \log_p(n)$$

Lemma: If $n \in \mathbb{N}$ then $v_2(n) \leq \lfloor \log_2(n) \rfloor$ with equality iff n is a power of 2.

Proof: Equality in the case of n being a power is trivial. Assume n isn't a power of 2 then:

$$n \geq 3 \cdot 2^{v_2(n)}$$

Hence:

$$\begin{aligned} \log_2(n) \geq& \log_2(3) + v_2(n) \\ \log_2(n) - \log_2(3) \geq& v_2(n) \\ \lfloor \log_2(n) - \log_2(3) \rfloor \geq& \lfloor v_2(n) \rfloor \end{aligned}$$

Since $\log_2(3) > 1$ we have:

$$\lfloor \log_2(n) \rfloor > \lfloor \log_2(n) - \log_2(3) \rfloor$$

And $v_2(n)$ is an integer, hence: $\lfloor \log_2(n) \rfloor > v_2(n) \quad \square$

Note that this can't be generalized to larger p since the logarithm being less than one isn't guaranteed. Compare with $p = 3$ with $v_3(6) = 1 = \lfloor \log_3(6) \rfloor$ because $\log_3(2) < 1$. Note to the previous note, in this case you can say $v_3(n) \leq \lfloor \log_p(n) \rfloor$ iff n is a power of 3 or even, which is still *kind of* cool.

Lemma: Let $S_n = \{k^{-1} | 1 \leq k \leq n\}$ and $k_0 = \lfloor \log_2 n \rfloor$. Then $2^{-k_0} \in S$ and satisfies:

$$v_2(S_n \backslash \{2^{-k_0}\}) > v_2(2^{-k_0})$$

Proof: $2^{k_0} = 2^{\lfloor \log_2 n \rfloor} \leq 2^{\log_2 n} = n$ hence $2^{-k_0} \in S$. Now consider $s \in S$ then $s^{-1} \in \mathbb{N}$ which from the previous lemma means:

$$v_2(s^{-1}) \leq \lfloor \log_2(s^{-1}) \rfloor \leq \lfloor \log_2(n) \rfloor = k_0$$

With equality iff s^{-1} is a power of 2. It's can't be a larger power than 2^{k_0} since $2^{\lfloor \log_2(n) \rfloor}$ is the largest power less than or equal to n. But from the definition of $S_n \backslash \{2^{-k_0}\}$ it can't be the same power. Hence s^{-1} can only be a lower power making the final inequality strict anyway. Hence:

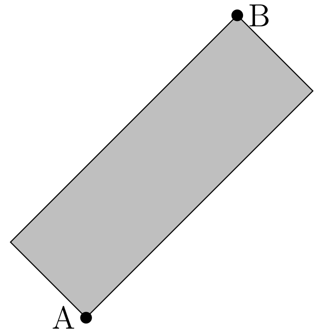
$$v_2(s^{-1}) < k_0 \Rightarrow v_2(s) > -k_0 = v_2(2^{-k_0}) \quad \square$$

Noting that $H_n = \sum_{s \in S_n} s$ means the proof that $v_2(H_n) = -\lfloor \log_2(n) \rfloor$ follows immediately from the first and last lemmas.

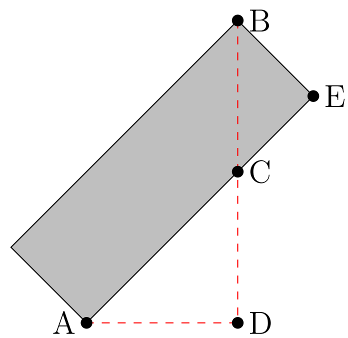
0.9 Causal Metric

0.9.1 Basic Geometry

Consider two points A and B that we wish to find the area of the rectangle between them in terms of their coordinates on an axis at a 45° angle.



Drop an altitude from B to line up horizontally with A and label the end-point D. The length |AD| and |BD| are the coordinates.



The triangles $\triangle$ACD and $\triangle$BCE are $45° - 90° - 45°$ triangles meaning:

$$
\begin{aligned}
|\mathrm{AC}| =& \sqrt{2}|\mathrm{AD}| \\
|\mathrm{BE}| =& \frac{1}{\sqrt{2}}|\mathrm{BC}| \\
=& \frac{1}{\sqrt{2}}(|\mathrm{BD}| - |\mathrm{CD}|) \\
=& \frac{1}{\sqrt{2}}(|\mathrm{BD}| - |\mathrm{AD}|)
\end{aligned}
$$

Hence the (signed) area of the rectangle is:

$$\begin{aligned}|\mathrm{BE}| \cdot |\mathrm{AE}| =& |\mathrm{BE}| \cdot (|\mathrm{AC}| + |\mathrm{CE}|)\\ =&(|\mathrm{BD}| - |\mathrm{AD}|)(|\mathrm{BD}| + |\mathrm{AD}|)\\ =&|\mathrm{BD}|^2 - |\mathrm{AD}|^2\end{aligned}$$

0.9.2 Causal Metric

This construction supplies some intuition for Minkowski Metric. Since if we interpret the plane as the set of events where the horizontal component is space-like and the vertical is time-like. The reason the rectangle is at a 45° is because that's the maximum speed of propagation. The rectangle's area is a measure of the amount of events in-between the two. And the sign of the area is the type of causal connection. Wether the points are in the way (space-like), or another events are a means by which the earlier effect the later (time-like).

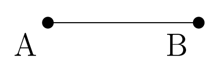
0.10 Quick Summary of Spaces

A space is a collection of elements, often called points, with some additional structure. There are four main types of spaces that form a nice hierarchy, that is that some types of spaces are always a subtype of other space.

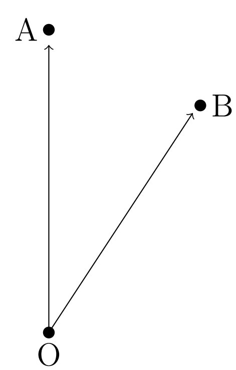
0.10.1 Hierarchy

Topological Space: Neighbourhood In a topological space the elements have a concept of "Neighbourhood", that is the points which are adjacent to each point.

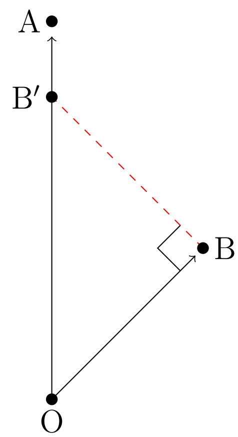
Metric Space: Distance In a metric space the element have a concept of distance to each other. You can induce a neighbourhood by saying all points within a certain threshold distance to a point are in the neighbourhood of that point.



Normed Space: Length In a normed space there is a concept of length of a point. Shown here as their distance to some origin. A distance can be induced by getting the length of an arrow going from A to B.



Inner-product Space: Projection An inner product space has a concept of projecting one element on another. You can induce a length by projecting an object onto itself.



0.10.2 Euclidean

Euclidean space is a normed space on $\mathbb{R}^n$ where the inner product is given by $\sum_k A_k B_k$. This naturally induces a metric on $\mathbb{R}^n$.

The reason we care about the metric of Euclidean space is because all normed spaces have a form of the Pythagorean theorem. Making metric space the most endowed space we can easily start thinking about non-Euclidean space.

This sucks because most talk about non-Euclidean space talks about "parallel lines" which makes one naturally think about angles and hence an inner product. But no, instead we have the concept of a geodesic. A geodesic is a curve that is locally distance minimizing. This locality comes from the topology, and means that for all points $\gamma(t_1)$ and $\gamma(t_0)$ on the curve γ such that they are in the same neighbourhood satisfy:

$$d(\gamma(t_1), \gamma(t_0)) = k|t_1 - t_2|$$

0.10.3 non-Euclidean

Quick rundown of some attempts of non-Euclidean geometries:

Pseudo-Euclidean: Like how we defined Euclidean Space was defined with an inner product we define a new structure with a different quadratic form.

Riemann Manifold: A Manifold is a space that is "locally Euclidean". And a Riemann Manifold is one where this inner-product in the local Euclidean space is always positive.

Pseudo-Riemann Manifold: Like A Riemann Manifold but the inner-product is only required to be non-degenerate. I think this includes Pseudo-Euclidean space, but haven't put much thought into it.

0.11 Wave Equation

0.11.1 Symmetry

Given some function $u(\vec{x}, t)$ on space and time we want to understand it's dynamics under the following assumptions:

1. **Reversible**, the dynamics should be symmetric in respect to reversing time.
2. **Isotropic**, the function should be symmetric in respect to all directions.
3. **Relative**, the absolute value doesn't matter, only changes.
4. **Perturbation**, the changes should be small.

The wave equation is a natural conclusion from from these constraints:

$$\frac{\partial^2}{\partial t^2}u = c\sum_i \frac{\partial^2}{\partial x_i^2}u$$

Because the change change is small we start by considering a Taylor expansion and try to get the lowest order terms. Because it's relative we ignore the 0^{th} order terms. Because it's reversible we ignore the 1^{st} order terms Giving:

$$\frac{\partial^2}{\partial t^2}u = \sum_{i,j} w_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}u$$

Because it's isotropic all the $\frac{\partial^2}{\partial x_i^2}$ coefficient need to be equal. And without loss of generality we can assume a basis where the cross terms vanish, giving:

$$\frac{\partial^2}{\partial t^2}u = \sum_i cu\frac{\partial^2}{\partial x_i^2} = c\sum_i \frac{\partial^2}{\partial x_i^2}u \quad \square$$

0.11.2 Old Symmetry Attempt

Lets start with with a function of space $u(\vec{x})$. Lets assume it is continuous such that:

$$u(\vec{x}+\vec{r}) = u(\vec{x}) + \sum_i r_i \frac{\partial}{\partial x_i}u(\vec{x}) + \frac{1}{2}\sum_{i,j} r_i r_j \frac{\partial^2}{\partial x_i \partial x_j}u(\vec{x})$$

Lets remove all terms that aren't reversible:

$$u(\vec{x}+\vec{r}) = u(\vec{x}) + \frac{1}{2}\sum_i r_i^2 \frac{\partial^2}{\partial x_i^2} u(\vec{x})$$

Lets impose localization:

$$c^2 r_0^2 = \sum_{i>0} r_i^2$$

Not sure exactly how this part works but we get the sigh we need one example is to only keep terms that have $c^2 r_0^2 = r_i^2$

0.11.3 1-D

I need to remind myself how to solve the 1-D case as a stepping stone for something else. Reminder that the 1-D form, with unitary speed, is:

$$\frac{\partial^2}{\partial t^2} u = \frac{\partial^2}{\partial x^2} u$$

Observe that plane waves where angular frequency and wave number have the same magnitude solve this:

$$\exp(ik(x \pm t))$$

This suggests the use of Fourier transform, where the $\pm$ is used to fit the solution to initial value and derivative:

$$u = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (a_+(k)e^{ikt} + a_-(k) - e^{-ikt})e^{ikx}\, dk$$

Assume $f(x) = u(x,0)$ and $h(x) = \left[\frac{\partial}{\partial t} u\right](x,0)$. Equating Fourier transforms gives:

$$\begin{aligned} \hat{f}(k) =& a_+(k) + a_-(k) \\ \hat{h}(k) =& ik(a_+(k) - a_-(k)) \end{aligned}$$

Hence:

$$u = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\hat{f}(k)\cos(kt) + \hat{h}(k)k^{-1}\sin(kt))e^{ikx}\, dk$$