Total: 100/100

ISyE 8803 Homework 1

Parisa Babolhavaeji

February 13, 2025

1 MGF method vs. moment bounds

It is natural to compare the best bound on the tails obtained via MGF and by bounding the moments. As it turns out, the moment bounds are sharper, even if we only use the integer moments.

(a) Show that if $X > 0$ a.s., then for any $u > 0$,

$$\inf_{\lambda>0} M_X(\lambda)e^{-\lambda u} \geq \inf_{k\in\mathbb{Z}_+} \frac{\mathbb{E}[X^k]}{u^k}.$$

Answer:

$$M_X(\lambda) = \mathbb{E}[e^{\lambda X}] = \sum_{k=0}^{\infty} \frac{\lambda^k \mathbb{E}[X^k]}{k!}. \tag{1}$$

Multiplying by $e^{-\lambda u}$:

$$M_X(\lambda)e^{-\lambda u} = \sum_{k=0}^{\infty} \frac{\lambda^k \mathbb{E}[X^k]}{k!} e^{-\lambda u}. \tag{2}$$

Let $t = \lambda u$. Substituting:

$\exists \lambda = \frac{t}{u}$

$$M_X(\lambda)e^{-\lambda u} = \sum_{k=0}^{\infty} \left(\frac{\mathbb{E}[X^k]}{u^k}\right) \cdot \frac{t^k e^{-t}}{k!}, \tag{3}$$

where $t = \lambda u > 0$. The term $\frac{t^k e^{-t}}{k!}$ corresponds to the Poisson probability mass function with parameter t, which sums to 1 over k.

The expression $M_X(\lambda)e^{-\lambda u}$ is a convex combination of $\frac{\mathbb{E}[X^k]}{u^k}$ (denoted $B(k)$) with Poisson weights. Specifically:

$$M_X(\lambda)e^{-\lambda u} = \sum_{k=0}^{\infty} B(k) \cdot \text{Poisson}(k;t), \tag{4}$$

where $\text{Poisson}(k;t) = \frac{t^k e^{-t}}{k!}$.

Since $M_X(\lambda)e^{-\lambda u}$ is an average of $B(k)$, it must satisfy:

$$M_X(\lambda)e^{-\lambda u} \geq \inf_{k\in\mathbb{Z}_+} B(k) \quad \text{for all } \lambda > 0. \tag{5}$$

(The weighted average of $B(k)$ can't be less than the smallest term in the sequence)

Taking the infimum over $\lambda > 0$ preserves the inequality:

$$\inf_{\lambda>0} M_X(\lambda)e^{-\lambda u} \geq \inf_{k\in\mathbb{Z}_+} \frac{\mathbb{E}[X^k]}{u^k}. \tag{6}$$

$\boxed{\text{Well done!}}$ (Consider pursuing career in theory)

(b) Show that if X is symmetric (i.e., X and $-X$ have the same distribution), then for any $u > 0$,

$$\inf_{\lambda>0} M_X(\lambda)e^{-\lambda u} \geq \frac{1}{2} \inf_{k\in\mathbb{Z}_+} \frac{\mathbb{E}[X^{2k}]}{u^{2k}}.$$

Answer:

Since X is symmetric, all its odd moments vanish, and the MGF has only even terms:

$$M_X(\lambda) \;=\; \mathbb{E}\big[e^{\lambda X}\big] \;=\; \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}\,\mathbb{E}[X^j] \;=\; \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!}\,\mathbb{E}[X^{2k}].$$

Hence for every fixed k,

$$M_X(\lambda) \;\geq\; \frac{\lambda^{2k}}{(2k)!}\,\mathbb{E}[X^{2k}], \quad \text{so} \quad M_X(\lambda)\,e^{-\lambda u} \;\geq\; \frac{\lambda^{2k}}{(2k)!}\,\mathbb{E}[X^{2k}]\,e^{-\lambda u}.$$

$$\frac{\mathbb{E}[X^{2k}]}{u^{2k}} \;\leq\; \frac{(2k)!}{(\lambda u)^{2k}}\,M_X(\lambda)\,e^{-\lambda u}.$$

Taking $\inf_{\lambda>0}$ on the right side:

$$\inf_{\lambda>0}\Big\{M_X(\lambda)\,e^{-\lambda u}\Big\} \;\geq\; \inf_{k\in\mathbb{Z}_+} \frac{\mathbb{E}[X^{2k}]}{u^{2k}} \times \Big(\text{constant factor}\Big)^{-1}.$$

this wouldn't do - but your previous method still works, if properly adjusted.

By symmetry,

$$\mathbb{P}(X \geq u) \;=\; \tfrac{1}{2}\,\mathbb{P}(|X| \geq u) \;\leq\; \tfrac{1}{2}\,\frac{\mathbb{E}[X^{2k}]}{u^{2k}}.$$

Thus:

$$\boxed{\inf_{\lambda>0} M_X(\lambda)e^{-\lambda u} \geq \frac{1}{2} \inf_{k\in\mathbb{Z}_+} \frac{\mathbb{E}[X^{2k}]}{u^{2k}}.}$$

+/-

Can you guess how?

(A.k.a., what is the averaging distribution?)

2 Convexity of the Cumulant-Generating Function

For any distribution X, the logarithm of the MGF

$$K_X(t) = \log \mathbb{E}[e^{tX}]$$

is called the cumulant-generating function, or the log-partition function of the distribution.

(a) Show that K_X is convex. Use Young's inequality: for $a, b \in \mathbb{R}^d$ and $p, q \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$$|a^\top b| \leq \|a\|_p \|b\|_q.$$

You can assume that X has a discrete distribution.

Answer:

Because

$$K_X(t) \;=\; \log \mathbb{E}\big[e^{tX}\big],$$

the required inequality is equivalent to

$$\mathbb{E}\Big[e^{(\theta t_1 + (1-\theta)\, t_2)\, X}\Big] \;\leq\; \Big(\mathbb{E}\big[e^{t_1 X}\big]\Big)^{\theta} \Big(\mathbb{E}\big[e^{t_2 X}\big]\Big)^{1-\theta}.$$

Assume X is discrete with $\Pr[X = x_j] = p_j$. Then the inequality becomes

$$\sum_j p_j \, \exp\Big((\theta t_1 + (1-\theta) t_2)\, x_j\Big) \;\leq\; \Big(\sum_j p_j \, e^{t_1 x_j}\Big)^{\theta} \Big(\sum_j p_j \, e^{t_2 x_j}\Big)^{1-\theta}.$$

Hölder's inequality:

$$e^{(\theta t_1 + (1-\theta) t_2)\, x_j} \;=\; e^{\theta t_1 x_j} \cdot e^{(1-\theta)\, t_2 x_j},$$

and choose exponents $p = \frac{1}{\theta}$, $q = \frac{1}{1-\theta}$ (so that $1/p + 1/q = 1$).

$$\sum_j p_j \, e^{(\theta t_1 + (1-\theta) t_2) x_j} \;\leq\; \Big(\sum_j p_j \, e^{t_1 x_j}\Big)^{\theta} \Big(\sum_j p_j \, e^{t_2 x_j}\Big)^{1-\theta}.$$

Taking the log:

$$K_X(\theta t_1 + (1-\theta) t_2) \;\leq\; \theta \, K_X(t_1) \;+\; (1-\theta)\, K_X(t_2),$$

which shows K_X is convex.

$+$

3 Gaussian Tails

3.1 Mills Ratio

Let $\phi(\cdot)$ be the p.d.f. of $\mathcal{N}(0,1)$, i.e., $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$. For any $u \geq 0$, let $\Phi(u) := \int_u^\infty \phi(t)dt$.

(a)

Prove the following bounds (holding for all $u > 0$):

$$\left(\frac{1}{u} - \frac{1}{u^3}\right)\phi(u) \leq \Phi(u) \leq \frac{1}{u}\phi(u).$$

Hint 1: Try to prove the upper bound first.
Hint 2: Integrate by parts – first to prove the upper bound, then again for the lower bound.

Answer:

Upper:

$$\Phi(u) \;=\; \int_u^\infty \phi(t)\,dt \;=\; \int_u^\infty \phi(t)\,\frac{u}{t}\,dt \;\leq\; \frac{1}{u}\int_u^\infty t\,\phi(t)\,dt \;=\; \frac{1}{u}\,\phi(u),$$

+

The last equality follows from the fact that $\int_u^\infty t\,\phi(t)\,dt = \phi(u)$.

Lower bound:

$$\Gamma(u) \;:=\; \Phi(u) \;-\; \left(\tfrac{1}{u} - \tfrac{1}{u^3}\right)\phi(u).$$

+

We show $\Gamma(u) \geq 0$. $\Phi'(u) = -\phi(u)$. Also let

$$g(u) \;=\; \left(\tfrac{1}{u} - \tfrac{1}{u^3}\right)\phi(u).$$

$$g'(u) \;=\; \phi(u)\left(\tfrac{3}{u^4} - 1\right), \quad \text{so} \quad \Gamma'(u) \;=\; \Phi'(u) - g'(u) \;=\; -\phi(u) \;-\; \phi(u)\left(\tfrac{3}{u^4} - 1\right) \;=\; -\tfrac{3}{u^4}\,\phi(u).$$

Since $\phi(u) > 0$ for $u > 0$, we have $\Gamma'(u) < 0$, so $\Gamma(u)$ is strictly decreasing in u. As $u \to \infty$, both $\Phi(u)$ and $\phi(u)$ vanish, so $\Gamma(u) \to 0$. Being a decreasing function with limit 0 at infinity forces $\Gamma(u) \geq 0$ for all $u > 0$, i.e.,

$$\Phi(u) \;\geq\; \left(\tfrac{1}{u} - \tfrac{1}{u^3}\right)\phi(u).$$

Combining both bounds proves the claim.

+

(b)

Capitalizing on the trick you have just figured out to get the lower bound from the upper bound, prove a new upper bound:

$$\Phi(u) \le \left(\frac{1}{u} - \frac{1}{u^3} + \frac{3}{u^5}\right)\phi(u).$$

Note that this bound is sharper than the previous one for large enough u.

Answer:

Let

$$\Gamma(u) \;:=\; \left(\tfrac{1}{u} - \tfrac{1}{u^3} + \tfrac{3}{u^5}\right)\phi(u) \;-\; \Phi(u).$$

We have $\Phi'(u) = -\phi(u)$. Set $h(u) = \frac{1}{u} - \frac{1}{u^3} + \frac{3}{u^5}$. Then

$$\Gamma'(u) = \left[h'(u)\phi(u) + h(u)\phi'(u)\right] - (-\phi(u)) = \phi(u)\left[h'(u) - u\,h(u) + 1\right].$$

$$h'(u) = -\tfrac{1}{u^2} + \tfrac{3}{u^4} - \tfrac{15}{u^6}, \quad u\,h(u) = 1 - \tfrac{1}{u^2} + \tfrac{3}{u^4} \implies h'(u) - u\,h(u) + 1 = -\tfrac{15}{u^6}.$$

+

Hence $\Gamma'(u) = -\frac{15}{u^6}\,\phi(u) < 0$, so Γ is strictly decreasing and $\Gamma(u) \to 0$ as $u \to \infty$. Therefore $\Gamma(u) \ge 0$ for all $u > 0$, i.e., $\Phi(u) \le \left(\frac{1}{u} - \frac{1}{u^3} + \frac{3}{u^5}\right)\phi(u)$.

Good Job!

+

3.2 Power Series for C.D.F.

Show that

$$\frac{1}{2} - \Phi(u) = \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k u^{2k+1}}{2^k k!(2k+1)}.$$

Hint: Change variable to remove u from the integration limits; differentiate in u under the integral.

Answer:

$$\tfrac{1}{2} - \Phi(u) \;=\; \int_0^u \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \quad \Big(\text{since} \int_0^{\infty} \phi(t)\, dt = \tfrac{1}{2}\Big).$$

Expanding the integrand in a power series:

$$e^{-t^2/2} \;=\; \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \left(\tfrac{t^2}{2}\right)^k \;=\; \sum_{k=0}^{\infty} \frac{(-1)^k\, t^{2k}}{2^k\, k!}.$$

Thus

$$\int_0^u e^{-t^2/2}\, dt \;=\; \int_0^u \sum_{k=0}^{\infty} \frac{(-1)^k\, t^{2k}}{2^k\, k!}\, dt \;=\; \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k\, k!} \int_0^u t^{2k}\, dt.$$

Integral yields $\int_0^u t^{2k}\, dt = \frac{u^{2k+1}}{2k+1}$. So

$$\int_0^u e^{-t^2/2}\, dt \;=\; \sum_{k=0}^{\infty} \frac{(-1)^k\, u^{2k+1}}{2^k\, k!\, (2k+1)}.$$

Factor out $\frac{1}{\sqrt{2\pi}}$:

$$\tfrac{1}{2} - \Phi(u) \;=\; \int_0^u \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \;=\; \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k\, u^{2k+1}}{2^k\, k!\, (2k+1)},$$

Well done! +

4 Paley-Zygmund and friends

(i) Prove the Paley-Zygmund inequality (it can be interpreted as a counterpart of Markov: a nonnegative random variable cannot be much smaller than its expectation):

If X is a non-negative random variable with $\mathbb{E}[X^2] < \infty$, then for any $t \in [0,1]$ one has

$$\mathbb{P}(X \geq (1-t)\mathbb{E}X) \geq t^2 \frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]}. \tag{1}$$

(ii) Under the same assumptions, strengthen to Cantelli's inequality:

$$\mathbb{P}(X \geq (1-t)\mathbb{E}X) \geq \frac{t^2(\mathbb{E}X)^2}{t^2(\mathbb{E}X)^2 + \mathrm{Var}[X]}.$$

This new inequality is sharp – give an example where it is attained.

(iii) Now: instead of $\mathbb{E}[X^2] < \infty$, assume that $\mathbb{E}[|X|^p] < \infty$ for some $p > 1$, and **generalize** (**??**) to

$$\mathbb{P}(X \geq (1-t)\mathbb{E}X) \geq \left(t^p \frac{(\mathbb{E}X)^p}{\mathbb{E}[|X|^p]} \right)^{\frac{1}{p-1}}.$$

Note that when $p > 2$, this gives an improvement over (**??**) *for small t, which is important in applications where X is itself the sample average of i.i.d. $Y_1, \ldots, Y_n$.*

Hint: Use Hölder's inequality: given $p, q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, and random variables U, V on the same sample space, one has $\mathbb{E}[|UV|] \leq (\mathbb{E}[|U|^p])^{1/p} \cdot (\mathbb{E}[|V|^q])^{1/q}$.

Answer:

i

Let $A = \{X \geq (1-t)\mathbb{E}[X]\}$. Split $\mathbb{E}[X]$:

$$\mathbb{E}[X] = \mathbb{E}[X\mathbf{1}_A] + \mathbb{E}[X\mathbf{1}_{A^c}].$$

Since $X \geq 0$, the second term satisfies:

$$\mathbb{E}[X\mathbf{1}_{A^c}] \leq (1-t)\mathbb{E}[X] \cdot \mathbb{P}(A^c).$$

Substituting into the first equation:

$$\mathbb{E}[X] \leq \mathbb{E}[X\mathbf{1}_A] + (1-t)\mathbb{E}[X] \cdot \mathbb{P}(A^c).$$

$$\mathbb{E}[X\mathbf{1}_A] \geq t\mathbb{E}[X].$$

Apply the Cauchy-Schwarz inequality to $\mathbb{E}[X\mathbf{1}_A]$:

$$\mathbb{E}[X\mathbf{1}_A] \leq \sqrt{\mathbb{E}[X^2]\mathbb{P}(A)}.$$

Combining the two inequalities:

$$t\mathbb{E}[X] \leq \sqrt{\mathbb{E}[X^2]\mathbb{P}(A)} \quad \Rightarrow \quad \mathbb{P}(A) \geq \frac{t^2(\mathbb{E}[X])^2}{\mathbb{E}[X^2]}.$$

Thus:

$$\mathbb{P}(X \geq (1-t)\mathbb{E}[X]) \geq \frac{t^2(\mathbb{E}[X])^2}{\mathbb{E}[X^2]}.$$

+

ii

Define $Y = X - (1-t)\mathbb{E}[X]$. Split $\mathbb{E}[Y\mathbf{1}_{Y\geq 0}]$:

$$\mathbb{E}[Y\mathbf{1}_{Y\geq 0}] = \mathbb{E}[X\mathbf{1}_A] - (1-t)\mathbb{E}[X]\mathbb{P}(A),$$

where $A = \{X \geq (1-t)\mathbb{E}[X]\}$. From part (i), $\mathbb{E}[X\mathbf{1}_A] \geq t\mathbb{E}[X]$, so:

$$\mathbb{E}[Y\mathbf{1}_{Y\geq 0}] \geq t\mathbb{E}[X] - (1-t)\mathbb{E}[X]\mathbb{P}(A).$$

Apply Cauchy-Schwarz to $\mathbb{E}[Y\mathbf{1}_{Y\geq 0}] \leq \sqrt{\mathbb{E}[Y^2]\mathbb{P}(A)}$:

$$t\mathbb{E}[X] - (1-t)\mathbb{E}[X]\mathbb{P}(A) \leq \sqrt{\mathbb{E}[Y^2]\mathbb{P}(A)}.$$

$$\mathbb{P}(A) \geq \frac{t^2(\mathbb{E}[X])^2}{t^2(\mathbb{E}[X])^2 + \mathrm{Var}(X)}.$$

Equality holds when X is a constant. If $X = c$ almost surely, then $\mathrm{Var}(X) = 0$, and: +

$$\mathbb{P}(X \geq (1-t)c) = 1 = \frac{t^2c^2}{t^2c^2 + 0}.$$

iii

Let $A = \{X \geq (1-t)\mathbb{E}[X]\}$. Split $\mathbb{E}[X]$ as in part (i) to get:

$$\mathbb{E}[X\mathbf{1}_A] \geq t\mathbb{E}[X].$$

Apply Hölder's inequality with $U = X, V = \mathbf{1}_A$, and conjugate exponents p, q:

$$\mathbb{E}[X\mathbf{1}_A] \leq (\mathbb{E}[|X|^p])^{1/p}(\mathbb{P}(A))^{1/q},$$

where $\frac{1}{p} + \frac{1}{q} = 1$. Substituting $\mathbb{E}[X\mathbf{1}_A] \geq t\mathbb{E}[X]$:

$$t\mathbb{E}[X] \leq (\mathbb{E}[|X|^p])^{1/p}(\mathbb{P}(A))^{1/q}.$$

Solve for $\mathbb{P}(A)$:

$$\mathbb{P}(A) \geq \left(\frac{t^p(\mathbb{E}[X])^p}{\mathbb{E}[|X|^p]}\right)^{\frac{1}{p-1}}.$$ +

5 Tail bound for χ_d^2

Let $X \sim \chi^2_{2d}$ (chi-squared distribution with $2d$ degrees of freedom), that is $X = \|Z\|^2 = Z_1^2 + \cdots + Z_{2d}^2$ where $Z \sim \mathcal{N}(0, I_d)$ (equivalently, $Z_i \sim \mathcal{N}(0,1)$ are i.i.d.). Define $M_{2d}(\cdot)$ as the MGF of $X \sim \chi^2_{2d}$,

$$M_{2d}(t) := \mathbb{E}[e^{tX}], \quad t \in \mathbb{R};$$

in particular, $M_2(t) = \mathbb{E}[e^{t(Z_1^2+Z_2^2)}]$. Our ultimate goal here is to prove that, with probability $\geq 1-\delta$,

$$X - 2d \leq \sqrt{Cd \log\left(\frac{1}{\delta}\right)} + c \log\left(\frac{1}{\delta}\right) \tag{2}$$

for some numerical constants $C, c > 0$. This bound is, in fact, optimal (see, e.g., [LM00, Lemma 1]).

(a) Derive the explicit form of $M_2(t)$:

$$M_2(t) = \frac{1}{1-2t}, \quad t < \frac{1}{2};$$

and $M_2 = +\infty$ for $t \geq \frac{1}{2}$. (To take the integral, pass to polar coordinates $(z_1, z_2) \mapsto (r, \theta)$ with $r = \sqrt{z_1^2 + z_2^2}$ and don't forget the Jacobian, which equals r.) Claim that, as a corollary,

$$M_{2d}(t) = \frac{1}{(1-2t)^d}, \quad t < \frac{1}{2}.$$

Answer:

$$M_2(t) \;=\; \mathbb{E}\big[e^{\,t(Z_1^2+Z_2^2)}\big].$$

Using polar coordinates (r,θ) with Jacobian r, the expectation becomes

$$M_2(t) \;=\; \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{(t-1/2)\,r^2}\, r\, dr\, d\theta.$$

- The angular integral $\int_0^{2\pi} d\theta = 2\pi$.
- The radial integral simplifies via the substitution $u = r^2$, so $du = 2r\,dr$.

$$M_2(t) \;=\; \int_0^{\infty} e^{(t-1/2)\,u}\, \frac{1}{2}\, du \;=\; \frac{1}{2} \int_0^{\infty} e^{(t-1/2)\,u}\, du.$$

This integral converges only if $t < \frac{1}{2}$, and in that case

$$\int_0^{\infty} e^{(t-1/2)u}\, du \;=\; \int_0^{\infty} e^{-\left(\frac{1}{2}-t\right)\,u}\, du \;=\; \frac{1}{\frac{1}{2}-t} \;=\; \frac{2}{1-2t}.$$

Therefore,

$$M_2(t) \;=\; \frac{1}{2} \cdot \frac{2}{1-2t} \;=\; \frac{1}{1-2t} \quad \text{for } t < \tfrac{1}{2}, \quad \text{and diverges } (+\infty) \text{ otherwise.}$$

For $X \sim \chi^2_{2d}$, X is the sum of $2d$ independent χ^2_1 random variables. The MGF of a single χ^2_1 is

$$M_{\chi^2_1}(t) = (1-2t)^{-1/2}, \quad t < \tfrac{1}{2}.$$

Since X is the sum of $2d$ such independent variables, the MGF is the product of their MGFs:

$$M_{2d}(t) \;=\; \Big[(1-2t)^{-1/2}\Big]^{2d} \;=\; \frac{1}{(1-2t)^d} \quad \text{for } t < \tfrac{1}{2}.$$

So:

$$M_2(t) \;=\; \frac{1}{1-2t} \quad \text{for } t < \tfrac{1}{2} \quad \text{and} \quad M_{2d}(t) \;=\; \frac{1}{(1-2t)^d} \quad \text{for } t < \tfrac{1}{2}.$$

(b) Using Chernoff's method, bound the tail function $\mathbb{P}(X > x)$, for any $x > 2d$, as follows:

$$\mathbb{P}(X > x) = \inf_{t<\frac{1}{2}} e^{-tx} M_{2d}(t) = \exp\left(d \log\left(\frac{x}{2d}\right) - \frac{x - 2d}{2}\right).$$

Hint: it is convenient to take the logarithm, and use that $u \mapsto \log(u)$ on $\mathbb{R}_+$ is increasing. Note that, in terms of the deviation $z = x - 2d > 0$, this is equivalent to

$$\mathbb{P}(X - 2d > z) = \exp\left(d \log\left(\frac{2d + z}{2d}\right) - \frac{z}{2}\right).$$

Answer:

For $t < \frac{1}{2}$, the tail probability is bounded by:

$$P(X > x) \leq \inf_{t<\frac{1}{2}} e^{-tx} M_{2d}(t),$$

where $M_{2d}(t) = \frac{1}{(1-2t)^d}$. Substituting $M_{2d}(t)$:

$$P(X > x) \leq \inf_{t<\frac{1}{2}} \frac{e^{-tx}}{(1 - 2t)^d}.$$

Define $f(t) = -tx - d\log(1 - 2t)$. Minimizing $e^{-tx}/(1 - 2t)^d$ is equivalent to minimizing $f(t)$. Set the derivative to zero:

$$f'(t) = -x + \frac{2d}{1 - 2t}.$$

$$\frac{2d}{1 - 2t^*} = x \implies t^* = \frac{1}{2}\left(1 - \frac{2d}{x}\right).$$

$$f(t^*) = -\frac{x}{2}\left(1 - \frac{2d}{x}\right) - d\log\left(\frac{2d}{x}\right).$$

$$f(t^*) = -\frac{x}{2} + d - d\log\left(\frac{2d}{x}\right) = d\log\left(\frac{x}{2d}\right) - \frac{x}{2} + d.$$

Let $x = 2d + z$. Substitute into $f(t^*)$:

$$f(t^*) = d\log\left(1 + \frac{z}{2d}\right) - \frac{z}{2}.$$

Thus, the tail bound becomes:

$$P(X > 2d + z) \leq \exp\left(d\log\left(1 + \frac{z}{2d}\right) - \frac{z}{2}\right).$$

4

(c) **Bonus.** *Bear with me, this part is a bit delicate – but we need it to reach the conclusion.*

(c.i) Show that

$$\mathbb{P}(X - 2d > z) \leq \begin{cases} \exp\left(-\frac{z^2}{16d}\right) & \text{for } 0 \leq z \leq 2d, \\ \exp\left(-\frac{z}{8}\right) & \text{for } z > 2d. \end{cases}$$

It is OK if you get some worse pair of constants $C > 16, c > 8$ leading to a weaker bound.
Hint: first show, using calculus, that

$$\log(1+u) \leq u - \frac{1}{4}\min\{u, u^2\}, \quad \forall u \geq 0.$$

Answer:

From (b) we know the bound of tail probability:

$$P(X - 2d > z) \leq \exp\left(d \log\left(1 + \frac{z}{2d}\right) - \frac{z}{2}\right).$$

Let $u = \frac{z}{2d} \geq 0$.

$$\log(1+u) \leq u - \frac{1}{4}\min\{u, u^2\}.$$

$$d \log(1+u) \leq d\left(u - \frac{1}{4}\min\{u, u^2\}\right).$$

$$P(X - 2d > z) \leq \exp\left(du - \frac{z}{2} - \frac{d}{4}\min\{u, u^2\}\right).$$

Substitute $u = \frac{z}{2d}$:

$$du = \frac{z}{2}, \quad \text{so} \quad du - \frac{z}{2} = 0.$$

+

The bound simplifies to:

$$P(X - 2d > z) \leq \exp\left(-\frac{d}{4}\min\left\{\frac{z}{2d}, \left(\frac{z}{2d}\right)^2\right\}\right).$$

Splitting based on cases;
Case 1: $0 \leq z \leq 2d$
Here, $\frac{z}{2d} \leq 1$, so $\min\left\{\frac{z}{2d}, \left(\frac{z}{2d}\right)^2\right\} = \left(\frac{z}{2d}\right)^2$.
Substitute into the bound:

$$P(X - 2d > z) \leq \exp\left(-\frac{d}{4}\frac{z^2}{4d^2}\right) = \exp\left(-\frac{z^2}{16d}\right).$$

Case 2: $z > 2d$
Here, $\frac{z}{2d} > 1$, so $\min\left\{\frac{z}{2d}, \left(\frac{z}{2d}\right)^2\right\} = \frac{z}{2d}$.
Substitute into the bound:

$$P(X - 2d > z) \leq \exp\left(-\frac{d}{4}\frac{z}{2d}\right) = \exp\left(-\frac{z}{8}\right).$$

So:

$$P(X - 2d > z) \leq \begin{cases} \exp\left(-\frac{z^2}{16d}\right) & \text{for } 0 \leq z \leq 2d, \\ \exp\left(-\frac{z}{8}\right) & \text{for } z > 2d. \end{cases}$$

+

(c.ii) *Reformulating the last bound as*

$$\mathbb{P}(X - 2d > z) \le \exp\left(-\min\left\{\frac{z^2}{16d}, \frac{z}{8}\right\}\right),$$

and letting $\mathbb{P}(X - 2d > z) = \delta$, *"invert" the last inequality to get (2) with* $C = 16$ *and* $c = 8$ *(or with some worse constants).*

Hint: $\max\{a, b\} \le a + b$ for $a, b \ge 0$.

Answer:

$$\mathbb{P}(X - 2d > z) \le \exp\left(-\min\left\{\frac{z^2}{16d}, \frac{z}{8}\right\}\right).$$

Set $\mathbb{P}(X - 2d > z) = \delta$:

$$\exp\left(-\min\left\{\frac{z^2}{16d}, \frac{z}{8}\right\}\right) \ge \delta.$$

$$-\min\left\{\frac{z^2}{16d}, \frac{z}{8}\right\} \ge \log(\delta).$$

$$\min\left\{\frac{z^2}{16d}, \frac{z}{8}\right\} \le -\log(\delta) = \log\left(\frac{1}{\delta}\right).$$

The inequality $\min\left\{\frac{z^2}{16d}, \frac{z}{8}\right\} \ge \log\left(\frac{1}{\delta}\right)$ must hold to ensure $\mathbb{P}(X - 2d > z) \le \delta$. So:

$$\frac{z^2}{16d} \ge \log\left(\frac{1}{\delta}\right) \quad \text{and} \quad \frac{z}{8} \ge \log\left(\frac{1}{\delta}\right).$$

Thus, z must satisfy:

$$z \ge 4\sqrt{d \log\left(\frac{1}{\delta}\right)} \quad \text{and} \quad z \ge 8 \log\left(\frac{1}{\delta}\right).$$

+

Use $\max\{a, b\} \le a + b$ for $a, b \ge 0$. Let:

$$a = 4\sqrt{d \log\left(\frac{1}{\delta}\right)}, \quad b = 8 \log\left(\frac{1}{\delta}\right).$$

Then:

$$\max\left\{4\sqrt{d \log\left(\frac{1}{\delta}\right)}, 8 \log\left(\frac{1}{\delta}\right)\right\} \le 4\sqrt{d \log\left(\frac{1}{\delta}\right)} + 8 \log\left(\frac{1}{\delta}\right).$$

Set:

$$z = \sqrt{16d \log\left(\frac{1}{\delta}\right)} + 8 \log\left(\frac{1}{\delta}\right).$$

Substituting z into $X - 2d \le z$, we get with probability $\ge 1 - \delta$:

$$X - 2d \le \sqrt{16d \log\left(\frac{1}{\delta}\right)} + 8 \log\left(\frac{1}{\delta}\right).$$

+

6 Stein's paradox

Consider the problem of estimating the mean μ in the multivariate Gaussian location family

$$P_\mu = \mathcal{N}(\mu, I_d), \quad \mu \in \mathbb{R}^d,$$

where I_d is the $d \times d$ identity matrix, from a single observation $X \sim P_\mu$. Note that here, X itself is the maximum likelihood estimator (MLE) for μ. Defining for any estimator $\hat{\mu} = \hat{\mu}(X)$ of μ the variance

$$\mathrm{Var}_\mu[\hat{\mu}] := \mathbb{E}_\mu[\|\hat{\mu} - \mathbb{E}[\hat{\mu}]\|^2]$$

and the quadratic risk

$$\mathrm{Risk}_\mu[\hat{\mu}] := \mathbb{E}_\mu[\|\hat{\mu} - \mu\|^2],$$

where $\|x\| := \left(\sum_{i=1}^d x_i^2\right)^{1/2}$ is the Euclidean norm of $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, we see that for any $\mu \in \mathbb{R}^d$,

$$\mathrm{Risk}_\mu[X] = \mathrm{Var}_\mu[X] = d.$$

Intuitively, one can suspect that no better estimator of X can be found: really, what can be done with only a single observation of the mean? Yet, this turns out to be false: one may improve over the MLE uniformly on the family (3) when $d > 2$. This celebrated result was established by James and Stein in 1976, and our goal is to reproduce it. But first, let us establish the terminology.

Definition 1. An estimator $\hat{\mu}$ is *dominated* by some other estimator $\hat{\mu}'$ if $\mathrm{Risk}_\mu[\hat{\mu}'] \le \mathrm{Risk}_\mu[\hat{\mu}]$ for any μ, and there exists a parameter value μ such that $\mathrm{Risk}_\mu[\hat{\mu}'] < \mathrm{Risk}_\mu[\hat{\mu}]$.

Definition 2. An estimator $\hat{\mu}$ is called *admissible* if it is not dominated by any other estimator. Otherwise, it is called *inadmissible*.

As statisticians, ideally we would like to compare two estimators over the whole family at once, without specifying a value of μ. Two admissible estimators cannot be compared this way, but at the very least we can rule out an inadmissible estimator, for it exists a uniformly better one.

You will show that the MLE is inadmissible when $d > 3$, by constructing a dominating estimator.
As statisticians, ideally we would like to compare two estimators over the whole family at once, without specifying a value of μ. Two admissible estimators cannot be compared this way, but at the very least we can rule out any *inadmissible* estimator, as for it there exists a uniformly better one.

(a) Consider *shrinkage estimators* $\hat{\mu} = sX$ with $s \in \mathbb{R}$, and compute their risks for any s. Show that one can restrict attention to $s \in [0, 1]$ (hence "shrinkage") by finding a dominating estimator for $\hat{\mu}$ with $s < 0$ or $s > 1$.

Answer:

$$\mathrm{Risk}_\mu[sX] = \mathbb{E}_\mu \left[\|sX - \mu\|^2 \right].$$

$$\|sX - \mu\|^2 = s^2\|X - \mu\|^2 + (1-s)^2\|\mu\|^2 + 2s(1-s)(X-\mu)^\top \mu.$$

$\mathbb{E}_\mu[X - \mu] = 0$– the cross-term vanishes:

$$\mathrm{Risk}_\mu[sX] = s^2 \cdot \mathbb{E}_\mu\left[\|X - \mu\|^2\right] + (1-s)^2\|\mu\|^2.$$

Since $X \sim N(\mu, I_d)$, $\|X - \mu\|^2$ follows a χ^2 distribution with $\mathbb{E}_\mu[\|X - \mu\|^2] = d$. Thus:

$$\mathrm{Risk}_\mu[sX] = s^2 d + (1-s)^2\|\mu\|^2.$$

For $s > 1$, compare sX with the MLE ($s = 1$):

$$\mathrm{Risk}_\mu[sX] - \mathrm{Risk}_\mu[X] = \left(s^2 d + (1-s)^2\|\mu\|^2\right) - d = (s^2 - 1)d + (1-s)^2\|\mu\|^2.$$

Since $s > 1$, both $(s^2 - 1) > 0$ and $(1-s)^2\|\mu\|^2 \geq 0$. Hence:

$$\mathrm{Risk}_\mu[sX] > \mathrm{Risk}_\mu[X] \quad \text{for all } \mu \in \mathbb{R}^d.$$

Thus, the MLE ($s = 1$) dominates any estimator with $s > 1$.
For $s < 0$, compare sX with the zero estimator ($s = 0$):

$$\mathrm{Risk}_\mu[sX] - \mathrm{Risk}_\mu[0] = \left(s^2 d + (1-s)^2\|\mu\|^2\right) - \|\mu\|^2 = s^2 d + s(s-2)\|\mu\|^2.$$

Since $s < 0$, $s(s-2) > 0$ (as s and $s - 2$ are both negative). Thus:

$$\mathrm{Risk}_\mu[sX] > \mathrm{Risk}_\mu[0] \quad \text{for all } \mu \in \mathbb{R}^d.$$

The zero estimator ($s = 0$) dominates any estimator with $s < 0$.
For $s \notin [0, 1]$, either $s = 1$ (for $s > 1$) or $s = 0$ (for $s < 0$) provides a uniformly lower risk. Therefore, attention can be restricted to $s \in [0, 1]$.

(b) Show that, for given μ, the best value of s—i.e., the one minimizing the risk—is given by

$$s^* = \frac{\|\mu\|^2}{d + \|\mu\|^2} = 1 - \frac{d}{d + \|\mu\|^2}.$$

$$\text{Risk}_\mu[sX] = s^2 d + (1-s)^2 \|\mu\|^2.$$

Answer:

$$\text{Risk}_\mu[sX] = s^2 d + \|\mu\|^2 - 2s\|\mu\|^2 + s^2\|\mu\|^2.$$

$$\text{Risk}_\mu[sX] = s^2\big(d + \|\mu\|^2\big) - 2s\|\mu\|^2 + \|\mu\|^2.$$

To find the min:

$$\frac{d}{ds}\text{Risk}_\mu[sX] = 2s\big(d + \|\mu\|^2\big) - 2\|\mu\|^2 = 0.$$

$$s = \frac{\|\mu\|^2}{d + \|\mu\|^2}.$$ +

The second derivative is:

$$\frac{d^2}{ds^2}\text{Risk}_\mu[sX] = 2\big(d + \|\mu\|^2\big) > 0,$$ +

So the critical point is definitely a minimum. Thus, the optimal shrinkage factor is:

$$s^* = \frac{\|\mu\|^2}{d + \|\mu\|^2} = 1 - \frac{d}{d + \|\mu\|^2}.$$ +

(c) Unfortunately, $\hat{\mu}^* = s^* X$ is not a proper estimator. (*Why?*) Instead of it, one may consider

$$\left(1 - \frac{d}{\|X\|^2}\right) X,$$

which is an actual estimator. Can you explain the heuristic motivation behind this estimator?

Answer:

It's invalid because it depends on μ, which we are trying to estimate. This creates a cycle. To resolve this circularity, we replace $\|\mu\|^2$ with an observable quantity derived from the data X.

Since $X \sim N(\mu, I_d)$, we know that:

$$\mathbb{E}[\|X\|^2] = \|\mu\|^2 + d.$$

$$\|\mu\|^2 \approx \|X\|^2 - d.$$

Precisely

Substituting this approximation into $s^* = \frac{\|\mu\|^2}{d+\|\mu\|^2}$:

$$s^* \approx \frac{\|X\|^2 - d}{\|X\|^2} = 1 - \frac{d}{\|X\|^2}.$$

+

$\frac{d}{\|X\|^2}$ adjusts the shrinkage. If $\|X\|^2$ is large (suggesting $\|\mu\|^2$ is large), the shrinkage factor $1 - \frac{d}{\|X\|^2}$ approaches 1, and the estimator behaves like the MLE; if it's small (suggesting $\|\mu\|^2$ is small), the shrinkage factor reduces X toward zero and lowers variance at the cost of bias.

It might be suboptimal compared to s^* but in high dimensions ($d > 2$), the aggregate risk reduction from shrinkage outweighs the bias uniformly across μ.

(d)

Assume $d \geq 2$. The James-Stein estimator is defined as:

$$\hat{\mu}^{JS} = \left(1 - \frac{d-2}{\|X\|^2}\right) X \tag{4}$$

This can be derived by minimizing, over $\delta \in \mathbb{R}$, the risk of the estimator:

$$\hat{\mu}^{\delta} = \left(1 - \frac{\delta}{\|X\|^2}\right) X$$

for a fixed μ.

Stein's Lemma To show that $R(\delta) = \text{Risk}_{\mu}[\hat{\mu}^{\delta}]$ is minimized at $\delta = d - 2$, use Stein's lemma:

Lemma 1. Let $X \sim \mathcal{N}(\mu, I)$ and $g(x)$ be a function on $\mathbb{R}^d$ that is differentiable almost everywhere, and satisfies:

$$\mathbb{E}_{\mu}\left[\frac{\partial}{\partial x_i} g(X)\right] < \infty, \quad \mathbb{E}_{\mu}\left[(X_i - \mu_i) g(X)\right] < \infty, \quad \forall i \in \{1, 2, \ldots, d\}.$$

Then:

$$\mathbb{E}_{\mu}[(X_i - \mu_i) g(X)] = \mathbb{E}_{\mu}\left[\frac{\partial}{\partial x_i} g(X)\right], \quad \forall i \in \{1, 2, \ldots, d\}.$$

Application of Stein's Lemma When applying Stein's lemma to the appropriate function $g(X)$, verify the absolute integrability conditions required by the lemma. Explain why the derivation fails when $d = 1$.

Convexity of $R(\delta)$ Finally, verify that $R(\delta)$ is strictly convex when $d \geq 3$, confirming that $\hat{\mu}^{JS}$ dominates the MLE.

Answer:

$$R(\delta) = \mathbb{E}_{\mu}\left[\|\hat{\mu}^{\delta} - \mu\|^2\right].$$

$$R(\delta) = \mathbb{E}_{\mu}\left[\left\|X - \mu - \frac{\delta}{\|X\|^2} X\right\|^2\right].$$

$$R(\delta) = \mathbb{E}_{\mu}\left[\|X - \mu\|^2\right] - 2\delta \mathbb{E}_{\mu}\left[\frac{X^{\top}(X - \mu)}{\|X\|^2}\right] + \delta^2 \mathbb{E}_{\mu}\left[\frac{\|X\|^2}{\|X\|^4}\right].$$

The MLE risk is d, so:

$$R(\delta) = d - 2\delta \mathbb{E}_{\mu}\left[\frac{X^{\top}(X - \mu)}{\|X\|^2}\right] + \delta^2 \mathbb{E}_{\mu}\left[\frac{1}{\|X\|^2}\right].$$

For $g(X) = \frac{X}{\|X\|^2}$, compute:

$$\mathbb{E}_{\mu}\left[\frac{X^{\top}(X - \mu)}{\|X\|^2}\right] = \sum_{i=1}^{d} \mathbb{E}_{\mu}\left[\frac{X_i(X_i - \mu_i)}{\|X\|^2}\right].$$

By Stein's Lemma:

$$\mathbb{E}_{\mu}\left[X_i(X_i - \mu_i)\right] = \mathbb{E}_{\mu}\left[\frac{\partial}{\partial X_i} g(X)\right].$$

Derivative:

$$\frac{\partial}{\partial X_i}\left(\frac{X_i}{\|X\|^2}\right) = \frac{1}{\|X\|^2} - \frac{2X_i^2}{\|X\|^4}.$$

Summing over i:

$$\sum_{i=1}^{d}\left(\frac{1}{\|X\|^2} - \frac{2X_i^2}{\|X\|^4}\right) = \frac{d-2}{\|X\|^2}.$$

Thus:

$$\mathbb{E}_\mu\left[\frac{X^\top(X-\mu)}{\|X\|^2}\right] = \mathbb{E}_\mu\left[\frac{d-2}{\|X\|^2}\right].$$

Substitute into $R(\delta)$:

$$R(\delta) = d - 2\delta(d-2)\mathbb{E}_\mu\left[\frac{1}{\|X\|^2}\right] + \delta^2\mathbb{E}_\mu\left[\frac{1}{\|X\|^2}\right].$$

Let $C = \mathbb{E}_\mu\left[\frac{1}{\|X\|^2}\right]$. Then:

$$R(\delta) = d - 2\delta(d-2)C + \delta^2 C.$$

$$\frac{dR}{d\delta} = -2(d-2)C + 2\delta C = 0 \implies \delta = d-2.$$

Thus, the optimal δ is $\delta = d-2$:

$$\hat{\mu}^{\text{JS}} = \left(1 - \frac{d-2}{\|X\|^2}\right)X.$$

Invalidity: for $d = 1$, $\|X\|^2 = X_1^2$, and $\mathbb{E}\left[\frac{1}{X_1^2}\right]$ is infinite if $X_1 \sim N(\mu, 1)$. The expectations required by Stein's Lemma do not exist. This violates the conditions.

The second derivative of $R(\delta)$ is:

$$\frac{d^2R}{d\delta^2} = 2C > 0 \quad (\text{since } C > 0 \text{ for } d \geq 3).$$

Thus, $R(\delta)$ is strictly convex, and $\delta = d-2$ is the unique global minimum. Therefore:

$$R(d-2) < R(1) = d \quad \text{for all } \mu \in \mathbb{R}^d,$$

proving that $\hat{\mu}^{\text{JS}}$ dominates the MLE when $d \geq 3$.

Exercise

Consider a similar location family as in the problem, but with some other noise distribution that is also rotation-invariant (you can normalize it as per $\mathbb{E}[\|X-\mu\|_2^2] = d$, as in the Gaussian case, but $\|X-\mu\|_2^2$ now doesn't have to be chi-squared, just any nonnegative random variable with expectation d). Show that for any such distribution, you have inadmissibility of MLE for $d > 3$ (strictly). This can be done by going up to Stein's formula but stopping short of applying it—as we're not Gaussian anymore—and doing something simple (and silly) instead.

Answer:

$$\hat{\mu}_{JS} = \left(1 - \frac{d-2}{\|X\|^2}\right) X$$

$$\text{Risk}(\hat{\mu}_{JS}) = \mathbb{E}\left[\left\|\left(1 - \frac{d-2}{\|X\|^2}\right) X - \mu\right\|^2\right] = \mathbb{E}\left[\left\| Z - \frac{(d-2)}{\|X\|^2} X\right\|^2\right]$$

Expand the squared norm:

$$= \mathbb{E}\|X-\mu\|^2 - 2(d-2)\mathbb{E}\left[\frac{(X-\mu)\cdot X}{\|X\|^2}\right] + (d-2)^2\mathbb{E}\left[\frac{1}{\|X\|^2}\right] \quad +$$

- Let $Z = X - \mu$, which is spherically symmetric around 0.
- $\mathbb{E}[\|X-\mu\|^2] = d$ (given).
- Compute $\mathbb{E}\left[\frac{Z\cdot X}{\|X\|^2}\right]$:

$$Z\cdot X = \|X\|^2 - \mu\cdot X \Rightarrow \mathbb{E}\left[\frac{Z\cdot X}{\|X\|^2}\right] = 1 - \mathbb{E}\left[\frac{\mu\cdot X}{\|X\|^2}\right]$$

By spherical symmetry, align μ along the first coordinate axis. Then: $X = \|\mu\| e_1 + Z$

$$\mathbb{E}\left[\frac{\mu\cdot X}{\|X\|^2}\right] = \|\mu\|^2\mathbb{E}\left[\frac{1}{\|X\|^2}\right] \quad +$$

Let $B = \mathbb{E}\left[\frac{1}{\|X\|^2}\right]$. Then:

$$\mathbb{E}\left[\frac{Z\cdot X}{\|X\|^2}\right] = 1 - \|\mu\|^2 B \quad +$$

$$\text{Risk}(\hat{\mu}_{JS}) = d - 2(d-2)(1 - \|\mu\|^2 B) + (d-2)^2 B \quad +$$

Simplify:

$$= d - 2(d-2) + [2(d-2)\|\mu\|^2 + (d-2)^2]B \quad +$$

From $\mathbb{E}[\|X\|^2] = \|\mu\|^2 + d$, apply the Cauchy-Schwarz inequality:

$$\mathbb{E}[\|X\|^2]\cdot\mathbb{E}\left[\frac{1}{\|X\|^2}\right] \geq 1 \Rightarrow B \geq \frac{1}{\|\mu\|^2 + d} \quad +$$

Substitute $B \geq \frac{1}{\|\mu\|^2+d}$ into the Risk:

Not obvious, let me check?...

$$\text{Risk}(\hat{\mu}_{JS}) \leq d - 2(d-2) + \frac{(d-2)(2\|\mu\|^2 + d - 2)}{\|\mu\|^2 + d}$$

$$= -d + 4 + \frac{(d-2)(2\|\mu\|^2 + d - 2)}{\|\mu\|^2 + d}$$

$(4-2d)(\|\mu\|^2+d) + (d-2)(2\|\mu\|^2+d-2)$

$= (d-2)[d-2-2d]$

$= -(d-2)(d+2) < 0$ when $d \geq 3$.

$$\text{Risk}(\hat{\mu}_{JS}) < d \quad \text{for all } \mu \in \mathbb{R}^d \text{ when } d \geq 3$$

Great Job! :)

Another exercise, probably a more interesting one, is to find a rotation-invariant distribution for which MLE is admissible in $d = 3$. I think it could be done, and the worst-case example might have somewhat heavier tails than chi-squared (for the squared norm). Possibly, an exponential distribution for the *norm* of the noise would do the trick. I've never run this calculation.

Answer:
I tried finding a rotation-invariant distribution with a singularity at zero but finite variance.

$$\hat{\mu}_{JS} = \left(1 - \frac{d-2}{||X||^2}\right) X \tag{7}$$

$$\text{Risk}(\hat{\mu}_{JS}) = \mathbb{E}\left[\left\|\left(1 - \frac{d-2}{||X||^2}\right) X - \mu\right\|^2\right] \tag{8}$$

$$\text{Risk} = \mathbb{E}[||X - \mu||^2] - 2(d-2)\mathbb{E}\left[\frac{(X-\mu)\cdot X}{||X||^2}\right] + (d-2)^2\mathbb{E}\left[\frac{1}{||X||^2}\right] \tag{9}$$

Consider a noise distribution where $Z = X - \mu$ follows:

- For $||z|| \leq 1$, let $f_Z(z) \propto ||z||^{-4}$. This creates a singularity at zero, ensuring $\mathbb{E}\left[\frac{1}{||Z||^2}\right] = \infty$.
- For $||z|| > 1$, let $f_Z(z) \propto e^{-||z||^2}$. This rapid decay ensures $\mathbb{E}[||Z||^2] = d = 3$.

The density is normalized such that:

$$\int_{\mathbb{R}^3} f_Z(z)dz = 1. \tag{10}$$

The tail decay $e^{-||z||^2}$ guarantees finite variance.
By construction, $\mathbb{E}[||Z||^2] = 3$, so:

$$\text{Risk}(X) = \mathbb{E}[||X - \mu||^2] = 3 < \infty. \tag{11}$$

I think you need 2 free params to match both normalization conditions, but the idea seems to work.

The JS estimator's risk involves $\mathbb{E}\left[\frac{1}{||X||^2}\right]$. For $\mu = 0$, we have $X = Z$, so:

Jacobian

$$\mathbb{E}\left[\frac{1}{||Z||^2}\right] = \int_0^1 \frac{1}{r^2}\cdot r^{-4}\cdot r^2 dr + \int_1^\infty \frac{1}{r^2}\cdot e^{-r^2}\cdot r^2 dr. \tag{12}$$

- The first integral diverges:

$$\int_0^1 r^{-4}dr = \infty. \tag{13}$$

- The second integral converges:

$$\int_1^\infty e^{-r^2}dr < \infty. \tag{14}$$

Thus,

$$\text{Risk}(\hat{\mu}_{JS}) = \infty, \tag{15}$$

This doesn't show that MLE is admissible, only that JS doesn't work to dominate MLE (and it's clear why). Still, impressive attempt!

7 Planar Venn diagrams

A (*congruent*) *Venn diagram* in $\mathbb{R}^d$ for n sets is the following object: you choose a "base" set $A \subset \mathbb{R}^d$ and n locations $a_1, \ldots, a_n \in \mathbb{R}^d$ such that the shifted sets $A_1, A_2, \ldots, A_n$, where $A_j := \{a + a_j : a \in A\}$, intersect in all possible combinations: for any subset of indices $I \subset \{1, 2, \ldots, n\}$, the set

$$A_I := \bigcap_{i \in I} A_i$$

must be nonempty. Prove the following result:

One cannot draw a planar ($d = 2$) Venn diagram for $n \geq 5$ sets by shifting a circle.

Use **Euler's formula**: any planar graph with V vertices, E edges, and F faces (subsets in which $\mathbb{R}^2$ is partitioned by the graph) satisfies

$$V - E + F = 2.$$

For example, in the case of a triangle $V = E = 3$ and $F = 2$.

Hint: Estimate V_n, E_n, F_n in a Venn diagram for n sets in terms of V_{n-1}, E_{n-1}, F_{n-1} respectively. [1]

Answer:

When adding the nth circle, it can intersect each of the previous $n-1$ circles in at most two points, which creates at most $2(n-1)$ new vertices, and each such new vertex can introduce at most one new face, giving $F_n \leq F_{n-1} + 2(n-1)$. Starting with $F_1 = 2$, $F_n \leq n^2 - n + 2$. For a complete n-set Venn diagram, one needs $2^n - 1$ nonempty regions, which exceeds $n^2 - n + 2$ when $n \geq 5$. Therefore, no arrangement of $n \geq 5$ congruent circles can form a planar Venn diagram.

Claim:
Each new vertex can introduce at most one new face.

Proof:

- Let G_{n-1} be the planar graph formed by $n-1$ circles. By Euler's formula, G_{n-1} partitions the plane into F_{n-1} faces.
- The n-th circle intersects each existing circle at 2 points, therefore creates $2(n-1)$ new vertices. These vertices divide the n-th circle into $2(n-1)$ arcs.
- Consider an arbitrary arc e of the n-th circle. By planarity, e does not cross any edges of G_{n-1} except at its endpoints (the two vertices where the n-th circle intersects existing circles).
- Thus, e lies entirely within a **single face** f of G_{n-1}.
- The arc e splits f into two sub-regions:
  - **Sub-region 1**: Bounded by e and existing edges of f.
  - **Sub-region 2**: The complement of Sub-region 1 within f.
- This splitting action increases the total number of faces by 1 (since one face becomes two).
- Since there are $2(n-1)$ arcs on the n-th circle, and each arc splits at most one face, the total number of new faces introduced is at most $2(n-1)$.

No Overcounting

- Each arc resides in a distinct face of G_{n-1}. This follows because:
  - The n-th circle is a simple closed curve.
  - Between consecutive arcs, the n-th circle transitions into a new face via an intersection vertex (ensuring no two arcs lie in the same face).

Well done!

- Hence, no two arcs split the same face, and each contributes uniquely to the face count.

Topology Justification

- Let f be a face of G_{n-1}. By the Jordan Curve Theorem, the arc e (a simple curve with endpoints on ∂f) divides f into two connected components.
- Formally, $f \setminus e$ has two path-connected components, which correspond to two new faces in G_n.
- *a simple curve connecting two boundary points of a face splits it into two faces.*