Math 6262: Statistical Estimation (Spring 2024)
Solutions to Homework 2

1^o: *Tail bounds for the Gaussian distribution.*

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

(a) Prove the following bounds (holding for all $u \geqslant 0$):

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

Hint 1: *Try to prove the upper bound first.*

Hint 2: *Integration by parts is the way here; use it first to prove the upper bound, and then for the lower bound.*

For the upper bound, note that

$$\sqrt{2\pi}\Phi(u) = \int_u^\infty e^{-\frac{t^2}{2}} dt \leqslant \frac{1}{u}\int_u^\infty te^{-\frac{t^2}{2}} dt = \frac{1}{u}\int_{\frac{u^2}{2}}^\infty e^{-v} dv = \frac{1}{u}e^{-\frac{u^2}{2}}.$$

and we are done. However, it is not clear how to get the lower bound from this (as we applied an estimate from above already in the first step.) So, let us try a slightly different route:

$$\sqrt{2\pi}\Phi(u) = \int_u^\infty e^{-\frac{t^2}{2}} dt = \int_u^\infty \underbrace{\frac{1}{t}}_{f_0(t)} \underbrace{te^{-\frac{t^2}{2}} dt}_{dg(t)}$$

where $g(t) = -e^{-\frac{t^2}{2}}$. Integrating by parts we get

$$\sqrt{2\pi}\Phi(u) = -\frac{1}{t}e^{-\frac{t^2}{2}}\Big|_u^\infty - \int_u^\infty \frac{1}{t^2}e^{-\frac{t^2}{2}} dt = \frac{1}{u}e^{-\frac{u^2}{2}} - \int_u^\infty \frac{1}{t^2}e^{-\frac{t^2}{2}} dt.$$

Now the upper bound follows by observing that the integral is non-negative. On the other hand, for the lower bound we have to *upper-bound* this integral, i.e. upper-bound

$$I_1(u) := \int_u^\infty \frac{1}{t^2}e^{-\frac{t^2}{2}} dt.$$

This can be done by the same trick as before when dealing with $I_0(u) := \int_u^\infty e^{-\frac{t^2}{2}} dt$, namely:

$$I_1(u) = \int_u^\infty \underbrace{\frac{1}{t^3}}_{f_1(t)} \underbrace{te^{-\frac{t^2}{2}} dt}_{dg(t)} = \frac{1}{u^3}e^{-\frac{u^2}{2}} - \int_u^\infty \frac{3}{t^4}e^{-\frac{t^2}{2}} dt.$$

We recognize the lower bound by noting that the subtracted integral is non-negative.

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

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

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

The new upper bound follows by estimating the integral $I_2(u) = \int_u^\infty \frac{1}{t^4} e^{-\frac{t^2}{2}} dt$ via the same method, i.e. by writing it as $I_2(u) = \int_u^\infty f_5(t) dg(t)$ with $f_2(t) = \frac{1}{t^5}$ and integrating by parts.

(c)* If we continue applying this approach iteratively, what bounds shall we get after k such "iterations?"

We have the following pattern. For any nonnegative integer k, define

$$I_k(u) := \int_u^\infty t^{-2k} e^{-\frac{t^2}{2}} dt \quad \text{and} \quad f_k(t) := t^{-(2k+1)}.$$

Then, in order to bound $I_0(u) = \sqrt{2\pi}\Phi(u)$, we observe that, through the same trick as before,

$$\begin{aligned} I_k(u) = \int_u^\infty f_k(t) dg(t) &= f_k(t) g(t) \Big|_u^\infty - (2k+1) I_{k+1}(u) \\ &= \frac{1}{u^{2k+1}} e^{-\frac{u^2}{2}} - (2k+1) I_{k+1}(u). \end{aligned}$$

From this it is easy to see that

$$e^{\frac{u^2}{2}} I_0(u) = \frac{1}{u} - \frac{1}{u^3} + \frac{1 \cdot 3}{u^5} - \frac{1 \cdot 3 \cdot 5}{u^7} + \ldots = \sum_{k=0}^\infty (-1)^k \frac{(2k-1)!!}{u^{2k+1}},$$

and that we get an upper (resp. lower) bound if we stop at any positive (resp. negative) term.

See also "Bounding Standard Gaussian Tail Probabilities" by L. Duembgen (2010).

2°: *Stein's paradox.* Consider the problem of estimating the mean μ in the multivariate Gaussian location family

$$\mathbb{P}_\mu = \mathcal{N}(\mu, I), \quad \text{for} \quad \mu \in \mathbb{R}^d, \tag{1}$$

from a single observation $X \sim \mathbb{P}_\mu$. 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 $\|x\| := (\sum_i x_i^2)^{1/2}$ is the Euclidean norm of $x = (x_1, ..., x_d) \in \mathbb{R}^d$, we see that for any $\mu \in \mathbb{R}^d$,

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

Intuitively, it is hard to suspect that one can find a more reasonable estimator of μ than X. Yet, this turns out to be the case: one may improve over the MLE uniformly on the family (**??**) 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* $\mathrm{Risk}_\mu[\hat{\mu}'] \leqslant \mathrm{Risk}_\mu[\hat{\mu}]$ *for any μ, and there exists a parameter value $\bar{\mu}$ such that* $\mathrm{Risk}_{\bar{\mu}}[\hat{\mu}'] < \mathrm{Risk}_{\bar{\mu}}[\hat{\mu}]$.

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 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.

You will show that the MLE is inadmissible when $d \geqslant 3$, by constructing a dominating estimator.

(i) 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 $s > 1$. (**Hint:** *look for an estimator* $\hat{\mu}' = s'X$ *with* $s' \in [0, 1]$.)

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

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

Define $L_\mu(s) := \mathrm{Risk}_\mu[sX]$ for any (fixed) $s \in \mathbb{R}$, then (denoting $\|\cdot\|$ the ℓ_2-norm on $\mathbb{R}^d$):

$$\begin{aligned} L_\mu(s) = \mathbb{E}[\|sX - \mu\|^2] &= \mathbb{E}[\|(1-s)\mu + s(X - \mu)\|^2] \\ &= (1-s)^2\|\mu\|^2 + s^2\mathbb{E}[\|X - \mu\|^2] \\ &= (1-s)^2\|\mu\|^2 + s^2 d. \end{aligned}$$

Clearly, $L_\mu(s) > L_\mu(0)$ for any $s < 0$ and $\mu \in \mathbb{R}^d$); thus, $\hat{\mu} = sX$ with such s is dominated by the trivial estimator $\hat{\mu} \equiv 0$. Similarly, $\hat{\mu} = sX$ with $s > 1$ is dominated by $\hat{\mu} = X$. As for (ii), it is clear that s^* is indeed the unique minimizer of $L_\mu(\cdot)$: by examining the first two derivatives we see that $L_\mu(\cdot)$ is strongly convex on $\mathbb{R}^d$.

(iii) 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?

The "ideal" estimator, as given by $s^* X$, is unavailable because s^* depends on the unknown quantity $\|\mu\|$ in the denominator. In such situations, a natural idea is to replace the unknown quantity by its unbiased estimator—that is, if we are lucky enough, and such an estimate can be inferred from observations at hand. This is the case here: $\mathbb{E}[\|X\|^2] = \|\mu\|^2 - d$, so we have an unbiased estimate of $\|\mu\|^2$ as given by $\|X\|^2 - d$. This is what motivates the proposed estimator.

Explanation (not required in the solution): However, this estimator does *not* dominate MLE (at least I am not aware of any proof). In order to transform it to a dominating estimator, we have to slightly modify it as suggested in (iv). Indeed, unbiased estimation of $\|\mu\|^2$, as suggested here, is simply a heuristic idea, and the suggested estimator does not have to be optimal. Moreover, it is not hard to guess—intuitively—why it will not be so: while $\|X\|^2$ is an unbiased estimate of $\|\mu\|^2 + d$, clearly $d/\|X\|^2$ is a *biased* estimate of the quantity $\frac{d}{\|\mu\|^2+d}$. In particular, one may verify that if $z \sim \chi^2_d$, then $\mathbb{E}[\frac{1}{z}] = \frac{1}{d-2}$ for $d > 2$ and $\mathbb{E}[\frac{1}{z}] = \infty$ for $d \leqslant 2$. (Try it!) The final step of the problem can be understood as a way of eliminating this bias.

(iv*) **Here is the most difficult step.** Assuming that $d \geqslant 2$, derive the *James-Stein estimator*

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

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. In order to show that $R(\delta) = \mathrm{Risk}_\mu[\hat{\mu}^\delta]$ is minimized at $d-2$, use Stein's lemma:

Lemma 1. *Let $X \sim \mathcal{N}(\mu, I)$ and $g(x)$ be a function on $\mathbb{R}^d$ differentiable almost everywhere, and such that $\mathbb{E}_\mu\left[|\frac{\partial}{\partial x_i} g(X)|\right] < \infty$ and $\mathbb{E}_\mu[|(X_i - \mu_i) g(X)|] < \infty$ for any $i \in [d] := \{1, 2, ..., d\}$. Then*

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

When applying Stein's lemma to the right function $g(X)$, please do check the absolute integrability conditions in its premise, and explain why the argument does not work for $d = 1$. Finally, verify that $R(\delta)$ is strictly convex when $d \geqslant 3$ (thus $\hat{\mu}^{JS}$ indeed dominates the MLE).

Let $R(\delta) = \mathrm{Risk}_\mu[\hat{\mu}^\delta]$. Observe that (writing $\mathbb{E}$ instead of $\mathbb{E}_\mu$ for brevity):

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

Clearly, $\mathbb{E}[\|X-\mu\|^2] = d$. However, the last two terms are not easily computed explicitly, and this is where Stein's lemma will help us. However, even before we proceed with it, observe that $R(\delta)$ is a strictly convex quadratic, and thus has a (unique) minimizer.

Now, our plan is to apply Stein's lemma with functions $g_i(x) = \frac{x_i}{\|x\|^2}$, for each $i \in [d]$, in the role of $g(x)$. Note that differentiable everywhere on $\mathbb{R}^d$ except the origin, and we compute

$$\frac{\partial}{\partial x_i} g_i(x) = \frac{\|x\|^2 - 2x_i^2}{\|x\|^4} = \frac{1}{\|x\|^2} - \frac{2x_i^2}{\|x\|^4}.$$

We'll check the conditions of the lemma later on; for now let's see what it implies. By writing $\langle X-\mu, \frac{1}{\|X\|^2}X\rangle$ as $\sum_{i\in[d]}(X_i-\mu_i)g_i(X)$ and combining Stein's lemma with the previous result, we get

$$\mathbb{E}\left[\langle X-\mu, \frac{1}{\|X\|^2}X\rangle\right] = (d-2)\mathbb{E}\left[\frac{1}{\|X\|^2}\right].$$

(The "-2" comes from summing the $-\frac{2x_i^2}{\|x\|^4}$ terms.) As a result,

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

and we find the minimizer $\delta^* = d-2$ without computing $\mathbb{E}\left[\frac{1}{\|X\|^2}\right]$. Strict convexity is clear by noting that $\mathbb{E}\left[\frac{1}{\|X\|^2}\right] > 0$. It remains to verify that the premise of the lemma indeed holds. (Note that we also used that $\mathbb{E}\left[\frac{1}{\|X\|^2}\right] < \infty$, but we shall prove it as well.) Note that

$$\mathbb{E}\left[\left|\frac{\partial}{\partial x_i} g_i(X)\right|\right] \leqslant 3\mathbb{E}\left[\frac{1}{\|X\|^2}\right]$$

and

$$\begin{aligned}\sum_{i\in[d]} \mathbb{E}[|(X_i-\mu_i)g_i(X)|] \leqslant \mathbb{E}\left[\sum_{i\in[d]} \frac{X_i^2}{\|X\|^2} + \frac{|\mu_i x_i|}{\|X\|^2}\right] &\leqslant 1 + \mathbb{E}\left[\sum_{i\in[d]} \frac{|\mu_i x_i|}{\|X\|^2}\right] \\ &\leqslant 1 + \frac{1}{2}\mathbb{E}\left[\sum_{i\in[d]} \frac{\mu_i^2 + x_i^2}{\|X\|^2}\right] \\ &= \frac{3}{2} + \frac{\|\mu\|^2}{2}\mathbb{E}\left[\frac{1}{\|X\|^2}\right].\end{aligned}$$

Thus, for $d > 2$ it suffices to verify that $\mathbb{E}\left[\frac{1}{\|X\|^2}\right] < \infty$. Let $Z = X - \mu \sim \mathcal{N}(0, I_d)$, then

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

By rotational invariance of the Euclidean norm and the distribution of Z, we can w.l.o.g. assume that $\mu = ae_1$, where $a = \|\mu\|$ and e_1 is the first canonical basis vector; this results in

$$\mathbb{E}\left[\frac{1}{\|X\|^2}\right] = \mathbb{E}\left[\frac{1}{(Z_1-a)^2 + \sum_{i\in[d-1]} Z_i^2}\right].$$

It remains to bound the right-hand side. First, there is a partial solution working when $d > 3$, but *not* for $d = 3$: to neglect the term depending on μ and note that

$$\mathbb{E}\left[\frac{1}{\|X\|^2}\right] \leqslant \mathbb{E}\left[\frac{1}{\sum_{i\in[d-1]} Z_i^2}\right].$$

Since $\sum_{i\in[d-1]} Z_i^2 \sim \chi^2_{d-1}$, and the expectation of the inverse of χ^2_d is $1/(d-2)$ for $d > 2$ (as we shall check in a moment), this would result in

$$\mathbb{E}\left[\frac{1}{\|X\|^2}\right] \leqslant \frac{1}{d-3},$$

which is finite for $d > 3$. For the full solution, we show that the case $a = 0$ is the hardest, namely

$$\mathbb{E}\left[\frac{1}{(Z_1 - a)^2 + \sum_{i\in[d-1]} Z_i^2}\right] \leqslant \mathbb{E}\left[\frac{1}{\|Z\|^2}\right] + c$$

for some $c < \infty$. (This is a crude bound, but it would suffice for our purposes.) Indeed, taking $d = 2$ w.l.o.g. (the case $d \geqslant 3$ is analogous, and is not needed anyway in view of the partial solution) we have

$$\mathbb{E}\left[\frac{1}{(Z_1 - a)^2 + Z_2^2}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{1}{(Z_1 - a)^2 + Z_2^2}\middle| Z_2\right]\right],$$

and it suffices to show that

$$\mathbb{E}\left[\frac{1}{(Z_1 - a)^2 + b^2}\right] \leqslant \mathbb{E}\left[\frac{1}{Z_1^2 + b^2} + c\right]$$

for any $a, b \geqslant 0$. To this end, observe that

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{(Z_1 - a)^2 + b^2}\right] \propto \int_{-\infty}^{+\infty} \frac{e^{-\frac{z^2}{2}}}{(z-a)^2 + b^2} dz &= e^{-\frac{a^2}{2}} \int_{-\infty}^{+\infty} \frac{e^{-\frac{u^2}{2}} e^{-au}}{u^2 + b^2} du \\
&= e^{-\frac{a^2}{2}} \left(\int_0^{+\infty} \frac{e^{-\frac{u^2}{2}} e^{-au}}{u^2 + b^2} du + \int_{-\infty}^0 \frac{e^{-\frac{u^2}{2}} e^{-au}}{u^2 + b^2} du \right) \\
&\leqslant e^{-\frac{a^2}{2}} \left(\int_0^{+\infty} \frac{e^{-\frac{u^2}{2}}}{u^2 + b^2} du + \int_{-\infty}^0 \frac{e^{-\frac{u^2}{2}} e^{-au}}{u^2 + b^2} du \right) \\
&\leqslant \frac{1}{2}\mathbb{E}\left[\frac{1}{Z_1^2 + b^2}\right] + e^{-\frac{a^2}{2}} \int_{-\infty}^0 \frac{e^{-\frac{u^2}{2}} e^{-au}}{u^2 + b^2} du,
\end{aligned}$$

and the last term is finite: $|au| \leqslant \frac{1}{2}a^2 + \frac{1}{2}u^2$ implies $e^{-\frac{a^2}{2}} \int_{-\infty}^0 \frac{e^{-\frac{u^2}{2}} e^{-au} du}{u^2+b^2} \leqslant \int_{-\infty}^0 \frac{du}{u^2+b^2} < \infty$. In order to validate our application of Stein's lemma, it only remains to verify that $\mathbb{E}[\frac{1}{\|Z\|^2}] < \infty$ if $Z \sim \mathcal{N}(0, I_d)$ with $d \geqslant 3$. We shall use the spherical coordinates: recall that the Jacobian determinant is r^{d-1} and denoting A_{d-1} the surface area of the sphere $\{z \in \mathbb{R}^d : \|z\| = 1\}$, then

$$\mathbb{E}\left[\frac{1}{\|Z\|^2}\right] = \frac{A_{d-1}}{(2\pi)^{d/2}} \int_0^{\infty} r^{d-3} e^{-r^2/2} dr.$$

It is easy to see that the integral is finite for any $d \geqslant 3$: this is the case for $d = 3$, and for $d > 3$ we can split the integral into $\int_0^1$ and on $\int_1^\infty$ and estimate them separately. Finally, for $d \leqslant 2$—in particular, for $d = 1$—the proposed argument does not work as the integral clearly diverges. ■