Received: 27 October 2023 | Revised: 29 January 2024 | Accepted: 30 January 2024

DOI: 10.1112/blms.13025

RESEARCH ARTICLE

Bulletin of the London Mathematical Society

# Most likely balls in Banach spaces: Existence and nonexistence

**Bernd Schmidt**

Institut für Mathematik, Universität Augsburg, Augsburg, Germany

**Correspondence**
Bernd Schmidt, Institut für Mathematik, Universität Augsburg, Universitätsstr. 14, 86159 Augsburg, Germany.
Email: bernd.schmidt@math.uni-augsburg.de

**Abstract**
We establish a general criterion for the existence of convex sets of fixed shape as, for example, balls of a given radius, of maximal probability on Banach spaces. We also provide counterexamples, showing that their existence may fail even in some common situations.

**MSC 2020**
60B11, 28C20, 62F10 (primary)

## 1 | INTRODUCTION

In this note, we address the natural question if for a (Borel) probability measure $\mu$ on a separable Banach space $X$ and for a given radius, there is always a "ball of maximum probability," that is, if the maximum of $\mu(B)$ is attained among all balls $B \subset X$ of radius $r > 0$. For small radii, such a maximizer — if existent — can be viewed as an approximation to and a regularization of a "mode" referring to a "point of maximum probability" for the given measure $\mu$ as $r$ becomes small. More generally, instead of balls of a fixed radius, our main existence theorem will also apply to the system of all translates of a fixed convex set $C$.

This issue has received considerable attention recently, notably in the area of Bayesian inverse problems (cf. [11]) where the problem arises to devise maximum a posteriori (MAP) estimators, in particular for Gaussian (and more general) priors on infinite dimensional (separable) Banach spaces, [1, 4–8, 10]. Indeed, seminal results in that area made implicit use of the existence of balls of maximum likelihood (cp. the discussion in [7]), whose existence could be established in specific situations, cf. [1, 9]. Only very recently, it has been noted that the question of their existence can be circumvented by considering the asymptotics of almost maximizers in order to obtain MAP estimators [7, 8]. Nevertheless, besides being a question of intrinsic interest, the problem remains relevant as the quest for the position of a (small) ball of positive radius with maximal probability amounts to solving a regularized point optimization problem. As such it typically enjoys improved stability properties and might even be favorable from a modeling perspective. In particular, this will be the case in situations when, possibly due to data uncertainties in the presence of noise, it is preferable to estimate parameter regions rather than single points.

The problem is addressed for general metric spaces in some detail in [9, Sect. 4.2]. There the authors provide a collection of technical sufficient conditions for the existence of maximum likelihood balls ("radius-$r$ modes" in their terminology). They also give counterexamples for some particular measures on specific spaces and certain ranges of $r$. The farthest reaching existence result for measures on Banach spaces known to date seems to be the following direct consequence of [9, Lem. SM1.6]: maximum likelihood balls do exist for measures whose canonical finite-dimensional projections do not charge any sphere. In particular, this applies to "radius-$r$ maximum likelihood a posteriori estimators for Gaussian priors" on $\ell^p$, $1 < p < \infty$, which are absolutely continuous with respect to a nondegenerate Gaussian. It appears that no counterexample on a Banach space is known to date.

The purpose of this note is twofold. First, we establish a general existence theorem for maximum likelihood convex shapes (and, in particular, balls of any radius) on Banach spaces. In particular, this will apply to every separable and reflexive space (and $\ell^1$), thus closing a gap in the seminal contribution [5]. Second, by way of various examples, we also show that existence may fail in some natural situations. In fact, we will provide a couple of counterexamples of probability measures on $c_0$ and the classical Wiener space, which might even be absolutely continuous with respect to a nondegenerate Gaussian measure, and which do not allow for balls of maximal probability for any value of radius $r$.

## 2 | A GENERAL EXISTENCE RESULT

Throughout we assume that $X$ is a separable (real) Banach space. By $B_r$ and $B_r^\circ$, we denote the closed and, respectively, open ball of radius $r$ in $X$ centered at 0. For $x \in X$, $C \subset X$, we write $x + C =: C(x)$.

**Theorem 2.1.** *Suppose that $X$ is the separable dual of a Banach space and $\mu$ is a Borel probability measure on $X$. Let $C \subset X$ be a bounded weak*-closed convex set $C \subset X$ with nonempty interior. Then, there exists $x_0 \in X$ such that*

$$\mu(C(x_0)) \geqslant \mu(C(x))$$

*for all $x \in X$.*

*Remark* 2.2.

1. In particular, this applies to every separable reflexive Banach space $X$.
2. An admissible choice for $C$ is $C = B_r$ for any $r > 0$.
3. As open balls have been considered in the literature as well, we notice that the result holds true, obviously, for $C = B_r^\circ$ in case $\mu$ does not charge any sphere of radius $r$, while it fails in

general. If, for example, $\mu$ is the uniform distribution on $\partial B_1$ in $\mathbb{R}^2$, then $\mu(B_1^\circ(x)) < \frac{1}{2}$ for any $x \in \mathbb{R}^2$ while $\lim_{x\to 0} \mu(B_1^\circ(x)) = \frac{1}{2}$.

*Proof.* Without loss of generality, we assume that 0 is an interior point of $C$. We consider a maximizing sequence of translates $C(x_n)$, that is,

$$\mu(C(x_n)) \to m_0 := \sup\{\mu(C(x)) : x \in X\}.$$

Since $X$ is separable and $C$ has nonempty interior, it is easy to see that $m_0 > 0$. Clearly, the sequence $(x_n)$ is bounded. (If $R > 0$ is chosen such that $\mu(X \setminus B_R) \leqslant \frac{m_0}{2}$, then $x_n \in B_{R+\mathrm{diam}(C)}$ eventually.) As $\mu$ is tight (on the polish space $X$), also the family $(\mu_n)$ of restrictions $\mu_n = \mu(\cdot \cap C(x_n))$ is tight. Thus, Prohorov's theorem (see, e.g., [3, Thm. 8.6.2]) implies $\mu_n \xrightarrow{w} \mu_0$ weakly (in duality with $C_b(X)$) for a (not relabeled) subsequence and a finite measure $\mu_0$. It follows that

$$\mu_0(X) = \lim_{n\to\infty} \mu_n(X) = \lim_{n\to\infty} \mu(C(x_n)) = m_0.$$

By assumption, $X$ has a separable predual $X_*$, so by Alaoglu's theorem, we may pass to a further subsequence (not relabeled) such that $x_n \overset{*}{\rightharpoonup} x_0$ weakly* in $X$ for some $x_0 \in X$. We will now prove that $\mu_0$ is supported on $C(x_0)$. To this end, we fix any $z \notin C(x_0)$. Since $C$ is weak*-closed, with the help of the Hahn–Banach theorem for the dual pairing $(X, X_*)$ (see, e.g., [2, Thms. 5.79 & 5.93]), we can choose an element $x_* \in X_*$ and then an $\varepsilon > 0$ such that

$$\sup\{\langle x_*, y\rangle : y \in B_\varepsilon(z)\} = \langle x_*, z\rangle + \varepsilon\|x_*\| < \inf\{\langle x_*, y\rangle : y \in C(x_0)\}. \tag{1}$$

On the other hand, the Portmanteau theorem (see, e.g., [3, Thm. 8.2.3]) implies

$$\mu_0(B_\varepsilon^\circ(z)) \leqslant \liminf_{n\to\infty} \mu_n(B_\varepsilon^\circ(z)) = \liminf_{n\to\infty} \mu(C(x_n) \cap B_\varepsilon^\circ(z)).$$

As a consequence, we conclude that, in case $\mu_0(B_\varepsilon^\circ(z)) > 0$, we have $C(x_n) \cap B_\varepsilon^\circ(z) \neq \emptyset$ for sufficiently large $n$, say $x_n + y_n \in C(x_n) \cap B_\varepsilon^\circ(z)$ (and so $y_n \in C$). Passing to yet another subsequence (not relabeled), we get $y_n \overset{*}{\rightharpoonup} y$ for some $y \in C$. It follows that $x_n + y_n \overset{*}{\rightharpoonup} x_0 + y \in C(x_0) \cap B_\varepsilon(z)$, which contradicts (1). So, we must have $\mu_0(B_\varepsilon^\circ(z)) = 0$. This proves that $\operatorname{supp} \mu_0 \subset C(x_0)$. It remains to observe that $\mu_0 \leqslant \mu$, which follows from the outer regularity of the Borel measure $\mu$ and from the fact that for any open subset $U \subset X$ the Portmanteau theorem gives

$$\mu_0(U) \leqslant \liminf_{n\to\infty} \mu_n(U) = \liminf_{n\to\infty} \mu(C(x_n) \cap U) \leqslant \mu(U).$$

Summarizing we find that

$$\mu(C(x_0)) \geqslant \mu_0(C(x_0)) = \mu_0(X) \geqslant m_0,$$

which, by definition of $m_0$, proves $\mu(C(x_0)) = m_0$. □

## 3 | EXAMPLES OF NONEXISTENCE

We discuss a number of concrete cases, where balls of maximum likelihood do not exist. There is a common underlying idea in all of them that would easily allow to generate further examples along these lines.

Our first two examples are on the Banach space $c_0$ of (real) null sequences equipped with the sup-norm, which is separable and even has a separable dual (namely, $\ell^1$), but is not a dual space itself.

**Example 3.1.** On $X = c_0$, we let $\mu = \bigotimes_{k\in\mathbb{N}} \mathrm{Exp}(k)$, where $\mathrm{Exp}(\lambda)$ denotes the exponential distribution on $\mathbb{R}$ with rate parameter $\lambda$ (and cumulative distribution function $x \mapsto 1 - \mathrm{e}^{-\lambda x^+}$). An application of the Borel–Cantelli lemma to $(\{x \in \mathbb{R}^{\mathbb{N}} : x_n \geqslant n^{-1/2}\})_{n\in\mathbb{N}}$ shows $\mu(X) = 1$. Let $r > 0$ arbitrary. For any ball $B_r(x)$, $x = (x_1, x_2, \ldots)$, one has

$$\mu(B_r(x)) = \prod_{k\in\mathbb{N}} \left(\mathrm{e}^{-k(x_k - r)^+} - \mathrm{e}^{-k(x_k + r)^+}\right).$$

Choosing $k_0$ such that $x_{k_0} < r$ and setting $x'_k = x_k$ for $k \neq k_0$, $x'_{k_0} = r$, we get $\mu(B_r(x')) > \mu(B_r(x))$ if $\mu(B_r(x)) > 0$. This shows that $x \mapsto \mu(B_r(x))$ does not have a maximizer. (Its supremum is $m_0 = \prod_{k\in\mathbb{N}} \left(1 - \mathrm{e}^{-2kr}\right)$, which can be seen by maximizing each factor separately and considering the maximizing sequence $(x_{\cdot,n}) \subset c_0$, $x_{k,n} = r$ for $k \leqslant n$, $x_{k,n} = 0$ for $k > n$.)

**Example 3.2.** In order to give an example where $\mu$ is absolutely continuous with respect to a Gaussian measure on $X = c_0$, we first let $\mu_0 = \bigotimes_{k\in\mathbb{N}} \mathcal{N}(0, k^{-2})$, where $\mathcal{N}(0, \sigma^2)$ denotes the Gaussian on $\mathbb{R}$ with mean 0, variance $\sigma^2$, and cumulative distribution function $X \mapsto \Phi(x/\sigma)$. An application of the Borel–Cantelli lemma to $(\{x \in \mathbb{R}^{\mathbb{N}} : |x_n| \geqslant n^{-1/2}\})_{n\in\mathbb{N}}$ shows $\mu_0(X) = 1$. We consider the (closed) set $A \subset X$ given by

$$A = \left\{x \in X : x_k \geqslant -1/\sqrt{k} \text{ for all } k \in \mathbb{N}\right\},$$

and note that $\mu_0(A)$ (the probability that the coordinate process does not pass the moving boundary $k \mapsto -1/\sqrt{k}$) is positive since $\mu_0(X \setminus A) \leqslant \sum_{k\in\mathbb{N}} \Phi(-\sqrt{k}) < 1$. We then define $\mu$ by conditioning on $A$, that is, we set $\mu = \frac{1}{\mu_0(A)} \mathbb{1}_A \mu_0$. A similar reasoning as above shows that, for any $r > 0$, balls of maximal probability do not exist (and the supremum is explicitly given as $m_0 = \frac{1}{\mu_0(A)} \prod_{k\in\mathbb{N}} [\Phi((kr - \sqrt{k})^+ + kr) - \Phi((kr - \sqrt{k})^+ - kr)]$), where $\mu_0(A) = \prod_{k\in\mathbb{N}} (1 - \Phi(-\sqrt{k}))$.

We now give some examples on the classical Wiener space $C_0[0,1] = \{\omega \in C[0,1] : \omega(0) = 0\}$ equipped, as usual, with the sup-norm in order to show that nonexistence of maximum likelihood balls is encountered in common situations in a continuous time setting. They follow the similar basic idea of the previous two examples.

In what follows we let $\mu_0$ be the Wiener measure on $C_0[0,1]$ so that the coordinate process $(\omega(t))_{t\in[0,1]}$ is Brownian motion.

**Example 3.3.** Similarly as in Example 3.1, we can consider typical processes that assume only nonnegative values as, for example, the running maximum of Brownian motion or the reflected

Brownian motion:

$$\omega_{\max}(t) = \max\{\omega(s) : 0 \leqslant s \leqslant t\}, \quad \text{respectively,} \quad |\omega|(t) = |\omega(t)|.$$

If $\mu$ denotes the corresponding distribution on $X = C_0[0,1]$, in both cases, the maximum of $\omega \mapsto \mu(B_r(\omega))$ is not attained for any $r > 0$. Indeed, as $\omega(t) \to 0$ for $t \to 0$ for any $\omega \in X$ and so $\omega(t) < r$ on $[0,s]$ for some $0 < s < 1$, it suffices to choose any $\omega' \in X$ such that $\omega < \omega' \leqslant r$ on $(0,s)$ and $\omega' = \omega$ on $[s,1]$ to get $\mu(B_r(\omega')) > \mu(B_r(\omega))$ if $\mu(B_r(\omega)) > 0$.

**Example 3.4.** Let $X = C_0[0,1]$. We choose a nonpositive $\rho \in X$ such that $\mu_0(\{\omega : \omega(t) \geqslant \rho(t) \text{ for all } t \in [0,1]\}) > 0$. Such a $\rho$ can be found with the help of Khinchin's law of the iterated logarithm: $\liminf_{t\to 0} \omega(t)/\sqrt{2t\log\log(1/t)} = -1$ for $\mu_0$-a.e. $\omega \in X$, which allows to choose $0 < t_0 < 1/\mathrm{e}$ such that

$$\mu_0\left(\left\{\omega(t) \geqslant -2\sqrt{2t\log\log(1/t)} \text{ for all } t \in [0,t_0]\right\}\right) > 0.$$

We now define $\rho \in X$ by $\rho(t) = -2\sqrt{2t\log\log(1/t)}$ for $t \leqslant t_0$ and then $\rho(t) = \rho(t_0)$ for $t > t_0$. The (closed) set $A \subset X$

$$A = \left\{\omega \in X : \omega(t) \geqslant \rho(t) \text{ for all } t \in [0,1]\right\} \subset X$$

will then have positive probability $\mu_0(A) > 0$. Conditioning on $A$, we define $\mu = \frac{1}{\mu_0(A)}\mathbb{1}_A\mu_0$. Let $r > 0$ arbitrary. For any ball $B_r(\omega)$, we have $\mu(B_r(\omega)) > 0$ if and only if $\omega + r > \rho$ on [0,1]. A construction as in the previous example shows that $\mu(B_r(\omega')) > \mu(B_r(\omega))$ for such $\omega$. So, again, the supremum of these values is not attained.

## ACKNOWLEDGMENTS

I am grateful to Philipp Wacker for drawing my attention to this problem and interesting conversations on the subject.

Open access funding enabled and organized by Projekt DEAL.

## JOURNAL INFORMATION

The *Bulletin of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

## ORCID

*Bernd Schmidt* https://orcid.org/0000-0003-4805-9055

## REFERENCES

1. S. Agapiou, M. Burger, M. Dashti, and T. Helin, *Sparsity-promoting and edge-preserving maximum* a posteriori *estimators in non-parametric Bayesian inverse problems*, Inverse Problems **34** (2018), no. 4, 045002, 37.
2. C. D. Aliprantis and K. C. Border, *Infinite dimensional analysis*, 3rd ed., Springer, Berlin, 2006.
3. V. I. Bogachev, *Measure theory. Vol. II*, Springer, Berlin, 2007.

4. C. Clason, T. Helin, R. Kretschmann, and P. Piiroinen, *Generalized modes in Bayesian inverse problems*, SIAM/ASA J. Uncertain. Quantif. **7** (2019), no. 2, 652–684.
5. M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss, *MAP estimators and their consistency in Bayesian nonparametric inverse problems*, Inverse Problems **29** (2013), no. 9, 095017, 27.
6. T. Helin and M. Burger, *Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems*, Inverse Problems **31** (2015), no. 8, 085009, 22.
7. I. Klebanov and P. Wacker, *Maximum a posteriori estimators in $\ell^p$ are well-defined for diagonal Gaussian priors*, Inverse Problems **39** (2023), no. 6, 065009, 27.
8. H. Lambley, *Strong maximum* a posteriori *estimation in Banach spaces with Gaussian priors*, Inverse Problems **39** (2023), no. 12, Paper No. 125010, 22.
9. H. Lambley and T. J. Sullivan, *An order-theoretic perspective on modes and maximum a posteriori estimation in Bayesian inverse problems*, SIAM/ASA J. Uncertain. Quantif. **11** (2023), no. 4, 1195–1224.
10. H. C. Lie and T. J. Sullivan, *Equivalence of weak and strong modes of measures on topological vector spaces*, Inverse Problems **34** (2018), no. 11, 115013, 22.
11. A. M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numer. **19** (2010), 451–559.