# Benefits and Pitfalls of the Exponential Mechanism with Applications to Hilbert Spaces and Functional PCA Supplementary Material

**Jordan Awan** [1] **Ana Kenney** [1] **Matthew Reimherr** [1] **Aleksandra Slavković** [1]

*Proof of Proposition 2.3.* The reverse direction is given in Remark 1 from Hall et al. (2013), though we provide the argument here again for completeness. Let $B \in \mathscr{F}$ and $X, X' \in \mathcal{X}^n$ be adjacent elements. Then

$$\mu_X(B) = \int_B f_X(b)\, d\nu(b) = \int_B \frac{f_{X'}(b)}{f_{X'(b)}} f_X(b)\, d\nu(b) \leq \int_B \exp(\epsilon) f_{X'}(b)\, d\nu(b) = \exp(\epsilon)\mu_{X'}(b),$$

which implies that $\mathscr{M}$ achieves $\epsilon$-DP.

Going in the other direction we will use a proof by contradiction. Assume that $\mathscr{M}$ is an $\epsilon$-DP mechanism. Recall that two measures are equivalent if they agree on the zero sets. Thus, as we have said, the measures in a DP mechanism must all be equivalent. So, we can assume that all of the measures have a density with respect to some common base measure, $\nu$, which, without loss of generality, we can take to be one of the elements of $\mathscr{M}$. Now assume that there exists a set $B$ and some adjacent databases $X, X'$ such that $f_X(b) > f_{X'}(b)\exp(\epsilon)$ for all $b \in B$ and that $\nu(B) > 0$. Then this would imply the strict inequality

$$\mu_X(B) = \int_B f_X(b)\, d\nu(b) > \exp(\epsilon) \int_B f_{X'}(b)\, d\nu(b) = \exp(\epsilon)\mu_{X'}(B),$$

which is a contradiction, and thus the claim holds. □

*Proof of Theorem 3.2.* The density of the exponential mechanism can be expressed as

$$f_X(b) = c_n^{-1} g(b) \exp\left\{\frac{\epsilon}{2\Delta}\xi_X(b)\right\},$$

where $c_n$ is the normalizing constant. Define the random variable $Z = \sqrt{n}(\tilde{b} - \hat{b})$, then its density is given by

$$f_n(z) = c_n^{-1} n^{-1/2} g(\hat{b} + z/\sqrt{n}) \exp\left\{\frac{\epsilon}{2\Delta}\xi_n(\hat{b} + z/\sqrt{n})\right\}.$$

We now aim to show that, for $z$ fixed, the density converges to a multivariate normal. Using a two term Taylor expansion, we have by Assumption (2) and (3) that

$$\xi_X(\hat{b} + z/\sqrt{n}) = [\xi_X(\hat{b}) + z^\top \xi_X'(\hat{b})/\sqrt{n} + z^\top \xi_X''(\hat{b}) z/2n] + o(1).$$

The first term will be absorbed into the constants, since it does not depend on $z$, while the second term is zero, since $\hat{b}$ minimizes $\xi_X$. So, only the third term contributes to the form of the density. Obviously $|g(\hat{b} + z/\sqrt{n}) - g(b^\star)| \to 0$, so the only remaining task is to show that the combined constants behave appropriately. The integrating constant is of the form

$$c_n n^{1/2} \exp\left\{-\frac{\epsilon}{2\Delta}\xi_n(\hat{b})\right\} = \int_{B_n} g(\hat{b} + z/\sqrt{n}) \exp\left\{\frac{\epsilon}{2\Delta}[\xi_n(\hat{b} + z/\sqrt{n}) - \xi_n(\hat{b})]\right\} dz.$$

By Assumption (1) we have that

$$\xi_X(\hat{b} + z/\sqrt{n}) - \xi_n(\hat{b}) \leq -\frac{\alpha}{2}\|z\|^2.$$

Since $\exp\{-\|z\|^2\}$ is integrable, we can apply the dominated convergence theorem to conclude that the constants converge to something nonzero as well.

Putting everything together, we conclude that

$$f_n(z) \to f(z) \propto \exp\left\{-\frac{\epsilon}{2\Delta} z^\top \Sigma^{-1} z/2\right\},$$

which is the density of the multivariate normal. Applying Scheffe's Theorem, we thus have both convergence in distribution as well as convergence in total variation:

$$\sqrt{n}(\tilde{b} - \hat{b}) \xrightarrow{D} N_p\left(0, \frac{2\Delta}{\epsilon}\Sigma\right).$$ □

[1]Department of Statistics, Pennsylvania State University, University Park, Pennsylvania. Correspondence to: Jordan Awan <awan@psu.edu>.

*Proceedings of the $36^{th}$ International Conference on Machine Learning*, Long Beach, California, PMLR 97, 2019. 

**Lemma 0.1.** *Suppose that $\Sigma$ and $C$ are nuclear positive-definite operators on $\mathcal{H}$ such that $\Sigma^{-1}C$ is Hilbert-Schmidt. Then $C^{1/2}\Sigma^{-1}C^{1/2}$ and $\Sigma^{-1/2}C\Sigma^{-1/2}$ are also Hilbert-Schmidt.*

*Proof.* Recall that $\Sigma^{-1}C$ is Hilbert-Schmidt is equivalent to $\|\Sigma^{-1}C\|_{HS} = \|C\Sigma^{-1}\|_{HS} < \infty$. Then

$$\begin{aligned}\infty &> \|\Sigma^{-1}C\|_{HS} \cdot \|C\Sigma^{-1}\|_{HS} \\ &\geq \langle C\Sigma^{-1}, \Sigma^{-1}C\rangle_{HS} \\ &= \operatorname{tr}\left(\Sigma^{-1}C\Sigma^{-1}C\right) \\ &= \operatorname{tr}\left(C^{1/2}\Sigma^{-1}C^{1/2}C^{1/2}\Sigma^{-1}C^{1/2}\right) \\ &= \|C^{1/2}\Sigma^{-1}C^{1/2}\|_{HS}^2,\end{aligned}$$

which implies that $C^{1/2}\Sigma^{-1}C^{1/2}$ is Hilbert-Schmidt. The same trick works for $\Sigma^{-1/2}C\Sigma^{-1/2}$. □

**Lemma 0.2.** *In the setting of Theorem 3.3, let $\Sigma$ and $C$ be nuclear positive-definite operators on $\mathcal{H}$ such that $\Sigma^{-1}C$ is Hilbert-Schmidt (with respect to the inner product of $\mathscr{H}$, $\Sigma^{-1}C$ is bounded with respect to the Cameron-Martin space (CMS) of $C$, and $\hat{b}_n$ lies in the CMS of $C$ for all $n$. Then*

1. *The Gaussian process on $\mathcal{H}$ with mean $\sqrt{n}\left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + \frac{1}{n}C^{-1}\right)^{-1}\left(\frac{1}{n}C^{-1}\right)\hat{b}$ and covariance $\left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + \frac{1}{n}C^{-1}\right)^{-1}$ is equivalent, as probability measures, to a Gaussian process with mean $-\sqrt{n}\hat{b}$ and covariance $nC$.*
2. *$\left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + \frac{1}{n}C^{-1}\right)^{-1}$ converges to $\Sigma$ in the space of nuclear operators.*
3. *$-n^{-1/2}\left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + \frac{1}{n}C^{-1}\right)^{-1}\left(\frac{1}{n}C^{-1}\right)\hat{b} \to 0$ in $\mathcal{H}$.*

*Proof.* 1. We first check that the covariances will induce equivalent measures (Corollary 6.4.11 Bogachev, 1998). Namely we first require that

$$\left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + \frac{1}{n}C^{-1}\right)^{1/2}(nC)^{1/2}$$

is invertible and bounded. This can be written in the form

$$\left(\frac{\epsilon n}{2\Delta}\Sigma^{-1}C + I\right)^{1/2}.$$

Since $\Sigma^{-1}C$ is Hilbert-Schmidt and $I$ is bounded, the combined quantity is bounded. Furthermore, the smallest eigenvalue is $\geq 1$, so it is invertible.

Second, we check that

$$(nC)^{-1/2}\left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + \frac{1}{n}C^{-1}\right)^{-1}(nC)^{-1/2} - I$$

is Hilbert-Schmidt. This can be rearranged as follows:

$$\begin{aligned}&\left(\frac{\epsilon n}{2\Delta}C^{1/2}\Sigma^{-1}C^{1/2} + I\right)^{-1} - I \\ &= \left(\frac{\epsilon n}{2\Delta}C^{1/2}\Sigma^{-1}C^{1/2} + I\right)^{-1}\left[\frac{-\epsilon n}{2\Delta}C^{1/2}\Sigma^{-1}C^{1/2}\right]\end{aligned}$$

At this point, we recall that $\Sigma^{-1}C$ being Hilbert-Schmidt implies that $C^{1/2}\Sigma^{-1}C^{1/2}$ is Hilbert Schmidt, by Lemma 0.1. So, we have a bounded operator multiplied by a Hilbert-Schmidt operator, which shows that the result is Hilbert-Schmidt.

Third and last, we verify that

$$\sqrt{n}\left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + \frac{1}{n}C^{-1}\right)^{-1}\left(\frac{1}{n}C^{-1}\right)\hat{b} - \sqrt{n}\hat{b}$$

lies in the CMS of $C$. We can express this difference as follows:

$$\begin{aligned}&\sqrt{n}\left[\left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + \frac{1}{n}C^{-1}\right)^{-1}\left(\frac{1}{n}C^{-1}\right) - I\right]\hat{b} \\ &= \sqrt{n}\left[\left(\frac{\epsilon n}{2\Delta}\Sigma^{-1}C + I\right)^{-1} - I\right]\hat{b} \\ &= \sqrt{n}\left(\frac{\epsilon n}{2\Delta}\Sigma^{-1}C + I\right)^{-1}\left[I - \left(\frac{\epsilon n}{2\Delta}\Sigma^{-1}C + I\right)\right]\hat{b} \\ &= -\sqrt{n}\left(\frac{\epsilon n}{2\Delta}\Sigma^{-1}C + I\right)^{-1}\left(\frac{\epsilon n}{2\Delta}\Sigma^{-1}C\right)\hat{b}.\end{aligned}$$

From this representation, we see that $-\sqrt{n}\left(\frac{\epsilon n}{2\Delta}\Sigma^{-1}C + I\right)^{-1}$ is a bounded operator, since $\Sigma^{-1}C$ is Hilbert-Schmidt. So, it suffices to show that

$$\langle \Sigma^{-1}C\hat{b}, C^{-1}\Sigma^{-1}C\hat{b}\rangle < \infty.$$

Equivalently, we may show that

$$\|\Sigma^{-1}C\hat{b}\|_C < \infty,$$

where $\|\cdot\|_C$ is the norm of the CMS of $C$. Since $\Sigma^{-1}C$ is bounded in the CMS of $C$, and since $\hat{b}$ lies in the CMS of $C$, the result holds.

2. Since $\Sigma^{-1/2}C\Sigma^{-1/2}$ a is symmetric, positive definite, and Hilbert-Schmidt there exists an orthonormal sequence $(u_i)_{i=1}^{\infty}$ in $\mathcal{H}$ and a sequence of real numbers $a_i \in \mathbb{R}^+$ such that

$$\Sigma^{-1/2}C\Sigma^{-1/2} = \sum_{i=1}^{\infty} a_i u_i \otimes u_i \qquad \sum_{i=1}^{\infty} a_i^2 < \infty.$$

Then

$$\begin{aligned}&\left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + \frac{1}{n}C^{-1}\right)^{-1} \\ &= \frac{2\Delta}{\epsilon}\Sigma^{1/2}\left(I + \frac{2\Delta}{n\epsilon}\Sigma^{1/2}C^{-1}\Sigma^{1/2}\right)^{-1}\Sigma^{1/2}.\end{aligned}$$

Using the eigen decomposition with the $u_i$ we have the inside term is given by

$$\begin{aligned}&\left(I + \frac{2\Delta}{n\epsilon}\Sigma^{1/2}C^{-1}\Sigma^{1/2}\right)^{-1}\\ &= \sum_{i=1}^{\infty}\left(1 + \frac{2\Delta a_i^{-1}}{n\epsilon}\right)^{-1} u_i \otimes u_i\\ &= \sum_{i=1}^{\infty}\frac{a_i}{a_i + 2\Delta/(n\epsilon)}u_i \otimes u_i.\end{aligned}$$

So then we can express the difference

$$\begin{aligned}&\frac{2\Delta}{\epsilon}\Sigma - \left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + \frac{1}{n}C^{-1}\right)^{-1}\\ &= \frac{2\Delta}{\epsilon}\sum_{i=1}^{\infty}\frac{2\Delta}{n\epsilon a_i + 2\Delta}\Sigma^{1/2} \circ (u_i \otimes u_i) \circ \Sigma^{1/2}.\end{aligned}$$

Notice that $\Sigma = \sum_{i=1} \Sigma^{1/2} \circ (u_i \otimes u_i) \circ \Sigma^{1/2}$, since $\sum u_i \otimes u_i$ is just the identity operator. Thus, since $\Sigma$ is nuclear, for any $\delta > 0$, we can choose $m$ such that

$$\left\|\frac{2\Delta}{\epsilon}\sum_{i=m+1}^{\infty}\frac{2\Delta}{n\epsilon a_i + 2\Delta}\Sigma^{1/2} \circ (u_i \otimes u_i) \circ \Sigma^{1/2}\right\| \leq \frac{\delta}{2},$$

in the nuclear norm. Finally, now that the sum is finite, we can choose $n$ such that

$$\left\|\frac{2\Delta}{\epsilon}\sum_{i=1}^{m}\frac{2\Delta}{n\epsilon a_i + 2\Delta}\Sigma^{1/2} \circ (u_i \otimes u_i) \circ \Sigma^{1/2}\right\| \leq \frac{\delta}{2},$$

as desired.

3. To see the convergence of

$$-n^{-1/2}\left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + \frac{1}{n}C^{-1}\right)^{-1}\left(\frac{1}{n}C^{-1}\right)\hat{b},$$

note that the largest (absolute) singular value of $\left(\frac{\epsilon n}{2\Delta}\Sigma^{-1}C + I\right)^{-1}$ is upper bounded by 1. So,

$$\begin{aligned}&\left\|-n^{-1/2}\left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + \frac{1}{n}C^{-1}\right)^{-1}\left(\frac{1}{n}C^{-1}\right)\hat{b}\right\|\\ &= \left\|n^{1/2}\left(\frac{\epsilon n}{2\Delta}\Sigma^{-1}C + I\right)^{-1}\hat{b}\right\|\\ &\leq \frac{1}{\sqrt{n}}\|\hat{b}\|\\ &\to 0.\end{aligned}$$

□

*Proof of Theorem 3.3.* The proof strategy is the same as for Theorem 3.2. However since the base measure is no longer Lebesgue, the effect of changing variables on the Gaussian base measure must be handled more carefully. Consider $Z = \sqrt{n}(\tilde{b} - \hat{b})$ and

$$\begin{aligned}P(\sqrt{n}(\tilde{b} - \hat{b}) \in A) &= \int_{\hat{b} + A/\sqrt{n}} f_X(b)\, d\nu(b)\\ &= n^{-1/2}\int_A f_X(\hat{b} + z/\sqrt{n})\, d\nu(\hat{b} + z/\sqrt{n}).\end{aligned}$$

The same Taylor expansion arguments from before still apply, however the base measure has now been shifted and scaled. In particular, if $d\tilde{\nu}(z) = d\nu(\hat{b} + z/\sqrt{n})$, then $\tilde{\nu}$ is the measure of a Gaussian process with mean $-\sqrt{n}\hat{b}$ and covariance operator $nC$. So we have that

$$\begin{aligned}&P(\sqrt{n}(\tilde{b} - \hat{b}) \in A)\\ &= c_n^{-1}\int_A \exp\left\{-\left\langle z, \frac{\epsilon}{2\Delta}\Sigma^{-1}z\right\rangle/2\right\}\, d\tilde{\nu}(z) + o(1),\end{aligned} \tag{1}$$

where $c_n$ is the normalizing constant. However, this is a Gaussian measure with covariance operator $(\frac{\epsilon}{2\Delta}\Sigma^{-1} + C^{-1}/n)^{-1}$ and mean $-n^{-1/2}(\frac{\epsilon}{2\Delta}\Sigma^{-1} + C^{-1}/n)^{-1}C^{-1}\hat{b}$. By part 1 of Lemma 0.2, we know that this Gaussian process is equivalent to $\tilde{v}$, which is a Gaussian process with mean $-\sqrt{n}\hat{b}$ and covariance $nC$, meaning that the density in (1) is well defined. By parts 2 and 3 of Lemma 0.2, we have that the following limits hold:

$$\left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + C^{-1}/n\right)^{-1} \to \frac{2\Delta}{\epsilon}\Sigma$$

$$-n^{-1/2}\left(\frac{\epsilon}{2\Delta}\Sigma^{-1} + C^{-1}/n\right)^{-1}C^{-1}\hat{b} \to 0,$$

where the first limit is in the space of nuclear operators, implying the sequence of measures is tight, and the second limit occurs in $\mathbb{H}$. We conclude that the characteristic functions of the measures converge and, since the sequence is also tight, this implies that $\sqrt{n}(\tilde{b} - \hat{b})$ converges in distribution to the specified Gaussian process. □

*Proof of Theorem 4.1.* Recall that we assumed that $\|X_i\| \leq 1$ for all $i = 1, \ldots, n$. So, $\|PX_i\|^2 \leq \|X_i\|^2 \leq 1$ for any $P \in \mathscr{P}_k$ and any $i = 1, \ldots, n$. Since we also have that $\|PX_i\|^2 \geq 0$, we see that $\Delta_\xi = 1$. Since $\sum_{i=1}^n \|PX_i\|^2 \leq n$, we have that $\exp\left(-\frac{\epsilon}{2\Delta}\sum_{i=1}^n \|PX_i\|^2\right)$ is a valid density with respect to any probability measure in $\mathscr{P}_k$. By Proposition 3.1, the mechanism $\mathscr{M}$ satisfies $\epsilon$-DP. □

*Proof of Theorem 4.2.* The proof is essentially the same as for Theorem 4.1. □

## References

Bogachev, V. I. *Gaussian measures*. Number 62. American Mathematical Soc., 1998.

Hall, R., Rinaldo, A., and Wasserman, L. Differential privacy for functions and functional data. *Journal of Machine Learning Research*, 14(1):703–727, February 2013. ISSN 1532-4435.