A note on exponential Efron-Stein inequalities

July 6, 2017

1 Introduction

For bounded functions of independent variables I give an entropy bound (Theorem 2 below) in terms of the operator V^{+} introduced in [1] together with some corollaries which slightly improve over some - now classical - results in the theory of concentration inequalities. I also improve on the recent Bernstein-type inequality in [6].

2 A bound on the thermal variance

Let $(\Omega, \mathcal{M}, \mu)$ be a probability space and $\mathcal{A}(\Omega)$ the algebra of bounded, measurable real valued functions on Ω. For $f \in \mathcal{A}(\Omega)$ and $\beta \in \mathbb{R}$ we define the thermal measure $\mu_{\beta f} = e^{\beta f} d\mu / E\left[e^{\beta f}\right]$, and the corresponding functionals of thermal expectation $E_{\beta f}[.]$ and thermal variance $\sigma^2_{\beta f}[.]$. We prove the

Lemma 1 *Let $0 \leq s \leq \beta$. Then*

$$\sigma^2_{sf}(f) \leq E_{x \sim \mu_{\beta f}}\left[E_{x' \sim \mu}\left[\left(f(x) - f(x')\right)^2_{+}\right]\right].$$

Proof. Let ψ be any real function. By direct computation

$$\frac{d}{d\beta} E_{\beta f}[\psi(f)] = E_{\beta f}[\psi(f) f] - E_{\beta f}[\psi(f)] E_{\beta f}[f]. \tag{1}$$

By Chebychev's association inequality $E_{\beta f}[\psi(f)]$ is nonincreasing (nondecreasing) in β if ψ is nonincreasing (nondecreasing). Now define $g : \mathbb{R}^2 \to \mathbb{R}$ by

$$g(s,t) = E_{x \sim \mu_{sf}}\left[E_{x' \sim \mu_{tf}}\left[\left(f(x) - f(x')\right)^2 1_{f(x) \geq f(x')}\right]\right],$$

so that

$$\sigma^2_{sf}(f) = \frac{1}{2} E_{x \sim \mu_{sf}}\left[E_{x' \sim \mu_{sf}}\left[\left(f(x) - f(x')\right)^2\right]\right] = g(s,s).$$

Now for fixed x the function $\left(f(x) - f(x')\right)^2 1_{f(x) \geq f(x')}$ is nonincreasing in $f(x')$, so $g(s,t)$ is nonincreasing in t. On the other hand, for fixed x', $\left(f(x) - f(x')\right)^2 1_{f(x) \geq f(x')}$

is nondecreasing in $f(x)$, so $g(s,t)$ is nondecreasing in s (this involves exchanging the two expectations in the definition of $g(s,t)$). So, since $\mu_{0f}=\mu$, we get from $0\leq s\leq\beta$ that

$$\sigma_{sf}^{2}(f)=g(s,s)\leq g(\beta,0)=E_{x\sim\mu_{\beta f}}\left[E_{x'\sim\mu}\left[\left(f(x)-f(x')\right)_{+}^{2}\right]\right].$$

■

Here is another way to write the conclusion: let $h\in\mathcal{A}(\Omega)$ be defined by $h(x)=E_{x'\sim\mu}\left[\left(f(x)-f(x')\right)_{+}^{2}\right]$. Then $\sigma_{sf}^{2}(f)\leq E_{\beta f}[h]$.

3 Some background material

The contents of this section are explained in more detail in [5]. Let $(\Omega,\mathcal{M},\mu)=\prod_{k=1}^{n}(\Omega_k,\mathcal{M}_k,\mu_k)$ be a product of probability spaces. For $k\in\{1,...,n\}$ and $y\in\Omega_k$ we define the substitution operator S_y^k on $\mathcal{A}(\Omega)$ by

$$\left(S_y^k f\right)(x_1,...,x_n)=f(x_1,...,x_{k-1},y,x_{k+1},...,x_n).$$

The conditional expectation operator E_k is defined by

$$E_k f=\int_{\Omega_k}S_y^k f\ d\mu_k.$$

For $\beta\in\mathbb{R}$ and $f\in\mathcal{A}(\Omega)$ and $k\in\{1,...,n\}$ the conditional thermal measure is $\mu_{k,\beta f}=e^{\beta f}d\mu_k/E_k\left[e^{\beta f}\right]$ and the conditional thermal expectations $E_{k,\beta f}[.]$ and variances $\sigma_{k,\beta f}^{2}$ are defined correspondingly. The entropy $S_f(\beta)$ of f at β is given by

$$S_f(\beta)=\beta E_{\beta f}[f]-\ln E\left[e^{\beta f}\right].$$

Again the conditional entropy $S_{k,f}(\beta)$ is the analogous member of $\mathcal{A}(\Omega)$, where the expectation E is replaced by E_k. The following three identities are obtained from straightforward computations (see [5])

$$\ln E\left[e^{\beta(f-Ef)}\right] = \beta\int_0^\beta\frac{S_f(\gamma)}{\gamma^2}d\gamma \tag{2}$$

$$S_f(\beta) = \int_0^\beta\int_t^\beta\sigma_{sf}^{2}(f)\,ds\ dt \tag{3}$$

$$S_{k,f}(\beta) = \int_0^\beta\int_t^\beta\sigma_{k,sf}^{2}(f)\,ds\ dt. \tag{4}$$

We also have the well known thermal subadditivity of entropy

$$S_f(\beta)\leq E_{\beta f}\left[\sum_{k=1}^{n}S_{k,f}(\beta)\right],$$

which, together with (4) gives

$$S_f(\beta)\leq E_{\beta f}\left[\sum_{k=1}^{n}\int_0^\beta\int_t^\beta\sigma_{k,sf}^{2}(f)\,ds\ dt\right]. \tag{5}$$

4 Exponential Efron Stein inequalities

Define two operators D and V^+ on $\mathcal{A}(\Omega)$ by

$$\begin{aligned} D(f) &= \sum_k \left(f - \inf_{y\in\Omega_k} S_y^k f \right)^2 \\ V^+(f) &= \sum_k E_{y\sim\mu_k}\left[\left(f - S_y^k f \right)_+^2 \right]. \end{aligned}$$

Clearly we have $V^+(f) \leq D(f)$ for any $f \in \mathcal{A}(\Omega)$.

Theorem 2 *For $\beta > 0$*

$$S_f(\beta) \leq \frac{\beta^2}{2} E_{\beta f}\left[V^+(f) \right].$$

Proof. For $k \in \{1, ..., n\}$ write $h_k = E_{y\sim\mu_k}\left[\left(f - S_y^k f \right)_+^2 \right]$, so that $V^+(f) = \sum_k h_k$. The conditional version of Lemma 1 then reads for $0 \leq s \leq \beta$ and $k \in \{1, ..., n\}$

$$\sigma_{k,sf}^2(f) \leq E_{k,\beta f}[h_k].$$

Substitution in (5) gives

$$\begin{aligned} S_f(\beta) &\leq \int_0^\beta \int_t^\beta \sum_k E_{\beta f}\left[\sigma_{k,sf}^2(f) \right] ds dt \\ &\leq \int_0^\beta \int_t^\beta \sum_k E_{\beta f}\left[E_{k,\beta f}[h_k] \right] ds dt \\ &= \int_0^\beta \int_t^\beta \sum_k E_{\beta f}[h_k] ds dt \\ &= \frac{\beta^2}{2} E_{\beta f}\left[V^+(f) \right], \end{aligned}$$

where we used the identity $E_{\beta f}\left[E_{k,\beta f}[h] \right] = E_{\beta f}[h]$ for $h \in \mathcal{A}(\Omega)$. ■

Spelling this out for comparison with Proposition 10 in [1] gives for $\beta \geq 0$

$$\beta E\left[f e^{\beta f} \right] - E\left[e^{\beta f} \right] \ln\left[e^{\beta f} \right] \leq \frac{\beta^2}{2} E\left[e^{\beta f} V^+(f) \right].$$

In the sequel some corollaries are given.

Corollary 3

$$\Pr\{f - Ef > t\} \leq \exp\left(\frac{-t^2}{2\left\| V^+(f) \right\|_\infty} \right).$$

Proof. By (2)

$$\ln E\left[e^{\beta(f-Ef)}\right]=\beta\int_0^\beta\frac{S_f(\gamma)}{\gamma^2}d\gamma\leq\frac{\beta}{2}\int_0^\beta E_{\gamma f}\left[V^+(f)\right]d\gamma\leq\frac{\beta^2}{2}\left\|V^+(f)\right\|_\infty.$$

The result then follows from a straightforward application of the exponential moment method. ■

This corollary improves on Theorem 1 (1) in [4] by using the tighter functional $\|V^+(.)\|_\infty$ instead of $\|D(f)\|_\infty$, and it improves the exponent in Corollary 3 in [1] by a factor of 2. In a similar way the following improves on Theorem 13 (1) in [4] (and Theorem 6.19 in [3]) and on Theorem 5 in [1].

Corollary 4 *Suppose there are positive constants a and b such that*

$$V^+(f)\leq af+b.$$

Then

$$\begin{aligned}\ln Ee^{\beta f} &\leq \frac{\beta E[f]}{1-\frac{1}{2}a\beta}+\frac{\beta^2 b/2}{1-\frac{1}{2}a\beta}\\ \textit{and}\ \Pr\{f-Ef>t\} &\leq \exp\left(\frac{-t^2}{2aE[f]+2b+at}\right).\end{aligned}$$

Proof. We start by bounding the log moment generating function as above

$$\begin{aligned}\ln Ee^{\beta(f-Ef)} &\leq \frac{\beta}{2}\int_0^\beta E_{\gamma f}\left[V^+(f)\right]d\gamma\leq\frac{a\beta}{2}\int_0^\beta E_{\gamma f}[f]\,d\gamma+\frac{\beta^2}{2}b\\ &= \frac{a\beta}{2}\ln Ee^{\beta f}+\frac{\beta^2}{2}b\\ &= \frac{a\beta}{2}\ln Ee^{\beta(f-Ef)}+\frac{\beta^2}{2}(aE[f]+b).\end{aligned}$$

Rearrangement gives for $\beta\in(0,2/a)$

$$\ln Ee^{\beta(f-Ef)}\leq\frac{\beta^2}{1-\frac{1}{2}a\beta}\left(\frac{a}{2}E[f]+\frac{b}{2}\right).$$

This implies the first conclusion and gives the second one by proceeding as in the proof of Theorem 13 in [4]. ■

Next I apply the V_+ bounds to the suprema of empirical processes. The proof uses the inequality

$$E_{\beta f}[g]\leq S_f(\beta)+\ln E\left[e^g\right], \tag{6}$$

which can be derived from Jensen's inequality.

Theorem 5 *Let $X_1, ..., X_n$ be independent with values in $\mathcal{X}$ with X_i distributed as μ_i, and let $\mathcal{F}$ be a finite class of functions $f : \mathcal{X} \to [-1, 1]$ with $E[f(X_i)] = 0$. Define $F : \mathcal{X}^n \to \mathbb{R}$ and $W : \mathcal{X}^n \to \mathbb{R}$ by*

$$\begin{aligned} F(\mathbf{x}) &= \sup_{f\in\mathcal{F}} \sum_i f(x_i) \quad \textit{and} \\ W(\mathbf{x}) &= \sup_{f\in\mathcal{F}} \sum_i \left(f^2(x_i) + E\left[f^2(X_i)\right]\right). \end{aligned}$$

Then for $t > 0$

$$\Pr\{F - E[F] > t\} \leq \exp\left(\frac{-t^2}{2E[W] + t}\right).$$

This improves over Theorem 12.2 in [3], since by the triangle inequality $E[W] \leq \Sigma^2 + \sigma^2$ and the constants in the denominator of the exponent are better by a factor of two, and optimal for the variance term. Furthermore the proof is more economical and elementary, relying exclusively on the LSI of Theorem 2.

Proof. Let $0 < \gamma \leq \beta < 2$. Using Theorem 2 and (6) we get

$$S_F(\gamma) \leq \frac{\gamma^2}{2\gamma} E_{\gamma F}\left[\gamma V^+(F)\right] \leq \frac{\gamma}{2}\left[S_F(\gamma) + \ln E e^{\gamma V^+(F)}\right].$$

Rearranging gives

$$S_F(\gamma) \leq \frac{\gamma}{2-\gamma} \ln E e^{\gamma V^+(F)}. \tag{7}$$

Fix some $\mathbf{x} \in \mathcal{X}^n$ and let $\hat{f} \in \mathcal{F}$ witness the maximum in the definition of $F(\mathbf{x})$. For $y \in \mathcal{X}$ we have $\left(F - S_y^k F\right)_+ \leq \left(\hat{f}(x_i) - \hat{f}(y)\right)_+$ and by the zero mean assumption

$$\begin{aligned} V_+(F)(\mathbf{x}) &= \sum_k E_{y\sim\mu_k}\left[\left(F(\mathbf{x}) - S_y^k F(\mathbf{x})\right)_+^2\right] \\ &\leq \sum_k E_{y\sim\mu_k}\left(\hat{f}(x_k) - \hat{f}(y)\right)_+^2 \\ &\leq \sum_k E_{y\sim\mu_k}\left(\hat{f}(x_k) - \hat{f}(y)\right)^2 \\ &= \sum_k \left(\hat{f}^2(x_k) + E\left[\hat{f}^2(X_k)\right]\right) \\ &\leq W(\mathbf{x}). \end{aligned}$$

So $V_+(F) \leq W$. Now let $\hat{f} \in \mathcal{F}$ (different from the previous $\hat{f}$, which we don't need any more) witness the maximum in the definition of $W(\mathbf{x})$. Then

$$\begin{aligned} V_+(W)(\mathbf{x}) &= \sum_k E_{y\sim\mu_k}\left(W(\mathbf{x}) - S_y^k W(\mathbf{x})\right)_+^2 \\ &\leq \sum_k E_{y\sim\mu_k}\left[\left(\hat{f}^2(x_k) - \hat{f}^2(y)\right)_+^2\right] \\ &\leq \sum_k \hat{f}^2(x_k) \\ &\leq W. \end{aligned}$$

If follows from (7), the fact that $V_+(F) \leq W$ and Corollary 4 above, that

$$\begin{aligned} S_F(\gamma) &\leq \frac{\gamma}{2-\gamma}\ln E e^{\gamma V^+(F)} \\ &\leq \frac{\gamma}{2-\gamma}\ln E\left[e^{\gamma W}\right] \\ &\leq \frac{\gamma}{2-\gamma}\left(\frac{\gamma E[W]}{1-\gamma/2}\right) \\ &= \frac{\gamma^2}{(1-\gamma/2)^2}\frac{E[W]}{2}. \end{aligned}$$

From the bound on the log moment generating function (2) we conclude that

$$\begin{aligned} \ln E e^{\beta(F-EF)} &= \beta\int_0^\beta \frac{S_F(\gamma)}{\gamma^2}d\gamma \leq \beta\int_0^\beta \frac{1}{(1-\gamma/2)^2}d\gamma\ \frac{E[W]}{2} \\ &= \frac{\beta^2}{1-\beta/2}\frac{E[W]}{2}. \end{aligned}$$

Using Lemma 12 in [4] it follows that

$$\begin{aligned} \Pr\{F - E[F] > t\} &\leq \inf_{\beta\in(0,2)}\exp\left(-\beta t + \frac{\beta^2}{1-\beta/2}\frac{E[W]}{2}\right) \\ &\leq \exp\left(\frac{-t^2}{2E[W]+t}\right). \end{aligned}$$

■

5 Softening the interaction functional

Another application of Theorem 2 is a subtle improvement of the interaction functional used in the Bernstein-type inequality in [6]. For $f \in \mathcal{A}(\Omega)$ define

$$J_\mu^+(f) = 2\left(\sup_{\mathbf{x}\in\Omega}\sum_l E_{z\sim\mu_l}\left[\sum_{k:k\neq l}\sigma_k^2\left(f - S_z^l f\right)1_{A_l}(z)\right]\right)^{1/2},$$

where $A_l = A_l(\mathbf{x})$ is the subset of Ω_l defined by

$$A_l = \left\{ z \in \Omega_l : S_z^l \Sigma^2(f) \leq \Sigma^2(f) \right\}.$$

A_l is a set-valued function depending on $\mathbf{x} \in \Omega$. Clearly $J_\mu^+(f) \leq J_\mu(f)$ for any f.

The modification works as follows. Thanks to Theorem 2 the operator D can simply be replaced by V^+ in Lemma 9, Lemma 10 and Proposition 14 in [6]. Proposition 15 in [6] then has to be replaced by the following.

Proposition 6 *We have* $V^+\left(\Sigma^2(f)\right) \leq J_\mu^+(f)^2 \; \Sigma^2(f)$ *for any* $f \in \mathcal{A}(\Omega)$.

Proof. Fix $\mathbf{x} \in \Omega$. For any $z \in \Omega_l$

$$S_z^l \Sigma^2(f) = \sum_k S_z^l \sigma_k^2(f) = \sigma_l^2(f) + \sum_{k:k \neq l} S_z^l \sigma_k^2(f),$$

where we used the fact that $S_{z_l}^l \sigma_l^2(f) = \sigma_l^2(f)$, because $\sigma_l^2(f) \in \mathcal{A}_l(\Omega)$. Then

$$\begin{aligned}
V^+\left(\Sigma^2(f)\right) &= \sum_l E_{z\sim\mu_l}\left[\left(\Sigma^2(f) - S_z^l\Sigma^2(f)\right)^2 1_{A_l}(z)\right] \\
&= \sum_l E_{z\sim\mu_l}\left[\left(\sum_k \sigma_k^2(f) - \sigma_l^2(f) - \sum_{k:k\neq l} S_z^l\sigma_k^2(f)\right)^2 1_{A_l}(z)\right] \\
&= \sum_l E_{z\sim\mu_l}\left[\left(\sum_{k:k\neq l}\left(\sigma_k^2(f) - S_z^l\sigma_k^2(f)\right)\right)^2 1_{A_l}(z)\right].
\end{aligned}$$

Using $2\sigma_k^2(f) = E_{(y,y')\sim\mu_k^2}\left(D_{y,y'}^k f\right)^2$ we get, similar to [6],

$$\begin{aligned}
&4V^+\left(\Sigma^2(f)\right) \\
&= \sum_l E_{z\sim\mu_l}\left[\left(\sum_{k:k\neq l} E_{(y,y')\sim\mu_k^2}\left(D_{y,y'}^k f\right)^2 - S_z^l E_{(y,y')\sim\mu_k^2}\left(D_{y,y'}^k f\right)^2\right)^2 1_{A_l}(z)\right] \\
&= \sum_l E_{z\sim\mu_l}\left[\left(\sum_{k\neq l} E_{(y,y')\sim\mu_k^2}\left[\left(D_{y,y'}^k f - D_{y,y'}^k S_z^l f\right)\left(D_{y,y'}^k f + D_{y,y'}^k S_z^l f\right)\right]\right)^2 1_{A_l}(z)\right] \\
&\leq \sum_l E_{z\sim\mu_l}\left[\sum_{k:k\neq l} E_{(y,y')\sim\mu_k^2}\left[D_{y,y'}^k\left(f - S_z^l f\right)\right]^2\right. \\
&\qquad\qquad \left.\times \sum_{k:k\neq l} E_{(y,y')\sim\mu_k^2}\left[D_{y,y'}^k f + D_{y,y'}^k S_z^l f\right]^2 1_{A_l}(z)\right]
\end{aligned}$$

by Cauchy-Schwarz. We use Hölder's inequality to bound this by

$$\sum_l E_{z\sim\mu_l}\left[\sum_{k:k\neq l} E_{(y,y')\sim\mu_k^2}\left[D^k_{y,y'}\left(f-S^l_z f\right)\right]^2 1_{A_l}(z)\right]\times$$
$$\times\sup_{z\in A_l}\sum_{k:k\neq l} E_{(y,y')\sim\mu_k^2}\left[D^k_{y,y'}f+D^k_{y,y'}S^l_z f\right]^2$$

We then bound the supremum by

$$\begin{aligned}
&\sup_{z\in A_l}\sum_{k:k\neq l} E_{(y,y')\sim\mu_k^2}\left[D^k_{y,y'}f+D^k_{y,y'}S^l_z f\right]^2\\
\leq\ &\sup_{z\in A_l}\sum_{k:k\neq l} E_{(y,y')\sim\mu_k^2}\left[2\left(D^k_{y,y'}f\right)^2+2\left(D^k_{y,y'}S^l_z f\right)^2\right]\\
=\ &4\sum_{k:k\neq l}\sigma_k^2(f)+4\sup_{z\in A_l}S^l_z\sum_{k:k\neq l}\sigma_k^2(f)\\
\leq\ &4\left(\Sigma^2(f)+\sup_{z\in A_l}S^l_z\Sigma^2(f)\right)\leq 8\Sigma^2(f),
\end{aligned}$$

where the last inequality follows from the definition of A_l. To conclude

$$\begin{aligned}
V^+\left(\Sigma^2(f)\right) &\leq 2\sum_l E_{z\sim\mu_l}\left[\sum_{k:k\neq l} E_{(y,y')\sim\mu_k^2}\left[D^k_{y,y'}\left(f-S^l_z f\right)\right]^2 1_{A_l}(z)\right]\Sigma^2(f)\\
&\leq 4\sup_{\mathbf{x}\in\Omega}\sum_l E_{z\sim\mu_l}\left[\sum_{k:k\neq l}\sigma_k^2\left(f-S^l_{z_l}f\right)1_{A_l}(z)\right]\Sigma^2(f)\\
&= \left(J_\mu^+\right)^2(f)\,\Sigma^2(f).
\end{aligned}$$

■

Substitution in the appropriately modified Proposition 14 of [6] then gives the main result in [6] with J_μ replaced by J_μ^+.

References

[1] S.Boucheron,G.Lugosi,P.Massart, Concentration Inequalities using the entropy method, *Annals of Probability* 31, Nr 3, 2003

[2] S.Boucheron, G.Lugosi, P.Massart, On concentration of self-bounding functions, *Electronic Journal of Probability* Vol.14 (2009), Paper no. 64, 1884–1899, 2009

[3] S. Boucheron, G. Lugosi, P. Massart. Concentration Inequalities, Oxford University Press (2013)

[4] A.Maurer. Concentration inequalities for functions of independent variables. *Random Structures and Algorithms* 29: 121–138, 2006

[5] A.Maurer, Thermodynamics and concentration. *Bernoulli* 18.2 (2012): 434-454.

[6] Maurer, A. (2017). A Bernstein-type inequality for functions of bounded interaction. arXiv preprint arXiv:1701.06191.