Improved concentration theorem for projections

Daniel Hsu <djhsu@cs.columbia.edu>

Let $\mathbf{x}$ be a random vector taking values in $\mathbb{R}^D$ with mean zero and finite second moments. For any $d \times D$ real matrix θ, let f_θ denote the probability distribution of $\theta\mathbf{x}/\sqrt{D}$. If $\mathbf{\Theta}$ is a $d \times D$ matrix of i.i.d. $\mathrm{N}(0,1)$ random variables, then for any Borel measurable set $S \subseteq \mathbb{R}^d$, the expected probability mass assigned by $f_{\mathbf{\Theta}}$ to S is

$$\overline{f}(S) := \mathbb{E}\, f_{\mathbf{\Theta}}(S) = \int \nu_\sigma(S)\mu(\mathrm{d}\sigma),$$

where ν_σ is the density of the d-dimensional spherical Gaussian $\mathrm{N}(0, \sigma^2 I_d)$, and μ is the probability distribution of $\|\mathbf{x}\|_2/\sqrt{D}$. That is, the expected measure $\overline{f}$ is a scale-mixture of spherical Gaussian distributions. Dasgupta et al. (2006) show the following probability concentration theorem.

Theorem 1 (Theorem 9 of Dasgupta, Hsu, and Verma, 2006)**.** *Pick any $0 < \epsilon < 1$ and $\sigma_\epsilon > 0$ such that $\mu\{\sigma \mid \sigma < \sigma_\epsilon\} \leq \epsilon$, and any open ball $B \subset \mathbb{R}^d$. Then*

$$\Pr\big[|f_{\mathbf{\Theta}}(B) - \overline{f}(B)| > 2\epsilon\big] \leq \exp\left(-\Omega\left(\frac{\epsilon^4 D}{d} \cdot \frac{\sigma_\epsilon^2}{\lambda_{\max}} \cdot \frac{1}{\ln(1/\epsilon)}\right)\right),$$

where $\lambda_{\max}$ is the largest eigenvalue of the covariance matrix of $\mathbf{x}$.

The probability bound in Theorem 1 can be improved to the following (which also immediately improves their Theorem 13 in the same way).

Theorem 2. *In the same setting as Theorem 1,*

$$\Pr\big[|f_{\mathbf{\Theta}}(B) - \overline{f}(B)| > 2\epsilon\big] \leq \exp\left(-\Omega\left(\epsilon^4 D \cdot \frac{\sigma_\epsilon^2}{\lambda_{\max}}\right)\right).$$

The main ingredient is the following improvement to Lemma 7 of Dasgupta et al. (2006).

Lemma 1. *Pick any $0 < \epsilon < 1$ and any $\sigma > 0$. If*

$$\Delta := \sqrt{\frac{\pi}{2}}\sigma\epsilon,$$

then $\nu_\sigma(B_\Delta) \leq \nu_\sigma(B) + \epsilon$ for any open ball B in $\mathbb{R}^d$. (Here B_Δ is the same ball as B with radius increased by Δ.)

Dasgupta et al. had previously obtained the conclusion of Lemma 1 only with a much smaller value of Δ, roughly $\sigma\epsilon/\sqrt{d}$. Lemma 1 is a direct consequence of Propositions 1 and 2, which together show that the probability density function of $\|\mathbf{z} - c\|$ for $\mathbf{z} \sim \mathrm{N}(0, I_d)$ is uniformly bounded, independent of the dimension d and the shift $c \in \mathbb{R}^d$.

Proposition 1. *Let $\mathbf{z} \sim \mathrm{N}(0, I_d)$. For any $c \in \mathbb{R}^d$, the probability density function for $\|\mathbf{z} - c\|_2$ is bounded above by $\sqrt{2/\pi}$. Consequently, for any $r, \delta \geq 0$,*

$$\Pr(r \leq \|\mathbf{z} - c\|_2 \leq r + \delta) \leq \sqrt{\frac{2}{\pi}}\delta.$$

Proof. The probability distribution of $\|\mathbf{z} - c\|_2^2$ is $\chi_d^2(\lambda)$, the non-central chi-squared distribution with d degrees-of-freedom and non-centrality parameter $\lambda := \|c\|_2^2$. Since $\chi_d^2(\lambda)$ has the mixture representation

$$\sum_{k=0}^{\infty} \Pr(\mathbf{K} = k) \cdot \chi_{d+2k}^2,$$

where χ_{d+2k}^2 denotes the (central) chi-squared distribution with $d + 2k$ degrees-of-freedom and $\mathbf{K} \sim \text{Poisson}(\lambda/2)$, it follows that the probability distribution of $\|\mathbf{z} - c\|_2$ has the mixture representation

$$\sum_{k=0}^{\infty} \Pr(\mathbf{K} = k) \cdot \chi_{d+2k},$$

where χ_{d+2k} denotes the (central) chi distribution with $d + 2k$ degrees-of-freedom. Hence, the probability density function for $\|\mathbf{z} - c\|_2$ is bounded above by

$$\sup\{p_{d+2k}(r) \mid k \in \mathbb{Z}, k \geq 0, r \geq 0\},$$

where p_{d+2k} denotes the probability density function for χ_{d+2k}. By Proposition 2, this supremum is at most $\sqrt{2/\pi}$. □

Proposition 2. *For every $k \in \mathbb{N}$, let p_k denote the probability density function for χ_k, the chi distribution with k degrees-of-freedom:*

$$p_k(r) = \frac{2^{1-k/2}}{\Gamma(k/2)} r^{k-1} e^{-r^2/2}, \quad r \geq 0.$$

Then

$$\sup\{p_k(r) \mid r \geq 0\} \leq \sqrt{\frac{2}{\pi}}.$$

Proof. We have $p_k(r) \propto \exp(g_k(r))$, where

$$g_k(r) := (k-1)\ln(r) - \frac{r^2}{2}, \quad r \geq 0.$$

Observe that

$$\frac{\mathrm{d}g_k}{\mathrm{d}r}(r) = \frac{k-1}{r} - r, \qquad \frac{\mathrm{d}^2 g_k}{\mathrm{d}r^2}(r) = -\frac{k-1}{r^2} - 1 < 0.$$

So g_k is strictly concave and maximized at $r_k := \sqrt{k-1}$. Therefore

$$M_k := \sup_{r \geq 0} p_k(r) = p_k(r_k) = \frac{2^{1-k/2}}{\Gamma(k/2)} (k-1)^{(k-1)/2} e^{-(k-1)/2}.$$

For $k \in \{1, 2\}$, we have

$$M_1 = \sqrt{\frac{2}{\pi}} = 0.7979\ldots \quad \text{and} \quad M_2 = e^{-1/2} = 0.6065\ldots,$$

so $M_2 < M_1$. We claim that $M_k \leq M_1$ for all $k \in \mathbb{N}$. Since we already have $M_2 < M_1$, it suffices to show that $M_{k+2} \leq M_k$ for all $k \in \mathbb{N}$. Using the identity $\Gamma(k/2+1) = (k/2)\Gamma(k/2)$, we have

$$\frac{M_{k+2}}{M_k} = \frac{1}{ek} \cdot \frac{(k+1)^{(k+1)/2}}{(k-1)^{(k-1)/2}},$$

so

$$\begin{aligned}
\ln\left(\frac{M_{k+2}}{M_k}\right) &= \frac{1}{2}\Big((k+1)\ln(k+1) - (k-1)\ln(k-1) - 2\Big) - \ln(k) \\
&= \frac{1}{2}\Big(\big((k+1)\ln(k+1) - (k+1)\big) - \big((k-1)\ln(k-1) - (k-1)\big)\Big) - \ln(k) \\
&= \frac{1}{2}\int_{k-1}^{k+1} \ln(t)\,\mathrm{d}t - \ln(k) \\
&\leq \ln\left(\frac{1}{2}\int_{k-1}^{k+1} t\,\mathrm{d}t\right) - \ln(k) \quad \text{(by Jensen's inequality)} \\
&= 0.
\end{aligned}$$

This implies $M_{k+2} \leq M_k$. □

In their choice of Δ, Dasgupta et al. consider the ratio $\mathrm{vol}(B_\Delta)/\mathrm{vol}(B)$, which is $(1+\Delta/r)^d$ when B has radius r, and then use this ratio with the point-wise Gaussian density ratio to control $\nu_\sigma(B_\Delta)$ in terms of $\nu_\sigma(B)$ up to a small additive tail term. When $r = \sigma\sqrt{d}$, their multiplicative bound is $1 + O(\epsilon)$ only for $\Delta = O(\sigma\epsilon/\sqrt{d})$. However, this bound is pessimistic: for balls centered at the origin, the rapid growth of the surface area near radius $\sigma\sqrt{d}$ is compensated by the decay of the ν_σ (Proposition 2), and the same is true for balls shifted away from the origin (Proposition 1).

References

Sanjoy Dasgupta, Daniel Hsu, and Nakul Verma. A concentration theorem for projections. In *Twenty-Second Conference on Uncertainty in Artificial Intelligence*, 2006.