Log-Sobolev inequalities and concentration

Sections 5.1, 5.2, 5.3, 5.4 from Boucheron et al. (2013)

Geelon So, agso@eng.ucsd.edu
Concentration inequalities reading group — April 21, 2021

Entropy of a general random variable

Definition (Entropy)

Let $\Phi(x) = x \log x$ *where* $0 \log 0 \equiv 0$. *Let* Y *be a nonnegative, integrable random variable, so that* $E[Y] < \infty$. *The* ***entropy*** *of* Y *is the Jensen's gap:*

$$\mathrm{Ent}(Y) = E\Phi(Y) - \Phi(EY).$$

Subadditivity of entropy

Theorem

Let $X_1, \ldots, X_n$ *be independent and* $Z = f(X_1, \ldots, X_n)$ *be a nonnegative measurable function so that* $\Phi(Z)$ *is integrable. Then:*

$$\mathrm{Ent}(Z) \leq E\left[\sum_{i=1}^{n} \mathrm{Ent}^{(i)}(Z)\right],$$

where $\mathrm{Ent}^{(i)}(Z) = E^{(i)}\Phi(Z^2) - \Phi(E^{(i)}Z)$.

- $E^{(i)}$ is expectation conditioned on $X^{(i)} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$.

Concentration on the binary hypercube

The binary hypercube

- Let $Q_n = \{-1, +1\}^n$ be the n-dimensional binary hypercube
- Let $X \in Q_n$ be a vertex drawn uniformly at random
- Let $f : Q_n \to \mathbb{R}$ be any function

The binary hypercube (cont.)

- Let $\mathrm{Ent}(f) = \mathrm{Ent}(f(X)) = E\big[f(X)\log f(X)\big] - Ef(X)\log Ef(X)$.
- Let $\mathcal{E}(f) = \frac{1}{2}E\left[\sum_{i=1}^{n}\big(f(X) - f(\tilde{X}^{(i)})\big)^2\right]$
  - $\tilde{X}^{(i)}$ replaces the ith coordinate X_i of X with a fresh draw X_i'.
  - Efron–Stein inequality: $\mathrm{Var}\big(f(X)\big) \leq \mathcal{E}(f)$.

Calculus on Q_n

We can define the **discrete gradient** of f at $x \in Q_n$ by:

$$\nabla_i f(x) = \frac{f(\overline{x}^{(i)}) - f(x)}{2}$$

- where $\overline{X}^{(i)} = (X_1, \ldots, -X_i, \ldots, X_n)$ flips the sign of the ith coordinate
- *read: rate of change in f while crossing the ith edge at x*

It follows that:

$$\mathcal{E}(f) = \frac{1}{2}E\left[\sum_{i=1}^{n}\big(f(X) - f(\tilde{X}^{(i)})\big)^2\right] = \frac{1}{4}E\left[\sum_{i=1}^{n}\big(f(X) - f(\overline{X}^{(i)})\big)^2\right] = E\|\nabla f(X)\|^2.$$

A log-Sobolev inequality

Theorem

Let $f : Q_n \to \mathbb{R}$ and X is uniform over Q_n. Then:

$$\mathrm{Ent}(f^2) \leq 2\mathcal{E}(f).$$

- We can also write this as $\mathrm{Ent}(f^2) \leq 2E\left[\|\nabla f(X)\|^2\right]$.

As a generalization of the edge-isopermetric inequality

If $f = \mathbf{1}_A$ for $A \subset Q_n$, then $\mathrm{Ent}(f^2) \leq 2\mathcal{E}(f)$ is the **edge-isoperimetric inequality**:

$$2P(A) \log \frac{1}{P(A)} \leq I(A).$$

Let $|A|$ be size of A and $|\partial_E A|$ be size of edge set $E(A, A^c)$.

- $P(A) = |A|/2^n$
- $I(A) = |\partial_E A|/2^{n-1}$ is the **influence** of A
  - $I(A) = \sum_{i=1}^{n} \Pr\left(f(X) \neq f(\overline{X}^{(i)})\right)$

As a generalization of Efron-Stein on hypercube

Problem (Problem 5.1)

If g is nonnegative, then $\mathrm{Var}\big(g(X)\big) \leq \mathrm{Ent}(g^2)$.

Problem (Problem 5.2)

Combining Problem 5.1 and the log-Sobolev inequality implies for any $f : Q_n \to \mathbb{R}$,

$$\mathrm{Var}(f) \leq \mathcal{E}(f).$$

Proof of log-Sobolev inequality, $\mathrm{Ent}(f^2) \leq 2\mathcal{E}(f)$

Claim: it suffices to show one-dimensional version.

- Recall that: $\mathcal{E}(f) = \frac{1}{4}\sum_{i=1}^{n} E\left[\left(f(X) - f(\overline{X}^{(i)})\right)^2\right]$.
- Let $Z = f(X)$. Subadditivity of entropy:

$$\mathrm{Ent}(Z^2) \leq E\left[\sum_{i=1}^{n} \mathrm{Ent}^{(i)}(Z^2)\right],$$

  where $\mathrm{Ent}^{(i)}(Z^2) = E^{(i)}\Phi(Z^2) - \Phi(E^{(i)}Z)$.
- Suffices to show:

$$\mathrm{Ent}^{(i)}(Z^2) \leq \frac{1}{2}E^{(i)}\left[\left(f(X) - f(\overline{X}^{(i)})\right)^2\right].$$

Proof of log-Sobolev inequality, one-dimensional version

Let $f(-1) = a$ and $f(1) = b$. Want to show:

$$\frac{a^2}{2}\log a^2 + \frac{b^2}{2}\log b^2 - \frac{a^2+b^2}{2}\log\frac{a^2+b^2}{2} \leq \frac{1}{2}(a-b)^2.$$

- Assume WLOG $a \geq b \geq 0$
  - by symmetry and $\frac{1}{2}(|a|-|b|)^2 \leq \frac{1}{2}(a-b)^2$.
- Inequality holds:
  - fix $b \geq 0$ and set $h_b(a) = \mathrm{Ent}(f^2) - 2\mathcal{E}(f)$
  - $h_b(b) = 0$ and $h_b'(b) = 0$
  - h_b is concave on $[b, \infty)$.

□

Tightness of log-Sobolev inequality

Problem (Problem 5.3)

Let $f : Q_n \to \mathbb{R}$ and X uniform over Q_n. Then

$$\mathrm{Ent}(f^2) \leq c\mathcal{E}(f)$$

does not hold for any $c < 2$.

A more general log-Sobolev inequality

Theorem

Let $f : Q_n \to \mathbb{R}$ *and* $X = (X_1, \ldots, X_n)$ *where* $X_i \sim \mathrm{Ber}(p)$. *Then:*

$$\mathrm{Ent}(f^2) \leq c(p)\mathcal{E}(f),$$

where $c(p) = \dfrac{1}{1-2p} \log \dfrac{1-p}{p}$ *is optimal.*[1]

[1] Proof relegated to section 14.3.

Herbst's argument: concentration from log-Sobolev inequality

Theorem

Let $f : Q_n \to \mathbb{R}$ *and* X *uniformly distributed on* Q_n*. Let* $v > 0$ *so that for all* $x \in Q_n$*:*

$$\sum_{i=1}^{n} \left(f(x) - f(\overline{x}^{(i)}) \right)_+^2 \leq v.$$

Then for all $t > 0$*, the random variable* $Z = f(X)$ *satisfies:*

$$\Pr\left(Z > EZ + t\right) \leq e^{-t^2/v} \quad \textit{and} \quad \Pr\left(Z < EZ - t\right) \leq e^{-t^2/v}.$$

- C.f. Efron–Stein: $\mathrm{Var}(Z) \leq v/2$. This theorem states that $Z - EZ$ is subgaussian. Here, we need pointwise control $\sum \left(f(x) - f(\overline{x}^{(i)}\right)_+^2$ while E-F just in expectation.

Herbst's argument: key ideas

- Recall that to show concentration, we would like to bound the log-MGF ψ_Z, since:

$$\begin{aligned}\Pr(Z > EZ + t) &\leq \inf_{\lambda>0} \exp\left(\psi_Z(\lambda) - \lambda EZ - \lambda t\right) \\ &= \inf_{\lambda>0} \exp\left\{\lambda\left(\frac{\psi_Z(\lambda)}{\lambda} - EZ - t\right)\right\}\end{aligned}$$

  - $\psi_Z(\lambda) = \log E[\exp(\lambda Z)]$
  - similar expression for left-tail bound
- Idea: bound the derivative of $\psi_Z(\lambda)/\lambda$,

$$\left(\frac{\psi_Z(\lambda)}{\lambda}\right)' = \frac{E\left[e^{\lambda Z}\log e^{\lambda Z}\right] - E\left[e^{\lambda Z}\right]\log E\left[e^{\lambda Z}\right]}{\lambda^2 E[e^{\lambda Z}]} = \frac{\mathrm{Ent}(e^{\lambda f})}{\lambda^2 E[e^{\lambda Z}]}.$$

- Log-Sobolev inequality implies upper bound $\mathrm{Ent}(e^{\lambda f}) \leq \frac{\nu\lambda^2}{4} E[e^{\lambda Z}]$.

Proof of concentration

- Log-Sobolev inequality implies:

$$\begin{aligned}\mathrm{Ent}(e^{\lambda f}) &\le 2\mathcal{E}(e^{\lambda f/2})\\ &= \frac{1}{2}\sum_{i=1}^{n} E\left[\left(e^{\lambda f(X)/2} - e^{\lambda f(\overline{X}^{(i)})/2}\right)^2\right]\\ &= \sum_{i=1}^{n} E\left[\left(e^{\lambda f(X)/2} - e^{\lambda f(\overline{X}^{(i)})/2}\right)_+^2\right]\end{aligned}$$

- Apply convexity of exponential function: $z > y$ implies $e^{z/2} - e^{y/2} \le \frac{1}{2}(z-y)e^{z/2}$,

$$2\mathcal{E}(e^{\lambda f/2}) \le \frac{\lambda^2}{2} E\left[\frac{1}{2}\sum_{i=1}^{n}\left(f(X) - f(\overline{X}^{(i)})\right)_+^2 \cdot e^{\lambda f(X)}\right] \le \frac{v\lambda^2}{4} E[e^{\lambda Z}].$$

Proof of concentration (cont.)

- Combining everything:

$$\mathrm{Ent}(e^{\lambda f}) \leq \frac{v\lambda^2}{4} E[e^{\lambda Z}].$$

- This implies:

$$\left(\frac{\psi_Z(\lambda)}{\lambda}\right)' = \frac{\mathrm{Ent}(e^{\lambda f})}{\lambda^2 E[e^{\lambda Z}]} \leq \frac{v}{4}.$$

- For $\lambda > 0$, integrating implies:

$$\frac{\psi_Z(\lambda)}{\lambda} \leq \lim_{h \downarrow 0} \frac{\psi_Z(h)}{h} + \frac{\lambda v}{4}.$$

- Apply l'Hospital and fact $\psi_Z'(0) = E[Z]$ to obtain:

$$\psi_Z(\lambda) \leq \lambda E[Z] + \frac{\lambda v}{4}.$$

Proof of concentration (cont.)

The upper bound $\psi_Z(\lambda) \leq \lambda E[Z] + \frac{\lambda v}{4}$ implies:

$$\begin{aligned}
\Pr(Z > EZ + t) &\leq \inf_{\lambda > 0} \exp\left\{\lambda\left(\frac{\psi_Z(\lambda)}{\lambda} - EZ - t\right)\right\} \\
&\leq \inf_{\lambda > 0} \exp\left\{\frac{\lambda^2 v}{4} - \lambda t\right\}
\end{aligned}$$

which is optimized at $\lambda = 2t/v$ so the RHS is $e^{-t^2/v}$. □

Hoeffding's as a variant of Herbst's

Problem (Problem 5.5)

Let $f : Q_n \to \mathbb{R}$ and X uniformly distributed on Q_n. Let $v > 0$ so that for all $x \in Q_n$,

$$\sum_{i=1}^{n} \left(f(x) - f(\overline{x}^{(i)}) \right)^2 \leq v.$$

Then for all $t > 0$, the random variable $Z = f(X)$ satisfies:

$$\Pr(Z > EZ + t) \leq e^{-2t^2/v}.$$

- Note: this time, v bounds not merely the squared positive parts.
- Hint: prove for all $z \geq y$ that $\left(e^{z/2} - e^{y/2}\right)^2 \leq \frac{(z-y)^2}{8}(e^z + e^y)$.

General technique

- Use log-Sobolev inequalities to prove a differential inequality:

$$\left(\frac{\psi_Z(\lambda)}{\lambda}\right)' \leq h(\lambda).$$

- Combine with Cramér-Chernoff bound:

$$\Pr(Z > EZ + t) \leq \inf_{\lambda>0} \exp\left\{\lambda\left(\frac{\psi_Z(\lambda)}{\lambda} - EZ - t\right)\right\}$$

Concentration for Gaussians

A log-Sobolev inequality for Gaussians

Theorem

Let $X = (X_1, \ldots X_n) \sim \mathcal{N}(0, I)$ *and* $f : \mathbb{R}^n \to \mathbb{R}$ *be continuously differentiable. Then:*

$$\mathrm{Ent}(f^2) \leq 2E\left[\|\nabla f(X)\|^2\right].$$

- Let's see an application before proving this.

Obtaining Gaussian concentration from log-Sobolev inequality

Theorem (Gaussian concentration inequality)

Let $X = (X_1, \dots X_n) \sim \mathcal{N}(0, I)$ *and* $f : \mathbb{R}^n \to \mathbb{R}$ *be L-Lipschitz. Then, for all* $t > 0$,

$$\Pr\left\{f(X) - Ef(X) \geq t\right\} \leq e^{-t^2/2L^2}.$$

Proof of Gaussian concentration

- If f is L-Lipschitz, then f is differentiable with $\|\nabla f\| \leq L$ almost everywhere.
- **Claim:** log-Sobolev inequality implies differential inequality:

$$\left(\frac{\psi_Z(\lambda)}{\lambda}\right)' = \frac{\mathrm{Ent}(e^{\lambda f})}{\lambda^2 E[e^{\lambda Z}]} \leq \frac{L^2}{2}.$$

- Combine with Cramér-Chernoff to obtain result.

Proof of claim. The log-Sobolev inequality states:

$$\begin{aligned}\mathrm{Ent}(e^{\lambda f}) &\leq 2E\left[\left\|\nabla e^{\lambda f(X)/2}\right\|^2\right] \\ &= \frac{\lambda^2}{2} E\left[e^{\lambda f(X)} \cdot \|\nabla f(X)\|^2\right] \leq \frac{\lambda^2 L^2}{2} E[e^{\lambda Z}].\end{aligned}$$

□

Application of Gaussian concentration

Let $X = (X_1, \ldots, X_n) \sim \mathcal{N}(0, \Sigma)$.

- There exists linear map $f : \mathbb{R}^n \to \mathbb{R}^n$ such that $X =_d f(Y)$ where $Y \sim \mathcal{N}(0, I)$.
- The map f has bounded operator norm,

$$L = \|f\|_{\text{op}} = \sup_{\|y\|_2 = 1} \|Ay\|_p.$$

- Gaussian concentration implies:

$$\Pr\left\{\|X\|_p > E\|X\|_p + t\right\} \leq e^{-t^2/2L^2}.$$

Proof of Gaussian log-Sobolev inequality, $\mathrm{Ent}(f^2) \leq 2E[\|\nabla f(X)\|^2]$

- It suffices to prove the one-dimensional version (c.f. argument for Q_n using subadditivity of entropy).
- Apply similar technique as proof of Gaussian–Poincaré inequality:
  - $X \sim \mathcal{N}(0,1)$ by sequence $Y_m \Rightarrow X$

    $$Y_m = \frac{1}{\sqrt{m}} \sum_{k=1}^{m} \varepsilon_{m,k},$$

    where $\varepsilon_{m,k}$ are independent Rademacher r.v.s (i.e. $\varepsilon_{m,k} = \pm 1$ uniformly)
  - Central limit theorem implies:

    $$\lim_{m \to \infty} \mathrm{Ent}\left[f(Y_m)^2\right] = \mathrm{Ent}\left[f(X)^2\right]$$

    for f continuous and uniformly bounded.

Proof of Gaussian log-Sobolev inequality (cont.)

- The log-Sobolev inequality on Q_m implies:

$$
\begin{aligned}
\operatorname{Ent}\left[f(Y_m)^2\right] &\leq \frac{1}{2} E\left[\sum_{k=1}^{m} \left| f(Y_m) - f\left(Y_m - \frac{2\varepsilon_{m,k}}{\sqrt{m}}\right)\right|^2\right] \\
&= \frac{m}{2} \cdot E\left[\left|\frac{f(Y_m) - f(Y_m \pm 2/\sqrt{m})}{\sqrt{m}/2}\right|^2 \cdot \frac{4}{m}\right]
\end{aligned}
$$

- Taking the limit $m \to \infty$,

$$
\operatorname{Ent}\left[f(X)^2\right] \leq 2E\left[|f'(X)|^2\right].
$$

□

References

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence.* Oxford university press, 2013.