Concentration Inequalities (cont.)

$X_1, \dots, X_n$ — indep. r.v.'s

$f(X_1, \dots, X_n)$: function that is "not too sensitive" to changes in each X_i

$$\mathbb{P}\{|f(X_1,\dots,X_n) - \mathbb{E}f(X_1,\dots,X_n)| \geq \varepsilon\} \leq K_1 e^{-K_2 n}$$

where $K_1 > 0$ is a small "absolute" constant

$$K_2 = K_2(f, \varepsilon) > 0$$

Recap: exponential Markov trick

$$\mathbb{P}\{X - \mathbb{E}X \geq t\} \leq \inf_{s \geq 0}\{e^{-st}\,\mathbb{E}[e^{s(X-\mathbb{E}X)}]\} \qquad t > 0$$

$$\mathbb{P}\{X - \mathbb{E}X \leq -t\} \leq \inf_{s \geq 0}\{e^{-st}\,\mathbb{E}[e^{s(X-\mathbb{E}X)}]\}$$

– need good bounds on $\psi(s) := \log \mathbb{E}[e^{s(X-\mathbb{E}X)}]$.

Hoeffding's Lemma. If $X \in [a,b]$ a.s., then

$$\log \mathbb{E}[e^{s(X-\mathbb{E}X)}] \leq \frac{s^2(b-a)^2}{8}.$$

Corollary: 1)

$$\mathbb{P}\{X \geq \mathbb{E}X + t\} \leq \exp\left(-\frac{2t^2}{(b-a)^2}\right)$$

$$\mathbb{P}\{X \leq \mathbb{E}X - t\} \leq \exp\left(-\frac{2t^2}{(b-a)^2}\right)$$

2) If $X_i \in [a_i, b_i]$ a.s., $1 \leq i \leq n$, and are independent, then

$$\mathbb{P}\left\{\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}X_i \geq t\right\} \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Proof of Hoeffding's Lemma

w.l.o.g. $\mathbb{E}X = 0$, $\quad \psi(s) = \log \mathbb{E}[e^{sX}]$

1) Properties of $\psi(s)$: $\quad \psi'(s) = \dfrac{\mathbb{E}[Xe^{sX}]}{\mathbb{E}[e^{sX}]}$

$$\psi''(s) = \frac{\mathbb{E}[X^2 e^{sX}]}{\mathbb{E}[e^{sX}]} - \left(\frac{\mathbb{E}[Xe^{sX}]}{\mathbb{E}[e^{sX}]}\right)^2$$

2) Twisting: $\quad X \sim P \longrightarrow U \sim \tilde{P}$

$$\int_{\mathbb{R}} f(u)\, \tilde{P}(du) = \frac{\int_{\mathbb{R}} f(x) e^{sx} P(dx)}{\int e^{sx} P(dx)} \qquad \forall f$$

$$\mathbb{P}[U \in A] = \frac{\mathbb{E}[\mathbb{1}_{\{X \in A\}} e^{sX}]}{\mathbb{E}[e^{sX}]}$$

3) $X \in [a,b]$ a.s. $\Rightarrow$ $U \in [a,b]$ a.s.

$$\mathbb{E}U = \frac{\mathbb{E}[Xe^{sX}]}{\mathbb{E}[e^{sX}]} = \psi'(s)$$

$$\mathrm{Var}(U) = \mathbb{E}U^2 - (\mathbb{E}U)^2 = \psi''(s)$$

4) Claim: if U is any r.v. that takes values in $[a,b]$ a.s., then

$$\mathrm{Var}[U] \le \frac{(b-a)^2}{4}.$$

5) For any s, $\quad \psi''(s) \le \dfrac{(b-a)^2}{4}$.

6) $\psi(s) = \displaystyle\int_0^s \int_0^t \psi''(r)\, dr\, dt$

$$\int_0^t \psi''(r)\,dr = \psi'(t) - \underbrace{\psi'(0)}_{=0} = \psi'(t)$$

$$\int_0^s \psi'(t)\,dt = \psi(s) - \underbrace{\psi(0)}_{=0} = \psi(s)$$

$$\Rightarrow \psi(s) = \int_0^s \int_0^t \underbrace{\psi''(r)}_{\le \frac{(b-a)^2}{4}}\,dr\,dt \le \frac{s^2 (b-a)^2}{8}.$$

$\blacksquare$

McDiarmid's inequality: concentration beyond sums

$X_1, \dots, X_n$ - indep. r.v.'s taking values in $\mathcal{X}$

$f : \mathcal{X}^n \to \mathbb{R}$

$$\mathbb{P}\{ | f(X_1,\dots,X_n) - \mathbb{E} f(X_1,\dots,X_n) | \ge t \} \le ?$$

- f is not too sensitive to each x_i

$$x^n = (x_1,\dots,x_n) = (\underbrace{x_1,\dots,x_{i-1}}_{x^{i-1}}, x_i, \underbrace{x_{i+1},\dots,x_n}_{x_{i+1}^n})$$

Discrete gradient:

$$\Delta_i f(x_1,\dots,x_n) = \sup_{x_i} f(x^{i-1}, x_i, x_{i+1}^n) - \inf_{x_i} f(x^{i-1}, x_i, x_{i+1}^n)$$

Then f has bounded differences when

$$\underbrace{\sup_{x^n \in \mathcal{X}^n} \Delta_i f(x_1,\dots,x_n)}_{\|\Delta_i f\|_\infty} \le c_i < \infty \quad \text{for all } i.$$

<u>Thm</u> (McDiarmid's inequality) If $f: \mathcal{X}^n \to \mathbb{R}$ has bounded differences, then, for $X_1, \dots, X_n$ independent,

$$\mathbb{P}\left\{ f(X^n) - \mathbb{E} f(X^n) \geq t \right\} \leq \exp\left(- \frac{2t^2}{\sum_{i=1}^n \|\Delta_i f\|_\infty^2} \right)$$

$$\mathbb{P}\left\{ f(X^n) - \mathbb{E} f(X^n) \leq -t \right\} \leq \exp\left(- \frac{2t^2}{\sum_{i=1}^n \|\Delta_i f\|_\infty^2} \right)$$

Note: $\|\Delta_i f\|_\infty$ measures the sensitivity of f to ith input.

Examples

1) $f(x_1, \dots, x_n) = x_1 + \dots + x_n \qquad a_i \leq x_i \leq b_i$

$$\Delta_i f(x^n) = \sup_{x_i \in [a_i, b_i]} (x_1 + \dots + x_i + \dots + x_n) - \inf_{x_i' \in [a_i, b_i]} (x_1 + \dots + x_i' + \dots + x_n)$$

$$= \sup_{x_i \in [a, b]} \sup_{x_i' \in [a_i, b_i]} (x_i - x_i') = b_i - a_i$$

$$\Rightarrow \mathbb{P}\left\{ \left| \sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E} X_i \right| \geq t \right\} \leq 2 \exp\left(- \frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right)$$

2) Uniform deviations of empirical means

$X_1, \dots, X_n \overset{iid}{\sim} P$

$$P_n(A) := \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \in A\}} \quad \text{— empirical frequency of } A$$

$$\mathbb{E} P_n(A) = P(A)$$

$0 \le P_n(A) \le 1$, function of $X_1, \dots, X_n$

$$\mathbb{P}\{|P_n(A) - P(A)| \geq t\} \le 2e^{-2nt^2} \text{ (Hoeffding)}$$

$$f(X_1, \dots, X_n) = \sup_{A \in \mathcal{A}} |P_n(A) - P(A)|$$

(see this later in binary classification)

$$= \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \in A\}} - P(A) \right|$$

Later: if $\mathcal{A}$ is not too 'rich,' $\mathbb{E} f(X^n) \sim \frac{C}{\sqrt{n}}$

$$\mathbb{P}\left\{ \sup_{A \in \mathcal{A}} |P_n(A) - P(A)| \geq \mathbb{E} \sup_{A \in \mathcal{A}} |P_n(A) - P(A)| + t \right\}$$

$$\le e^{-2nt^2}.$$

<u>Claim:</u> $f(x^n) = \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n 1_{\{x_i \in A\}} - P(A) \right|$

$\| \Delta_i f \|_\infty \le 1/n \quad \forall i$

$$\sum_{i=1}^n \| \Delta_i f \|_\infty^2 \le \frac{1}{n^2} \cdot n = \frac{1}{n}$$

$$\Rightarrow \exp\left(- \frac{2t^2}{\sum_{i=1}^n \| \Delta_i f \|_\infty^2} \right) \le \exp(-2nt^2)$$

$$\Delta_i f(x^n) = \sup_{x_i} \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{j=1}^{n} 1_{\{x_j \in A\}} - P(A) \right|$$

$$- \inf_{x_i} \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{j=1}^{n} 1_{\{x_j \in A\}} - P(A) \right|$$

$$\left| \frac{1}{n} \sum_{j=1}^{n} 1_{\{x_j \in A\}} - P(A) \right|$$

$$= \left| \frac{1}{n} \sum_{j \neq i} 1_{\{x_j \in A\}} + \frac{1}{n} 1_{\{x_i \in A\}} - P(A) \right|$$

$$\left| \frac{1}{n} \sum_{j \neq i} 1_{\{x_j \in A\}} + \frac{1}{n} 1_{\{x_i \in A\}} - P(A) \right|$$

$$- \left| \frac{1}{n} \sum_{j \neq i} 1_{\{x_j \in A\}} + \frac{1}{n} 1_{\{x_i' \in A\}} - P(A) \right|$$

$|a| - |b| \leq |a - b|$

$$\leq \frac{1}{n} \left| 1_{\{x_i \in A\}} - 1_{\{x_i' \in A\}} \right| \leq \frac{1}{n}.$$