SDS 384 Homework 2

Qiaohui Lin

Feb 2019

1 Question 1

Question: Remember Hoeffding's Lemma? We proved it with a weaker constant in class using a symmetrization type argument. Now we will prove the original version. Let X be a bounded r.v. in $[a, b]$ such that $E[X] = \mu$. Let $f(\lambda) = \log E[e^{\lambda(X-\mu)}]$. Show that $f''(\lambda) \le (b-a)^2/4$. Now use the fundamental theorem of calculus to write $f(\lambda)$ in terms of $f''(\lambda)$ and finish the argument.

Part 1: Show that $f''(\lambda) \le (b-a)^2/4$.

Proof: $f(\lambda) = M_x(\lambda) = E(e^{\lambda(x-\mu)})$, Thus, taking first and second derivative, we have

$$f'(\lambda) = \frac{M_x'(\lambda)}{M_x(\lambda)} = \frac{E[(x-\mu)e^{\lambda(x-mu)}]}{E(e^{\lambda(x-\mu)})}$$

$$\begin{aligned}
f''(\lambda) &= \frac{M_x''(\lambda)(M_x(\lambda) - (M_x'(\lambda)^2))}{(M_x(\lambda))^2} = \frac{M_x''(\lambda)}{M_x(\lambda)} - (\frac{M_x'(\lambda)}{M_x(\lambda)})^2 \\
&= \frac{E[(x-\mu)^2 e^{\lambda(x-\mu)}]}{E(e^{\lambda(x-\mu)})} - \left(\frac{E[(x-\mu)e^{\lambda(x-\mu)}]}{E(e^{\lambda(x-\mu)})}\right)^2 \\
&= E\left[(x-\mu)^2 \frac{e^{\lambda(x-\mu)}}{E(e^{\lambda(x-\mu)})}\right] - \left(E\left[(x-\mu)\frac{e^{\lambda(x-\mu)}}{E(e^{\lambda(x-\mu)})}\right]\right)^2
\end{aligned}$$

if we change the measure to $dQ = \frac{e^{\lambda x}}{Ee^{\lambda x}}dP$, which means, $E(\frac{f(x)e^{\lambda x}}{E(e^{\lambda(x)})}) = E_Q(f(x))$, i.e., the above becomes

$$\begin{aligned}
f''(\lambda) &= var_Q(x) \le E_Q(x-t)^2 \text{for any t} \\
&= E_Q(x - \frac{a+b}{2})\text{take } t = \frac{a+b}{2} \\
&\le \frac{(b-a)^2}{4}
\end{aligned}$$

Part 2 use the fundamental theorem of calculus to write $f(\lambda)$ in terms of $f''(\lambda)$.
Proof :

$$\begin{aligned}
f(\lambda) &= \int_0^\lambda \int_0^t f''(\rho) d\rho dt \\
&\leq \int_0^\lambda \int_0^t \frac{(b-a)^2}{4} d\rho dt \\
&= \int_0^\lambda \frac{(b-a)^2}{4} \rho \Big|_0^t dt \\
&= \int_0^\lambda \frac{(b-a)^2}{4} t dt \\
&= \frac{(b-a)^2}{8} t^2 \Big|_0^\lambda \\
&= \frac{(b-a)^2}{8} \lambda^2
\end{aligned}$$

Thus, $E(e^{\lambda(x-\mu)}) \leq e^{\frac{\lambda^2(b-a)^2}{8}}$, which is the Hoeffding's Lemma.

2 Question 2

Question: Bernstein's inequality for bounded i.i.d sequences of random variables $\{X_i\}$ with $|X_i| \leq M$ gives: $P(|\sum_i (X_i - E[X_i])| \geq t) \leq 2 \exp\left(\frac{-t^2/2}{\sum_i \mathrm{var}(X_i) + Mt/3}\right)$. There is another better inequality called Bennett's inequality, which we will prove here.
Part 1 Consider zero mean r.v.s X_i such that $|X_i| \leq b$ and $\mathrm{var}(X_i) = \sigma_i^2$. Prove that

$$\log E[\exp(\lambda X_i)] \leq \sigma_i^2 \lambda^2 \left(\frac{e^{\lambda b} - 1 - \lambda b}{(\lambda b)^2}\right) \quad \forall \lambda \in \mathbb{R}.$$

Proof:

We first use the common equality $\log(1+x) \leq x$ (since $1+x \leq e^x$), Thus,

$$
\begin{aligned}
log(E^{\lambda x_i}) &\leq E(\lambda x_i) - 1 \\
&= \sum_{k=0} \frac{\lambda^k E(x_i)^k}{k!} - 1 \\
&= \frac{\lambda^2 \sigma_i^2}{2} + \sum_{k\geq 3} \frac{\lambda^k E(x_i)^k}{k!} \\
&= \frac{\lambda^2 \sigma_i^2}{2} + \lambda^2 \sigma_i^2 \sum_{k\geq 3} \frac{\lambda^{k-2} E(x_i)^{k-2}}{k!} \\
&\leq \frac{\lambda^2 \sigma_i^2}{2} + \lambda^2 \sigma_i^2 \sum_{k\geq 3} \frac{(\lambda b)^{k-2}}{k!} (since |x_i| < b, E(x_i)^k < E|x_i|^k < b^k) \\
&= \lambda^2 \sigma_i^2 \sum_{k\geq 2} \frac{(\lambda b)^{k-2}}{k!} \\
&= \lambda^2 \sigma_i^2 \frac{\sum_{k\geq 2} \frac{(\lambda b)^k}{k!}}{\lambda^2 b^2} \\
&= \lambda^2 \sigma_i^2 \frac{\sum_{k\geq 0} \frac{(\lambda b)^k}{k!} - 1 - \lambda b}{\lambda^2 b^2} \\
&= \lambda^2 \sigma_i^2 \frac{e^{\lambda b} - 1 - \lambda b}{\lambda^2 b^2}
\end{aligned}
$$

Which concludes $log(E^{\lambda x_i}) \leq \lambda^2 \sigma_i^2 \frac{e^{\lambda b} - 1 - \lambda b}{\lambda^2 b^2}$.

Part 2 Given independent r.v.s $X_i, i = 1, \ldots, n$ satisfying the above condition prove

$$
\text{(Bennett's inequality)} \qquad P\left(\sum_i X_i \geq n\delta\right) \leq \exp\left(-\frac{n\sigma^2}{b^2} h(b\delta/\sigma^2)\right),
$$

where $n\sigma^2 = \sum_i \sigma_i^2$ and $h(t) := (1+t)\log(1+t) - t$ for $t \geq 0$.
Proof:

$$
\begin{aligned}
P(\sum x_i \geq n\delta) &= P(\lambda \sum x_i \geq \lambda n\delta) \\
&= P(e^{\lambda \sum x_i} \geq e^{\lambda n\delta}) \\
&\leq \frac{E(e^{\lambda} \sum x_i)}{e^{\lambda\delta n}} \\
&= \frac{\prod_i^n E(\lambda x_i)}{e^{\lambda\delta n}} \\
&\leq \frac{\prod_i^n e^{\lambda^2 \sigma_i^2 \frac{e^{\lambda b} - 1 - \lambda}{\lambda^2 b^2}}}{e^{\lambda\delta n}} \\
&= e^{\sum \sigma_i^2 \lambda^2 * \frac{e^{\lambda b} - 1 - \lambda b}{\lambda^2 b^2} - \lambda\delta n} \\
&= e^{n\sigma^2 \frac{e^{\lambda b} - 1 - \lambda b}{b^2} - \lambda\delta n}
\end{aligned}
$$

To tight the bound, we need to find $\inf_{\lambda} e^{n\sigma^2 \frac{e^{\lambda b} - 1 - \lambda b}{b^2} - \lambda\delta n}$. Thus, we take the deravative with respect to λ.

$$
\frac{\partial e^{n\sigma^2 \frac{e^{\lambda b} - 1 - \lambda b}{b^2} - \lambda\delta n}}{\partial \lambda} = \frac{n\sigma^2}{b}(be^{\lambda b} - b) - \delta n = 0
$$

$$
\lambda = \frac{1}{b} \log(1 + \frac{\delta b}{\sigma^2})
$$

Plug λ in, the bound becomes

$$
\begin{aligned}
&exp(\frac{n\sigma^2}{b^2}(\frac{\delta b}{\sigma^2} - \log(1 + \frac{\delta b}{\sigma^2})) - \frac{\delta n}{b} \log(1 + \frac{\delta b}{\sigma^2})) \\
&= exp(\frac{n\sigma^2}{b^2}((1 + \frac{\delta b}{\sigma^2}) \log(1 + \frac{\delta b}{\sigma^2}) - \frac{b\delta}{\sigma^2})) \\
&= exp(-\frac{n\sigma^2}{b^2} h(\frac{\delta b}{\sigma^2}))
\end{aligned}
$$

Thus, we have, $P(\sum x_i \geq n\delta) \leq exp(-\frac{n\sigma^2}{b^2} h(\frac{\delta b}{\sigma^2}))$

Part 3 Show that Bennett's inequality is at least as good as Bernstein's inequality.

Proof: We know
Bennet: $P(\sum x_i \geq n\delta) \leq exp(-\frac{n\sigma^2}{b^2}h(\frac{\delta b}{\sigma^2}))$
Bernestein: $P(\sum x_i \geq n\delta) \leq exp(-\frac{\frac{n\sigma^2}{2}}{n\sigma^2+\frac{bn\delta}{3}})$
If we want to show Bennet is always better than Bernstein, we want to show

$$-\frac{n\sigma^2}{b^2}h(\frac{\delta b}{\sigma^2}) \leq -\frac{\frac{n\sigma^2}{2}}{n\sigma^2+\frac{bn\delta}{3}}$$

always holds, which means we want to show that

$$\frac{\sigma^2}{b^2}h(\frac{\delta b}{\sigma^2}) - \frac{3\delta^2}{6\sigma^2+2b\delta} \geq 0$$

always holds.
To show this, we define $t = \frac{b\delta}{\sigma^2}$, thus, $\delta^2 = \frac{t^2\sigma^4}{b^2}$, which means we want to show,

$$\frac{\sigma^2}{b^2}h(t) - \frac{3 * \frac{\sigma^4 t^2}{b^2}}{(6+2t)\sigma^2} = \frac{\sigma^2}{b^2}h(t) - \frac{3\sigma^2 t^2}{b^2(6+2t} = \frac{\sigma^2}{b^2}(h(t) - \frac{3t^2}{6+2t}) \geq 0$$

always holds.
Thus, we define $g(t) = h(t) - \frac{3t^2}{6+2t}$, we want to show $g(t) \geq 0$ when $t \geq 0$.

$$g(t) = h(t) - \frac{3t^2}{6+2t} = (1+t)\log(1+t) - t - \frac{3t^2}{6+2t}$$

$$g'(t) = \log(1+t) - \frac{6t^2+36t}{(6+2t)^2}$$

$$g''(t) = \frac{1}{1+t} - \frac{27}{(3+t)^3} = \frac{t^3+9t^2}{(1+t)(t+3)^3} \geq 0$$

when $t \geq 0$.

$g'(0) = 0$, $g''(t) > 0$ when $t > 0$, thus, $g'(t) \geq 0$ always holds.
Similarly, $g(0) = 0$, $g'(t) > 0$ when $t > 0$, thus, $g(t) \geq 0$ always holds.
We have proved, $g(t) \geq 0$ always holds, which means, $-\frac{n\sigma^2}{b^2}h(\frac{\delta b}{\sigma^2}) \leq -\frac{\frac{n\sigma^2}{2}}{n\sigma^2+\frac{bn\delta}{3}}$ always holds, Bennet is always better than Bernstein.

3 Question 3

Question : Given a scalar random variable X, suppose that there are positive constants c_1, c_2 such that,

$$P(X - E[X] \geq t) \leq c_1 \exp(-c_2 t^2) \qquad \forall t \geq 0.$$

Part 1
Prove that $\mathrm{var}(X) \leq \frac{c_1}{c_2}$
Proof:

$$\begin{aligned}
var(x) &= E(x-\mu)^2 = E|x-\mu|^2 \\
&= \int_0^\infty P(|x-\mu|^2 > t)dt \\
&= \int_0^\infty P(|x-\mu| > t^{\frac{1}{2}})dt \\
&\leq \int_0^\infty c_1 e^{-c_2 t} dt \\
&= -\frac{c_1}{c_2} e^{-c_2 t}\Big|_0^\infty \\
&= \frac{c_1}{c_2}
\end{aligned}$$

Part 2
A median m_X is any number such that $P(X \geq m_X) \geq 1/2$ and $P(X \leq m_X) \geq 1/2$. Show by example that the median does not need to be unique.
Proof:
$x \sim Bernoulli(\frac{1}{2})$,

$$x = \begin{cases} 0 & w.p \quad \frac{1}{2} \\ 1 & w.p. \quad \frac{1}{2} \end{cases}$$

Thus, anything between $(0,1)$ m,

$$P(x \geq m) = P(x \leq m) = \frac{1}{2}$$

Thus, anything between $(0,1)$ is a medium, the medium is not unique.

Part 3
Show that if the mean concentration bound $P(X - E[X] \geq t) \leq c_1 \exp(-c_2 t^2)$ holds, then for any median m_X, $\exists$ some positive constant c_3, c_4 such that

$$P(|X - m_X| \geq t) \leq c_3 \exp(-c_4 t^2),$$

Proof:
set a t_0 such that $P(|X - EX| \geq t_0) \leq c_1 exp(-c_2 t_0^2) = \frac{1}{2}$. Then we have,

$$P(x \in (EX - t_0, EX + t_0)) \geq \frac{1}{2}$$

Then based on the definition of m_x, $P(x \geq m) = P(x \leq m) = \frac{1}{2}$, we have

$$m_x \in (EX - t_0, EX + t_0)$$

Thus,

$$\begin{aligned}
P(|X - m_x| \geq t) &\leq P(|X - EX| \geq |t - t_0|) \\
&\leq c_1 exp(-c_2 (t - t_0)^2) \\
&\text{now we want to show there exits some } c_3 \text{ and } c_4 \text{ that this holds} \\
&\leq c_3 exp(-c_4 t^2)
\end{aligned}$$

To show the conditions for c_3 and c_4 we discuss the two cases $t < 2t_0$ and $t > 2t_0$.
When $t < 2t_0$,

$$c_1 exp(-c_2 (t - t_0)^2) \geq c_1 exp(-c_2 t_0^2) = \frac{1}{2}$$

But when $t \to 0$, we need the the bound of $P(|X - m_x| \geq t)$ to ≥ 1, as an always-hold trivial bound, thus we multiply both sides by 2,

$$2c_1 exp(-c_2 (t - t_0)^2) \geq 1$$

Thus, we choose $c_3 = 2c_1$.
When $t > 2t_0$, i.e., $t_0 \leq \frac{t}{2}$

$$c_1 exp(-c_2 (t - t_0)^2) \leq c_1 exp(-c_2 (\frac{t}{2})^2) \leq c_3 exp(-\frac{c_2}{4} t^2)$$

which means we can choose $c_4 = \frac{c_2}{4}$.
We have proved $P(|X - m_X| \geq t)$ can be bounded by $c_3 \exp(-c_4 t^2)$ with, $c_3 = 2c_1$, $c_4 = \frac{c_2}{4}$

Part 4
Conversely, show that whenever the above median concentration holds, then mean concentration holds with $c_1 = 2c_3$ and $c_2 = c_4/8$.

Proof:

Given
$$P(|X - m_x| \geq t) \leq c_3 e^{-c_4 t^2}$$

We first make an i.i.d copy of X, i.e., Y has the same distribution as X and $Y \perp\!\!\!\perp X$.

First, we have

$$\begin{aligned} P(|X - Y| \geq t) &\leq P(|X - m_x| \geq \frac{t}{2}) + P(|Y - m_Y| \geq \frac{t}{2} \\ &\leq c_3 e^{-c_4 (\frac{t}{2})^2} + c_3 e^{-c_4 (\frac{t}{2})^2} \\ &= 2c_3 e^{-\frac{c_4}{4}^2 t^2} \end{aligned}$$

Call $2c_3 = K_1$, $\frac{c_4}{4} = K2$,

$$P(|X - Y| \geq t) \leq K_1 e^{-K_2 t^2}$$

Now we begin the real proof,

$$\begin{aligned} P(|X - EX| \geq t) &= P(exp(\lambda^2 (X - EX)^2) \geq exp(\lambda^2 t^2)) \\ &\leq \frac{E(exp(\lambda^2 (X - EX)^2))}{e^{\lambda^2 t^2}} \end{aligned}$$

And because Y is an i.i.d copy of X,,

$$\begin{aligned} E[exp(\lambda^2 (X - EX)^2)] &= E[exp(\lambda^2 (X - EY)^2)] \\ &= E_X[exp(\lambda^2 E_Y (X - Y)^2)] \\ &\text{Using Jensen's equality, } exp(E_Y (X - Y)^2) \leq E_Y (exp(X - Y)^2) \\ &\leq E_{X,Y} exp(\lambda^2 (X - Y)^2) \end{aligned}$$

Use the moment and tail probability relationship,

$$
\begin{aligned}
E_{X,Y}exp(\lambda^2(X-Y)^2) &= \int_0^\infty P(\lambda^2(X-Y)^2 \geq \log t)dt \\
&= \int_0^1 P((X-Y)^2 > \frac{\log t}{\lambda^2})dt + \int_1^\infty P(|X-Y| \geq \sqrt{\frac{\log t}{\lambda^2}})dt \\
&= 1 + \int_1^\infty P(|X-Y| \geq \sqrt{\frac{\log t}{\lambda^2}})dt \\
&\text{change of variable, } \epsilon = \sqrt{\frac{\log t}{\lambda^2}},\ t = e^{\lambda^2\epsilon^2},\ \frac{dt}{d\epsilon} = 2\lambda^2\epsilon e^{\lambda^2\epsilon^2} \\
&= 1 + \int_1^\infty P(|X-Y| \geq \epsilon)2\lambda^2\epsilon e^{\lambda^2\epsilon^2}d\epsilon \\
&\text{Plug in what we proved before } P(|X-Y| \geq \epsilon) \leq K_1 e^{-K_2\epsilon^2} \\
&\leq 1 + \int_1^\infty K_1 e^{-K_2\epsilon^2} 2\lambda^2\epsilon e^{\lambda^2\epsilon^2}d\epsilon \\
&\leq 1 + \int_0^\infty K_1 e^{-K_2\epsilon^2} 2\lambda^2\epsilon e^{\lambda^2\epsilon^2}d\epsilon \\
&\text{take } \lambda = \sqrt{\frac{K_2}{2}} \\
&= 1 + \int_0^\infty K_1 K_2 \epsilon e^{-\frac{K_2}{2}\epsilon^2}d\epsilon \\
&\text{change of variable, } m = \epsilon^2 \\
&= 1 + \int_0^\infty \frac{1}{2}K_1 K_2 \epsilon e^{-\frac{K_2}{2}m}dm \\
&= 1 - K_1 e^{-\frac{k_2}{2}m}\Big|_0^\infty \\
&= 1 + K_1
\end{aligned}
$$

We have proved,

$$
\begin{aligned}
P(|X-EX| \geq t) &\leq \frac{E(exp(\lambda^2(X-EX)^2))}{e^{\lambda^2 t^2}} \\
&\leq \frac{1+K_1}{e^{\frac{K_2}{2}t^2}} \\
&= (1+K_1)e^{\frac{-K_2}{2}t^2}
\end{aligned}
$$

Recall, we have proved in the first step $K_1 = 2c_3$, $K_2 = \frac{c_4}{4}$, plug in,

$$
P(|X-EX| \geq t) = (1+2c_3)e^{\frac{-c_4}{8}t^2}
$$

4 Question 4

Question: Given a positive semidefinite matrix $Q \in \mathbb{R}^{n\times n}$, consider $Z = \sum_{i,j} Q_{ij} X_i X_j$. When $X_i \sim N(0,1)$, prove the Hanson-Wright inequality.

$$P\left(Z \geq \mathrm{trace}(Q) + t\right) \leq 2\exp\left(-\min\left\{c_1 t/\|Q\|_{op}, c_2 t^2/\|Q\|_F^2\right\}\right),$$

where $\|Q\|_{op}$ and $\|Q\|_F$ denote the operator and frobenius norms respectively. *Hint: The rotation-invariance of the Gaussian distribution and sub-exponential nature of χ^2-variables could be useful.*
Proof:

$$Z = \sum_{ij} Q_{ij} x_i x_j = x^T Q x$$

$$\begin{aligned}
E(Z) &= E(\sum_{ij} Q_{ij} x_i x_j =) = E(\sum_i Q_{ii} x_i^2) + E(\sum_{j\neq i} Q_{ij} x_i x_j) \\
&= \sum_i Q_{ii} E(x_i^2) + \sum_{j\neq i} Q_{ij} E(x_i x_j) \\
&= \sum_i Q_{ii} \quad (since E(x_i^2) = 0, \quad E(x_i x_j) = 0) \\
&= trace(Q)
\end{aligned}$$

Thus, to prove the bound for $P\left(Z \geq \mathrm{trace}(Q) + t\right)$, we are proving the bound for $P\left(Z - E(Z) \geq t\right)$, the concerntration equality for random variable Z.

We first do an eigen decomposition of Q,

$$Q = P^T \Lambda P$$

$$Z = x^T Q x = (P^T x)^T \Lambda (P^T x) = y^T \Lambda y = \sum \lambda_i y_i^2$$

if we let $y = P^T x$, P is orthonormal, y is still multivariate normal with $N(0, I)$ due to the rotation invarariance of x,thus, $y_i \sim N(0,1)$.

y_i^2 is χ_1^2, is sub-exponential (2,4), meaning $E(e^{ty_i}) \leq e^{2t^2 y_i^2}$ when $t < \frac{1}{4}$, i.e.,

$$E(e^{t\lambda_i y_i}) \leq e^{2t^2 \lambda_i^2 y_i^2} \quad \text{when } t < \frac{1}{4\max_i |\lambda_i|}$$

which means, $\lambda_i y_i^2 \sim Sub - Exp(2\lambda_i^2, 4\max_i|\lambda_i|)$.

$$Z = \sum \lambda_i y_i^2 \sim Sub - Exp(2\sqrt{\sum_i (\lambda_i^2)}, max4|\lambda_i|)$$

We also know,

$$\begin{aligned} ||Q||_F^2 &= tr(Q^TQ) = tr(P^T \Lambda P P^T \Lambda P^T \\ &= tr(P^T\Lambda^2 P) = tr(\Lambda^2 PP^T) \\ &= \sum_i \lambda_i^2 \end{aligned}$$

$$||Q||_2 = \sup_{||x||_2=1} ||Qx||_2 = max_i|\lambda_i|$$

Thus,

$$Z \sim Sub - Exp(2||Q||_F, 4||Q||_2)$$

$P(Z - EZ > t) < exp(-\frac{t^2}{2*4*||Q||_F^2}) = exp(-\frac{t^2}{8||Q||_F^2})$ when $t < \frac{4||Q||_F^2}{4||Q||_2}$, and $P(Z - EZ > t) < exp(-\frac{t}{2*4*||Q||_2}) = exp(-\frac{t}{8||Q||_2})$ otherwise.

$$P(Z - EZ > t) < \max(exp(-\frac{t^2}{8||Q||_F^2}), exp(-\frac{t}{8||Q||_2}))$$

Therefore,

$$P(Z - EZ > t) < (exp(-\min(\frac{c_1 t^2}{||Q||_F^2}), exp(-\frac{c_2 t}{||Q||_2})))$$