Solution 2: 26 February 2019
CS-526 Learning Theory

1. For every $\alpha \in [0,1]$, a convex function f satisfies

$$f(\alpha a + (1-\alpha)b) \leq \alpha f(a) + (1-\alpha) f(b).$$

Substituting $f(X) = e^{\lambda X}$ and $\alpha = \frac{b-X}{b-a} \in [0,1]$ we get

$$e^{\lambda X} \leq \frac{b-X}{b-a} e^{\lambda a} + \frac{X-a}{b-a} e^{\lambda b}.$$

Taking the expectation on both sides and using $\mathbb{E}[X] = 0$ we have

$$\mathbb{E}[e^{\lambda X}] \leq \frac{b}{b-a} e^{\lambda a} - \frac{a}{b-a} e^{\lambda b}.$$

2. With $p = -a/(b-a)$ and $h = \lambda(b-a)$, we have

$$\begin{aligned}
\log(\frac{b}{b-a} e^{\lambda a} - \frac{a}{b-a} e^{\lambda b}) &= \log(e^{\lambda a}) + \log(\frac{b}{b-a} - \frac{a}{b-a} e^{\lambda(b-a)}) \\
&= \lambda a + \log(1 + \frac{a}{b-a} - \frac{a}{b-a} e^{\lambda(b-a)}) \\
&= -hp + \log(1 - p + pe^h).
\end{aligned}$$

3. Let $\theta = \frac{pe^h}{1-p+pe^h}$. One can compute

$$L'(h) = -p + \theta, \qquad L''(h) = \theta(1-\theta) = -(\theta - \frac{1}{2})^2 + \frac{1}{4} \leq \frac{1}{4}.$$

One can also verify $L(0) = L'(0) = 0$. Using these remarks on the equation $L(h) = L(0) + hL'(0) + (h^2/2)L''(\xi)$, we obtain $L(h) \leq h^2/8$. Combining with the previous steps implies

$$\mathbb{E}[e^{\lambda X}] \leq e^{L(\lambda(b-a))} \leq e^{-\lambda^2(a-b)^2/8}.$$

4. Let $X_i = Z_i - \mu$ and $\bar{X} = \frac{1}{m}\sum_{i=1}^m X_i$. Using the monotonicity of the exponent function and Markov's inequality, we have

$$\mathbb{P}\big(\bar{X} \geq \epsilon\big) = \mathbb{P}(e^{\lambda \bar{X}} \geq e^{\lambda \epsilon}) \leq e^{-\lambda \epsilon}\, \mathbb{E}[e^{\lambda \bar{X}}].$$

As X_i are independent, we have $\mathbb{E}[e^{\lambda \bar{X}}] = \prod_{i=1}^m \mathbb{E}[e^{\lambda X_i/m}]$. Also, the previous exercise provides $\mathbb{E}[e^{\lambda X_i/m}] \leq e^{-\lambda^2(a-b)^2/(8m^2)}$. So we conclude

$$\mathbb{P}(\bar{X} \geq \epsilon) \leq \exp\big(-\lambda\epsilon + \frac{\lambda^2(b-a)^2}{8m}\big).$$

5. The exponent $-\lambda\epsilon + \frac{\lambda^2(b-a)^2}{8m}$ is a quadratic (convex) function of λ. It is minimized when $\lambda = 4m\epsilon/(b-a)^2$. This optimization gives the desired bound.