Lemma 2.2.1 (Hoeffding)**.** Let $U \in \mathbb{R}$ be a random variable, such that $U \in [a, b]$ a.s. for some finite $a < b$. Then, for every $t \in \mathbb{R}$,

$$\mathbb{E}\left[\exp\big(t(U - \mathbb{E}U)\big)\right] \le \exp\left(\frac{t^2(b-a)^2}{8}\right). \tag{2.2.3}$$

Proof. For every $p \in [0, 1]$ and $\lambda \in \mathbb{R}$, let us define the function

$$H_p(\lambda) \triangleq \ln\left(pe^{\lambda(1-p)} + (1-p)e^{-\lambda p}\right). \tag{2.2.4}$$

Let $\xi = U - \mathbb{E}U$, where $\xi \in [a - \mathbb{E}U, b - \mathbb{E}U]$. Using the convexity of the exponential function, we can write

$$\begin{aligned}\exp(t\xi) &= \exp\left(\frac{U-a}{b-a} \cdot t(b - \mathbb{E}U) + \frac{b-U}{b-a} \cdot t(a - \mathbb{E}U)\right) \\ &\le \left(\frac{U-a}{b-a}\right) \exp\big(t(b - \mathbb{E}U)\big) + \left(\frac{b-U}{b-a}\right) \exp\big(t(a - \mathbb{E}U)\big).\end{aligned}$$

Taking expectations of both sides, we get

$$\begin{aligned}\mathbb{E}[\exp(t\xi)] &\le \left(\frac{\mathbb{E}U-a}{b-a}\right) \exp\big(t(b - \mathbb{E}U)\big) + \left(\frac{b-\mathbb{E}U}{b-a}\right) \exp\big(t(a - \mathbb{E}U)\big) \\ &= \exp\big(H_p(\lambda)\big) \qquad (2.2.5)\end{aligned}$$

where we have let

$$p = \frac{\mathbb{E}U - a}{b-a} \qquad \text{and} \qquad \lambda = t(b-a).$$

In the following, we show that for every $\lambda \in \mathbb{R}$

$$H_p(\lambda) \le \frac{\lambda^2}{8}, \quad \forall\, p \in [0,1]. \tag{2.2.6}$$

From (2.2.4), we have

$$H_p(\lambda) = -\lambda p + \ln\big(pe^{\lambda} + (1-p)\big), \tag{2.2.7}$$

$$H_p'(\lambda) = -p + \frac{pe^{\lambda}}{pe^{\lambda} + 1 - p}, \tag{2.2.8}$$

$$H_p''(\lambda) = \frac{p(1-p)e^{\lambda}}{\big(pe^{\lambda} + (1-p)\big)^2}. \tag{2.2.9}$$

From (2.2.7)–(2.2.9), we have $H_p(0) = H_p'(0) = 0$, and

$$\begin{aligned}H_p''(\lambda) &= \frac{1}{4}\, \frac{pe^{\lambda} \cdot (1-p)}{\left(\frac{pe^{\lambda} + (1-p)}{2}\right)^2} \\ &\le \frac{1}{4}, \quad \forall\, \lambda \in \mathbb{R},\ p \in [0,1]\end{aligned}$$

where the last inequality holds since the geometric mean is less than or equal to the arithmetic mean. Using a Taylor's series expansion, there exists an intermediate value $\theta \in [0, \lambda]$ (or $\theta \in [\lambda, 0]$ if $t < 0$) such that

$$H_p(\lambda) = H_p(0) + H_p'(0)\lambda + \frac{1}{2}\, H_p''(\theta)\, \lambda^2$$

so, consequently, (2.2.6) holds. Substituting this bound into (2.2.5) and using the above definitions of p and λ, we get (2.2.3). □