Hoeffding 不等式的推导基于概率论和一些基本的数学不等式（如 Chernoff 界限和鞅技巧）。下面是一个较为清晰的推导过程，分为主要步骤：

---

1. **问题背景**

考虑 n 个独立随机变量 $X_1, X_2, \ldots, X_n$，其中每个 X_i 都满足 $a_i \leq X_i \leq b_i$。我们要研究 $S_n = \sum_{i=1}^n X_i$ 偏离其期望 $\mu = \mathbb{E}[S_n]$ 的概率。

目标是给出：

$$\mathbb{P}(S_n - \mu \geq t) \quad 和 \quad \mathbb{P}(S_n - \mu \leq -t)$$

的上界。

---

2. **引入** Chernoff **界限**

利用 Chernoff 界限, 通过指数函数放大偏差的影响, 首先处理 $\mathbb{P}(S_n - \mu \geq t)$:

$$\mathbb{P}(S_n - \mu \geq t) = \mathbb{P}(e^{\lambda(S_n - \mu)} \geq e^{\lambda t}),$$

其中 $\lambda > 0$ 是自由选择的参数。

利用 Markov 不等式（后有证明）:

$$\mathbb{P}(e^{\lambda(S_n - \mu)} \geq e^{\lambda t}) \leq \frac{\mathbb{E}[e^{\lambda(S_n - \mu)}]}{e^{\lambda t}}.$$

因此，有:

$$\mathbb{P}(S_n - \mu \geq t) \leq e^{-\lambda t}\mathbb{E}[e^{\lambda(S_n - \mu)}].$$

---

3. **计算期望** $\mathbb{E}[e^{\lambda(S_n - \mu)}]$

由于 $S_n = \sum_{i=1}^n X_i$，且 X_i 是独立的，因此:

$$\mathbb{E}[e^{\lambda(S_n - \mu)}] = \prod_{i=1}^n \mathbb{E}[e^{\lambda(X_i - \mathbb{E}[X_i])}].$$

---

4. **处理单个随机变量** X_i

因为 $a_i \le X_i \le b_i$，定义 X_i 的偏移值 $Z_i = X_i - \mathbb{E}[X_i]$。则有：

$$\mathbb{E}[e^{\lambda Z_i}] \le \exp\left(\frac{\lambda^2(b_i - a_i)^2}{8}\right),$$

这里用到了 Hoeffding 引理（详见附录）。

---

5. **结合结果**

将每个随机变量的期望结果代入总的期望中，有：

$$\mathbb{E}[e^{\lambda(S_n-\mu)}] \le \prod_{i=1}^{n} \exp\left(\frac{\lambda^2(b_i - a_i)^2}{8}\right) = \exp\left(\frac{\lambda^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2\right).$$

因此：

$$\mathbb{P}(S_n - \mu \ge t) \le \exp\left(-\lambda t + \frac{\lambda^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2\right).$$

---

6. **优化参数** λ

为了使右侧的上界尽可能小，对 λ 求最优值。取：

$$\lambda = \frac{4t}{\sum_{i=1}^{n}(b_i - a_i)^2}.$$

代入后，上界变为：

$$\mathbb{P}(S_n - \mu \ge t) \le \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

---

7. **得到 Hoeffding 不等式**

类似地，通过对 $\mathbb{P}(S_n - \mu \le -t)$ 的推导，可以得到完全对称的结果。最终，Hoeffding 不等式为：

$$\mathbb{P}(|S_n - \mu| \ge t) \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

---

8. 附录：Hoeffding 引理

Hoeffding 引理是推导的关键步骤，用于处理单个有界随机变量的偏差。设随机变量 X 满足 $a \leq X \leq b$，则对于任何 $\lambda > 0$：

$$\mathbb{E}[e^{\lambda(X-\mathbb{E}[X])}] \leq \exp\left(\frac{\lambda^2(b-a)^2}{8}\right).$$

这个结果可以通过凸函数的 Jensen 不等式和有界变量的性质证明。

Markov 不等式

1. 指标函数的定义

我们用指标函数 $1_{X\geq a}$ 来表示事件 $X \geq a$，它的定义是：

$$1_{X\geq a} = \begin{cases} 1, & \text{如果} X \geq a; \\ 0, & \text{如果} X < a. \end{cases}$$

这表示：

- 当 $X \geq a$ 时，指标函数 $1_{X\geq a} = 1$，说明事件 $X \geq a$ 发生；
- 当 $X < a$ 时，指标函数 $1_{X\geq a} = 0$，说明事件 $X \geq a$ 没有发生。

通过这个定义可以看出，指标函数在 $X < a$ 时自然等于 0。

---

2. 不等式如何成立

为了利用指标函数进行上界处理，我们需要分析 $1_{X\geq a} \leq \frac{X}{a}$ 这个不等式是如何成立的。

(1) 当 $X \geq a$ 时

如果 $X \geq a$，则指标函数 $1_{X\geq a} = 1$。
而 $\frac{X}{a} \geq 1$（因为 X 比 a 更大)。
因此，在这种情况下：

$$1_{X\geq a} = 1 \leq \frac{X}{a}.$$

(2) 当 $X < a$ 时

如果 $X < a$，则指标函数 $1_{X \geq a} = 0$。
而 $\frac{X}{a} > 0$（因为 $X \geq 0$，且 $a > 0$）。
因此，在这种情况下:

$$1_{X \geq a} = 0 \leq \frac{X}{a}.$$

---

3. 整体结论

无论 $X \geq a$ 还是 $X < a$，都有:

$$1_{X \geq a} \leq \frac{X}{a}.$$

---

4. 不等式用于期望

对两边取期望后，利用随机变量的性质和指标函数的定义，得到 Markov 不等式。