Lecture 5: Concentration Inequalities and Uniform Convergence

From Planning to Learning

In the previous lectures, we studied *planning*: given complete knowledge of an MDP (the transition function $\mathcal{P}$ and reward function R), we showed how to compute optimal policies via Value Iteration and Policy Iteration. In practice, however, the MDP model is rarely known. Instead, the agent must *learn* from data—samples of transitions and rewards collected through interaction with the environment.

The Learning Challenge: When we replace exact expectations with sample averages, errors are introduced. To analyze learning algorithms rigorously, we need tools that quantify how much an empirical estimate can deviate from its true (population) value. These tools are **concentration inequalities** and **uniform convergence** bounds.

This lecture develops the probabilistic machinery that will be used throughout the remainder of the course. We begin with concentration inequalities for sums of independent random variables, then extend to martingale differences (which naturally arise in RL), and finally develop uniform convergence results that allow us to reason about function classes rather than individual functions.

Concentration Inequalities for Independent Random Variables

The simplest statistical question is: given n independent samples $X_1, \ldots, X_n$ of a random variable, how close is the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$ to the true mean $\mathbb{E}[X]$? We know from the Law of Large Numbers that the sample mean converges to the true mean as $n \to \infty$, but we need *quantitative*, *finite-sample* bounds.

Markov's Inequality

The most basic tail bound follows from a simple observation about nonnegative random variables.

Lemma 1 (Markov's Inequality)**.** *Let X be a nonnegative random variable. Then for all $a > 0$,*

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Proof. From the definition of expectation:

$$\mathbb{E}[X] \geq \Pr(X < a) \cdot 0 + \Pr(X \geq a) \cdot a = \Pr(X \geq a) \cdot a.$$

Rearranging gives the result. □

Markov's inequality is very general but also very loose. By applying it to carefully chosen transformations of the random variable, we can obtain much sharper bounds.

Sub-Gaussian Random Variables

The key ingredient for deriving Hoeffding's inequality is the notion of sub-Gaussian random variables, which captures the idea that a random variable's tails decay at least as fast as those of a Gaussian.

Definition 1 (Sub-Gaussian Random Variable)**.** *A random variable X is σ-*sub-Gaussian *if for all $\lambda \in \mathbb{R}$,*

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

The parameter σ is called the sub-Gaussian parameter*.*

Remark 1. *Any Gaussian random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is σ-sub-Gaussian (with equality in the definition). The name "sub-Gaussian" reflects that the moment generating function is bounded by that of a Gaussian.*

The following lemma shows that bounded random variables are sub-Gaussian, which is the bridge to Hoeffding's inequality.

Lemma 2 (Hoeffding's lemma, Bounded Implies Sub-Gaussian)**.** *If X is a random variable with $X \in [a, b]$ almost surely, then X is $\frac{b-a}{2}$-sub-Gaussian. That is,*

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

Proof. Let $Z = X - \mathbb{E}[X]$ and define the *cumulant generating function* $\psi(\lambda) = \log \mathbb{E}[e^{\lambda Z}]$. By

Taylor's theorem, for some λ' between 0 and λ:

$$\psi(\lambda) = \psi(0) + \lambda\psi'(0) + \frac{\lambda^2}{2}\psi''(\lambda').$$

We compute: $\psi(0) = \log \mathbb{E}[e^0] = 0$. For the first derivative, $\psi'(\lambda) = \frac{\mathbb{E}[Z\, e^{\lambda Z}]}{\mathbb{E}[e^{\lambda Z}]}$, so $\psi'(0) = \frac{\mathbb{E}[Z\cdot 1]}{\mathbb{E}[1]} = \mathbb{E}[Z] = 0$ (since Z is centered).

For the second derivative, a direct calculation gives:

$$\psi''(\lambda') = \frac{\mathbb{E}[Z^2 e^{\lambda' Z}]}{\mathbb{E}[e^{\lambda' Z}]} - \left(\frac{\mathbb{E}[Z\, e^{\lambda' Z}]}{\mathbb{E}[e^{\lambda' Z}]}\right)^2.$$

Let $c = a - \mathbb{E}[X]$, $d = b - \mathbb{E}[X]$, so $Z \in [c, d]$ with $d - c = b - a$. Since $(Z - c)(d - Z) \geq 0$ pointwise, we have $Z^2 \leq (c + d)Z - cd$. Multiplying both sides by $e^{\lambda' Z} > 0$, taking expectations, and dividing by $\mathbb{E}[e^{\lambda' Z}] > 0$ (both steps preserve the inequality):

$$\frac{\mathbb{E}[Z^2 e^{\lambda' Z}]}{\mathbb{E}[e^{\lambda' Z}]} \leq (c + d)\, m - cd, \qquad \text{where } m := \frac{\mathbb{E}[Z\, e^{\lambda' Z}]}{\mathbb{E}[e^{\lambda' Z}]}.$$

Therefore $\psi''(\lambda') \leq (c + d)\, m - cd - m^2 = (m - c)(d - m) \leq \left(\frac{d-c}{2}\right)^2 = \frac{(b-a)^2}{4}$, where the last inequality uses $uv \leq \left(\frac{u+v}{2}\right)^2$.

Therefore $\psi(\lambda) \leq \frac{\lambda^2(b-a)^2}{8}$, which gives $\mathbb{E}[e^{\lambda Z}] \leq e^{\lambda^2(b-a)^2/8}$. □

Hoeffding's Inequality

We are now ready to state and prove the main concentration result for bounded independent random variables.

Theorem 3 (Hoeffding's Inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in [a_i, b_i]$ almost surely. Let $R_i = b_i - a_i$. Then for all $t \geq 0$,*

$$\Pr\left(\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} R_i^2}\right).$$

Proof. The proof uses the "Chernoff method": exponentiating, applying Markov's inequality, then optimizing.

Step 1: Exponential Markov. For any $\lambda > 0$:

$$\Pr\left(\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq t\right) = \Pr\left(e^{\lambda \sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])} \geq e^{\lambda t}\right)$$
$$\leq \frac{\mathbb{E}\left[e^{\lambda \sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])}\right]}{e^{\lambda t}}. \qquad \text{(Markov's inequality)}$$

Step 2: Factor by independence. Since $X_1, \ldots, X_n$ are independent:

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

Step 3: Apply sub-Gaussian bound. By Lemma 2, each factor satisfies $\mathbb{E}[e^{\lambda(X_i - \mathbb{E}[X_i])}] \leq e^{\lambda^2 R_i^2/8}$. Therefore:

$$\Pr\left(\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq t\right) \leq e^{-\lambda t + \frac{\lambda^2}{8}\sum_{i=1}^{n} R_i^2}.$$

Step 4: Optimize over λ. The exponent $-\lambda t + \frac{\lambda^2}{8}\sum R_i^2$ is a quadratic in λ, minimized at $\lambda^\star = \frac{4t}{\sum R_i^2}$. Substituting:

$$\Pr\left(\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} R_i^2}\right). \qquad \square$$

We can reformulate Hoeffding's inequality in the more convenient "with high probability" form.

Corollary 4 (Hoeffding, High-Probability Form)**.** *Under the conditions of Theorem 3, with probability at least $1 - \delta$,*

$$\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \leq \sqrt{\frac{\sum_{i=1}^{n} R_i^2}{2} \log \frac{1}{\delta}}.$$

Proof. Set the right-hand side of Theorem 3 equal to δ and solve for t. $\square$

For i.i.d. random variables, the bound takes a particularly clean form.

Corollary 5 (Hoeffding for i.i.d. Variables)**.** *Let $X_1, \ldots, X_n$ be i.i.d. with $X_i \in [a, b]$ and*

$R = b - a$. Then with probability at least $1-\delta$,

$$\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]\right| \le \frac{R}{\sqrt{n}}\sqrt{\frac{1}{2}\log\frac{2}{\delta}} = \widetilde{O}\left(\frac{R}{\sqrt{n}}\right).$$

Here and throughout, $\widetilde{O}(\cdot)$ hides logarithmic factors (e.g., in $1/\delta$ and n).

Proof. Apply Corollary 4 with $R_i = R$ and divide by n. To get the two-sided bound, apply the one-sided bound to both $\sum(X_i - \mathbb{E}[X_i])$ and $\sum(\mathbb{E}[X_i] - X_i)$, then take a union bound (using $\delta/2$ for each side). □

The $1/\sqrt{n}$ Rate: The key message of Hoeffding's inequality is that the sample mean concentrates around the true mean at a rate of $O(R/\sqrt{n})$ (ignoring logarithmic factors). This is the canonical rate for estimating expectations from independent samples, and matches the Central Limit Theorem scaling.

Bernstein's Inequality

Hoeffding's inequality uses only the *range* of the random variables. When the variance is much smaller than the range, Bernstein's inequality provides a sharper bound by incorporating variance information.

Theorem 6 (Bernstein's Inequality). *Let $X_1, \ldots, X_n$ be independent random variables with $|X_i - \mathbb{E}[X_i]| \le R$ almost surely and $\mathrm{Var}(X_i) \le \sigma_i^2$. Then for all $t \ge 0$,*

$$\Pr\left(\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \ge t\right) \le \exp\left(-\frac{t^2}{2\sum_{i=1}^{n}\sigma_i^2 + \frac{2}{3}Rt}\right).$$

Corollary 7 (Bernstein, High-Probability Form). *Under the conditions of Theorem 6, with probability at least $1-\delta$,*

$$\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \le O\left(\sqrt{\sum_{i=1}^{n}\sigma_i^2 \cdot \log\frac{1}{\delta}} + R\log\frac{1}{\delta}\right).$$

Corollary 8 (Bernstein for i.i.d. Variables). *Let $X_1, \ldots, X_n$ be i.i.d. with $|X_i - \mathbb{E}[X_i]| \le R$ and $\mathrm{Var}(X_i) \le \sigma^2$. Then with probability at least $1-\delta$,*

$$\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]\right| \le \widetilde{O}\left(\frac{\sigma}{\sqrt{n}} + \frac{R}{n}\right).$$

Comparison with Hoeffding. Hoeffding gives a bound of $\widetilde{O}(R/\sqrt{n})$, while Bernstein gives $\widetilde{O}(\sigma/\sqrt{n} + R/n)$. When $\sigma \ll R$ (small variance relative to the range), Bernstein is significantly sharper. The first term $\sigma/\sqrt{n}$ is the "variance term" (dominant for moderate n) and the second term R/n is the "range term" (dominant for small n or when σ is very small).

Example: Bernoulli Random Variables. Consider estimating the bias p of a coin from n i.i.d. tosses $X_i \sim \text{Bernoulli}(p)$. Here $R = 1$ and $\sigma^2 = p(1-p) \le p$. Hoeffding gives $|\frac{1}{n}\sum X_i - p| \le \widetilde{O}(1/\sqrt{n})$. Bernstein gives $|\frac{1}{n}\sum X_i - p| \le \widetilde{O}(\sqrt{p/n} + 1/n)$, which is much tighter when p is small.

Concentration for Martingale Differences

In reinforcement learning, the independence assumption is almost never satisfied: the agent's actions depend on previously observed states and rewards, creating dependencies between successive random variables. We need concentration inequalities that work under weaker conditions.

Motivation: Why Independence Fails in RL

Consider a Multi-Armed Bandit (MAB) problem with K arms, where arm j yields reward $R_j^{(i)} \sim \text{Bernoulli}(p_j)$. At each round i, the learner pulls arm a_i and observes reward $X_i = R_{a_i}^{(i)} - p_{a_i}$ (centered). The arm a_i is chosen based on the entire history of past observations:

$$a_i = f(r_i, X_{i-1}, a_{i-1}, \ldots, X_1, a_1)$$

where r_i denotes internal randomness. Since X_i depends on a_i, which depends on $X_1, \ldots, X_{i-1}$, the sequence $\{X_i\}$ is *not* independent.

However, observe that conditioned on $\mathcal{F}_{i-1}$ (the history up to round $i-1$) *and* a_i, the reward $R_{a_i}^{(i)}$ is an independent draw with mean p_{a_i}, so $\mathbb{E}[R_{a_i}^{(i)} - p_{a_i} \mid a_i, \mathcal{F}_{i-1}] = 0$. By the tower property:

$$\mathbb{E}[X_i \mid \mathcal{F}_{i-1}] = \mathbb{E}\big[\mathbb{E}[R_{a_i}^{(i)} - p_{a_i} \mid a_i, \mathcal{F}_{i-1}] \,\big|\, \mathcal{F}_{i-1}\big] = 0.$$

This is the *martingale difference* property, which turns out to be sufficient for deriving concentration bounds.

Martingale Differences

Definition 2 (Martingale Difference Sequence)**.** *Let $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n$ be a filtration (an increasing sequence of σ-algebras). A sequence $\{X_i - \mathbb{E}[X_i]\}_{i=1}^n$ is a* martingale difference sequence *with respect to $\{\mathcal{F}_i\}$ if X_i is $\mathcal{F}_i$-measurable and*

$$\mathbb{E}[X_i - \mathbb{E}[X_i] \mid \mathcal{F}_{i-1}] = 0, \qquad \textit{equivalently,} \qquad \mathbb{E}[X_i \mid \mathcal{F}_{i-1}] = \mathbb{E}[X_i].$$

Key Property: Conditioned on the past, the "surprise" $X_i - \mathbb{E}[X_i]$ has zero mean. This does *not* require X_i to be independent of $\mathcal{F}_{i-1}$—the conditional *distribution* of X_i can depend on the history, as long as the conditional *mean* equals the unconditional mean.

The partial sums $S_t = \sum_{i=1}^t (X_i - \mathbb{E}[X_i])$ form a *martingale*: $\mathbb{E}[S_{t+1} \mid \mathcal{F}_t] = S_t$. Intuitively, no matter what history we condition on, the expected value of the next increment is zero.

Azuma–Hoeffding Inequality

The following is the martingale analogue of Hoeffding's inequality. The proof technique is essentially the same—the Chernoff method—but uses the tower property of conditional expectations in place of the independence factorization.

Theorem 9 (Azuma–Hoeffding Inequality)**.** *Suppose $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n$ is a filtration and $X_1, \ldots, X_n$ are random variables such that X_i is $\mathcal{F}_i$-measurable. If:*

(i) $\mathbb{E}[X_i - \mathbb{E}[X_i] \mid \mathcal{F}_{i-1}] = 0$ *(martingale difference condition),*

(ii) $X_i - \mathbb{E}[X_i] \in [a_i, b_i]$ *almost surely, with* $R_i := b_i - a_i$ *(bounded range),*

then for all $t \geq 0$,

$$\Pr\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n R_i^2}\right).$$

Proof. The proof follows the same Chernoff method as Hoeffding's inequality, but replaces the independence factorization with the tower property.

For any $\lambda > 0$, by Markov's inequality:

$$\Pr\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \geq t\right) \leq e^{-\lambda t} \cdot \mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right)\right].$$

We bound the moment generating function using the tower property. Let $Z_i = X_i - \mathbb{E}[X_i]$.

Then:

$$\begin{aligned}\mathbb{E}\left[e^{\lambda\sum_{i=1}^{n}Z_i}\right] &= \mathbb{E}\left[\mathbb{E}\left[e^{\lambda Z_n}\cdot e^{\lambda\sum_{i=1}^{n-1}Z_i}\;\middle|\;\mathcal{F}_{n-1}\right]\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[e^{\lambda Z_n}\mid\mathcal{F}_{n-1}\right]\cdot e^{\lambda\sum_{i=1}^{n-1}Z_i}\right].\end{aligned}$$

In the second step, we used the fact that $\sum_{i=1}^{n-1} Z_i$ is $\mathcal{F}_{n-1}$-measurable, so it can be taken outside the conditional expectation.

Now, conditioned on $\mathcal{F}_{n-1}$, Z_n satisfies the martingale difference condition $\mathbb{E}[Z_n \mid \mathcal{F}_{n-1}] = 0$ and is bounded in an interval of length R_n. By the same sub-Gaussian argument as in Lemma 2 (applied conditionally):

$$\mathbb{E}[e^{\lambda Z_n}\mid\mathcal{F}_{n-1}] \leq \exp\left(\frac{\lambda^2 R_n^2}{8}\right).$$

Therefore:

$$\mathbb{E}\left[e^{\lambda\sum_{i=1}^{n}Z_i}\right] \leq \exp\left(\frac{\lambda^2 R_n^2}{8}\right)\mathbb{E}\left[e^{\lambda\sum_{i=1}^{n-1}Z_i}\right].$$

Applying this peeling argument inductively from $i = n$ down to $i = 1$:

$$\mathbb{E}\left[e^{\lambda\sum_{i=1}^{n}Z_i}\right] \leq \exp\left(\frac{\lambda^2\sum_{i=1}^{n}R_i^2}{8}\right).$$

Optimizing over λ as in the proof of Theorem 3 yields the result. □

Azuma–Bernstein Inequality

The variance-sensitive version also extends to the martingale setting.

Theorem 10 (Azuma–Bernstein Inequality)**.** *Suppose $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n$ is a filtration and $X_1, \ldots, X_n$ are random variables such that X_i is $\mathcal{F}_i$-measurable. If:*

(i) $\mathbb{E}[X_i - \mathbb{E}[X_i] \mid \mathcal{F}_{i-1}] = 0$,

(ii) $\mathrm{Var}(X_i \mid \mathcal{F}_{i-1}) \leq \sigma_i^2$,

(iii) $X_i - \mathbb{E}[X_i] \leq R$ *almost surely,*

then for all $t \geq 0$,

$$\Pr\left(\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq t\right) \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^{n}\sigma_i^2 + \frac{2}{3}Rt}\right).$$

Remark 2 (Two-Sided Bound)**.** *Theorem 10 gives an upper tail bound under the one-sided condition $X_i - \mathbb{E}[X_i] \le R$. To obtain a two-sided bound, apply the theorem to $-X_i$ (which satisfies $-(X_i - \mathbb{E}[X_i]) \le R$ when $|X_i - \mathbb{E}[X_i]| \le R$), then take a union bound: with probability at least $1-\delta$,*

$$\left|\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])\right| \le O\left(\sqrt{\sum_{i=1}^{n}\sigma_i^2 \cdot \log\frac{1}{\delta}} + R\log\frac{1}{\delta}\right).$$

This recovers the form of Bernstein's inequality (Corollary 7) when $X_1, \ldots, X_n$ are independent.

Comparison. The Azuma–Hoeffding and Azuma–Bernstein inequalities reduce to Hoeffding's and Bernstein's inequalities, respectively, when $X_1, \ldots, X_n$ are independent. For i.i.d. variables with range R and variance σ^2:

Inequality	**Bound on** $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])$ **(w.p.** $1-\delta$**)**
Azuma–Hoeffding	$\widetilde{O}\left(\frac{R}{\sqrt{n}}\right)$
Azuma–Bernstein	$\widetilde{O}\left(\frac{\sigma}{\sqrt{n}} + \frac{R}{n}\right)$

The Azuma–Bernstein bound is strictly better when $\sigma \ll R$.

Uniform Convergence

So far, our concentration bounds apply to a *fixed* function or random variable. In learning problems, however, the function we care about (e.g., the learned policy, the estimated value function) is itself *chosen based on the data*. We need bounds that hold *simultaneously* for all functions in a class.

The Problem

Suppose $X_1, \ldots, X_n$ are i.i.d. random variables taking values in a finite set $\mathcal{S}$ with $|\mathcal{S}| = S$. Let $f : \mathcal{S} \to [0, 1]$ be a function.

Case 1: f is fixed (independent of data). By Hoeffding's inequality:

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]\right| \ge t\right) \le 2\exp\left(-2nt^2\right).$$

Case 2: f depends on data. Suppose $\hat{f} = \operatorname{argmin}_{f \in \mathcal{F}} \ell(f, \{X_i\}_{i=1}^n)$ for some loss ℓ and function class $\mathcal{F}$. Since $\hat{f}$ depends on the data, we cannot directly apply Hoeffding. However, we can *relax* to a uniform bound:

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n}\hat{f}(X_i)-\mathbb{E}[\hat{f}(X)]\right|\geq t\right)\leq \Pr\left(\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}f(X_i)-\mathbb{E}[f(X)]\right|\geq t\right).$$

The right-hand side is a *uniform deviation bound*—it bounds the worst-case estimation error over the entire function class.

Union Bound

The key tool for converting pointwise bounds to uniform bounds is the union bound.

Lemma 11 (Union Bound)**.** *For any events $A_1, A_2, \ldots, A_m$,*

$$\Pr\left(\bigcup_{j=1}^{m}A_j\right)\leq\sum_{j=1}^{m}\Pr(A_j).$$

Finite Function Classes

When $\mathcal{F}$ is a finite set, we can directly apply the union bound.

Theorem 12 (Uniform Convergence for Finite Classes)**.** *Let $X_1, \ldots, X_n$ be i.i.d. and let $\mathcal{F}$ be a finite class of functions $f : \mathcal{S} \to [0, 1]$. Then with probability at least $1 - \delta$,*

$$\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}f(X_i)-\mathbb{E}[f(X)]\right|\leq\sqrt{\frac{1}{2n}\log\frac{2|\mathcal{F}|}{\delta}}=O\left(\sqrt{\frac{\log|\mathcal{F}|}{n}}\right).$$

Proof. For each fixed $f \in \mathcal{F}$, Hoeffding's inequality gives:

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n}f(X_i)-\mathbb{E}[f(X)]\right|\geq t\right)\leq 2\exp(-2nt^2).$$

Taking a union bound over all $f \in \mathcal{F}$:

$$\Pr\left(\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}f(X_i)-\mathbb{E}[f(X)]\right|\geq t\right)\leq\sum_{f\in\mathcal{F}}2\exp(-2nt^2)=2|\mathcal{F}|\exp(-2nt^2).$$

Setting $2|\mathcal{F}|\exp(-2nt^2) = \delta$ and solving for t:

$$t = \sqrt{\frac{1}{2n}\log\frac{2|\mathcal{F}|}{\delta}}. \qquad \square$$

The Price of Uniformity: Compared to a single-function bound ($\sqrt{\frac{1}{2n}\log\frac{2}{\delta}}$), the uniform bound has an extra $\log|\mathcal{F}|$ factor. This is the "cost" of ensuring the bound holds simultaneously for all $f \in \mathcal{F}$. Crucially, the dependence on $|\mathcal{F}|$ is only *logarithmic*, so even exponentially large function classes lead to manageable bounds.

Infinite Function Classes via ε-Nets

When $\mathcal{F}$ is infinite (e.g., $\mathcal{F} = \{f \mid f : \mathcal{S} \to [0,1]\}$), the union bound cannot be applied directly. The idea is to *discretize* $\mathcal{F}$ into a finite set and control the discretization error.

Definition 3 (ε-Net)**.** *An* ε-net *of a function class $\mathcal{F}$ (with respect to $\|\cdot\|_\infty$) is a finite subset $\mathcal{F}_\varepsilon \subset \mathcal{F}$ such that for every $f \in \mathcal{F}$, there exists $f_\varepsilon \in \mathcal{F}_\varepsilon$ with*

$$\sup_{x\in\mathcal{S}} |f(x) - f_\varepsilon(x)| \le \varepsilon.$$

The covering number $\mathcal{N}(\mathcal{F}, \varepsilon, \|\cdot\|_\infty)$ *is the minimum size of an ε-net.*

Theorem 13 (Uniform Convergence via Covering Numbers)**.** *Let $\mathcal{F}$ be a class of functions $f : \mathcal{S} \to [0,1]$. Let $X_1, \ldots, X_n$ be i.i.d. taking values in $\mathcal{S}$. Then for any $\varepsilon > 0$, with probability at least $1-\delta$,*

$$\sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)] \right| \le \sqrt{\frac{\log\mathcal{N}(\mathcal{F},\varepsilon,\|\cdot\|_\infty) + \log(2/\delta)}{2n}} + 2\varepsilon.$$

Proof. **Step 1: Reduce to a finite class.** Let $\mathcal{F}_\varepsilon \subset \mathcal{F}$ be an ε-net of $\mathcal{F}$ with $|\mathcal{F}_\varepsilon| = \mathcal{N}(\mathcal{F}, \varepsilon, \|\cdot\|_\infty)$.

Step 2: Apply uniform convergence on $\mathcal{F}_\varepsilon$. By Theorem 12, with probability at least $1-\delta$:

$$\sup_{f_\varepsilon\in\mathcal{F}_\varepsilon} \left| \frac{1}{n}\sum_{i=1}^n f_\varepsilon(X_i) - \mathbb{E}[f_\varepsilon(X)] \right| \le \sqrt{\frac{1}{2n}\log\frac{2|\mathcal{F}_\varepsilon|}{\delta}} = \sqrt{\frac{\log\mathcal{N}(\mathcal{F},\varepsilon,\|\cdot\|_\infty) + \log(2/\delta)}{2n}}.$$

Step 3: Control the discretization error. For any $f \in \mathcal{F}$, let $f_\varepsilon \in \mathcal{F}_\varepsilon$ be its nearest neighbor in the ε-net. Then:

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]\right| &= \left|\frac{1}{n}\sum_{i=1}^{n} f_\varepsilon(X_i) - \mathbb{E}[f_\varepsilon(X)] + \frac{1}{n}\sum_{i=1}^{n}(f - f_\varepsilon)(X_i) - \mathbb{E}[(f - f_\varepsilon)(X)]\right| \\ &\le \underbrace{\left|\frac{1}{n}\sum_{i=1}^{n} f_\varepsilon(X_i) - \mathbb{E}[f_\varepsilon(X)]\right|}_{\text{finite-class uniform convergence}} + \underbrace{2\varepsilon}_{\text{discretization error}}.\end{aligned}$$

The discretization error bound uses $|f(x) - f_\varepsilon(x)| \le \varepsilon$ for all x, so $|\frac{1}{n}\sum_{i=1}^{n}(f - f_\varepsilon)(X_i)| \le \varepsilon$ and $|\mathbb{E}[(f - f_\varepsilon)(X)]| \le \varepsilon$.

Taking the supremum over $f \in \mathcal{F}$ and combining Steps 2–3 gives the result. $\square$

Corollary 14 (Bounded Functions on Finite Domains)**.** *Let $\mathcal{S}$ be a finite set with $|\mathcal{S}| = S$, and let $\mathcal{F} = \{f \mid f : \mathcal{S} \to [0, 1]\}$. Then with probability at least $1 - \delta$,*

$$\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]\right| = O\left(\sqrt{\frac{S\log n}{n}} + \sqrt{\frac{\log(1/\delta)}{n}}\right).$$

Proof. We bound the covering number of $\mathcal{F}$. Define the grid $\mathcal{G}_\varepsilon = \{\varepsilon, 2\varepsilon, \ldots, \lfloor 1/\varepsilon\rfloor\varepsilon\}$ and the finite function class $\mathcal{F}_\varepsilon = \{f \mid f : \mathcal{S} \to \mathcal{G}_\varepsilon\}$. For every $f \in \mathcal{F}$, rounding each value $f(x)$ to the nearest grid point produces some $f_\varepsilon \in \mathcal{F}_\varepsilon$ with $\|f - f_\varepsilon\|_\infty \le \varepsilon$. Since $|\mathcal{G}_\varepsilon| = \lfloor 1/\varepsilon\rfloor \le 1/\varepsilon$ and each function in $\mathcal{F}_\varepsilon$ independently assigns one of $|\mathcal{G}_\varepsilon|$ values to each of the S elements of $\mathcal{S}$:

$$\mathcal{N}(\mathcal{F}, \varepsilon, \|\cdot\|_\infty) \le |\mathcal{F}_\varepsilon| = |\mathcal{G}_\varepsilon|^S \le (1/\varepsilon)^S.$$

Substituting into Theorem 13 and choosing $\varepsilon = 1/\sqrt{n}$:

$$\sqrt{\frac{S\log\sqrt{n} + \log(2/\delta)}{2n}} + \frac{2}{\sqrt{n}} = O\left(\sqrt{\frac{S\log n}{n}} + \sqrt{\frac{\log(1/\delta)}{n}}\right).$$ $\square$

Interpretation for RL: When $\mathcal{S}$ is the state space and $\mathcal{F}$ is the set of all bounded value functions on $\mathcal{S}$, Corollary 14 tells us that $n = O(S\log(S/\varepsilon^2)/\varepsilon^2)$ samples suffice to estimate *any* value function uniformly to accuracy ε (solving the implicit inequality $n \ge S\log n/\varepsilon^2$ via $n = O(A\log A)$ for $A = S/\varepsilon^2$). The factor $S = |\mathcal{S}|$ is the price of not knowing which value function we need to estimate in advance; the $\log(S/\varepsilon^2)$ factor is the price of discretization via the ε-net.

Summary

Key Results:

- **Hoeffding's Inequality:** For bounded independent r.v.'s, $\frac{1}{n}\sum X_i$ concentrates at rate $\widetilde{O}(R/\sqrt{n})$.
- **Bernstein's Inequality:** Sharper bound $\widetilde{O}(\sigma/\sqrt{n} + R/n)$ when variance $\sigma^2 \ll R^2$.
- **Azuma–Hoeffding/Bernstein:** Same bounds under the weaker martingale difference condition (essential for RL where data is dependent).
- **Uniform Convergence (Finite $\mathcal{F}$):** Concentration holds for all $f \in \mathcal{F}$ simultaneously at cost $\sqrt{\log|\mathcal{F}|/n}$ (union bound).
- **Uniform Convergence (Infinite $\mathcal{F}$):** Via ε-nets, achieves $\widetilde{O}(\sqrt{S/n})$ for functions on a state space of size S.