For a fixed $f \in \mathcal{F}$, if we observe $\frac{1}{n} \sum_{i=1}^{n} I\left(f(X_i) \neq Y_i\right)$ is small, can we say that $\mathbb{P}\left(f(X) \neq Y\right)$ is small? By the Law of Large Numbers,

$$\frac{1}{n} \sum_{i=1}^{n} I\left(f(X_i) \neq Y_i\right) \to \mathbb{E} I(f(X) \neq Y) = \mathbb{P}\left(f(X) \neq Y\right).$$

The Central Limit Theorem says

$$\frac{\sqrt{n}\left(\frac{1}{n} \sum_{i=1}^{n} I\left(f(X_i) \neq Y_i\right) - \mathbb{E} I(f(X) \neq Y)\right)}{\sqrt{\mathrm{Var} I}} \to \mathcal{N}(0, 1).$$

Thus,

$$\frac{1}{n} \sum_{i=1}^{n} I\left(f(X_i) \neq Y_i\right) - \mathbb{E} I(f(X) \neq Y) \sim \frac{k}{\sqrt{n}}.$$

Let $Z_1, \cdots, Z_n \in \mathbb{R}$ be i.i.d. random variables. We're interested in bounds on $\frac{1}{n} \sum Z_i - \mathbb{E} Z$.

(1) Jensen's inequality: If ϕ is a convex function, then $\phi(\mathbb{E} Z) \leq \mathbb{E} \phi(X)$.

(2) Chebyshev's inequality: If $Z \geq 0$, then $\mathbb{P}\left(Z \geq t\right) \leq \frac{\mathbb{E} Z}{t}$.

    Proof:

$$\mathbb{E} Z = \mathbb{E} Z I(Z < t) + \mathbb{E} Z I(Z \geq t) \geq \mathbb{E} Z I(Z \geq t)$$

$$\geq \mathbb{E} t I(Z \geq t) = t \mathbb{P}\left(Z \geq t\right).$$

(3) Markov's inequality: Let Z be a signed r.v. Then for any $\lambda > 0$

$$\mathbb{P}\left(Z \geq t\right) = \mathbb{P}\left(e^{\lambda Z} \geq e^{\lambda t}\right) \leq \frac{\mathbb{E} e^{\lambda Z}}{e^{\lambda t}}$$

    and therefore

$$\mathbb{P}\left(Z \geq t\right) \leq \inf_{\lambda > 0} e^{-\lambda t} \mathbb{E} e^{\lambda Z}.$$

Theorem 5.1 (Bennett). *Assume* $\mathbb{E} Z = 0$, $\mathbb{E} Z^2 = \sigma^2$, $|Z| < M = const$, $Z_1, \cdots, Z_n$ *independent copies of* Z, *and* $t \geq 0$. *Then*

$$\mathbb{P}\left(\sum_{i=1}^{n} Z_i \geq t\right) \leq \exp\left(-\frac{n\sigma^2}{M^2} \phi\left(\frac{tM}{n\sigma^2}\right)\right),$$

where $\phi(x) = (1 + x) \log(1 + x) - x$.

1

Proof. Since Z_i are i.i.d.,

$$\mathbb{P}\left(\sum_{i=1}^{n} Z_i \geq t\right) \leq e^{-\lambda t}\mathbb{E}e^{\lambda \sum_{i=1}^{n} Z_i} = e^{-\lambda t}\prod_{i=1}^{n}\mathbb{E}e^{\lambda Z_i} = e^{-\lambda t}\left(\mathbb{E}e^{\lambda Z}\right)^n.$$

Expanding,

$$
\begin{aligned}
\mathbb{E}e^{\lambda Z} &= \mathbb{E}\sum_{k=0}^{\infty}\frac{(\lambda Z)^k}{k!} = \sum_{k=0}^{\infty}\lambda^k\frac{\mathbb{E}Z^k}{k!} \\
&= 1 + \sum_{k=2}^{\infty}\frac{\lambda^k}{k!}\mathbb{E}Z^2 Z^{k-2} \leq 1 + \sum_{k=2}^{\infty}\frac{\lambda^k}{k!}M^{k-2}\sigma^2 \\
&= 1 + \frac{\sigma^2}{M^2}\sum_{k=2}^{\infty}\frac{\lambda^k M^k}{k!} = 1 + \frac{\sigma^2}{M^2}\left(e^{\lambda M} - 1 - \lambda M\right) \\
&\leq \exp\left(\frac{\sigma^2}{M^2}\left(e^{\lambda M} - 1 - \lambda M\right)\right)
\end{aligned}
$$

where the last inequality follows because $1 + x \leq e^x$.

Combining the results,

$$
\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{n} Z_i \geq t\right) &\leq e^{-\lambda t}\exp\left(\frac{n\sigma^2}{M^2}\left(e^{\lambda M} - 1 - \lambda M\right)\right) \\
&= \exp\left(-\lambda t + \frac{n\sigma^2}{M^2}\left(e^{\lambda M} - 1 - \lambda M\right)\right)
\end{aligned}
$$

Now, minimize the above bound with respect to λ. Taking derivative w.r.t. λ and setting it to zero:

$$-t + \frac{n\sigma^2}{M^2}\left(Me^{\lambda M} - M\right) = 0$$

$$e^{\lambda M} = \frac{tM}{n\sigma^2} + 1$$

$$\lambda = \frac{1}{M}\log\left(1 + \frac{tM}{n\sigma^2}\right).$$

2

Lecture 05
One dimensional concentration inequalities.
Bennett's inequality.
18.465

The bound becomes

$$
\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{n} Z_i \geq t\right) \;\leq\;& \exp\left(-\frac{t}{M}\log\left(1+\frac{tM}{n\sigma^2}\right)+\frac{n\sigma^2}{M^2}\left(\frac{tM}{n\sigma^2}+1-\log\left(1+\frac{tM}{n\sigma^2}\right)\right)\right) \\
=\;& \exp\left(\frac{n\sigma^2}{M^2}\left(\frac{tM}{n\sigma^2}-\log\left(1+\frac{tM}{n\sigma^2}\right)-\frac{tM}{n\sigma^2}\log\left(1+\frac{tM}{n\sigma^2}\right)\right)\right) \\
=\;& \exp\left(\frac{n\sigma^2}{M^2}\left(\frac{tM}{n\sigma^2}-\left(1+\frac{tM}{n\sigma^2}\right)\log\left(1+\frac{tM}{n\sigma^2}\right)\right)\right) \\
=\;& \exp\left(-\frac{n\sigma^2}{M^2}\phi\left(\frac{tM}{n\sigma^2}\right)\right)
\end{aligned}
$$

$\square$