Awesome Concentration Inequalities

Jingfeng Wu

Created: May 2020
Last updated: July 24, 2020

1 Concentration of one-dimensional random variables

Theorem 1 (Markov's inequality)**.** *Assume that* $X \geq 0$ *almost surely. Then* $\mathbb{P}(X \geq \epsilon) \leq \frac{\mathbb{E}[X]}{\epsilon}$.

Theorem 2 (Chebychev's inequality)**.** *Assume* X *has finite expectation and non-zero variance. Then*

1. $\mathbb{P}(|X - \mathbb{E}[X]| \geq \epsilon) \leq \frac{\mathrm{Var}[X]}{\epsilon^2}$.
2. *With probability at least* $1-\delta$,

$$|X - \mathbb{E}[X]| \leq \sqrt{\frac{\mathrm{Var}[X]}{\delta}}.$$

Remark 1. Let $\bar{X} = \frac{1}{m}\sum_{i=1}^m X_i$, then $\sqrt{\mathrm{Var}[\bar{X}]} \sim \mathcal{O}\left(\frac{1}{\sqrt{m}}\right)$. Then Chebychev's inequality guarantees $\bar{X}$ converges in rate $\mathcal{O}\left(\frac{1}{\sqrt{m}}\right)$. However, this is not a "high probability" result, since δ is not in a logarithm term — in this case one cannot do a union bound on exponential many such random variables.

Theorem 3 (Chernoff bound)**.** $\mathbb{P}(X \geq \epsilon) \leq \frac{\mathbb{E}[e^{tX}]}{e^{t\epsilon}}$.

Theorem 4 (Hoeffding's inequality)**.** *Let* $Z_1, \dots, Z_m$ *be independent random variables. Let* $S_m = \sum_{i=1}^m Z_i$. *Assume that* $a_i \leq Z_i \leq b_i$ *holds almost surely for* $i \geq 1$. *Then*

1. $\mathbb{P}(S_m - \mathbb{E}[S_m] \geq \epsilon) \leq \exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^m (b_i - a_i)^2}\right)$.
2. *With probability at least* $1-\delta$,

$$S_m - \mathbb{E}[S_m] \leq \sqrt{\frac{\sum_{i=1}^m (b_i - a_i)^2}{2} \log\frac{1}{\delta}}.$$

Remark 2. According to Hoeffding's inequality, $\frac{S_m}{m}$ converges in rate $\widetilde{\mathcal{O}}\left(\frac{1}{\sqrt{m}}\right)$ with high probability.

Theorem 5 (Bernstein's inequality)**.** *Let* $Z_1, \dots, Z_m$ *be independent random variables. Let* $S_m = \sum_{i=1}^m Z_i$. *Assume that* $|Z_i| \leq M$ *holds almost surely for* $i \geq 1$. *Then*

1. $\mathbb{P}(S_m - \mathbb{E}[S_m] \geq \epsilon) \leq \exp\left(\frac{-\epsilon^2/2}{\sum_{i=1}^m \mathrm{Var}[Z_i] + M\epsilon/3}\right)$.
2. *With probability at least* $1-\delta$,

$$\begin{aligned} S_m - \mathbb{E}[S_m] &\leq \frac{M}{3}\log\frac{1}{\delta} + \sqrt{\frac{M^2}{9}\log^2\frac{1}{\delta} + 2\sum_{i=1}^m \mathrm{Var}[Z_i]\log\frac{1}{\delta}} \\ &\leq \frac{2M}{3}\log\frac{1}{\delta} + \sqrt{2\sum_{i=1}^m \mathrm{Var}[Z_i]\log\frac{1}{\delta}}. \end{aligned}$$

Remark 3. According to Bernstein's inequality, $\frac{S_m}{m}$ converges in rate $\widetilde{\mathcal{O}}\left(\frac{C_1}{m} + \frac{\mathrm{Var}[Z_i]}{\sqrt{m}}\right)$ with high probability. Bernstein's inequality is very useful for eliminating some square root dependence in the convergence rate, if one can properly bound the variance term.

Theorem 6 (Empirical Bernstein's inequality). *Let $Z_1, \dots, Z_m$ be independent random variables. Let $S_m = \sum_{i=1}^m Z_i$. Assume that $|Z_i| \le M$ holds almost surely for $i \ge 1$. Then With probability at least $1-\delta$,*

$$S_m - \mathbb{E}\left[S_m\right] \le \frac{7M}{3} \log \frac{1}{\delta} + \sqrt{2 \sum_{i=1}^m \widehat{\mathrm{Var}}\left[Z_i\right] \log \frac{1}{\delta}},$$

where $\widehat{\mathrm{Var}}\left[Z_i\right] := \frac{1}{m(m-1)} \sum_{i<j} (Z_i - Z_j)^2$ is the empirical variance.

Remark 4. A proof comes from [1]. For practical applications, we usually do not have access to the population variance, and empirical Bernstein's inequality enables us to analyze the concentration phenomena in these cases.

Theorem 7 (Azuma's inequality). *Let $\{X_0, X_1, \dots\}$ be a martingale with respect to filtration $\{\mathcal{F}_0, \mathcal{F}_1, \dots\}$. Assume that $A_i \le X_i - X_{i-1} \le B_i$ holds almost surely for $i \ge 1$. Then*

1. $\mathbb{P}\left(X_m - X_0 \ge \epsilon\right) \le \exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^m (B_i - A_i)^2}\right)$.

2. *With probability at least $1-\delta$,*

$$X_m - X_0 \le \sqrt{\frac{\sum_{i=1}^m \left(B_i - A_i\right)^2}{2} \log \frac{1}{\delta}}.$$

Remark 5. Azuma's inequality improves Hoeffding's inequality by replacing the independence assumption with a more general condition, *martingale*. A typical application of Azuma's inequality is in the analysis of SGD.

2 Concentration of distributions

Theorem 8 (Pinsker's inequality). *Let P and Q be two probability distributions over a measurable space (X, Σ). Then*

1. $\|P - Q\|_\infty \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \,\|\, Q)}$.

2. $\|P - Q\|_1 \le \sqrt{2 D_{\mathrm{KL}}(P \,\|\, Q)}$.

Theorem 9 (ℓ_1-deviation of the empirical distribution). *Let P be a probability distribution over a finite discrete measurable space (X, Σ). Let $\widehat{P}_m$ be the empirical distribution of P estimated from m observations. Then with probability at least $1-\delta$,*

$$\left\|\widehat{P}_m - P\right\|_1 \le \sqrt{\frac{2\,|X|}{m} \log \frac{1}{\delta}}.$$

Remark 6. A proof comes from [2].

References

[1] Andreas Maurer and Massimiliano Pontil. Empirical bernstein bounds and sample variance penalization. *arXiv preprint arXiv:0907.3740*, 2009.

[2] Tsachy Weissman, Erik Ordentlich, Gadiel Seroussi, Sergio Verdu, and Marcelo J Weinberger. Inequalities for the l1 deviation of the empirical distribution. *Hewlett-Packard Labs, Tech. Rep*, 2003.