Common Concentration Inequalities

Beier Zhu

Markov's inequality: basis for the rest inequalities

Theorem 1. ***(Markov's inequality).*** *If X is a non-negative r.v. and $\mu = \mathbb{E}[X]$, then $\forall t > 0$:*

$$\mathbb{P}[X \geq t] \leq \frac{\mu}{t} \tag{1}$$

Polynomial tail bounds $\mathcal{O}(t^{-k})$

Derivation insight: Markov's inequality + bounded k-th central moment

Theorem 2. ***(Polynomial variant of markov's inequality).*** *If X is a r.v. with mean μ and finite k-th central moment $\mathbb{E}[|X-\mu|^k]$, then $\forall t > 0$,*

$$\mathbb{P}[|X-\mu| \geq t] \leq \frac{\mathbb{E}[|X-\mu|^k]}{t^k} \tag{2}$$

Specifically, when $k = 2$ and denote the variance as σ^2, we have:

Theorem 3. ***(Chebyshev's inequality).*** *If X is a r.v. with mean μ and variance σ^2, then $\forall t > 0$*

$$\mathbb{P}[|X-\mu| \geq t] \leq \frac{\sigma^2}{t^2} \tag{3}$$

Theorem 4. ***(Chebyshev's inequality for the sample mean).*** *Let $X_1, ..., X_n$ be i.i.d with mean μ and variance σ^2. Define $\nu = \frac{1}{n}\sum_{i=1}^n X_i$. Then, $\forall t > 0$*

$$\mathbb{P}[|\nu-\mu| \geq t] \leq \frac{\sigma^2}{nt^2} \tag{4}$$

Theorem 5. ***(Cantelli's inequality, a.k.a. one-sided chebyshev's inequality).*** *If X is a r.v. with mean μ and variance σ^2, then $\forall t > 0$*

$$\mathbb{P}[X-\mu \geq t] \leq \frac{\sigma^2}{\sigma^2+t^2}, \quad \mathbb{P}[X-\mu \leq -t] \leq \frac{\sigma^2}{\sigma^2+t^2} \tag{5}$$

Exponential tail bounds $\mathcal{O}(\exp\{-t^2\})$

Derivation insight: Markov inequality + bounded moment generating function

Theorem 6. ***(Chernoff's inequality).*** *For a r.v. X with finite moment generating function $M_X(\lambda)$, we have*

$$\mathbb{P}[X-\mu \geq t] \leq \inf_{\lambda \geq 0} M_X(\lambda)\exp\{-\lambda(t+\mu)\} \tag{6}$$

Theorem 7. ***(Tail bound for sub-Gaussian r.v.)*** *If a r.v. X with finite mean μ is σ-sub-Gaussian, then $\forall t > 0$*

$$\mathbb{P}[|X-\mu| \geq t] \leq 2\exp\{-\frac{t^2}{2\sigma^2}\} \tag{7}$$

Theorem 8. ***(Hoeffding's inequality).*** *Let $X_1, ..., X_n$ be independent real-valued r.v.s drawn from some distribution, such that $a_i \leq X_i \leq b_i$ almost surely. Define $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, and let*

$\mu = \mathbb{E}[\bar{X}]$. *Then* $\forall t > 0$,

$$\mathbb{P}[|\bar{X} - \mu| \geq t] \leq 2\exp\{-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\} \tag{8}$$

Corollary 1. ***(Hoeffding's inequality for the sample mean).*** *Let* $X_1, ..., X_n$ *be i.i.d r.v.s with* $a \leq X_i \leq b$. *Then* $\forall t > 0$,

$$\mathbb{P}[|\bar{X} - \mu| \geq t] \leq 2\exp\{-\frac{2nt^2}{(b-a)^2}\} \tag{9}$$

Specifically, when $n = 1$*: If* X *is a r.v. with* $a \leq X \leq b$. *Then* $\forall t > 0$,

$$\mathbb{P}[|X - \mu| \geq t] \leq 2\exp\{-\frac{2t^2}{(b-a)^2}\} \tag{10}$$

Motivation: Sub-Gaussian tails scale with the variance, while Hoeffding's bound depends only on the range of the variables. For bounded variables with **small variance**, this suggests we can obtain sharper concentration bounds than Hoeffding's bound.

Theorem 9. ***(Bernstein's inequality).*** *Let* $X_1, ..., X_n$ *be i.i.d r.v.s with* $a \leq X_i \leq b$ *and* $\mathbb{V}[X] = \sigma^2$. *Then* $\forall t > 0$

$$\mathbb{P}[|\bar{X} - \mu| \geq t] \leq 2\exp\{-\frac{nt^2}{2(\sigma^2 + (b-a)t)}\} \tag{11}$$

Theorem 10. *(*χ^2 ***tail bound).*** *Suppose that* $X_1, ..., X_n \sim \mathcal{N}(0, 1)$, *then* $\forall t \in (0, 1)$*:*

$$\mathbb{P}[|\frac{1}{n}\sum_{i=1}^n X_i^2 - 1| \geq t] \leq 2\exp\{-\frac{nt^2}{8}\} \tag{12}$$

Concentrations of functions of r.v.s

Motivation: So far we have focused on the contraction of averages. A natural question is whether other functions of i.i.d r.v.s also show exponential concentration.

Theorem 11. ***(McDiarmid's inequality).*** *Suppose* $f : \mathbb{R}^n \to \mathbb{R}$ *satisfies the* ***bounded difference condition:*** *there exist constants* $c_1, ..., c_n \in \mathbb{R}$ *such that for all real numbers* $x_1, ..., x_n$ *and* x_i',

$$|f(x_1, ..., x_n) - f(x_1, ..., x_{i-1}, x_i', x_{i+1}, ..., x_n)| \leq c_i \tag{13}$$

Intuitively, Eq. (13) *states that* f *is not overly sensitive to arbitrary changes in a single coordinates. Then for any independent random variables* $X_1, ..., X_n$,

$$\mathbb{P}\{f(X_1, ..., X_n) - \mathbb{E}[f(X_1, ..., X_n)] \geq t\} \leq \exp\left[-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right]. \tag{14}$$

Moreover, $f(X_1, ..., X_n)$ *is* $\mathcal{O}(\sqrt{\sum_{i=1}^n c_i^2})$*-sub-Gaussian.*

Motivation: The bounded difference in McDiarmid's inequality is often satisfied by bounded r.v.s or a bounded function. To get similar concentration inequalities for unbounded r.v.s like Gaussians, we need some other special conditions.

Theorem 12. *Suppose* $f : \mathbb{R}^n \to \mathbb{R}$ *is L-Lipschitz w.r.t* ℓ_2*-norm, and* $X_1, ..., X_n$ *drawn i.i.d. from* $\mathcal{N}(0, 1)$. *Then,* $\forall t \in \mathbb{R}$,

$$\mathbb{P}[|f(X_1, ..., X_n) - \mathbb{E}[f(X_1, ..., X_n)]|] \leq 2\exp\left(-\frac{t^2}{2L^2}\right) \tag{15}$$

1 Proofs

1.1 Proof of Theorem 1: Markov's inequality

$$t\mathbb{P}[X \geq t] = t\int_t^\infty \mathbb{P}(x)\mathrm{d}x = \int_t^\infty t\mathbb{P}(x)\mathrm{d}x \tag{16}$$
$$\leq \int_t^\infty x\mathbb{P}(x)\mathrm{d}x \leq \int_0^\infty x\mathbb{P}(x)\mathrm{d}x \tag{17}$$
$$= \mathbb{E}[X] \tag{18}$$

1.2 Proof of Theorem 2: Polynomial variant of Markov's inequality

$$\mathbb{P}[|X-\mu| \geq t] = \mathbb{P}[|X-\mu|^k \geq t^k] \tag{19}$$
$$\leq \frac{\mathbb{E}[|X-\mu|^k]}{t^k} \qquad \text{[Markov's Inequality]} \tag{20}$$

1.3 Proof of Theorem 4: Chebyshev's inequality for the sample mean

$$\mathbb{V}[\nu] = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}[X_i] = \frac{\sigma^2}{n}. \tag{21}$$

According to Theorem 3:

$$\mathbb{P}[|\nu-\mu| \geq t] \leq \frac{\mathbb{V}[\nu]}{t^2} = \frac{\sigma^2}{nt^2} \tag{22}$$

1.4 Proof of Therem 5: Cantelli's inequality

Let $Y = X - \mu$, then $\mathbb{E}[Y] = 0$ and $\mathbb{V}[Y] = \sigma^2$. For any λ s.t. $t + \lambda > 0$, we have:

$$\mathbb{P}[Y \geq t] = \mathbb{P}[Y + \lambda \geq t + \lambda] \tag{23}$$
$$= \mathbb{P}[\frac{Y+\lambda}{t+\lambda} \geq 1] \qquad [t+\lambda > 0] \tag{24}$$
$$\leq \mathbb{P}[(\frac{Y+\lambda}{t+\lambda})^2 \geq 1] \tag{25}$$
$$\leq \mathbb{E}[(\frac{Y+\lambda}{t+\lambda})^2] \qquad \text{[Markov's inequality]} \tag{26}$$
$$= \frac{\sigma^2 + t^2}{(\lambda+t)^2} \tag{27}$$

We pick λ to minimize the R.H.S, which is $\lambda = \frac{\sigma^2}{t} > 0$. That proves the theorem.

1.5 Proof of Theorem 6: Chernoff's inequality

Define $\mu = \mathbb{E}[X]$. For any $\lambda > 0$, we have

$$\mathbb{P}[X - \mu \geq t] = \mathbb{P}[X \geq \mu + t] \tag{28}$$
$$= \mathbb{P}[\exp\{\lambda X)\} \geq \exp\{\lambda(t+\mu)\}] \tag{29}$$
$$\leq \exp\{-\lambda(t+\lambda)\}\mathbb{E}\{\exp\{\lambda X)\} \qquad \text{[Markov's inequality]} \tag{30}$$
$$= \exp\{-\lambda(t+\lambda)\}M_X(\lambda) \tag{31}$$

Now λ is a parameter we can choose to get a tight upper bound.

1.6 Proofs for Theorem 7 and 8: tail bound for sub-Gaussian r.v. and Hoeffding's inequalities

See last document.