INQUALITIES

MINSEOK SONG

Markov Inequality. Start with $P(Z \geqslant a) \leqslant \frac{E(Z)}{a}$
It follows that $P(Z > 1-a) \geqslant \frac{\mu-(1-a)}{a}$. The idea is that if you know that Z is less than or equal to 1, then you use Markov inequality to $1-Z$; then we get the opposite inequality direction from Markov inequality.

Hoeffding's inequality. Assume that Z_i's are i.i.d. samples and $P[a \leqslant Z_i \leqslant b] = 1$ for every i. Further let us say $E(Z_i) = \mu$. Then we have $P(\bar{Z} - \mu) \leqslant 2\exp(\frac{-2m\epsilon^2}{(b-a)^2})$.

Remark 1. Hoeffding's inequality provides a decay rate of deviation (it is exponentially fast).

Proof. This follows directly from (after taking exponential on both sides and using Markov inequality) Heoffding's Lemma, which states that

$$\mathbb{E}[e^{\lambda X}] \leqslant e^{\frac{\lambda^2(b-a)^2}{8}}$$

□

- Proof of Hoeffding's Lemma: use the convexity of exponential, then use the fact that $b > a$ and $\mathbb{E}(X) = 0$ in the assumption.

Central limit theorem. Central limit theorem states that $\sqrt{n}(\bar{X_n} - \mu)$ converges in distribution to $\mathcal{N}(0, \sigma^2)$

Remark 2. CLT gives the rate of convergence of law of large number, which is $\frac{1}{\sqrt{n}}$.

- We can relax this to mutually independent random variables.
- Some variations:
  (1) Lyapunov CLT: this applies for when we do not necessarily have the identical distributions. Some conditions have to be met.
  (2) Martingale CLT
  (3) Levy's convergence theorem

Chernoff Bound. Let $X = \sum_{i=1}^n X_i$, where X_i's all being Bernoulli(p_i) iid. Let $\mu = \mathbb{E}(X) = \sum_{i=1}^n p_i$. Then we have

$$\mathbb{P}(X \geqslant (1+\delta)\mu) \leqslant e^{\frac{-\delta^2}{2+\delta}\mu} \text{ for all } \delta > 0 \tag{1}$$

$$\mathbb{P}(X \leqslant (1-\delta)\mu) \leqslant e^{-\mu\delta^2/2} \text{ for all } 0 < \delta < 1 \tag{2}$$

Generic Chernoff Bound.

$$\mathbb{P}(X \geqslant a) \leqslant \inf_{t>0} M(t)e^{-ta} \qquad (t > 0) \tag{3}$$

$$\mathbb{P}(X \leqslant a) \leqslant \inf_{t<0} M(t)e^{-ta} \qquad (t < 0) \tag{4}$$

where $M(t) = \mathbb{E}(e^{-ta})$

Remark 3. • Markov inequality is the best we hope for, given no other information. The idea being that we use Markov inequality, along with the moment generating function, which is already known.

- This explains that the probability of "bad event" is very low.
- The real power of the Chernoff bound is when we apply to the sum of independent random variables. Chernoff bound gives an exponentially decreasing bound on the probability of X deviating from its expectation.
- Another advantage is that we can use this regardless of the sign.

Concentration of χ^2 Variables. Let $Z \sim \chi_k^2$. Then, for all $\epsilon \in (0, 3)$ we have

$$\mathbb{P}[(1-\epsilon)k \leqslant Z \leqslant (1+\epsilon)k] \geqslant 1 - 2e^{-\epsilon^2 k/6}$$

Remark 4.

- With small ϵ and reasonably small k, we have a decent chance that χ^2 is aronud its mean.