Concentration

11.1 Motivation

Suppose $\widehat{\theta}(X_1, \ldots, X_n)$ is an estimator for θ with $X_i \overset{i.i.d.}{\sim} \mathcal{D}_\theta$. How good is $\widehat{\theta}$?

Things we can demand are the following:

1. $\widehat{\theta}_n \to \theta$ (in $P, L_p, a.s., \ldots$) as $n \to \infty$.
2. $\mathbb{P}\left[\left\|\widehat{\theta} - \theta\right\| > \delta\right] < \epsilon$.
3. $\mathbb{E}\left[\left\|\widehat{\theta} - \theta\right\|\right] < \epsilon$.
4. what's the difference of the norm of $\widehat{\theta}_n - \theta$.

The next question is that how do we find such δ, ϵ? θ is a R.V. however we don't know or can't find its distribution. It incentives us to look random variables.

11.2 Starting Point

Let Z be non-negative. Then the following theorem holds.

Theorem 11.1 (Markov's inequality)**.**

$$\mathbb{P}[Z \geq \delta] \leq \frac{\mathbb{E}[Z]}{\delta}.$$

Proof. I

$$\begin{aligned} \mathbb{E}[Z] &= \textstyle\int_0^\infty zp(z)dz \\ &= \textstyle\int_0^\delta zp(z)dz + \int_\delta^\infty zp(z)dz \\ &\geq \textstyle\int_\delta^\infty zp(z)dz \geq \delta \int_\delta^\infty p(z)dz = \delta\mathbb{P}[Z \geq \delta], \end{aligned}$$

where $p(z)$ is the pdf of R.V. Z. ■

Here is another proof.

Proof. II Since $Z \geq \delta\mathbb{1}(Z \geq \delta)$ always holds, take expectation for both sides and we will have $\mathbb{E}[Z] \geq \delta \cdot \mathbb{E}[\mathbb{1}(Z \geq \delta)] = \delta\mathbb{P}[Z \geq \delta]$. ■

Actually, we can apply Markov's inequality to more general cases using one trick. Let ϕ be an non-decreasing, non-negative function on support of Z. Then it holds that

$$\mathbb{P}[Z \geq \delta] = \mathbb{P}[\phi(Z) \geq \phi(\delta)] \leq \frac{\mathbb{E}[\phi(Z)]}{\phi(Z)}.$$

Let $\phi(t) = t^2$ and $Y = |Z - \mathbb{E}[Z]\,|$. We will get the following theorem.

Theorem 11.2 (Chebyshev's inequality)**.**

$$\mathbb{P}[|Z - \mathbb{E}[Z]\,| \geq \delta] \leq \frac{\mathrm{Var}[Z]}{\delta^2}.$$

11.3 Cramér-Chernoff method

Let $\phi(t) = e^{\lambda t}$ for some $\lambda > 0$. We will have $\mathbb{P}[Z > \delta] \leq e^{-\lambda\delta}\mathbb{E}\left[e^{\lambda Z}\right]$, where we can find that $\mathbb{E}\left[e^{\lambda Z}\right]$ is the moment generating function of Z. Since this inequality holds for any $\lambda > 0$, it holds that

$$\mathbb{P}[Z > \delta] \leq \inf_{\lambda > 0} e^{-\lambda\delta}\mathbb{E}\left[e^{\lambda Z}\right].$$

Define $\psi_Z(\lambda) = \log \mathbb{E}\left[e^{\lambda Z}\right], \forall \lambda > 0$. Then $\inf_{\lambda>0} e^{-\lambda\delta}\mathbb{E}\left[e^{\lambda Z}\right] = \inf_{\lambda>0} \exp(-\lambda\delta + \psi_Z(\lambda))$. Further define

$$\psi_Z^*(\delta) = \sup_{\lambda > 0}(\lambda\delta - \psi_Z(\lambda)).$$

Theorem 11.3 (Jensen's inequality)**.** *If ϕ is convex, it holds that*

$$\phi(\mathbb{E}[Z]) \leq \mathbb{E}[\phi(Z)].$$

We have the following three comments:

1. Since $\psi_Z(0) = 0$, it holds that for all Z, ψ_Z^* is non-negative;
2. If $\mathbb{E}[Z]$ exists, by convexity and Jensen's inequality, we have $\psi_Z(\lambda) \geq \delta\mathbb{E}[Z]$;
3. If $\lambda < 0$, $\lambda\delta - \psi_Z(\lambda) \leq 0$. Thus sup occurs when $\lambda > 0$. We can extend the definition of $\psi_Z^*(\delta)$ to the following $\psi_Z^*(\delta) = \sup(\lambda\delta - \psi_Z(\lambda))$, which is the conjugate function of $\psi_Z(\lambda)$.

Theorem 11.4 (Chernoff inequality)**.**

$$\mathbb{P}[Z \geq \delta] \leq \exp(-\psi_Z^*(\delta)).$$

The theorem becomes trivial if $\psi_Z^*(\delta) = 0$. If Z is centered, then ψ_Z is continuously differentiable on $(0, b)$ for $0 < b \leq \infty$ and $\psi_Z'(0) = \psi_Z(0) = 0$. Also ψ_Z is convex (strictly) unless $Z = C$ w.p. 1.

We define $\lambda_\delta = (\psi_Z')^{-1}(\delta)$.

Example 11.5. *We have $Z \sim \mathcal{N}(0, \sigma^2)$. We can find that $\psi_Z(\lambda) = \frac{\lambda^2\delta^2}{2}$. Hence*

$$\psi_Z^*(\delta) = \sup_{\lambda} \lambda\delta - \frac{\lambda^2\delta^2}{2} = \frac{\delta^2}{2\sigma^2}.$$

Therefore, we have

$$\mathbb{P}[|Z| > \delta] \leq 2\exp\left(-\frac{\delta^2}{2\sigma^2}\right).$$

There is also a similar inequality regarding the normal random variable (i.e. $Z \sim \mathcal{N}(0,1)$).

Theorem 11.6 (Mill's inequality)**.**

$$\mathbb{P}[|Z| \geq \delta] \leq \frac{2}{\sqrt{2\pi}} \exp(-\delta^2/2)/\delta.$$

Example 11.7. *We have $Y \sim Poisson(v)$. Define $Z = Y - v$. Hence, we have*

$$\mathbb{E}\left[e^{\lambda Z}\right] = e^{-\lambda v - v} e^{v e^{\lambda}}.$$

Therefore, $\psi_Z(\lambda) = v(e^{\lambda} - \lambda - 1)$ and $\psi_Z^*(\lambda) = vh(\delta/v)$ where $h(x) = (1+x)\log(1+x) - x$ for $x \geq -1$. So

$$\mathbb{P}[Z > \delta] \leq \exp(-vh(\delta/v))].$$

Chernoff bound is nice because we can apply to sums or means. Suppose $Z = X_1 + \ldots X_n$ where $X_i = X \overset{i.i.d.}{\sim} \mathcal{D}$. Also, we have $\psi_{X_i}(\lambda) = \log \mathbb{E}\left[e^{\lambda X_i}\right]$. Then $\psi_Z(\lambda) = \log \mathbb{E}\left[e^{\lambda \sum_i X_i}\right] = \log \Pi_i \mathbb{E}\left[e^{\lambda X_i}\right] = n\psi_X(\lambda)$. So

$$\psi_Z^*(\delta) = n\psi_X^*(\delta/n).$$

Example 11.8. *We have $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$. It holds that*

$$\mathbb{P}[|Z| > \delta] \leq 2\exp\left(-\frac{\delta^2}{2n\sigma^2}\right).$$

Let $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. Then $\psi_{\overline{X}} = \log \mathbb{E}\left[e^{\frac{1}{n}\sum_i X_i}\right] = n\psi_X(\lambda/n)$. So $\psi_{\overline{X}}^*(\delta) = n\psi_X^*(\delta)$. Then we will have the following "tight" inequality.

$$\mathbb{P}[|\overline{X}| > \delta] \leq exp\left(-\frac{\delta^2}{2\sigma^2}\right).$$