DAILY COMPUTER ΔC

Tags: Probability *Written By: Nimitt*

Concentration Inequalities

We bound the probability of a random variable being far from the expectation.

Theorem 1 (Markov Inequality)**.** *Let $X \geq 0$ be a random variable. Then, for any $a > 0$,*

$$\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$$

Proof. Case (I) Discrete.
Let $X \geq 0$ be discrete random variable taking values in $\{x_i\}$ and $p(x)$ be probability mass function of X.

$$\begin{aligned}\mathbb{E}[X] = \sum_{x_i} x_i p(x_i) &= \sum_{x_i<a} x_i p(x_i) + \sum_{x_i \geq a} x_i p(x_i) \\ &\leq \sum_{x_i \geq a} a p(x_i) = a\Pr[X \geq a]\end{aligned}$$

By rearranging, we get the inequality.

Case (II) Continuous.
Let $X \geq 0$ be continuous random variable and $f(x)$ be probability density function of X.

$$\begin{aligned}\mathbb{E}[X] = \int_0^\infty x f(x) dx &= \int_0^a x f(x) dx + \int_a^\infty x f(x) dx \\ &\leq \int_a^\infty a f(x) dx = a\Pr[X \geq a]\end{aligned}$$

By rearranging, we get the inequality. ☐

Note that Markov is a first order inequality. We can use Markov inequality to get similar higher order inequalities (higher order inequalities give tighter bounds on probability). Particularly, we apply Markov inequality to $\phi(|X - \mathbb{E}[X]|)$ where $\phi(t)$ is an increasing function in t.

Theorem 2 (Chebyshev Inequality)**.** *Let X be a random variable. Then, for any $a > 0$,*

$$\Pr[|X - \mathbb{E}[X]| \geq a] \leq \frac{\mathrm{Var}[X]}{a^2}$$

Proof. Let X be a random variable with $\mathbb{E}[X] = \mu$. Let $\phi(t) = t^2$. Using Markov inequality,

$$\begin{aligned}\Pr[|X-\mu| \geq a] = \Pr[\phi(|X-\mu|) \geq \phi(a)] &\leq \mathbb{E}[\phi(|X-\mu|)]/\phi(a) \\ &\leq \mathbb{E}[|X-\mu|^2]/a^2 = \mathrm{Var}[X]/a^2\end{aligned}$$

☐

As mentioned earlier, with different ϕ we can get different higer order inequalities. Using $\phi(t) = e^{\lambda t}$ gives us chernoff inequality. Since, $e^x = 1 + x/1! + x^2/2! + x^3/3! \cdots$, chernoff captures inequalities for every order. This is the same reason why Gaussian kernel is used in SVM.

Theorem 3 (Chernoff Inequality)**.** *Let X be a random variable. Then, for $t > 0$,*

$$\Pr[X \geq t] \leq e^{-\psi_X^*(t)}$$

where ψ_X^* is the cramer transform of X (See Proof).

Proof. Let X be a random variable and $a > 0$. Using Markov inequality and $\phi(t) = e^{\lambda t}$,

$$\begin{aligned}\Pr[X \geq a] &= \Pr[\phi(X) \geq \phi(a)] \leq \mathbb{E}[\phi(X)]/\phi(a) \\ &\leq \mathbb{E}[e^{\lambda X}]/e^{\lambda a} = \exp\left(-(\lambda a - \ln\left(\mathbb{E}[e^{\lambda X}]\right)\right)\end{aligned}$$

Let $\psi_X(\lambda) = \ln(\mathbb{E}[e^{\lambda X}])$ (ψ_X is referred as cumulant function of X and $\mathbb{E}[e^{\lambda X}]$ is the Moment Generating Function (MGF)). Because, above holds for any λ, we find the λ that gives tightest bound.

$$\begin{aligned}\Pr[X \geq a] &\leq \exp\left(-(\lambda a - \psi_X(\lambda)\right) \\ &\leq \exp\left(-\sup_{\lambda}(\lambda a - \psi_X(\lambda))\right)\end{aligned}$$

Define $\psi_X^*(a) = \sup_\lambda(\lambda a - \ln\left(\mathbb{E}[e^{\lambda X}]\right))$ as the Cramer transform of X. Then,

$$\Pr[X \geq a] \leq e^{-\psi_X^*(a)}$$

□

Example. Chernoff bound for gaussian distribution. Let $X \sim \mathcal{G}(\mu, \sigma^2)$. Using Gaussian intergral, MGF, $\mathbb{E}[e^{\lambda X}] = \exp(\lambda^2\sigma^2/2 + \mu\lambda)$. Thus, $\psi_X(\lambda) = \mu\lambda + \sigma^2\lambda^2/2$. Minimizing $\lambda a - \psi_X(\lambda)$ over λ, we get $\psi_X^*(a) = (\mu - a)^2/2\sigma^2$. Thus,

$$\Pr[X \geq a] \leq e^{-(\mu-a)^2/2\sigma^2}$$