Two Proofs for Chernoff-Hoeffding Theorem

Xiao Yunxuan

March 28^{th}, 2018

1 Chernoff-Hoeffding Theorem

Suppose $X_1, ..., X_n$ are random variables, taking values in $\{0, 1\}$. Let $p = E[X_i]$ and $t > 0$. Then

$$P\left(\left|\frac{S_n}{n} - p\right| \geq t\right) \leq e^{-nh_+(t)} + e^{-nh_-(t)}$$

where:

$$h_+(t) = D_{KL}(p + t \| p)$$
$$h_-(t) = D_{KL}(p - t \| p)$$

D_{KL} is the Kullback–Leibler divergence between Bernoulli distributed random variables with parameters x and y respectively.

$$D_{KL}(x\|y) = x \ln \frac{x}{y} + (1 - x) \ln \left(\frac{1 - x}{1 - y}\right)$$

Proof1: Suppose S_n is a random variable that denotes the total number of 1s in Bernoulli distributed random variables$\{X_1, X_2, \cdots, X_n\}$.For any $\lambda > 0$ and through Chebyshev's Inequality, we have

$$\begin{aligned} P(S_n \geq (p+t)n) =& P(\lambda S_n \geq \lambda(p+t)n) = P(e^{\lambda S_n} \geq e^{\lambda(p+t)n}) \\ \leq& \frac{E[e^n]}{e^{\lambda(p+t)n}} = \frac{pe^{\lambda} + 1 - p}{e^{\lambda(p+t)}}^{n} = g(\lambda) \end{aligned}$$

The lower bound of the inequality can be determined by calculating the zero point $\lambda^{'}$ of the derivative of $g(\lambda)$

$$\frac{d(g(\lambda))}{d(\lambda)} = n\left[\frac{pe^{\lambda} + 1 - p}{e^{\lambda(p+t)}}\right]^{n-1} \frac{p(1-p-t)e^{\lambda} - (p+t)(1-p)}{e^{2\lambda(p+t)}} = 0$$

$$p(1-p-t)e^{\lambda'} - (p+t)(1-p) = 0$$
$$e^{\lambda'} = \left(\frac{1-p}{p}\right)\left(\frac{p+t}{1-p-t}\right)$$

Then we have the lower bound $g(\lambda^{'})$

$$
\begin{aligned}
g(\lambda^{'}) &= \left(\frac{pe^{\lambda^{'}}+1-p}{e^{\lambda^{'}(p+t)}}\right)^n \\
&= \left[\frac{p\left(\frac{1-p}{p}\right)\left(\frac{p+t}{1-p-t}\right)+1-p}{\left[\left(\frac{1-p}{p}\right)\left(\frac{p+t}{1-p-t}\right)\right]^{(p+t)}}\right]^n \\
&= \left[\left(\frac{p}{p+\varepsilon}\right)^{p+\varepsilon}\left(\frac{1-p}{1-p-\varepsilon}\right)^{1-p-\varepsilon}\right]^n \quad = e^{-nh_+(t)}
\end{aligned}
$$

hence we get

$$P\left(\frac{S_n}{n}-p\geq t\right)\leq e^{-nh_+(t)}$$

Similarly, by setting t to $-t$ and applying the same derivation process, it can also be proved that

$$P\left(\frac{S_n}{n}-p\leq -t\right)\leq e^{-nh_-(t)}$$

Therefore,

$$
\begin{aligned}
P\left(\left|\frac{S_n}{n}-p\right|\geq t\right) =& P\left(\frac{S_n}{n}-p\geq t\right)+P\left(\frac{S_n}{n}-p\leq -t\right) \\
\leq& e^{-nh_+(t)}+e^{-nh_-(t)}
\end{aligned}
$$

Proof 2(Encoding Arguments): Let D be a probability distribution on $\{0,1\}^n$ that assign to each element x a probabilityP_x, let ω be a non-negative weight function such that $\sum_{x\in\{0,1\}^n} \omega(x) \leq 1$

First, we prove a Lemma:

$$For\ any\ s \leq 1, P_{x\sim D}[\omega(x) \geq sP_x] \leq \frac{1}{s} \tag{$*$}$$

Let $Z_s = x|\omega(x) \geq sP_x$, in terms of Markov Inequality, we have

$$\begin{aligned} P_{x\sim D}[\omega(x) \geq sP_x] \leq & \frac{E[\omega(x)]}{sP_x} \\ \leq & \frac{\sum_{x\in Z_s} P_x\omega(x)}{sP_x} = \frac{1}{s}\sum_{x\in Z_s}\omega(x) \\ \leq & \frac{1}{s} \end{aligned}$$

Now consider $X = X_1X_2\cdots X_n$ is an encoded 0/1 string. Suppose that p denotes the probability of $X_i = 1$ ($p \geq 1/2$), k_x is the total number of 1s in string X. So

$$P_x = p^{k_x}(1-p)^{n-k_x}$$

Then we construct a weight function $\omega_{k_x}(x)$

$$\omega_{k_x}(x) = (p+t)^{k_x}(1-p-t)^{n-k_x}$$

Notice that $\omega(x)$ is a monotonically increasing function:

$$\frac{d\omega_{k_x}(x)}{dk_x} = \left(ln\frac{p+t}{1-p-t}\right)\omega(x) > 0$$

When $k_x = n(p+t)$:

$$\begin{aligned} P_x \cdot e^{nh_+(t)} &= p^{n(p+t)}(1-p)^{n(1-p-t)}\left[\left(\frac{p+t}{p}\right)^{(p+t)}\left(\frac{1-p-t}{1-p}\right)^{(1-p-t)}\right]^n \\ &= (p+t)^{(p+t)n}(1-p-t)^{(1-p-t)n} \\ &= \omega_{(p+t)n}(x) \end{aligned} \tag{$**$}$$

Thus we obtain the probability:

$$\begin{aligned}P(\frac{S_n}{n} - p \geq t) =& P(S_n \geq (p+t)n) = P_{x\sim D}[k_x \geq (p+t)n] \\ =& P_{x\sim D}[\omega_{k_x}(x) \geq \omega_{(p+t)n}(x)]\end{aligned}$$

And from lemma (*) and equation (**) we obtain:

$$P(\frac{S_n}{n} - p \geq t) = P_{x\sim D}[\omega(x) \geq P_x \cdot e^{nh_+(t)}] \leq e^{-nh_+(t)}$$

Similarly

$$P(\frac{S_n}{n} - p \leq -t) = P_{x\sim D}[\omega(x) \leq P_x \cdot e^{nh_-(t)}] \leq e^{-nh_-(t)}$$

We now have our desired result, that

$$P\left(\left|\frac{S_n}{n} - p\right| \geq t\right) \leq e^{-nh_+(t)} + e^{-nh_-(t)}$$