Sub-Gaussian random variable and its properties

Shengming Luo

6 Sep 2017

0.1 Recap

Last time we've defined $X \in SG(\sigma^2)$ if $\mathbb{E}[e^{\lambda(X-\mathbb{E}(X))}] \leq e^{\lambda^2\sigma^2/2}$, where the RHS is mgf of $N(0,\sigma^2)$.

We're also given the tail bound as $\mathbb{P}(|X-\mu| \geq t) \leq 2\exp\{-\frac{t^2}{2\sigma^2}\}$. Some other properties of $X \in SG(\sigma^2)$.

- $V[X] \leq \sigma^2$
- $a \leq X - \mu \leq b$, then $X \in SG((\frac{b-a}{2})^2)$
- $X \in SG(\sigma^2) \Longrightarrow aX \in SG(a^2\sigma^2)$
- $X \in SG(\sigma^2)$ and $Y \in SG(\tau^2)$, then $X+Y \in SG((\sigma+\tau)^2)$. Additionally, if X and Y are independent, $X+Y \in SG(\sigma^2+\tau^2)$, which is a sharper bound.

Question: let $\sigma(X) = \inf\{\sigma > 0 : \mathbb{E}[e^{\lambda(X-\mathbb{E}X)}]\} \leq e^{\frac{\lambda^2\sigma^2}{2}}\}$, is $V[X]=\sigma^2$? Not in general. Let

$$X = \begin{cases} 1, & \text{with prob } \frac{1-p}{2}, \\ 0, & \text{with prob } p, \\ -1, & \text{with prob } \frac{1-p}{2}, \end{cases}$$

Then, $\mathbb{E}[e^{\lambda X}] > \exp\{\frac{\lambda^2\mathbb{E}X^2}{2}\}$ for all λ small enough.

Proof of property (4): W.L.O.G., $\mathbb{E}[X]=\mathbb{E}[Y]=0$. Then, $\forall\lambda \in \mathbb{R}$,

$$\mathbb{E}[e^{\lambda(X+Y)}] \leq \left(\mathbb{E}[e^{\lambda pX}]\right)^{\frac{1}{p}}\left(\mathbb{E}[e^{\lambda qY}]\right)^{\frac{1}{q}} = \exp\left\{\frac{\lambda^2\sigma^2p^2}{2}\frac{1}{p} + \frac{\lambda^2\tau^2q^2}{2}\frac{1}{q}\right\},$$

where the inequality is Holder's inequality, $\frac{1}{p}+\frac{1}{q}=1$.

Minimize RHS over all pairs of p,q, we get, by Cauchy inequality,

$$\frac{\lambda^2\sigma^2p}{2} + \frac{\lambda^2\sigma^2q}{2} = (\frac{\lambda^2\sigma^2p}{2} + \frac{\lambda^2\tau^2q}{2})(\frac{1}{p}+\frac{1}{q}) \geq \frac{\lambda^2(\sigma+\tau)^2}{2}.$$

1 Hoeffding inequality

Let $X_1,\cdots,X_n$ be independent with $\mu_i = \mathbb{E}[X_i], X_i \in SG(\sigma_i^2)$ for all $1 \leq i \leq n$. Then,

$$\mathbb{P}\big(|\frac{\sum(X_i-\mu_i)}{n}| \geq t\big) \leq 2\exp\big\{-\frac{n^2t^2}{2\sum_{i=1}^n\sigma_i^2}\big\}.$$

If $\sigma_i^2=\sigma^2$, then RHS becomes $\leq 2\exp\{-\frac{nt^2}{2\sigma^2}\}$.

Example:

Let $X_1, \cdots, X_n$ be independent $Bernoulli(p_i)$, where $\pi_i \in (0,1)$. Then $X_i \in SG(\frac{1}{4})$.

As a result, $\mathbb{P}(|\bar{X} - \bar{p}_n| \geq t) \leq 2\exp\{-2t^2 n\}$. Let RHS=$\delta$ and solve for t, we get $|\bar{X} - \bar{p}_n| \leq \sqrt{\frac{1}{2n}\log(\frac{1}{\delta})}$ with prob at least $1-\delta$.

Let $\delta = n^{-c}$, then $|\bar{X} - \bar{p}_n| = O(\sqrt{\frac{\log(n)}{n}})$ with prob at least $1 - n^{-c}$.

1.1 Is Hoeffding the sharpest bound?

The answer is No! For Bernoulli, Chernoff bound gives sharper result.

Consider following multiplicative bounds,

$$\mathbb{P}\big(\sum X_i \geq (1+\epsilon)\sum p_i\big) \leq \begin{cases} \exp\big\{-\frac{\epsilon^2\mu}{3}\big\}, & \epsilon \in (0,1), \\ \exp\big\{-\frac{\epsilon^2\mu}{2+\mu}\big\}, & \epsilon > 1, \end{cases}$$

Hoeffding vs Multiplicative bound:

Suppose $X_1, \cdots, X_n \overset{iid}{\sim} Bern(p)$, then Hoeffding bound would give us:

$$\mathbb{P}(p - \frac{\sum X_i}{n} \geq t) \leq \exp\{-2nt^2\} \Longrightarrow p - \bar{X}_n \leq \sqrt{\frac{1}{2n}\log(1/\delta)}, \text{ with probs } \geq 1-\delta.$$

However, using multiplicative bound,

$$\mathbb{P}(p - \frac{\sum X_i}{n} \geq t) \leq \exp\{-np\epsilon^2/2\} \Longrightarrow p - \bar{X}_n \leq \sqrt{\frac{2p}{n}\log(1/\delta)}, \text{ with probs } \geq 1-\delta.$$

If $p \leq 1/4$, multiplicative bound is sharper. If $p \to 0$, multiplicative bounds is much better (e.g. See paper minimax classification with optimal decision tree.)

2 Equivalent characterizations of SubGaussian random variables

Assuming $\mathbb{X} = 0$, we've told the definition of Sub gaussian random variable is 1) $\mathbb{E}[e^{\lambda X}] \leq e^{\frac{\lambda^2\sigma^2}{2}}, \forall \lambda \in \mathbb{R}$.

Equivalent definitions: 2) there exists $c > 0$ and $Z \in N(0, \tau^2)$, s.t.

$$\mathbb{P}(|X| > t) \leq c\mathbb{P}(|Z| \geq \tau).$$

3) $\mathbb{E}[e^{\frac{\lambda X^2}{2\sigma^2}}] \leq \frac{1}{\sqrt{1-\lambda}}$, where $\lambda \in (0,1)$.

4) $\mathbb{E}[e^{aX^2}] \leq 2$ for some $a > 0$.

5) Finally, if $X \in SG(\sigma^2)$ and $\forall p > 0$,

$$\mathbb{E}[|X|^p] \leq p2^{p/2}\sigma^p\Gamma(p/2) \leq c\sigma\sqrt{p},$$

where $\Gamma(x) = \int_0^\infty e^{-\mu}\mu^{p/2-1}d\mu$.

Proof of (5):

$$\begin{aligned}\mathbb{E}[|X|^p] &= \int_0^\infty \mathbb{P}(|X|^p \geq t)dt \\ &\leq 2\int_0^\infty \exp\{-\frac{t^{2/p}}{2\sigma^2}\}dt = (2\sigma^2)^{p/2}p\Gamma(p/2).\end{aligned}$$

3 Sub-Exponential random variables

$X \sim$ Laplacian$(b), b > 0$. Pdf of X is $p(x) = \frac{b}{2}e^{-|x|b}$, i.e. the tail probability does not decay as fast as sub-gaussian random variables (thicker tails). In fact, the MGF of X equals $\frac{1}{1-\lambda^2 b^2}$, where $|\lambda| < \frac{1}{|b|}$.

Thus, we might need consider a another class of random variables with slower tail probability decay rate. That motivates us to the definition of sub exponential random variables.

Definition: Suppose $v, \alpha > 0$, then $X \in SE(v^2, \alpha)$ when $\mathbb{E}[e^{\lambda(X-\mathbb{E}[X])}] \leq e^{\frac{\lambda^2 v^2}{2}}$ if $|\lambda| < \frac{1}{\alpha}$.

If $X \in SG(\sigma^2)$, we have $X \in SE(v^2, 0)$.

Example: If $Z \sim N(0,1)$, $X = Z^2 \sim \mathcal{X}_1^2$. $\mathbb{E}[X] = 1$.

$$\begin{aligned}
\mathbb{E}[e^{\lambda(X-1)}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\lambda(z^2-1)} e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} e^{-\lambda} \int_{-\infty}^{+\infty} e^{\frac{z^2}{2}(1-2\lambda)} dz \\
&= \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \leq \exp\{\frac{\lambda^2}{1-2\lambda}\},
\end{aligned}$$

where the last inequality is due to $-\log(1-u) - u \leq \frac{u^2}{1-u}$

Hence, $\mathbb{E}[e^{\lambda(X-1)}] \leq \exp\{\frac{\lambda^2}{1-2\lambda}\} \leq \exp\{\frac{4\lambda^2}{2}\} \Longrightarrow Z^2 \sim SE(4,4) \Longrightarrow \mathbb{P}(Z^2 - 1 \geq 2t + 2\sqrt{t}) \leq e^{-t}, \forall t > 0$.

Properties:

1) If $X \in SG(\sigma^2)$, then $X^2 \in SE(256\sigma^2, 16\sigma)$.

2) If $V[X] = \sigma^2$ and $|X - \mathbb{E}[X]| \leq b, a.s.$ Then $X \in SE(2\sigma^2, 2b)$.