Concentration of mass around equator for an $n$-sphere: We have seen that most of the mass of a unit ball is at its boundary. A student asked in the previous lecture whether it is true that the mass is concentrated around an equator for an $n$-sphere

$$S^{n-1} = \{ x \in \mathbb{R}^n \mid \| x \| = 1 \}$$

and Ale answered that in the following.

Let $\sigma_{n-1}(A)$ denote the surface area of a subset $A$ of $S^{n-1}$. For concreteness, define an equator $E$, a band $E_r$ around it for $r > 0$ and a spherical cap $C_r$ as follows:

$$E = \{ x \in S^{n-1}, x_1 = 0 \},$$

$$E_r = \{ x \in S^{n-1}, |x_1| < r \} \text{ and }$$

$$C_r = \{ x \in \mathbb{S}^n, x_1 > 0 \} \cap E_r^c.$$

Consider the convex cone $D_r$ generated by $C_r$ and the center of the sphere and note that the surface area of $C_r$,

$$\sigma_{n-1}(C_r) = \frac{\text{Volume of the cone } D_r}{V_n}$$

where $V_n$ is the volume of unit ball. The ball constructed with the base of the cap $C_r$ as its equator contains $D_r$ and has a volume of $(1 - r^2)^{n/2}V_n$. So

$$\sigma_{n-1}(C_r) \leq (1 - r^2)^{n/2} \leq e^{-nr^2/2}.$$

This shows that

$$\sigma_{n-1}(E_r) \geq 1 - 2e^{-nr^2/2}.$$

It may seem puzzling that the mass is concentrated at any equator of the sphere. This may be understood in the following manner. If $\sum_{i=1}^n x_i^2 = 1$ and $x_i$ are identically distributed then we expect all of them to be small. $x_i$ being small is the same thing as saying that $(x_1, \cdots, x_n)$ lies close to an equator.

Recall from the previous lecture that we derived the Chernoff bound

$$\Pr(Z \geq z) \leq \exp(-\psi_Z^*(z)).$$

Also recall that

- For $Z \sim \mathcal{N}(0, \sigma^2)$, the log mgf function $\psi_Z(\lambda) = \lambda^2\sigma^2/2$ and its conjugate $\psi_Z^*(x) = x^2/2\sigma^2$. 

3-1
• For $X \sim \text{Poisson}(\nu)$, where $\nu > 0$ considering $Z = X - \nu$

$$\psi_Z(\lambda) = \nu(e^\lambda - 1), \psi^*_Z(x) = \nu h(x/\nu)$$

where $h(u) = (1 + u) \log(1 + u) - u$ for $u \geq -1$. And for $x \leq \nu$,

$$\psi^*_Z(x) = \nu h(-x/\nu).$$

• If $X \sim \text{Bernoulli}(p)$, then considering $Z = X - p$, for $0 < x < 1 - p$,

$$\psi^*_Z(x) = (1 - p - x) \log \frac{1 - p - x}{1 - p} + (p + x) \log \frac{p + x}{p}.$$

which may be recognized as $\text{KL} (\text{Bernoulli}(x + p), \text{Bernoulli}(p))$.

**Sums of independent random variables:** Suppose $X_i - \mathbb{E}X_i$, $i = 1, \cdots, n$ are i.i.d with log mgf $\psi_X$ and let $Z = \sum_{i=1}^{n} X_i - \mathbb{E}X_i$. Then

$$\psi_Z(\lambda) = \log \mathbb{E}[e^{\lambda Z}] = n \psi_X(\lambda)$$

which implies $\psi^*_Z(x) = n \psi^*_X(x/n)$.

**Example:** Suppose $X_1, X_2, \cdots, X_n \sim \mathcal{N}(0, \sigma^2)$ are independent. Then the log mgf of $Z = \sum_i X_i$ is

$$\psi^*_Z(x) = n(x/n)^2/2\sigma^2 = x^2/2n\sigma^2.$$

So for $t > 0$,

$$P(\bar{X}_n \geq t) = P(Z \geq nt) \leq \exp -\psi^*_Z(nt) = e^{-nt^2/2\sigma^2}.$$

**Subgaussian Random Variables**

**Definition:** A random variable $X$ is sub-Gaussian with variance factor $\nu^2$ if

$$\psi_X(\lambda) \leq \lambda^2 \nu^2/2 \quad \forall \lambda \in \mathbb{R}.$$  

We will write this as $X \in G(\nu)$. The log mgf of $X \in G(\nu)$ is upper bounded by that of a normal with mean 0 and variance $\nu^2$.

**Example:** If $X \sim \mathcal{N}(0, \nu^2)$ then $\psi_X(\lambda) = \lambda^2 \nu^2/2$ and hence $X \in G(\nu)$.

Note that if $X \in G(\nu)$, then for $x > 0$, by the exponentiation technique used in Chernoff bounds, we can write

$$P(X \geq x) \leq \inf_{\lambda > 0} \mathbb{E}e^{\lambda(X-x)} \leq \exp \left( \frac{1}{2} \lambda^2 \nu^2 - \lambda x \right) = e^{-x^2/2\nu^2}.$$  

Similarly it can be shown that $P(X \leq -x) \leq e^{-x^2/2\nu^2}$ and hence $P(|X| > x) \leq 2e^{-x^2/2\nu^2}$.

We remark that sub-Gaussian behaviour results in Gaussian concentration due to these bounds.

**Theorem 3.1** If $X \in G(\nu)$, then $\mathbb{E}X = 0$ and $\text{Var}(X) =: V(X) \leq \nu^2$.  

Proof: As $X \in G(\nu)$, we should have $0 \leq \mathbb{E}e^{\lambda X} \leq e^{\lambda^2 \nu^2}$ for all $\lambda \in \mathbb{R}$. After subtracting 1 and dividing by $\lambda > 0$, taking limits as $\lambda \to 0^+$, we get $\mathbb{E}X = 0$. To get the bound on variance, we again start from the same inequality and write

$$-\frac{1}{\lambda^2} \leq \mathbb{E} \left[ \frac{e^{\lambda X} - 1 - \lambda X}{\lambda^2} \right] \leq \frac{e^{\lambda^2 \nu^2} - 1}{\lambda^2}.$$

Taking limits as $\lambda \to 0$, we obtain $\mathbb{E}X^2 \leq \nu^2$ which gives the desired result.

Example: Suppose $X$ is a Radamacher random variable, that is it takes $-1$ or $+1$ with probability $1/2$ each. Then $X \in G(1)$ because

$$\mathbb{E}e^{\lambda X} = \frac{1}{2}(e^{-\lambda} + e^{\lambda}) = \cosh(\lambda) \leq e^{\lambda^2/2}.$$

Example: Suppose $X \sim \text{Uniform}[-a, a]$, where $a > 0$. Then $X \in G(a)$ because

$$\mathbb{E}e^{\lambda X} = \int_{-a}^{a} e^{\lambda x} \frac{1}{2a} \, dx = \frac{1}{2a} \lambda (e^{\lambda a} - e^{-\lambda a}) = \frac{\sinh(\lambda a)}{\lambda a} \leq \frac{e^{\lambda^2 a^2/2}}{2}.$$

for nonzero $\lambda$ and the inequality holds for $\lambda = 0$.

Lemma 3.2 1. $X \in G(\nu) \Rightarrow \alpha X \in G(|\alpha|\nu)$ for all $\alpha \in \mathbb{R}$.

2. If $X_1 \in G(\nu_1)$, $X_2 \in G(\nu_2)$, then $X_1 + X_2 \in G(\nu_1 + \nu_2)$.

3. If $X_1 \in G(\nu_1)$, $X_2 \in G(\nu_2)$ and further $X_1$ and $X_2$ are independent, then $X_1 + X_2 \in G(\sqrt{\nu_1^2 + \nu_2^2})$.

Proof: (1) and (3) are trivial to show. For the second part, we use Holder’s inequality as follows, with $1/p = \nu_2/(\nu_1 + \nu_2)$ and $1/q = 1 - 1/p$:

$$\mathbb{E}e^{\lambda(X_1 + X_2)} = \mathbb{E}[e^{\lambda X_1} e^{\lambda X_2}] \leq (\mathbb{E}e^{p\lambda X_1})^{1/p} (\mathbb{E}e^{q\lambda X_2})^{1/q} \leq e^{p\lambda \nu_1^2/2} e^{q\lambda \nu_2^2/2} \leq e^{\lambda^2/2(\nu_1 + \nu_2)^2}.$$

Note that $\sqrt{\nu_1^2 + \nu_2^2} \leq \nu_1 + \nu_2$ for positive $\nu_1, \nu_2$ and hence in the third part, we get a tighter bound with the additional assumption of independence.

Characterization of Sub-Gaussianity: If $\mathbb{E}X = 0$, then the following are equivalent.

1. $\exists \nu > 0 \text{ s.t } \mathbb{E}e^{\lambda X} \leq e^{\lambda^2 \nu^2/2}$ for all $\lambda \in \mathbb{R}$.

2. $\exists c > 0 \text{ s.t } \mathbb{P}(|X| > \lambda) \leq 2e^{-c\lambda^2}$ for all $\lambda > 0$.

3. $\exists a > 0 \text{ s.t } \mathbb{E}e^{\alpha X^2} \leq 2$.

4. $\forall p \geq 1$, $(\mathbb{E}|X|^p)^{1/p} \leq B\nu \sqrt{p}$ for some $B > 0$.

Note that linear combinations of sub-Gaussians are sub-Gaussian. Also, if $X$ is such that $\mathbb{E}X = 0, a \leq X \leq b$ almost surely, then $X \in G((b - a)/2)$. 
Hoeffding’s inequality

**Theorem 3.3 (Hoeffding’s inequality)** Let $X_1, X_2, \ldots, X_n$ be independent random variables such that $E X_i = 0, a_i \leq X_i \leq b_i$ almost surely. Then letting $S_n = \sum_{i=1}^{n} X_i$, for $t > 0$

$$P(S_n \geq t) \leq \exp \left( -\frac{2t^2}{\sum_{i=1}^{n} (b_i - a_i)^2} \right).$$

There are several proofs for the inequality. Most of them use the fact that $X_i \in G(\frac{(b_i - a_i)^2}{8})$. Generally, the proof shows that

$$\psi_{X_i, EX_i} \leq \frac{\lambda^2(b_i - a_i)^2}{8}, \quad \psi'_{X_i, EX_i} \leq \frac{(b_i - a_i)^2}{4}$$

**Example:** Let $X_1, \ldots, X_n$ be independent and $X_i \sim \text{Bernoulli}(p_i)$ where $p_i \in (0, 1)$, for $i = 1, \ldots, n$. Then from Hoeffding’s inequality, denoting $p = \sum_{i=1}^{n} p_i$,

$$P(|\bar{X}_n - p| > t) \leq 2e^{-2nt^2}.$$ 

In other words, for $\delta \in (0, 1)$, the following holds with probability at least $1 - \delta$:

$$|\bar{X}_n - p| \leq \sqrt{\frac{1}{2n \log \frac{2}{\delta}}}.$$ 

If $\delta = \delta_n = n^{-c}$ where $c > 0$, then the statement holds with probability at least $1 - n^{-c}$ for an rhs that is $O\left(\sqrt{\frac{\log n}{n}}\right)$.

Using Chernoff bounds,

$$P(\bar{X}_n - p \geq t) \leq \exp \left( -n H_p(p + t) \right) \text{ for } 0 < t < 1 - p$$

$$P(\bar{X}_n - p \leq -t) \leq \exp \left( -n H_{1-p}(1 - p + t) \right) \text{ for } 0 < t < p$$

where $H_p(x) = x \log(\frac{x}{p}) + (1 - x) \log(\frac{1-x}{1-p})$. This bound is tighter than Hoeffding’s bound.

There is also a multiplicative version of concentration inequality, namely:

$$P\left( \sum X_i \geq (1 + \epsilon)\mu \right) \leq e^{-\epsilon^2\mu/3}$$

$$P\left( \sum X_i \leq (1 - \epsilon)\mu \right) \leq e^{-\epsilon^2\mu/2}$$

where $\mu = np$. Multiplicative bounds can also be better than Hoeffding too. Let $X_1, \ldots, X_n$ be iid Bernoulli($p$). Then Hoeffding and the multiplicative bounds give respectively,

$$P\left( p - \bar{X}_n \geq t \right) \leq e^{-2nt^2}$$

$$P\left( p - \bar{X}_n \geq \epsilon p \right) \leq e^{-np\epsilon^2/2}$$

which lead to (respectively)

$$P\left( p - \bar{X}_n \geq \sqrt{\frac{1}{2n \log \frac{2}{\delta}}} \right) \leq \delta$$

$$P\left( p - \bar{X}_n \geq \sqrt{\frac{2p}{n \log \frac{1}{\delta}}} \right) \leq \delta.$$
The second one is better than the first one if $p \leq 1/4$, and gets much better as $p \to 0$.

Hoeffding’s inequality can be sharpened to take into account where $E_X$ falls with respect to its bounds $a, b$. In the above case, as $p \to 0$, $E_X$ goes closer to the lower bound. $E_X = \frac{a+b}{2}$ is best for Hoeffding. In the next lecture we see how Berend and Kantorovich overcome asymmetric situations that is handicapping the Hoeffding inequality.