LECTURER: PROF. MCDONALD SCRIBE: LEI DING

More details about this lecture can be found on Boucheron et al. (2013) in chapter 2.

12.1 Sub-gamma Random Variables

Definition 12.1 (Sub-gamma random variables). *A real-valued centered random variable X is said to be* ***sub-gamma on the right tail with variance factor*** σ ***and scale parameter*** c *if*

$$\psi_X(\lambda) = log\mathbb{E}\Big[e^{\lambda X}\Big] \leq \frac{\lambda^2\sigma}{2(1-c\lambda)} \qquad for\ \forall\lambda \in (0, \frac{1}{c}) \tag{12.1}$$

We denote as $X \sim \Gamma_+(\sigma, c)$. Similarly, X is said to be ***sub-gamma on the left tail with variance factor*** σ ***and scale parameter*** c *if $-X$ is sub-gamma on the right tail with variance factor σ and scale parameter c. We denote as $X \sim \Gamma_-(\sigma, c)$. Finally, X is said to be* ***sub-gamma with variance factor*** σ ***and scale parameter*** c *if X is sub-gamma both on the right and left tails with the same variance factor σ and scale parameter c. We denote as $X \sim \Gamma(\sigma, c)$.*

Observe that $\Gamma(\sigma, 0) = \mathcal{G}(\sigma)$.

Example 12.2 (centered gamma variable is a typical example of a sub-gamma variable). *Let $Y \sim Gamma(a, b)$, that is Y has density*

$$f(y) = \frac{y^{a-1}e^{-y/b}}{\Gamma(a)b^a}, \qquad y \geq 0 \tag{12.2}$$

Also, we have $\mathbb{E}[Y] = ab$ and $\mathrm{Var}[Y] = ab^2$. Let$X = Y - \mathbb{E}[Y]$. We first consider the right tail of X. Then for all $0 < \lambda < 1/b$,

$$\psi_X(\lambda) = a(-log(1-\lambda b) - \lambda b) \leq \frac{\lambda^2 ab^2}{2(1-b\lambda)} \tag{12.3}$$

that is $X \sim \Gamma_+(ab^2, b)$. Similarly we can prove that for the left tail $X \sim \Gamma_-(ab^2, 0)$. Thus $X \sim \Gamma(ab^2, b)$.

Theorem 12.3. *Let X be a centered random variable.Then for some $v > 0$, the following statements are equivalent*

1. $\mathbb{E}\big[X^{2q}\big] \leq q!(8v)^q + (2q)!(4c)^{2q}$ *for every integer* $q \geq 1$
2. $X \in \Gamma(4(8v + 16c^2), 8c)$
3. $\mathbb{P}\Big[X > \sqrt{8(8v+16c^2)t} + 8ct\Big] \vee \mathbb{P}\Big[-X > \sqrt{8(8v+16c^2)t} + 8ct\Big] \leq e^{-t}$ *for every* $t > 0$

12.2 Maximal Inequalities

Intuition: If we know the information about the Cramér transform of random variables in a finite collection, how can we use this information to bound the expected maximum of these random variables?

For example, if $Z_1, \cdots, Z_n$ follows i.i.d. σ-sub-Gaussian, we want to get a upper bound for $\mathbb{E}\Big[\max_{i=1,\cdots,n} Z_i\Big]$. By Jensen's inequality

$$\exp(\lambda \mathbb{E}\Big[\max_{i=1,\cdots,n} Z_i\Big]) \leq \mathbb{E}\Big[\exp(\lambda \max_{i=1,\cdots,n} Z_i)\Big] \tag{12.4}$$

$$= \mathbb{E}\Big[\max_{i=1,\cdots,n} e^{\lambda Z_i}\Big] \tag{12.5}$$

$$\leq \sum_{i=1}^{n} \mathbb{E}\Big[e^{\lambda Z_i}\Big] \tag{12.6}$$

$$\leq n e^{\lambda^2 \sigma^2/2} \tag{12.7}$$

Take logarithms on both sides, we have

$$\mathbb{E}\Big[\max_{i=1,\cdots,n} Z_i\Big] \leq \frac{\log n}{\lambda} + \frac{\lambda \sigma^2}{2} \tag{12.8}$$

for any $\lambda > 0$. This upper bound is minimized for $\lambda = \sqrt{2\log n/\sigma^2}$, which yields

$$\mathbb{E}\Big[\max_{i=1,\cdots,n} Z_i\Big] \leq \sigma\sqrt{2\log n} \tag{12.9}$$

Theorem 12.4. *Let $Z_1, \cdots, Z_n$ be independent random variables such that for any $\lambda \in (0, b)$ and $i = 1, \cdots, n$, $\psi_{Z_i}(\lambda) \leq \psi(\lambda)$, where ψ is a convex and continuously differentiable function on $[0, b)$, with $0 < b \leq \infty$ and $\psi(0) = \psi'(0) = 0$. Then*

$$\mathbb{E}\Big[\max_{i=1,\cdots,n} Z_i\Big] \leq \psi^{*-1}(\log n) \tag{12.10}$$

where

$$\psi^*(t) = \sup_{\lambda \in (0,b)} (\lambda t - \psi(\lambda)) \tag{12.11}$$

Useful results:

1. If $Z_i \in \Gamma_+(v, c)$, then $\mathbb{E}\Big[\max_{i=1,\cdots,n} Z_i\Big] \leq \sqrt{2v\log n} + c\log n$
2. If $Z_i \sim \chi^2(p) - p$, then $Z_i \in \Gamma_+(2p, 2)$

12.3 Inequalities for sum of independent random variables

Theorem 12.5 (Hoeffding's Inequality)**.** *Let $X_1, \cdots, X_n$ be independent random variables such that X_i takes its values in $[a_i, b_i]$ almost surely for all $i = 1, \cdots, n$. Then for every $\delta > 0$,*

$$\mathbb{P}\left[\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq \delta\right] \leq \exp(-\frac{2\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}) \tag{12.12}$$

Furthermore, if $X_1, \cdots, X_n$ are i.i.d. random variables and each X_i takes its values in $[a, b]$ almost surely, then for every $\delta > 0$,

$$\mathbb{P}\Big[\overline{X} - \mathbb{E}[X] \geq \delta\Big] \leq \exp(-\frac{2n\delta^2}{(b-a)^2}) \tag{12.13}$$

Theorem 12.6 (Bennett's Inequality). *Let $X_1, \cdots, X_n$ be independent random variables with finite variance such that $X_i \leq b$ for some $b > 0$ almost surely for all $i = 1, \cdots, n$. Let $v = \sum_{i=1}^n \mathbb{E}\big[X_i^2\big]$. Then for every $\delta > 0$,*

$$\mathbb{P}\left[\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \geq \delta\right] \leq \exp\left(-\frac{v}{b^2} h\left(\frac{b\delta}{v}\right)\right) \tag{12.14}$$

where

$$h(u) = (1+u)\log(1+u) - u \tag{12.15}$$

for $u > 0$.

Corollary 12.7 (Bernstein's Inequality). *Let $X_1, \cdots, X_n$ be independent random variables with finite variance such that $X_i \leq b$ for some $b > 0$ almost surely for all $i = 1, \cdots, n$. Let $v = \sum_{i=1}^n \mathbb{E}\big[X_i^2\big]$. Then for every $\delta > 0$,*

$$\mathbb{P}\left[\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \geq \delta\right] \leq \exp\left(-\frac{\delta^2}{2(v + b\delta/3)}\right) \tag{12.16}$$

Proof hint: use $h(u) = (1+u)\log(1+u) - u \geq \frac{u^2}{2(1+u/3)}$

Example 12.8 (Gaussian chaos of order two). *Let $X \sim N_n(0, I)$, A be a symmetric $n \times n$ matrix with 0 along the diagonal. Let $Z = X^t AX$. Then for every $\delta > 0$,*

$$\mathbb{P}[Z > \delta] \leq \exp\left(-\frac{\delta^2}{4\|A\|_F^2 + \|A\|\,\delta}\right) \tag{12.17}$$

Proof:(sketch)

1. *$Z \sim \sum_{i=1}^n u_i(X_i^2 - 1)$, where u_i are eigenvalues of A*
2. *$\psi_{X_i^2-1}(\lambda) = \frac{1}{2}(-\log(1-2\lambda) - 2\lambda) \leq \frac{\lambda^2}{1-2\lambda}$, for all $\lambda < 1/2$*
3. *$\psi_Z(\lambda) = \sum_{i=1}^n \frac{1}{2}(-\log(1-2u_i\lambda) - 2u_i\lambda) \leq \sum_{i=1}^n \frac{u_i^2\lambda^2}{1-2(u_i)_+\lambda} \leq \frac{\lambda^2\|A\|_F^2}{1-2\lambda\|A\|}$, for all $\lambda \in (0, 1/(2max_i u_i))$*

Example 12.9 (application to ecdf). *Let $X_1, \cdots, X_n$ be i.i.d. random variables. Consider the empirical cumulative distribution function (ecdf) $\widehat{F}(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \leq x)$. Then for any $\epsilon > 0$,*

$$\mathbb{P}\left[\left|\widehat{F}(x) - F(x)\right| > \epsilon\right] \leq 2e^{-2n\epsilon^2} \tag{12.18}$$

Meanwhile, we have DKW inequality

$$\mathbb{P}\left[\sup_x \left|\widehat{F}(x) - F(x)\right| > \epsilon\right] \leq 2e^{-2n\epsilon^2} \tag{12.19}$$

Proof:(sketch) Define $Y_i = I(X_i \leq x)$, then $Y_i \in \{0, 1\}$, $\mathbb{E}[Y_i] = \mathbb{E}\big[I(X_i \leq x)\big] = F(X)$. By applying Hoeffding's Inequality in Equation 12.13 with $b = 1$, $a = 0$, we can prove the first inequality.

References

Boucheron, S., Lugosi, G., and Massart, P. (2013), *Concentration Inequalities*, Oxford University Press, Oxford, UK.