CS 388R: Randomized Algorithms, Fall 2019 October 24th, 2019

Lecture 17: Subgamma variables and Johnson-Lindenstrauss

Prof. Eric Price *Scribe: Sebastian Oberhoff*

NOTE: THESE NOTES HAVE NOT BEEN EDITED OR CHECKED FOR CORRECTNESS

1 Subgamma Variables

Definition 1. *A random variable X is* subgamma *with variance proxy σ^2 and exponential scale c if:*

$$\text{(I)} \qquad \mathbb{E}\left[e^{\lambda(X-\mathbb{E}[X])}\right] \leq e^{\frac{\lambda^2\sigma^2}{2}} \qquad \forall\, |\lambda| \leq \frac{1}{c}$$

This definition implies:

(II)

$$\mathbb{P}\big[|X-\mathbb{E}[X]| \geq t\big] \leq 2 \cdot \max\left(e^{-\frac{t^2}{2\sigma^2}}, e^{-\frac{2}{2c}}\right).$$

(III) With probability $\geq 1-\delta$,

$$|X-\mathbb{E}[X]| \leq \sqrt{2\sigma^2 \ln\left(\frac{2}{\delta}\right)} + 2c\ln\left(\frac{2}{\delta}\right).$$

Either of the latter two properties also implies the definition (I), with a loss in parameters: $\sigma^2 \mapsto O(\sigma^2 + c^2)$ and $c \mapsto O(c)$.

This is a generalization of the subgaussian random variables we considered in last class, with the introduction of c. In particular, Subgaussian(σ^2) = Subgamma$(\sigma^2, 0)$.

Example

Let $Z \sim \mathcal{N}(0,1)$ and $X = Z^2$.

Then

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}}\, dz = 1\,. \qquad \text{(split the integrand into } z \cdot ze^{-z^2/2} \text{ and use integration by parts)}$$

And the centered MGF is

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

$$
\begin{aligned}
&= e^{-\lambda}\sigma \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2\sigma^2}}\, dz && \left(\lambda - \frac{1}{2} = -\frac{1}{2\sigma^2}\right) \\
&= e^{-\lambda}\sigma \\
&= \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \\
&\leq e^{\frac{4\lambda^2}{2}}. && \left(\text{if } |\lambda| \leq \frac{1}{4}\right)
\end{aligned}
$$

So X is Subgamma$(4, 4)$.

Alternatively, if Z is mean 0 and Subgaussian(σ^2) and $X = Z^2$ then we claim that X is

$$\text{Subgamma}\big(O(\sigma^4), O(\sigma^2)\big)\,.$$

To show this note that from the definition of Subgaussian random variables we know that with probability $1 - \delta$

$$Z = O\left(\sigma\sqrt{\log\left(\frac{2}{\delta}\right)}\right).$$

So by squaring both sides, we have

$$Z^2 = O\left(\sigma^2 \log\left(\frac{2}{\delta}\right)\right)$$

which fits (III) if $\sigma^2 = 0$ and $c = O(\sigma^2)$. Thus, after conversion we end up with variance proxy $O(\sigma^4)$ and exponential scale $O(\sigma^2)$.

2 Basic Properties

- If X_1, X_2 are independent and are (σ_1^2, c_1)- and (σ_2^2, c_2)-Subgamma, then $X_1 + X_2$ is $\big(\sigma_1^2 + \sigma_2^2, \max(c_1, c_2)\big)$-Subgamma.
- If X is (σ^2, c)-Subgamma, then αX is $(\alpha^2\sigma^2, \alpha c)$-Subgamma for any constant α.

3 Johnson-Lindenstrauss-Lemma (84)

Theorem 2 (Johnson-Lindenstrauss-Lemma)**.** *Let $X_1, \ldots, X_n \in \mathbb{R}^d$. Then there exist $y_1, \ldots, y_n \in \mathbb{R}^m$ (m "small") such that for all i, j:*

$$\|y_i - y_j\|_2 = (1 \pm \epsilon)\|x_i - x_j\|_2$$

Theorem 3 (Distributional Johnson-Lindenstrauss-Lemma)**.** *There exists a random linear map $A \in \mathbb{R}^{m \times d}$ (entries of $A \sim \mathcal{N}(0, 1/m)$) such that $\forall x \in \mathbb{R}^d$*

$$\|Ax\|_2 = (1 \pm \epsilon)\|x\|_2 \qquad \textit{with probability } 1 - 2e^{-\Omega(\epsilon^2 m)}$$

(or with probability $1 - \delta$ if $m = O\big(\frac{1}{\epsilon^2}\log(\frac{2}{\delta})\big)$).

From The Distributional To the Standard Johnson-Lindenstrauss-Lemma

Set $y_i = Ax_i$ with $\delta = \frac{1}{n^3}$. Then with probability $1 - \frac{1}{n}$ we have for all i, j:

$$\underbrace{\|A(x_i - x_j)\|}_{=\|y_i - y_j\|} = (1 \pm \epsilon)\|x_i - x_j\|$$

(Probability > 0 certainly implies existence.)

Proving The Distributional Johnson-Lindenstrauss-Lemma

Proof. Select the entries of A according to $\mathcal{N}(0, 1/m)$. Then, denoting the ith row of A as A_i, we have

$$\forall x: \qquad y_i = A_i x \sim \sum_{j=1}^{m} \mathcal{N}(0, 1/m) x_j = \mathcal{N}\Big(0, \underbrace{\sum_{j=1}^{m} x_j^2}_{=\|x\|_2^2} / m\Big)$$

meaning y_i is normally distributed with mean 0 and variance $\frac{1}{m}\|x\|_2^2$.

Therefore,

$$\mathbb{E}\big[\|y\|_2^2\big] = \sum i = 1^m \, \mathbb{E}[y_i^2] = m \cdot \mathrm{Var}[y_i] = \|x\|_2^2$$

All we need then is

$$\mathbb{P}\Big[\|y\|_2^2 \text{ is far from } \mathbb{E}\big[\|y\|_2^2\big]\Big] \leq \text{ something small.}$$

Suppose $y_i \sim \mathcal{N}(0, 1)$. How does $\sum_{i=1}^{m} y_i^2$ concentrate about the expected value m?

$\sum y_i^2$ is Subgamma$(4m, 4)$. Hence,

$$\begin{aligned}
\mathbb{P}\Big[\Big|\sum y_i^2 - m\Big| \geq t\Big] &\leq 2 \cdot \max\Big(e^{-\frac{t^2}{8m}}, e^{-\frac{t}{8}}\Big) \\
&\leq 2 \cdot \max\Big(e^{-\frac{\epsilon^2 m}{8}}, e^{-\frac{\epsilon m}{8}}\Big) && (t = \epsilon m) \\
&= 2e^{-\frac{\epsilon^2 m}{8}}. && (\text{if } \epsilon < 1)
\end{aligned}$$

□

4 Bernstein-type Bound

If $|X - \mu| \leq m$ with probability 1—in other words, if X is restricted to a finite interval—then X is Subgamma$\big(2 \cdot \mathrm{Var}[X], 2m\big)$.

Example 1

X_i is a coin with bias p_i towards 0. And $X = \sum_i X_i$. Then

$$\mathbb{E}[X] = \sum_i p_i = \mu \quad \text{and} \quad \mathrm{Var}[X] \le \mu\,.$$

Therefore, X is Subgamma$(2\mu, 2)$ which implies

$$\mathbb{P}\big[|X - \mu| \ge \epsilon\mu\big] \le 2e^{-\min(\epsilon^2,\epsilon)\mu/4}\,. \qquad \text{(multiplicative Chernoff)}$$

Example 2

In the coupon collector problem we had

$$T_i \sim \mathrm{geom}(p_i) \qquad p_i = \frac{n+1-i}{n}$$

where T_i was the arrival time of the ith item.

Consequently,

$$\mathbb{E}[T_i] = \frac{n+1-i}{n}$$

and

$$\mathbb{E}[T] = \sum_i \mathbb{E}[T_i] = nH_n = O(n\log(n))\,.$$

By the Bernstein bound we have that T_i is Subgamma$\Big(O\big(\frac{1}{p_i^2}\big), O\big(\frac{1}{p_i}\big)\Big)$. Therefore, $T = \sum T_i$ is

$$\mathrm{Subgamma}\left(O\left(\sum \frac{1}{p_i^2}\right), O\left(\max \frac{1}{p_i}\right)\right) = \mathrm{Subgamma}(n^2, n)\,.$$

Hence, with probability $1 - \delta$

$$\begin{aligned} T &\le \mathbb{E}[T] + O\left(\sqrt{n^2 \log\left(\frac{2}{\delta}\right)} + n\log\left(\frac{1}{\delta}\right)\right) \\ &= O\left(n\log\left(\frac{n}{\delta}\right)\right). \end{aligned}$$