TTIC 31150/CMSC 31150
Mathematical Toolkit (Spring 2023)

Avrim Blum and Ali Vakilian

Lecture 14: Probability over uncountably-infinite spaces, Gaussian RVs, Johnson-Lindenstrauss Lemma

Recap

- Chernoff-Hoeffding bounds
- Use in randomized algorithm for routing to minimize congestion.
- Randomized complexity classes **RP** and **BPP**, connections to **P/poly**.

Probability over uncountably-infinite spaces

- In finite or countable probability spaces, we could think of the probability distribution ν as a function from Ω to $[0,1]$, assigning a probability to each element of Ω.

- In uncountably-infinite spaces, like $\Omega = \mathbb{R}$, this is problematic:

  - At most n points x can have $\nu(x) \geq 1/n$.

  - Only countably many points x can have $\nu(x) > 0$. (Any such x must have $\nu(x) \geq 1/n$ for some integer n).

To resolve, will only talk about probabilities of events from an allowed set of events known as a σ-algebra or σ-field.

Probability over uncountably-infinite spaces

Definition 1.1 *Let 2^Ω denote the set of all subsets of Ω. A set $\mathcal{F} \subseteq 2^\Omega$ is called a σ-*field *(or σ-algebra) if*

1. $\emptyset \in \mathcal{F}$.
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ *(where $A^c = \Omega \setminus A$).*
3. *For a (countable) sequence $A_1, A_2, \ldots$ such that each $A_i \in \mathcal{F}$, we have $\cup_i A_i \in \mathcal{F}$.*

It is also closed under countable intersections (by De Morgan's laws)

The sets in $\mathcal{F}$ are the allowed events that may have probabilities (the *measurable* sets).

Definition 1.2 *Given a σ-field $\mathcal{F} \subseteq 2^\Omega$, a function $\nu : \mathcal{F} \to [0, 1]$ is known as a* probability measure *if*

1. $\nu(\emptyset) = 0$.
2. $\nu(E^c) = 1 - \nu(E)$ *for all $E \in \mathcal{F}$.*
3. *For a (countable) sequence of disjoint sets $E_1, E_2, \ldots$ such that all $E_i \in \mathcal{F}$, we have*

$$\nu(\cup_i E_i) = \sum_i \nu(E_i).$$

Not necessarily for uncountably-infinite unions

Probability over uncountably-infinite spaces

Definition 1.1 *Let 2^Ω denote the set of all subsets of Ω. A set $\mathcal{F} \subseteq 2^\Omega$ is called a σ-field (or σ-algebra) if*

1. $\emptyset \in \mathcal{F}$.
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ *(where $A^c = \Omega \setminus A$).*
3. *For a (countable) sequence $A_1, A_2, \ldots$ such that each $A_i \in \mathcal{F}$, we have $\cup_i A_i \in \mathcal{F}$.*

The Borel σ-algebra is the smallest σ-algebra on $\mathbb{R}$ that contains all intervals.

A real-valued **random variable** X is a measurable function over $(\Omega, \mathcal{F}, \nu)$: a function from Ω to $\mathbb{R}$ such that for every Borel set B, the set $X^{-1}(B) = \{\omega: X(\omega) \in B\}$ is a measurable set (in $\mathcal{F}$).

Equivalently, for any $c \in \mathbb{R}$, $\{\omega: X(\omega) \leq c\}$ is a measurable set, and so has a well-defined probability.

Often, we will think of a random variable as just a probability distribution on its range.

Random variables

- Given a R.V. X, we define its cumulative distribution function $F_X(z) = \mathbb{P}[X \leq z]$.
- Can observe that F_X is a non-decreasing function. If it is differentiable, then its derivative f is the density function of X, and we typically refer to X as a continuous R.V.
- $\mathbb{E}[X] = \int_\Omega X(\omega) d\nu = \int_{-\infty}^{\infty} x f(x) dx$.
- For discrete RVs, we had $\mathbb{E}[X] = \sum_\omega X(\omega)\nu(\omega) = \sum_a a \cdot \mathbb{P}(X = a)$.

Gaussian Random variables

- A Gaussian Random Variable is an R.V. with density $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, for some μ and σ^2 which are its mean and variance respectively.
- Notationally, we write $X \sim N(\mu, \sigma^2)$.

Claim: $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1.$

Proof: $\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx\right)^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-x^2/2} dx \int_{-\infty}^{\infty} e^{-y^2/2} dy$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy = \frac{1}{2\pi} \int_{0}^{\infty} \int_{0}^{2\pi} e^{-r^2/2} r dr d\theta$$

$$= \int_{0}^{\infty} e^{-r^2/2} r dr = -e^{-r^2/2} \Big|_{0}^{\infty} = 1.$$

Gaussian Random variables

A couple more useful facts we'll need:

- For $X \sim N(0,1), \lambda \in (0,1/2)$, $\mathbb{E}\left[e^{\lambda X^2}\right] = 1/\sqrt{1-2\lambda}$.

- Let $Z = c_1 X_1 + c_2 X_2$ where $X_1, X_2 \sim N(0,1)$ are independent. Then $Z \sim N(0, c_1^2 + c_2^2)$.

  - One way to think of this: consider taking an inner-product between the vector $c = (c_1, c_2)$ and the vector (X_1, X_2). Because a d-dimensional Gaussian is spherically-symmetric, we can instead choose an orthogonal basis where one basis vector is $\hat{c} = {}^{c}/_{\|c\|}$ and the others are orthogonal (and so can be ignored). So, we just have a value taken from a single Gaussian, stretched by $|c|$.

Dimensionality Reduction and the Johnson-Lindenstrauss Lemma

Imagine you have n data points in a d-dimensional space, where d is large.

The JL lemma says that no matter how large d is, if you randomly project the data down to a space of dimension $k = O\left(\frac{\log n}{\epsilon^2}\right)$, then whp you will approximately preserve the relative distances between points up to a $1 \pm \epsilon$ factor.

So, if all you care about are approximate distances, then you can wlog assume your data is in a not-too-high dimensional space.

How to randomly project? Choose k random vectors $G_1, \ldots, G_k$ from spherical Gaussian, and project by inner-product: $v \to (\langle G_1, v\rangle, \ldots, \langle G_k, v\rangle)$. I.e., $v \to Gv$.

Dimensionality Reduction and the Johnson-Lindenstrauss Lemma

The JL Lemma: Let $v_1, \ldots, v_n \in \mathbb{R}^d$. Choose a random matrix $G \in \mathbb{R}^{k \times d}$ for $k = \frac{8 \ln n}{\epsilon^2/2 - \epsilon^3/2}$, with each $G_{ij} \sim N(0,1)$ independently. Consider $\varphi(v) = Gv/\sqrt{k}$.

With probability at least $1 - 1/n$, for all pairs v_i, v_j we have:

$$(1-\epsilon)\|v_i - v_j\|^2 \le \|\varphi(v_i) - \varphi(v_j)\|^2 \le (1+\epsilon)\|v_i - v_j\|^2.$$

Note that since φ is linear, $\varphi(v_i) - \varphi(v_j) = \varphi(v_i - v_j)$. So, it suffices to prove that for a single vector $w = v_i - v_j$, with probability at least $1 - 1/n^3$ we have:

$$(1-\epsilon)\|w\|^2 \le \|\varphi(w)\| \le (1+\epsilon)\|w\|^2.$$

And then apply a union bound.

Dimensionality Reduction and the Johnson-Lindenstrauss Lemma

Claim: Let $w \in \mathbb{R}^d$. Choose a random matrix $G \in \mathbb{R}^{k\times d}$ for $k = \frac{8 \ln n}{\epsilon^2/2 - \epsilon^3/2}$, with each $G_{ij} \sim N(0,1)$ independently. With probability at least $1 - 1/n^3$ we have:

$$(1-\epsilon)\|w\|^2 \leq \left\|Gw/\sqrt{k}\right\|^2 \leq (1+\epsilon)\|w\|^2.$$

Proof:

- Consider $\frac{(Gw)_i}{\|w\|} = \frac{G_i w}{\|w\|} = \frac{1}{\|w\|}\sum_j G_{ij} w_j$. This is a Gaussian RV $X_i \sim N(0,1)$.
- So, $\frac{\|Gw\|^2}{\|w\|^2} = \sum_{i=1}^k X_i^2$ where X_i are independent. $\mathbb{E}\left[\sum_i X_i^2\right] = k$.
- Just need to show that for $Z = \sum_i X_i^2$, whp, $(1-\epsilon)k \leq Z \leq (1+\epsilon)k$.

(In other words, need tail bound for sum of independent squared-Gaussian R.V.s)

Dimensionality Reduction and the Johnson-Lindenstrauss Lemma

Other direction is similar

$$
\begin{aligned}
\mathbb{P}\left[Z \geq (1+\varepsilon)k\right] &\leq \mathbb{P}\left[e^{\lambda Z} \geq e^{\lambda\cdot(1+\varepsilon)k}\right] \\
&\leq \frac{\mathbb{E}\left[e^{\lambda\cdot Z}\right]}{e^{\lambda\cdot(1+\varepsilon)k}} && \text{(by Markov's inequality)} \\
&= \frac{\mathbb{E}\left[e^{\lambda\cdot\sum_{i=1}^{k} X_i^2}\right]}{e^{\lambda\cdot(1+\varepsilon)k}} = \frac{\prod_{i=1}^{k}\mathbb{E}\left[e^{\lambda\cdot X_i^2}\right]}{e^{\lambda\cdot(1+\varepsilon)k}} && \text{(by the independence of } X_1,\ldots,X_k\text{)} \\
&= \frac{\prod_{i=1}^{k}\frac{1}{\sqrt{1-2\lambda}}}{e^{\lambda\cdot(1+\varepsilon)k}} && \text{(by Lemma 2.3)} \\
&\leq \left(\frac{e^{-2(1+\varepsilon)\lambda}}{1-2\lambda}\right)^{k/2} && \text{(assume } \lambda < 1/2\text{)} \\
&\leq \left(e^{-\varepsilon}(1+\varepsilon)\right)^{k/2} && \text{(let } \lambda = \frac{\varepsilon}{2(1+\varepsilon)}\text{)} \\
&\leq \left(\left(1-\varepsilon+\frac{\varepsilon^2}{2}\right)(1+\varepsilon)\right)^{k/2} && \text{(by Taylor expansion of } e^{-x}\text{)} \\
&\leq e^{-\left(\frac{\varepsilon^2}{2}-\frac{\varepsilon^3}{2}\right)\frac{k}{2}} && \text{(by } 1+x \leq e^x\text{)}
\end{aligned}
$$

Finally, for $k = \frac{8\ln n}{\epsilon^2/2-\epsilon^3/2}$, this is sufficiently small

Dimensionality Reduction and the Johnson-Lindenstrauss Lemma

Conclusion: if you only care about approximate distances, approximate angles, etc, then can assume wlog that data lies in a space of dimension no greater than $O(\frac{\log n}{\epsilon^2})$.

Use for: approximate nearest-neighbor, streaming algorithms, …