CSE 392: Matrix and Tensor Algorithms for Data

Instructor: Shashanka Ubaru

University of Texas, Austin
Spring 2024

Lecture 2: Probability Review

Outline

1. Probability review

2. Concentration inequalities

This lecture

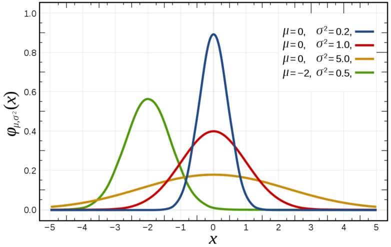
Topics to be covered today

- Probability and properties.
- Concentration measures.
  - Markov and Chebyshev inequality
  - CLT and tail bounds

Probability review

Let x be a random variable taking value in some set $\mathbb{S}$.
For continuous random variable, it might be $\mathbb{S} = \mathbb{R}$.

- **Expectation:** $\mathbb{E}[\mathrm{x}] = \sum_{\mathrm{s} \in \mathbb{S}} \mathrm{s} \cdot \Pr[\mathrm{x} = \mathrm{s}]$
  For continuous case, $\mathbb{E}[\mathrm{x}] = \int_{\mathrm{s} \in \mathbb{S}} \mathrm{s} \cdot \Pr[\mathrm{x} = \mathrm{s}] ds$
- **Variance:** $\mathrm{Var}[\mathrm{x}] = \mathbb{E}\left[(\mathrm{x} - \mathbb{E}[\mathrm{x}])^2\right] = \mathbb{E}[\mathrm{x}^2] - \mathbb{E}[\mathrm{x}]^2$



Excerise 1: For any scalar α, show that $\mathbb{E}[\alpha \mathrm{x}] = \alpha \mathbb{E}[\mathrm{x}]$ and $\mathrm{Var}[\alpha \mathrm{x}] = \alpha^2 \mathrm{Var}[\mathrm{x}]$.

Probability review

Let A and B be random events. Then,

- **Joint Probability:** $\Pr(A \cap B)$ - The probability that both events happen.
- **Conditional Probability:** $\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$. Probability A happens conditioned on the event that B happens.
- **Independence:** A and B are independent events if: $\Pr(A \mid B) = \Pr(A)$. For independent events, we also have that

$$\Pr(A \cup B) = \Pr(A) \cdot \Pr(B)$$

- **Mutually exclusive** events : $\Pr(A \cap B) = 0$.

Probability review

Random sampling can be:

- *with replacement*
- *without replacement*

Question: Which of the above event is independent?
Example: What is the probability that for two independent dice rolls taking values uniformly in $\{1, 2, 3, 4, 5, 6\}$, the first roll comes up even and the second is < 4?

Expectation

For random variables x and y,

- **Linearity of expectation:** For constants $c_1, c_2 \in \mathbb{R}$,

$$\mathbb{E}[c_1 \mathrm{x} + c_2 \mathrm{y}] = c_1 \mathbb{E}[\mathrm{x}] + c_2 \mathbb{E}[\mathrm{y}].$$

  Result holds irrespective of the dependence between x and y.

- **Law of Total Expectation:** If the sample space is the disjoint union of events $A_1, A_2, \ldots,$ then

$$\mathbb{E}[\mathrm{x}] = \sum_i \mathbb{E}[\mathrm{x} \mid A_i] \Pr(A_i).$$

- **Product of expectation:** For any two independent random variables x and y,

$$\mathbb{E}[\mathrm{x} \cdot \mathrm{y}] = \mathbb{E}[\mathrm{x}] \cdot \mathbb{E}[\mathrm{y}]$$

  also $\mathrm{Var}[\mathrm{x} + \mathrm{y}] = \mathrm{Var}[\mathrm{x}] + \mathrm{Var}[\mathrm{y}]$.

Norms of random variables

- **Moment norm:** For a real random variable x and $p \geq 1$, let

$$|||\mathrm{x}|||_p = \mathbb{E}[|\mathrm{x}|^p]^{1/p}.$$

  We use $|||\cdot|||$ to distinguish from matrix/vector norm.
- For real random variables x, y and $p \geq 1$,
  (Minkowski) $|||\mathrm{x}+\mathrm{y}|||_p \leq |||\mathrm{x}|||_p + |||\mathrm{y}|||_p$,
  and for $\alpha \in \mathbb{R}$, $|||\alpha \mathrm{x}|||_p = |\alpha| \, |||\mathrm{x}|||_p$.
- **Centered random variables:** Random variable $\mathrm{x} \in \mathbb{R}$ is centered if $\mathbb{E}[\mathrm{x}] = 0$.
- **Tail from norms:** For $t > 0$, for centered x,

$$\Pr\{|\mathrm{x}| \geq t\} \leq |||\mathrm{x}|||_p^p / t^p$$

- For centered x, $\|\|x\|\|_2^2 = \mathbb{E}[x^2] = \mathrm{Var}[x]$. So, $\|\|x\|\|_2 = \mathrm{sd}[x]$.
- We know that for two independent random variables x, y,

$$\mathrm{Var}[x+y] = \mathrm{Var}[x] + \mathrm{Var}[y].$$

So, if they are also centered, then

$$\|\|x+y\|\|_2 = \sqrt{\|\|x\|\|_2 + \|\|y\|\|_2} \leq \|\|x\|\|_2 + \|\|y\|\|_2.$$

- **Sub-Guassian norms:** For a real random variable x,

$$\|\|x\|\|_{\psi_2} \equiv \sup_{p\geq 1} \|\|x\|\|_p / \sqrt{p}$$

If $\|\|x\|\|_{\psi_2}$ is bounded, we call x *sub-Gaussian*.

Concentration inequalities

One of the key tools in analyzing randomized algorithms.
How likely a random variable x deviates a certain amount from its expectation $\mathbb{E}[x]$.

We will learn three fundamental concentration inequalities:

- **Markov's Inequality** - Applies to *non-negative* random variables.
- **Chebyshev's Inequality** - For random variables with *bounded variance*.
- **Hoeffding/Bernstein/Chernoff bounds** - For *sums* of *independent* random variables.

Markov's Inequality

For any random variable x which only takes *non-negative* values, and any positive t,

$$\Pr[x \geq t] \leq \frac{\mathbb{E}[x]}{t}.$$

Equivalently, $\Pr[x \geq \alpha \cdot \mathbb{E}[x]] \leq \frac{1}{\alpha}$.

Markov's Inequality

For any random variable x which only takes *non-negative* values, and any positive t,

$$\Pr[\mathrm{x} \geq t] \leq \frac{\mathbb{E}[\mathrm{x}]}{t}.$$

Equivalently, $\Pr[\mathrm{x} \geq \alpha \cdot \mathbb{E}[\mathrm{x}]] \leq \frac{1}{\alpha}$.

Proof: We have to show that $\mathbb{E}[\mathrm{x}] \geq t \cdot \Pr[\mathrm{x} \geq t]$:

$$\begin{aligned}
\mathbb{E}[\mathrm{x}] &= \sum_k k \cdot Pr(\mathrm{x} = k) \\
&\geq \sum_{k \geq t} k \cdot \Pr(\mathrm{x} = k) \\
&\geq \sum_{k \geq t} t \cdot \Pr(\mathrm{x} = k) \\
&= t \cdot \sum_{k \geq t} \Pr(\mathrm{x} = k)
\end{aligned}$$

Example

A coin is weighted so that its probability of landing on heads is 20%. Suppose the coin is flipped 20 times. Find a bound for the probability it lands on heads at least 16 times.

Example

A coin is weighted so that its probability of landing on heads is 20%. Suppose the coin is flipped 20 times. Find a bound for the probability it lands on heads at least 16 times.

Binomial distribution - $n = 20, p = 0.2$

$$\mathbb{E}[\mathrm{x}] = n \cdot p = 20 * 0.2 = 4.$$

Let us use **Markov's**:

$$\Pr[\mathrm{x} \geq 16] \leq \frac{\mathbb{E}[\mathrm{x}]}{16} = 0.25.$$

Is this a good estimate?

Popular applications: k-frequent items, hash functions, and others.

Union Bound

Union Bound

For any random events $A_1, \ldots, A_k$:

$$\Pr[A_1 \cup A_2 \cup \ldots \cup A_k] \leq \Pr[A_1] + \Pr[A_2] + \ldots + \Pr[A_k]$$

Union Bound

Union Bound

For any random events $A_1, \ldots, A_k$:

$$\Pr[A_1 \cup A_2 \cup \ldots \cup A_k] \leq \Pr[A_1] + \Pr[A_2] + \ldots + \Pr[A_k]$$

Proof: Choose $\mathrm{x}_i = \mathbb{1}[A_i]$, and we apply Markov's to $S = \sum_{i=1}^{k} \mathrm{x}_i$.
Hint: Express the union event $A_1 \cup A_2 \cup \ldots \cup A_k$ in terms of S. What is $\mathbb{E}[\mathrm{x}_i] = ?$

Chebyshev's Inequality

Let x be a random variable, then for any $k > 0$,

$$\Pr(|\mathrm{x} - \mathbb{E}[\mathrm{x}]| \geq k) \leq \frac{\mathrm{Var}[\mathrm{x}]}{k^2}.$$

Chebyshev's Inequality

Let x be a random variable, then for any $k > 0$,

$$\Pr(|\mathrm{x} - \mathbb{E}[\mathrm{x}]| \geq k) \leq \frac{\mathrm{Var}[\mathrm{x}]}{k^2}.$$

Proof: Note that

$$\Pr(|\mathrm{x} - \mathbb{E}[\mathrm{x}]| \geq k) = \Pr((\mathrm{x} - \mathbb{E}[\mathrm{x}])^2 \geq k^2).$$

Applying Markov's inequality to the random variable $(\mathrm{x} - \mathbb{E}[\mathrm{x}])^2$ gives us the result.

- Alternatively, for any $c > 0$,

$$\Pr(|\mathrm{x} - \mathbb{E}[\mathrm{x}]| \geq c \cdot \sigma_{\mathrm{x}}) \leq \frac{1}{c^2},$$

  where $\sigma_{\mathrm{x}} = \sqrt{\mathrm{Var}[\mathrm{x}]} = \sqrt{\mathbb{E}[(\mathrm{x} - \mathbb{E}[\mathrm{x}])^2]}$, is the *standard deviation* of x.

Properties of Chebyshev's inequality

- x need not be non-negative.
- It is a two-sided bound, gives the probability that $|\mathrm{x} - \mathbb{E}[\mathrm{x}]|$ is large or not.
  I.e., x is not too far above or below its expectation.
  Markov's only bounded probability that x exceeds $\mathbb{E}[\mathrm{x}]$.
- Probability of x being c times away from σ.
- We need a bound on the variance of x.

It is worst case bound, may not be tight in many cases.

Gaussian concentration

For $\mathrm{x} \sim \mathcal{N}(\mu, \sigma^2)$, we have:

$$\Pr[\mathrm{x} = \mu \pm x] \sim \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$

Gaussian Tail Bound:

For $\mathrm{x} \sim \mathcal{N}(\mu, \sigma^2)$,

$$\Pr[|\mathrm{x} - \mu| \geq k \cdot \sigma] \leq e^{-k^2/2}$$

Where as, using Chebyshevs's inequality we get $\Pr[|\mathrm{x} - \mu| \geq k \cdot \sigma] \leq 1/k^2$

Gaussian random variables concentrate *much tighter* around their expectation than what Chebyshevs's inequality predicts.

Central limit theorem

Lindeberg–Levy CLT:

Suppose $\{\mathrm{x}_1, \ldots, \mathrm{x}_n\}$ is a sequence of i.i.d. random variables with $\mathbb{E}[\mathrm{x}_i] = \mu$ and $\mathrm{Var}[\mathrm{x}_i] = \sigma^2 < \infty$. Then, as n approaches infinity, the random variables $\sqrt{n}(\bar{\mathrm{x}}_n - \mu)$, where $\bar{\mathrm{x}}_n = \sum_{i-1}^{n} \mathrm{x}_i/n$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$:

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{a} \mathcal{N}\left(0, \sigma^2\right).$$

CLT can be made rigorous to obtain tighter tail bounds than Chebyshevs's inequality.

- Chernoff bound
- Bernstein bound
- Hoeffding bound

Different assumptions on random variables (e.g. binary vs. bounded), different forms (additive vs. multiplicative error), etc.

Chernoff Bound

Chernoff Bounds

Let $X = \sum_{i=1}^{n} \mathrm{x}_i$, where $\mathrm{x}_i = 1$ with probability p_i and $\mathrm{x}_i = 0$ with probability $1 - p_i$, and all x_i are independent. Let $\mu = \mathbb{E}(X) = \sum_{i=1}^{n} p_i$. Then

- **Upper Tail:**. $\Pr(X \geq (1+\delta)\mu) \leq e^{-\frac{\delta^2}{2+\delta}\mu}$ for all $\delta > 0$;
- **Lower Tail:** $\Pr(X \leq (1-\delta)\mu) \leq e^{-\frac{\delta^2}{2}\mu}$ for all $0 < \delta < 1$;

Idea of proof:
Based on applying Markov's inequality to moment generating function $\mathbb{E}[e^{s|X-\mathbb{E}[X]|}]$.

Bernstein Inequality

Bernstein Inequality

Let $\mathrm{x}_1, \ldots, \mathrm{x}_k$ be independent random variables with each $\mathrm{x}_i \in [-1, 1]$. Let $\mathbb{E}[\mathrm{x}_i] = \mu_i$ and $\mathrm{Var}[\mathrm{x}_i] = \sigma_i^2$. Let $\mu = \sum_i \mu_i$ and $\sigma^2 = \sum_i \sigma_i^2$. Then, for $k \leq \frac{1}{2}\sigma, S = \sum_i \mathrm{x}_i$ satisfies

$$\Pr[|S - \mu| \geq k \cdot \sigma] \leq 2e^{-k^2/4}.$$

Idea of proof:
Based on applying Markov's inequality to $e^{\lambda \sum_i \mathrm{x}_i}$ for suitable choice of the parameter $\lambda > 0$.

Hoeffding Inequality

Hoeffding Inequality

Let $\mathrm{x}_1, \ldots, \mathrm{x}_k$ be independent random variables with each $\mathrm{x}_i \in [a_i, b_i]$. Let $\mathbb{E}[\mathrm{x}_i] = \mu_i$ and $\mathrm{Var}[\mathrm{x}_i] = \sigma_i^2$. Let $\mu = \sum_i \mu_i$ and $\sigma^2 = \sum_i \sigma_i^2$. Then, for and $\alpha > 0, S = \sum_i \mathrm{x}_i$ satisfies

$$\Pr[|S - \mu| \geq \alpha] \leq 2e^{-\frac{\alpha^2}{\sum_i (a_i - b_i)^2}}.$$

Idea of proof: Similar to Chernoff bounds. We use that for a real random variable $\mathrm{x} \in [a, b]$ almost surely,

$$\mathrm{E}\left[e^{s(\mathrm{x} - \mathrm{E}[\mathrm{x}])}\right] \leq \exp\left(\frac{1}{8} s^2 (b-a)^2\right).$$

Example

Coin flip application

We are given a biased coin which lands heads with probability p. How many k times should we flip to ensure

$$\Pr[|\#heads - p \cdot k| \geq \epsilon k] \leq \delta.$$

Setup: Let $\mathrm{x}_i = \mathbb{1}[i^{th} \text{ flip is heads}]$. We want bound probability that $S = \sum_{i=1}^{k} \mathrm{x}_i$ deviates from the expectation.

Using Chebyshev: $k \geq ?$
Using Chernoff: $k \geq ?$

Reference

Recommended reading:
A good reference for introduction and proofs of the various concentration inequalities, see Dr. Karthik Sridharan's article:

A Gentle Introduction to Concentration Inequalities.

Questions?