Probability

John Augustine (IIT Madras)
Krishna Palem (RICE University)

Reference for the content in these slides:
Probability and Computing Randomized Algorithms and Probabilistic Analysis
Michael Mitzenmacher, Eli Upfal

Acknowledgement:
These slides are prepared with the help of : S. Sumathi (IIT Madras)

List of Topics

1. Probability Spaces
2. Basics of Random Variables

The Basics of Probability

A random experiment is a procedure that can be repeated over an infinite number of *trials* and has a set of well defined outcomes. The probability space of such an experiment has 3 components

1. A Sample Space Ω, which is the set of all possible outcomes of the experiment.
2. A family of sets F representing the set of allowable events where each event or member of the family F is denoted by the symbol E.
   - Each set in F is a subset of the sample space Ω.
3. A probability measure or function $\mathbb{P}: F \rightarrow R$.
   - Satisfies additional conditions.

Tossing a Coin

- Consider the experiment of tossing a coin into the air. The two outcomes or events (E_i) will be either 'Heads' or 'Tails'.
- Therefore our sample space will $\Omega = \{H, T\}$
- F the set of allowable events will also be $\{H, T\}$
- Now if the coin is unbiased,
  - 50% chance that the outcome is heads and
  - a 50% chance the outcome will be tails.
- Therefore our probability measure will be

$$\mathbb{P}(H) = \mathbb{P}(T) = \frac{1}{2}$$

Probability Space (Contd.)

A *probability function* is any function $\mathbb{P}: F \rightarrow R$ that satisfies the following conditions

1. For any event $E, 0 \leq \mathbb{P} \leq 1$;
2. $\mathbb{P}(\Omega) = 1$: and
3. For any finite or countable finite sequence of pairwise mutually disjoint events $E_1, E_2, E_3, \ldots,$

$$\mathbb{P}\left(\bigcup_{i\geq 1} E_i\right) = \Sigma_{i\geq 1}\mathbb{P}(E_i)$$

What are Random Variables?

- Let's play a game with the coin from our example.
  - If the outcome is heads heads then A wins.
  - and B wins if the outcome is tails.
- Let X be the number of times A wins or loses in one toss.
- $X = \begin{cases} +1 \;\; if\ outcome\ is\ heads \quad (\boldsymbol{A\ wins}) \\ -1 \;\; if\ outcome\ is\ \ tails \quad (\boldsymbol{B\ wins\ or\ A\ loses}) \end{cases}$

How much can A win in one round?

- Using an unbiased coin, informally, there is a 50/50 chance that A wins or loses.
- Therefore, A is ``expected to win''
  - This intuitively is the expected value of random variable X.

$$\frac{1}{2}.(1) + \frac{1}{2}.(-1) = 0 \text{ per round.}$$

Random variables or R.V. more generally?

- The quantity we are interested in is how many times A wins per round of the game .
- The R.V. is called discrete if it takes only a finite or countably infinite number of values.
  - X is a discrete random variable in our example.
  - Integers form a countable set.
- It is continuous if it takes uncountably many infinite many values.
  - Real numbers are uncountable sets.

Expectation of a Random Variable

- Formally for a discrete random variable X, which takes value x_i with probability p_i , the ***expectation*** is defined to be

$$\mathbb{E}[X] = \sum_i x_i p_i$$

Linearity of Expectation

- Let $X_1, X_2, \dots, X_n$ be n random variables and $\lambda_1, \dots \lambda_n$ be n constants. Then

$$E(\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_n X_n) = \lambda_1 E[X_1] + \cdots \lambda_n E[X_n]$$

- The above result is called linearity of expectation

Variance of a Random Variable

- For both Discrete and continuous R.V. the ***variance*** $Var(X)$ is
$$Var(X) = E[X^2] - [E[X]]^2$$
- Intuitively, the variance indicated how far the random variable is from the average.

Some Probability Distribution Functions

- Consider again a coin, and let the probability of heads being the outcome be p, while the probability of tails being the toucome be $1 - p$.
- Let random variable $X = \begin{cases} 1\ if\ heads\ (success) \\ 0\ if\ tails\ (failure) \end{cases}$
- Note that for X, the expectation is simply
  $$\mathbb{E}[X] = 1.\mathbb{P}(X = 1) + 0 = \mathbb{P}(X = 1) = p$$
- Consider now a sequence of n coin flips.
  - What is the distribution or number of heads across the sequence?

Binomial Distribution and R.V.

- The sequence of coin flips can be considered to be n independent experiments each with probability p of success.
- If we let X be the *number of success* in n experiments, then X has a binomial distribution.
- **Definition** : A binomial R.V. X with parameter n and p, is define by the following probability distribution on $j = 0,1,2,\dots,n$.

$$\mathbb{P}(X = j) = \binom{n}{j} p^j (1-p)^{n-j}$$

Probability of $n - j$ tails

Probability of j heads

Number of ways in which n coin tosses can have j heads

Expectation of Binomial Distribution

- The expected number of successes (i.e., heads for the coin example) is np.

Tail bounds

Tail Bounds of Distributions

- What is the probability that a random variable will deviate from its expectation?
- One useful insight to start with: **Markov's inequality**.

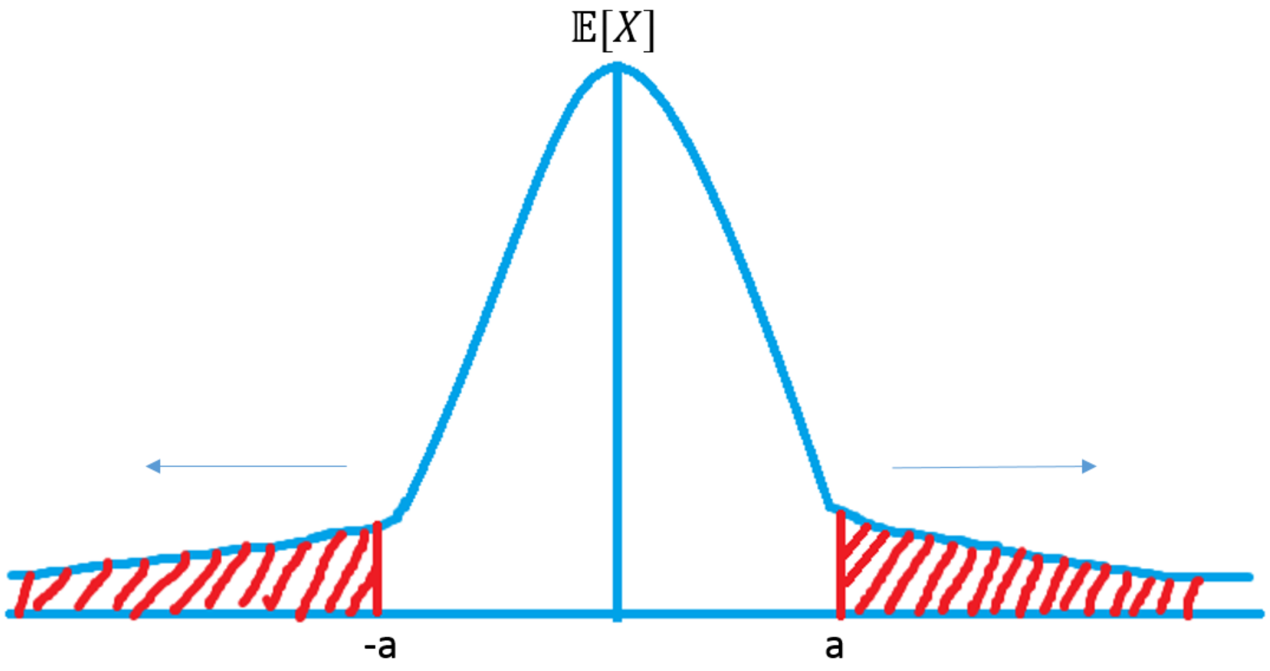
Let X be a nonnegative random variable. Then

$$\Pr(|X| \geq t) \leq \frac{\mathbb{E}[|X|]}{t} \text{ for all } t > 0$$

Meaning of Markov's Inequality

- Informally, Markov's inequality states that the probability of X taking a value much larger than the expectation is very small.
- For eg: Consider a random variable whose values are distributed normally as shown in the next slide

Illustration of Markov's Inequality:



Area under the curve denotes values X can take

Deriving Markov's Inequality

- Let X be a nonnegative random variable. Then

$$\Pr(|X| \geq t) \leq \frac{\mathbb{E}[|X|]}{t} \text{ for all } t>0$$

- **Proof:** Set up a new random variable I such that

$$I = \begin{cases} 1 & if\ X \geq t \\ 0 & otherwise \end{cases}$$

Deriving Markov's Inequality (contd):

- Now we know that $t.I \leq X$ and $\mathbb{E}[I] = \mathbb{P}(I = 1) = \mathbb{P}(X \geq t)$

$$\Rightarrow \Pr(X \geq \mathrm{t}) \leq \mathbb{E}\left[\frac{X}{t}\right] = \frac{E[X]}{t}$$

Because, from linearity $\mathbb{E}[\mathrm{tX}] = \mathrm{t}\mathbb{E}[X]$

Refer to: "Probability and Computing: Randomized Algorithms and Probabilistic Analysis" Chapter: 3 Page: 44

Chernoff Bound

- The Chernoff bounds give exponentially decreasing bound on tail distribution.
- The most commonly used one is for n *independent Poisson trials*.
- Poisson trials are 0-1 independent R.V., where each R.V. does not necessarily have the same distribution.
  - Indicator random variables are used to model Poisson trials.

Chernoff Bounds Contd.

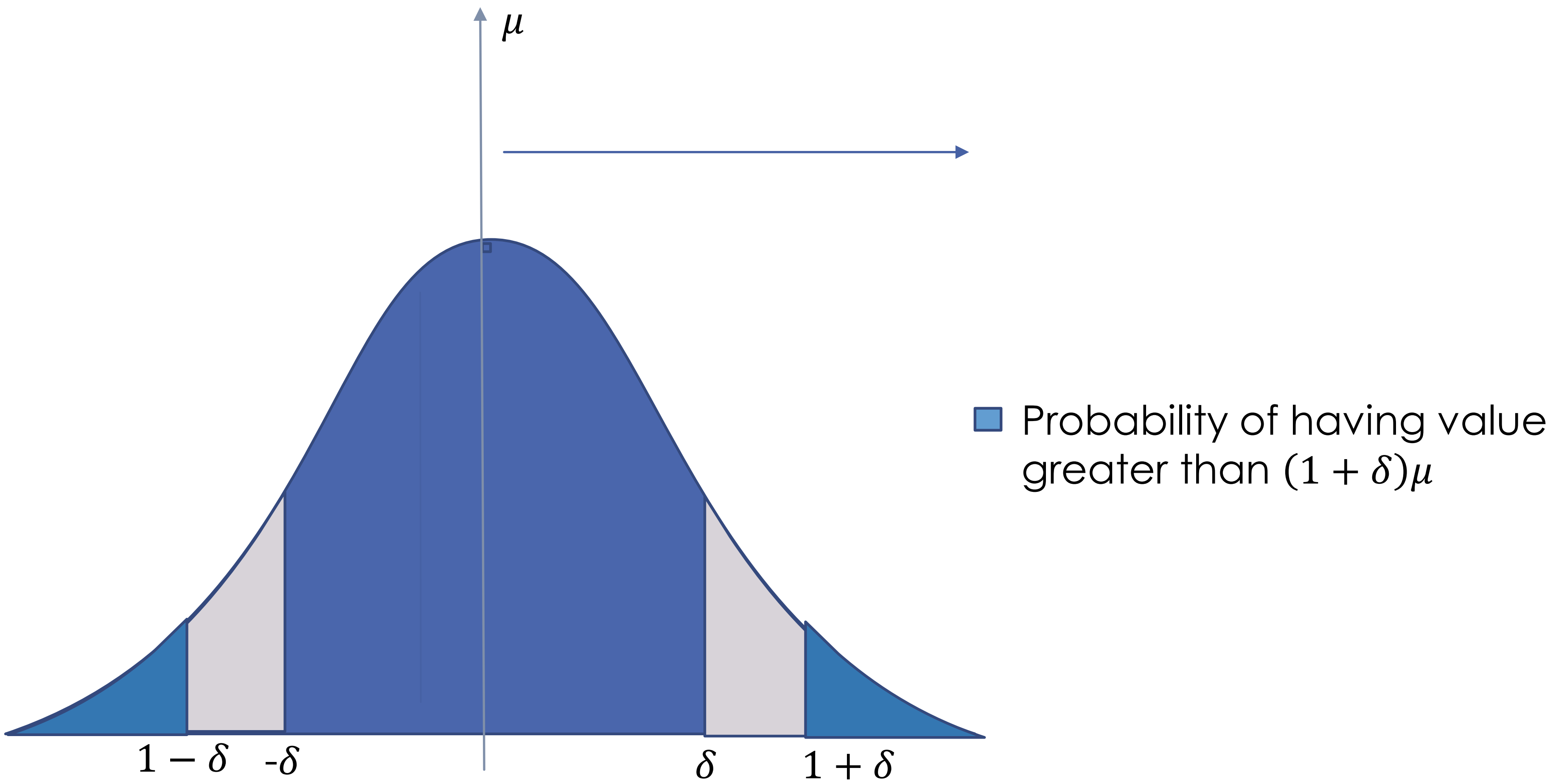
- Let $\boldsymbol{X_1, \ldots, X_n}$ be the independent Poisson trials such that $\mathbb{P}(X_i = 1) = p_i$ and let

$$X = \sum_{i=1}^{n} X_i \text{ and } \mu = \mathbb{E}[X].$$

Then the following Chernoff bounds hold:

1. For any $\delta > 0, \mathbb{P}(X \geq (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$
2. For $0 < \delta \leq 1, \mathbb{P}(X \geq (1+\delta)\mu) < e^{\frac{-\mu\delta^2}{3}}$
3. For $\delta > 0, \mathbb{P}(|X - \mu| \geq \delta) \leq 2e^{\frac{n\delta^2}{4}}$

Illustration of Chernoff Bounds:



The proof for the first form

Chernoff Bounds Contd.

- Let $\boldsymbol{X_1, \ldots, X_n}$ be the independent Poisson trials such that $\mathbb{P}(X_i = 1) = p_i$ and let

$$X = \sum_{i=1}^{n} X_i \text{ and } \mu = \mathbb{E}[X].$$

Then the following Chernoff bounds hold:

1. For any $\delta > 0, \mathbb{P}(X \geq (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$

Moment Generating Function

- To Prove the concept of Chernoff Bounds, we will first need the concept of moment generating functions.
- Let $X_1, X_2 \ldots X_n$ be n Poisson trials, with $\mathbb{P}(X_i = 1) = p_i$

  $\mu = \mathbb{E}[X] = \Sigma_{i=1}^{n} \mathbb{E}[X_i] = \Sigma_{i=1}^{n} p_i$
- Let $M_{X_i}(t) = \mathbb{E}[e^{tX_i}] = p_i(e^t) + (1 - p_i)$

  $= 1 + p_i(e^t - 1)$

  $\leq e^{p_i(e^t - 1)}$ because $1 + x \leq e^x$

this is called the **moment generating function** of X_i

Refer to: "Probability and Computing: Randomized Algorithms and Probabilistic Analysis" Chapter :4 Page: 64

Moment Generating Function: Contd.

- Now, for X, $M_{X(t)} = \Pi_{i=1}^{n} M_{X_i}(t)$

$$\leq \Pi_{i=1}^{n} e^{p_i(e^t - 1)}$$
$$= \exp\{\Sigma_{i=1}^{n}\left(p_i(e^t - 1)\right)$$
$$= e^{(e^t - 1)\mu}$$

- Now we will prove the 1^{st} Chernoff bound, and we leave the other forms as possible exercises.

Proof of Chernoff Bound:

- $\mathbb{P}(X \geq (1+\delta)\mu) = \mathbb{P}(e^{tX} \geq e^{(t(1+\delta)\mu)})$

$$\leq \frac{\mathbb{E}[e^{tx}]}{e^{t(1+\delta)\mu}}$$ By Markov's Inequality

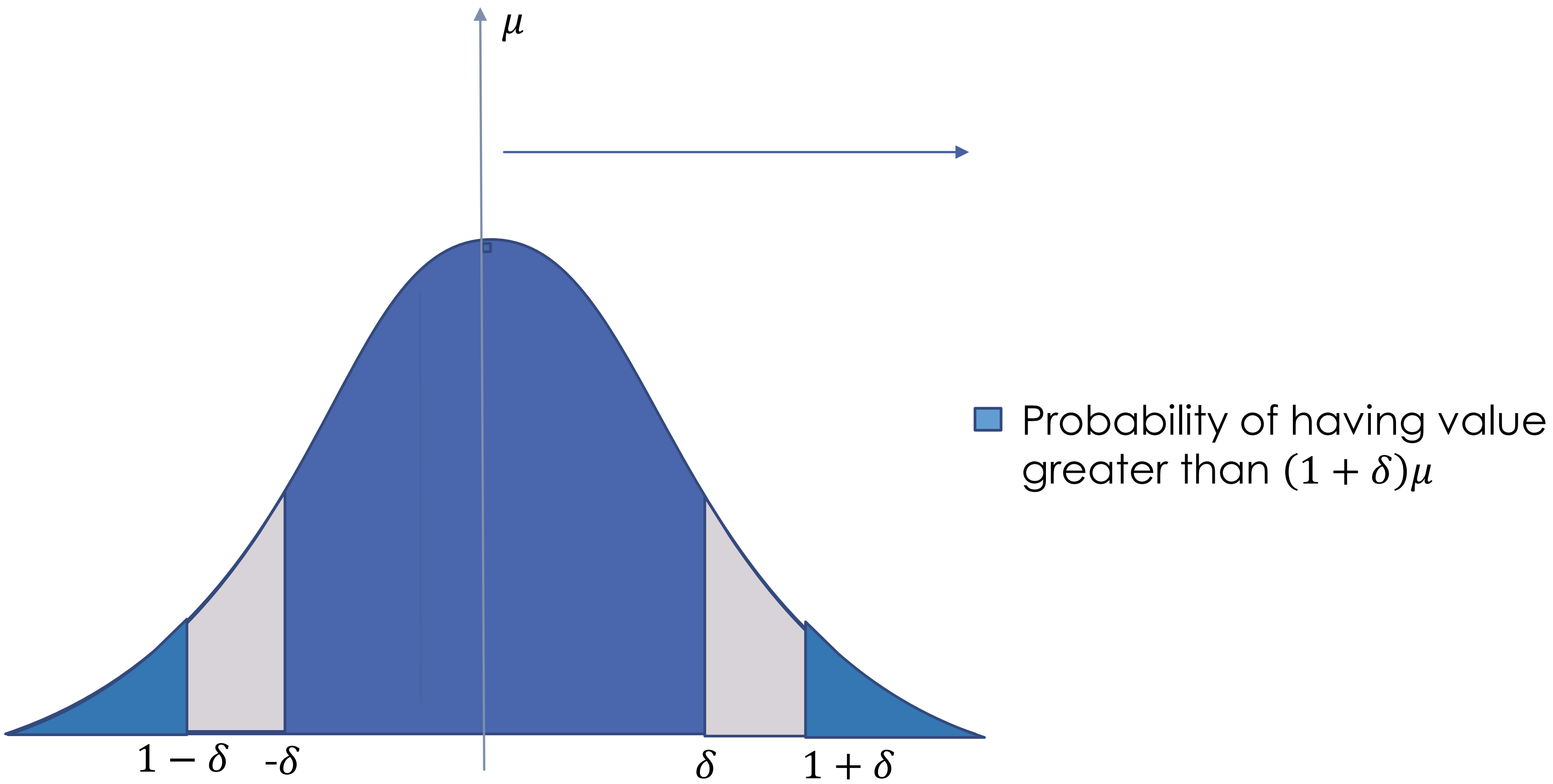
$$\leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}}$$

- Now for $\delta > 0$, set $t = \log(1+\delta)$ and you get

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$$

Refer to: "Probability and Computing: Randomized Algorithms and Probabilistic Analysis" Chapter :4 Page: 66

Illustration of Chernoff Bounds:



Thank You!