Tail Inequalities

- We study three ways to estimate the tail probabilities of random variables.
- It will be noted that, the more information we know about the random variable the better the estimate we can derive about a given tail probability.

Tail Inequalities

- Markov Inequality: If X is a non-negative valued random variable with an expectation of μ, then Pr[X ≥ cμ] ≤ 1/c.
- Proof of Markov inequality:

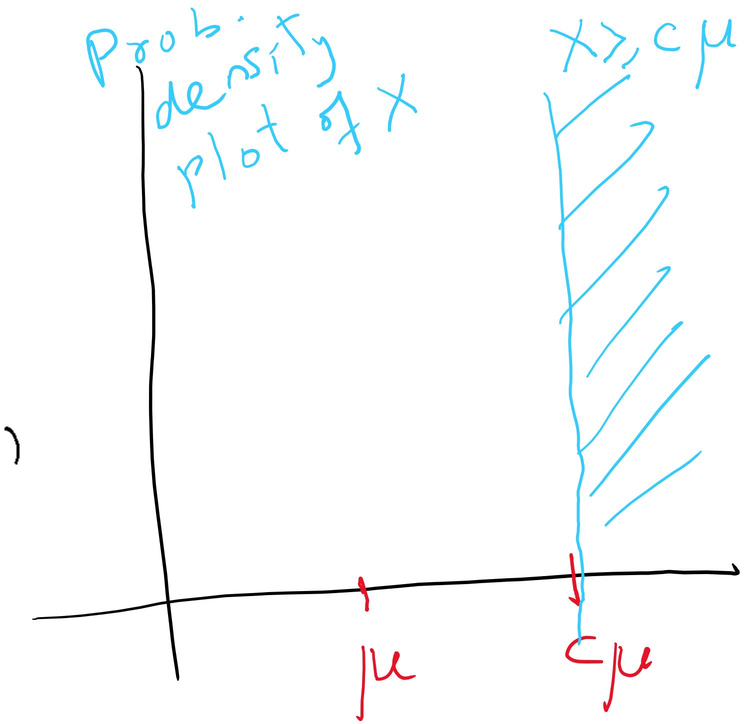
$$\mu = \sum_{a} a \cdot Pr(X = a)$$

$$= \sum_{a < c\mu} a \cdot Pr(X = a) + \sum_{a \geq c\mu} a \cdot Pr(X = a)$$

$$\geq 0 + \sum_{a \geq c\mu} a \cdot Pr(X = a)$$

$$\geq c\mu \cdot \sum_{a \geq c\mu} Pr(X = a)$$

$$= c\mu \cdot Pr(X \geq c\mu)$$



Tail Inequalities

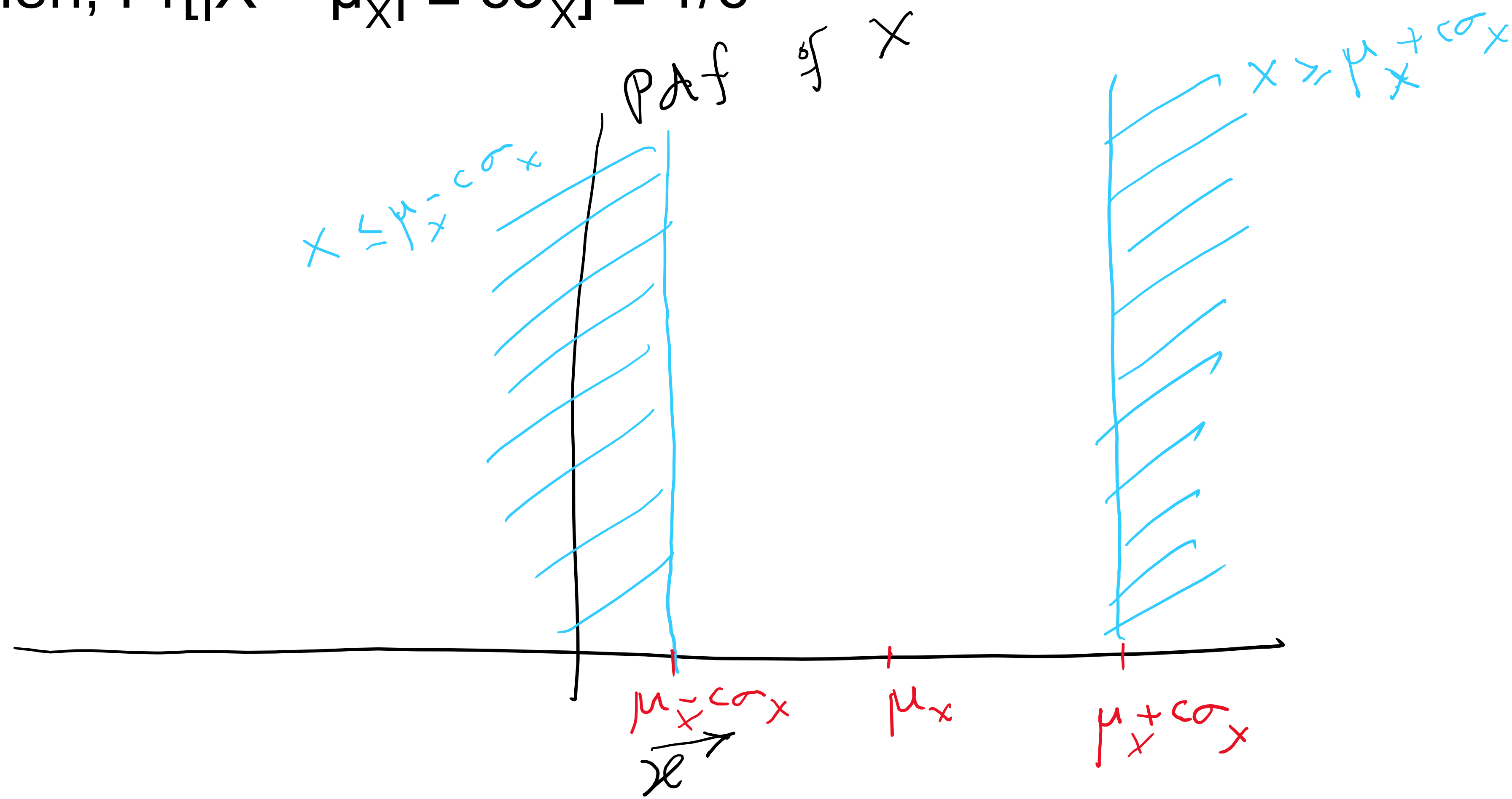
- Markov Inequality: If X is a non-negative valued random variable with an expectation of μ, then Pr[X ≥ cμ] ≤ 1.
- Applying this inequality tells us that the randomized quick sort algorithm has a run time of more than twice its expectation with a probability of 1/2.
- The run time is n^2 with probability of nearly log n /n.

Tail Inequalities

- Chebychev Inequality : We first define the terms standard deviation and variance of a random variable X.
- Let X be a random variable with an expectation of μ. The variance of X, denoted by var(X), is defined as $var(X) = E[(X - \mu)^2]$. The standard deviation of X, denoted by σ_X, is defined as $\sigma_X = var(X)$.
- Note that by definition, $var(x) = E[(X-\mu)^2] = E[X^2-2X\mu+\mu^2] = E[X^2]-\mu^2$.
- The second equality follows from the linearity of expectations.

Tail Inequalities

- Chebychev inequality: Let X be a random variable with expectation μ_X and standard deviation σ_X.
- Then, $\Pr[|X - \mu_X| \geq c\sigma_X] \leq 1/c^2$



Tail Inequalities

- Chebychev inequality: Let X be a random variable with expectation μ_X and standard deviation σ_X.
- Then, $\Pr[|X - \mu_X| \geq c\sigma_X] \leq 1$
- Proof. Let random variable $Y = (X - \mu_X)^2$. Then,

$$E[Y] = E[(X - \mu_X)^2] = \sigma_X^2 \text{ by definition}$$

$$\text{Now, } Pr[|X - \mu_X| \geq c\sigma_X] = Pr[(X - \mu_X)^2 \geq c^2\sigma_X^2]$$

$$= Pr[Y \geq c^2\sigma_X^2].$$

Applying Markov Inequality to the random variable Y

$$Pr[Y \geq c^2\sigma_X^2] = Pr[Y \geq c^2\mu_Y] \leq \frac{1}{c^2}$$

Tail Inequalities

- Better tail inequalitites can be obtained by the powerful technique called Chernoff bounds.
- However, applicability is a little restricted too.
- Let us study the most popular version to start with.

Tail Inequalities

- Let X be a random variable defined as the sum of n independent and identically distributed random variables $X_1, X_2, ..., X_n$.
  - In other words, $X = \Sigma_i X_i$
  - Short form i.i.d.
- Let us assume that each X_i is a Bernoulli random variable.
  - In other words, each X_i takes values in {0, 1}.
- Let $Pr(X_i = 1) = p$ and hence $Pr(X_i = 0) = 1 – p$.
- Finally, let $E[X] = \mu$.
  - Notice that $E[X] = \Sigma_i E[X_i] = n. (1.p + (1-p).0)) = np$.

Tail Inequalities

- Several settings relate to the above statements.
- Consider throwing a (biased) coin over n trials.
- Each trial, the probability of Heads is p.
- So, each X_i corresponds to the fact that the ith trial results in a Heads.
- Let us count the number of Heads over the n trails. Indeed, $X = \Sigma_i X_i$ captures this count as a random variable.
- Note that the expected number of Heads over n trails is exactly np.

Tail Inequalities

- Finally, to the theorem.
- Given the earlier conditions, it holds that for any $\delta > 0$,

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

- The normal strategy employed to prove tail estimates of sums of independent random variables is to make use of exponential moments.
- While proving Chebychev inequality, we made use of the second-order moment. It can be observed that using higher order moments would generally improve the bound on the tail inequality.
- But using exponential moments would result in a vast improvement.

Tail Inequalities

- Given the earlier conditions, it holds that for any $\delta > 0$,

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

- While proving Chebychev inequality, we made use of the second-order moment of a random variable.
- It can be observed that using higher order moments would generally improve the bound on the tail inequality
- But using exponential moments would result in a vast improvement in the bound.
- An exponential moment of a random variable X is the expectation of functions of X such as e^X.

Tail Inequalities

- Given the earlier conditions, it holds that for any $\delta > 0$,

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

- For each i, we define the random variable $Y_i = e^{tX_i}$ for a real number $t > 0$ that will be chosen later.
- Notice that

1) Y_i is a positive valued random variable.

2) $$E[Y_i] = E[e^{tX_i}] = p_i \cdot e^{t} + (1-p_i) \cdot e^{0}$$
$$= p_i e^{t} + 1 - p_i$$

Tail Inequalities

- Given the earlier conditions, it holds that for any $\delta > 0$,

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

- Now define another random variable $Y = Y_1.Y_2....Y_n$.
- Now, we can note that

$$E[Y] = E[Y_1 \cdot Y_2 \cdots Y_n]$$
$$= \prod_{i=1}^{n} E[Y_i] = (p_i \cdot e^t + 1 - p_i)^n$$

- Why does the above calculation hold?

Tail Inequalities

- Given the earlier conditions, it holds that for any $\delta > 0$,

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

- The next step we do is do apply Markov inequality on Y as follows.
- First, notice that $Y = e^{tX}$ as

$$Y = Y_1 \cdot Y_2 \cdots Y_n = e^{tX_1} \cdot e^{tX_2} \cdots e^{tX_n}$$

$$= e^{t(X_1 + X_2 + \cdots + X_n)} = e^{tX}$$

Tail Inequalities

- Given the earlier conditions, it holds that for any $\delta > 0$,

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

- First, notice that $Y = e^{tX}$. And,

- Further, $X \geq \mu(1+\delta) \Leftrightarrow e^{tX} \geq e^{t\mu(1+\delta)}$

$$\Rightarrow \quad Y \geq e^{t\mu(1+\delta)}$$

- So, we are interested in the event $Y \geq e^{t\mu(1+\delta)}$.

Tail Inequalities

- Given the earlier conditions, it holds that for any $\delta > 0$,

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

- So, we are interested in the event $Y \geq t\mu(1+\delta)$. We proceed as:

$$\Pr(Y \geq e^{t\mu(1+\delta)}) \leq \frac{E[Y]}{e^{t\mu(1+\delta)}} = \frac{\prod(1-p_i+p_i e^t)}{e^{t\mu(1+\delta)}}$$

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

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

Tail Inequalities

- Given the earlier conditions, it holds that for any $\delta > 0$,

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

- So, we are interested in the event $Y \geq t\mu(1+\delta)$. We proceed as:

$$\Pr(Y \geq e^{t\mu(1+\delta)}) \leq \frac{E[Y]}{e^{t\mu(1+\delta)}} = \frac{\prod (1-p_i+p_i e^t)}{e^{t\mu(1+\delta)}}$$

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

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

Tail Inequalities

- Given the earlier conditions, it holds that for any $\delta > 0$,

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

- So, $\Pr\left(Y \geq e^{t\mu(1+\delta)}\right) = e^{-\mu(1-e^{t}) - t\mu(1+\delta)}$
- Since t is a free parameter in the above, we can find a t that minimizes the right hand side.
- To simplify, let $f(t) = \ln e^{-\mu(1-e^{t}) - t\mu(1+\delta)}$

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

Tail Inequalities

- Given the earlier conditions, it holds that for any $\delta > 0$,

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

- To simplify, let $f(t) = \ln e^{-\mu(1-e^{t}) - t\mu(1+\delta)}$

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

- Differentiating f(t) wrt t, we get $f'(t) = \mu e^{t} - \mu(1+\delta)$

- So, f'(t) = 0 at $t = \ln(1+\delta)$

- Verify that the above t corresponds to a minima. OFFLINE.

Tail Inequalities

- Given the earlier conditions, it holds that for any δ > 0,

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

- With t = ln(1+δ), we get that

$$\Pr(X \geq \mu(1+\delta)) \leq \frac{e^{-\mu(1-(1+\delta))}}{(1+\delta)^{\mu(1+\delta)}} = \frac{e^{\mu\delta}}{(1+\delta)^{1+\delta}}$$

$$= \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

- completing the proof.