Analysis of Nakamoto Consensus

Ling Ren

University of Illinois Urbana-Champaign
renling@illinois.edu

1 Introduction

The famed Bitcoin white paper presented an unconventional (at the time) Byzantine fault tolerant consensus algorithm that is now known as the Nakamoto consensus [6]. Nakamoto consensus centers around the proof-of-work (PoW) mechanism and the "longest-chain-win" rule. It is extremely simple and can be described very succinctly: at any time, an honest node adopts the longest PoW chain to its knowledge and attempts to mine a new block that extends this longest chain; a block is committed when buried sufficiently deep in the chain. Such a simple algorithm deserves a simple analysis, which is what this paper aims to provide.

2 Model and Overview

We assume the readers are familiar with how Nakamoto consensus works and we review its basics only to introduce notations. Transactions in Nakamoto consensus are batched into blocks. Each block is linked to a unique predecessor block via PoW, thus forming a PoW chain. A block's height is its predecessor block's height plus one. Upon adopting a new longest chain, either through mining or by receiving from other nodes, a node broadcasts and mines on top of the new longest chain. Ties are broken arbitrarily. An honest node considers a block B committed if B is buried at least k blocks deep in its adopted chain.

We will prove that Nakamoto consensus guarantees safety and liveness with overwhelming probability.

- **Safety**. Honest nodes do not commit different blocks at the same height.
- **Liveness** Every transaction is eventually committed by all honest nodes.

The liveness property of Nakamoto consensus is phrased as chain growth and chain quality in recent literature [2, 7]. We use the conventional liveness definition since it separates the problem formulation (state machine replication) from the specific solution (Nakamoto consensus).

Assumptions. We assume PoW mining are modeled by homogeneous Poisson point processes. Let α and β denote the collective mining rate of honest nodes and malicious nodes, respectively. We assume the network delay between any pair of honest nodes is upper bounded by Δ. If two honest nodes are connected via multiple hops, Δ is an upper bound on the end-to-end network delay between them.

Main theorem (informal). *Let $g = e^{-\alpha\Delta}$. Let δ be any positive constant. Nakamoto consensus with the k-confirmation commit rule guarantees safety and liveness except for $e^{-\Omega(\delta^2 g^2 k)}$ probability if*

$$g^2\alpha > (1+\delta)\beta.$$

The above condition has a clear and intuitive interpretation: it is the "honest majority" condition after taking network delay into account. Malicious nodes can coordinate their actions (mining strategies) perfectly with zero network delay, so they can extend a chain at their collective mining rate β. Between honest nodes, however, the (bounded) network delay will lead to a (bounded) loss in their collective mining rate. The crux of the proof is to bound the honest mining rate loss. We will show that the loss is at most g for liveness and at most g^2 for safety. We assume $0 < \delta < 1$ throughout the paper unless otherwise stated.

Δ	upper bound on network delay
α	collective honest mining rate
β	collective malicious mining rate
g	$= e^{-\alpha\Delta}$, a discount factor of honest mining rate due to network delay
k	number of confirmations needed to commit

3 Proofs

3.1 Preliminary

In a Poisson process with rate λ, the number of event arrivals (blocks mined) in a time interval (t_1, t_2) is independent of other non-overlapping internals and follows a Poisson distribution with rate $\lambda(t_2 - t_1)$. The interarrival times in a Poisson process follow independent exponential distributions with the same rate parameter λ, whose cumulative distribution function is $\Pr[T \leq t] = 1 - e^{-\lambda t}$.

The following tail bounds will be frequently used. The Chernoff bound is well known. We give its proof in the appendix for completeness. The Poisson tail bound and proof are almost identical to the Chernoff bound. This is as expected because the Poisson distribution is a limiting case of the binomial distribution.

Lemma 1 (Chernoff)**.** *Let $X = \sum_{i=1}^{n} X_i$ be the sum of n independent Boolean random variables and μ be the expectation of X. For $0 < \delta < 1$, $\Pr[X \leq (1-\delta)\mu] < e^{-\frac{\delta^2\mu}{2}}$. For $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] < e^{-\frac{\delta^2\mu}{2+\delta}}$.*

Lemma 2 (Poisson tail)**.** *Let X be a Poisson random variable with rate μ (which is also its expectation). For $0 < \delta < 1$, $\Pr[X \leq (1-\delta)\mu] < e^{-\frac{\delta^2\mu}{2}}$. For $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] < e^{-\frac{\delta^2\mu}{2+\delta}}$.*

3.2 Non-tailgaters and Loners

We call a block mined by an honest (resp. malicious) node an honest (resp. malicious) block. Let us put all honest blocks on a time axis based on when they are mined.

Definition 3 (Non-tailgaters and loners)**.** *Suppose an honest block B is mined at time t. If no other honest block is mined between time $t - \Delta$ and t, then B is a non-tailgater (otherwise, B is a tailgater). If no other honest block is mined between time $t - \Delta$ and $t + \Delta$, then B is a loner.*

In other words, a *non-tailgater* is an honest block that is mined at least Δ time after the previous honest block. A *loner* (called a "convergence opportunity" in [7]) is an honest block that does not tailgate and is not tailgated. We emphasize that these notions apply to honest blocks only. An honest block is potentially "wasted" if it tailgates. On the other hand, the next lemma establishes the key property that makes non-tailgaters and loners contribute to liveness and safety.

Lemma 4. *(i) Non-tailgaters have different heights. (ii) A loner is the only honest block at its height.*

Proof. It suffices to show that if two honest blocks do not tailgate one another, then they have different heights. Let the two blocks be B and B'. Without loss of generality, assume B is mined first. B reaches all honest nodes within Δ time, which is before B' is mined (otherwise B' tailgates B). Upon receiving B, an honest node will attempt to extend B and will only mine at a height greater than B. □

The next two lemmas show that there are abundant non-tailgaters and loners, both in raw numbers and relative to malicious blocks. While the conclusions of these two lemmas are intuitive, their proofs turn out to be quite subtle. Readers seeking a proof sketch may skip their proofs if content with the intuition below.

We can easily show that each honest block has a probability $g = e^{-\alpha\Delta}$ to be a non-tailgater and a probability g^2 to be a loner. Lemma 6 states that the number of non-tailgaters and loners in an interval of duration t cannot be much smaller than their respective expectations of $g\alpha t$ and $g^2\alpha t$. Lemma 7 states that, with $g^2\alpha > (1+\delta)\beta$, there will be more loners than malicious blocks in all time intervals covering a sufficiently long interval $(r, s]$. Lemma 7 is first proved in [5] and we give an alternative proof.

Definition 5. *For any time interval $(r, s]$, define $X_{r,s}$, $Y_{r,s}$, and $Z_{r,s}$ to be the number of non-tailgaters, loners, and malicious blocks mined in the interval, respectively.*

Lemma 6. *For any time interval $(r, r+t]$, (i) $\Pr[X_{r,r+t} \leq (1-\delta)g\alpha t] < e^{-\Omega(\delta^2 g\alpha t)}$, and (ii) $\Pr[Y_{r,r+t} \leq (1-\delta)g^2\alpha t] < e^{-\Omega(\delta^2 g^2\alpha t)}$.*

Proof. Pick $\delta_1 \in (\frac{\delta}{3}, \frac{2\delta}{3})$ such that $n = (1-\delta_1)\alpha t$ is an integer. For sufficiently large t, such an δ_1 exists. Consider the first $n+1$ honest blocks mined since the beginning of the interval r (but not necessarily before the end of interval $r+t$). Number these blocks $1, 2, \ldots, n+1$. Define block 0 to be the last honest block mined prior to the interval. Let $X_i = 1$ if block i is a non-tailgater, and 0 otherwise. Let $Y_i = 1$ if block i is a loner, and 0 otherwise. Define $X = \sum_{i=1}^{n} X_i$ and $Y = \sum_{i=1}^{n} Y_i$.

Let A_X be the event that $X > (1-\delta)g\alpha t$. Recall that interarrival times in a Poisson process follow i.i.d. exponential distributions with the same rate parameter. Thus, X_i's are i.i.d. and $\Pr[X_i = 1] = e^{-\alpha\Delta} = g$. Let $\delta_2 = \delta - \delta_1$; we have $\delta_2 \in (\frac{\delta}{3}, \frac{2\delta}{3})$ and $(1-\delta_2)(1-\delta_1) > (1-\delta)$. Then, by Chernoff bound.

$$\Pr[\overline{A_X}] = \Pr[X \leq (1-\delta)g\alpha t] \leq \Pr[X \leq (1-\delta_2)gn] < e^{-\delta_2^2 gn/2} = e^{-\Omega(\delta^2 g\alpha t)}.$$

Let A_0 be the event that more than n honest blocks are mined in the interval. By Lemma 2, $\Pr[\overline{A_0}] < e^{-\Omega(\delta^2\alpha t)}$. $A_0 \cap A_X$ is the event that "more than n honest blocks are mined in the interval *and* more than $(1-\delta)g\alpha t$ among the first n blocks mined are non-tailgaters", which implies $X_{r,r+t} > (1-\delta)g\alpha t$. Thus,

$$\Pr[X_{r,r+t} \leq (1-\delta)g\alpha t] \leq \Pr[\overline{A_0} \cup \overline{A_X}] < e^{-\Omega(\delta^2 g\alpha t)}.$$

Let A_Y be the event that $Y > (1-\delta)g^2\alpha t$. We have $Y_i = X_i X_{i+1}$, so $\Pr[Y_i = 1] = g^2$. But Y_i and Y_{i+1} are dependent (both depend on X_{i+1}), so we cannot directly invoke Chernoff bounds. Luckily, Y_i and Y_{i+2} are independent. Thus, Y can be broken up into two summations of independent Boolean random variables $Y = \sum_{\text{odd}} Y_i + \sum_{\text{even}} Y_i$. Applying Chernoff bounds on the two summations separately yields $\Pr[\overline{A_Y}] = \Pr[Y \leq (1-\delta)g^2\alpha t] < e^{-\Omega(\delta^2 g^2\alpha t)}$. The rest of the proof for part (ii) is identical to part (i). □

Lemma 7. *Suppose $g^2\alpha > (1+\delta)\beta$. For two points in time $r < s$, define $G_{r,s}$ to be the following event: for all $u \leq r$ and $v \geq s$, $Y_{u+\Delta,v-\Delta} > Z_{u,v}$. Then, $\Pr[\overline{G_{r,s}}] < e^{-\Omega(\delta^2 g^2\alpha(s-r))}$.*

Proof. Let $w = (r+s)/2$. We define two events: $G^-_{r,s}$ is the event that for all $u \leq r$, $Y_{u-\Delta,w} > Z_{u,w}$, and $G^+_{r,s}$ is the event that for all $v \geq s$, $Y_{w,v-\Delta} > Z_{w,v}$. Since $Y_{u+\Delta,v-\Delta} = Y_{u+\Delta,w} + Y_{w,v-\Delta}$ and $Z_{u,v} = Z_{u,w} + Z_{w,v}$, $G^-_{r,s} \cap G^+_{r,s}$ implies $G_{r,s}$. Hence, $\Pr[\overline{G_{r,s}}] \leq \Pr[\overline{G^-_{r,s}}] + \Pr[\overline{G^+_{r,s}}]$. Due to the symmetry over the mid-point w, it suffices to bound $\Pr[\overline{G^+_{r,s}}]$.

Pick $\delta_1 = \delta/4$. Let S_i be the time the i-th malicious block after w is mined. Let $s_i = w + \frac{i}{(1-\delta_1)g^2\alpha} + \Delta$. Let F_i be the event that $S_i \geq s_i$ and $Y_{w,s_i-\Delta} > i$, i.e., since time w, the i-th malicious block is mined after s_i and more than i loners have been mined by $s_i - \Delta$. Let $y_0 = (1-\delta_1)g^2\alpha(s-w-\Delta)$. Note that $y_0 = \Omega(g^2\alpha(s-r))$. Let A_Y be the event that $Y_{w,s-\Delta} > y_0$. Observe that $\left(\bigcap_{i>y_0} F_i\right) \cap A_Y$ implies $G^+_{r,s}$. (There is no need to include F_i for smaller i because event A_Y ensures that more than y_0 loners are mined by time $s-\Delta$.) Thus, $\overline{G^+_{r,s}}$ implies $\left(\bigcup_{i>y_0} \overline{F_i}\right) \cup \overline{A_Z}$ and $\Pr[\overline{G^+_{r,s}}] \leq \Pr[\overline{A_Z}] + \sum_{i>y_0} \Pr[\overline{F_i}]$.

$\overline{F_i}$ is the event that either $S_i < s_i$ or $Y_{w,s_i-\Delta} \leq i$. The former means that at least i malicious blocks are mined in $\frac{i}{(1-\delta_1)g^2\alpha} + \Delta$ time. When $s-w$ is sufficiently large (a sufficient condition is $s - w > 2\Delta + \frac{\Delta}{\delta}$), for all $i > y_0$, the above duration is shorter than $\frac{i}{(1+\delta_1)\beta}$, and we have $\Pr[S_i < s_i] < e^{-\Omega(\delta^2 i)}$ by Lemma 2. The latter means that at most i loners are mined in $\frac{i}{(1-\delta_1)g^2\alpha}$ time, which happens with $e^{-\Omega(\delta^2 i)}$ probability by Lemma 6(ii). Thus, when $s-w$ is sufficiently large, for all $i > y_0$, $\Pr[\overline{F_i}] < e^{-\Omega(\delta^2 i)}$. Summing over i, $\sum_{i>y_0} \Pr[\overline{F_i}] < e^{-\Omega(\delta^2 y_0)}$. In addition, $\Pr[\overline{A_Y}] < e^{-\Omega(\delta^2 y_0)}$ by Lemma 6(ii). Plugging them into $\Pr[\overline{G^+_{r,s}}]$ yields $\Pr[\overline{G_{r,s}}] \leq 2\Pr[\overline{G^+_{r,s}}] < e^{-\Omega(\delta^2 y_0)} = e^{-\Omega(\delta^2 g^2\alpha(s-r))}$. □

3.3 Liveness

Theorem 8 (Liveness)**.** *Suppose $g\alpha > (1+\delta)\beta$. At time t, except for $e^{-\Omega(\delta^2 g\alpha t)}$ probability, every honest node commits at least $\frac{\delta}{6}g\alpha t - k - 1$ honest blocks.*

Proof. At time t, each honest node receives at least $X_{0,t-\Delta}$ non-tailgaters, so its chain is at least that long by Lemma 4(i). Its chain contains at most $Z_{0,t}$ malicious blocks. The last k blocks are not committed. Further note that $X_{0,t} \le X_{0,t-\Delta} + 1$ because there can be at most one non-tailgater within Δ time by definition. Thus, every honest node commits at least $L_t \ge X_{0,t} - Z_{0,t} - k - 1$ honest blocks at time t.

Pick $\delta_1 = \delta/3$. By Lemma 4(i) and 6(i), $X_{0,t} > (1 - \frac{\delta_1}{2})g\alpha t$ except for $e^{-\Omega(\delta^2 g\alpha t)}$ probability. By Lemma 2, $Z_{0,t} < (1-\delta_1)g\alpha t$ except for $e^{-\frac{\varepsilon^2\beta t}{2+\varepsilon}}$ probability where $\varepsilon = \frac{(1-\delta_1)g\alpha}{\beta} - 1$. The exponent $\frac{\varepsilon^2\beta t}{2+\varepsilon}$ can be shown to be $\Omega(\delta^2 g\alpha t)$. Therefore, $L_t > \frac{\delta_1}{2}g\alpha t - k - 1 = \frac{\delta}{6}g\alpha t - k - 1$, except for $e^{-\Omega(\delta^2 g\alpha t)}$ probability. □

3.4 Safety

Recall that with the k-confirmation commit rule, at any time t, a block B is considered committed by an honest node if B is buried at least k blocks deep in that node's adopted chain.

Theorem 9 (Safety)**.** *Suppose $g^2\alpha > (1+\delta)\beta$. Consider any time t and any block B that is considered committed by some honst node at time t. Except for $e^{-\Omega(\delta^2 g^2 k)}$ probability, for all time $t' \ge t$, no honest node commits a block $B' \ne B$ at the height of B.*

Proof. Let T_1 be the first time since t that some honest node commits such a block B'. Thus, right before T_1 (or at T_1 if $T_1 = t$), an honest node (potentially the same one) adopts a chain that extends B. Let these two diverging chains end at B_1 and B_1' respectively. Let block B_0' be the last common ancestor and B_0 be the last *honest* common ancestor of the two diverging chains. Note that if B_0' is mined by an honest node, then B_0 and B_0' are the same block; otherwise, the blocks between B_0 (excluded) and B_0' (included) are all malicious blocks. Let H_0, H_0', H_1, H_1' be the height of block B_0, B_0', B_1, B_1', respectively. Let T_0 be the time B_0 is mined. Without loss of generality, assume $H_1 \le H_1'$. Figure 1 illustrate the scenario. Note that T_0 and T_1 are random variables (as they depend on random mining outcomes), so the proof cannot invoke properties of Poisson for intervals defined by them.

Lemma 10. $Z_{t_0,t_1} \ge Y_{t_0+\Delta,t_1-\Delta}$.

Proof. We omit the subscripts and write the two sets as Z and Y for short. First, we show that every loner $y \in Y$ has height $H_y \in (H_0, H_1]$. By time $T_0 + \Delta$, all honest nodes have received B_0 with height H_0 and will never again mine on height H_0 or lower. If any $y \in Y$ has height $H_y > H_1$, then by time T_1, it arrives at all honest nodes and no honest node will adopt a chain ending at B_1 with height H_1.

Each $y \in Y$ can now be paired with a distinct malicious block in Z as follows. If y has height $H_y \in (H_0, H_0']$, then it is paired with the height-H_y block that buries B_0, which is a malicious block by the definition of B_0. If y has height $H_y \in (H_0', H_1]$, then it is paired with the height-H_y block on one of the two diverging chains, which must be a malicious block because a loner does not share height with other honest blocks due to Lemma 4(ii). In either case, the paired malicious block belongs to Z (i.e., mined during $(T_0, T_1]$) because it extends B_0 (mined at T_0) and is known to some honest node by time T_1. Lastly, these malicious blocks have distinct heights (because the loners have distinct heights) and are thus distinct. □

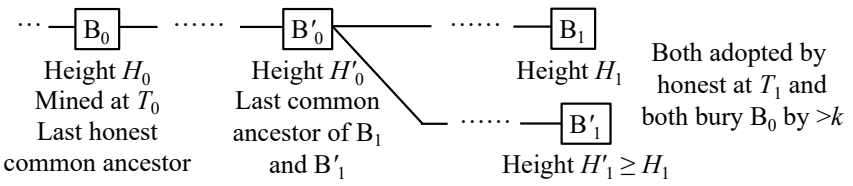

Figure 1: The scenario analyzed in the proof of Theorem 9, which is similar to the one studied in [2].

Recall from Lemma 7 that event $G_{t-\frac{k}{2\alpha},t}$ contradicts with $Z_{T_0,T_1} \geq Y_{T_0+\Delta,T_1-\Delta}$ if $T_0 \leq t - \frac{k}{2\alpha}$. Thus, for such a safety violation to happen, either $G_{t-\frac{k}{2\alpha},t}$ does not occur or $T_0 > t - \frac{k}{2\alpha}$. By Lemma 7, $\Pr[\overline{G_{t-\frac{k}{2\alpha},t}}] < e^{-\Omega(\delta^2 g^2 \alpha \frac{k}{2\alpha})} = e^{-\Omega(\delta^2 g^2 k)}$. Since B extends B_0 and is considered committed by some honest node at time t, at least k blocks are mined during $(T_0, t]$. On the other hand, since the combined mining rate of honest and malicious nodes is $\alpha + \beta < \alpha(1 + \frac{1}{1+\delta})$, by Lemma 2, fewer than k blocks are mined during $(t - \frac{k}{2\alpha}, t]$ except for $e^{-\Omega(\delta^2 k)}$ probability. Hence, $\Pr[T_0 > t - \frac{k}{2\alpha}] < e^{-\Omega(\delta^2 k)}$. Therefore, such a safety violation happens with $e^{-\Omega(\delta^2 g^2 k)}$ probability. □

4 Remarks

Related work. Garay et al. [2] show that Nakamoto consensus works under the "honest majority" assumption in the lock-step synchrony model which abstracts away network delays. Pass et al. [7, 8] extended the analysis to non-lock-step synchrony. Translated into our terminology, they show that Nakamoto consensus works if $(2g - 1)\alpha > (1 + \delta)\beta$. When $g \approx 1$, $2g - 1 \approx g^2$; but when g is noticeably smaller than 1 (implying that the network delay cannot be ignored), their result is looser than ours. The reason why their result is looser is that they undercount loners (called convergence opportunities in their papers). Kiffer et al. [4] gave a method to count loners tightly using Markov chains. Aside from the extra complexity from advanced tools, their result has a non-closed form that involves three levels of nested summations. A concurrent work Zhao [9] shows that the nested summation, in fact, equals g^2. The main source of simplification in this paper is a much simpler way to count loners tightly.

Modeling PoW mining as Poisson processes is by no means new, as it is used in Nakamoto's original paper [6]. Other papers analyzing Nakamoto consensus [7, 8, 4] assume mining happens in discrete rounds. We prefer the continuous model because the discrete rounds seem artificial and various quantities in the continuous world take on simpler forms than their discrete counterparts.

Several papers referred to the non-lock-step synchrony model as "asynchroy" [7, 9] or "partial synchrony" [4]. We strongly advise against it because those two terms are well established in the literature and they describe much weaker models that Nakamoto consensus cannot handle. For example, the partial synchrony model assumes the network delay bound Δ is "unknown", meaning that the adversary can choose Δ after all the other protocol parameters are fixed [1]. Then, it can break the safety of Nakamoto consensus even with zero mining rate: pick a very large Δ to keep the network partitioned for sufficiently long until two forks of length k are created. This is reflected in our result: the error probability contains g in the exponent, so if the adversary can pick Δ after k is fixed, it can make the error probability unacceptable.

Open problems. Whether the safety condition $g^2\alpha > (1 + \delta)\beta$ is tight remains open. The best attack we know is still the simplest "private chain" attack described by Nakamoto [6] and it requires at least $\beta > g\alpha$.

Another aspect that is quite loose in all existing analysis is the precise error probability of the k-confirmation commit rule. Because of this, we still do not have a theoretically sound way to recommend the k needed for a target error probability in Nakamoto consensus.

We have assumed that there is no mining difficulty adjustment. Garay et al. has extended their lock-step analysis [2] to variable difficulty [3]. Such an analysis under non-lock-step synchrony remains open.

Acknowledgement The author is grateful to Chen Feng, Dongning Guo, and Jing Li who pointed out subtle flaws in previous versions of this paper and suggested fixes. The author thanks Jiantao Jiao, Kartik Nayak, and Elaine Shi, and David Tse for helpful discussions and feedback.

References

[1] Cynthia Dwork, Nancy Lynch, and Larry Stockmeyer. Consensus in the presence of partial synchrony. *Journal of the ACM*, 35(2):288–323, 1988.

[2] Juan Garay, Aggelos Kiayias, and Nikos Leonardos. The Bitcoin backbone protocol: Analysis and applications. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 281–310. Springer, 2015.

[3] Juan Garay, Aggelos Kiayias, and Nikos Leonardos. The Bitcoin backbone protocol with chains of variable difficulty. In *Annual International Cryptology Conference*, pages 291–323. Springer, 2017.

[4] Lucianna Kiffer, Rajmohan Rajaraman, et al. A better method to analyze blockchain consistency. In *ACM SIGSAC Conference on Computer and Communications Security*, pages 729–744. ACM, 2018.

[5] Jing Li and Dongning Guo. Continuous-time analysis of the bitcoin and prism backbone protocols. *arXiv preprint arXiv:2001.05644*, 2020.

[6] Satoshi Nakamoto et al. Bitcoin: A peer-to-peer electronic cash system, 2008.

[7] Rafael Pass, Lior Seeman, and Abhi Shelat. Analysis of the blockchain protocol in asynchronous networks. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 643–673. Springer, 2017.

[8] Rafael Pass and Elaine Shi. Rethinking large-scale consensus. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 115–129. IEEE, 2017.

[9] Jun Zhao. An analysis of blockchain consistency in asynchronous networks: Deriving a neat bound. *arXiv preprint 1909.06587*, 2019.

A Proofs of Tail Bounds

Lemma 1 (Chernoff)**.** *Let $X = \sum_{i=1}^{n} X_i$ be the sum of n independent Boolean random variables and μ be the expectation of X. For $0 < \delta < 1$, $\Pr[X \leq (1-\delta)\mu] < e^{-\frac{\delta^2\mu}{2}}$. For $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] < e^{-\frac{\delta^2\mu}{2+\delta}}$.*

Proof. We show the $X \geq (1+\delta)\mu$ side. The other side is similar. First, by Markov Inequality, for all $t > 0$,

$$\Pr[X \geq (1+\delta)\mu] = \Pr[e^{tX} \geq e^{t(1+\delta)\mu}] \leq \frac{E[e^{tX}]}{e^{t(1+\delta)\mu}}. \tag{1}$$

For Bernoulli random variable X_i, let $p_i = \Pr[X_i = 1] = E[X_i]$. Then, $E[e^{tX_i}] = p_i e^t + 1 - p_i = 1 + (e^t - 1)p_i < e^{(e^t-1)p_i}$ where the last step uses the inequality $1 + x < e^x$. Thus, we have

$$E[e^{tX}] = E[\Pi_i e^{tX_i}] = \Pi_i E[e^{tX_i}] < \Pi_i e^{(e^t-1)p_i} = e^{(e^t-1)\sum_i p_i} = e^{(e^t-1)\mu},$$

where the second equality is by independence. Plugging into (1) gives

$$\Pr[X > (1+\delta)\mu] \leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}}.$$

Setting $t = \ln(1+\delta)$, the right-hand side becomes $e^{[\delta-(1+\delta)\ln(1+\delta)]\mu}$. Lastly, it is not hard to show that $\ln(1+\delta) > \frac{2\delta}{\delta+2}$ and hence $(1+\delta)\ln(1+\delta) > \delta + \frac{\delta^2}{2+\delta}$, which completes the proof of the upper tail bound. The other side is similar. □

Lemma 2 (Poisson tail)**.** *Let X be a Poisson random variable with rate μ (which is also its expectation). For $0 < \delta < 1$, $\Pr[X \leq (1-\delta)\mu] < e^{-\frac{\delta^2\mu}{2}}$. For $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] < e^{-\frac{\delta^2\mu}{2+\delta}}$.*

Proof. Recall that the moment generating function of a Poisson random variable is $E[e^{tX}] = e^{(e^t-1)\mu}$. By Markov Inequality, we have, for all $t > 0$,

$$\Pr[X \geq (1+\delta)\mu] = \Pr[e^{tX} \geq e^{t(1+\delta)\mu}] \leq \frac{E[e^{tX}]}{e^{t(1+\delta)\mu}} = \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}}.$$

The rest of the proof is identical to that of the Chernoff bound. □