kurtosis κ. A reasonable estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^{n} X_t^2 \,. \tag{2}$$

Clearly this estimator is unbiased and has variance

$$\mathbf{V}[\hat{\sigma}^2] = \frac{\mathbf{E}[X^4] - \mathbf{E}[X^2]^2}{n} = \frac{\sigma^4 (\kappa - 1)}{n} \,.$$

Therefore, if we are to expect good estimation of σ^2, then the kurtosis should be finite. Note that if σ^2 is estimated by (2), then the central limit theorem combined with finite kurtosis is enough for an estimation error of $O(\sigma^2((\kappa - 1)/n)^{1/2})$ *asymptotically*. For bandits, however, finite-time bounds are required, which are not available using (2) without additional moment assumptions (for example, on the moment generating function). An example demonstrating the necessity of the limit in the standard central limit theorem is as follows. Suppose that $X_1, \ldots, X_n$ are Bernoulli with bias $p = 1/n$, then for large n the distribution of the sum is closely approximated by a Poisson distribution with parameter 1, which is very different to a Gaussian. Finite kurtosis alone *is* enough if the classical empirical estimator is replaced by a robust estimator such as the median-of-means estimator [Alon et al., 1996] or Catoni's estimator [Catoni, 2012]. Of course, if the kurtosis were not known, then you could try and estimate it with assumptions on the eighth moment, and so on. Is there any justification to stop here? The main reason is that this seems like a *useful* place to stop. Large classes of distributions have known bounds on their kurtosis (see table) and the independence of scale is a satisfying property.

Distribution	**Parameters**	**Kurtosis**
Gaussian	$\mu \in \mathbf{R}, \sigma^2 > 0$	3
Bernoulli	$\mu \in [0, 1]$	$\frac{1-3\mu(1-\mu)}{\mu(1-\mu)}$
Exponential	$\lambda > 0$	9
Laplace	$\mu \in \mathbf{R}, b > 0$	9
Uniform	$a < b \in \mathbf{R}$	$9/5$

Table 2: Kurtosis

Contributions The main contribution is the new assumption, algorithm, and the proof of Theorem 2 (see §2). The upper bound is also complemented by an asymptotic lower bound (§3) that applies to all strategies with sub-polynomial regret and all bandit problems with bounded kurtosis.

Additional notation Let $T_i(t) = \sum_{s=1}^{t} \mathbb{1}\{A_s = i\}$ be the number of times arm i has been played after round t. For measures P, Q on the same probability space, $\mathrm{KL}(P, Q)$ is the relative entropy between P and Q and $\chi^2(P, Q)$ is the χ^2 distance. The following lemma is well known.

Lemma 3. *Let X_1, X_2 be independent random variables with X_i having variance σ_i^2 and kurtosis $\kappa_i < \infty$ and skewness $\gamma_i = \mathbf{E}[(X_i - \mathbf{E}[X_i])^3/\sigma_i^3]$, then:*

$$\textit{(a)} \quad \mathrm{Kurt}[X_1 + X_2] = 3 + \frac{\sigma_1^4(\kappa_1 - 3) + \sigma_2^4(\kappa_2 - 3)}{(\sigma_1^2 + \sigma_2^2)^2} \qquad \textit{(b)} \quad \gamma_1 \leq \sqrt{\kappa_1 - 1}\,.$$

2 Algorithm and upper bound

Like the robust upper confidence bound algorithm by Bubeck et al. [2013], the new algorithm makes use of the robust median-of-means estimator.

Median-of-means estimator Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of independent and identically distributed random variables. The median-of-means estimator first partitions the data into m blocks of equal size (up to rounding errors). The empirical mean of each block is then computed and the estimate is the median of the means of each of the blocks. The number of blocks depends on the desired confidence level and should be $O(\log(1/\delta))$. The median-of-means estimator at confidence level $\delta \in (0, 1)$ is denoted by $\widehat{\mathrm{MM}}_\delta((Y_t)_{t=1}^n)$.

Lemma 4 (Bubeck et al. 2013)**.** *Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of independent and identically distributed random variables with mean μ and variance $\sigma^2 < \infty$.*

$$\mathbf{P}\left(\left|\widehat{\mathrm{MM}}_\delta\left((Y_t)_{t=1}^n\right) - \mu\right| \geq C_1 \sqrt{\frac{\sigma^2}{n} \log\left(\frac{C_2}{\delta}\right)}\right) \leq \delta\,,$$