APPENDIX

A RELATED WORKS

We now present a survey of related works in DP accounting.

Moments Accountant and Rényi DP. Abadi et al. (2016) proposed "moments accountant" that uses Rényi DP (Mironov, 2017) to give an upper bound for the DP guarantee of composition of DP algorithms. With the help of moments accountant, Abadi et al. (2016) proposed the differentially private stochastic gradient descent (DP-SGD) algorithm, whose privacy loss can be bounded effectively. However, as mentioned before, Rényi DP can only yield lossy conversion to (ε, δ)-DP, making the upper bound often impractical to use. The runtime of the accountant is independent of m, the number of composition, for DP-SGD, and is $O(m)$ for the composition of general algorithms.

Numerical Composition via FFT. Another line of work (Koskela et al., 2020; Gopi et al., 2021) approximated the privacy loss of compositions using fast Fourier transform to the convolutions of privacy-loss random variables (PRVs). This notion is closely related to our definition of PLLRs. Though both definitions allow for computing compositions via understanding convolutions of random variables, we note that the two concepts stem from a different analysis framework. Specifically, PRV amounts to finding a pair of random variables that reparameterizes the privacy curve, which is dual to the trade-off function. On the other hand, PLLRs are defined naturally from the f-DP's hypothesis testing perspective, hence the random variables have a direct decomposition into sum of the log-likelihood ratios. As a result, our Proposition 3.2 is a strict generalization which encompasses their Theorem 3.2 as a special case when $m = 1$. Note that their FFT accountant is the first algorithm that can approximate the privacy loss up to arbitrary precision, and the runtime of their algorithm is $\widetilde{O}(\sqrt{m})$ for DP-SGD and $O(m^{2.5})$ for general compositions.

Analytical Composition via Characteristic Functions. Recently, Zhu et al. (2021) proposed using characteristic function to analytically compute composition of privacy algorithms. Their algorithm, Analytical Fourier Accountant, yields tight privacy accounting but fails to perform efficient computation for the sub-sampled mechanisms. Their time complexity is $O(1)$ if the characteristic function of their dominating PLD of m-fold is simple enough for closed-form composition, and is at least $\Omega(m^2)$ when no closed-form solution is available.

f-DP accountant via Edgeworth expansion. It is worth mentioning that Zheng et al. (2020) also uses Edgeworth expansion for DP guarantees. Specifically, they use Edgeworth approximation as a refinement to the CLT to better approximate the f-DP trade-off curve. The most important difference between the two approaches is that we provide a finite-sample error bound that allows for an exact DP accountant, while they focus solely on an asymptotic approximation to the trade-off curves. Also, we use Edgeworth approximation on PLLRs to get an estimate of the exact characterization of (ε, δ)-DP (the lower path in Figure 1), while they directly approximate trade-off function f (the same as GDP, using the upper path in Figure 1). Therefore, we focus more on the practical side (finite-sample guarantee), and interpretability (directly deal with (ε, δ)-DP).

B ANALYSIS OF NOISYSGD

We present the algorithms we considered in Section 1.1. To start with, suppose we have a neural network h that is governed by weights $\mathbf{w}$, with samples $\mathbf{x}_i$ and labels y_i ($i = 1, ..., n$). The prediction for each example is $h(\mathbf{x}_i, \mathbf{w})$, and the per-sample loss is given by $\ell(h(\mathbf{x}_i, \mathbf{w}), y_i)$ for some loss function ℓ. We define the objective function L to be the average of per-sample losses

$$L(\mathbf{w}) = \frac{1}{n} \sum_{i=1}^{n} \ell(h(\mathbf{x}_i, \mathbf{w}), y_i).$$

Stochastic Gradient Descent (SGD) algorithm uses a mini-batch as a proxy to this objective function. To control the potential privacy leak in each step of SGD, we need to clip the gradients to control the sensitivity, after which a Gaussian noise is added to it. The details of the algorithm is shown below.

Algorithm 1 NoisySGD (with local or global flat per-sample clipping)

Parameters: initial weights $\mathbf{w}_0$, learning rate η_t, subsampling probability p, number of iterations m, noise scale σ, gradient norm bound R.

for $t = 0, \ldots, m-1$ **do**
  Subsample a batch $I_t \subseteq \{1, \ldots, n\}$ from training set with probability p
  for $i \in I_t$ **do**
    $v_t^{(i)} \leftarrow \nabla_{\mathbf{w}} \ell(f(\mathbf{x}_i, \mathbf{w}_t), y_i)$
    $\bar{v}_t^{(i)} \leftarrow \min\left\{1, R/\|v_t^{(i)}\|_2\right\} \cdot v_t^{(i)}$ ▷ Clip the gradient
  $\bar{V}_t \leftarrow \sum_{i \in I_t} \bar{v}_t^{(i)}$ ▷ Sum over batch
  $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \frac{\eta_t}{|I_t|}\left(\bar{V}_t + \sigma R \cdot \mathcal{N}(0, I)\right)$ ▷ Apply Gaussian mechanism and descend
Output $\mathbf{w}_m$

Recall that in Section 3.1, in order to transfer the bounds from CDF approximations to privacy parameters, we need to find a range that contains all possible roots of $\delta = g^+(\varepsilon), \delta = g^-(\varepsilon)$. Here we showcase how to find such bound in the case of NoisySGD.

Remark B.1. For NoisySGD, we can express such range analytically. Specifically, for any $\alpha \in \{1, 2\}$ (the index of the sequence of PLLRs), we focus on finding roots in the range $[0, C]$ for $\varepsilon^{(\alpha)+}$ and $\varepsilon^{(\alpha)-}$, where C is the smallest value of ε such that

$$C \geq \sup_{S \subset \mathbb{R}, \mathbb{P}(Y^{(\alpha)} \in S) \geq \delta} \log\left(\frac{\mathbb{P}(Y^{(\alpha)} \in S)}{\mathbb{P}(X^{(\alpha)} \in S)}\right).$$

This is clearly a (sub-optimal) upper bound. A loose bound of the range can be easily proved to be

$$C = \min\left\{m \log\left(p \frac{\delta}{1 - \Phi(z_\delta + \mu)}\right), \log\left(\frac{\delta}{1 - \Phi\left(\frac{z_{\delta/\sqrt{m}} + \mu}{\sqrt{m}}\right)}\right)\right\},$$

where z_δ is the upper δ quantile of a standard normal distribution. We can find $0 < \varepsilon^{(\alpha)-} \leq \varepsilon^{(\alpha)+}$ in the range defined above.

C Full definition of AEA and EEAI

Recall that in Section 4.1, the AEA and EEAI are defined for a specific trade-off function $f^{(\alpha)}$. This is only for the simplicity of notations. We now demonstrate how to generalize the definitions to the general trade-off function of the form $\left(\inf_\alpha f^{(\alpha)}\right)^{**}$.

Definition C.1 (AEA for general trade-off function)**.** The k-th order AEA of $\left(\inf_\alpha f^{(\alpha)}\right)^{**}$-DP that defines $\delta(\varepsilon)$ for $\varepsilon > 0$ is given by $\delta(\cdot) = \sup_\alpha \delta^{(\alpha)}(\cdot)$, where

$$\delta^{(\alpha)}(\varepsilon) = 1 - G_{m,k,Y^{(\alpha)}}(\varepsilon) - e^\varepsilon(1 - G_{m,k,X^{(\alpha)}}(\varepsilon)), \tag{C.1}$$

for any α.

Definition C.2 (EEAI for general trade-off function)**.** The k-th order EEAI of $\left(\inf_\alpha f^{(\alpha)}\right)^{**}$-DP associated with privacy parameter $\delta(\varepsilon)$ for $\varepsilon > 0$ is given by $[\delta^-, \delta^+]$, where $\delta^-(\cdot) = \sup_\alpha \delta^{(\alpha)-}(\cdot)$, $\delta^+(\cdot) = \sup_\alpha \delta^{(\alpha)+}(\cdot)$, and

$$\begin{aligned}
\delta^{(\alpha)-}(\varepsilon) &\equiv 1 - G_{m,k,Y^{(\alpha)}}(\varepsilon) - \Delta_{m,k,Y^{(\alpha)}}(\varepsilon) - e^\varepsilon(1 - G_{m,k,X^{(\alpha)}}(\varepsilon) + \Delta_{m,k,X^{(\alpha)}}(\varepsilon)), \\
\delta^{(\alpha)+}(\varepsilon) &\equiv 1 - G_{m,k,Y^{(\alpha)}}(\varepsilon) + \Delta_{m,k,Y^{(\alpha)}}(\varepsilon) - e^\varepsilon(1 - G_{m,k,X^{(\alpha)}}(\varepsilon) - \Delta_{m,k,X^{(\alpha)}}(\varepsilon)).
\end{aligned} \tag{C.2}$$

For completeness, we also provide the formal definition of general k-th order Edgeworth Expansion $E_{m,k,X}$. More details can be found, for example, in Hall (2013).

Definition C.3 (Definition of k-th order Edgeworth Expansion)**.** For any sequence of m distributions $X_1, ..., X_m$, let $S_m = \frac{\sum_{i=1}^m X_i - \sum_{i=1}^m \mathbb{E} X_i}{\bar{B}_m}$ be the standardized sum, where $\bar{B}_m = \sqrt{\mathrm{Var} \sum_{i=1}^m X_i}$.

Define $\lambda_r = \frac{1}{m}\frac{\sum_{j=1}^{m}\kappa_{r,j}}{B_m^r}$, where $\kappa_{r,j}$ is the r-th cumulant of X_j. Assume X_i's have $(k+2)$-th cumulant. We define the k-th Edgeworth expansion $E_{m,k,X}$ as

$$E_{m,k,X}(x) = \Phi(x) + \sum_{r=1}^{k} \frac{1}{n^{r/2}} \frac{P_r(-D)}{D}\phi(x), \tag{C.3}$$

where D is the differential operator, and $P_r(-D)$ is a polynomial of degree $3r$. The explicit form of $P_r(-D)$ can be written as

$$P_r(-D) = \sum \left(\prod_i \frac{1}{k_i!} \left(\frac{\lambda_{i+2}}{(i+2)!} \right)^{k_i} (-D)^{k_i(i+2)} \right),$$

where the summation is over all the integer partitions of m such that $\sum_i ik_i = m$.

D Implementation of Edgeworth Accountant

We now present the detailed implementation of AEA and EEAI.

Algorithm 2 AEA

Parameters: m general mechanisms $M_1, ..., M_m$. An epsilon $\varepsilon \geq 0$, and an order $k \geq 1$.
  for $i = 1, \ldots, m$ **do**
    Analytically encode all the corresponding PLLRs for M_i, $\{(X_i^{(\alpha)}, Y_i^{(\alpha)})\}_\alpha$ for all α.
    Numerically calculate the cumulants up to order $k+2$ for $X_i^{(\alpha)}$ and $Y_i^{(\alpha)}$ for all α.
  Calculate $G_{m,k,X^{(\alpha)}}(\varepsilon)$ and $G_{m,k,Y^{(\alpha)}}(\varepsilon)$ for each α using k-th order Edgeworth expansion.
  Calculate $\delta^{(\alpha)}(\varepsilon)$ for each α by (C.1).
  Output $\sup_\alpha \delta^{(\alpha)}(\varepsilon)$.

And similarly, we present the algorithm for the general EEAI.

Algorithm 3 EEAI

Parameters: m general mechanisms $M_1, ..., M_m$. An epsilon $\varepsilon \geq 0$, and fix the order $k = 1$.
  for $i = 1, \ldots, m$ **do**
    Analytically encode all the corresponding PLLRs for M_i, $\{(X_i^{(\alpha)}, Y_i^{(\alpha)})\}_\alpha$ for all α.
    Numerically calculate the cumulants up to order 4 for $X_i^{(\alpha)}$, and $Y_i^{(\alpha)}$ for all α.
  Calculate $G_{m,1,X^{(\alpha)}}(\varepsilon)$ and $G_{m,1,Y^{(\alpha)}}(\varepsilon)$ for each α using first order Edgeworth expansion.
  Calculate $\Delta_{m,1,X^{(\alpha)}}(\varepsilon)$ and $\Delta_{m,1,Y^{(\alpha)}}(\varepsilon)$ for each α using Lemma 4.3 or Theorem 1.
  Calculate $\delta^{(\alpha)+}(\varepsilon)$ and $\delta^{(\alpha)-}(\varepsilon)$ for each α by (C.2).
  Output $[\sup_\alpha \delta^{(\alpha)-}(\varepsilon), \sup_\alpha \delta^{(\alpha)+}(\varepsilon)]$.

Note that Algorithm 2 (Algorithm 3) is an algorithm to find an estimate (bounds) of δ given an ε. And both algorithms run in constant/linear time for m identical/general compositions. In practice, people often would like to find an estimate or bounds on ε given an δ. To get such an estimate of ε given δ, we can directly inverse the Algorithm 2[4]. And to get upper and lower bounds of ε given δ, we can use the inversion method discussed in Section 3.1, and specifically, the equations in (3.2).

E Additional Experiments

In this section, we perform more numerical experiment. Specifically, we found that in the current implementation of the FFT method[5], they may suffer from numerical issues from heterogeneous

[4]Note that if we substitute Edgeworth approximation with the true CDF of PLLRs , it is direct to show (by taking derivative) that $\delta^{(\alpha)}(\varepsilon)$ is always a decreasing function of ε, and the supremum of decreasing functions is still a decreasing function. Therefore, we can always take inversion.

[5]In all experiments, we use the update-to-date package (v0.2.0) at the time of this submission

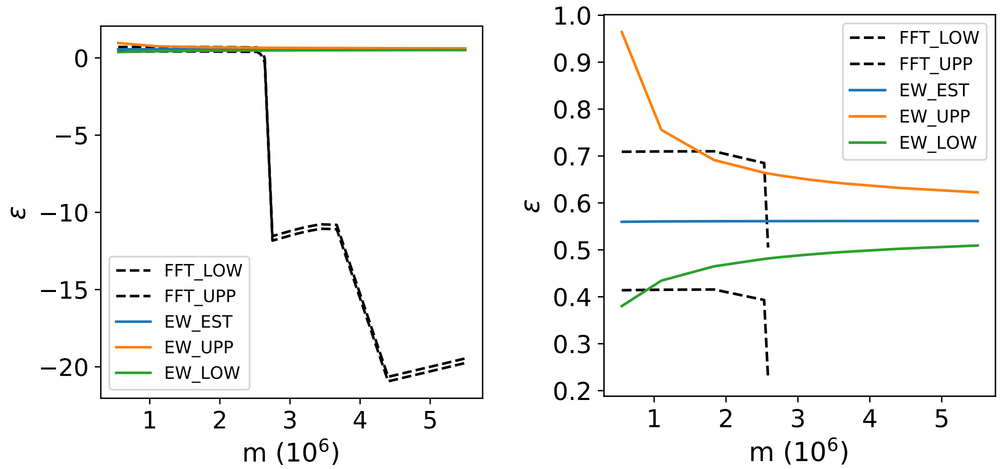

Figure 4: With $\delta = 0.1, \sigma = 0.8$, $p_1 = 0.35/\sqrt{m_1}$, $p_2 = 0.02/\sqrt{m_2}$, $m_2 = 10m_1$, and $m = m_1 + m_2$, the FFT method suffers from numerical issues and fails to correctly evaluate the ε for large m. The left subplot demonstrates that as m grows larger, the FFT method provides negative ε bounds. And the right subplot truncates all epsilon bounds to be positive, showing that our EEAI provides stable bounds even for such large m. This numerical instability may be inherent to the fact that FFT has polynomial dependence on m, and as m grows very large, the numerical stability is an important issue to address. In this case, our EEAI is still numerically stable and useful even when m is very large.

composition with large m. The issue could either prevents the code to run at all, or, in the case that the code successfully executed, have incorrect privacy guarantee. We demonstrate one such example in Figure 4, where we find the FFT method could succeed in running, yet producing incorrect results. We consider two different subsampled Gaussian mechanisms: M_1 with noise multiplier $\sigma = 0.8$ and subsample rate p_1, and M_2 with noise multiplier $\sigma = 0.8$ and subsample rate p_2, we consider the case when we need to compose m_1 times of M_1 with m_2 times of M_2. Similar to the setting in Figure 3, we still set the same precision of ε of FFT method as 0.1 (and if we tune down the precision further, even more severe numerical issue will occur). The privacy guarantees computed by our method and their method are shown in Figure 4. Since the FFT method requires discrete convolution for each mechanism, it can be numerically unstable as m grows larger. In contrast, our analytical finite-sample Edgeworth bounds could be more stable in those scenarios.

F Proofs in Section 3

F.1 Proof of Proposition 3.2

We present the proof of Proposition 3.2 in this section. The proof relies on two Lemmas that are of self-interest and we first present the lemmas. The proof of Proposition 3.2 is straightforward from results of Lemmas. Recall that the trade-off functions $f_i = T(P_i, Q_i)$ we consider are realized by the two following hypotheses:

$$H_{0,i} : w_i \sim P_i \text{ vs. } H_{1,i} : w_i \sim Q_i,$$

where P_i, Q_i are two distributions. To evaluate the trade-off function $f = \bigotimes_{i=1}^m f_i$, we are essentially distinguishing between the two composite hypotheses

$$H_0 : \boldsymbol{w} \sim P_1 \times P_2 \times \cdots \times P_m \quad \text{vs.} \quad H_1 : \boldsymbol{w} \sim Q_1 \times Q_2 \times \cdots \times Q_m, \tag{F.1}$$

where $\boldsymbol{w} = (w_1, ..., w_m)$ is the concatenation of all the w_i's. The following lemma shows how to connect PLLRs of each f_i to the trade-off function f.

Lemma F.1. *Let $X_1, \ldots, X_m$ be the PLLR under the null hypothesis and, likewise, $Y_1, \ldots, Y_m$ be the PLLR under the alternative. Let $F_{X,m}, F_{Y,m}$ be the CDFs of $x \equiv X_1 + \ldots + X_m$ and $Y \equiv Y_1 + \cdots + Y_m$, respectively. Then we have the following relationship between privacy* *parameters and privacy-loss log-likelihood ratios*

$$\begin{aligned}\varepsilon &= \log \frac{F'_{Y,m}(c)}{F'_{X,m}(c)},\\ \delta &= \frac{F'_{X,m}(c)\,(1-F_{Y,m}(c)) - F'_{Y,m}(c)\,(1-F_{X,m}(c))}{F'_{X,m}(c)},\end{aligned} \tag{F.2}$$

where c *is some constant.*

Proof of Lemma F.1. To distinguish between $H_0 : P_1 \times P_2 \times ... \times P_m$ vs. $H_1 : Q_1 \times Q_2 \times ... \times Q_m$, By the Neyman-Pearson lemma, we know that each point of the trade-off function f is realized by a likelihood ratio test (cut-off at some threshold c). So, the trade-off function takes a parametric form $f(\alpha) = \beta$, where α is the type-I error of the test, and β is type-II error of the test:

$$\alpha = \mathbb{P}_{H_0}\left(\log\left(\frac{dP_1 \times P_2 \times \cdots \times P_m}{dQ_1 \times Q_2 \times \cdots \times Q_m}(\boldsymbol{w})\right) > c\right)$$
$$\beta = \mathbb{P}_{H_1}\left(\log\left(\frac{dP_1 \times P_2 \times \cdots \times P_m}{dQ_1 \times Q_2 \times \cdots \times Q_m}(\boldsymbol{w})\right) \le c\right)$$

Note that under H_0, we have

$$\begin{aligned}\log\left(\frac{dP_1 \times P_2 \times \cdots \times P_m}{dQ_1 \times Q_2 \times \cdots \times Q_m}(\boldsymbol{w})\right) &= \log\left(\frac{dP_1}{dQ_1}(w_1) \times \cdots \times \frac{dP_m}{dQ_m}(w_m)\right)\\ &= \log\left(\frac{dP_1}{dQ_1}(w_1)\right) + \cdots + \log\left(\frac{dP_m}{dQ_m}(w_m)\right)\\ &= X_1 + \cdots + X_m = X.\end{aligned}$$

and similaly under H_1,

$$\log\left(\frac{dP_1 \times P_2 \times \cdots \times P_m}{dQ_1 \times Q_2 \times \cdots \times Q_m}(\boldsymbol{w})\right) = Y_1 + \cdots + Y_m = Y.$$

So, we can simplfy the parametric form of f by $f(\alpha) = \beta$, where

$$\begin{aligned}\alpha &= \mathbb{P}\left(X_1 + \cdots + X_m > c\right) = 1 - F_{X,m}(c)\\ \beta &= \mathbb{P}\left(Y_1 + \cdots + Y_m \le c\right) = F_{Y,m}(c).\end{aligned}$$

This allows us to simply write

$$f(\alpha) = F_{Y,m} \circ F_{X,m}^{-1}(1-\alpha).$$

For a point (α, β) on the trade-off function f, where

$$\beta = F_{Y,m}(c) = F_{Y,m}\left(F_{X,m}^{-1}(1-\alpha)\right),$$

and α is small. By the equivalence given in Proposition 2.5 in Dong et al. (2022), we know that the slope of the tangent line passing through (α, β) (for small α) is given by

$$-\mathrm{e}^{\varepsilon} = \frac{df}{d\alpha}(\alpha) = F'_{Y,m}\left(F_{X,m}^{-1}(1-\alpha)\right) \cdot \frac{1}{F'_{X,m}\left(F_{X,m}^{-1}(1-\alpha)\right)} \cdot (-1) = -\frac{F'_{Y,m}\left(F_{X,m}^{-1}(1-\alpha)\right)}{F'_{X,m}\left(F_{X,m}^{-1}(1-\alpha)\right)},$$

which gives

$$\varepsilon = \log \frac{F'_{Y,m}\left(F_{X,m}^{-1}(1-\alpha)\right)}{F'_{X,m}\left(F_{X,m}^{-1}(1-\alpha)\right)} = \log \frac{F'_{Y,m}(c)}{F'_{X,m}(c)}.$$

The equation of the tangent line takes the form of

$$y = -\frac{F'_{Y,m}(c)}{F'_{X,m}(c)}\left(x - 1 + F_{X,m}(c)\right) + F_{Y,m}(c).$$

Using the same proposition,we know the intercept of the line is $1-\delta$, so we should have

$$1-\delta = \frac{F'_{Y,m}(c)\,(1-F_{X,m}(c))}{F'_{X,m}(c)} + F_{Y,m}(c) = \frac{F'_{Y,m}(c)\,(1-F_{X,m}(c)) + F'_{X,m}(c)F_{Y,m}(c)}{F'_{X,m}(c)},$$

which gives

$$\begin{aligned}\delta &= 1 - \frac{F'_{Y,m}(c)\,(1-F_{X,m}(c)) + F'_{X,m}(c)F_{Y,m}(c)}{F'_{X,m}(c)} \\ &= \frac{F'_{X,m}(c)\,(1-F_{Y,m}(c)) - F'_{Y,m}(c)\,(1-F_{X,m}(c))}{F'_{X,m}(c)}.\end{aligned}$$

Therefore, ε and δ takes the following parametric form as in the statement of the lemma,

$$\begin{aligned}\varepsilon &= \log\frac{F'_{Y,m}(c)}{F'_{X,m}(c)} \\ \delta &= \frac{F'_{X,m}(c)\,(1-F_{Y,m}(c)) - F'_{Y,m}(c)\,(1-F_{X,m}(c))}{F'_{X,m}(c)}.\end{aligned}$$

□

To simplify the relation in (F.2), we observe the following interesting lemma about PLLRs.

Lemma F.2. *Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_m$ and $F_{X,m}, F_{Y,m}$ be defined as in Lemma F.1. Let $f_{X,m}, f_{Y,m}$ be the PDFs of $\sum_{i=1}^m X_i$ and $\sum_{i=1}^m Y_i$. Then we have for any $c \in \mathbb{R}$,*

$$c = \log\frac{f_{Y,m}(c)}{f_{X,m}(c)}. \tag{F.3}$$

Proof of Lemma F.2. We use induction on m, the number of compositions, to prove this Lemma.

Base Case: $m=1$. When $m=1$, we write out the forms of X and Y explicitly as

$$\begin{aligned}X &= \log\frac{Q_1(w_1)}{P_1(w_1)} \quad \text{where } w_1 \sim P_1, \\ Y &= \log\frac{Q_1(w_1)}{P_1(w_1)} \quad \text{where } w_1 \sim Q_1.\end{aligned}$$

As a result, for any measurable function $g : \mathbb{R} \to \mathbb{R}$, we have

$$\begin{aligned}\mathbb{E}_Y[g(Y)] &= \mathbb{E}_{w_1\sim Q_1}\left[g\left(\log\frac{Q_1(w_1)}{P_1(w_1)}\right)\right] \\ &= \mathbb{E}_{w_1\sim P_1}\left[g\left(\log\frac{Q_1(w_1)}{P_1(w_1)}\right)\frac{Q_1(w_1)}{P_1(w_1)}\right] \\ &= \mathbb{E}_{w_1\sim P_1}\left[g\left(\log\frac{Q_1(w_1)}{P_1(w_1)}\right)e^X\right] \\ &= \mathbb{E}_X[g(X)e^X].\end{aligned}$$

Since the above equality holds for all g, we must have that their exists a version of both PDFs such that $f_{Y,1}(t) = f_{X,1}(t)e^t$. This shows that for $m=1$,

$$c = \log\frac{f_{Y,1}(c)}{f_{X,1}(c)}.$$

Induction Step: Suppose the result is true for m, we now show that it is also true for $m+1$ compositions. We now claim the following Lemma.

Lemma F.3. *Let A_1, A_2, B_1, B_2 be four random variables. Denote the PDFs of A_1, A_2, B_1, B_2 by $f_{A_1}, f_{A_2}, f_{B_1}, f_{B_2}$ respectively. Suppose further that*

$$f_{B_1}(t) = g(t)f_{A_1}(t), \quad f_{B_2}(t) = g(t)f_{A_2}(t) \quad \textit{for all } t,$$

for some function g satisfying $g(x+y) = g(x)g(y)$. Let $f_{A,2}, f_{B,2}$ denote the density function for $A_1 + A_2, B_1 + B_2$. Then

$$f_{B,2}(t) = g(t)f_{A,2}(t) \quad \textit{for all } t.$$

Proof of Lemma F.3 will be given at the end of the proof. Applying Lemma F.3 on random variables $A_1 = \sum_{i=1}^m X_i, A_2 = X_{m+1}$ and $B_1 = \sum_{i=1}^m Y_i, B_2 = Y_{m+1}$, we will show that we get the desired relationship for $m+1$ compositions. By induction hypothesis we know that $f_{B_1}(t) = g(t) f_{A_1}(t)$ for $g(t) = e^t$. Since $f_{Y_{m+1}}(t) = g(t) f_{X_{m+1}}(t)$ by the base case in induction, and $g(x+y) = g(x)g(y)$, we have $f_{Y,m+1}(t) = g(t) f_{X,m+1}(t)$ for all t. This indicates that we have

$$c = \log \frac{f_{Y,m+1}(c)}{f_{X,m+1}(c)}$$

for any c. Hence we have completed the induction step and concluded the proof.

□

Proof of Lemma F.3. We use the convolution formula on B_1, B_2 and obtain

$$\begin{aligned} f_{B,2}(t) &= \int_{-\infty}^{\infty} f_{B_1}(t-u) f_{B_2}(u) du \\ &= \int_{-\infty}^{\infty} g(t-u) g(u) f_{A_1}(t-u) f_{A_2}(u) du \\ &= \int_{-\infty}^{\infty} g(t) f_{A_1}(t-u) f_{A_2}(u) du \\ &= g(t) f_{A,2}(t) \end{aligned}$$

by convolution formula on A_1, A_2.

□

Proof of Lemma 3.3. Define

$$h(x) = \left(\inf_{\alpha \in \mathcal{I}} f^{(\alpha)} \right)^* (x),$$

which is convex and lower semi-continuous by definition of convex conjugate. By Fenchel–Moreau theorem, we have $h^{**} = h$. Denote $(\varepsilon, \delta(\varepsilon))$ to be the equivalent dual relationship to $f = \inf_\alpha \{f^{(\alpha)}\}^{**}$-DP. From Dong et al. (2022) we know that

$$\delta(\varepsilon) = 1 + \left(\inf_{\alpha \in \mathcal{I}} f^{(\alpha)} \right)^{***} (-e^\varepsilon) = 1 + h^{**}(-e^\varepsilon) = 1 + h(-e^\varepsilon).$$

By order reversing property of convex conjugate, we have

$$\begin{aligned} \delta(\varepsilon) &= 1 + h(-e^\varepsilon) \\ &= 1 + \left(\inf_{\alpha \in \mathcal{I}} f^{(\alpha)} \right)^* (-e^\varepsilon) \\ &= 1 + \sup_{\alpha \in \mathcal{I}} f^{(\alpha)*}(-e^\varepsilon) \\ &= \sup_{\alpha \in \mathcal{I}} \left(1 + f^{(\alpha)*}(-e^\varepsilon) \right) \\ &= \sup_{\alpha \in \mathcal{I}} \delta^{(\alpha)}(\varepsilon) \end{aligned}$$

where we used dual relationship for each $\alpha \in \mathcal{I}$ again in the last step. And the other direction follows directly from the duality of f-DP and $(\varepsilon, \delta(\varepsilon))$-DP, meaning that if the mechanism satisfies $(\varepsilon, \sup_{\alpha \in \mathcal{I}} \delta^{(\alpha)})$-DP, then it also satisfies $f = \left(\inf_\alpha \{f^{(\alpha)}\} \right)^{**}$-DP.

□

G PROOFS IN SECTION 4

G.1 PROOF OF THEOREM 1

Proof of Theorem 1. We first briefly introduce the idea of the proof. The main idea is to construct a random variable $\widetilde{X}_i$ by choosing an $a \geq 0$, such that it stochastically dominates X_i (that is, $X_i \leq \widetilde{X}_i$

a.s.), and satisfies $\mathbb{E}(\widetilde{X}_i) < 0$. We then choose $\eta(a) = -\mathbb{E}(\widetilde{X}_i)$ which is a positive number. In what follows, we will explicitly construct $\widetilde{X}_i$ so that $\widetilde{X}_i$ can be decomposed into the sum of two sub-Gaussian random variables with parameters σ_A^2, σ_B^2. Then since $X_i \leq \widetilde{X}_i$ a.s., we deduce that

$$\mathbb{P}\left(\sum_{i=1}^{m} X_i \geq \varepsilon\right) \leq \mathbb{P}\left(\sum_{i=1}^{m} \widetilde{X}_i \geq \varepsilon\right) = \mathbb{P}\left(\sum_{i=1}^{m} \widetilde{X}_i - \sum_{i=1}^{m} \mathbb{E}(\widetilde{X}_i) \geq \varepsilon + m\eta\right),$$

The final conclusion, which will be proved at the end, follows from the sub-Gaussian bounds.

For notation-wise convenience, we first define a quantity depending on the value of ξ_i, where

$$\Delta(\xi_i) := X_i - (\xi_i\mu - \frac{1}{2}\mu^2) = \log\left(p + \frac{1-p}{e^{\mu\xi_i - \frac{\mu^2}{2}}}\right).$$

It is obvious that $\Delta(\xi_i)$ is a strictly decreasing function of the value of ξ_i. Now, we construct the random variable $\widetilde{X}_i$ as follows. Define

$$\widetilde{X}_i = A_i + B_i$$

where

$$A_i = \begin{cases} X_i & \text{if } \xi_i < a, \\ a^+\mu - \frac{1}{2}\mu^2 + \Delta(a) & \text{if } \xi_i \geq a. \end{cases} \tag{G.1}$$

and

$$B_i = \begin{cases} 0 & \text{if } \xi_i < a, \\ \xi_i\mu - a^+\mu & \text{if } \xi_i \geq a. \end{cases} \tag{G.2}$$

Here, we define $a^+ = \frac{\phi(a)}{1-\Phi(a)}$. Note that $a^+ > a$ for any $a > 0$ by bounds on Mills ratio. To shed light on this decomposition, we first show that $\widetilde{X}_i$ stochastically dominates X_i. Since $\Delta(\xi_i)$ is a decreasing function in ξ_i, hence when $\xi_i \geq a$, we have $\Delta(\xi_i) \leq \Delta(a)$. As a result, when $\xi_i > a$,

$$\widetilde{X}_i = \xi_i\mu - \frac{1}{2}\mu^2 + \Delta(a) \geq \xi_i\mu - \frac{1}{2}\mu^2 + \Delta(\xi_i) = X_i.$$

This guarantees that $\widetilde{X}_i \geq X_i$ a.s.. Another good property of this decomposition, we observe that A_i is sub-Gaussian due to Lemma G.1, and B_i is a mean-zero sub-Gaussian random variable from Lemma G.2. Proof of the two Lemmas is postponed to the next section.

Note that the above construction is valid for any a. Now we show that there exists some $a > 0$ such that $\mathbb{E}(\widetilde{X}_i) = -\eta(a) < 0$ where $\eta(a)$ only depends on a. Note that

$$\mathbb{E}(\widetilde{X}_i) - \mathbb{E}(X_i) = \int_a^\infty (\Delta(a) - \Delta(\xi))\phi(\xi)d\xi,$$

where $\phi(x)$ is the density for standard Normal random variable. Also recall that $\mathbb{E}(X_i) < 0$. Then

$$\begin{aligned}\mathbb{E}(\widetilde{X}_i) &= \mathbb{E}(X_i) + \int_a^\infty (\Delta(a) - \Delta(\xi))\phi(\xi)d\xi \\ &= \mathbb{E}(X_i) + e(a),\end{aligned}$$

where $e(a)$ satisfies that $\lim_{a\to\infty} e(a) = 0$. This is because by construction $\int_a^\infty (\Delta(a) - \Delta(\xi))\phi(\xi)d\xi \geq 0$ and that

$$e(a) = \int_a^\infty \Delta(a)\phi(\xi)d\xi - \int_a^\infty \Delta(\xi)\phi(\xi)d\xi.$$

The second term vanishes as $\xi \to \infty$ since $\Delta(\xi)$ is integrable. For the first term, if $\Delta(a) < 0$ the integral is already negative. If $\Delta(a) > 0$ we have $\int_a^\infty \Delta(a)\phi(\xi)d\xi < \int_a^\infty \Delta(0)\phi(\xi)d\xi$ which also vanishes. As a result, we have shown that $\lim_{a\to\infty} e(a) \leq 0$. Combined with what we have above, we deduce that $\lim_{a\to\infty} e(a) = 0$ as required. Then we can pick a large enough such that $e(a) = -\frac{1}{2}\mathbb{E}(X_i)$ and we have $\mathbb{E}(\widetilde{X}_i) = \frac{1}{2}\mathbb{E}(X_i) < 0$.

Now we can combine the previous results and prove the tail bound of $\sum_{i=1}^m X_i$. Recall we have constructed random variable $\widetilde{X}_i = A_i + B_i$ such that $\widetilde{X}_i \geq X_i$ a.s. with $\mathbb{E}(\widetilde{X}_i) = -\eta(a) < 0$. Moreover, A_i, B_i are both sub-Gaussian with parameters σ_A^2 and σ_B^2. Then we have

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^m X_i \geq \varepsilon\right) &\leq \mathbb{P}\left(\sum_{i=1}^m \widetilde{X}_i \geq \varepsilon\right) = \mathbb{P}\left(\sum_{i=1}^m \widetilde{X}_i - \sum_{i=1}^m \mathbb{E}(\widetilde{X}_i) \geq \varepsilon + m\eta\right) \\ &\leq \mathbb{P}\left(\sum_{i=1}^m A_i - \sum_{i=1}^m \mathbb{E}(A_i) + \sum_{i=1}^m B_i - \sum_{i=1}^m \mathbb{E}(B_i) \geq \varepsilon + m\eta\right) \\ &\leq \mathbb{P}\left(\sum_{i=1}^m A_i - \sum_{i=1}^m \mathbb{E}(A_i) \geq \frac{\varepsilon + m\eta}{2}\right) + \mathbb{P}\left(\sum_{i=1}^m B_i - \sum_{i=1}^m \mathbb{E}(B_i) \geq \frac{\varepsilon + m\eta}{2}\right),\end{aligned}$$

where the last inequality follows from the union bound. Finally, since A_i, B_i are both sub-Gaussian, we know that their sum $\sum_{i=1}^m A_i, \sum_{i=1}^m B_i$ are still sub-Gaussian. Hence

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^m A_i - \sum_{i=1}^m \mathbb{E}(A_i) \geq \frac{\varepsilon + m\eta}{2}\right) &\leq \exp\left(-\frac{(\varepsilon + m\eta)^2}{8m\sigma_A^2}\right), \\ \mathbb{P}\left(\sum_{i=1}^m B_i - \sum_{i=1}^m \mathbb{E}(B_i) \geq \frac{\varepsilon + m\eta}{2}\right) &\leq \exp\left(-\frac{(\varepsilon + m\eta)^2}{8m\sigma_B^2}\right).\end{aligned}$$

As a result,

$$\mathbb{P}\left(\sum_{i=1}^m X_i \geq \varepsilon\right) \leq \exp\left(-\frac{(\varepsilon + m\eta)^2}{8m\sigma_A^2}\right) + \exp\left(-\frac{(\varepsilon + m\eta)^2}{8m\sigma_B^2}\right) \leq 2\exp\left(-\frac{(\varepsilon + m\eta)^2}{8m\tau^2}\right),$$

where

$$\begin{aligned}\tau^2 &= \max\{\sigma_A^2, \sigma_B^2\} \\ &= \max\left\{\frac{(\log(1 - p + pe^{\mu a - \frac{1}{2}\mu^2}) + \mu(a^+ - a) - \log(1-p))^2}{4}, \mu^2, \frac{(a^+ - a)^2\mu^2}{2\log(\Phi(a^+) - \Phi(a))}\right\}.\end{aligned}$$

□

G.2 Technical Lemmas

Lemma G.1. *The random variable A_i, defined in* (G.1) *is sub-Gaussian random variable with parameter σ_A^2 where* $\sigma_A^2 = \dfrac{(\log(1 - p + pe^{\mu a - \frac{1}{2}\mu^2}) + \mu(a^+ - a) - \log(1-p))^2}{4}$.

Proof of Lemma G.1. The proof of Lemma G.1 is straightforward, we show that A_i is bounded and thus sub-Gaussian by Hoeffding's inequality. Note that when $\xi_i < a$, we have $A_i = X_i = \log(1 - p + pe^{\mu\xi_i - \frac{1}{2}\mu^2}) < \log(1 - p + pe^{\mu a - \frac{1}{2}\mu^2})$. Moreover, since X_i is bounded below by $\log(1-p)$, we deduce that when $\xi_i < a$, we have

$$\log(1-p) < A_i < \log(1 - p + pe^{\mu a - \frac{1}{2}\mu^2}),$$

which is bounded as desired.

On the other hand, when $\xi_i > a$, by definition of $\Delta(\xi_i)$,

$$A_i = a^+\mu - \frac{1}{2}\mu^2 + \log\left(p + \frac{1-p}{e^{\mu a - \frac{\mu^2}{2}}}\right) = \log(1 - p + pe^{\mu a - \frac{1}{2}\mu^2}) + \mu(a^+ - a),$$

which is a constant. Since $a^+ > a$, in this case the above constant is greater than $\log(1 - p + pe^{\mu a - \frac{1}{2}\mu^2})$. Combine the above two settings, we deduce that $A_i \in (\log(1-p), \log(1 - p + pe^{\mu a - \frac{1}{2}\mu^2}) + \mu(a^+ - a))$ is a bounded random variable. By Hoeffding's inequality, it is sub-Gaussian with parameter defined in the Lemma.

□

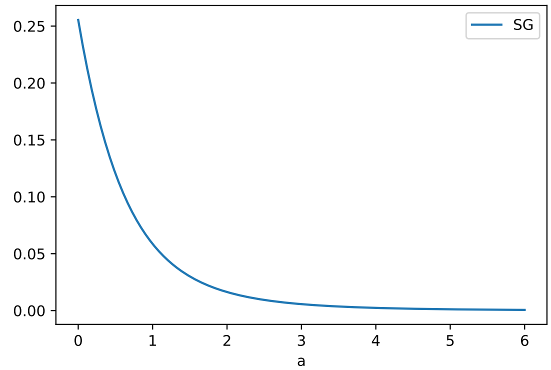

Figure 5: The value of $\frac{(a^+-a)^2}{2\log(\Phi(a^+)-\Phi(a))}$ when $a > 0$.

Lemma G.2. *The random variable B_i defined in Equation* (G.2) *is a mean-zero sub-Gaussian random variable with parameter $\sigma_B^2 = \mu^2 \max\left\{1, \frac{(a^+-a)^2}{2\log(\Phi(a^+)-\Phi(a))}\right\}$.*

Remark G.3. We note that as a function of a, $\frac{(a^+-a)^2}{2\log(\Phi(a^+)-\Phi(a))}$ is in fact a decreasing function, and is always less than 1 as $a > 0$. Its plot can be found in Figure 5. Therefore, by truncating normal at a, we essentially loss nothing, since B_i is still a sub-Gaussian random variable with parameter μ.

Proof of Lemma G.2. Recall the definition in Equation (G.2), we can re-write B_i as a mixture random variable

$$B_i = \begin{cases} 0 & \text{w.p. } \mathbb{P}(\xi_i < a), \\ \widetilde{\xi}_i\mu - a^+\mu & \text{w.p. } \mathbb{P}(\xi_i \geq a). \end{cases}$$

where $\widetilde{\xi}_i = \xi_i \big| \xi_i > 0$ is the normal $\mathcal{N}(0,1)$ truncated at $a > 0$, whose probability density function is given by

$$f(t) = \frac{\phi(t)}{1-\Phi(a)}, \text{ for } t > a.$$

From Lemma G.4, we know the expectation of B_i is

$$\mathbb{E}[B_i] = 0 + (1-\Phi(a))(\mathbb{E}[\widetilde{\xi}_i] - a^+)\mu = 0.$$

Therefore, to prove that the mean-zero variable B_i is sub-Gaussian, we only need to bound the probability of $\mathbb{P}(B_i > t)$ and $\mathbb{P}(B_i < -t)$ for any $t > 0$ with the form of $\exp(-\frac{t^2}{2\sigma^2})$ for some $\sigma > 0$.

We will first prove the part for $\mathbb{P}(B_i > t)$ Note that

$$\begin{aligned}
\mathbb{P}(B_i > t) &= (1-\Phi(a))\mathbb{P}(\widetilde{\xi}_i\mu - a^+\mu > t) \\
&= (1-\Phi(a))\mathbb{P}(\widetilde{\xi}_i\mu - a^+\mu > t) \\
&= (1-\Phi(a))\mathbb{P}(\widetilde{\xi}_i > a^+ + t/\mu) \\
&= (1-\Phi(a))\left(1 - \frac{(\Phi(a^+ + t/\mu) - \Phi(a))}{(1-\Phi(a))}\right) \\
&= 1 - \Phi(a^+ + t/\mu) \\
&\leq 1 - \Phi(t/\mu) \leq \exp\left(-\frac{t^2}{2\mu^2}\right),
\end{aligned}$$

where the fourth equality is due to (2) in Lemma G.4, the first inequality is due to the fact that $a^+ \geq a > 0$, and the last inequality is due to the fact that $\mathcal{N}(0, \mu^2)$ is sub-Gaussian with parameter μ^2.

We now prove the other side. Observe that $B_i > \mu(a^+ - a)$, so for $t > \mu(a^+ - a)$ we have $\mathbb{P}(B_i < -t) = 0$. Therefore, we only need to bound $\mathbb{P}(B_i < -t)$ for $0 < t <= \mu(a^+ - a)$. Note that in this range,

$$\begin{aligned}\mathbb{P}(B_i < -t) \leq \mathbb{P}(B_i < 0) &= (1 - \Phi(a))\mathbb{P}(\widetilde{\xi}_i \mu - a^+ \mu < 0)\\ &= (1 - \Phi(a))\mathbb{P}(\widetilde{\xi}_i < a^+)\\ &= \Phi(a^+) - \Phi(a),\end{aligned}$$

where the last equality is again due to (2) of Lemma G.4. On the other hand, we have for any $\sigma > 0$,

$$\exp\left(-\frac{t^2}{2\sigma^2}\right) \geq \exp\left(-\frac{(\mu(a^+ - a))^2}{2\sigma^2}\right), \text{ for any } 0 < t \leq \mu(a^+ - a).$$

And with the choice of $\sigma^2 = \mu^2 \frac{(a^+ - a)^2}{2\log(\Phi(a^+) - \Phi(a))}$, we have

$$\exp\left(-\frac{t^2}{2\sigma^2}\right) \geq \exp\left(-\frac{(\mu(a^+ - a))^2}{2\sigma^2}\right) = \Phi(a^+) - \Phi(a) \geq \mathbb{P}(B_i < -t)$$

holds for all $0 < t \leq \mu(a^+ - a)$.

Therefore, combining the two sides, we know that B_i is sub-Gaussian with parameter $\max\{\mu^2, \mu^2 \frac{(a^+ - a)^2}{2\log(\Phi(a^+) - \Phi(a))}\}$, which concludes the proof.

□

Lemma G.4. *For a truncated normal distribution $\widetilde{\xi}_i$ with density $f(t) = \frac{\phi(t)}{1-\Phi(a)}$,for $t > a$ we have*

1. $\mathbb{E}(\widetilde{\xi}_i) = \frac{\phi(a)}{1-\Phi(a)}$.
2. $\mathbb{P}(\widetilde{\xi}_i \leq t) = \frac{\Phi(t) - \Phi(a)}{1-\Phi(a)}$.

Proof of Lemma G.4. This is based on several well-known truncated normal properties, and is easy to prove from the density function. Therefore we omit the proof here. □

H Details of Edgeworth approximation error

The following discussion is largely adapted from Derumigny et al. (2021) to be self-contained. For a distribution P, let f_P denote its characteristic function; similarly, for a random variable X, we denote by f_X its characteristic function. We recall that $f_{\mathcal{N}(0,1)}(t) = e^{-t^2/2}$. Some constants are used in the definition.

- Denote by χ_1 the constant $\chi_1 := \sup_{x>0} x^{-3} \left|\cos(x) - 1 + x^2/2\right| \approx 0.099162$ (Shevtsova, 2010),
- Denote by θ_1^* the unique root in $(0, 2\pi)$ of the equation $\theta^2 + 2\theta\sin(\theta) + 6(\cos(\theta) - 1) = 0$,
- Denote by $t_1^* := \theta_1^*/(2\pi) \approx 0.635967$ (Shevtsova, 2010).

H.1 Details of first-order Edgeworth expansion

We now provide details on the first-order Edgeworth expansion in Lemma 4.3. The main narrative is adapted from Derumigny et al. (2021).

We first define the reminder term $r_{1,m}$. To this end, we define

$$\Psi(t) := \frac{1}{2}\left(1 - |t| + i\left[(1 - |t|)\cot(\pi t) + \frac{\text{sign}(t)}{\pi}\right]\right)\mathbb{1}\{|t| \leq 1\}$$

where i is the imaginary number. Note that from Prawitz (1975) we have the following bound for function Ψ:

$$|\Psi(t)| \leq \frac{1.0253}{2\pi|t|} \text{ and } \left|\Psi(t) - \frac{i}{2\pi t}\right| \leq \frac{1}{2}\left(1 - |t| + \frac{\pi^2}{18}t^2\right).$$

We further define

$$I_{3,1}(T) := \frac{2}{T}\int_0^{\sqrt{2\varepsilon}(m/K_{4,m})^{1/4}} |\Psi(u/T)| \left| f_{S_m}(u) - e^{-u^2/2}\left(1 - \frac{iu^3\lambda_{3,m}}{6\sqrt{m}}\right)\right| du$$
$$I_{3,2}(T) := \frac{2}{T}\int_{\sqrt{2\varepsilon}(m/K_{4,m})^{1/4}}^{t_0 T} |\Psi(u/T)| \left| f_{S_m}(u) - e^{-u^2/2}\right| du$$
$$I_{3,3}(T) := \frac{2}{T}\frac{|\lambda_{3,m}|}{6\sqrt{m}}\int_{\sqrt{2\varepsilon}(m/K_{4,m})^{1/4}}^{t_0 T} |\Psi(u/T)| e^{-u^2/2}|u|^3 du,$$

and $r_{1,m}$ is defined to be

$$r_{1,m} := \frac{(14.1961 + 67.0415)\widetilde{K}_{3,m}^4}{16\pi^4 m^2} + \frac{|\lambda_{3,m}| \exp\left(-2m^2/\widetilde{K}_{3,m}^4\right)}{3\pi\sqrt{m}} + I_{3,2}(T) + I_{3,3}(T)$$
$$+ \frac{1.0253}{\pi}\int_0^{\sqrt{2\varepsilon}(n/K_{4,m})^{1/4}} ue^{-u^2/2} R_{1,m}(u,\varepsilon) du. \tag{H.1}$$

For $\varepsilon \in (0, 1/3)$ and $t \geq 0$, we further define

$$R_{1,m}(t,\varepsilon) := \frac{U_{1,1,m}(t) + U_{1,2,m}(t)}{2(1-3\varepsilon)^2} + e_1(\varepsilon)\left(\frac{t^8 K_{4,m}^2}{2m^2}\left(\frac{1}{24} + \frac{P_{1,m}(\varepsilon)}{2(1-3\varepsilon)^2}\right)^2 + \frac{|t|^7 |\lambda_{3,m}| K_{4,m}}{6m^{3/2}}\left(\frac{1}{24} + \frac{P_{1,m}(\varepsilon)}{2(1-3\varepsilon)^2}\right)\right),$$
$$P_{1,m}(\varepsilon) := \frac{144 + 48\varepsilon + 4\varepsilon^2 + \left\{96\sqrt{2\varepsilon} + 32\varepsilon + 16\sqrt{2}\varepsilon^{3/2}\right\} \mathbb{1}\left\{\exists i \in \{1,\dots,m\} : \mathbb{E}\left[(X_i - \mu_i)^3\right] \neq 0\right\}}{576},$$
$$e_1(\varepsilon) := \exp\left(\varepsilon^2\left(\frac{1}{6} + \frac{2P_{1,m}(\varepsilon)}{(1-3\varepsilon)^2}\right)\right),$$
$$U_{1,1,m}(t) := \frac{t^6}{24}\left(\frac{K_{4,m}}{m}\right)^{3/2} + \frac{t^8}{24^2}\left(\frac{K_{4,m}}{m}\right)^2,$$
$$U_{1,2,m}(t) := \left(\frac{|t|^5}{6}\left(\frac{K_{4,m}}{m}\right)^{5/4} + \frac{t^6}{36}\left(\frac{K_{4,m}}{m}\right)^{3/2} + \frac{|t|^7}{72}\left(\frac{K_{4,m}}{m}\right)^{7/4}\right) \mathbb{1}\left\{\exists i \in \{1,\dots,m\} : \mathbb{E}\left[(X_i - \mu_i)^3\right] \neq 0\right\}.$$

Observe the bound from Lemma 4.3 is a bound of leading order $O(1/\sqrt{m})$, which is due to the fact that the variables in the sequence may not be identical since we may encounter non-identical compositions, and we do not require any continuous property of the densities (and their existence as well). When we have i.i.d. distribution of absolute continuous density, we can guarantee to have an $O(1/m)$ explicit bound of the difference as

$$\Delta_{m,1} \leq \frac{0.195K_{4,m} + 0.038\lambda_{3,m}^2}{m} + \frac{1.0253}{\pi}\int_{a_m}^{b_m} \frac{|f_{S_m}(t)|}{t} dt + r_{2,m},$$

where $a_m := 2t_1^*\pi\sqrt{m}/\widetilde{K}_{3,m}, b_n := 16\pi^4 m^2/\widetilde{K}_{3,m}^4$, and $r_{2,m}$ is a remainder term that depends only on $K_{3,m}, K_{4,m}$ and $\lambda_{3,m}$. Specifically, the term $r_{2,m}$ is defined by

$$r_{2,m} := \frac{1.2533\widetilde{K}_{3,m}^4}{16\pi^4 m^2} + \frac{0.3334\,|\lambda_{3,m}|\,\widetilde{K}_{3,m}^4}{16\pi^4 m^{5/2}} + \frac{14.1961\widetilde{K}_{3,m}^{16}}{16^4\pi^{16} m^8} + \frac{|\lambda_{3,m}| \exp\left(-128\pi^6 m^4/\widetilde{K}_{3,m}^8\right)}{3\pi\sqrt{m}}$$
$$+ I_{5,2}(T) + I_{5,3}(T) + I_{5,4}(T) + J_3(T) + J_5(T)$$
$$+ \frac{1.0253}{\pi}\int_0^{\sqrt{2\varepsilon}(m/K_{4,m})^{1/4}} ue^{-u^2/2} R_{1,m}(u,\varepsilon) du.$$

Here,

$$I_{5,2}(T) := E_{1,m}\frac{|\lambda_{3,m}|}{3T\sqrt{m}}\int_{\sqrt{2\varepsilon}(m/K_{4,m})^{1/4}}^{T^{1/4}/\pi}|\Psi(u/T)|u^3e^{-u^2/2}du,$$

$$I_{5,3}(T) := E_{1,m}\frac{2}{T}\int_{\sqrt{2\varepsilon}(m/K_{4,m})^{1/4}}^{T^{1/4}}|\Psi(u/T)|\left|f_{S_m}(u)-e^{-u^2/2}\right|du,$$

$$I_{5,4}(T) := E_{2,m}\frac{|\lambda_{3,m}|}{3T\sqrt{m}}\int_{T^{1/4}/\pi}^{T/\pi}|\Psi(u/T)||u|^3e^{-u^2/2}du,$$

where $E_{1,m} := 1_{\left\{\sqrt{2\varepsilon}\,(m/K_{4,m})^{1/4}<T^{1/4}/\pi\right\}}$ and $E_{2,m} := 1_{\left\{T^{1/4}<T\right\}}$. Further, $T = 16\pi^4m^2/\widetilde{K}_{3,m}^4$. Note that if $T^{1/4} > T$ or $\sqrt{2\varepsilon}\,(m/K_{4,m})^{1/4} > T^{1/4}/\pi$, our bounds are still valid and can even be improved in the sense that the corresponding integrals can be removed. Further, we have the following bound for the terms $I_{5,2}$, $I_{5,3}$ and $I_{5,4}$:

$$\begin{aligned}I_{5,2}(T) &\le \frac{|\lambda_{3,m}|}{3\sqrt{m}}\int_{\sqrt{2\varepsilon}(m/K_{4,m})^{1/4}}^{T^{1/4}/\pi}\frac{1.0253}{2\pi}u^2e^{-u^2/2}du\\ &= \frac{1.0253\,|\lambda_{3,m}|}{3\pi\sqrt{2}\sqrt{m}}\left(\Gamma\left(3/2,\varepsilon\,(m/K_{4,m})^{1/2}\right)-\Gamma\left(3/2,T^{1/2}/2\pi^2\right)\right),\end{aligned}$$

$$\begin{aligned}I_{5,3}(T) &\le \frac{2}{T}\int_{\sqrt{2\varepsilon}(m/K_{4,m})^{1/4}}^{T^{1/4}/\pi}|\Psi(u/T)|\frac{K_{3,m}}{6\sqrt{m}}|t|^3\exp\left(-\frac{t^2}{2}+\frac{\chi_1|t|^3\widetilde{K}_{3,m}}{\sqrt{m}}+\frac{t^2\sqrt{K_{4,m}}}{2\sqrt{m}}\right)du\\ &= \frac{K_{3,m}}{3\sqrt{m}}J_2\left(3,\sqrt{2\varepsilon}/\left(mK_{4,m}\right)^{1/4},T^{1/4}/\pi,\widetilde{K}_{3,m},K_{4,m},T,m\right)\end{aligned}$$

and

$$I_{5,4}(T) = \frac{1.0253\,|\lambda_{3,m}|}{3\pi\sqrt{2}\sqrt{m}}\left(\Gamma\left(3/2,T^{1/2}/2\pi^2\right)-\Gamma\left(3/2,T^2/2\pi^2\right)\right),$$

and all the terms converge exponentially fast to zero. Here $\Gamma(a,x)$ is the incomplete Gamma function and can be numerically evaluated.

For the other terms, we have

$$\begin{aligned}J_3(T) &:= \frac{2}{T}\int_{T^{1/4}/\pi}^{t_1^*T^{1/4}}|\Psi(u/T)|\,|f_{S_m}(u)|\,du = \frac{2}{T^{3/4}}\int_{1/\pi}^{t_1^*}\left|\Psi\left(v/T^{3/4}\right)\right|\left|f_{S_m}\left(T^{1/4}v\right)\right|dv,\\ J_4(T) &:= 1_{\left\{t_1^*T^{1/4}<T/\pi\right\}}\frac{2}{T}\int_{t_0^*T^{1/4}}^{T/\pi}|\Psi(u/T)|\,|f_{S_m}(u)|\,du,\\ J_5(T) &:= \frac{2}{T}\int_{T^{1/4}/\pi}^{T/\pi}|\Psi(u/T)|e^{-u^2/2}du.\end{aligned}$$

Obviously, now all the above bounds are real integrations, and can be calculated numerically.

H.2 Extension to higher-order Edgeworth expansion

We now briefly state that how we can extend the current first-order Edgeworth bound to higher-orders. We essentially need to upper bound the approximation error by a careful decomposition. For example, when extending to the second-order Edgeworth expansion, we have the following new smoothing Lemma.

Lemma H.1. *For every $t_0 \in (0,1]$ and every $T > 0$, we have*

$$\Delta_{m,2} \le \Omega_1\left(t_0,T\right)+\Omega_2\left(t_0,T\right)+\Omega_3\left(t_0,T\right),$$

where

$$\begin{aligned}
\Omega_1\left(t_0, T\right) :=& 2 \int_0^{t_0}\left|\Psi(t)-\frac{i}{2 \pi t}\right| e^{-(T t)^2 / 2}\left(\left|1+\frac{\left|\lambda_{4, m}\right||T t|^4}{24 m}-\frac{\lambda_{3, m}^2|T t|^6}{72 m}\right|+\frac{\left|\lambda_{3, m}\right||T t|^3}{6 \sqrt{m}}\right) d t \\
& +\frac{1}{\pi} \int_{t_0}^{+\infty} \frac{e^{-(T t)^2 / 2}}{t}\left(\left|1+\frac{\left|\lambda_{4, m}\right||T t|^4}{24 m}-\frac{\lambda_{3, m}^2|T t|^6}{72 m}\right|+\frac{\left|\lambda_{3, m}\right||T t|^3}{6 \sqrt{m}}\right) d t, \\
\Omega_2\left(t_0, T\right) :=& 2 \int_{t_0}^1|\Psi(t)|\left|f_{S_m}(T t)\right| d t, \\
\Omega_3\left(t_0, T\right) :=& 2 \int_0^{t_0}|\Psi(t)|\left|f_{S_m}(T t)-e^{-(T t)^2 / 2}\left(1-\frac{\lambda_{3, m} i(T t)^3}{6 \sqrt{m}}+\frac{\lambda_{4, m}(T t)^4}{24 m}-\frac{\lambda_{3, m}^2(T t)^6}{72 m}\right)\right| d t .
\end{aligned}$$

Using such bound we can numerically compute $\Omega_1\left(t_0, T\right), \Omega_2\left(t_0, T\right), \Omega_3\left(t_0, T\right)$ for suitably chosen t_0, T and get the uniform bound on Edgeworth expansion of different orders. It is expected that the order of approximation error decays as we increase the order of Edgeworth expansion. In practice however, we notice that first-order Edgeworth expansion already yields accurate results.