The following pages contain the expanded versions of the proofs for Theorems 2 and 4.

APPENDIX

Theorem 2. *Let $\ell(\theta,x,y)$ be a loss function, $\lambda>0$ a variance penalization factor, and $\gamma\in\mathbb{R}$ the variance exponent. We then have the following bounds for the two stochastic optimization problems:*

$$\max_{\mathbb{Q}_n\in\mathcal{A}} \mathbb{E}_{\mathbb{Q}_n}\left[\ell(\theta,x,y)\right] \le \mathbb{E}_{\mathbb{P}_n}[\ell(\theta,x,y)] + \lambda\mathbb{V}_{\mathbb{P}_n}[\ell(\theta,x,y)]^{\gamma}$$
$$\min_{\mathbb{Q}_n\in\mathcal{A}} \mathbb{E}_{\mathbb{Q}_n}\left[\ell(\theta,x,y)\right] \ge \mathbb{E}_{\mathbb{P}_n}[\ell(\theta,x,y)] - \lambda\mathbb{V}_{\mathbb{P}_n}[\ell(\theta,x,y)]^{\gamma}$$

where $\mathcal{A}=\left\{\mathbb{Q}_n : D_{\chi^2}(\mathbb{Q}_n\,\|\,\mathbb{P}_n)\le \frac{\lambda^2}{n^2}\mathbb{V}_{\mathbb{P}_n}[\ell(\theta,x,y)]^{2\gamma-1}\right\}$.

Proof.

Let Z be a random variable with $Z_i=\ell(\theta,x_i,y_i)$.

1. The expectation of the two stochastic optimization problems can be written as:
$$\sum_{i=1}^{n} q_i Z_i \quad \text{s.t.} \quad q_i\ge 0, \quad \sum_{i=1}^{n} q_i = 1, \quad \frac{1}{n}\sum_{i=1}^{n}\left(nq_i-1\right)^2 \le \frac{\lambda^2}{n^2}\mathbb{V}_{\mathbb{P}_n}[Z]^{2\gamma-1}.$$

2. Writing q_i in terms of a new variable u_i with $q_i=\frac{1}{n}+u_i$, we have:
$$\frac{1}{n}\sum_{i=1}^{n} Z_i + \sum_{i=1}^{n} u_i Z_i \quad \text{s.t.} \quad u_i\ge -\frac{1}{n}, \quad \sum_{i=1}^{n} u_i = 0, \quad \frac{1}{n}\sum_{i=1}^{n} u_i^2 \le \frac{\lambda^2}{n^2}\mathbb{V}_{\mathbb{P}_n}[Z]^{2\gamma-1}.$$

3. $\frac{1}{n}\sum_{i=1}^{n} Z_i + \sum_{i=1}^{n} u_i Z_i \quad \Leftrightarrow \quad \mathbb{E}_{\mathbb{P}_n}[Z] + n\mathbb{E}_{\mathbb{P}_n}\left[u\big(Z-\mathbb{E}_{\mathbb{P}_n}[Z]\big)\right]$

   Proof: For the second term of the sum we have:
$$\begin{aligned}
\sum_{i=1}^{n} u_i Z_i &= n\mathbb{E}_{\mathbb{P}_n}\left[uZ\right] \\
&= n\mathbb{E}_{\mathbb{P}_n}\left[u\big(Z-\mathbb{E}_{\mathbb{P}_n}[Z]\big)\right] + n\mathbb{E}_{\mathbb{P}_n}\left[u\mathbb{E}_{\mathbb{P}_n}[Z]\right] \\
&= n\mathbb{E}_{\mathbb{P}_n}\left[u\big(Z-\mathbb{E}_{\mathbb{P}_n}[Z]\big)\right] + n\mathbb{E}_{\mathbb{P}_n}\left[u\right]\mathbb{E}_{\mathbb{P}_n}[Z] \\
&= n\mathbb{E}_{\mathbb{P}_n}\left[u\big(Z-\mathbb{E}_{\mathbb{P}_n}[Z]\big)\right]
\end{aligned}$$
   Where on penultimate step we use the fact that $\mathbb{E}_{\mathbb{P}_n}[u]=0$ since $\sum_{i=1}^{n} u_i = 0$.

4. $n\left|\mathbb{E}_{\mathbb{P}_n}\left[u\big(Z-\mathbb{E}_{\mathbb{P}_n}[Z]\big)\right]\right| \le \lambda\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma}$

   Proof: Using the Cauchy–Schwarz inequality we have:
$$\begin{aligned}
n\left|\mathbb{E}_{\mathbb{P}_n}\left[u\big(Z-\mathbb{E}_{\mathbb{P}_n}[Z]\big)\right]\right| &\le n\sqrt{\mathbb{E}_{\mathbb{P}_n}\left[u^2\right]\mathbb{E}_{\mathbb{P}_n}\left[\big(Z-\mathbb{E}_{\mathbb{P}_n}[Z]\big)^2\right]} \\
&\le n\sqrt{\frac{\lambda^2}{n^2}\mathbb{V}_{\mathbb{P}_n}[Z]^{2\gamma-1}\mathbb{V}_{\mathbb{P}_n}[Z]} \\
&\le \lambda\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma}
\end{aligned}$$
   Where for the second inequality we used the constraint from step 2 on u^2.

5. $n\left|\mathbb{E}_{\mathbb{P}_n}\left[u\big(Z-\mathbb{E}_{\mathbb{P}_n}[Z]\big)\right]\right| = \lambda\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma}$ for $u_i=\pm\frac{\lambda}{n}\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}\big(Z_i-\mathbb{E}_{\mathbb{P}_n}[Z]\big)$.

   Proof: From Cauchy–Schwarz, the equality holds if u_i and $\big(Z_i-\mathbb{E}_{\mathbb{P}_n}[Z]\big)$ are linearly dependent. Thus, we achieve equality for
$$u_i = \pm\frac{\frac{\lambda}{n}\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma}}{\mathbb{V}_{\mathbb{P}_n}[Z]}\big(Z_i-\mathbb{E}_{\mathbb{P}_n}[Z]\big) = \pm\frac{\lambda}{n}\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}\big(Z_i-\mathbb{E}_{\mathbb{P}_n}[Z]\big)$$
   given that the constraint $u_i\ge -\frac{1}{n}$ is satisfied.

6. When the constraint $u_i \geq -\frac{1}{n}$ is violated, $n\left|\mathbb{E}_{\mathbb{P}_n}\left[u(Z-\mathbb{E}_{\mathbb{P}_n}[Z])\right]\right| < \lambda\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma}$.

   Proof: To avoid negative probabilities we clamp u_i with values smaller than $-\frac{1}{n}$ to $-\frac{1}{n}$, however, as a result u_i and $\big(Z_i - \mathbb{E}_{\mathbb{P}_n}[Z]\big)$ are no longer linearly dependent.

   In addition we have $u_i = \pm\frac{\alpha\lambda}{n}\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}(Z-\mathbb{E}_{\mathbb{P}_n}[Z]-\mu)$ where we set $u_i \leq -\frac{1}{n}$ to $-\frac{1}{n}$, and $\mu > 0$ and $\alpha > 1$ are two constants selected to satisfy the constraints from step (ii).

7. $\max\limits_{\mathbb{Q}_n\in\mathcal{A}} \mathbb{E}_{\mathbb{Q}_n}[Z] \leq \mathbb{E}_{\mathbb{P}_n}[Z] + \lambda\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma}$

   $\min\limits_{\mathbb{Q}_n\in\mathcal{A}} \mathbb{E}_{\mathbb{Q}_n}[Z] \geq \mathbb{E}_{\mathbb{P}_n}[Z] - \lambda\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma}$.

   Proof: Combining the results of steps 3 to 6 with equality achieved when the constraint $u_i \geq -\frac{1}{n}$ is satisfied.

$\square$

Theorem 4. *Let $\ell(\theta,x,y)$ be an unbounded and non-negative loss function, and q_i the solution to $\min_{\mathbb{Q}_n\in\mathcal{A}} \mathbb{E}_{\mathbb{Q}_n}[\ell(\theta,x,y)]$ with $\mathcal{A} = \left\{\mathbb{Q}_n : D_{\chi^2}(\mathbb{Q}_n \,\|\, \mathbb{P}_n) \leq \frac{\lambda^2}{n^2}\mathbb{V}_{\mathbb{P}_n}[\ell(\theta,x,y)]^{2\gamma-1}\right\}$, then:*

$$0 \leq \sum_{j=1}^{C} q_i\ell(\theta,x_i,j) \leq \tfrac{C[1+\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[\ell(\theta,x,y)]^{\gamma-1}(\mathbb{E}_{\mathbb{P}_n}[\ell(\theta,x,y)]+\mu)]^2}{4n\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[\ell(\theta,x,y)]^{\gamma-1}}$$

where $\alpha \geq 1$ and $\mu \geq 0$ are two constants that emerge when finding the distribution $\mathbb{Q}_n$.

Proof.

Let Z be a random variable with the realizations $Z_i = \ell(\theta,x_i,y_i)$ then the probabilities q_i from Theorem 2 are calculated as $q_i = \frac{1}{n}\left[1-\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}\big(Z_i-\mathbb{E}_{\mathbb{P}_n}[Z]-\mu\big)\right]_+$ where $\alpha \geq 1$ and $\mu \geq 0$ are constants found when solving the stochastic optimization problem $\min_{\mathbb{Q}_n\in\mathcal{A}} \mathbb{E}_{\mathbb{Q}_n}[\ell(\theta,x,y)]$. The operator $[\cdot]_+$ truncates negative values to 0 and enforces the non-negativity constraint for q_i.

1. $0 \leq q_iZ_i$.

   Proof: The two terms are non-negative and thus their product is also non-negative.

2. $\dfrac{\partial}{\partial Z_i}(q_iZ_i) = \dfrac{1}{n}\left[1-\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}\big(2Z_i-\mathbb{E}_{\mathbb{P}_n}[Z]-\mu\big)\right]$.

   Proof: To find the maximum of q_iZ_i, the probability q_i must be positive and thus we drop the non-negativity constraint. Of note, since we are interested to find the distance from the mean where we achieve the maximum value given $\mathbb{E}_{\mathbb{P}_n}[Z]$ and $\mathbb{V}_{\mathbb{P}_n}[Z]$, we will we treat them as constants.

$$\begin{aligned}
\frac{\partial}{\partial Z_i}(q_iZ_i) &= q_i\frac{\partial}{\partial Z_i}Z_i + Z_i\frac{\partial}{\partial Z_i}q_i \\
&= \frac{1}{n}\left[1-\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}\big(Z_i-\mathbb{E}_{\mathbb{P}_n}[Z]-\mu\big)\right] - \frac{1}{n}\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}Z_i \\
&= \frac{1}{n}\left[1-\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}\big(2Z_i-\mathbb{E}_{\mathbb{P}_n}[Z]-\mu\big)\right].
\end{aligned}$$

3. $\dfrac{\partial}{\partial Z_i}(q_iZ_i) = 0$ for $Z_i = \dfrac{1}{2\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}} + \dfrac{\mathbb{E}_{\mathbb{P}_n}[Z]+\mu}{2}$.

   Proof:

$$\begin{aligned}
\frac{\partial}{\partial Z_i}(q_iZ_i) &= 0 \\
\frac{1}{n}\left[1-\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}\big(2Z_i-\mathbb{E}_{\mathbb{P}_n}[Z]-\mu\big)\right] &= 0 \\
2\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}Z_i &= 1+\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}(\mathbb{E}_{\mathbb{P}_n}[Z]+\mu) \\
Z_i &= \frac{1}{2\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}} + \frac{\mathbb{E}_{\mathbb{P}_n}[Z]+\mu}{2}
\end{aligned}$$

4. $q_i Z_i \leq \dfrac{[1+\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}(\mathbb{E}_{\mathbb{P}_n}[Z]+\mu)]^2}{4n\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}}$.

Proof: Since $\dfrac{\partial^2}{\partial Z_i^2}(q_i Z_i) = -2\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}$ is negative, the product $q_i Z_i$ is maximized when using for Z_i the value found at the previous step. Thus, we have:

$$
\begin{aligned}
q_i Z_i &\leq \frac{1}{n}\left[1-\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}\left(\frac{1}{2\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}}+\frac{\mathbb{E}_{\mathbb{P}_n}[Z]+\mu}{2}-\mathbb{E}_{\mathbb{P}_n}[Z]-\mu\right)\right]\left(\frac{1}{2\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}}+\frac{\mathbb{E}_{\mathbb{P}_n}[Z]+\mu}{2}\right)\\
&\leq \frac{1}{n}\left[1-\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}\left(\frac{1}{2\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}}-\frac{\mathbb{E}_{\mathbb{P}_n}[Z]+\mu}{2}\right)\right]\left(\frac{1}{2\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}}+\frac{\mathbb{E}_{\mathbb{P}_n}[Z]+\mu}{2}\right)\\
&\leq \frac{1}{n}\left[\frac{1}{2\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}}+\frac{\mathbb{E}_{\mathbb{P}_n}[Z]+\mu}{2}-\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}\left(\frac{1}{4(\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1})^2}-\frac{(\mathbb{E}_{\mathbb{P}_n}[Z]+\mu)^2}{4}\right)\right]\\
&\leq \frac{1}{n}\left[\frac{1}{2\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}}+\frac{\mathbb{E}_{\mathbb{P}_n}[Z]+\mu}{2}-\frac{1}{4\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}}+\frac{\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}(\mathbb{E}_{\mathbb{P}_n}[Z]+\mu)^2}{4}\right]\\
&\leq \frac{1}{n}\left[\frac{1}{4\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}}+\frac{\mathbb{E}_{\mathbb{P}_n}[Z]+\mu}{2}+\frac{\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}(\mathbb{E}_{\mathbb{P}_n}[Z]+\mu)^2}{4}\right]\\
&\leq \frac{1+2\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}(\mathbb{E}_{\mathbb{P}_n}[Z]+\mu)+(\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1})^2(\mathbb{E}_{\mathbb{P}_n}[Z]+\mu)^2}{4n\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}}\\
&\leq \frac{[1+\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}(\mathbb{E}_{\mathbb{P}_n}[Z]+\mu)]^2}{4n\lambda\alpha\mathbb{V}_{\mathbb{P}_n}[Z]^{\gamma-1}}
\end{aligned}
$$

□