where $\gamma_n := \mathbb{E}_q[\lambda_n^{-1}]$. Using the definition of $\zeta_n := \ell - y_n \mathbf{w}^\top \tilde{\mathbf{x}}_n$ and ignoring the constants, we have the simplified objective function (again without the ℓ_2-regularizer):

$$\begin{aligned}
\mathcal{L}_{[\mathbf{w}]} &= \sum_{n=1}^{N} \mathbb{E}_p \left[\frac{c^2}{2} \gamma_n \mathbf{w}^\top \tilde{\mathbf{x}}_n \tilde{\mathbf{x}}_n^\top \mathbf{w} - (c + \ell c^2 \gamma_n) y_n \mathbf{w}^\top \tilde{\mathbf{x}}_n \right] \\
&= \frac{c^2}{2} \sum_{n=1}^{N} \gamma_n \mathbb{E}_p \left[\mathbf{w}^\top \tilde{\mathbf{x}}_n \tilde{\mathbf{x}}_n^\top \mathbf{w} - 2 y_n^h \mathbf{w}^\top \tilde{\mathbf{x}}_n \right] \\
&= \frac{c^2}{2} \sum_{n=1}^{N} \gamma_n \mathbb{E}_p \left[(\mathbf{w}^\top \tilde{\mathbf{x}}_n - y_n^h)^2 \right], \qquad (24)
\end{aligned}$$

where $y_n^h := (\frac{1}{c\gamma_n} + \ell) y_n$ is the re-weighted label.

We now derive the equations to compute γ_n. Let x be a random variable and $y = f(x)$ is a function of x. Then, we have the transformation rule of probability distributions, $p(x) = p(f(x))|\frac{df(x)}{dx}|$. For our case, let $x = \lambda_n$, and $f(x) = \frac{1}{\lambda_n}$, we have $q(\lambda_n) = \frac{1}{\lambda_n^2} q(\frac{1}{\lambda_n})$. Then

$$\begin{aligned}
\mathbb{E}_{q(\lambda_n)}[\lambda_n^{-1}] &= \int_0^\infty q(\lambda_n) \frac{1}{\lambda_n} d\lambda_n \\
&= \int_0^\infty q\left(\frac{1}{\lambda_n}\right) \frac{1}{\lambda_n^3} d\lambda_n \\
&= \int_\infty^0 q(\mu_n) \mu_n^3 d\mu_n^{-1} \ \ (\text{define } \mu_n = \frac{1}{\lambda_n}) \\
&= \int_0^\infty q(\mu_n) \mu_n d\mu_n \\
&= \mathbb{E}_{q(\lambda_n^{-1})}[\lambda_n^{-1}]. \qquad (25)
\end{aligned}$$

Since $q(\lambda_n^{-1})$ is an inverse Gaussian distribution as shown in Eq. (11), it is easy to get

$$\mathbb{E}_{q(\lambda_n)}[\lambda_n^{-1}] = \mathbb{E}_{q(\lambda_n^{-1})}[\lambda_n^{-1}] = \frac{1}{c\sqrt{\mathbb{E}[\zeta_n^2]}}. \qquad (26)$$

Combining the above results finishes the proof of Lemma 1. □