Revisions needed for Xintong's thesis (Ben's notes)

1. Add sufficient details to **every** numerical experiments. (suggested by Emil)
   (a) Dynamic asset model (e.g., Black-Scholes, regime-switching, etc.)
   (b) All model parameters (i.e., risk-free rates, volatilities, correlations in all regimes, etc.).
   (c) Computing platforms, if applicable (i.e., personal laptop, desktop, university servers, etc.)
2. Add an appendix to summarize all acronyms (suggested by Emil)
   (a) I think at least two summary tables are needed: One for different variable annuity riders (i.e., GMWB, GMMB, GLWB, etc.) and one for different machine learning models (i.e., QPR, RNN, LSTM, etc.).
   (b) You might want to add a third table to summarize other acronyms that don't fall into the above two categories.
3. Clarify expected values, sample average as an estimator of the expected value, and pathwise sample of an expected value. (suggested by Yuying). This is **a major issue and a high-priority item** that must be properly addressed. Tony has offered some helpful insights in his notes. I give mine too below.
   (a) Use GMMB as an example to explain the missing expectation that Yuying had in mind: An insurer's time-t liability for a GMMB contract

$$V_t = V(\boldsymbol{S}_t) = E[e^{-r(T-t)}(G_T - F_T)^+ - \sum_{s=t+1}^{T} e^-r(T-s)F_s\eta_n|\boldsymbol{S}_t]$$

   where
      i. $\boldsymbol{S}_t$ is the (outer) stock path up to time t
      ii. The (conditional) expectation is taken with respect to the inner sample path $\widetilde{S}_{t+1}, \ldots, \widetilde{S}_T$ given the outer path $\boldsymbol{S}_t$

   By removing the (conditional) expectation, I believe you meant the time-t liability for one simulate inner path. But the value for one simulated path does not equal to the (conditional) expected value.

   Please do a global check. When revising the thesis, please try to understand readers' perspective. A reader who has not gone through this project would easily be confused. So your writing needs to be very clear in every detail.
   (b) Similar idea for the pathwise delta

$$\begin{aligned}
\Delta_t(\widetilde{S}_{t+1}, \ldots, \widetilde{S}_T|\boldsymbol{S}_t) &= \frac{\partial V_t}{\partial S_t} \\
&= \frac{\partial}{\partial S_t} V(\boldsymbol{S}_t) \\
&= \frac{\partial}{\partial S_t} E[e^{-r(T-t)}(G_T - F_T)^+ - \sum_{s=t+1}^{T} e^-r(T-s)F_s\eta_n|\boldsymbol{S}_t] \\
&\overset{*}{=} E[\frac{\partial}{\partial S_t} e^{-r(T-t)}(G_T - F_T)^+ - \sum_{s=t+1}^{T} e^-r(T-s)F_s\eta_n|\boldsymbol{S}_t]
\end{aligned}$$

   A few things needs to be clarified
      i. The value $V(\boldsymbol{S}_t)$ is a function of $\boldsymbol{S}_t = (S_1, \ldots, S_t)$, but you are only taking derivative with respect to the last value S_t

ii. You are passing through the partial derivative inside the expectation. This is the technique called infinitesimal perturbation analysis (IPA). This technique allows you to take derivative for each inner sample path first (i.e., the pathwise derivative) then use the sample average of the pathwise derivative to approximate the (conditional) expected value.

(c) In my opinions, you should also clarify how the recursive pathwise derivatives (on page 73) are derived, i.e., based on the account evolution on page 72.

4. Clarify "What is the neural network approximating?" (suggested by Yuying).

   (a) The neural network approximates L in Equation (3.4), where

   $$L = L(\boldsymbol{S}_T) = L(\Delta_0, \ldots, \Delta_{T-1}, S_0, \ldots, S_T).$$

   (b) The loss L depends on all $\Delta_0, \ldots, \Delta_{T-1}$ and each Δ_t depends on $\boldsymbol{S}_t$, as discussed above.

   (c) The loss also depends on all $S_0, \ldots, S_T$ directly, too.

   (d) Since $L(\boldsymbol{S}_T)$ is a function of the **entire** outer path, the neural network $\widehat{L}(\boldsymbol{S}_T)$ approximates L and is also a function of the entire outer path.

   (e) Statistically speaking, $L(\boldsymbol{S}_T)$ is a random variable because $\boldsymbol{S}_T$ is random. Nonetheless, we can see $L(\boldsymbol{S}_T)$ as a function of the underlying random path $\boldsymbol{S}_T$ and approximate this function.

   (f) You should revise not only the math equations, but also paragraphs leading up to the equations and those after the equations.

   (g) Please do a careful global revision rather than only the two places mentioned above.

5. Remove reference for regression or use a different reference. (suggested by Mary Hardy).

   (a) WILL ADD MORE DETAILS LATER. I do not remember which specific reference she was referring to.

6. Rewrite the Multi-level Monte Carlo sections (suggested by Mary Hardy)

   (a) Clarify the descriptions for multi-level Monte Carlo

   (b) Add more details in the numerical experiments for readers to replicate the results

7. Add a line between the following inequality explaining the use of Cauchy-Schwarz inequality:

   $$E[(\hat{\rho}_{M,N} - \rho)^2] \leq 2E[(\hat{\rho}_{M,N} - \rho_M)^2] + 2E[(\rho_M - \rho)^2].$$

   One way to express the Cauchy-Schwarz inequality is (you will need some reference)

   $$(x_1 y_1 + x_2 y_2)^2 \leq (x_1^2 + x_2^2)(y_1^2 + y_2^2)$$

   So setting $x_1 = \hat{\rho}_{M,N} - \rho_M$, $x_2 = \rho_M - \rho$, $y_1 = y_2 = 1$, we get

   $$\begin{aligned}
   (\hat{\rho}_{M,N} - \rho)^2 &= [(\hat{\rho}_{M,N} - \rho_M) + (\rho_M - \rho)]^2 \\
   &= (x_1 y_1 + x_2 y_2)^2 \\
   &\leq (x_1^2 + x_2^2)(y_1^2 + y_2^2) \\
   &= 2(\hat{\rho}_{M,N} - \rho_M)^2 + 2(\rho_M - \rho)^2
   \end{aligned}$$

8. Variable z needs to be clearly defined when applying Taylor expansion. (suggested by Chengguo)

   (a) Chengguo mentioned Equation (2.16), but you should do a global check to make sure all Taylor expansions are clearly explained.

9. Rewrite section 2.3.1 and re-evaluate the contributions of Theorem 1, specially in light of the well-know statistical property that L2-convergence implies convergence in probability. This was a main issue raised by Chengguo during the defence. Questions about this part were not answered well. Thus this is a **high-priority item**.
   (a) Chengguo suggested (1) clarifying the definitions (4 & 5), (2) cleaning up the proof by applying existing statistical results when possible, (3) re-evaluating the theoretical contributions of Theorem 1 and reposition it (i.e., tone down) if appropriate.

10. Minor items
    (a) Move copyright to first page.
    (b) Add a paragraph under the "sole-author" declaration that part of this thesis has been published in a WSC proceeding.