Derivation of the Gelman-Rubin Variance Estimator

The derivation follows these steps:

1. **Equation (2)** shows that the sample variance s_i^2 is biased for the target variance σ^2:
$$E_F[s_i^2] = \frac{n}{n-1}\left(\sigma^2 - \mathrm{Var}_F(\bar{X}_{i\cdot})\right) \tag{1}$$

2. **Key insight**: For correlated MCMC samples, $\mathrm{Var}_F(\bar{X}_{i\cdot})$ is much larger than σ^2/n (which would be the case for independent samples). This means s_i^2 systematically underestimates σ^2.

3. **Rearranging equation (2)** to solve for σ^2:
$$\sigma^2 = \frac{n-1}{n}E_F[s_i^2] + \mathrm{Var}_F(\bar{X}_{i\cdot}) \tag{2}$$

4. **Equation (3)** provides an estimator for $\mathrm{Var}_F(\bar{X}_{i\cdot})$:
$$\frac{B}{n} = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{X}_{i\cdot} - \hat{\mu})^2 \tag{3}$$

   This is the sample variance of the m chain means, which estimates the variance of $\bar{X}_{i\cdot}$.

5. **Final estimator**: Substituting the sample quantities:

   - Use s^2 (the average of the m sample variances) to estimate $E_F[s_i^2]$
   - Use B/n to estimate $\mathrm{Var}_F(\bar{X}_{i\cdot})$

   This gives:
$$\boxed{\hat{\sigma}^2 = \frac{n-1}{n}s^2 + \frac{B}{n}} \tag{4}$$

Intuition: Gelman-Rubin corrects for the downward bias in s^2 by adding back an estimate of the between-chain variance (B/n), which captures the additional variability due to correlation in the Markov chains.