Derivation of Expected Sample Variance for One Chain

Goal: Find $\mathbb{E}\left[\frac{1}{T-1}\sum_{t=1}^{T}(X_{m,t} - \bar{X}_m)^2\right]$

Setup

- $X_{m,t}$ = sample t from chain m
- $\bar{X}_m = \frac{1}{T}\sum_{t=1}^{T} X_{m,t}$ = sample mean
- $\mu = \mathbb{E}[X_{m,t}]$ = true mean
- $\sigma^2 = \text{Var}(X_{m,t})$ = true variance

Step 1: Rewrite the deviation from sample mean

$$\sum_{t=1}^{T}(X_{m,t} - \bar{X}_m)^2 = \sum_{t=1}^{T}\left[(X_{m,t} - \mu) - (\bar{X}_m - \mu)\right]^2 \tag{1}$$

Step 2: Expand the square

$$= \sum_{t=1}^{T}\left[(X_{m,t} - \mu)^2 - 2(X_{m,t} - \mu)(\bar{X}_m - \mu) + (\bar{X}_m - \mu)^2\right] \tag{2}$$

$$= \sum_{t=1}^{T}(X_{m,t} - \mu)^2 - 2(\bar{X}_m - \mu)\sum_{t=1}^{T}(X_{m,t} - \mu) + T(\bar{X}_m - \mu)^2 \tag{3}$$

Step 3: Simplify the middle term

Note that:

$$\sum_{t=1}^{T}(X_{m,t} - \mu) = \sum_{t=1}^{T} X_{m,t} - T\mu \tag{4}$$

$$= T\bar{X}_m - T\mu \tag{5}$$

$$= T(\bar{X}_m - \mu) \tag{6}$$

Therefore:

$$\sum_{t=1}^{T}(X_{m,t} - \bar{X}_m)^2 = \sum_{t=1}^{T}(X_{m,t} - \mu)^2 - 2T(\bar{X}_m - \mu)^2 + T(\bar{X}_m - \mu)^2 \tag{7}$$

$$= \sum_{t=1}^{T}(X_{m,t} - \mu)^2 - T(\bar{X}_m - \mu)^2 \tag{8}$$

1

Step 4: Take expectations

$$\mathbb{E}\left[\sum_{t=1}^{T}(X_{m,t} - \bar{X}_m)^2\right] = \mathbb{E}\left[\sum_{t=1}^{T}(X_{m,t} - \mu)^2\right] - T \cdot \mathbb{E}\left[(\bar{X}_m - \mu)^2\right] \quad (9)$$

$$= \sum_{t=1}^{T} \mathbb{E}\left[(X_{m,t} - \mu)^2\right] - T \cdot \operatorname{Var}(\bar{X}_m) \quad (10)$$

$$= T\sigma^2 - T \cdot \operatorname{Var}(\bar{X}_m) \quad (11)$$

$$= T\left(\sigma^2 - \operatorname{Var}(\bar{X}_m)\right) \quad (12)$$

Step 5: Divide by $(T-1)$

$$\boxed{\mathbb{E}\left[\frac{1}{T-1}\sum_{t=1}^{T}(X_{m,t} - \bar{X}_m)^2\right] = \frac{T}{T-1}\left(\sigma^2 - \operatorname{Var}(\bar{X}_m)\right)} \quad (13)$$

Key Insight

For i.i.d. samples: $\operatorname{Var}(\bar{X}_m) = \sigma^2/T$, which gives:

$$\frac{T}{T-1}\left(\sigma^2 - \frac{\sigma^2}{T}\right) = \frac{T}{T-1} \cdot \frac{(T-1)\sigma^2}{T} = \sigma^2 \quad (14)$$

So the sample variance is unbiased.

For MCMC samples: Due to autocorrelation, $\operatorname{Var}(\bar{X}_m) > \sigma^2/T$, which means:

$$\mathbb{E}\left[\frac{1}{T-1}\sum_{t=1}^{T}(X_{m,t} - \bar{X}_m)^2\right] < \sigma^2 \quad (15)$$

Therefore, W **underestimates the true variance** σ^2.

Derivation of $\mathbb{E}[B]$

Given:

$$B = \frac{1}{M-1}\sum_{m=1}^{M}(\bar{X}_m - \bar{X}_{..})^2 \quad (16)$$

where:

- $\bar{X}_m = \frac{1}{T}\sum_{t=1}^{T} X_{m,t}$ = mean of chain m

- $\bar{X}_{..} = \frac{1}{M}\sum_{m=1}^{M} \bar{X}_m$ = overall mean across all chains

- $\mu = \mathbb{E}[X_{m,t}]$ = true mean

Step 1: Rewrite deviations from overall mean

$$\sum_{m=1}^{M} (\bar{X}_m - \bar{X}_{..})^2 = \sum_{m=1}^{M} \left[(\bar{X}_m - \mu) - (\bar{X}_{..} - \mu) \right]^2 \tag{17}$$

Step 2: Expand the square

$$= \sum_{m=1}^{M} \left[(\bar{X}_m - \mu)^2 - 2(\bar{X}_m - \mu)(\bar{X}_{..} - \mu) + (\bar{X}_{..} - \mu)^2 \right] \tag{18}$$

$$= \sum_{m=1}^{M} (\bar{X}_m - \mu)^2 - 2(\bar{X}_{..} - \mu) \sum_{m=1}^{M} (\bar{X}_m - \mu) + M(\bar{X}_{..} - \mu)^2 \tag{19}$$

Step 3: Simplify the middle term

Note that:

$$\sum_{m=1}^{M} (\bar{X}_m - \mu) = \sum_{m=1}^{M} \bar{X}_m - M\mu \tag{20}$$

$$= M\bar{X}_{..} - M\mu \tag{21}$$

$$= M(\bar{X}_{..} - \mu) \tag{22}$$

Therefore:

$$\sum_{m=1}^{M} (\bar{X}_m - \bar{X}_{..})^2 = \sum_{m=1}^{M} (\bar{X}_m - \mu)^2 - 2M(\bar{X}_{..} - \mu)^2 + M(\bar{X}_{..} - \mu)^2 \tag{23}$$

$$= \sum_{m=1}^{M} (\bar{X}_m - \mu)^2 - M(\bar{X}_{..} - \mu)^2 \tag{24}$$

Step 4: Take expectations

$$\mathbb{E}\left[\sum_{m=1}^{M} (\bar{X}_m - \bar{X}_{..})^2 \right] = \mathbb{E}\left[\sum_{m=1}^{M} (\bar{X}_m - \mu)^2 \right] - M \cdot \mathbb{E}\left[(\bar{X}_{..} - \mu)^2 \right] \tag{25}$$

$$= \sum_{m=1}^{M} \mathbb{E}\left[(\bar{X}_m - \mu)^2 \right] - M \cdot \mathbb{E}\left[(\bar{X}_{..} - \mu)^2 \right] \tag{26}$$

$$= M \cdot \mathrm{Var}(\bar{X}_m) - M \cdot \mathrm{Var}(\bar{X}_{..}) \tag{27}$$

(assuming all chains have the same variance of their means)

Step 5: Find $\text{Var}(\bar{X}_{..})$

Since $\bar{X}_{..} = \frac{1}{M} \sum_{m=1}^{M} \bar{X}_m$ and assuming chains are independent:

$$\text{Var}(\bar{X}_{..}) = \text{Var}\left(\frac{1}{M} \sum_{m=1}^{M} \bar{X}_m \right) \tag{28}$$

$$= \frac{1}{M^2} \sum_{m=1}^{M} \text{Var}(\bar{X}_m) \tag{29}$$

$$= \frac{1}{M^2} \cdot M \cdot \text{Var}(\bar{X}_m) \tag{30}$$

$$= \frac{1}{M} \text{Var}(\bar{X}_m) \tag{31}$$

Step 6: Substitute back

$$\mathbb{E}\left[\sum_{m=1}^{M} (\bar{X}_m - \bar{X}_{..})^2 \right] = M \cdot \text{Var}(\bar{X}_m) - M \cdot \frac{1}{M} \text{Var}(\bar{X}_m) \tag{32}$$

$$= M \cdot \text{Var}(\bar{X}_m) - \text{Var}(\bar{X}_m) \tag{33}$$

$$= (M-1) \text{Var}(\bar{X}_m) \tag{34}$$

Step 7: Divide by $(M-1)$

$$\boxed{\mathbb{E}[B] = \frac{1}{M-1} \cdot (M-1) \text{Var}(\bar{X}_m) = \text{Var}(\bar{X}_m)} \tag{35}$$

Conclusion

This confirms what's stated on slide 6: $\mathbb{E}[B] = \textbf{Var}(\bar{X}_m)$

The between-chain variance B is an unbiased estimator of the variance of the chain means!

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:

$$E_F[s_i^2] = \frac{n}{n-1} \left(\sigma^2 - \text{Var}_F(\bar{X}_{i\cdot}) \right) \tag{36}$$

2. **Key insight**: For correlated MCMC samples, $\text{Var}_F(\bar{X}_{i\cdot})$ 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:

$$\sigma^2 = \frac{n-1}{n} E_F[s_i^2] + \text{Var}_F(\bar{X}_{i\cdot}) \tag{37}$$

4. **Equation (3)** provides an estimator for $\text{Var}_F(\bar{X}_{i\cdot})$:

$$\frac{B}{n} = \frac{1}{m-1} \sum_{i=1}^{m} (\bar{X}_{i\cdot} - \hat{\mu})^2 \tag{38}$$

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 $\text{Var}_F(\bar{X}_{i\cdot})$

   This gives:

$$\boxed{\hat{\sigma}^2 = \frac{n-1}{n} s^2 + \frac{B}{n}} \tag{39}$$

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.