

Formula Sheet for the Midterm

ELG 5218 Uncertainty Evaluation in
Engineering Measurements and Machine
Learning

Instructor: Miodrag Bolić, University of Ottawa

1 Core Bayesian Identities

1.1 Bayes' Theorem

$$p(\theta | D) = \frac{p(D | \theta) p(\theta)}{p(D)}, \quad (1)$$

$$p(D) = \int p(D | \theta) p(\theta) d\theta. \quad (2)$$

Posterior is often written up to proportionality:

$$p(\theta | D) \propto p(D | \theta) p(\theta). \quad (3)$$

1.2 Likelihood for IID Data

For IID data $D = \{x_i\}_{i=1}^N$:

$$p(D | \theta) = \prod_{i=1}^N p(x_i | \theta), \quad (4)$$

$$\log p(D | \theta) = \sum_{i=1}^N \log p(x_i | \theta). \quad (5)$$

1.3 MLE and MAP

Maximum Likelihood Estimate (MLE):

$$\hat{\theta}_{\text{MLE}} = \arg \max_{\theta} \log p(D | \theta). \quad (6)$$

Maximum A Posteriori (MAP):

$$\hat{\theta}_{\text{MAP}} = \arg \max_{\theta} \log p(D | \theta) + \log p(\theta). \quad (7)$$

1.4 Posterior Predictive Distribution

$$p(x_* | D) = \int p(x_* | \theta) p(\theta | D) d\theta. \quad (8)$$

A plug-in (approximate) predictive:

$$p(x_* | D) \approx p(x_* | \hat{\theta}). \quad (9)$$

2 Beta–Binomial Model

2.1 Model

$$k | \theta \sim \text{Binomial}(n, \theta), \quad (10)$$

$$\theta \sim \text{Beta}(\alpha, \beta). \quad (11)$$

2.2 Posterior

$$\theta | k, n \sim \text{Beta}(\alpha + k, \beta + n - k). \quad (12)$$

2.3 Moments and MAP

Posterior mean:

$$\mathbb{E}[\theta | k, n] = \frac{\alpha + k}{\alpha + \beta + n}. \quad (13)$$

Posterior variance:

$$\text{Var}[\theta | k, n] = \frac{(\alpha + k)(\beta + n - k)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)}. \quad (14)$$

MAP estimator (for $\alpha, \beta > 1$):

$$\hat{\theta}_{\text{MAP}} = \frac{k + \alpha - 1}{n + \alpha + \beta - 2}. \quad (15)$$

3 Gaussian Distributions and Identities

3.1 Scalar Gaussian

$$p(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \quad (16)$$

Precision form with $\lambda = 1/\sigma^2$:

$$p(x) \propto \sqrt{\lambda} \exp\left(-\frac{\lambda}{2}(x - \mu)^2\right). \quad (17)$$

3.2 Multivariate Gaussian

$$p(z) = \mathcal{N}(z; \mu, \Sigma) \quad (18)$$

$$= \frac{1}{(2\pi)^{D/2} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(z - \mu)^\top \Sigma^{-1} (z - \mu)\right). \quad (19)$$

Precision form with $\Lambda = \Sigma^{-1}$:

$$p(z) \propto |\Lambda|^{1/2} \exp\left(-\frac{1}{2}(z - \mu)^\top \Lambda (z - \mu)\right). \quad (20)$$

3.3 Partitioned Gaussian: Marginal and Conditional

Joint Gaussian:

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}\right). \quad (21)$$

Marginal of x :

$$x \sim \mathcal{N}(\mu_x, \Sigma_{xx}). \quad (22)$$

3.4 Theorem 2: Conditional of x given y

$$p(\mathbf{x} | \mathbf{y}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{x|y}, \boldsymbol{\Sigma}_{x|y}), \quad (23)$$

$$\boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} (\mathbf{y} - \boldsymbol{\mu}_y), \quad (24)$$

$$\boldsymbol{\Sigma}_{x|y} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yx}. \quad (25)$$

This is Theorem 2 (conditioning) expressed with (\mathbf{x}, \mathbf{y}) .

3.5 Theorem 3: Affine Transformation

Assume

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x), \quad (26)$$

$$p(\mathbf{y} | \mathbf{x}) = \mathcal{N}(\mathbf{y}; \mathbf{A}\mathbf{x} + \mathbf{b}, \boldsymbol{\Sigma}_{y|x}), \quad (27)$$

where the conditional $\mathbf{y} | \mathbf{x}$ is a linear function of \mathbf{x} plus Gaussian noise. Then the stacked vector $(\mathbf{x}, \mathbf{y})^\top$ is jointly Gaussian:

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}; \begin{pmatrix} \boldsymbol{\mu}_x \\ \mathbf{A}\boldsymbol{\mu}_x + \mathbf{b} \end{pmatrix}, \mathbf{R}\right), \quad (28)$$

with block covariance

$$\mathbf{R} = \begin{pmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_x \mathbf{A}^\top \\ \mathbf{A}\boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{y|x} + \mathbf{A}\boldsymbol{\Sigma}_x \mathbf{A}^\top \end{pmatrix}. \quad (29)$$

Corollary 2: Marginal of y (Predictive Distribution)

$$p(\mathbf{y}) = \int p(\mathbf{y} | \mathbf{x}) p(\mathbf{x}) d\mathbf{x} = \mathcal{N}(\mathbf{y}; \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y), \quad (30)$$

where

$$\boldsymbol{\mu}_y = \mathbf{A}\boldsymbol{\mu}_x + \mathbf{b}, \quad (31)$$

$$\boldsymbol{\Sigma}_y = \boldsymbol{\Sigma}_{y|x} + \mathbf{A}\boldsymbol{\Sigma}_x \mathbf{A}^\top. \quad (32)$$

Corollary 1: Posterior of xx given yy

$p(\mathbf{x} | \mathbf{y})$:

$$p(\mathbf{x} | \mathbf{y}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{x|y}, \boldsymbol{\Sigma}_{x|y}), \quad (33)$$

with

$$\boldsymbol{\Sigma}_{x|y} = \left(\boldsymbol{\Sigma}_x^{-1} + \mathbf{A}^\top \boldsymbol{\Sigma}_{y|x}^{-1} \mathbf{A} \right)^{-1}, \quad (34)$$

$$\boldsymbol{\mu}_{x|y} = \boldsymbol{\Sigma}_{x|y} \left(\boldsymbol{\Sigma}_x^{-1} \boldsymbol{\mu}_x + \mathbf{A}^\top \boldsymbol{\Sigma}_{y|x}^{-1} (\mathbf{y} - \mathbf{b}) \right). \quad (35)$$

3.6 Product and Ratio of Scalar Gaussians

Product: if

$$\mu_1(x) \propto \mathcal{N}(x; m_1, \sigma_1^2), \quad (36)$$

$$\mu_2(x) \propto \mathcal{N}(x; m_2, \sigma_2^2), \quad (37)$$

then

$$\mu_1(x)\mu_2(x) \propto \mathcal{N}(x; m, \sigma^2), \quad (38)$$

$$\sigma^2 = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)^{-1}, \quad (39)$$

$$m = \sigma^2 \left(\frac{m_1}{\sigma_1^2} + \frac{m_2}{\sigma_2^2} \right). \quad (40)$$

Ratio: with same μ_1, μ_2 ,

$$\frac{\mu_1(x)}{\mu_2(x)} \propto \mathcal{N}(x; \tilde{m}, \tilde{\sigma}^2), \quad (41)$$

$$\tilde{\sigma}^2 = \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} \right)^{-1}, \quad (42)$$

$$\tilde{m} = \tilde{\sigma}^2 \left(\frac{m_1}{\sigma_1^2} - \frac{m_2}{\sigma_2^2} \right). \quad (43)$$

4 Gaussian Models and Conjugate Updates

4.1 4.1 Known Variance, Unknown Mean (Scalar)

Model: $x_i | \mu, \sigma^2 \mathcal{N}(\mu, \sigma^2)$, known σ^2 , precision $\lambda = 1/\sigma^2$.

Uniform Prior $p(\mu) \propto 1$:

$$\mu | x_{1:n} \sim \mathcal{N}\left(\bar{x}, \frac{\sigma^2}{n}\right), \quad (44)$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

Normal Prior $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$ with precision $\lambda_0 = 1/\sigma_0^2$:

$$\lambda_n = \lambda_0 + n\lambda, \quad (45)$$

$$\mu_n = \frac{\lambda_0 \mu_0 + n\lambda \bar{x}}{\lambda_n}, \quad (46)$$

$$\mu | x_{1:n} \sim \mathcal{N}\left(\mu_n, \frac{1}{\lambda_n}\right). \quad (47)$$

4.2 Sequential (Online) Update for Mean

Let μ_{n-1}, λ_{n-1} be posterior parameters after $n-1$ points; add x_n .

$$\lambda_n = \lambda_{n-1} + \lambda, \quad (48)$$

$$w_n = \frac{\lambda}{\lambda_n}, \quad (49)$$

$$\mu_n = \mu_{n-1} + w_n(x_n - \mu_{n-1}). \quad (50)$$

4.3 Unknown Mean and Variance: Normal–Gamma Prior

Use precision $\lambda = 1/\sigma^2$.

$$\mu | \lambda \sim \mathcal{N}(\mu_0, (\kappa_0 \lambda)^{-1}), \quad (51)$$

$$\lambda \sim \text{Gamma}(\alpha_0, \beta_0). \quad (52)$$

Posterior hyperparameters after n observations with sample mean \bar{x} :

$$\kappa_n = \kappa_0 + n, \quad (53)$$

$$\mu_n = \frac{\kappa_0 \mu_0 + n \bar{x}}{\kappa_n}, \quad (54)$$

$$\alpha_n = \alpha_0 + \frac{n}{2}, \quad (55)$$

$$\beta_n = \beta_0 + \frac{1}{2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 + \frac{\kappa_0 n}{\kappa_0 + n} (\bar{x} - \mu_0)^2 \right). \quad (56)$$

Posterior:

$$p(\mu, \lambda | x_{1:n}) = \mathcal{N}(\mu; \mu_n, (\kappa_n \lambda)^{-1}) \text{ Gamma}(\lambda; \alpha_n, \beta_n). \quad (57)$$

4.4 Student-t Distribution from Normal–Gamma

If μ, λ have Normal–Gamma posterior as above, the marginal for μ is Student- t :

$$\mu | x_{1:n} \sim t_{2\alpha_n} \left(\mu_n, \sqrt{\frac{\beta_n}{\alpha_n \kappa_n}} \right). \quad (58)$$

Posterior predictive (new x_*):

$$x_* | x_{1:n} \sim t_{2\alpha_n} \left(\mu_n, \sqrt{\frac{\beta_n}{\alpha_n} \left(1 + \frac{1}{\kappa_n} \right)} \right). \quad (59)$$

5 Logistic Regression and Laplace Approximation

5.1 Logistic Model

For binary labels $y_n \in \{0, 1\}$ and features x_n :

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \quad (60)$$

$$p(y_n = 1 | x_n, w) = \sigma(w^\top x_n), \quad (61)$$

$$p(y_n = 0 | x_n, w) = 1 - \sigma(w^\top x_n). \quad (62)$$

Likelihood (Bernoulli factorization):

$$p(y | X, w) = \prod_{n=1}^N \sigma(w^\top x_n)^{y_n} (1 - \sigma(w^\top x_n))^{1-y_n}. \quad (63)$$

Gaussian prior on weights:

$$p(w) = \mathcal{N}(w; 0, \alpha^{-1} I). \quad (64)$$

Posterior (up to proportionality):

$$p(w | y, X) \propto p(y | X, w) p(w). \quad (65)$$

5.2 Log-Posterior, Gradient, Hessian

Let $\mu_n = \sigma(w^\top x_n)$ and define $\mu = (\mu_1, \dots, \mu_N)^\top$. Log-posterior (up to constant):

$$\begin{aligned} \log p(w | D) &= -\frac{\alpha}{2} w^\top w + \sum_{n=1}^N \left[y_n \log \mu_n \right. \\ &\quad \left. + (1 - y_n) \log(1 - \mu_n) \right] + C. \end{aligned} \quad (66)$$

Gradient:

$$\nabla_w \log p(w | D) = -\alpha w + \sum_{n=1}^N (y_n - \mu_n) x_n \quad (67)$$

$$= -\alpha w + X^\top (y - \mu). \quad (68)$$

Hessian:

$$H = \nabla_w^2 \log p(w | D) \quad (69)$$

$$= -\alpha I - \sum_{n=1}^N \mu_n (1 - \mu_n) x_n x_n^\top \quad (70)$$

$$= -(\alpha I + X^\top S X), \quad (71)$$

where S is diagonal with $S_{nn} = \mu_n(1 - \mu_n)$.

5.3 Laplace Approximation for Posterior

Mode (MAP):

$$w_{\text{MAP}} = \arg \max_w \log p(w | D). \quad (72)$$

Quadratic (Laplace) approximation around w_{MAP} :

$$p(w | D) \approx \mathcal{N}(w; w_{\text{MAP}}, \Sigma_N), \quad (73)$$

$$\Sigma_N^{-1} = -H|_{w=w_{\text{MAP}}} = \alpha I + X^\top S X. \quad (74)$$

5.4 Predictive Probability (Approximate)

Exact predictive for new x_* :

$$p(y_* = 1 | x_*, D) = \int \sigma(w^\top x_*) p(w | D) dw. \quad (75)$$

Under Laplace approximation $p(w | D) \approx \mathcal{N}(w_{\text{MAP}}, \Sigma_N)$ this is a 1D Gaussian integral in $w^\top x_*$ (often evaluated numerically or via standard approximations).

5.5 Potential Energy for Logistic Regression

Define unnormalized posterior $\pi(w) \propto p(y | X, w)p(w)$. Potential energy:

$$U(w) = -\log \pi(w) \quad (76)$$

$$\begin{aligned} &= -\sum_{n=1}^N [y_n \log \mu_n + (1 - y_n) \log(1 - \mu_n)] \\ &\quad + \frac{\alpha}{2} w^\top w + C. \end{aligned} \quad (77)$$

Gradient of $U(w)$:

$$\nabla U(w) = X^\top (\mu - y) + \alpha w. \quad (78)$$

6 Langevin Dynamics and MALA

6.1 Langevin SDE

For a target density $p(x)$ with score $\nabla_x \log p(x)$, the Langevin stochastic differential equation is

$$dX_t = \frac{1}{2} \nabla_x \log p(X_t) dt + dW_t, \quad (79)$$

where W_t is standard Brownian motion.

6.2 Potential Form

For a target $\pi(\theta) \propto e^{-U(\theta)}$:

$$d\Theta_t = -\frac{1}{2} \nabla U(\Theta_t) dt + dW_t. \quad (80)$$

6.3 Unadjusted Langevin Algorithm (ULA)

Euler–Maruyama discretization with step size Δt :

$$\theta_{k+1} = \theta_k - \frac{1}{2} \nabla U(\theta_k) \Delta t + \sqrt{\Delta t} Z_k, \quad (81)$$

$$Z_k \sim \mathcal{N}(0, I). \quad (82)$$

6.4 Metropolis-Adjusted Langevin Algorithm (MALA)

Langevin proposal:

$$q(\tilde{\theta} | \theta) = \mathcal{N}\left(\tilde{\theta}; \theta - \frac{1}{2} \nabla U(\theta) \Delta t, \Delta t I\right). \quad (83)$$

Metropolis–Hastings acceptance probability:

$$\alpha(\theta \rightarrow \tilde{\theta}) = \min\left(1, \frac{\pi(\tilde{\theta}) q(\theta | \tilde{\theta})}{\pi(\theta) q(\tilde{\theta} | \theta)}\right) \quad (84)$$

$$= \min\left(1, \exp(-U(\tilde{\theta}) + U(\theta)) \frac{q(\theta | \tilde{\theta})}{q(\tilde{\theta} | \theta)}\right). \quad (85)$$

7 Markov Chain Monte Carlo (MCMC)

7.1 Metropolis–Hastings Algorithm

Given target density $\pi(\theta)$ and proposal $q(\theta' | \theta)$:

$$r(\theta \rightarrow \theta') = \frac{\tilde{\pi}(\theta') q(\theta | \theta')}{\tilde{\pi}(\theta) q(\theta' | \theta)}, \quad (86)$$

$$\alpha(\theta \rightarrow \theta') = \min(1, r(\theta \rightarrow \theta')), \quad (87)$$

where $\tilde{\pi}$ is any unnormalized version of π . If $\pi(\theta) \propto p(x | \theta)p(\theta)$, then

$$r(\theta \rightarrow \theta') = \frac{p(x | \theta') p(\theta') q(\theta | \theta')}{p(x | \theta) p(\theta) q(\theta' | \theta)}, \quad (88)$$

so the evidence cancels.

7.2 Random-Walk Metropolis (RWM)

Symmetric proposal:

$$q(\theta' | \theta) = \mathcal{N}(\theta'; \theta, \sigma_{\text{prop}}^2 I), \quad (89)$$

so acceptance probability simplifies to

$$\alpha(\theta \rightarrow \theta') = \min \left(1, \frac{\pi(\theta')}{\pi(\theta)} \right). \quad (90)$$

7.3 Detailed Balance Condition

A sufficient condition for stationarity of π is

$$\pi(\theta)P(\theta' | \theta) = \pi(\theta')P(\theta | \theta') \quad (91)$$

for all θ, θ' . Metropolis–Hastings transition kernel satisfies detailed balance.

7.4 Monte Carlo Estimation

Given samples $\{\theta^{(s)}\}_{s=1}^S$ from $\pi(\theta)$, an expectation

$$\mathbb{E}_\pi[f(\theta)] = \int f(\theta) \pi(\theta) d\theta \quad (92)$$

can be approximated as

$$\mathbb{E}_\pi[f(\theta)] \approx \frac{1}{S} \sum_{s=1}^S f(\theta^{(s)}). \quad (93)$$

Standard error decays as $\sim 1/\sqrt{S}$ (modulo autocorrelation).

8 Uncertainty Quantification and Intervals

8.1 Credible Interval (Equal-Tailed)

For parameter θ with posterior $p(\theta | D)$, a $(1 - \alpha)$ central credible interval is

$$\text{CrI}_{1-\alpha} = [q_{\alpha/2}, q_{1-\alpha/2}], \quad (94)$$

where q_p is the p -quantile of $p(\theta | D)$.

8.2 Highest Posterior Density (HPD) Region

HPD region H with mass $(1 - \alpha)$ satisfies

$$\int_H p(\theta | D) d\theta = 1 - \alpha, \quad (95)$$

$$p(\theta | D) \geq c, \quad \forall \theta \in H, \quad (96)$$

for some threshold c chosen to achieve mass $1 - \alpha$.

9 Bayesian Filtering (General State-Space Model)

Density form: $quadp(z_k | z_{k-1}), \quad p(y_k | z_k)$.

Prediction (Chapman–Kolmogorov):

$$p(z_k | y_{1:k-1}) = \int p(z_k | z_{k-1}) p(z_{k-1} | y_{1:k-1}) dz_{k-1}. \quad (97)$$

Update (Bayes' rule):

$$p(z_k | y_{1:k}) = \frac{p(y_k | z_k) p(z_k | y_{1:k-1})}{p(y_k | y_{1:k-1})}, \quad (98)$$

$$p(y_k | y_{1:k-1}) = \int p(y_k | z_k) p(z_k | y_{1:k-1}) dz_k. \quad (99)$$

Joint factorization (SSM):

$$p(z_{0:T}, y_{1:T}) = p(z_0) \prod_{k=1}^T p(z_k | z_{k-1}) \prod_{k=1}^T p(y_k | z_k). \quad (100)$$

9.1 Linear–Gaussian State Space Model (SSM)

Dynamics: $z_t = F_t z_{t-1} + B_t u_t + b_t + q_t, \quad q_t \sim \mathcal{N}(0, Q_t)$.

$$p(z_t | z_{t-1}, u_t) = \mathcal{N}(z_t; F_t z_{t-1} + B_t u_t + b_t, Q_t). \quad (101)$$

Measurement: $y_t = H_t z_t + D_t u_t + d_t + r_t, \quad r_t \sim \mathcal{N}(0, R_t)$.

$$p(y_t | z_t, u_t) = \mathcal{N}(y_t; H_t z_t + D_t u_t + d_t, R_t). \quad (102)$$

10 Kalman Filter (Complete Algorithm)

Initialization: $p(z_0) = \mathcal{N}(z_0; \mu_0, \Sigma_0)$. **For** $t=1, \dots, T$:

Prediction:

$$\begin{aligned} \mu_{t|t-1} &= F_t \mu_{t-1|t-1} + B_t u_t + b_t, \\ \Sigma_{t|t-1} &= F_t \Sigma_{t-1|t-1} F_t^\top + Q_t. \end{aligned} \quad (103)$$

Update:

$$\hat{y}_t = H_t \mu_{t|t-1} + D_t u_t + d_t, \quad S_t = H_t \Sigma_{t|t-1} H_t^\top + R_t, \quad (104)$$

$$K_t = \Sigma_{t|t-1} H_t^\top S_t^{-1}, \quad \mu_{t|t} = \mu_{t|t-1} + K_t(y_t - \hat{y}_t), \quad (105)$$

$$\Sigma_{t|t} = (I - K_t H_t) \Sigma_{t|t-1}. \quad (106)$$

Innovation (residual):

$$\tilde{y}_t = y_t - \hat{y}_t, \quad \tilde{y}_t \sim \mathcal{N}(0, S_t) \text{ (consistency check).}$$

10.1 Kalman Filter (1D Scalar Special Case)

Model:

$$z_t = z_{t-1} + q_t, \quad q_t \sim \mathcal{N}(0, q_t), \quad y_t = z_t + r_t, \quad r_t \sim \mathcal{N}(0, r_t). \quad (107)$$

Prediction:

$$\mu_{t|t-1} = \mu_{t-1|t-1}, \quad \sigma_{t|t-1}^2 = \sigma_{t-1|t-1}^2 + q_t. \quad (108)$$

Update:

$$\hat{y}_t = \mu_{t|t-1}, \quad S_t = \sigma_{t|t-1}^2 + r_t, \quad K_t = \frac{\sigma_{t|t-1}^2}{\sigma_{t|t-1}^2 + r_t}, \quad (109)$$

$$\mu_{t|t} = \mu_{t|t-1} + K_t(y_t - \hat{y}_t), \quad \sigma_{t|t}^2 = (1 - K_t)\sigma_{t|t-1}^2. \quad (110)$$

10.2 Monte Carlo and Importance Sampling

Monte Carlo expectation:

$$\mathbb{E}[f(Z)] = \int f(z)p(z) dz \approx \frac{1}{M} \sum_{m=1}^M f(z^{(m)}), \quad (111)$$

$$z^{(m)} \sim p(z). \quad (112)$$

Importance sampling identity:

$$\mathbb{E}[f(Z)] = \int f(z) \frac{p(z)}{q(z)} q(z) dz \approx \sum_{m=1}^M \tilde{w}^{(m)} f(z^{(m)}),$$

$$z^{(m)} \sim q(z), \quad (113)$$

$$w^{(m)} = \frac{p(z^{(m)})}{q(z^{(m)})}, \quad \tilde{w}^{(m)} = \frac{w^{(m)}}{\sum_{j=1}^M w^{(j)}}. \quad (114)$$

10.3 Sequential Importance Sampling / Particle Approximation

Empirical approximation:

$$p(z_k | y_{1:k}) \approx \sum_{m=1}^M w_k^{(m)} \delta(z_k - z_k^{(m)}). \quad (115)$$

General SMC weight update (observation-dependent proposal):

$$w_k^{(m)} \propto w_{k-1}^{(m)} \frac{p(y_k | z_k^{(m)}) p(z_k^{(m)} | z_{k-1}^{(m)})}{q(z_k^{(m)} | z_{k-1}^{(m)}, y_k)}. \quad (116)$$

Bootstrap / prior proposal ($q = p(z_k | z_{k-1})$):

$$w_k^{(m)} \propto w_{k-1}^{(m)} p(y_k | z_k^{(m)}). \quad (117)$$

10.4 Resampling (SIR / PF)

Resampling distribution: sample indices from

$$\sum_{j=1}^M w_k^{(j)} \delta(z_k - z_k^{(j)}), \quad (118)$$

then set new weights to $1/M$:

$$\{\bar{z}_k^{(m)}\}_{m=1}^M \sim \sum_{j=1}^M w_k^{(j)} \delta(z_k - z_k^{(j)}), \quad \bar{w}_k^{(m)} = \frac{1}{M}. \quad (119)$$

State estimate (posterior mean):

$$\hat{z}_k = \sum_{m=1}^M w_k^{(m)} z_k^{(m)}. \quad (120)$$

10.5 Optimal Proposal (Variance-Minimizing)

Optimal proposal:

$$q_k^{\text{opt}}(z_k | z_{k-1}, y_k) = p(z_k | z_{k-1}, y_k). \quad (121)$$

Weight simplification under q^{opt} :

$$w_k^{(m)} \propto w_{k-1}^{(m)} p(y_k | z_{k-1}^{(m)}), \quad (122)$$

$$p(y_k | z_{k-1}) = \int p(y_k | z_k) p(z_k | z_{k-1}) dz_k. \quad (123)$$