Lecture 11: Conjugate Priors and Bayesian Regression

STA702

Merlise Clyde
Duke University

Semi-Conjugate Priors in Linear Regression

• Regression Model (Sampling model)

$$\mathbf{Y} \mid oldsymbol{eta}, \phi \sim \mathsf{N}(\mathbf{X}oldsymbol{eta}, \phi^{-1}\mathbf{I}_n)$$

• Semi-Conjugate Prior Independent Normal Gamma

$$oldsymbol{eta} \sim \mathsf{N}(\mathbf{b}_0, oldsymbol{\Phi}_0^{-1}) \ \phi \sim \mathsf{Gamma}(
u_0/2, \mathsf{SS}_0/2)$$

- lacksquare Conditional Normal for $oldsymbol{eta}\mid\phi,\mathbf{Y}$ and
- Conditional Gamma $\phi \mid \mathbf{Y}, \boldsymbol{\beta}$
- requires Gibbs sampling or other Metropolis-Hastings algorithms

Conjugate Priors in Linear Regression

• Regression Model (Sampling model)

$$\mathbf{Y} \mid oldsymbol{eta}, \phi \sim \mathsf{N}(\mathbf{X}oldsymbol{eta}, \phi^{-1}\mathbf{I}_n)$$

ullet Conjugate Normal-Gamma Model: factor joint prior $p(oldsymbol{eta},\phi)=p(oldsymbol{eta}\mid\phi)p(\phi)$

$$eta \mid \phi \sim \mathsf{N}(\mathbf{b}_0,\phi^{-1}\mathbf{\Phi}_0^{-1}) \qquad p(oldsymbol{eta}\mid \phi) = rac{|\phi\mathbf{\Phi}_0|^{1/2}}{(2\pi)^{p/2}}e^{\left\{-rac{\phi}{2}(oldsymbol{eta}-\mathbf{b}_0)^T\mathbf{\Phi}_0(oldsymbol{eta}-\mathbf{b}_0)
ight\}} \ \phi \sim \mathsf{Gamma}(v_0/2,\mathsf{SS}_0/2) \qquad p(\phi) = rac{1}{\Gamma(
u_0/2)} igg(rac{\mathsf{SS}_0}{2}igg)^{
u_0/2} \phi^{
u_0/2-1}e^{-1} \ \Rightarrow (oldsymbol{eta},\phi) \sim \mathsf{NG}(\mathbf{b}_0,\mathbf{\Phi}_0,
u_0,\mathsf{SS}_0)$$

- Normal-Gamma distribution indexed by 4 hyperparameters
- Note Prior Covariance for $oldsymbol{eta}$ is scaled by $\sigma^2=1/\phi$

Finding the Posterior Distribution

ullet Likelihood: $\mathcal{L}(eta,\phi) \propto \phi^{n/2} e^{-rac{\phi}{2}(\mathbf{Y}-\mathbf{X}oldsymbol{eta})^T(\mathbf{Y}-\mathbf{X}oldsymbol{eta})}$

$$egin{aligned} p(oldsymbol{eta},\phi\mid\mathbf{Y})&\propto\phi^{rac{n}{2}}e^{-rac{\phi}{2}(\mathbf{Y}-\mathbf{X}oldsymbol{eta})^T(\mathbf{Y}-\mathbf{X}oldsymbol{eta})} imes\ \phi^{rac{
u_0}{2}-1}e^{-\phirac{\mathsf{SS}_0}{2}} imes\phi^{rac{p}{2}}e^{-rac{\phi}{2}(oldsymbol{eta}-\mathbf{b}_0)^Toldsymbol{\Phi}_0(oldsymbol{eta}-\mathbf{b}_0)} \end{aligned}$$

Quadratic in Exponential

$$\exp\left\{-rac{\phi}{2}(oldsymbol{eta}-\mathbf{b})^Toldsymbol{\Phi}(oldsymbol{eta}-\mathbf{b})
ight\}=\exp\left\{-rac{\phi}{2}(oldsymbol{eta}^Toldsymbol{\Phi}oldsymbol{eta}-2oldsymbol{eta}^Toldsymbol{\Phi}\mathbf{b}+\mathbf{b}^Toldsymbol{\Phi}\mathbf{b})
ight\}$$

- Expand quadratics and regroup terms
- Read off posterior precision from Quadratic in β
- Read off posterior mean from Linear term in β
- will need to complete the quadratic in the posterior mean due to ϕ

http://localhost:6565/resources/slides/11-bayes-regression.html?print-pdf=#/formal-posterior-distribution

Expand and Regroup

$$p(\boldsymbol{\beta}, \phi \mid \mathbf{Y}) \propto \phi^{\frac{n}{2}} e^{-\frac{\phi}{2} (\mathbf{Y}^T \mathbf{Y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta})} \times \\ \phi^{\frac{\nu_0}{2} - 1} e^{-\phi \frac{\mathsf{SS}_0}{2}} \times \phi^{\frac{p}{2}} e^{-\frac{\phi}{2} (\boldsymbol{\beta} \boldsymbol{\Phi}_0 \boldsymbol{\beta} - 2\boldsymbol{\beta}^T \boldsymbol{\Phi}_0 \mathbf{b} + \mathbf{b}_0^T \boldsymbol{\Phi}_0 \mathbf{b}_0)} \times \\ p(\boldsymbol{\beta}, \phi \mid \mathbf{Y}) \propto \phi^{\frac{n+p+\nu_0}{2} - 1} e^{-\frac{\phi}{2} (\mathsf{SS}_0 + \mathbf{Y}^T \mathbf{Y} + \mathbf{b}_0^T \boldsymbol{\Phi}_0 \mathbf{b}_0)} \times \\ e^{-\frac{\phi}{2} (\boldsymbol{\beta}^T (\mathbf{X}^T \mathbf{X}) \boldsymbol{\beta} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta} \boldsymbol{\Phi}_0 \boldsymbol{\beta} - 2\boldsymbol{\beta}^T \boldsymbol{\Phi}_0 \mathbf{b})} \times \\ = \phi^{\frac{n+p+\nu_0}{2} - 1} e^{-\frac{\phi}{2} (\mathsf{SS}_0 + \mathbf{Y}^T \mathbf{Y} + \mathbf{b}_0^T \boldsymbol{\Phi}_0 \mathbf{b}_0)} \times \\ e^{-\frac{\phi}{2} (\boldsymbol{\beta}^T (\mathbf{X}^T \mathbf{X} + \boldsymbol{\Phi}_0) \boldsymbol{\beta})} \times \\ e^{-\frac{\phi}{2} (-2\boldsymbol{\beta}^T (\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \boldsymbol{\Phi}_0 \mathbf{b}_0)})$$

Complete the Quadratic

$$\begin{split} p(\boldsymbol{\beta}, \boldsymbol{\phi} \mid \mathbf{Y}) &\propto \boldsymbol{\phi}^{\frac{n+p+\nu_0}{2}-1} e^{-\frac{\boldsymbol{\phi}}{2}(\mathsf{SS}_0 + \mathbf{Y}^T \mathbf{Y} + \mathbf{b}_0^T \mathbf{\Phi}_0 \mathbf{b}_0)} \times \\ &e^{-\frac{\boldsymbol{\phi}}{2} \left(\boldsymbol{\beta}^T (\mathbf{X}^T \mathbf{X} + \mathbf{\Phi}_0) \boldsymbol{\beta} \right)} \times \qquad \qquad \mathbf{\Phi}_n \equiv \mathbf{X}^T \mathbf{X} + \mathbf{\Phi}_0 \\ &e^{-\frac{\boldsymbol{\phi}}{2} \left(-2\boldsymbol{\beta}^T \mathbf{\Phi}_n \mathbf{\Phi}_n^{-1} (\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{\Phi}_0 \mathbf{b}_0) \right)} \times \qquad \qquad \mathbf{b}_n \equiv \mathbf{\Phi}_n^{-1} (\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{\Phi}_0 \mathbf{b}_0) \\ &e^{-\frac{\boldsymbol{\phi}}{2} (\mathbf{b}_n^T \mathbf{\Phi}_n \mathbf{b}_n - \mathbf{b}_n^T \mathbf{\Phi}_n \mathbf{b}_n)} \times \\ &e^{-\frac{\boldsymbol{\phi}}{2} (\mathbf{b}_n^T \mathbf{\Phi}_n \mathbf{b}_n - \mathbf{b}_n^T \mathbf{\Phi}_n \mathbf{b}_n)} \times \\ &e^{\frac{\boldsymbol{\phi}}{2} - 1} e^{-\frac{\boldsymbol{\phi}}{2} ((\mathbf{S}\mathbf{S}_0 + \mathbf{Y}^T \mathbf{Y} + \mathbf{b}_0^T \mathbf{\Phi}_0 \mathbf{b}_0 - \mathbf{b}_n^T \mathbf{\Phi}_n \mathbf{b}_n)} \times \\ &e^{\frac{\boldsymbol{\phi}}{2} e^{-\frac{\boldsymbol{\phi}}{2} ((\boldsymbol{\beta}^T - \mathbf{b}_n)^T \mathbf{\Phi}_n (\boldsymbol{\beta} - \mathbf{b}_n))} \times \\ &\propto \boldsymbol{\phi}^{\frac{n+\nu_0}{2} - 1} e^{-\frac{\boldsymbol{\phi}}{2} ((\mathbf{S}\mathbf{S}_0 + \mathbf{Y}^T \mathbf{Y} + \mathbf{b}_0^T \mathbf{\Phi}_0 \mathbf{b}_0 - \mathbf{b}_n^T \mathbf{\Phi}_n \mathbf{b}_n)} \times \\ &|\boldsymbol{\phi} \mathbf{\Phi}_n|^{\frac{1}{2}} e^{-\frac{\boldsymbol{\phi}}{2} ((\boldsymbol{\beta}^T - \mathbf{b}_n)^T \mathbf{\Phi}_n (\boldsymbol{\beta} - \mathbf{b}_n))} \end{split}$$

Posterior Distributions

Posterior density (up to normalizing contants) $p(m{eta}, \phi \mid \mathbf{Y}) = p(\phi \mid \mathbf{Y}) p(m{eta} \mid \phi \mathbf{Y})$

$$p(\phi\mid \mathbf{Y})p(oldsymbol{eta}\mid \phi\mathbf{Y})\propto \phi^{rac{n+
u_0}{2}-1}e^{-rac{\phi}{2}(\mathsf{SS}_0+\mathbf{Y}^T\mathbf{Y}+\mathbf{b}_0^Toldsymbol{\Phi}_0\mathbf{b}_0-\mathbf{b}_n^Toldsymbol{\Phi}_n\mathbf{b}_n)} imes \ (2\pi)^{-rac{p}{2}}|\phi\Phi_n|^{rac{1}{2}}e^{-rac{\phi}{2}(oldsymbol{eta}-\mathbf{b}_n)^Toldsymbol{\Phi}_n(oldsymbol{eta}-\mathbf{b}_n)}$$

Marginal

$$egin{aligned} p(\phi\mid\mathbf{Y})&\propto\phi^{rac{n+
u_0}{2}-1}e^{-rac{\phi}{2}(\mathsf{SS}_0+\mathbf{Y}^T\mathbf{Y}+\mathbf{b}_0^T\mathbf{\Phi}_0\mathbf{b}_0-\mathbf{b}_n^T\mathbf{\Phi}_n\mathbf{b}_n)} imes \ &\int_{\mathbb{R}^p}(2\pi)^{-rac{p}{2}}|\phi\mathbf{\Phi}_n|^{rac{1}{2}}e^{-rac{\phi}{2}(oldsymbol{eta}-\mathbf{b}_n)^T\mathbf{\Phi}_n(oldsymbol{eta}-\mathbf{b}_n)\,doldsymbol{eta} \ &=\phi^{rac{n+
u_0}{2}-1}e^{-rac{\phi}{2}(\mathsf{SS}_0+\mathbf{Y}^T\mathbf{Y}+\mathbf{b}_0^T\mathbf{\Phi}_0\mathbf{b}_0-\mathbf{b}_n^T\mathbf{\Phi}_n\mathbf{b}_n) \end{aligned}$$

- Conditional Normal for $oldsymbol{eta} \mid \phi, \mathbf{Y}$ and marginal Gamma for $\phi \mid \mathbf{Y}$
- No need for Gibbs sampling!

NG Posterior Distribution

$$m{eta} \mid \phi, \mathbf{Y} \sim \mathsf{N}(\mathbf{b}_n, (\phi \mathbf{\Phi}_n)^{-1}) \ \phi \mid \mathbf{Y} \sim \mathsf{Gamma}(rac{
u_n}{2}, rac{\mathsf{SS}_n}{2}) \ (m{eta}, \phi) \mid \mathbf{Y} \sim \mathsf{NG}(\mathbf{b}_n, \mathbf{\Phi}_n,
u_n, \mathsf{SS}_n)$$

Hyperparameters:

$$\begin{aligned} \mathbf{\Phi}_n &= \mathbf{X}^T \mathbf{X} + \mathbf{\Phi}_0 & \mathbf{b}_n &= \mathbf{\Phi}_n^{-1} (\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{\Phi}_0 \mathbf{b}_0) \\ \nu_n &= n + \nu_0 & \mathsf{SS}_n &= \mathsf{SS}_0 + \mathbf{Y}^T \mathbf{Y} + \mathbf{b}_0^T \mathbf{\Phi}_0 \mathbf{b}_0 - \mathbf{b}_n^T \mathbf{\Phi}_n \mathbf{b}_n \\ \mathsf{SS}_n &= \mathsf{SS}_0 + \|\mathbf{Y} - \mathbf{X} \mathbf{b}_n\|^2 + (\mathbf{b}_0 - \mathbf{b}_n)^T \mathbf{\Phi}_0 (\mathbf{b}_0 - \mathbf{b}_n) \\ &= \mathsf{SS}_0 + \|\mathbf{Y} - \mathbf{X} \mathbf{b}_n\|^2 + \|\mathbf{b}_0 - \mathbf{b}_n\|_{\mathbf{\Phi}_0}^2 \end{aligned}$$

- ullet Inner product induced by prior precision $\langle u,v
 angle_A\equiv u^TAv$
- ullet $\|\mathbf{b}_0 \mathbf{b}_n\|_{oldsymbol{\Phi}_0}^2$ mismatch of prior and posterior mean under prior

Marginal Distribution

▼ Theorem: Student-t

Let $m{ heta} \mid \phi \sim \mathsf{N}(m, rac{1}{\phi} m{\Sigma})$ and $\phi \sim \mathsf{Gamma}(
u/2,
u \hat{\sigma}^2/2)$.

Then $oldsymbol{ heta}\left(p imes1
ight)$ has a p dimensional multivariate t distribution

$$oldsymbol{ heta} \sim t_
u(m, \hat{\sigma}^2 oldsymbol{\Sigma})$$

with location m, scale matrix $\hat{\sigma}^2 \Sigma$ and density

$$p(oldsymbol{ heta}) \propto \left[1 + rac{1}{
u} rac{(oldsymbol{ heta} - m)^T oldsymbol{\Sigma}^{-1} (oldsymbol{ heta} - m)}{\hat{\sigma}^2}
ight]^{-rac{p+
u}{2}}$$

Note - true for prior or posterior given ${f Y}$

Derivation

Marginal density $p(oldsymbol{ heta}) = \int_0^\infty p(oldsymbol{ heta} \mid \phi) p(\phi) \, d\phi$

$$egin{aligned} p(oldsymbol{ heta}) &\propto \int |oldsymbol{\Sigma}/\phi|^{-1/2} e^{-rac{\phi}{2}(oldsymbol{ heta}-m)^Toldsymbol{\Sigma}^{-1}(oldsymbol{ heta}-m)} \phi^{
u/2-1} e^{-\phirac{
u\phi^2}{2}} \, d\phi \ &\propto \int \phi^{p/2} \phi^{
u/2-1} e^{-\phirac{(oldsymbol{ heta}-m)^Toldsymbol{\Sigma}^{-1}(oldsymbol{ heta}-m)+
u\hat{\sigma}^2}{2}} \, d\phi \ &\propto \int \phi^{rac{p+
u}{2}-1} e^{-\phirac{(oldsymbol{ heta}-m)^Toldsymbol{\Sigma}^{-1}(oldsymbol{ heta}-m)+
u\hat{\sigma}^2}{2}} \, d\phi \ &= \Gamma((p+
u)/2) igg(rac{(oldsymbol{ heta}-m)^Toldsymbol{\Sigma}^{-1}(oldsymbol{ heta}-m)+
u\hat{\sigma}^2}{2} igg)^{-rac{p+
u}{2}} \ &\propto ig((oldsymbol{ heta}-m)^Toldsymbol{\Sigma}^{-1}(oldsymbol{ heta}-m)+
u\hat{\sigma}^2igg)^{-rac{p+
u}{2}} \ &\propto igg(1+rac{1}{
u}rac{(oldsymbol{ heta}-m)^Toldsymbol{\Sigma}^{-1}(oldsymbol{ heta}-m)}{\hat{\sigma}^2} igg)^{-rac{p+
u}{2}} \end{aligned}$$

Marginal Posterior Distribution of $oldsymbol{eta}$

$$m{eta} \mid \phi, \mathbf{Y} \sim \mathsf{N}(\mathbf{b}_n, \phi^{-1}m{\Phi}_n^{-1}) \ \phi \mid \mathbf{Y} \sim \mathsf{Gamma}\left(rac{
u_n}{2}, rac{\mathsf{SS}_n}{2}
ight)$$

- Let $\hat{\sigma}^2 = \mathsf{SS}_n/\nu_n$ (Bayesian MSE)
- The marginal posterior distribution of $oldsymbol{eta}$ is multivariate Student-t

$$oldsymbol{eta} \mid \mathbf{Y} \sim t_{
u_n}(\mathbf{b}_n, \hat{\sigma}^2 \mathbf{\Phi}_n^{-1})$$

• Any linear combination $\lambda^T \beta$ has a univariate t distribution with \mathbf{v}_n degrees of freedom

$$\lambda^T oldsymbol{eta} \mid \mathbf{Y} \sim t_{
u_n}(\lambda^T \mathbf{b}_n, \hat{\sigma}^2 \lambda^T \Phi_n^{-1} \lambda)$$

• use for individual $m{eta}_j$, the mean of $Y, \mathbf{x}^T m{eta}$, at \mathbf{x} , or predictions $Y^* = \mathbf{x}^{*T} m{eta} + \epsilon_i^*$

Predictive Distributions

Suppose ${\bf Y}^* \mid {m \beta}, \phi \sim {\sf N}_s({\bf X}^*{m \beta}, {\bf I}_s/\phi)$ and is conditionally independent of ${\bf Y}$ given ${m \beta}$ and ϕ

- What is the predictive distribution of $\mathbf{Y}^* \mid \mathbf{Y}$?
- Use the representation that $\mathbf{Y}^* \stackrel{\mathrm{D}}{=} \mathbf{X}^* m{eta} + m{\epsilon}^*$ and $m{\epsilon}^*$ is independent of \mathbf{Y} given ϕ

$$\mathbf{X}^*oldsymbol{eta} + oldsymbol{\epsilon}^* \mid \phi, \mathbf{Y} \sim \mathsf{N}(\mathbf{X}^*\mathbf{b}_n, (\mathbf{X}^*oldsymbol{\Phi}_n^{-1}\mathbf{X}^{*T} + \mathbf{I}_s)/\phi) \ \mathbf{Y}^* \mid \phi, \mathbf{Y} \sim \mathsf{N}(\mathbf{X}^*\mathbf{b}_n, (\mathbf{X}^*oldsymbol{\Phi}_n^{-1}\mathbf{X}^{*T} + \mathbf{I}_s)/\phi) \ \phi \mid \mathbf{Y} \sim \mathsf{Gamma}\left(rac{
u_n}{2}, rac{\hat{\sigma}^2
u_n}{2}
ight)$$

• Use the Theorem to conclude that

$$\mathbf{Y}^* \mid \mathbf{Y} \sim t_{
u_n}(\mathbf{X}^*\mathbf{b}_n, \hat{\sigma}^2(\mathbf{I} + \mathbf{X}^*\mathbf{\Phi}_n^{-1}\mathbf{X}^T))$$

Choice of Conjugate (or Semi-Conjugate) Prior

- need to specify Normal prior mean ${f b}_0$ and precision ${f \Phi}_0$
- need to specify Gamma shape (u_o prior df) and rate (estimate of σ^2)
- hard in higher dimensions!
- default choices?
 - Jeffreys' prior
 - unit-information prior
 - Zellner's g-prior
 - ridge priors
 - mixtures of conjugate priors

Jeffreys' Prior

• Jeffreys prior is invariant to model parameterization of $m{ heta}=(m{eta},\phi)$

$$p(oldsymbol{ heta}) \propto |\mathcal{I}(oldsymbol{ heta})|^{1/2}$$

• $\mathcal{I}(\boldsymbol{\theta})$ is the Expected Fisher Information matrix

$$\mathcal{I}(heta) = -\mathsf{E}[\left[rac{\partial^2 \log(\mathcal{L}(oldsymbol{ heta}))}{\partial heta_i \partial heta_j}
ight]]$$

· log likelihood expressed as function of sufficient statistics

$$\log(\mathcal{L}(oldsymbol{eta},\phi)) = rac{n}{2} \log(\phi) - rac{\phi}{2} \|(\mathbf{I}_n - \mathbf{P_x})\mathbf{Y}\|^2 - rac{\phi}{2} (oldsymbol{eta} - \hat{oldsymbol{eta}})^T (\mathbf{X}^T \mathbf{X}) (oldsymbol{eta} - \hat{oldsymbol{eta}})$$

• projection matrix $\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$

Information matrix

$$egin{aligned} rac{\partial^2 \log \mathcal{L}}{\partial oldsymbol{ heta} \partial oldsymbol{ heta} \partial oldsymbol{ heta} \partial oldsymbol{ heta} & -(\mathbf{X}^T \mathbf{X}) (oldsymbol{eta} - \hat{oldsymbol{eta}}) \ -(oldsymbol{eta} - \hat{oldsymbol{eta}})^T (\mathbf{X}^T \mathbf{X}) & -rac{n}{2} rac{1}{\phi^2} \end{aligned} \end{bmatrix} \ \mathsf{E}[rac{\partial^2 \log \mathcal{L}}{\partial oldsymbol{ heta} \partial oldsymbol{ heta} \partial oldsymbol{ heta}^T}] = egin{bmatrix} -\phi(\mathbf{X}^T \mathbf{X}) & \mathbf{0}_p \ \mathbf{0}_p^T & -rac{n}{2} rac{1}{\phi^2} \end{bmatrix}$$

$$\mathcal{I}((oldsymbol{eta},\phi)^T) = egin{bmatrix} \phi(\mathbf{X}^T\mathbf{X}) & \mathbf{0}_p \ \mathbf{0}_p^T & rac{n}{2}rac{1}{\phi^2} \end{bmatrix}$$

Jeffreys' Prior (don't use!)

$$p_J(oldsymbol{eta},\phi) \propto |\mathcal{I}((oldsymbol{eta},\phi)^T)|^{1/2} = |\phi \mathbf{X}^T \mathbf{X}|^{1/2} igg(rac{n}{2} rac{1}{\phi^2}igg)^{1/2} \propto \phi^{p/2-1} |\mathbf{X}^T \mathbf{X}|^{1/2} \ \propto \phi^{p/2-1}$$

Recommended Independent Jeffreys Prior

- Treat $\boldsymbol{\beta}$ and ϕ separately (orthogonal parameterization)
- $ullet \ p_{IJ}(oldsymbol{eta}) \propto |\mathcal{I}(oldsymbol{eta})|^{1/2}$ and $p_{IJ}(\phi) \propto |\mathcal{I}(\phi)|^{1/2}$

$$\mathcal{I}((oldsymbol{eta},\phi)^T) = egin{bmatrix} \phi(\mathbf{X}^T\mathbf{X}) & \mathbf{0}_p \ \mathbf{0}_p^T & rac{n}{2}rac{1}{\phi^2} \end{bmatrix}$$

$$egin{align} p_{IJ}(oldsymbol{eta}) &\propto |\phi\mathbf{X}^T\mathbf{X}|^{1/2} \propto 1 \ p_{IJ}(\phi) &\propto \phi^{-1} \ p_{IJ}(eta,\phi) &\propto p_{IJ}(oldsymbol{eta}) p_{IJ}(\phi) = \phi^{-1} \ \end{aligned}$$