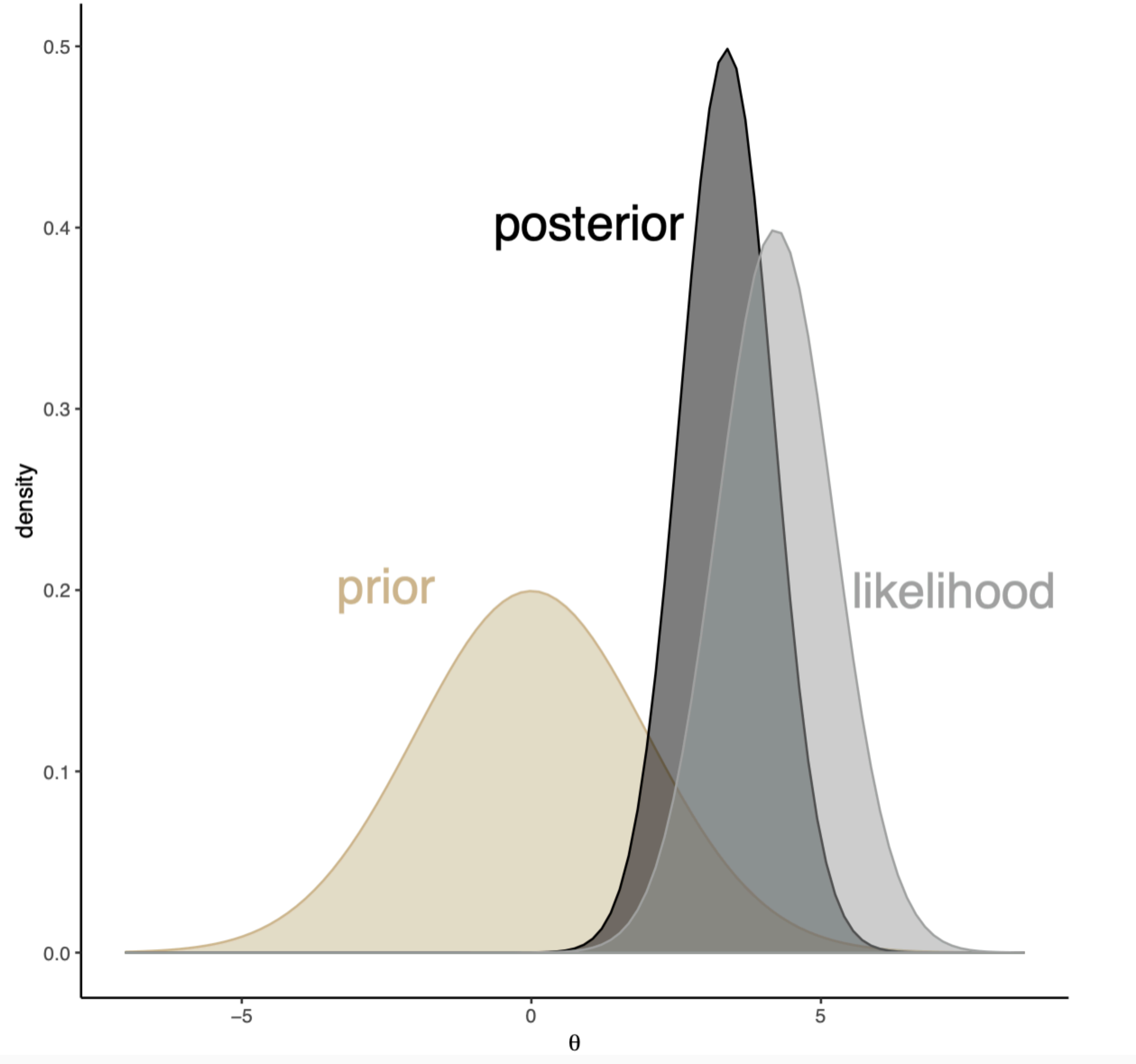
Unit #3: Bayesian inference on multiple parameters

Bayesian Statistical Methods, 2.1.6, 2.1.7, 4.1.2, 4.2.1, 4.2.2

Secondary reference: Bayesian Data Analysis, 3.1 - 3.3

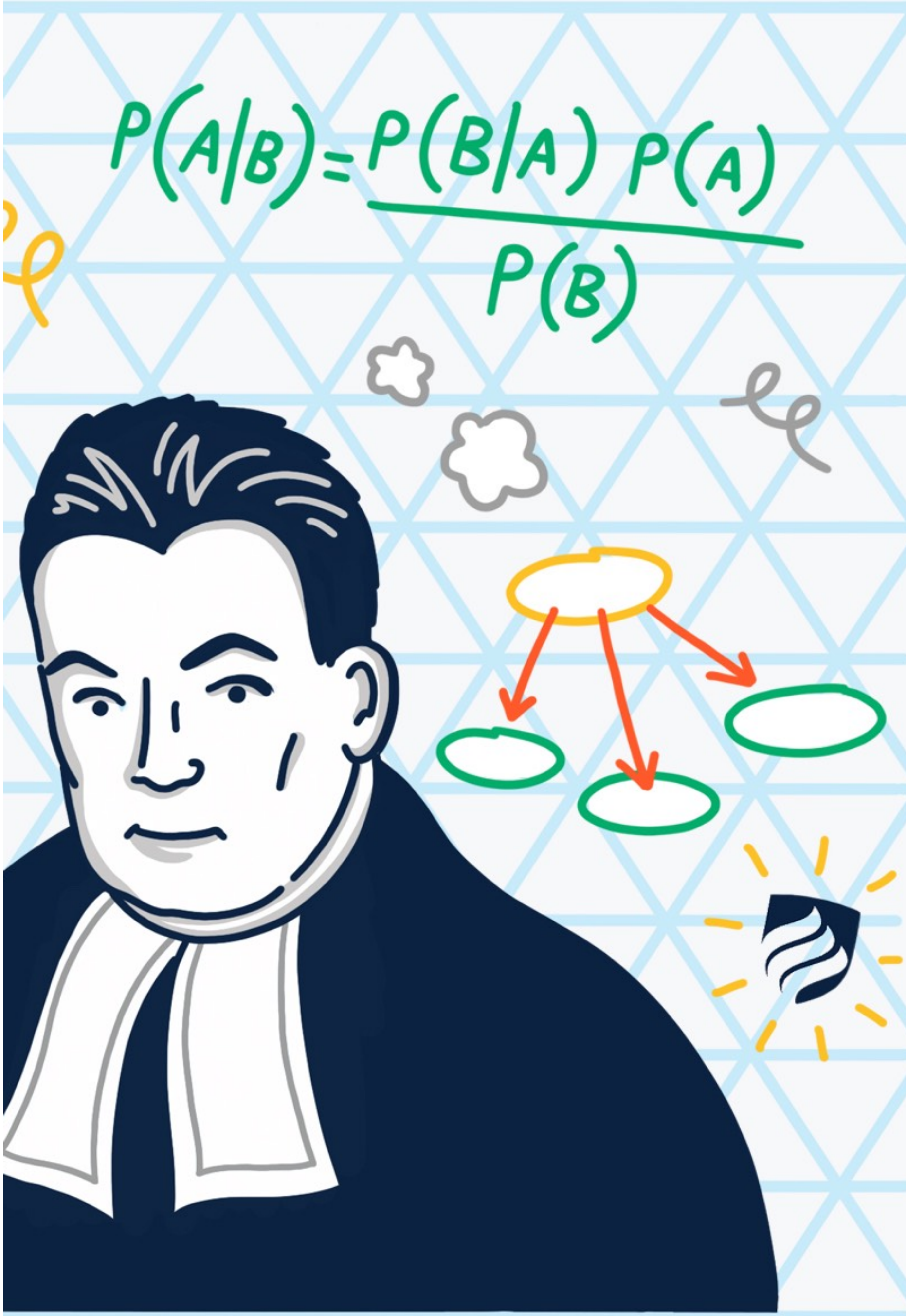


So far, we've mainly looked at single-parameter Bayesian inference.

But often, we will be interested in problems where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$.

In many problems inference is only required on a small number of the unknown parameters, even though other parameters are present in the model. Parameters of this kind are often called **nuisance parameters**.

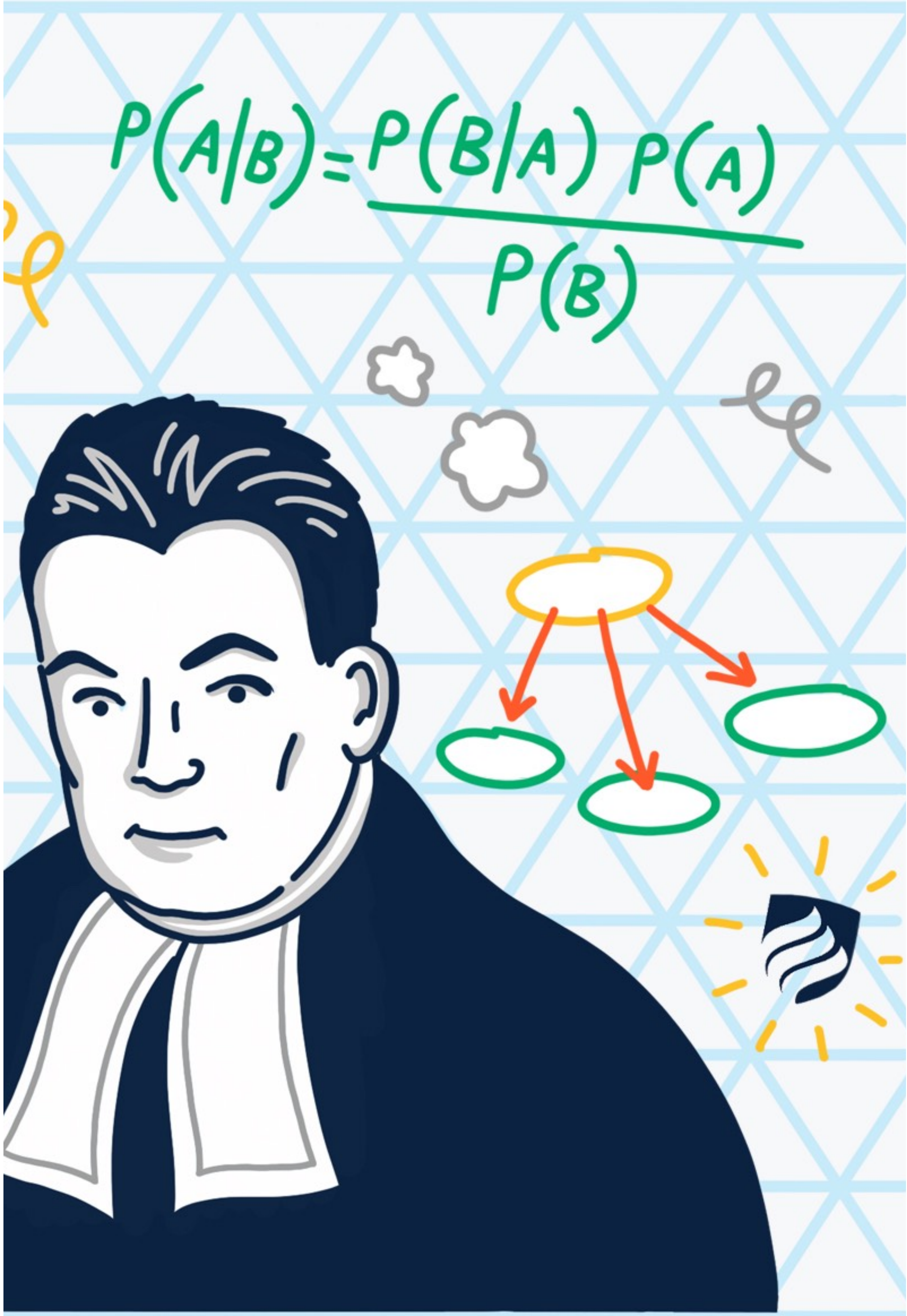
Example: Often, we have the model $X_1, \ldots, X_n \mid \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$, where both μ, σ^2 are unknown, but we are only interested in inferences about μ.



From: https://bit.ly/3gTjlBg

Let $X_1, \ldots, X_n \overset{iid}{\sim} f(\mathbf{x} \mid \boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$. Let $\pi(\boldsymbol{\theta})$ be the joint prior on the parameter vector $\boldsymbol{\theta}$. Then:

If we are mainly concerned with inferences on θ_1, and θ_2 is a nuisance parameter, then we can find the marginal posterior for θ_1 given the data:



From: https://bit.ly/3gTjlBg

Posterior distributions can be computed by marginal and conditional simulation, first drawing θ_2 from its marginal posterior distribution and then θ_1 from its conditional posterior distribution, given the drawn value of θ_2.


Photo by Toby Elliott on Unsplash

Example: Let $X_1, \ldots, X_n \,\Big|\, \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$. Let $\pi_\mu(\mu) = 1$ and $\pi_{\sigma^2}(\sigma^2) = 1/\sigma^2$.

Find the posterior distribution of $\mu, \sigma^2 \mid \mathbf{x}$.

- Prior:

- Likelihood:

- Posterior:

The conditional posterior distribution, $\pi(\mu \mid \sigma^2, \mathbf{x})$:

The marginal posterior distribution, $\pi(\sigma^2 \mid \mathbf{x})$:

Photo by Universal Eye on Unsplash

Modeling the variance:

The **inverse-χ^2 distribution** with $v_0 > 0$ degrees of freedom is equivalent to the inverse-Γ distribution with $\alpha = \dfrac{v_0}{2}$ and $\beta = \dfrac{1}{2}$.

The **scaled-inverse-χ^2 distribution** is useful for σ^2 in a normal model. The scaled-inverse-χ^2 distribution is equivalent to the inverse-Γ distribution with $\alpha = \dfrac{v_0}{2}$ and $\beta = \dfrac{v_0\sigma_0^2}{2}$.

For simulation purposes: to obtain a random draw from the scaled -inverse-χ^2 distribution, i.e., $\theta \sim$ scaled-inverse-$\chi^2(v_0, \sigma_0^2)$:

- Draw $\gamma \sim \chi^2(v_0) \equiv \Gamma\left(\dfrac{v_0}{2}, \dfrac{1}{2}\right)$
- Let $\theta = v_0\sigma_0^2 / \gamma$.

Normal data with a conjugate prior distribution

Likelihood/data: $X_1, \ldots, X_n \,\Big|\, \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$

Priors:

1. $\mu \,|\, \sigma^2 \sim N(\mu_0, \sigma^2 / k_0)$

2. $\sigma^2 \sim \text{inv-}\Gamma\left(\dfrac{v_0}{2}, \dfrac{v_0 \sigma_0^2}{2}\right) \equiv \text{Scale-Inv-}\chi^2\left(v_0, \sigma_0^2\right)$

3. $(\mu, \sigma^2) \sim \text{normal-inverse-}\Gamma\left(\mu, \sigma^2 \,|\, \mu_0,\ \sigma_0^2 / k_0; v_0, \sigma_0^2\right)$

$\pi(\mu, \sigma^2) =$

Normal data with a conjugate prior distribution

Likelihood/data: $X_1, \ldots, X_n \,\Big|\, \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$

Posterior:

$$\mu, \sigma^2 \mid \mathbf{x} \sim \text{normal-inverse-}\Gamma\left(\mu, \sigma^2 \mid \mu_n, \sigma_n^2/k_n; v_n, \sigma_n^2\right),$$

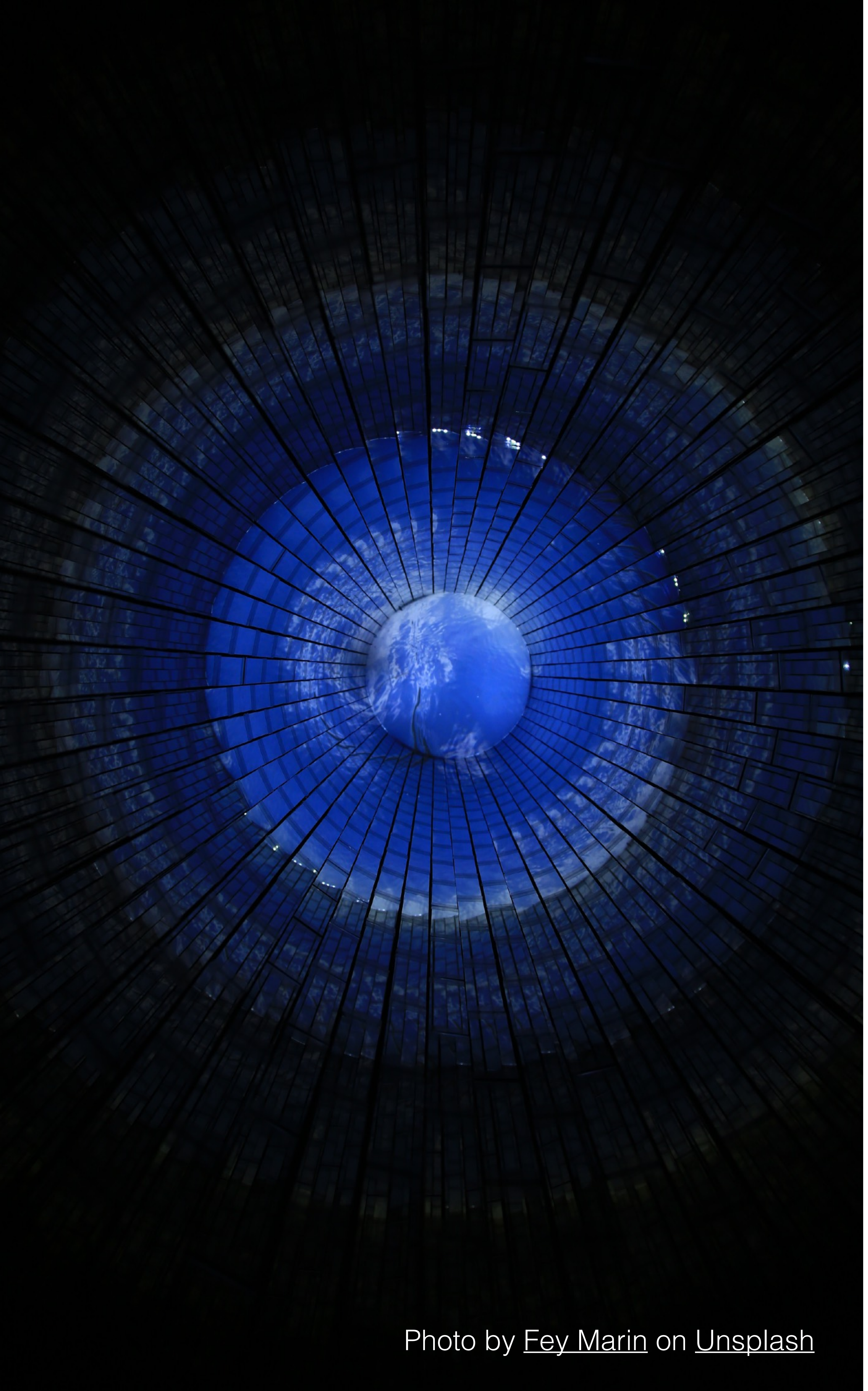
$$k_n = k_0 + n$$

$$\mu_n = \frac{k_0}{k_n}\mu_0 + \frac{n}{k_n}\bar{x}$$

$$v_n = v_0 + n$$

$$\sigma_n^2 = \frac{v_0\sigma_0^2 + (n-1)S^2 + \frac{k_0 n}{k_n}(\bar{x} - \mu_0)^2}{v_n}$$

Proof for HW!



Suppose that $\mathbf{Y} \mid \boldsymbol{\beta} \sim N\left(X\boldsymbol{\beta}, \Sigma_n\right)$, where X is an $n \times p$ design matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters, and Σ_n is a variance-covariance matrix. Let $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}, \Sigma_p)$. Find $\pi(\boldsymbol{\beta} \mid \mathbf{Y})$.

- Prior:

- Likelihood:

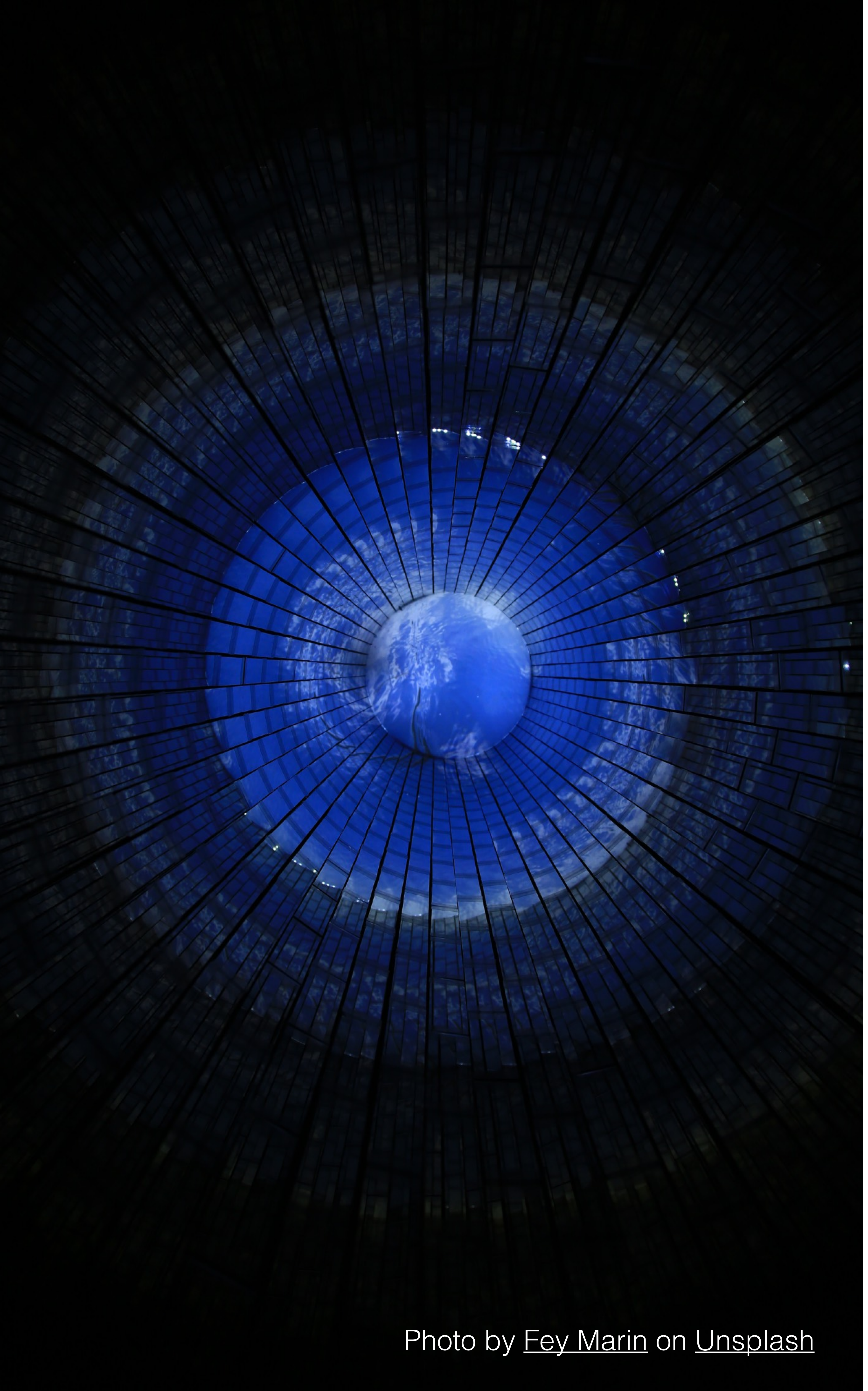


Photo by Fey Marin on Unsplash

Suppose that $\mathbf{Y} \mid \boldsymbol{\beta} \sim N\left(X\boldsymbol{\beta}, \Sigma_n\right)$, where X is an $n \times p$ design matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters, and Σ_n is a variance-covariance matrix. Let $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}, \Sigma_p)$. Find $\pi(\boldsymbol{\beta} \mid \mathbf{Y})$.

- Posterior:

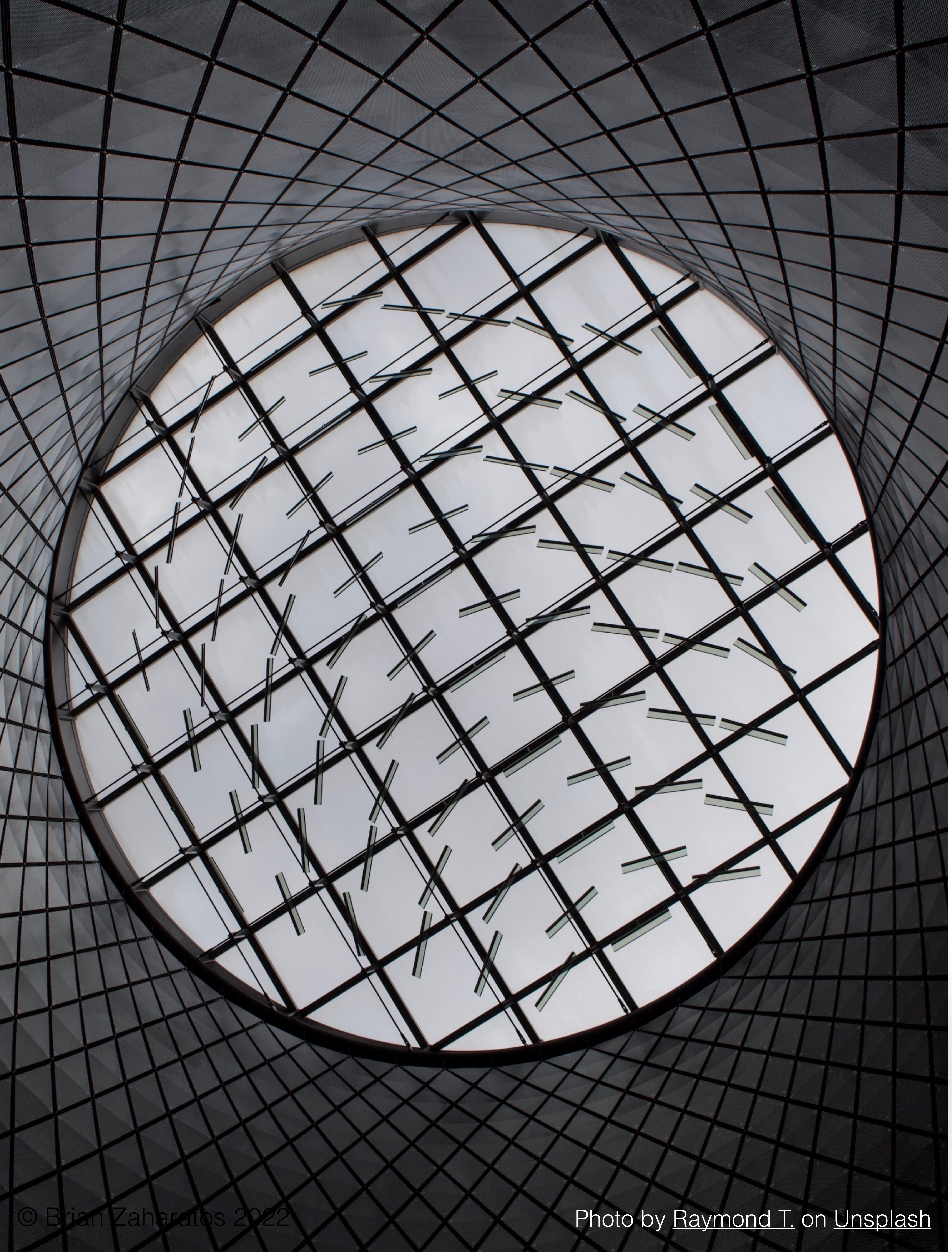


Photo by Raymond T. on Unsplash

A review of Frequentist regression assumptions:

1. $E\left(\varepsilon_i\right) = 0$ for all $i = 1,\ldots,n$.

2. $E\left(Y_i\right) = \mathbf{x_i}^T\boldsymbol{\beta}$ for all $i = 1,\ldots,n$.

3. $Cov\left(\varepsilon_i, \varepsilon_j\right) = \begin{cases} 0 & i \neq j \\ \sigma^2 & i = j \end{cases}$

4. $\left(X^TX\right)^{-1}$ exists.

5. $Y_i \overset{iid}{\sim} N\left(\mathbf{x_i}^T\boldsymbol{\beta}, \sigma^2\right)$

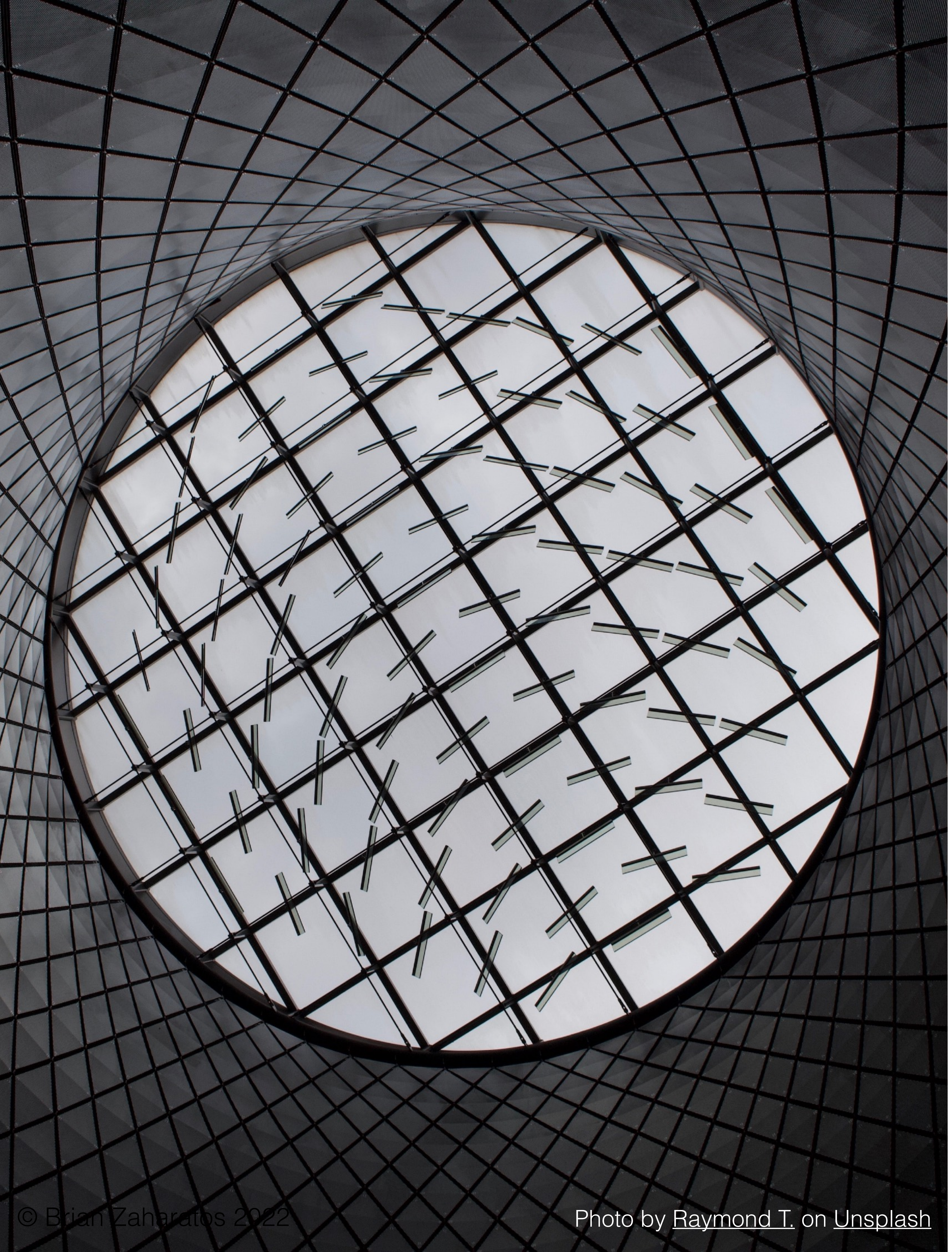


Photo by Raymond T. on Unsplash

Recall that the least squares (frequentist) estimator in regression is $\widehat{\boldsymbol{\beta}} = \underbrace{\left(X^T X\right)^{-1} X^T}_{\text{constant}} \mathbf{Y}$.

This estimator is a linear combination of normal random variables. Thus, it's normal with parameters:

What about the Bayes case?

Let $\mathbf{X}_i = (X_1, \ldots, X_p)^T \sim N_p(\mathbf{0}, \Sigma_p)$, for $i = 1,...,n$. Let's consider the multivariate problem of estimating Σ_p, the variance-covariance matrix of the data.

Note that Σ_p is the same across each $\mathbf{X}_i$. Let $D = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ be a random sample from $N_p(\mathbf{0}, \Sigma_p)$.

Photo by Universal Eye on Unsplash

Modeling the variance-covariance matrix:

The **inverse Wishart distribution** with v degrees of freedom and symmetric positive definite $p \times p$ scale matrix R is a generalization of the scaled-inverse-χ^2 distribution. It is useful for modeling Σ in a multivariate normal model.

Let $\mathbf{X}_i = (X_1, \ldots, X_p)^T \sim N_p(\mathbf{0}, \Sigma_p)$, for $\boldsymbol{i} = \mathbf{1}, \ldots, \boldsymbol{n}$. We'd like to estimate (i.e., find a posterior distribution for) Σ_p. Note that Σ_p is the same across each $\mathbf{X}_i$. Let $D = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ be a random sample from $N_p(\mathbf{0}, \Sigma_p)$.

$M_{p \times p}$ has a **Wishart distribution** $M \sim W_p(v, R)$ where:

1. $v > p - 1$ is the degrees of freedom.

2. R is a symmetric positive definite scale matrix ($p \times p$).

The pdf of M is given by:

$$f_M(\mathbf{m}) = \left(\frac{1}{2^{vp/2} \left| R \right|^{v/2} \Gamma_p(v/2)} \right) |\, \mathbf{m} \,|^{(v-p-1)/2} \exp \left\{ -\frac{1}{2} tr \left(R^{-1} \mathbf{m} \right) \right\}$$

A prior on Σ_p

$Q = M_{p\times p}^{-1}$ has an **Inverse Wishart distribution** $Q \sim InvW_p(v, R)$ with a pdf of Q is given by:

$$f_M(\mathbf{q}) = \left(\frac{1}{2^{vp/2} \left| R \right|^{-v/2} \Gamma_p(v/2)} \right) |\mathbf{q}|^{-(v+p+1)/2} \exp\left\{ -\frac{1}{2} tr\left(\mathbf{q}^{-1}\mathbf{R} \right) \right\}$$

with $v > p - 1$. The prior concentration around $E(Q) = \dfrac{R}{v - p - 1}$ increases with v.

Inference on a covariance matrix

Let $\mathbf{X}_i = (X_1, \ldots, X_p)^T \sim N_p(\mathbf{0}, \Sigma_p)$, for $i = 1,...,n$. We'd like to estimate (i.e., find a posterior distribution for) Σ_p. Note that Σ_p is the same across each $\mathbf{X}_i$.

Let $D = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ be a random sample from $N_p(\mathbf{0}, \Sigma_p)$. The Likelihood function is:

So, let's find $\pi(\Sigma_p \mid D)$.