常用概率分布

伯努利分布&贝塔分布

1. 基本概念

伯努利分布（Bernoulli Distribution）

- **定义**：描述单次二元实验结果的概率分布，结果只有两种可能（成功/失败），记为 $X \sim \text{Bernoulli}(p)$，其中 $p \in [0,1]$ 是成功的概率。
- **概率质量函数（PMF）**：

$$P(X=1)=p, \quad P(X=0)=1-p.$$

贝塔分布（Beta Distribution）

- **定义**：定义在区间 [0,1] 上的连续概率分布，常用于建模未知概率的分布（如伯努利实验的成功概率 p）。记为 $p \sim \text{Beta}(\alpha,\beta)$，其中 $\alpha > 0, \beta > 0$ 是形状参数。
- **概率密度函数（PDF）**：

$$f(p;\alpha,\beta)=\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)},$$

  其中 $B(\alpha,\beta)$ 是贝塔函数，作为归一化常数。

2. 核心关系：共轭先验（Conjugate Prior）

在贝叶斯统计中，**贝塔分布是伯努利分布的共轭先验**。这意味着：

- **先验假设**：若我们假设伯努利分布的参数 p 服从贝塔分布（即 $p \sim \text{Beta}(\alpha,\beta)$），
- **后验更新**：在观测到伯努利数据后，后验分布 $p \mid \text{数据}$ 仍服从贝塔分布。

数学推导

1. **似然函数**：假设观测到 k 次成功（即 $X=1$）和 $n-k$ 次失败（即 $X=0$），则似然函数为：

$$L(p;k,n-k)=p^k(1-p)^{n-k}.$$

2. **先验分布**：假设先验为 $\text{Beta}(\alpha,\beta)$，
3. **后验分布**：根据贝叶斯定理，后验正比于先验与似然的乘积：

$$\begin{aligned} f(p \mid k,n-k) &\propto f(p;\alpha,\beta)\cdot L(p;k,n-k) \\ &\propto p^{\alpha-1}(1-p)^{\beta-1}\cdot p^k(1-p)^{n-k} \\ &= p^{\alpha+k-1}(1-p)^{\beta+n-k-1}. \end{aligned}$$

   因此，后验分布为：

$$p \mid k,n-k \sim \text{Beta}(\alpha+k,\beta+n-k).$$

参数意义：α 和 β 可解释为先验中“成功”和“失败”的伪计数（pseudo-counts），例如：

- 若 $\alpha = \beta = 1$，先验为均匀分布，表示无先验知识；
- 若 $\alpha = 2, \beta = 2$，等价于先验中隐含了 1 次成功和 1 次失败的经验。

3. 应用场景

例：抛硬币实验。假设我们想估计硬币正面朝上的概率 p。

- 先验：$p \sim \text{Beta}(1,1)$（均匀分布）；
- 观测到 3 次正面、2 次反面（即 $k = 3, n = 5$）；
- 后验：$p \sim \text{Beta}(1+3, 1+2) = \text{Beta}(4,3)$。

参数更新过程：每次观测后，只需更新贝塔分布的参数：

- 成功次数 k 累加到 α
- 失败次数 $n - k$ 累加到 β

可视化解释：贝塔分布的形状随参数变化：

- $\alpha > \beta$：分布右偏（支持 $p > 0.5$）
- $\alpha = \beta$：对称分布（如 $\text{Beta}(1,1)$）
- $\alpha < \beta$：左偏（支持 $p < 0.5$）

4. 为什么选择贝塔分布？

- **数学便利性**：共轭性使得后验解析解易得，无需复杂数值积分。
- **灵活性**：通过调整 α 和 β，可表达各种先验信念（如保守估计、专家经验等）。
- **概率的自然建模**：贝塔分布定义在 [0,1] 区间，适合描述概率参数的不确定性。

分类分布&狄利克雷分布

分类分布（Categorical Distribution）和狄利克雷分布（Dirichlet Distribution）之间的关系与伯努利分布和贝塔分布的关系非常类似，核心在于**狄利克雷分布是分类分布的共轭先验**。这种关系在贝叶斯统计中具有重要意义，尤其在处理多类别概率建模时。以下是详细分析：

1. 基本概念

分类分布（Categorical Distribution）

- **定义**：描述单次多类别实验结果的概率分布，结果有 K 种可能（如掷骰子的 6 个面）。记为 $\boldsymbol{X} \sim \text{Categorical}(\boldsymbol{p})$，其中 $\boldsymbol{p} = (p_1, p_2, \ldots, p_K)$ 是每个类别的概率，满足 $\sum_{k=1}^{K} p_k = 1$。
- **概率质量函数（PMF）**：

$$P(\boldsymbol{X} = \mathbf{e}_k) = p_k,$$

  其中 $\mathbf{e}_k$ 是第 k 个类别的独热编码向量（One-Hot Vector）。

狄利克雷分布（Dirichlet Distribution）

- **定义**：定义在 K 维单纯形（Simplex）上的连续概率分布，用于建模 K 个非负概率参数的联合分布（如分类分布的参数 $\boldsymbol{p}$）。记为 $\boldsymbol{p} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha})$，其中 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K)$ 是正实数参数。
- **概率密度函数（PDF）**：

$$f(\boldsymbol{p}; \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^{K} p_k^{\alpha_k - 1},$$

其中 $B(\boldsymbol{\alpha})$ 是多变量贝塔函数，定义为：

$$B(\boldsymbol{\alpha}) = \frac{\prod_{k=1}^{K} \Gamma(\alpha_k)}{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)}.$$

---

2. 核心关系：共轭先验（Conjugate Prior）

在贝叶斯统计中，**狄利克雷分布是分类分布的共轭先验**。这意味着：

- **先验假设**：若我们假设分类分布的参数 $\boldsymbol{p}$ 服从狄利克雷分布（即 $\boldsymbol{p} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha})$），
- **后验更新**：在观测到分类数据后，后验分布 $\boldsymbol{p} \mid \text{数据}$ 仍服从狄利克雷分布。

数学推导

1. **似然函数**：假设观测到 N 次独立分类实验，其中第 k 类出现 n_k 次（$\sum_{k=1}^{K} n_k = N$），则似然函数为：

$$L(\boldsymbol{p}; \boldsymbol{n}) = \prod_{k=1}^{K} p_k^{n_k}.$$

2. **先验分布**：假设先验为 $\mathrm{Dirichlet}(\boldsymbol{\alpha})$，
3. **后验分布**：根据贝叶斯定理，后验正比于先验与似然的乘积：

$$\begin{aligned} f(\boldsymbol{p} \mid \boldsymbol{n}) &\propto f(\boldsymbol{p}; \boldsymbol{\alpha}) \cdot L(\boldsymbol{p}; \boldsymbol{n}) \\ &\propto \left(\prod_{k=1}^{K} p_k^{\alpha_k - 1}\right) \cdot \left(\prod_{k=1}^{K} p_k^{n_k}\right) \\ &= \prod_{k=1}^{K} p_k^{\alpha_k + n_k - 1}. \end{aligned}$$

因此，后验分布为：

$$\boldsymbol{p} \mid \boldsymbol{n} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha} + \boldsymbol{n}),$$

即每个参数 α_k 被更新为 $\alpha_k + n_k$。

参数意义：α_k 可解释为先验中第 k 类的伪计数（Pseudo-counts）。例如：

- 若 $\alpha_k = 1$ 对所有 k，先验为均匀分布，表示无先验知识；
- 若 α_k 较大，表示对第 k 类的强先验信念。

3. 与多项分布的扩展关系

- **分类分布**是单次多类别实验，**多项分布**（Multinomial Distribution）是 N 次独立分类实验的计数分布。
- 狄利克雷分布同样是多项分布参数 $\boldsymbol{p}$ 的共轭先验：

$$\boldsymbol{p} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha}), \quad \boldsymbol{X} \sim \mathrm{Multinomial}(N, \boldsymbol{p}) \Rightarrow \boldsymbol{p} \mid \boldsymbol{X} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha} + \boldsymbol{n}).$$

4. 应用场景

例：掷骰子实验。假设我们想估计骰子的公平性（6 个面）。

- 先验：$\boldsymbol{p} \sim \text{Dirichlet}(1,1,1,1,1,1)$（均匀分布）；
- 观测到 60 次实验中，各面出现次数为 $\boldsymbol{n} = (10,10,10,10,10,10)$；
- 后验：$\boldsymbol{p} \sim \text{Dirichlet}(11,11,11,11,11,11)$。

参数更新过程：每次观测后，只需更新狄利克雷分布的参数：

- 每个类别的观测次数 n_k 累加到对应的 α_k。

可视化解释：狄利克雷分布的形状随参数变化：

- 所有 α_k 相等：对称分布（如 $\boldsymbol{\alpha} = (1,1,1)$）；
- 某个 α_k 特别大：分布集中在该类别的概率附近；
- 所有 $\alpha_k < 1$：分布集中在单纯形的顶点（稀疏性）。

5. 与伯努利-贝塔关系的对比

特性	伯努利分布 + 贝塔分布	分类分布 + 狄利克雷分布
问题类型	二元结果（成功/失败）	多类别结果（K 种可能）
先验分布	贝塔分布 $\text{Beta}(\alpha, \beta)$	狄利克雷分布 $\text{Dirichlet}(\boldsymbol{\alpha})$
后验更新规则	$\alpha + k, \beta + n - k$	$\boldsymbol{\alpha} + \boldsymbol{n}$
参数维度	一维（单个概率 p）	多维（K 个概率 $p_1, \ldots, p_K$）
应用场景	A/B 测试、点击率预测	文本分类、主题建模、多类别概率估计

6. 为什么选择狄利克雷分布?

- **数学便利性**：共轭性使得后验解析解易得，无需复杂数值积分。
- **灵活性**：通过调整 $\boldsymbol{\alpha}$，可表达各种先验信念（如均匀分布、稀疏性、偏好某些类别）。
- **概率的自然建模**：狄利克雷分布定义在单纯形上，适合描述多类别概率参数的不确定性。

总结

分类分布与狄利克雷分布的关系是伯努利-贝塔关系的多维扩展：

- **分类分布**描述单次多类别实验的结果；
- **狄利克雷分布**作为其共轭先验，提供了一种动态更新多类别概率参数信念的机制。
- 这种关系简化了贝叶斯推断的计算，使得在观测数据后，后验分布仍能保持解析形式，便于实际应用。

一元正态分布&正态逆伽马分布

一元正态分布（Normal Distribution）与正态逆伽马分布（Normal-Inverse-Gamma Distribution）之间的关系是贝叶斯统计中的重要概念。**正态逆伽马分布是单变量正态分布（均值和方差均未知）的共轭先验分布**，这一关系在贝叶斯推断中用于同时估计正态分布的均值（μ）和方差（σ^2）。以下是详细解析：

1. 基本概念

一元正态分布（Normal Distribution）

- **定义**：描述连续随机变量的对称钟形分布，参数为均值 μ 和方差 σ²（即 $X \sim \mathcal{N}(\mu, \sigma^2)$）。
- **概率密度函数（PDF）**：

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

正态逆伽马分布（Normal-Inverse-Gamma Distribution）

- **定义**：联合分布，用于建模正态分布的均值 μ 和方差 σ² 的不确定性，记为 $(\mu, \sigma^2) \sim \mathcal{N}\text{-}\Gamma^{-1}(\mu_0, \lambda, \alpha, \beta)$。
- **参数含义**：
  - μ_0：均值的先验中心；
  - $\lambda > 0$：控制均值的先验精度（与样本量类似）；
  - $\alpha > 0$：逆伽马分布的形状参数；
  - $\beta > 0$：逆伽马分布的尺度参数。
- **概率密度函数（PDF）**：

$$f(\mu, \sigma^2; \mu_0, \lambda, \alpha, \beta) = \underbrace{\mathcal{N}\left(\mu; \mu_0, \frac{\sigma^2}{\lambda}\right)}_{\text{正态分布（均值的条件分布）}} \cdot \underbrace{\Gamma^{-1}\left(\sigma^2; \alpha, \beta\right)}_{\text{逆伽马分布（方差的边缘分布）}},$$

  其中：
  - 正态分布部分：$\mu \mid \sigma^2 \sim \mathcal{N}(\mu_0, \sigma^2/\lambda)$；
  - 逆伽马分布部分：$\sigma^2 \sim \Gamma^{-1}(\alpha, \beta)$。

2. 核心关系：共轭先验（Conjugate Prior）

在贝叶斯统计中，**正态逆伽马分布是单变量正态分布（均值和方差均未知）的共轭先验**。这意味着：

- **先验假设**：若我们假设正态分布的参数 (μ, σ^2) 服从正态逆伽马分布，
- **后验更新**：在观测到正态数据后，后验分布 $(\mu, \sigma^2) \mid \text{数据}$ 仍服从正态逆伽马分布。

数学推导

1. **似然函数**：假设观测到独立同分布的正态数据 $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$，则似然函数为：

$$L(\mu, \sigma^2; \boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right).$$

   简化后：

$$L(\mu, \sigma^2; \boldsymbol{x}) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\left[n(\mu-\bar{x})^2 + \sum_{i=1}^{n}(x_i-\bar{x})^2\right]\right),$$

   其中 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ 是样本均值。

2. **先验分布**：假设先验为 $(\mu, \sigma^2) \sim \mathcal{N}\text{-}\Gamma^{-1}(\mu_0, \lambda, \alpha, \beta)$，
   - 条件正态分布：$\mu \mid \sigma^2 \sim \mathcal{N}(\mu_0, \sigma^2/\lambda)$；
   - 边缘逆伽马分布：$\sigma^2 \sim \Gamma^{-1}(\alpha, \beta)$。
3. **后验分布**：根据贝叶斯定理，后验正比于先验与似然的乘积：

$$f(\mu, \sigma^2 \mid \boldsymbol{x}) \propto f(\mu, \sigma^2) \cdot L(\mu, \sigma^2; \boldsymbol{x}).$$

通过代数推导，后验分布为：

$$(\mu, \sigma^2) \mid \boldsymbol{x} \sim \mathcal{N}\text{-}\Gamma^{-1}(\mu_n, \lambda_n, \alpha_n, \beta_n),$$

其中参数更新规则为：

- $\lambda_n = \lambda + n,$
- $\mu_n = \frac{\lambda\mu_0 + n\bar{x}}{\lambda_n},$
- $\alpha_n = \alpha + \frac{n}{2},$
- $\beta_n = \beta + \frac{1}{2}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{\lambda n(\bar{x} - \mu_0)^2}{\lambda_n}\right).$

参数意义

- **均值更新**：μ_n 是先验均值 μ_0 和样本均值 $\bar{x}$ 的加权平均，权重由 λ 和 n 决定。
- **方差更新**：
  - α_n 反映数据量 n 对方差不确定性的修正；
  - β_n 包含样本方差和先验均值与样本均值差异的联合贡献。

3. 与正态伽马分布的区别

- **正态逆伽马分布**：以方差 σ^2 为参数，适用于正态分布方差的建模；
- **正态伽马分布**：以精度 $\tau = 1/\sigma^2$ 为参数，适用于精度的建模（此时联合分布为正态伽马分布）。

4. 应用场景

例：估计某地区居民身高的均值和方差。

- 先验：$\mu \sim \mathcal{N}(170, \sigma^2/1)$，$\sigma^2 \sim \Gamma^{-1}(2, 10)$；
- 观测到 100 人的身高数据，均值 $\bar{x} = 172$，样本方差 $s^2 = 25$；
- 后验参数更新：
  - $\lambda_n = 1 + 100 = 101,$
  - $\mu_n = \frac{1\cdot170 + 100\cdot172}{101} \approx 171.98,$
  - $\alpha_n = 2 + 50 = 52,$
  - $\beta_n = 10 + \frac{1}{2}\left(99 \cdot 25 + \frac{1\cdot100\cdot(172-170)^2}{101}\right) \approx 10 + 1237.5 + 1.96 \approx 1249.46$。

参数更新过程：每次观测后，只需更新正态逆伽马分布的四个参数：

- λ 增加样本量 n；
- μ_n 结合先验均值和样本均值；
- α 增加 $n/2$；
- β 增加与样本方差和均值差异相关的项。

5. 为什么选择正态逆伽马分布？

- **数学便利性**：共轭性使得后验解析解易得，无需复杂数值积分；
- **灵活性**：通过调整 $\mu_0, \lambda, \alpha, \beta$，可表达各种先验知识（如保守估计、专家经验）；

- **联合建模**：同时处理均值和方差的不确定性，避免固定参数带来的偏差。

总结

一元正态分布与正态逆伽马分布的关系是贝叶斯统计中的关键工具：

- **正态分布**描述单变量连续数据；
- **正态逆伽马分布**作为其共轭先验，提供了一种动态更新均值和方差联合信念的机制；
- 这种关系简化了贝叶斯推断的计算，使得在观测数据后，后验分布仍能保持解析形式，便于实际应用。

多元正态分布&正态逆维希特分布

多元正态分布（Multivariate Normal Distribution）与正态逆维希特分布（Normal-Inverse-Wishart Distribution）之间的关系是贝叶斯统计中多变量参数建模的核心工具。**正态逆维希特分布是多元正态分布（均值向量和协方差矩阵均未知）的共轭先验分布**，这一关系允许在观测数据后，通过简单的参数更新规则得到后验分布。以下是详细分析：

1. 基本概念

多元正态分布（Multivariate Normal Distribution）

- **定义**：描述多维连续随机变量的联合分布，参数为均值向量 $\boldsymbol{\mu}$ 和协方差矩阵 $\mathbf{\Sigma}$（即 $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$）。
- **概率密度函数（PDF）**：

$$f(\mathbf{x}; \boldsymbol{\mu}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),$$

  其中 d 是变量维度。

逆维希特分布（Inverse-Wishart Distribution）

- **定义**：定义在正定矩阵上的分布，用于建模协方差矩阵 $\mathbf{\Sigma}$ 的不确定性，记为 $\mathbf{\Sigma} \sim \mathcal{W}^{-1}(\mathbf{\Psi}, \nu)$，其中：
  - $\mathbf{\Psi}$：尺度矩阵（正定矩阵）；
  - $\nu > d-1$：自由度参数。
- **概率密度函数（PDF）**：

$$f(\mathbf{\Sigma}; \mathbf{\Psi}, \nu) = \frac{|\mathbf{\Psi}|^{\nu/2}}{2^{\nu d/2}\Gamma_d(\nu/2)}|\mathbf{\Sigma}|^{-(\nu+d+1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi}\mathbf{\Sigma}^{-1})\right),$$

  其中 $\Gamma_d(\cdot)$ 是多变量伽马函数。

正态逆维希特分布（Normal-Inverse-Wishart Distribution）

- **定义**：联合分布，用于建模多元正态分布的均值向量 $\boldsymbol{\mu}$ 和协方差矩阵 $\mathbf{\Sigma}$ 的不确定性，记为 $(\boldsymbol{\mu}, \mathbf{\Sigma}) \sim \mathcal{N}\text{-}\mathcal{W}^{-1}(\boldsymbol{\mu}_0, \lambda, \mathbf{\Psi}, \nu)$。
- **参数含义**：
  - $\boldsymbol{\mu}_0$：均值向量的先验中心；
  - $\lambda > 0$：控制均值的先验精度；
  - $\mathbf{\Psi}$：协方差矩阵的先验尺度矩阵；
  - ν：逆维希特分布的自由度。

- **概率密度函数（PDF）**：

$$f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \underbrace{\mathcal{N}\left(\boldsymbol{\mu}; \boldsymbol{\mu}_0, \frac{1}{\lambda}\boldsymbol{\Sigma}\right)}_{\text{正态分布（均值的条件分布）}} \cdot \underbrace{\mathcal{W}^{-1}(\boldsymbol{\Sigma}; \boldsymbol{\Psi}, \nu)}_{\text{逆维希特分布（协方差的边缘分布）}}.$$

2. 核心关系：共轭先验（Conjugate Prior）

在贝叶斯统计中，**正态逆维希特分布是多元正态分布（均值和协方差矩阵均未知）的共轭先验**。这意味着：

- **先验假设**：若我们假设多元正态分布的参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 服从正态逆维希特分布，
- **后验更新**：在观测到多元正态数据后，后验分布 $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \mid \text{数据}$ 仍服从正态逆维希特分布。

数学推导

1. **似然函数**：假设观测到独立同分布的多元正态数据 $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$，则似然函数为：

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathcal{D}) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right).$$

   简化后：

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathcal{D}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left(-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\left[n(\boldsymbol{\mu} - \bar{\mathbf{x}})(\boldsymbol{\mu} - \bar{\mathbf{x}})^\top + \mathbf{S}\right]\right)\right),$$

   其中：

   - $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$ 是样本均值；
   - $\mathbf{S} = \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$ 是样本协方差矩阵。

2. **先验分布**：假设先验为 $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \sim \mathcal{N}\text{-}\mathcal{W}^{-1}(\boldsymbol{\mu}_0, \lambda, \boldsymbol{\Psi}, \nu)$，
   - 条件正态分布：$\boldsymbol{\mu} \mid \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}/\lambda)$；
   - 边缘逆维希特分布：$\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}, \nu)$。

3. **后验分布**：根据贝叶斯定理，后验正比于先验与似然的乘积：

$$f(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathcal{D}) \propto f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \cdot L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathcal{D}).$$

   通过代数推导，后验分布为：

$$(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \mid \mathcal{D} \sim \mathcal{N}\text{-}\mathcal{W}^{-1}(\boldsymbol{\mu}_n, \lambda_n, \boldsymbol{\Psi}_n, \nu_n),$$

   其中参数更新规则为：

   - $\lambda_n = \lambda + n,$
   - $\boldsymbol{\mu}_n = \frac{\lambda\boldsymbol{\mu}_0 + n\bar{\mathbf{x}}}{\lambda_n},$
   - $\nu_n = \nu + n,$
   - $\boldsymbol{\Psi}_n = \boldsymbol{\Psi} + \mathbf{S} + \frac{\lambda n}{\lambda_n}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top.$

参数意义

- **均值更新**：$\boldsymbol{\mu}_n$ 是先验均值 $\boldsymbol{\mu}_0$ 和样本均值 $\bar{\mathbf{x}}$ 的加权平均，权重由 λ 和 n 决定。
- **协方差更新**：
  - ν_n 反映数据量 n 对协方差不确定性的修正；
  - $\boldsymbol{\Psi}_n$ 包含样本协方差矩阵 $\mathbf{S}$、先验均值与样本均值差异的联合贡献。

3. 应用场景

例：估计某地区居民身高和体重的联合分布参数（均值向量和协方差矩阵）。

- 先验：$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}/\lambda)$，$\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}, \nu)$；
- 观测到 n 个样本后，更新后验参数 $\boldsymbol{\mu}_n, \lambda_n, \boldsymbol{\Psi}_n, \nu_n$。

参数更新过程：每次观测后，只需更新正态逆维希特分布的四个参数：

- λ 增加样本量 n；
- $\boldsymbol{\mu}_n$ 结合先验均值和样本均值；
- ν 增加 n；
- $\boldsymbol{\Psi}_n$ 更新包含样本协方差矩阵和均值差异的项。

4. 与一元正态逆伽马关系的对比

特性	一元正态 + 正态逆伽马	多元正态 + 正态逆维希特
参数类型	均值 μ 和方差 σ^2	均值向量 $\boldsymbol{\mu}$ 和协方差矩阵 $\boldsymbol{\Sigma}$
先验分布	正态逆伽马 $\mathcal{N}$-Γ^{-1}	正态逆维希特 $\mathcal{N}$-$\mathcal{W}^{-1}$
后验更新规则	解析更新均值和方差参数	解析更新均值向量和协方差矩阵参数
参数维度	标量和一维参数	向量和矩阵参数
应用场景	单变量数据分析	多变量数据分析（如金融、生物统计）

5. 为什么选择正态逆维希特分布?

- **数学便利性**：共轭性使得后验解析解易得，无需复杂数值积分；
- **灵活性**：通过调整 $\boldsymbol{\mu}_0, \lambda, \boldsymbol{\Psi}, \nu$，可表达各种先验知识（如均匀分布、稀疏协方差）；
- **联合建模**：同时处理均值向量和协方差矩阵的不确定性，避免固定参数带来的偏差。

总结

多元正态分布与正态逆维希特分布的关系是贝叶斯统计中多变量参数建模的核心：

- **多元正态分布**描述多维连续数据；
- **正态逆维希特分布**作为其共轭先验，提供了一种动态更新均值向量和协方差矩阵联合信念的机制；
- 这种关系简化了贝叶斯推断的计算，使得在观测数据后，后验分布仍能保持解析形式，便于实际应用（如金融建模、机器学习等）。

共轭性

共轭性（Conjugacy）是贝叶斯统计中一个核心概念，其本质是**先验分布与后验分布属于同一概率分布族**，使得贝叶斯推断的计算变得简洁高效。通过共轭先验（Conjugate Prior），我们可以在观测数据后，直接通过参数更新规则得到后验分布，而无需进行复杂的积分运算。以下结合你之前提到的概率分布例子，深入浅出地解释这一概念。

1. 共轭性的定义

- **数学定义**：若先验分布 $P(\theta)$ 和似然函数 $P(X \mid \theta)$ 的乘积导致后验分布 $P(\theta \mid X)$ 与先验属于同一分布族，则称该先验为似然函数的**共轭先验**。
- **贝叶斯公式**：

$$\underbrace{P(\theta \mid X)}_{\text{后验}} \propto \underbrace{P(X \mid \theta)}_{\text{似然}} \cdot \underbrace{P(\theta)}_{\text{先验}}$$

  若后验与先验同分布族，则称为共轭关系。

2. 核心思想：先验与后验的“家族传承”

共轭性的本质是：**先验分布与后验分布共享相同的数学形式**，只是参数被观测数据更新了。这种“家族传承”的特性使得贝叶斯推断的计算变得高效，尤其适合迭代更新（如在线学习、实时数据处理）。

3. 经典共轭对的例子

(1) 伯努利分布 + 贝塔分布

- **问题**：抛硬币实验，估计正面朝上的概率 p。
- **共轭关系**：
  - **先验**：$p \sim \text{Beta}(\alpha, \beta)$（表示先验知识，如“成功次数 α 和失败次数 β”）；
  - **似然**：$X \sim \text{Bernoulli}(p)$；
  - **后验**：$p \mid X \sim \text{Beta}(\alpha + k, \beta + n - k)$，其中 k 是观测到的成功次数，n 是总试验次数。
- **直观理解**：贝塔分布像一个“计数器”，先验参数 α, β 是伪计数，观测数据 $k, n-k$ 直接累加到参数上。

(2) 分类分布 + 狄利克雷分布

- **问题**：掷骰子实验，估计每个面的概率 $\boldsymbol{p} = (p_1, \ldots, p_K)$。
- **共轭关系**：
  - **先验**：$\boldsymbol{p} \sim \text{Dirichlet}(\boldsymbol{\alpha})$（伪计数向量 $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$）；
  - **似然**：$X \sim \text{Categorical}(\boldsymbol{p})$；
  - **后验**：$\boldsymbol{p} \mid X \sim \text{Dirichlet}(\boldsymbol{\alpha} + \boldsymbol{n})$，其中 $\boldsymbol{n} = (n_1, \ldots, n_K)$ 是观测到的各类别计数。
- **直观理解**：狄利克雷分布是贝塔分布的多维扩展，每个维度的伪计数 α_k 直接加上观测次数 n_k。

(3) 一元正态分布 + 正态逆伽马分布

- **问题**：估计身高数据的均值 μ 和方差 σ^2。
- **共轭关系**：
  - **先验**：$(\mu, \sigma^2) \sim \mathcal{N}\text{-}\Gamma^{-1}(\mu_0, \lambda, \alpha, \beta)$；
  - **似然**：$X \sim \mathcal{N}(\mu, \sigma^2)$；
  - **后验**：参数更新规则为：

$$\begin{aligned}
\lambda_n &= \lambda + n, \\
\mu_n &= \frac{\lambda \mu_0 + n\bar{x}}{\lambda_n}, \\
\alpha_n &= \alpha + \frac{n}{2}, \\
\beta_n &= \beta + \frac{1}{2}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{\lambda n(\bar{x} - \mu_0)^2}{\lambda_n}\right).
\end{aligned}$$

- **直观理解**：正态逆伽马分布联合建模均值和方差，观测数据通过更新参数 $\lambda, \mu, \alpha, \beta$ 来修正先验。

(4) 多元正态分布 + 正态逆维希特分布

- **问题**：估计身高和体重的联合分布参数（均值向量 $\boldsymbol{\mu}$ 和协方差矩阵 $\boldsymbol{\Sigma}$）。
- **共轭关系**：
  - **先验**：$(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \sim \mathcal{N}\text{-}\mathcal{W}^{-1}(\boldsymbol{\mu}_0, \lambda, \boldsymbol{\Psi}, \nu)$；
  - **似然**：$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$；
  - **后验**：参数更新规则为：

$$\begin{aligned} \lambda_n &= \lambda + n, \\ \boldsymbol{\mu}_n &= \frac{\lambda \boldsymbol{\mu}_0 + n\bar{\mathbf{x}}}{\lambda_n}, \\ \nu_n &= \nu + n, \\ \boldsymbol{\Psi}_n &= \boldsymbol{\Psi} + \mathbf{S} + \frac{\lambda n}{\lambda_n}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top. \end{aligned}$$

- **直观理解**：正态逆维希特分布是正态逆伽马的多维扩展，联合更新均值向量和协方差矩阵。

4. 共轭性的核心优势

- **解析解**：后验分布的数学形式已知，无需数值积分或近似方法（如MCMC）。
- **参数更新**：只需调整分布参数，而非重新拟合整个分布。
- **灵活性**：通过调整先验参数，可表达不同强度的先验知识（如保守估计、专家经验）。
- **迭代性**：适合在线学习，每次观测后动态更新后验。

5. 类比：共轭性就像“积木搭积木”

想象你在搭积木：

- **先验分布** 是一块基础积木（如贝塔分布）；
- **观测数据** 是新的积木块（如伯努利实验的结果）；
- **共轭性** 保证新积木可以完美叠加在旧积木上，形成更高但形状相同的结构（后验分布仍是贝塔分布）。

6. 为什么共轭性重要?

- **简化计算**：避免复杂积分，直接通过参数更新得到后验。
- **可解释性**：参数更新规则直观（如伪计数、权重平均）。
- **应用广泛**：适用于A/B测试、推荐系统、金融建模、自然语言处理等领域。

7. 总结

共轭性是贝叶斯统计中“先验与后验同家族”的特性，其核心价值在于：

1. **数学便利**：后验分布形式已知，仅需更新参数；
2. **直观解释**：参数更新规则对应观测数据的累积效应；
3. **实用高效**：适合实时更新和大规模数据分析。