
# 贝叶斯数据分析

## 贝叶斯数据分析笔记

作者：韩方成

组织：中国海洋大学水产学院

时间：Dec 5, 2021

别回头，跟着光

# 目录

# 第 1 章 概率和推理

## 1.1 贝叶斯数据分析的三大步骤

贝叶斯数据分析的过程可以被分为一下三步：

1. 建立全概率模型，即一个问题中所有可观测和不可观测的量的联合概率密度。
2. 计算后验概率：在给定已知数据下未知参数的分布。
3. 评估模型的拟合程度。

## 1.2 贝叶斯推理

### 1.2.1 概率符号

我们定义变异系数为 $\mathrm{sd}(\theta)/E(\theta)$，几何平均数为 $\exp(E[\log(\theta)])$，并且几何标准差为 $\exp(\mathrm{sd}[\log(\theta)])$。

### 1.2.2 贝叶斯法则

贝叶斯法则可以写为：

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)} = \frac{p(\theta)p(y|\theta)}{\int_\theta p(\theta)p(y|\theta)d\theta} \tag{1.1}$$

因为在给定 $y$ 的条件下 $p(y)$ 为常数，与 $\theta$ 无关，所以我们常将1.1写为式1.2所示的非标准化后验分布：

$$p(\theta|y) \propto p(\theta)p(y|\theta) \tag{1.2}$$

### 1.2.3 预测

为了对一个未知的可观测数据做预测，即预测推断。在数据 $y$ 被考虑之前，未知但是可观测的 $y$ 的分布为

$$p(y) = \int p(y,\theta)d\theta = \int p(\theta)p(y|\theta)d\theta \tag{1.3}$$

这被称为先验预测分布：先验是因为它没有以任何先前的观测为条件，预测是因为他是一个可观测数据的分布。

当数据 $y$ 被观测到后，我们可以用同样的方式来预测一个未知可观测的 $\tilde{y}$。$\tilde{y}$ 的分布被称为后验预测分布，后验是因为它以观测 $y$ 为条件，预测是因为他是可观测数据 $\tilde{y}$ 的预测：

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y},\theta|y)d\theta \\ &= \int p(\tilde{y}|\theta,y)p(\theta|y)d\theta \\ &= \int p(\tilde{y}|\theta)p(\theta|y)d\theta \end{aligned} \tag{1.4}$$

### 1.2.4 似然和比率

在给定模型下在点 $\theta_1$ 和 $\theta_2$ 的后验密度 $p(\theta|y)$ 比值称为 $\theta_1$ 相对于 $\theta_2$ 的后验几率：

$$\frac{p(\theta_1|y)}{p(\theta_2|y)} = \frac{p(\theta_1)p(y|\theta_1)/p(y)}{p(\theta_2)p(y|\theta_2)/p(y)} = \frac{p(\theta_1)}{p(\theta_2)}\frac{p(y|\theta_1)}{p(y|\theta_2)} \tag{1.5}$$

由此可以看出后验比率是先验比率和似然比例的乘积。

## 1.3 概率论中的一些有用的结果

我们经常用下列方式表示随机变量 $u$ 的期望：

$$E(u)=E(E(u|v)) \tag{1.6}$$

公式1.6很好证明：

$$\iint up(u,v)dudv=\iint u\,p(u|v)du\,p(v)dv=\int E(u|v)p(v)dv \tag{1.7}$$

方差包括两项：

$$\operatorname{var}(u)=E(\operatorname{var}(u|v))+\operatorname{var}(E(u|v)) \tag{1.8}$$

式1.8的证明为：

$$\begin{aligned}
\mathrm{E}(\operatorname{var}(u\mid v))+\operatorname{var}(\mathrm{E}(u\mid v)) &= \mathrm{E}\left(\mathrm{E}\left(u^2\mid v\right)-(\mathrm{E}(u\mid v))^2\right)+\mathrm{E}\left((\mathrm{E}(u\mid v))^2\right)-(\mathrm{E}(\mathrm{E}(u\mid v)))^2\\
&=\mathrm{E}\left(u^2\right)-\mathrm{E}\left((\mathrm{E}(u\mid v))^2\right)+\mathrm{E}\left((\mathrm{E}(u\mid v))^2\right)-(\mathrm{E}(u))^2\\
&=\mathrm{E}\left(u^2\right)-(\mathrm{E}(u))^2\\
&=\operatorname{var}(u)
\end{aligned}$$

### 1.3.1 变量的变换

如果 $p_u$ 为一个离散分布，并且 $f$ 为一个一对一的函数，而 $v=f(u)$，则 $v$ 的概率密度为

$$p_v(v)=p_u(f^{-1}(v))$$

如果 $f$ 是一个多对一函数，那么会对 $p_v(v)$ 有一个求和。

如果 $p_u$ 为一个连续分布且 $v=f(u)$ 是一个一对一的函数，则

$$p_v(v)=|J|p_u(f^{-1}(v))$$

其中 $|J|$ 为变换 $u=f^{-1}(v)$ 的雅可比行列式，如果 $f$ 为多对一函数，则 $p_v(v)$ 上有积分或求和项。

当处理一维变量时，我们经常利用对数函数来将参数空间从 $(0,\infty)$ 转换到 $(-\infty,\infty)$。当处理唯一开单位区间 $(0,1)$ 上的参数时，我们用 logistic 变换：

$$\operatorname{logit}(u)=\log\left(\frac{u}{1-u}\right) \tag{1.9}$$

它的逆变换为：

$$\operatorname{logit}^{-1}(v)=\frac{e^v}{1+e^v}$$

**定理 1.1**

如果 $X$ 是一个连续随机变量，并且它的累积概率密度函数为 $F_X$，则 $U=F_X(X)\sim U(0,1)$ ♡

**证明** 在证明之前我们先定义：

$$F_X^{-1}(u)=\inf\{x:F_X(x)=u\},0<u<1$$

如果随机变量 $U\sim U(0,1)$，则对于所有 $x\in\mathbb{R}$：

$$\begin{aligned}
P\left(F_X^{-1}(U)\le x\right)&=P\left(\inf\{t:F_X(t)=U\}\le x\right)\\
&=P\left(U\le F_X(x)\right)\\
&=F_U\left(F_X(x)\right)\\
&=F_X(x)
\end{aligned}$$

得证。

# 第 A 章 Gamma,Beta 分布

以下内容基于 LDA 的数学八卦。

## A.1 Gamma 函数

**定义 A.1 (Gamma 函数)**

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt \tag{A.1}$$

通过分布积分的方法，可以推导出这个函数有如下的递归性质：

$$\Gamma(x+1) = x\Gamma(x)$$

于是很容易证明，$\Gamma(x)$ 函数可以当成是阶乘在实数集上的延拓，具有如下性质：

$$\Gamma(n) = (n-1)!$$

♣

那么 Gamma 函数是怎么来的呢？这个来源于哥德巴赫在将阶乘推广到实数域的问题上，欧拉解决了这个问题，他发现 $n!$ 可以用如下的一个无穷乘积表达：

$$\left[\left(\frac{2}{1}\right)^n \frac{1}{n+1}\right]\left[\left(\frac{3}{2}\right)^n \frac{2}{n+2}\right]\left[\left(\frac{4}{3}\right)^n \frac{3}{n+3}\right]\cdots = n! \tag{A.2}$$

用极限形式，这个式子整理后可以写为

$$\lim_{m\to\infty} \frac{1\cdot 2\cdot 3\cdots m}{(1+n)(2+n)\cdots(m+n)}(m+1)^n = n! \tag{A.3}$$

左边可以整理为：

$$\begin{aligned}
&\frac{1\cdot 2\cdot 3\cdots m}{(1+n)(2+n)\cdots(m+n)}(m+1)^n \\
=&1\cdot 2\cdot 3\cdots n\cdot \frac{(n+1)(n+2)m}{(1+n)(2+n)\cdots m}\cdot \frac{(m+1)^n}{(m+1)(m+2)\cdots(m+n)} \\
=&n!\frac{(m+1)^n}{(m+1)(m+2)\cdots(m+n)} \\
=&n!\prod_{k=1}^{n}\frac{m+1}{m+k} \to n! \quad (m\to\infty)
\end{aligned}$$

所以上式成立。

之后欧拉对得到的式子进行了处理，得到了积分的形式：

$$n! = \int_0^1 (-\log t)^n dt$$

如果我们令 $t = e^{-u}$，就可得我们常见的 Gamma 函数形式：

$$n! = \int_0^\infty u^n e^{-u} du$$

于是，利用上式把阶乘延拓到实数集上，我们就得到了 Gamma 函数的一般形式：

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$$

Gamma 函数不仅可以用来计算 (1/2)!，还可以拓展到许多其他的数学概念。比如导数，我们可以把导数的定义拓展到实数集，从而可以计算 1/2 阶导数。我们考虑 $x^n$ 的导数，我们知道，$x^n$ 的 $k$ 阶导数为：

$$n(n-1)(n-2)\cdots(n-k+1)x^{n-k} = \frac{n!}{(n-k)!}x^{n-k}$$

由于 $k$ 阶导数可以用阶乘表达，于是我们用 Gamma 函数表达式为：

$$\frac{\Gamma(n+1)}{\Gamma(n-k+1)}x^{n-k}$$

于是基于上式，我们可以把导数的阶从整数延拓到实数集。例如，取 $n=1, k=\frac{1}{2}$，我们可以计算 $x$ 的 $\frac{1}{2}$ 阶导数为：

$$\frac{\Gamma(1+1)}{\Gamma(1-1/2+1)}x^{1-1/2}=\frac{2\sqrt{x}}{\sqrt{\pi}}$$

Gamma 函数的图像如图A.1所示，由此可以看出 Gamma 函数为一个凸函数。不仅如此，$\log(\Gamma(x))$ 也是一个凸函数。我们可以用以下定理证明：

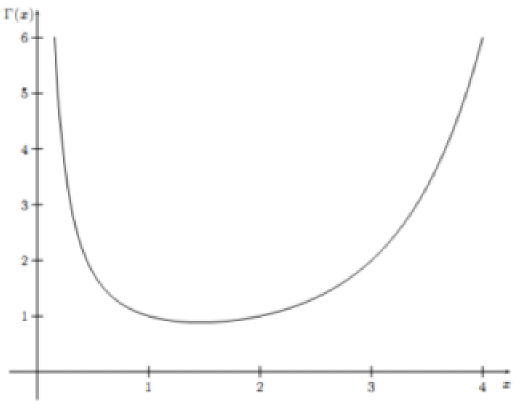


图 A.1: $\Gamma(x)$

**定理 A.1 (Bohr-Mullerup)**

如果 $f:(0,\infty)\to(0,\infty)$，且满足

1. $f(1)=1$
2. $f(x+1)=xf(x)$
3. $\log f(x)$ 是凸函数

那么 $f(x)=\Gamma(x)$，也就是 $\Gamma(x)$ 是唯一满足以上条件的函数。

♡

如下函数被称为 Digamma 函数，

$$\Psi(x)=\frac{d\log\Gamma(x)}{dx}$$

这也是一个很重要的函数，在涉及求 Dirichlet 分布相关的参数的极大似然估计时，往往需要用到这个函数。Digamma 函数具有如下一个漂亮的性质：

$$\Psi(x+1)=\Psi(x)+\frac{1}{x}$$

## A.2 从二项分布到 Gamma 分布

对 Gamma 函数的定义做一个变形，就可以得到如下式子：

$$\int_0^\infty \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}dx=1$$

于是，取积分中的函数作为概率密度，就得到一个形式最简单的 Gamma 分布的密度函数：

$$\mathrm{Gamma}(x|\alpha)=\frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}$$

如果做一个变换 $x = \beta t$，就得到 Gamma 分布的更一般形式

$$\text{Gamma}(t|\alpha, \beta) = \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}$$

其中 $\alpha$ 为 shape parameter，主要决定分布曲线的形状，而 $\beta$ 称为 rate parameter 或者 inverse scale parameter($\frac{1}{\beta}$ 被称为 scale parameter)，主要决定曲线有多陡。如图A.2所示。

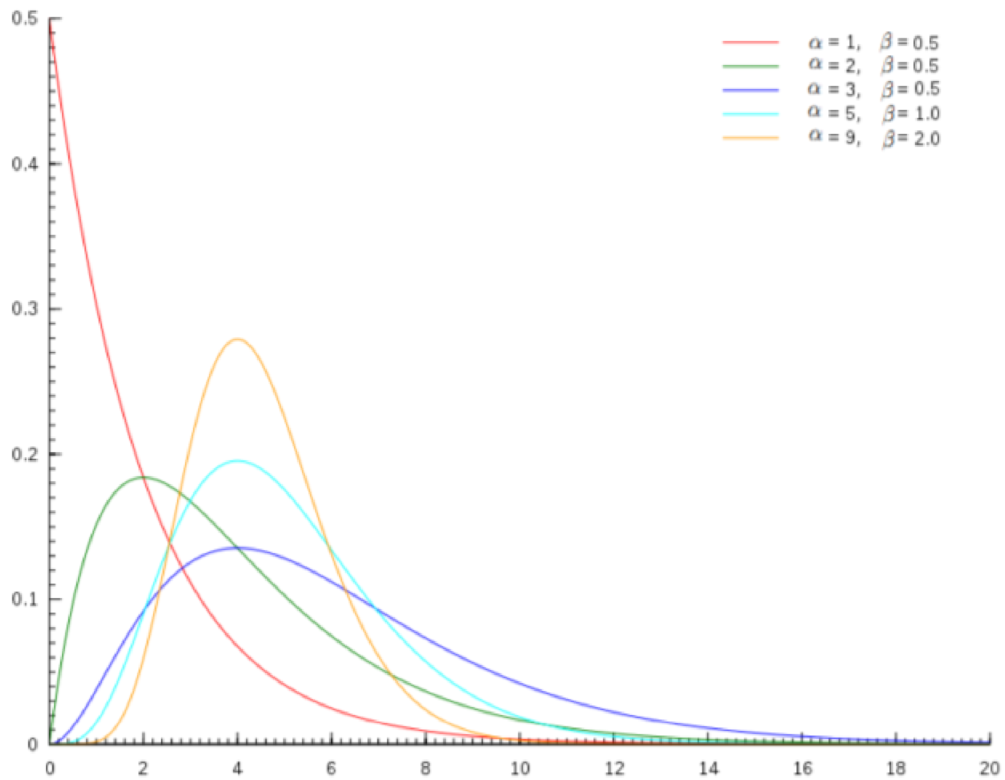


图 A.2: $\Gamma$ 分布

下面我们先讨论 $\beta = 1$ 情况。Gamma 分布首先与 Poisson 分布、Poisson 过程发生密切的联系。参数为 $\lambda$ 的 Poisson 分布，概率写为：

$$\text{Poisson}(X = k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

在 Gamma 分布的密度函数取 $\alpha = k + 1$ 得到

$$\text{Gamma}(x|\alpha = k + 1) = \frac{x^k e^{-x}}{k!}$$