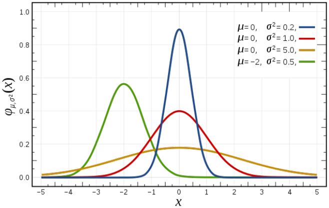
Cannlytics

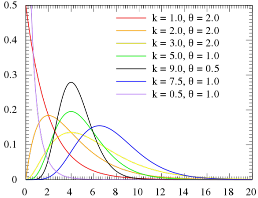
Cannabis Data Science

Saturday Morning Statistics #13

February 26th, 2022

Pertinent Distributions





Normal Distribution

Notation: $\mathcal{N}(\mu, \sigma^2)$

Parameters: mean $\mu \in \mathbb{R}$,
variance $\sigma^2 \in \mathbb{R}_{>0}$

PDF:$p(x|\mu, \sigma^2) =$
$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

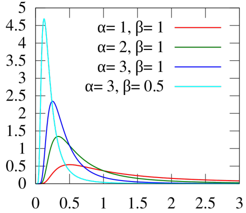
Gamma Distribution

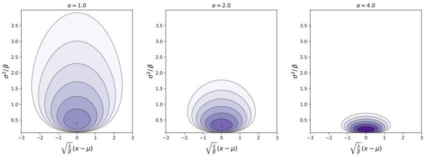
Notation: $Y \sim G(\alpha, \beta)$

Parameters: shape $\alpha > 0$, rate $\beta > 0$

PDF:$f_\gamma(y|\alpha, \beta) \equiv$
$$\begin{cases} (\beta^\alpha \Gamma(\alpha))^{-1} y^{\alpha-1} \exp(-y/\beta) & \text{if } 0 < y < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Pertinent Distributions - Continued





Inverse-gamma Distribution

Notation: $Y \sim \mathcal{IG}(\alpha, \beta)$

Parameters: shape $\alpha > 0$, scale $\beta > 0$

PDF: $p(y|\alpha, \beta) = [\Gamma(\alpha)\beta^{\alpha}]^{-1} y^{-(\alpha+1)} \exp(-1/[y\beta])$

Normal-inverse-gamma Distribution

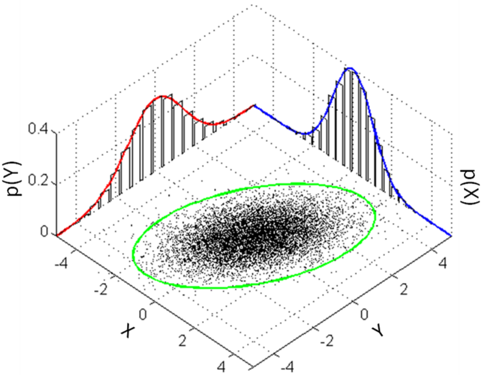
Notation: $(x, \sigma^2) \sim \mathcal{NIG}(\mu, \lambda, \alpha, \beta)$

Parameters: mean μ, variance σ^2/λ, shape $\alpha > 0$, scale $\beta > 0$

PDF: $f(x, \sigma^2|\mu, \lambda, \alpha, \beta) = \frac{\sqrt{\lambda}}{\sqrt{2\pi\sigma^2}} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left(-\frac{2\beta+\lambda(x-\mu)^2}{2\sigma^2}\right)$

Credit: IkamusumeFan (Inverse-gamma distribution figure), Peter.komar.hu (Normal-inverse-gamma distribution figure)

Multivariate Normal Distribution



Notation: $Y \sim \mathcal{N}(\mu, \Sigma)$

Parameters: mean $\mu \in \mathbb{R}^k$, variance $\Sigma \in \mathbb{R}^{k \times k}$

PDF*: $\phi(y|\mu, \Sigma) = \frac{1}{2\pi^{\frac{k}{2}}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right]$

* The PDF only exists where Σ is positive definite.

Linear Regressions

Given $i = 1, ..., N$ observations, the linear regression assumes a linear relationship between the dependent variable, y_i, and a vector of k explanatory variables, written as

$$y = X\beta + \epsilon,$$

where $y = [y_1 \ y_2 \ \cdots \ y_N]'$ is a $N \times 1$ matrix,

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

is a $N \times K$ matrix, and $\epsilon = [\epsilon_1 \ \epsilon_2 \ \cdots \ \epsilon_N]'$ is a $N \times 1$ matrix.

The Classical Linear Regression

One of the classical assumption is that errors are **normally distributed**

$$\epsilon \sim \mathcal{N}(0_N, \sigma^2 I_N),$$

Assuming a **multivariate normal distribution** between Y and X, then

$$p(y|\beta, h) = \phi(y; X\beta, h^{-1} I_N),$$

where $h \equiv \sigma^{-2}$. Assumptions about ϵ and X define the likelihood function

$$L(\beta, h) = \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} \exp\left[-\frac{h}{2}(y - X\beta)'(y - X\beta)\right].$$

The parameters can be estimated by **maximum likelihood** or with the **ordinary least squares (OLS)** quantities

$$\hat{\beta} = (X'X)^{-1}X'y.$$

The Bayesian Approach to Linear Regressions

The data is assumed to be sampled from a distribution:

$$y \sim \mathcal{N}(\beta^T X, h^{-1} I_N)$$

The posterior probability is:

$$P(\beta, h | y, X) = \frac{P(y|\beta, h, X) P(\beta, h|X)}{P(y|X)}$$

We can derive the posterior:

$$P(\beta, h|y, X) \propto L(\beta, h) P(\beta, h|X)$$

Deriving the Bayesian Linear Regression Posterior

Substituting in the p.d.f.s of the likelihood and our priors allows us to find a closed-form solution

$$P(\beta, h|y) \propto L(\beta, h)P(\beta, h)$$
$$P(\beta, h|y) = \phi(y; X\beta, h^{-1}I_N)\phi(\beta; \underline{\beta}, h^{-1}\underline{Q})f_\gamma(h|\underline{s}^{-2}, \underline{v})$$

where

- $\underline{\beta}$, $\underline{Q}$, $\underline{s}$, and $\underline{v}$ are hyperparameters of our priors,
- $\phi()$ is the multivariate normal p.d.f.,
- $f_\gamma()$ is the gamma density.

The resulting posterior has a **normal-gamma** density, we have established *c*onjugacy. Our posterior has the same distribution as our prior and our priors are **conjugate priors**.

Solving the Bayesian Linear Regression

After substituting the p.d.f.s, our **priors** can be found and written in terms of ordinary least squares (OLS) quantities

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$\overline{Q} = (\underline{Q}^{-1} + X'X)^{-1}$$

$$\overline{\beta} = \overline{Q}(\underline{Q}^{-1}\underline{\beta} + X'X\hat{\beta})$$

$$\overline{vs}^2 = \underline{vs}^2 + (y - X\hat{\beta})'(y - X\hat{\beta}) + (\hat{\beta} - \underline{\beta})'X'X\overline{Q}\underline{Q}^{-1}(\hat{\beta} - \underline{\beta})$$

The **posterior** distributions are established to be **normal-gamma**

$$\beta|y, h \sim \mathcal{N}(\overline{\beta}, h^{-1}\overline{Q}^{-1})$$

$$h|y \sim \gamma(\overline{s}^{-2}, \overline{v}).$$

Estimating the Bayesian Linear Regression Solution

An alternative **posterior** parametrization is with a **normal-inverse-gamma** distribution*

$$\beta|\sigma^2, y \sim \mathcal{N}(\overline{\mu}_\beta, \overline{V}_\beta)$$

$$\sigma^2|\beta, y \sim \mathcal{IG}(\overline{a}, \overline{b}).$$

with hyperparameters μ_β, V_β, a, and b and where

$$\overline{V}_\beta = \left(\frac{X'X}{\sigma^2} + V_\beta^{-1}\right)^{-1}$$

$$\overline{\mu}_\beta = \overline{V}_\beta \left(\frac{X'Y}{\sigma^2} + V_\beta^{-1}\mu_\beta\right)$$

$$\overline{a} = a + \frac{n}{2}$$

$$\overline{b} = \left(\frac{1}{b} + \frac{1}{2}(Y - X\beta)'(Y - X\beta)\right)^{-1}.$$

*A theorem states that the inverted gamma distribution has the property that, if Y has an inverted gamma distribution, then $1/Y$ has a gamma distribution.

Gibbs Sampling

Implementing Gibbs sampling to get a random sample from your **posterior** entails:

1. Start with an initial value for σ^2, the desired number of posterior draws, R, and the number of burn in draws, R_0, sufficient to converge to the posterior distribution.
2. Sample β^* from the posterior $\beta|\sigma^2, y$.
3. Sample σ^{*2} from the posterior $\sigma^2|\beta = \beta^* y$ given the draw, β^*.
4. Repeat steps (2) and (3) until you have draws for β^* and σ^{*2} equal to $R + R_0$, the number of desired posterior draws plus the number of burn in draws.
5. Drop the first number of burn in draws, R_0, for β^* and σ^{*2}.

🙏 Thank you for coming.

Lesson of the Day

- Bayesian analysis allows us to estimate a distribution for our parameters of interest.

References

Bayesian Econometric Methods, Koop, Poirier, and Tobias (2007).