Stuff I Should Know for Bayesian Statistics

D. Aaron Hillegass
hillegass@gatech.edu

February 25, 2019

Normal Distribution

A normal distribution has two parameters: the mean (θ) and the variance (σ^2). x can go from $-\infty$ to ∞. Its density is given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$$

The conjugate prior of the mean of a normal distribution is another normal distribution. Specifically, if $x_1, \ldots, x_n$, are identical and independent with a normal distribution $\mathcal{N}(\theta, \sigma^2)$ and the prior is $\theta \sim \mathcal{N}(\theta_0, \sigma_0^2)$,

$$\theta | x_1, \ldots, x_n \sim \mathcal{N}\left(\frac{\frac{\sigma^2}{n}\theta_0 + \sigma_0^2 \bar{x}}{\frac{\sigma^2}{n} + \sigma_0^2}, \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \right)^{-1} \right)$$

The conjugate prior of the variance of a normal distribution is an inverse gamma distribution.

Gamma Distribution

The gamma distribution has two parameters: α and β. x can go from 0 to ∞ Its density is given by

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad \text{for } x > 0 \text{ and } \alpha, \beta > 0,$$

The mean of the distribution is

$$E[x] = \frac{\alpha}{\beta}$$

If $\alpha \geq 1$, the mode is

$$\frac{\alpha - 1}{\beta}$$

Inverse Gamma Distribution

An inverse gamma distribution has two parameters: α and β. x can go from 0 to ∞. Its density is given by

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}(1/x)^{\alpha+1}e^{\left(-\frac{\beta}{x}\right)}$$

If $\alpha > 1$, the mean of the distribution is

$$E[x] = \frac{\beta}{\alpha - 1}$$

The mode is

$$\frac{\beta}{\alpha + 1}$$

Binomial Distribution

The binomial distribution is a discrete distribution that takes one parameter: π is the probability of one success in one try. The probability of getting exactly k successes in n trials is given by

$$p(k) = \binom{n}{k}\pi^k(1-\pi)^{n-k}$$

The mean is $n\pi$.

The conjugate prior of π is the Beta distribution.

Beta Distribution

The beta distribution takes two parameters: α and β. Its density is given by

$$p(x) = \frac{1}{Beta(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

The mode is

$$\frac{\alpha-1}{\alpha+\beta-2}$$

Negative Binomial Distribution

The negative binomial distribution is a discrete probability distribution that takes two parameters: r and p. For any $k \in 0, 1, 2, \ldots$,

$$p(k) = \binom{k+r-1}{k}(1-p)^r p^k$$

The mean is

$$\mathbb{E}\left(p(k)\right) = \frac{pr}{1-p}$$

If $r \leq 1$, the mode is 0. Otherwise it is $\lfloor \frac{p(r-1)}{1-p} \rfloor$

The variance is

$$\frac{pr}{(1-p)^2}$$

Exponential Distribution

The exponential distribution takes one parameter λ. Its probability density is

$$p(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Note that this is a special case of the gamma distribution.

Its cumulative distribution function is

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Its mean is $\frac{1}{\lambda}$. Its median is $\frac{\log 2}{\lambda}$. Its mode is 0. Its variance is $\frac{1}{\lambda^2}$.

The conjugate prior for λ in the exponential distribution is a gamma distribution.

Poisson Distribution

The Poisson Distribution is a discrete distribution that takes one parameter: λ. Its distribution is given by

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x \in \{0, 1, \ldots\}$$

λ is both its mean and its variance.

The congugate prior for λ is the gamma distribution

Pareto Distribution

The Type 1 Pareto Distribution has two parameters α and x_{m}, where x_{m} is the minimum possible value. Both x_{m} and α must be positive. Its density is

$$p(x) = \begin{cases} \frac{\alpha x_{\mathrm{m}}^{\alpha}}{x^{\alpha+1}} & x \geq x_{\mathrm{m}}, \\ 0 & x < x_{\mathrm{m}}. \end{cases}$$

The mean of the distribution is

$$\mathrm{E}(X) = \begin{cases} \infty & \alpha \leq 1, \\ \frac{\alpha x_{\mathrm{m}}}{\alpha - 1} & \alpha > 1. \end{cases}$$

Its cumulative distribution function is

$$F(x) = \begin{cases} 1 - \left(\frac{x_{\mathrm{m}}}{x}\right)^{\alpha} & x \geq x_{\mathrm{m}}, \\ 0 & x < x_{\mathrm{m}}. \end{cases}$$

Gamma Function

The gamma function is a generalization of factorial. That is, if n is a positive integer

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

But the gamma function is defined:

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

Beta Function

The beta function is

$$Beta(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\, dt = \frac{\Gamma(x)\,\Gamma(y)}{\Gamma(x+y)}$$

Choose Function

The choose function is:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Linear Regression

If you are trying to use p inputs $(x_1, \ldots, x_p)$ to predict an ouput y, you could create a vector of coefficients $(\beta_0, \ldots, \beta_n)$ such that

$$y \approx \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

So how do you figure out what the β is? You use the n data points you have: The column vector $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)'$ is the outcomes. The matrix $\boldsymbol{X}$ is the inputs. Row i of $\boldsymbol{X}$ starts with 1 and has the inputs $(x_{i,1}, x_{i,2}, \ldots, x_{i,p})$ that correspond to y_i. Thus X has n rows and $p+1$ columns.

Given the data you have, the sum of the squares of the errors for a column vector β (of length $p+1$) is given by

$$\boldsymbol{y}^T\boldsymbol{y} - 2\beta^T\boldsymbol{X}^T\boldsymbol{y} + \beta^T\boldsymbol{X}^T\boldsymbol{X}\beta$$

To minimize that, you set β to be:

$$\beta = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

Logarithmic Identities

$$\log\left(e^x\right) = x \qquad e^{\log x} = x$$

$$\log\left(ab\right) = \log a + \log b \qquad \log\left(\frac{a}{b}\right) = \log a - \log b$$

$$\log\left(a^b\right) = b\log a \qquad \log\left(\sqrt[b]{a}\right) = \frac{\log a}{b}$$

$$a^{\log b} = b^{\log a} \qquad x^{\frac{\log(a)}{\log(x)}} = a$$

$$\frac{d}{dx}\log x = \frac{1}{x} \qquad \log x = \int_1^x \frac{1}{t}dt$$

Stuff Bayesians Like To Calculate

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

$$\mathbb{E}(\theta|y) = \int \theta\ p(\theta|y)d\theta = \frac{\int \theta\ p(y|\theta)p(\theta)d\theta}{\int p(y|\theta)p(\theta)d\theta}$$

$$var(\theta|y) = \mathbb{E}(\theta^2|y) - \left(\mathbb{E}(\theta|y)\right)^2$$

$$p(\theta \in A \mid y) = \int_A p(\theta|y)d\theta$$

The posterior predictive distribution:

$$p(y \mid data) = \int p(y \mid \theta)p(\theta \mid data)d\theta$$