Learning about the World Through Bayesian Models

Tuesday, January 27, 2015 12:19 PM

- Objective of Bayesian analysis: to learn about characteristics of the world by combining our prior information with evidence in a rational way
- Today's objective: show multiple examples of how this can be done
    - Probabilities (binomial events)
    - Counts
    - Simple means and variances
    - Linear relationships (Bayesian regression)

Binomial events

Tuesday, January 27, 2015 12:49 PM

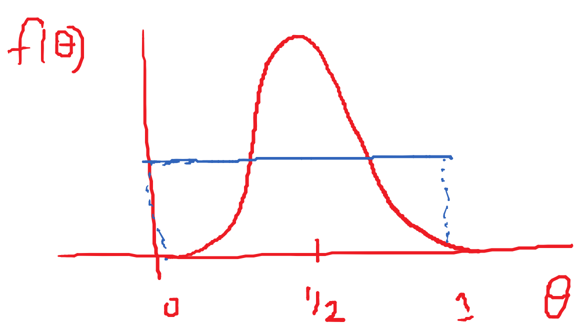
- Binomial event $x \in \{0,1\}$ occurs with probability $\theta \in [0,1]$

$\theta = pr(x=1)$

- Bayes rule:
  - $f(\theta|data) = \dfrac{f(data|\theta)f(\theta)}{\sum_\theta f(data|\theta)f(\theta)} \propto f(data|\theta)f(\theta)$

$\int f(\theta|data) = 1$

- Likelihood:
  - $f(data|\theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$ where there are k many $x = 1$ observations.
  - The binomial coefficient, which tells us how many possible combinations of binomial events could produce a pattern of k successes, is subsumed in the constant; we can ignore it when sampling.
- Prior: for conjugacy, use the beta
  - $f(\theta) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\boxed{\theta^{\alpha-1}(1-\theta)^{\beta-1}}$
  - α and β are hyperparameters indicating the equivalent number of observed successes and failures in the prior
  - The gamma functions are also subsumed into the prior; we can ignore when sampling.
- Posterior:
  - $f(\theta|data) \propto \theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}$
- Could use "flat" priors to imply no information about the location of θ
  - $f(\theta|data) \propto \theta^k(1-\theta)^{n-k}$



Count data

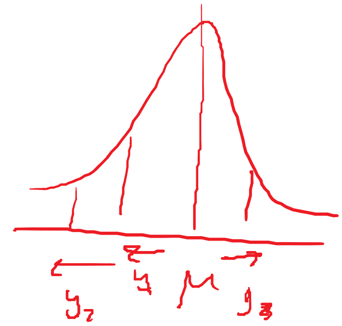
Tuesday, January 27, 2015 1:22 PM

- Count event $x \in \{0,1,2,\dots\}$ occurs; want to know cumulative and probability densities of y, the number of times that some event has occurred
- Likelihood: the Poisson function
  - $f(data|\lambda) = \frac{\exp(-n\lambda)\lambda^S}{\prod_{i=1}^n y_i!}$ where there are n many counts, each of which has y_i many events, and $S = \sum_{i=1}^n y_i$
  - Mean count = λ
- Prior: for conjugacy, use the gamma density
  - $f(\lambda) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}\exp(-b\lambda)$
  - a and b are hyperparameters; a is the prior "sum of counts", and b is the prior "sample size"
  - The ratio involving the gamma function is subsumed into the prior; we can ignore when sampling.
- Posterior:
  - $f(\theta|data) \propto \lambda^{S+a-1}\exp(-\lambda(n+b))$

Means and Variances

Tuesday, January 27, 2015 12:18 PM

- Continuous data $y \in (-\infty, \infty)$ occurs with average value of μ and a known standard deviation of σ
- Likelihood:
  - $f(data|\mu) = \prod_{i=1}^{n} \phi\left(\frac{y_i-\mu}{\sigma}\right) \propto \prod_{i=1}^{n} \exp\left[\frac{-(y_i-\mu)^2}{2\sigma^2}\right]$



- Prior: for conjugacy, use the normal. (Convenient!)
  - Hyperparameters: μ_0 and σ_0, prior belief about mean and standard deviation
- Posterior:
  - $f(\theta|data) \propto \phi(\mu^\star, \sigma^\star)$ with $\mu^\star = \frac{\mu_0\sigma_0^{-2} + \bar{y}\frac{n}{\sigma^2}}{\sigma_0^{-2} + \left(\frac{n}{\sigma^2}\right)}$ and ${\sigma^2}^\star = \left(\sigma_0^{-2} + \frac{n}{\sigma^2}\right)^{-1}$

$\bar{y}$ = sample avg. y

σ^2 = sample var of y.

Bayesian Regression

Saturday, September 01, 2012 9:33 PM

- A similar technique can be applied to regression analysis
  - we can specify prior densities for β and σ^2 (predictors of the conditional mean of $y|x$, and the variance of error around that mean)
  - run a regression analysis to determine the likelihood function (builds in the typical assumptions of a regression analysis)
  - combine the two to form posterior belief distribution
- $f(\beta|\sigma^2, y, x) = \Phi(b_1, \sigma^2 B_1)$
  $f(\sigma^2|y, x) = inv.gamma\left(\frac{v_1}{2}, \frac{v_1\sigma_1^2}{2}\right)$
  $f(\beta|y, x) = \int f(\beta|\sigma^2, y, x) f(\sigma^2|y, x) d\sigma^2 = t(b_1, \sigma_1^2 B_1, df = v_1)$
  - $b_1 = (B_0^{-1} + X'X)^{-1}(B_0^{-1} b_0 + X'X\hat{\beta})$
  - $B_1 = (B_0^{-1} + X'X)^{-1}$
  - $v_1 = v_0 + n$
  - $v_1\sigma_1^2 = v_0\sigma_0^2 + S + r$
  - S = sum of squared errors of the regression
  - r = $(b_0 - \hat{\beta})'(B_0 + (X'X)^{-1})^{-1}(b_0 - \hat{\beta})$

$f(\beta_x|y,x) = t(b_1[x], \sigma_1^2 B_1[x], df = v_1)$

- Bottom line: choose four things
  - b_0: a *k*x1 vector of prior means for a regression with *k* parameters
  - B_0: a *kxk* matrix so that $\sigma^2 B_0$ is the prior VCV of β
    - Prior density for β: $\Phi(b_0, \sigma^2 B_0)$
  - $v_0 > 0$, a degrees of freedom parameter for the prior
  - $\sigma_0^2 > 0$, a prior variance
    - Prior density for σ^2: $inv.gamma(v_0, \sigma_0^2 * v_0)$
- This will integrate with the information contained in the data set to create a closed form posterior density
  - For β: will be a multivariate Student's t density.