Curve fitting

Ramkumar R

August 19, 2009

- Random variables
- Probability functions
  - Discrete PMF
  - CDF

$$F_X(x) = Prob\{X \leq x\} \tag{1}$$

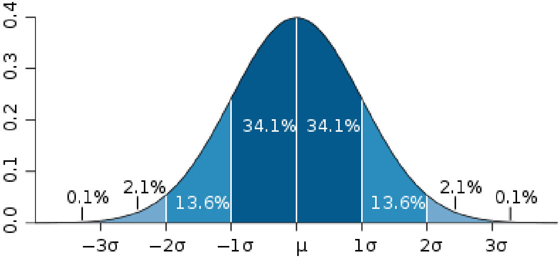
  - PDF

$$f_X(x) = F'(x) \tag{2}$$

Moments

- Central moment — $\langle (x-\mu)^m \rangle = \langle (x - \langle x \rangle)^m \rangle$
  - Zeroth and First — 1 and 0
  - Second — Variance $\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2$
  - Third and Fourth — Skewness and Kurtosis
- Non-central moment — $\langle X^m \rangle$

Gaussian (Normal)



$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{3}$$

Chi-square PDF

$$Q = \sum_{i=1}^{k} X_i^2 \tag{4}$$

$$f_Q(x;k) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} & ; \ x \geq 0 \\ 0 & ; \ x < 0 \end{cases} \tag{5}$$

where X_i's are normallly distributed random variables with 0 mean and variance 1.

Chi-square CDF

After integration, we get the Chi-square CDF in terms of an lower incomplete γ function defined as

$$\gamma(s, x) = \int_0^x t^{s-1} e^{-t} \, dt \tag{6}$$

Then the Chi-square CDF equals

$$\frac{\gamma(\frac{k}{2}, \frac{\chi^2}{2})}{\Gamma(\frac{k}{2})} \tag{7}$$

What does curve fitting mean?

- Fitting data to a model with adjustable parameters
- Design figure-of-merit function
- Obtain best fit parameters by adjusting parameters to achieve min in merit function
- More details to take care of
  - Assess appropriateness of model; goodness-of-fit
  - Accuracy of parameter determination
  - Merit function may not be unimodal

Least squares fitting

Maximizing the product

$$P \propto \prod_{i=1}^{N} \exp\left[-\frac{1}{2}\left(\frac{y_i - y_{i(t)}}{\sigma}\right)^2\right]\Delta y \tag{8}$$

ie. Minimizing its negative logarithm

$$\left[\sum_{i=1}^{N} \frac{[y_i - y_{i(t)}]^2}{2\sigma^2}\right] - N\log\Delta y \tag{9}$$

Minimizing this sum over $a_1, a_2, ..a_M$, we get the final form:

$$\sum_{i=1}^{N}[y_i - y_{i(t)}]^2 \tag{10}$$

Chi-square fitting

Modifying equation 8 to replace the σ by σ_i and going through the same process, we get

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - y_{i(t)}}{\sigma_i} \right)^2 \tag{11}$$

where

$$y_{i(t)} = f(a_1, a_2, ..a_M) \tag{12}$$

For models that are linear in the a's, however, it turns out that the probability distribution for different values of chi-square at its minimum can nevertheless be derived analytically, and is the chi-sqaure distribution for `N-M` degrees of freedom.

gnuplot

```
gnuplot> fit f(x) 'foo.data' u 1:2:3:4 via a, b
```

- Uses WSSR: Weighted Sum of Squared Residuals
- Marquardt-Levenberg algorithm to find parameters to use in next iteration
- After fitting, gnuplot reports *stdfit*, the standard deviation of the fit

- Numerical Receipies in C
- The gnuplot manual
- Miscellanous books on elementary probability
- Presentation created using LaTeX and Beamer

Presentation source code available on github.com/artagnon/curve-fitting