Massachusetts Institute of Technology

Department of Electrical Engineering & Computer Science

6.041/6.431: Probabilistic Systems Analysis

(Spring 2008)

Tutorial 8
April 17/18, 2008

1. For this problem we will need to compute the 1st, 2nd, 3rd, and 4th moments of the standard normal distribution. To facilitate this, we will find the moment generating function:

$$\begin{aligned}\mathbf{E}[e^{rx}] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{rx}\cdot e^{-\frac{1}{2}x^2}dx \\ &= \frac{1}{\sqrt{2\pi}}\cdot e^{\frac{1}{2}r^2}\cdot\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-r)^2}dx \\ &= e^{\frac{1}{2}r^2}\end{aligned}$$

the second equality above following from completing the square in the exponent, and the third following because the Gaussian density function integrates to 1. Therefore we now take derivatives w.r.t. r, and easily find the first four moments:

$$\mathbf{E}[X]=0, \qquad \mathbf{E}[X^2]=1, \qquad \mathbf{E}[X^3]=0, \qquad \mathbf{E}[X^4]=3.$$

We know that the correlation coefficient is given by:

$$\rho(X,Y)=\frac{\mathrm{cov}(X,Y)}{\sigma_X\sigma_Y}.$$

We first compute the covariance:

$$\begin{aligned}\mathrm{cov}(X,Y) &= \mathbf{E}[XY]-\mathbf{E}[X]\mathbf{E}[Y] \\ &= \mathbf{E}[aX+bX^2+cX^3]-\mathbf{E}[X]\mathbf{E}[Y] \\ &= a\mathbf{E}[X]+b\mathbf{E}[X^2]+c\mathbf{E}[X^3] \\ &= b.\end{aligned}$$

Now $\mathrm{var}(X)=1$ therefore $\sigma_X=1$ so we have left to find $\sigma_Y=\sqrt{\mathrm{var}(Y)}$.

$$\begin{aligned}\mathrm{var}(Y) &= \mathrm{var}(a+bX+cX^2) \\ &= \mathbf{E}[(a+bX+cX^2)^2]-\mathbf{E}[a+bX+cX^2]^2 \\ &= (a^2+2ac+b^2+3c^2)-(a^2+c^2+2ac) \\ &= b^2+2c^2\end{aligned}$$

and therefore we find:

$$\rho(X,Y)=\frac{b}{\sqrt{b^2+2c^2}}.$$

To find the best linear estimator, note that $E[Y]=a+c$. Use this and $\rho(X,Y)$ found above in the standard linear estimation formula to get:

$$\begin{aligned}\hat{Y}_{Linear}(X) &= E[Y]+\rho(X,Y)\frac{\sigma_Y}{\sigma_X}(X-E[X]) \\ &= a+c+\frac{b}{\sqrt{b^2+2c^2}}\sqrt{b^2+2c^2}X \\ &= a+c+bX\end{aligned}$$

2. (Problem 6.6 from the new chapter) By Bayes' rule we have

$$p_{\Theta|X}(\theta \mid x) = \frac{p_{X|\Theta}(x \mid \theta) p_\Theta(\theta)}{p_X(x)},$$

So it follows that

$$\begin{aligned} p_{\Theta|X}(\theta \mid x) &= \frac{p_{X|\Theta}(x \mid \theta) p_\Theta(\theta)}{p_X(x)} \\ &= \frac{p_{X|\Theta}(x \mid \theta) p_\Theta(\theta)}{\sum_{i=1}^{100} p_{X|\theta}(x \mid i) p_\Theta(i)} \\ &= \frac{\frac{1}{\theta} \cdot \frac{1}{100}}{\sum_{i=x}^{100} \frac{1}{i} \cdot \frac{1}{100}} \\ &= \begin{cases} \frac{\frac{1}{\theta}}{\sum_{i=x}^{100} \frac{1}{i}}, & \text{for } \theta = x, x+1, \ldots, 100, \\ 0, & \text{for } \theta = 1, 2, \ldots, x-1. \end{cases} \end{aligned}$$

The posterior probability is maximized at $\hat{\theta} = x$, which is the MAP estimate of Θ given x. The least squares estimator is

$$\hat{\Theta} = \mathbf{E}[\Theta \mid X] = \sum_{\theta=1}^{100} \theta p_{\theta|X}(\theta \mid X) = \frac{101 - X}{\sum_{i=X}^{100} \frac{1}{i}}.$$

Figure 1 plots the MAP and least squares estimates of Θ as a function of X.