M4056 Bayes Estimators. **October 20, 2010**

Background on joint distributions. Recall that if $f_{X,Y}(x,y)$ is a joint density, then the marginal and conditional density functions are defined as follows:

$$\text{marginals:} \quad f_X(x) := \int_{\infty}^{\infty} f_{X,Y}(x,y)\,dy \quad f_Y(y) := \int_{\infty}^{\infty} f_{X,Y}(x,y)\,dx,$$

$$\text{conditionals:} \quad f_{X|Y}(x|Y=y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad f_{Y|X}(y|X=x) := \frac{f_{X,Y}(x,y)\,dy}{f_X(x)}.$$

The "$Y = y$" may be shortened: $f_{X|Y}(x|y) = f_{X|Y}(x|Y = y)$. Similarly for X.

We also have marginals and conditionals in case X and Y are discrete. Suppose X has values in $\{1,\ldots,m\}$ and Y in $\{1,\ldots,n\}$. Each joint outcome can written as a pair (i,j), and we can picture the set of all pairs as a rectangular array. Denote its probability $p_{X,Y}(i,j)$. The marginal *pmf*s in this case are the row and column sums:

$$p_X(i) = \sum_{j=1}^{n} p(i,j) \quad \text{and} \quad p_Y(j) = \sum_{i=1}^{m} p(i,j).$$

The conditional *pmf*s are the normalized rows and columns themselves:

$$p_{X|Y}(i|j) := \frac{p_{X,Y}(i,j)}{p_Y(j)} \quad \text{and} \quad p_{Y|X}(j|i) := \frac{p_{X,Y}(i,j)}{f_X(i)}.$$

The Bayesian approach to parameter estimation treats the parameter as a random variable, not an unknown constant. We start with a presumed density for the parameter, which we call the *prior*, and we update the prior using the data, arriving at a revised density for the parameter, which we call the *posterior*. The details follow.

We introduce the symbol Θ to stand for the random variable that represents the parameter, and we use θ to stand for a fixed value. The *prior* is a density function for Θ. (In your book, this is called $\pi(\theta)$, but I will call it $f_\Theta(\theta)$.) The sample density is given as before by a function of the data $\vec{x}$ and the parameter θ , but now we think of this as a conditional density

$$f_{\vec{X}}(\vec{x}|\theta) = f_{\vec{X}|\Theta}(\vec{x}|\theta).$$

According to the formula for the conditional, the joint distribution of $\vec{X}$ and Θ is:

$$f_{\vec{X},\Theta}(\vec{x},\theta) = f_{\vec{X}|\Theta}(\vec{x}|\theta) f_\Theta(\theta).$$

The marginal of $\vec{X}$ is

$$f_{\vec{X}}(x) = \int_{\infty}^{\infty} f_{\vec{X},\Theta}(x,\theta)\,d\theta = \int_{\infty}^{\infty} f_{\vec{X}|\Theta}(x|\theta)\, f_\Theta(\theta)\,d\theta,$$

and the *posterior* density of Θ is

$$f_{\Theta|\vec{X}}(\theta|\vec{x}) = \frac{f_{\vec{X},\Theta}(x,\theta)}{f_{\vec{X}}(x)} = \frac{f_{\vec{X}|\Theta}(\vec{x}|\theta) f_\Theta(\theta)}{\int_{\infty}^{\infty} f_{\vec{X}|\Theta}(x|\theta)\, f_\Theta(\theta)\,d\theta}.$$

Exercise 7.22. Assume $X_1, \ldots, X_n$ i.i.d. normal(θ, σ^2) and that Θ has normal(μ, τ^2) prior. Then $\overline{X}$ is normal$(\theta, \sigma^2/n)$. For notational convenience, let $U = \overline{X}$.

$$\begin{aligned} f_{U,\Theta}(u,\theta) &= f_{U|\Theta}(u|\theta) f_\Theta(\theta) \\ &= \frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma} e^{-n(u-\theta)^2/(2\sigma^2)} \frac{1}{\sqrt{2\pi}\,\tau} e^{-(\theta-\mu)^2/(2\tau^2)} \\ &= \frac{\sqrt{n}}{2\pi\sigma\tau} \exp\Big(-\frac{n\tau^2(u-\theta)^2 + \sigma^2(\theta-\mu)^2}{2\sigma^2\tau^2} \Big). \end{aligned}$$

The (negative of the) exponent is quadratic in θ:

$$\frac{n\tau^2(u-\theta)^2 + \sigma^2(\theta-\mu)^2}{2\sigma^2\tau^2} = \frac{(n\tau^2+\sigma^2)\theta^2 - 2(n\tau^2 u + \mu\sigma^2)\theta + (n\tau^2 u^2 + \mu^2\sigma^2)}{2\sigma^2\tau^2}.$$

The posterior density of θ is a constant multiple of $f_{U|\Theta}(u|\theta) f_\Theta(\theta)$.

Given any function of the form $f(\theta) = k\, e^{-(a\theta^2+2b\theta+c)}$, there is a unique constant multiple of it which is a *pdf*. Since

$$a\,\theta^2 + b\,\theta + c = a\Big(\theta - \frac{-b}{2a}\Big)^2 + \Big(c - \frac{b^2}{4a}\Big),$$

we deduce

$$\frac{1}{2\sigma^2}(x-\mu)^2 = a\Big(x - \frac{-b}{2a}\Big)^2$$

so

$$\frac{-b}{2a} = \mu \quad \text{and} \quad \frac{1}{2a} = \sigma^2$$