University of Pittsburgh
CS 2750 Machine Learning
Professor Milos Hauskrecht

Handout 7
Feb 7, 2007

Solutions to Problem set 2

1 Problem 1

(a)

$$\sum_{x=0}^{\infty} p(x|\lambda) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

since

$$e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

then

$$\sum_{x=0}^{\infty} p(x|\lambda) = e^{-\lambda} e^{\lambda} = 1$$

(b)

$$E(x|\lambda) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = \lambda .1 = \lambda$$

(c) ML estimate of Poisson distribution is formulated in the most general form as

$$\lambda_{ML} = \arg\max_{\lambda} P(D|\lambda) = \arg\max_{\lambda} \prod_{i=1}^{n} P(x_i|\lambda)$$

$$l(\lambda; D) = \ln \prod_{i=1}^{n} P(x_i|\lambda) = \ln \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$= \sum_{i=1}^{n} \ln \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = -n\lambda + \sum_{i=1}^{n} (x_i \ln \lambda - x_i \ln x_i)$$

Now we differentiate with the respect to the parameter λ

$$\begin{aligned}
\frac{\partial l(\lambda; D)}{\partial \lambda} &= 0 \\
-n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i &= 0 \\
\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i &
\end{aligned}$$

(d) If the conjugate prior is Gamma distribution, the posterior probability $p(\lambda|y)$, where the prior $p(\lambda|a,b)$ is updated only by one data sample y, can be expressed as

$$\begin{aligned}
p(\lambda|y) &\propto p(y|\lambda)p(\lambda|a,b) \\
&= \frac{e^{-\lambda}\lambda^{y}}{y!}\frac{1}{b^{a}\Gamma(a)}\lambda^{a-1}e^{-\frac{\lambda}{b}} \\
&= \frac{1}{y!}\frac{1}{b^{a}\Gamma(a)}\lambda^{a+y-1}e^{-\frac{\lambda}{\frac{b}{b+1}}} \\
&\propto p(\lambda|a+y, b/(b+1))
\end{aligned}$$

The final step assumes that the expression can be renormalized, so the result is a distribution. Parameters a and b of prior Gamma distribution are updated to $a+y$ and $b/(b+1)$. This approach can be generalized for n inputs, so

$$\begin{aligned}
a_{new} &= a + \sum_{i=1}^{n} y_i \\
b_{new} &= \frac{b}{nb+1}
\end{aligned}$$

(e) Any function from the exponential family can be expressed in the form

$$f(x,\eta) = \frac{1}{Z(\eta)}h(x)\exp\left[\eta^{T}t(x)\right]$$

Poisson distribution is a member of this family because it can be rewritten as

$$p(x|\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$\begin{aligned}
\eta &= \ln\lambda \\
t(x) &= x \\
Z(\eta) &= e^{\lambda} \\
h(x) &= \frac{1}{x!}
\end{aligned}$$

(f) Figure 1 shows the plots of the probability functions for Poisson distributions.

(g) Maximum likelihood estimate λ_{ML} is computed as the mean of the input sequence.

$$\lambda_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i = 5.24$$

(h) The prior distributions of λ as gamma functions are plotted in figure 2.

(i) The posterior distributions of λ as gamma functions are plotted in figure 3. Both posterior distributions changed a lot comparing to their priors, which can be explained by the amount of data that were used to refine our first estimates. As the learning was govern by data, both posteriors distribution are closer to each other than the priors were.

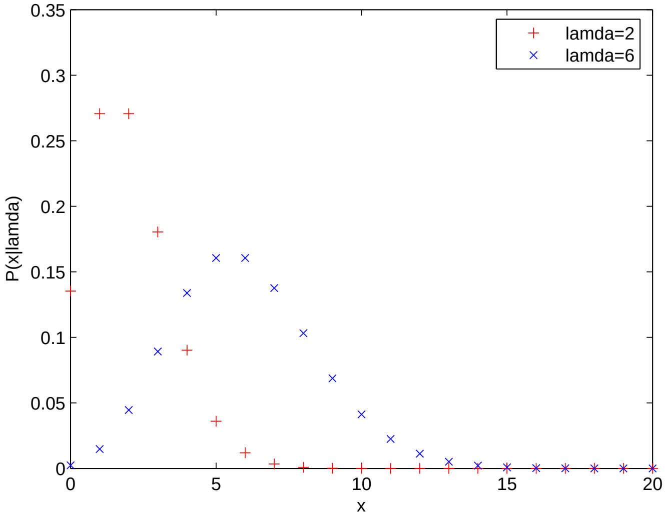

Figure 1: Poisson distributions with parameters $\lambda = 2$ and $\lambda = 6$.

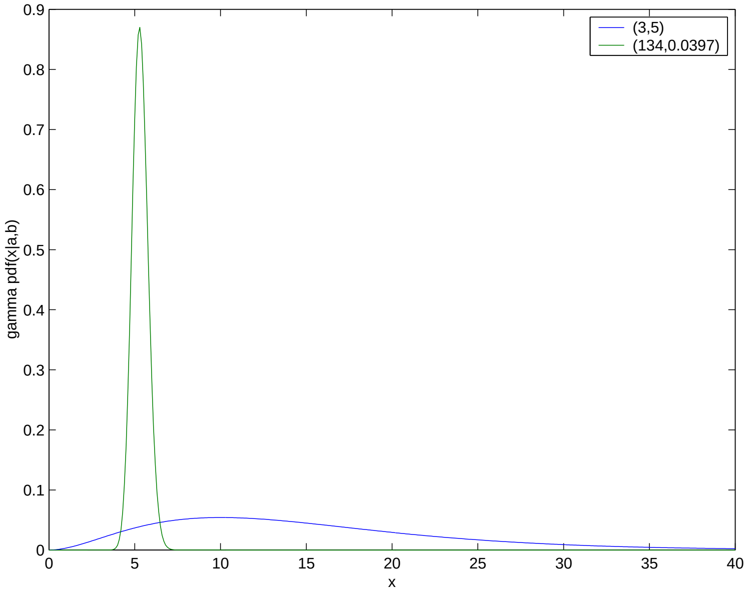

Figure 2: Prior gamma distributions with parameters $(a = 1, b = 2)$ and $(a = 3, b = 5)$.

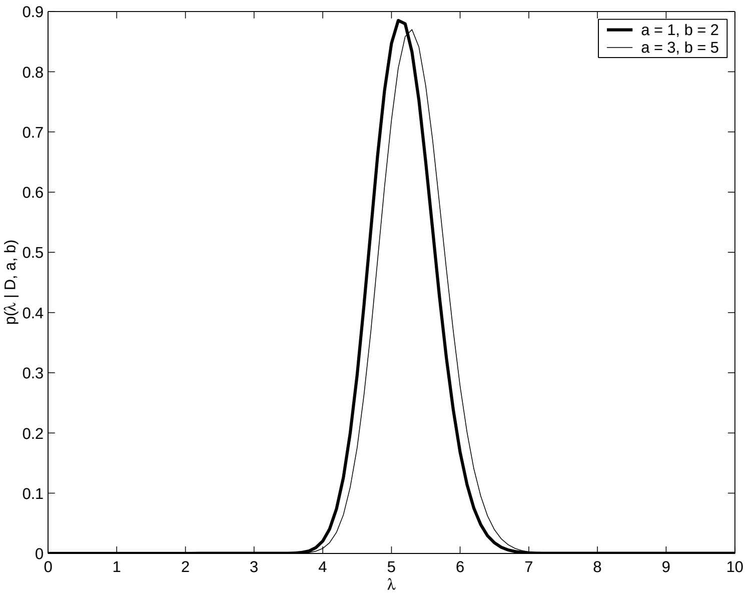

Figure 3: Posterior gamma distributions with parameters $(a = 1, b = 2)$ and $(a = 3, b = 5)$.