Matt Dickenson `mcd31`
STA561/CS571 — Fall 2013 **Homework 1** Due: 11 September, 2013

Homework Notes: I worked with Josh Cutler on this homework. To look up density functions I used `http://en.wikipedia.org/wiki/Poisson_distribution` and `http://en.wikipedia.org/wiki/Uniform_distribution_(continuous)`.

Problem 1

For two dice X_1 and X_2, the probability of getting a 5 is $P(X_i = 5) = \frac{1}{6}$. We can get the probability of getting a 5 on either die by $P(X_1 = 5 \cup X_2 = 5) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \mathbf{\frac{11}{36}}$ (where the subtraction term eliminates the double-counted portion).

Problem 2

Let d denote the disease state, nd denote "not disease", + a positive test result, and − a negative test result.

$$
\begin{aligned}
\pi_1(d|+) &= \frac{p(+|d)\pi_0(d)}{p(+)} && (1)\\
&= \frac{p(+|d)\pi_0(d)}{p(+|d)\pi_0(d) + p(+|nd)} && (2)\\
&= \frac{0.9 \times 0.001}{0.9 \times 0.001 + 0.2 \times 0.999} && (3)\\
&= 0.00448 && (4)
\end{aligned}
$$

Problem 3

a) For the probability desnity function, x has a density of $\frac{1}{\frac{1}{2}-0} = 2$ if $x \in [0, \frac{1}{2}]$ and 0 otherwise.

b) The density of the pdf at $x = 0.00027$ given $a = 0, b = \frac{1}{2}$ is 2.

c) $\int_{0.00027}^{0.00027} 2dx = 2 - 2 = 0$. For a continuous distribution the probability at any point has probability ≈ 0.

Problem 4

a) Exponential family distributions can be written in the form:

$$p(x|\lambda) = h(x)\exp(\eta^T T(x) - A(\eta))$$

If we take the Poisson parameterization as

$$\frac{\lambda^x}{x!} \times e^{-\lambda}$$

then we can write out its exponential form using

$$\begin{aligned} \eta &= \log\lambda & (5) \\ h(x) &= \frac{1}{x!} & (6) \\ T(x) &= x & (7) \\ A(\eta) &= e^\eta = \lambda & (8) \end{aligned}$$

as

$$\begin{aligned} p(x|\lambda) &= \frac{1}{x!}\exp((\log\lambda)x - \exp\eta) & (9) \\ &= \frac{1}{x!}\exp((\log\lambda)x - \lambda) & (10) \\ &= \frac{1}{x!}\lambda^x\exp(-\lambda) & (11) \end{aligned}$$

which is equivalent to our original parameterization.

b) As shown above, $T(x) = x$.

c) As shown above, $A(\eta) = e^\eta = \lambda$.

d) For our response function (to translate from η to λ), we use $\lambda = e^\eta$ (the inverse of $\eta = \log\lambda$).

e) To get $E[X]$ we can differentiate the log partition function $A(\eta) = e^\eta$:

$$\begin{aligned} E[X] &= \frac{\partial}{\partial\eta}e^\eta & (12) \\ &= e^\eta & (13) \\ &= \lambda & (14) \end{aligned}$$

For $Var(X)$ we take the second derivative:

$$\begin{aligned}
Var(X) &= \frac{\partial^2}{\partial \eta^2} e^\eta && (15)\\
&= \frac{\partial}{\partial \eta} e^\eta && (16)\\
&= e^\eta && (17)\\
&= \lambda && (18)
\end{aligned}$$

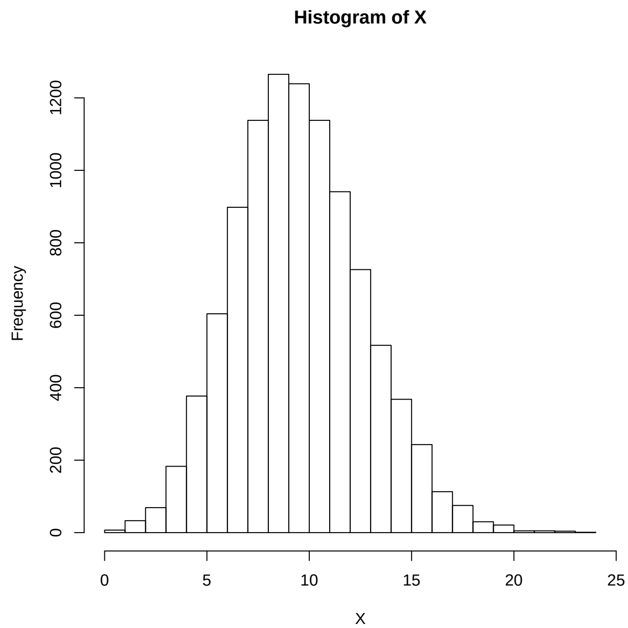
f) We can obtain the MLE estimate of λ, $\hat{\lambda}$, by taking the first derivative of the exponential familiy parameterization and setting it to zero:

$$\begin{aligned}
0 &= \frac{\partial}{\partial \eta}(\log h(x) + \eta \sum_{i=1}^{N} T(X_i) - ne^\eta) && (19)\\
&= 0 + \sum_{i=1}^{N} T(X_i) - ne^\eta && (20)\\
&= \sum_{i=1}^{N} x_i - n\lambda && (21)\\
n\hat{\lambda} &= \sum_{i=1}^{N} x_i && (22)\\
\hat{\lambda} &= \frac{1}{n}\sum_{i=1}^{N} x_i && (23)\\
&= \bar{x} && (24)
\end{aligned}$$

g) For $\lambda \sim Ga(\alpha_0, \beta_0)$ and $X \sim Pois(\lambda)$, the posterior hyperparameters on λ are $\alpha_1 = \alpha_0 + \sum_{i=1}^{n} x_i, \beta_1 = \beta_0 + n$ (thanks to the conjugacy of the Gamma and Poisson distributions). We can find the MAP of λ using the mode of this gamma posterior: $\frac{\alpha_0 + \sum_{i=1}^{n} x_i - 1}{\beta_0 + n}$. As a sanity check, it can be seen that with small prior hyperparamters, as we get more data the MAP estimate converges to the MLE estimate in part (f).

Problem 5

a) The histogram of data with 25 bins is shown below:



b) $\hat{\lambda} = 10.0171$

c) $\alpha = 1, \beta = 1$, $\text{MAP}(\lambda) = 10.0161$

d) $\alpha = 100, \beta = 1$, $\text{MAP}(\lambda) = 10.026$

e) $\alpha = 10, \beta = 1$, $\text{MAP}(\lambda) = 10.017$

f) Given the relatively large amount of data we have (relative to the magnitude of the hyperparameters), the α and β values have only a minor influence on our MAP estimate (at the tenths or hundreths place). That being said, I prefer the $\alpha = 10, \beta = 1$ parameterization because it is more informative than $\alpha = \beta = 1$ and also closest of the three options to the MLE estimate.