

Problem Set 11 Solutions

1. Check book solutions.
2. (a) To find the MAP estimate, we need to find the value x that maximizes the conditional density $f_{X|Y}(x | y)$ by taking its derivative and setting it to 0.

$$\begin{aligned} f_{X|Y}(x | y) &= \frac{p_{Y|X}(y | x) \cdot f_X(x)}{p_Y(y)} \\ &= \frac{e^{-x} x^y}{y!} \cdot \mu e^{-\mu x} \cdot \frac{1}{p_Y(y)} \\ &= \frac{\mu}{y! p_Y(y)} \cdot e^{-(\mu+1)x} x^y \end{aligned}$$

$$\begin{aligned} \frac{d}{dx} f_{X|Y}(x | y) &= \frac{d}{dx} \left(\frac{\mu}{y! p_Y(y)} \cdot e^{-(\mu+1)x} x^y \right) \\ &= \frac{\mu}{y! p_Y(y)} x^{y-1} e^{-(\mu+1)x} (y - x(\mu + 1)) \end{aligned}$$

Since the only factor that depends on x which can take on the value 0 is $(y - x(\mu + 1))$, the maximum is achieved at

$$\hat{x}_{\text{MAP}}(y) = \frac{y}{1 + \mu}$$

It is easy to check that this value is indeed maximum (the first derivative changes from positive to negative at this value).

- (b) i. To show the given identity, we need to use Bayes' rule. We first compute the denominator, $p_Y(y)$

$$\begin{aligned} p_Y(y) &= \int_0^\infty \frac{e^{-x} x^y}{y!} \mu e^{-\mu x} dx \\ &= \frac{\mu}{y! (1 + \mu)^{y+1}} \int_0^\infty (1 + \mu)^{y+1} x^y e^{-(1+\mu)x} dx \\ &= \frac{\mu}{(1 + \mu)^{y+1}} \end{aligned}$$

Then, we can substitute into the equation we had derived in part (a)

$$\begin{aligned} f_{X|Y}(x | y) &= \frac{\mu}{y! p_Y(y)} x^y e^{-(\mu+1)x} \\ &= \frac{\mu (1 + \mu)^{y+1}}{y! \mu} x^y e^{-(\mu+1)x} \\ &= \frac{(1 + \mu)^{y+1}}{y!} x^y e^{-(\mu+1)x} \end{aligned}$$

Thus, $\lambda = 1 + \mu$.

ii. We first manipulate $xf_{X|Y}(x | y)$:

$$\begin{aligned} xf_{X|Y}(x | y) &= \frac{(1 + \mu)^{y+1}}{y!} x^{y+1} e^{-(\mu+1)x} \\ &= \frac{y + 1}{1 + \mu} \frac{(1 + \mu)^{y+2}}{(y + 1)!} x^{y+1} e^{-(\mu+1)x} \\ &= \frac{y + 1}{1 + \mu} f_{X|Y}(x | y + 1) \end{aligned}$$

Now we can find the conditional expectation estimator:

$$\begin{aligned} \hat{x}_{\text{CE}}(y) &= \mathbf{E}[X|Y = y] = \int_0^\infty xf_{X|Y}(x | y) dx \\ &= \int_0^\infty \frac{y + 1}{1 + \mu} f_{X|Y}(x | y + 1) dx = \frac{y + 1}{1 + \mu} \end{aligned}$$

(c) The conditional expectation estimator is always higher than the MAP estimator by $\frac{1}{1+\mu}$.

3. (a) The likelihood function is

$$\prod_{i=1}^k P_{T_i}(T_i = t_i | Q = q) = q^k (1 - q)^{\sum_{i=1}^k t_i - k}.$$

To maximize the above probability we set its derivative with respect to q to zero

$$kq^{k-1}(1 - q)^{\sum_{i=1}^k t_i - k} - \left(\sum_{i=1}^k t_i - k\right)q^k(1 - q)^{\sum_{i=1}^k t_i - k - 1} = 0,$$

or equivalently

$$k(1 - q) - \left(\sum_{i=1}^k t_i - k\right)q = 0,$$

which yields $\hat{Q}_k = \frac{k}{\sum_{i=1}^k t_i}$. This is not different from the MAP estimate found before. Since the MAP estimate is calculated using a uniform prior, the likelihood function is a ‘scaled’ version of posterior probability and they can be maximized at the same value of q .

(b) Since $\frac{1}{\hat{Q}_k} = \frac{\sum_{i=1}^k T_i}{k}$, and that each T_i is independent identically distributed, it follows that $\frac{1}{\hat{Q}_k}$ is actually a sample mean estimator. The weak law of large numbers says that, when the number of samples increases to infinity, the sample mean estimator converges to the actual mean, which is $\frac{1}{q^*}$ in this case. So we can write the limit of probability as

$$\lim_{k \rightarrow \infty} \mathbf{P} \left(\left| \frac{1}{\hat{Q}_k} - \frac{1}{q^*} \right| > \epsilon \right) = \lim_{k \rightarrow \infty} \mathbf{P} \left(\left| \frac{\sum_{i=1}^k T_i}{k} - \mathbf{E}[T_1] \right| > \epsilon \right) = 0.$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
 Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
 (Fall 2010)

(c) Chebyshev inequality states that

$$\mathbf{P} \left(\left| \frac{\sum_{i=1}^k T_i}{k} - \mathbf{E}[T_1] \right| \geq \epsilon \right) \leq \frac{\text{var}(T_1)}{k\epsilon^2}.$$

So we have

$$\begin{aligned} \mathbf{P} \left(\left| \frac{1}{\widehat{Q}_k} - \frac{1}{q^*} \right| \leq 0.1 \right) &= \mathbf{P} \left(\left| \frac{\sum_{i=1}^k T_i}{k} - \frac{1}{q^*} \right| \leq 0.1 \right) \\ &= 1 - \mathbf{P} \left(\left| \frac{\sum_{i=1}^k T_i}{k} - \mathbf{E}[T_1] \right| \geq 0.1 \right) \geq 1 - \frac{\text{var}(T_1)}{k * 0.1^2} \end{aligned}$$

To ensure the above probability to be greater than 0.95, we need that

$$1 - \frac{\text{var}(T_1)}{k * 0.1^2} = 1 - \frac{1 - q}{k * 0.1^2} \geq 0.95,$$

or

$$k \geq 2000\text{var}(T_1) = 2000 \frac{1 - q}{q^2}$$

The number of observations k needed depends on the variance of T_1 . For q close to 1, the variance is close to 0, and the required number of observations is very small (close to 0). For $q = 1/2$, the variance is maximum ($\text{var}(T_1) = 2$), and we require $k = 4000$. Thus, to guarantee the required accuracy and confidence for all q , we need that,

$$k \geq 4000.$$

4. (a) Normalization of the distribution requires:

$$1 = \sum_{k=0}^{\infty} p_K(k; \theta) = \sum_{k=0}^{\infty} \frac{e^{-\frac{k}{\theta}}}{Z(\theta)} = \frac{1}{Z(\theta)} \sum_{k=0}^{\infty} e^{-\frac{k}{\theta}} = \frac{1}{Z(\theta) \cdot (1 - e^{-\frac{1}{\theta}})},$$

so $Z(\theta) = \frac{1}{1 - e^{-\frac{1}{\theta}}}$.

(b) Rewriting $p_K(k; \theta)$ as:

$$p_K(k; \theta) = \left(e^{-\frac{1}{\theta}} \right)^k \left(1 - e^{-\frac{1}{\theta}} \right), \quad k = 0, 1, \dots$$

the probability distribution for the photon number is a geometric probability distribution with probability of success $p = 1 - e^{-\frac{1}{\theta}}$, and it is shifted with 1 to the left since it starts with $k = 0$. Therefore the photon number expectation value is

$$\mu_K = \frac{1}{p} - 1 = \frac{1}{1 - e^{-\frac{1}{\theta}}} - 1 = \frac{1}{e^{\frac{1}{\theta}} - 1}$$

and its variance is

$$\sigma_K^2 = \frac{1 - p}{p^2} = \frac{e^{-\frac{1}{\theta}}}{(1 - e^{-\frac{1}{\theta}})^2} = \mu_K^2 + \mu_K.$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
 Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
 (Fall 2010)

(c) The joint probability distribution for the k_i is

$$p_K(k_1, \dots, k_n; \theta) = \frac{1}{Z(\theta)^n} \prod_{i=1}^n e^{-k_i/\theta} = \frac{1}{Z(\theta)^n} e^{-\frac{1}{\theta} \sum_{i=1}^n k_i}.$$

The log likelihood is $-n \cdot \log Z(\theta) - 1/\theta \sum_{i=1}^n k_i$.

We find the maxima of the log likelihood by setting the derivative with respect to the parameter θ to zero:

$$\frac{d}{d\theta} \log p_K(k_1, \dots, k_n; \theta) = -n \cdot \frac{e^{-\frac{1}{\theta}}}{\theta^2(1 - e^{-\frac{1}{\theta}})} + \frac{1}{\theta^2} \sum_{i=1}^n k_i = 0$$

or

$$\frac{1}{e^{\frac{1}{\theta}} - 1} = \frac{1}{n} \sum_{i=1}^n k_i = s_n.$$

For a hot body, $\theta \gg 1$ and $\frac{1}{e^{\frac{1}{\theta}} - 1} \approx \theta$, we obtain

$$\theta \approx \frac{1}{n} \sum_{i=1}^n k_i = s_n.$$

Thus the maximum likelihood estimator $\hat{\Theta}_n$ for the temperature is given in this limit by the sample mean of the photon number

$$\hat{\Theta}_n = \frac{1}{n} \sum_{i=1}^n K_i.$$

(d) According to the central limit theorem, the sample mean approaches for large n a Gaussian distribution with standard deviation our root mean square error

$$\sigma_{\hat{\Theta}_n} = \frac{\sigma_K}{\sqrt{n}}.$$

To allow only for 1% relative root mean square error in the temperature, we need $\frac{\sigma_K}{\sqrt{n}} < 0.01\mu_K$. With $\sigma_K^2 = \mu_K^2 + \mu_K$ it follows that

$$\sqrt{n} > \frac{\sigma_K}{0.01\mu_K} = 100 \frac{\sqrt{\mu_K^2 + \mu_K}}{\mu_K} = 100 \sqrt{1 + \frac{1}{\mu_K}}.$$

In general, for large temperatures, i.e. large mean photon numbers $\mu_K \gg 1$, we need about 10,000 samples.

(e) The 95% confidence interval for the temperature estimate for the situation in part (d), i.e.

$$\sigma_{\hat{\Theta}_n} = \frac{\sigma_K}{\sqrt{n}} = 0.01\mu_K,$$

is

$$[\hat{K} - 1.96\sigma_{\hat{K}}, \hat{K} + 1.96\sigma_{\hat{K}}] = [\hat{K} - 0.0196\mu_K, \hat{K} + 0.0196\mu_K].$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Fall 2010)

5. (a) The sample mean estimator $\hat{\Theta}_n = \frac{W_1 + \dots + W_n}{n}$ in this case is

$$\hat{\Theta}_{1000} = \frac{2340}{1000} = 2.34.$$

From the standard normal table, we have $\Phi(1.96) = 0.975$, so we obtain

$$\mathbf{P} \left(\frac{|\hat{\Theta}_{1000} - \mu|}{\sqrt{\text{var}(W_i)/1000}} \leq 1.96 \right) \approx 0.95.$$

Because the variance is less than 4, we have

$$\mathbf{P} \left(\hat{\Theta}_{1000} - \mu \leq 1.96 \sqrt{\text{var}(W_i)/1000} \right) \leq \mathbf{P} \left(\hat{\Theta}_{1000} - \mu \leq 1.96 \sqrt{4/1000} \right),$$

and letting the right-hand side of the above equation ≈ 0.95 gives a 95% confidence, i.e.,

$$\left[\hat{\Theta}_{1000} - 1.96 \sqrt{4/1000}, \hat{\Theta}_{1000} + 1.96 \sqrt{4/1000} \right] = \left[\hat{\Theta}_{1000} - 0.124, \hat{\Theta}_{1000} + 0.124 \right] = [2.216, 2.464]$$

- (b) The likelihood function is

$$f_W(w; \theta) = \prod_{i=1}^n f_{W_i}(w_i; \theta) = \prod_{i=1}^n \theta e^{-\theta w_i},$$

And the log-likelihood function is

$$\log f_W(w; \theta) = n \log \theta - \theta \sum_{i=1}^n w_i,$$

The derivative with respect to θ is $\frac{n}{\theta} - \sum_{i=1}^n w_i$, and by setting it to zero, we see that the maximum of $\log f_W(w; \theta)$ over $\theta \geq 0$ is attained at $\hat{\theta}_n = \frac{n}{\sum_{i=1}^n w_i}$. The resulting estimator is

$$\hat{\Theta}_n^{mle} = \frac{n}{\sum_{i=1}^n W_i}.$$

In this case,

$$\hat{\Theta}_n^{mle} = \frac{1000}{2340} = 0.4274.$$

6. (a) Using the regression formulas of Section 9.2, we have

$$\hat{\theta}_1 = \frac{\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^5 (x_i - \bar{x})^2}, \quad \hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{x},$$

where

$$\bar{x} = \frac{1}{5} \sum_{i=1}^5 x_i = 4.94, \quad \bar{y} = \frac{1}{5} \sum_{i=1}^5 y_i = 134.38.$$

The resulting ML estimates are

$$\hat{\theta}_1 = 40.53, \quad \hat{\theta}_0 = -65.86.$$

(b) Using the same procedure as in part (a), we obtain

$$\hat{\theta}_1 = \frac{\sum_{i=1}^5 (x_i^2 - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^5 (x_i^2 - \bar{x})^2}, \quad \hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{x},$$

where

$$\bar{x} = \frac{1}{5} \sum_{i=1}^5 x_i^2 = 33.60, \quad \bar{y} = \frac{1}{5} \sum_{i=1}^5 y_i = 134.38.$$

which for the given data yields

$$\hat{\theta}_1 = 4.09, \quad \hat{\theta}_0 = -3.07.$$

Figure 1 shows the data points (x_i, y_i) , $i = 1, \dots, 5$, the estimated linear model

$$y = 40.53x - 65.86,$$

and the estimated quadratic model

$$y = 4.09x^2 - 3.07.$$

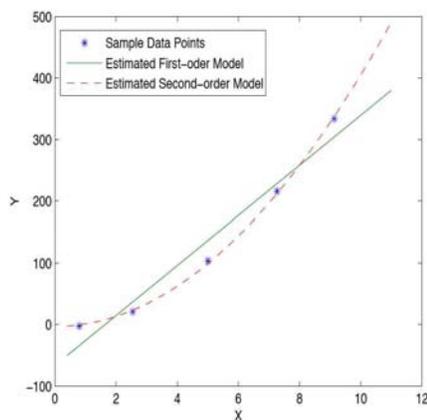


Figure 1: Regression Plot

MIT OpenCourseWare
<http://ocw.mit.edu>

6.041 / 6.431 Probabilistic Systems Analysis and Applied Probability
Fall 2010

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.